Latin American Data in Science, v.2, n.1
datainscience.com.br
ISSN: 2763-9290

Database of Active Pharmaceutical Ingredients (APIs) present in the Brazilian Pharmacopeia.

Recebido: 11/11/2021 | Aceito: 07/03/2022 | Publicado: 31/03/2022
https://doi.org/10.53805/lads.v2i1.35

Caio Felipe de A. R. Cheohen[*1,2]; Bruce V. Andriolo[2], Manuela L. da Silva [1,2]

ABSTRACT

Active Pharmaceutical Ingredients (APIs) are molecules that can be used for the manufacturing of Drugs. The Brazilian Health Regulatory Agency (Agência Nacional de Vigilância Sanitária) is responsible for the oversight and approval of production and distribution of all drugs in Brazilian territory. Four different databases, including 336 out of the 346 total APIs, were created from the APIs text information available in the 6th edition of the Brazilian pharmacopeia, volume 2, "Active Pharmaceutical Ingredients" monography. Each database is composed by a different extension file, they are: two-dimensional *sdf*, three-dimensional *sdf*, *mol2* and *PDBQT* file extensions, converted by using OpenBabel and Marvin Sketch software.

Keywords: Active Pharmaceutical Ingredients (APIs); Anvisa; Brazilian pharmacopeia; In silico; Virtual screening.

DATA IMPORTANCE

- Supplies four databases based on APIs present in the Brazilian pharmacopoeia and approved by Brazilian Health Regulatory Agency (Agência Nacional de Vigilância Sanitária - ANVISA);
- Decreases the time taken to construct the three-dimensional structure of the APIs and correction of protonation states for physiological pH;
- Allows different *in silico* analysis to be performed using the given databases, like pharmacokinetic and toxicity predictions, molecular docking, virtual screening, molecular dynamics, and others.

[1] Universidade Federal do Rio de Janeiro, Macaé, Brasil. caiocheohen@gmail.com.
[2] Instituto Nacional de Metrologia, Qualidade e Tecnologia – Inmetro, Duque de Caxias, Brasil.

