Latin American Data in Science, v.3, n.1
datainscience.com.br
ISSN: 2763-9290

Development of a Database of Peptides with Potential for Pharmacological Intervention in Human Pathogen Molecular Targets.

Recebido: 06/06/2023 I Aceito: 09/08/2023 I Publicado: 02/10/2023
https://doi.org/10.53805/lads.v3i1.63

Micael da S. P. Gonzalez[1,2], Caio F. de A. R. Cheohen[3], Bruce V. Andriolo[1], Manuela L. da Silva[1,3*]

ABSTRACT

Peptides are polymeric chains used as research objects in the search for new drugs with greater efficacy and fewer side effects. Therefore, we created three databases of antimicrobial peptides (GONZALEZ et al., 2023) using PubChem and ChEMBL. First we acquired the Simplified Molecular-Input Line-Entry System (SMILES) of several peptides belonging to different types of pathogens, namely bacteria, viruses, parasites, and fungi. Using the OpenBabel software, these SMILES had their file formats and structures converted to create: one database in one dimension SMI format, and two with three-dimensional MOL2 and PDB file formats. In total the three databases consists of 718 peptides that have been shown to possess inhibtory activity on molecular targets of clinically important pathogens.

Keywords: Database; Peptides; Pathogens; In silico; Virtual Screening.

DATA IMPORTANCE

- Streamline the scientific community's access to qualitative and reliable information about peptides with antipathogenic potential;
- Allow analysis of biochemical interactions between possible drugs and receptors;
- Reduction of time spent in traditional drug discovery through in silico methods.

[1] Instituto Nacional de Metrologia, Qualidade e Tecnologia - Duque de Caxias, RJ, Brazil. manuleal2@gmail.com
[2] Universidade Católica de Petrópolis - Petrópolis, RJ, Brazil.
[3] Instituto de Biodiversidade e Sustentabilidade NUPEM, Universidade Federal do Rio de Janeiro - Macaé, RJ, Brazil.

MATERIALS AND METHODS

Literature search

The first step involved conducting a search in the literature on PubMed for information on biologically active peptides selecting molecular targets in human pathogens and their mechanisms of biological action, both in silico and in vitro assays, to validate the compounds to be included in the database

With regards to keywords, the same search strategy was applied for all antimicrobial peptides. The filter selected was to search for free full-text articles. The terms used in the advanced search were as follows:

- "(virus OR viral) AND (peptide OR peptides) AND (inhibit OR block)";
- "(bacteria OR bacterial) AND (peptide OR peptides) AND (inhibit OR block)";
- "(fungus OR fungic) AND (peptide OR peptides) AND (inhibit OR block)";
- "(parasitic OR parasitical) AND (peptide OR peptides) AND (inhibit OR block)".

After retrieving the results of the literature search, the next step was to search if these peptides would be available in Pubchem. The search for these peptides were conducted in two ways: by searching for the common names of the peptides in Pubchem and using any direct link made available in the papers. Articles that lacked a relationship with Pubchem, had no peptide sequence, or lacked experimental data were excluded from further consideration.

Collection of compounds using databases

Once the sources referring to the target compounds were gathered, started to search for each of the peptides found in the databases linked to the servers of the National Institutes of Health (NIH) or in ChEMBL, so that it could have access to the desired data.

The collected items were grouped in a Microsoft Office Excel (2019) spreadsheet with the names of the peptides, their SMILES notation and their International Union of Pure and Applied Chemistry (IUPAC) nomenclatures allowing for different ways for identification. Additional information such as organisms which the peptides demonstrated activity and the molecular targets were also included to help future users of the worksheet. Finally, one cannot fail to mention the insertion of all the references where it is possible to find the reports of the use of the peptide. The research has a total of found 718 peptides.

Data Transformation

The first database, called "antivirals.smi", is made up of files in SMI format, generated through manual collection in the PubChem and ChEMBL databases.

The second database "antivirals.mol2" is composed of files in MOL2 format. Data conversion was performed by accessing the OpenBabel software version 2.3.1 (O'BOYLE et al., 2011) via a Linux terminal using the command "obabel *. smi -O *.mol2 –gen3d". This step is responsible for originating three-dimensional structure files.

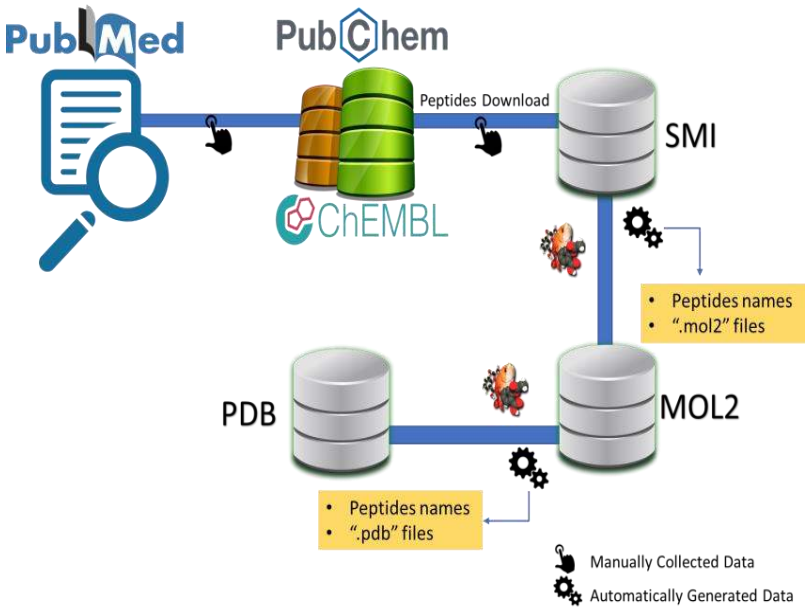
The last conversion step consisted of creating a database called "antivirals.pdb". Still in three-dimensional form, it was transformed directly from MOL2 format to PDB format using the command "obabel *.mol 2 -O *.pdb" via Linux terminal.

In the same way, following the steps (fig. 1), all four databases were produced: antiviral, antibacterial, antifungal and antiparasitic. In case it is desirable to conduct the steps we did, all the structures were available in formats with X coordinates (1D format), and XYZ coordinates (3D formats), together with an Excel spreadsheet containing information about the peptides.

Both SYBYL MOL2 and PDB formats are capable of recording atoms, bonds, vectors, 3D

coordinates etc., as well as storing information about flexibility, which is important for in silico simulations, depending on the programs used and the context in which they are tested (CHEOHEN; ANDRIOLO; SILVA, 2022; DALBY et al., 1992).

Figure 1.- Schematic representation of the steps in the creation of the databases. The process starts with manual extraction of data from literature obtained through a comprehensive search on PubMed and retrieval of peptides from the PubChem and ChEMBL platforms. The data is then downloaded in the form of SMILES notation (smi) and converted into MOL2 and PDB formats using OpenBabel software.



DATA DESCRIPTION

Peptides are polymers composed of peptide bonds of amino acids through a dehydration reaction. Such monomers now called, amino acid residues, have many distinct structures, twenty of which are essential for life. The amino acids are composed of an mino group, a carboxylic group and a side chain R, the latter differs from the basic structure and confer characteristic chemical properties to that molecule. The combination of these structures with each other will determine the function of the product generated, existing countless possibilities of combinations (NELSON; COX, 2014; ROSE, 2019; LOPEZ; MOHIUDDIN, 2022).

Peptides have been proposed as a solution for controlling infections caused by pathogens, as they have potential for clinical application. However, the limited availability and excessive cost of these compounds pose a challenge for their widespread use. Peptide-based drugs have several advantages over conventional drugs, including greater efficiency, lower toxicity (or absence thereof), and fewer side effects (or absence thereof) (CASTEL et al.,2011).

The Brazilian Health Regulatory Agency (ANVISA) is responsible for inspecting and approving the production and distribution of all medicines in Brazil and classifies peptides as biochemical products that can be used for the manufacture of vaccines and medicines in solid and liquid forms, semi-solid and gaseous.

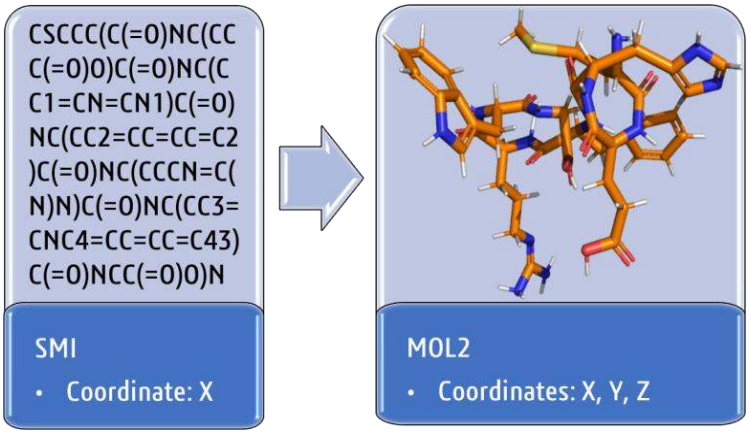
Peptides-based drugs also called biological drugs and determined by ANVISA, in its Collegiate Board Resolution (RDC 55/2010), as complex structures of high molecular weight obtained from biological fluids, tissues of animal origin or biotechnological procedures through manipulation, insertion of other genetic material (recombinant DNA technology) or alteration of

genes that occurs due to irradiation, chemical products or forced selection (ANVISA, 2010).