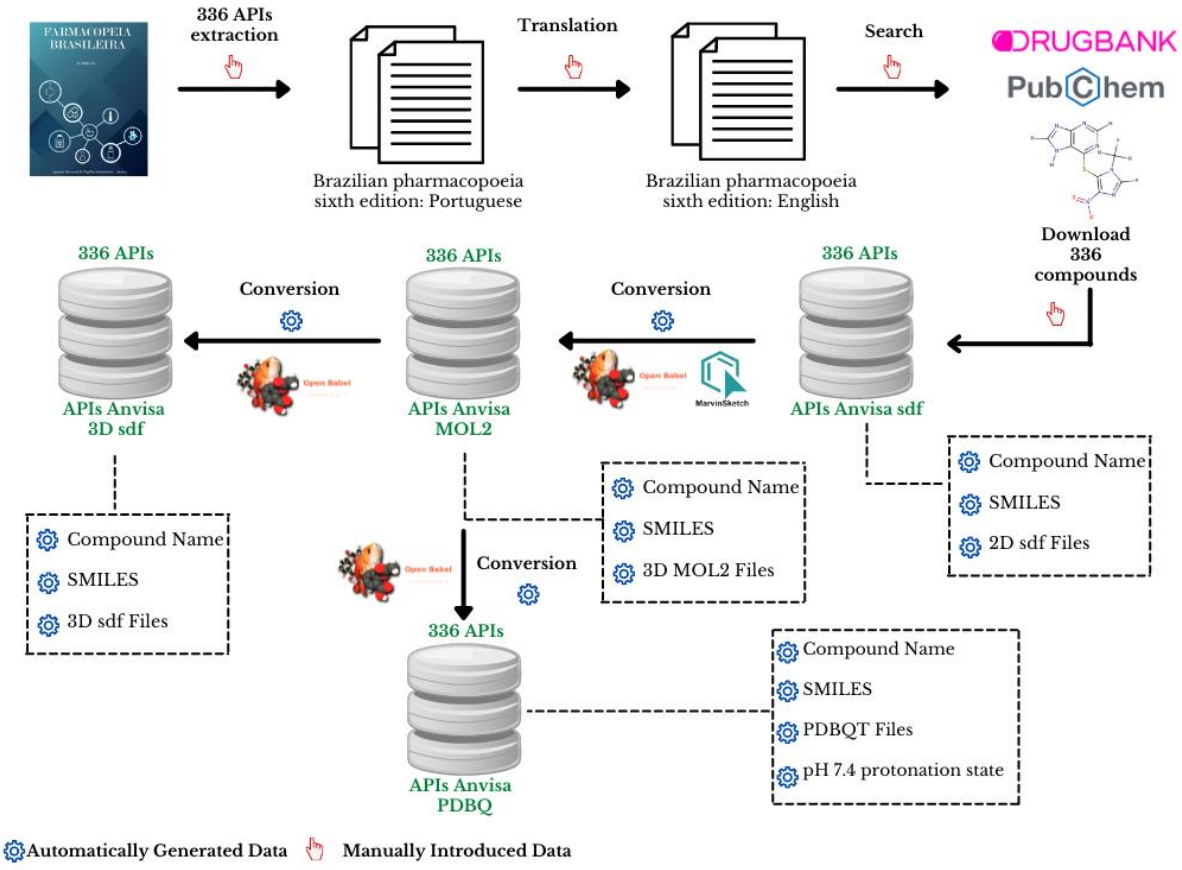
MATERIALS AND METHODS

Data extraction

Four different databases were created from the APIs text information available in the 6th edition of the Brazilian pharmacopeia, volume 2, "Active Pharmaceutical Ingredients" monography. The written document was originally published on August 12, 2019. The pharmacopeia owns a total of 346 APIs, including different formulation types, such as creams, tablets, injection solutions, oral suspensions, powder for solutions, powder for suspensions, powder for oral suspensions, capsules, suppositories, pills, shampoos, ophthalmic solutions, gels, ointments, elixirs, chewable tablets and diluted solutions.

The names of the APIs were extracted from the original file and sent to an editable text file where they were later translated from Portuguese into English. At this stage, the APIs that were repeated, or the same API, however, in different formulations, were excluded. After this exclusion, a total of 336 different molecules remained, which were entered into the databases (Fig. 1).

Figure 1. Schematic representation of the hierarchical process of data entry for all databases. This workflow is divided in two steps. First, the manually introduced data part, consisting in the download of the 6th edition of the Brazilian pharmacopeia, the extraction of the APIs compound names, translation from Portuguese to English, and the search for compounds and download the three-dimensional structures in *sdf* format at Drugbank or PubChem. Second, the automatically generated data part, including, molecules conversion, from 2D *sdf* to *mol2* format, using OpenBabel and Marvin sketch software, *mol2* to *3Dsdf format, mol2* to *PDBQT* format and correction of the protonation state to pH 7.4 by using the OpenBabel software.



Data conversion

The structures names were searched on the Drugbank (https://go.drugbank.com/) website, where the files having the two-dimensional coordinates in *structure-data file (sdf)* format were downloaded. From the 346 molecules searched in the Drugbank and PubChem databases, only 336 had a two-dimensional structure in the *sdf* format, this value being the total contained in all the created databases (Supplementary Table 1). A total of 10 structures were left out of the databases: Sterile water for irrigation; water for injectables; purified water; starch; aluminum chlorhydroxide, being aluminum and chlorine atoms; paraffins white and liquid petrolatum; dyes, red amaranth aluminum lacquer and ponceau red aluminum lacquer.

The first database, named "APIs Anvisa sdf" is composed of 2D *sdf* format files, obtained from PubChem or DrugBank databases, where the download of the molecules was made manually and individually.

The second database, named "APIs Anvisa MOL2" is composed of *mol2* format files. The data conversion was made by using the OpenBabel software (O'BOYLE et al., 2011) via linux terminal comand "obabel -isdf *.sdf -omol2 -O *.mol2 –gen3D", and Marvin sketch software, version number 21.10, was used for drawing, displaying and characterizing chemical structures (http://www.chemaxon.com/). The molecules were converted to *mol2* format, creating a three-dimensional structure file.

Third database "APIs Anvisa 3Dsdf" having three-dimensional structures in *sdf* format was created from "APIs Anvisa MOL2" by using OpenBabel via linux terminal comand "obabel -imol2 *.mol2 -osdf -O *.sdf".

Fourth, "APIs Anvisa PDBQT" is composed of PDBQT format files by using OpenBabel via linux terminal comand "obabel -imol2 *.mol2 -opdbqt -O *.pdbqt -r -p 7.4".

Each database makes available to users, in addition to the 2D and 3D structures of the APIs, a table containing all the information of the structures contained in the databases, namely, the identification (ID) of the API in PubChem or Drugbank; the compound name, which is previously translated from the original in Portuguese to English; the SMILES (Simplified molecular-input line-entry system) notation; and the illustrative figures of the two-dimensional and three-dimensional structures of the APIs, as shown in the example in Figure 2.

The *structure-data file* (*sdf*), the *SYBYL MOL2 format* (*mol2*) and the *PDBQT* (Protein Data Bank, Partial Charge (Q), & Atom Type (T)) are capable of recording the atoms, bonds, vectors, 3D coordinates, charges etc. However, the *sdf* format own the capability of storing multiple molecules at once that can be extracted by the program reding it. The *mol2* and *PDBQT* format can store information about flexibility (DALBY et al., 1992), which is important for *in silico* simulations, depending on the programs used, and on the situation tested.

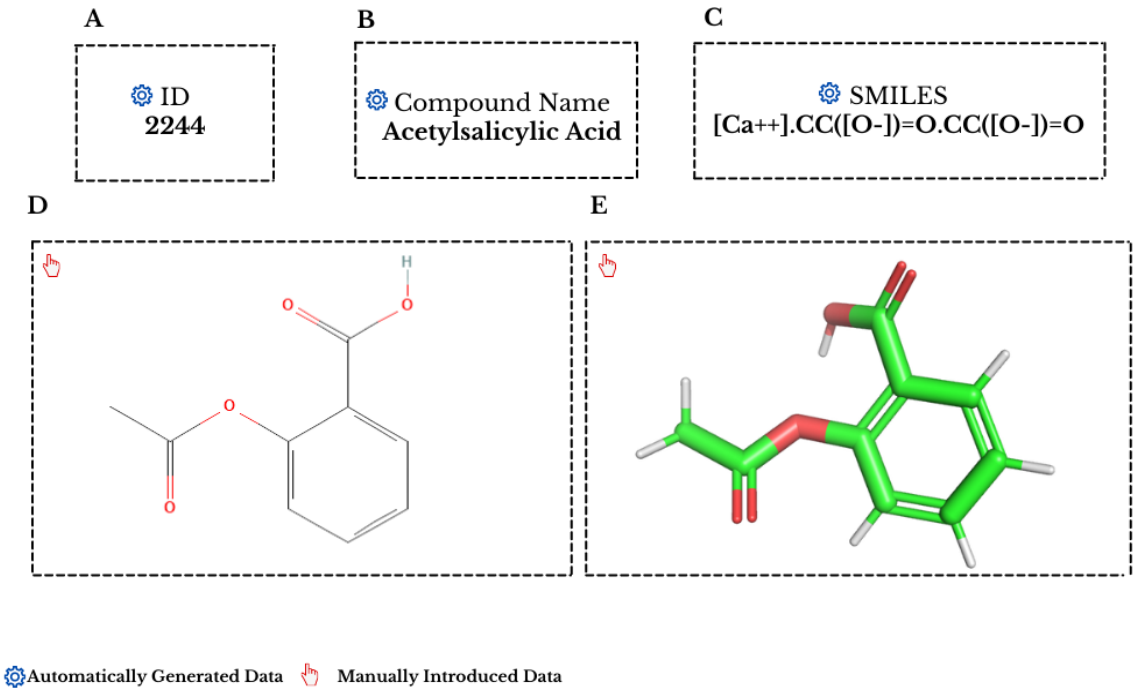
DATA DESCRIPTION

Active Pharmaceutical Ingredients (APIs) are the active ingredients of the medicine, and they are also the first representatives when it comes to drug production in the pharmaceutical industry. Being chemical substances, drugs or raw materials that have pharmacological properties with a medicinal purpose. APIs can be applied on diagnosis, relief or treatment, used to change or explore physiological systems or pathological states (ANVISA, 2018).

The APIs are molecules that can be used for the manufacturing of Drugs, but also vaccines, creams, pills, gels, among others, depending on the use, and to guarantee their safety use by the Brazilian population, they are thoroughly regulated, with strict rules to prove the identity of the API, as their purity, and characteristics. But they are not medicinal plants, they have their own

division. The API can be synthetic or have a plant origin (ANVISA, 2019).

Figure 2. Representation of the information presented into the supplementary table I. **A:** Reference database entry identification code; **B:** English compound name as described and manually translated from the 6th edition of the Brazilian pharmacopeia, volume 2, "Active Pharmaceutical Ingredients" monography. **C:** SMILES notation. **D:** two-dimensional compound image. **E:** Three-dimensional compound image.



The Brazilian Health Regulatory Agency (Agência Nacional de Vigilância Sanitária, Anvisa for short) is responsible for the oversight and approval of production and distribution of all drugs in Brazilian territory (ANVISA, 2019). Therefore, the official document, the Brazilian pharmacopeia, describes these drugs, called Active Pharmaceutical Ingredients (APIs), and how to correctly identify them and how to perform quality control tests (ANVISA, 2019). Unfortunately, this information is only available in Portuguese and in written format, causing a lack of scientific papers based on this document. Willing to use the Brazilian pharmacopeia, it would require the manual extraction of labels information, translation of APIs names from Portuguese to other languages (I.e., English), the search of the tridimensional structures in different databases (i.e., DrugBank and PubChem), download and preparation of the data, increasing the workload considerably, making the work impractical several times.

Anvisa is not only responsible for the regulation of all APIs that can be produced and distributed in Brazil. It is also responsible for the regulation of restaurants (to confirm that they are clean and won't spread diseases), medical devices, cosmetics, pesticides, drugs, medicinal plants, sanitizing products, tobacco products, Import and export, ports and airports sanitary conditions and control measures in case of outbreaks. This way, it can protect the Brazilian citizens from harm, like diseases and faulty products. Anvisa is independent and has influence throughout Brazil, sometimes working directly with the Ministry of Health, the Ministry of Planning, Development and Management and the Ministry of Agriculture, Livestock and Supply (ANVISA, 2019).

Dataset

The first database named as *'APIs Anvisa sdf'* has the two-dimensional structures in *sdf* format, and does not have hydrogen atoms directly related to the pH whose simulation wanted by the user can be performed (SACHDEVA et al., 2020).

The second database named as *'APIs Anvisa mol2',* was created from the first database. All hydrogens were added to the 3D structures, as a pattern for the *mol2* file type.

The third database, '*APIs Anvisa 3Dsdf',* derives from the second database and holds de 3D *sdf* format file structures. Differently when compared to the earlier database, this database does not have all the hydrogens into the 3D files.

The fourth database, "APIs Anvisa PDBQT", also derives from '*APIs Anvisa mol2',* and the protonation states for physiological pH (7.4) were defined as standard pattern. The molecules on PDBQT format are ready for molecular docking simulations, once that they contain only the largest compound found in each archive file. The need to define the correct protonation states come from the fact that one wants to simulate the physiological environment (GUEDES; MAGALHÃES; DARDENNE, 2014; MORRIS; LIM-WILBY, 2006), so APIs must be in a state that we would find inside cells, where we expect them to act. With the defined protonation state, the presence or absence of hydrogens will change the way these molecules interact chemically with the target protein, aiming to generate inhibition.

SUPPLEMENTARY MATERIALS

Repository name: Mendeley Data
DOI of the dataset (when available):

DOI:10.17632/snzrkfrg7x.3
DOI:10.17632/dp3f425kkf.2

Link to access the data:
https://data.mendeley.com/datasets/snzrkfrg7x/3
https://data.mendeley.com/datasets/dp3f425kkf/2

ACKNOWLEDGEMENTS

The authors thanks CNPq, CAPES and FAPERJ for financial support.

REFERENCES

ANVISA - AGÊNCIA NACIONAL DE VIGILÂNCIA SANITÁRIA. FARMACOPEIA BRASILEIRA. 6a ed. Brasília - DF: Agência Nacional de Vigilância Sanitária, 2019.

ANVISA - AGÊNCIA NACIONAL DE VIGILÂNCIA SANITÁRIA. Institutional. Disponível em: <https://www.gov.br/anvisa/pt-br/english/introduction>. Accessed October 5th, 2021.

ANVISA - AGÊNCIA NACIONAL DE VIGILÂNCIA SANITÁRIA. PERGUNTAS E RESPOSTAS, Assunto: Insumos Farmacêuticos Ativos. p. 21, 2018.

DALBY, A. et al. Description of Several Chemical Structure File Formats Used by Computer Programs Developed at Molecular Design Limited. Journal of Chemical Information and Computer Sciences, v. 32, n. 3, p. 244–255, 1992.

GUEDES, I. A.; MAGALHÃES, C. S. DE; DARDENNE, L. E. Atracamento Molecular. In: VERLI, H. (Ed.). Bioinformática: da Biologia à Flexibilidade Molecular. 1a Edição ed. São Paulo: SBBq, 2014. p. 188–208.

MORRIS, G. M.; LIM-WILBY, M. Molecular Docking. In: KUKOL, A. (Ed.). Methods in Molecular Biology. Totowa: [s.n.]. v. 443p. 365–382.

O'BOYLE, N. M. et al. Open Babel: An open chemical toolbox. Journal of Cheminformatics, v. 3, n. 33, p. 1–14, 2011.

SACHDEVA, C. et al. In silico Potential of Approved Antimalarial Drugs for Repurposing against COVID-19. OMICS A Journal of Integrative Biology, v. 24, n. 10, p. 568–580, 2020.