Dataset

The first database, called “antivirals.smi” has one-dimensional structures in smi format (SMILES) and differs from other extensions by providing a simple set of representations that are suitable as labels for chemical data and as a compact memory identifier for exchanging data among researchers (WEININGER, 1988). In addition, SMILES and its extensions serve as descriptive codes that allow the rapid generation of graphical objects that can be searched for chemical structures with tools such as Open Babel (O'BOYLE, 2011).

Figure 2.- Differentiation between the extensions present in the database. On the left, the one-dimensional format file; on the right, the file containing the three-dimensional peptide structure.



The second database, called “antivirals.mol2”, originates from the first database (fig. 2). MOL2 file format is capable of recording information regarding protonation state, force fields and multiple structures (CLARK et al., 1989). It is used as input structure by a wide range of prediction and simulation tools, such as GOLD, CGenFF and Discovery Studio to name a few, for example.

The third database, named “antivirals.pdb”, is derived from the second database and contains files that include a large "header" section of text that summarizes the protein, citation information, and solution details of the structure, followed by the sequence and a long list of atoms and their coordinates, this format (pdb) is also useful because it is easily read by most bioinformatics softwares (BERMAN; WESTBROOK; FENG, 2000). Such compounds are finalized to be used by molecular docking software for predicting a peptide-target contact, for example (MORRIS; RUTH; LINDSTROM, 2009).

SUPPLEMENTARY MATERIALS

The spreadsheet (GONZALEZ et al., 2023) is a Microsoft Excel file consisting of four separate tabs according to the performance in pathogenic classes, being divided into “Antivirals”, “Antibacterials”, “Antifungicals” and “Antiparasiticals” and seven columns, namely:

- Peptides Name: Peptide nomenclature as reported in PubChem.
- Smiles: SMILES notation for all peptides.
- IUPAC: Condensed IUPAC nomenclature of each peptide.

- Affected Organisms: The range of pathogens that are reported as potential inhibitors for those peptides.
- Targets: Molecular target that is hit by that peptide.
- Useful Reference: Links that redirect to the article directly in PubMed for each reference used in the peptides.
- Origin: The genesis of each peptide, be it natural, synthetic, or semi-synthetic.

ACKNOWLEDGEMENTS

This study is supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico, the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior and the Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro.

REFERENCES

ANVISA — NATIONAL HEALTH SURVEILLANCE AGENCY. BRAZILIAN PHARMACOPEIA. 6th ed. Brasília, DF: National Health Surveillance Agency, 2019.

ANVISA - NATIONAL HEALTH SURVEILLANCE AGENCY. Institutional. Available in: <https://www.gov.br/anvisa/pt-br/english/introduction >.

BRAZIL. Resolution No. 55, of December 16, 2010. Approves the regulation on the registration of new biological products and biological products and other measures. ANVISA - National Health Surveillance Agency. Available at: <http://portal.anvisa.gov.br/legislacao/?inheritRedirect=true#/visualizar/28623>.

BERMAN, H. M., WESTBROOK, J., FENG, Z. et al. The Protein Data Bank, Nucleic Acids Research, Volume 28, Issue 1, 1 January 2000, Pages 235–242. DOI: https://doi.org/10.1093/nar/28.1.235.

CASTEL, G. et al. Phage display of combinatorial peptide libraries: application to antiviral research, Molecules, 2011, vol. 16 (p. 3499-3518).

CLARK, M.; CRAMER III, RICHARD, D.; VAN OPDENBOSCH, N. Validation of the general purpose tripos 5.2 force field. Journal of computational chemistry, v. 10, n. 8, p. 982-1012, 1989.

DALBY, A. et al. Description of Several Chemical Structure File Format Used by Computer Programs Developed at Molecular Design Limited. Journal of Chemical Information and Computer Sciences, v. 32, no. 3, p. 244-255, 1992.

CHEOHEN, C. F. A. R.; ANDRIOLO, B. V.; SILVA, M. L. Database of Active Pharmaceutical Ingredients (APIs) present in the Brazilian Pharmacopeia. Latin American Data in Science, [S. l.], v. 2, no. 1, p. 7–12, 2022. DOI: 10.53805/lads. v2i1.35. Available in: <https://ojs.datainscience.com.br/index.php/lads/article/view/35>.

GUEDES, I. A.; MAGALHÃES, C. S.; DARDENNE, L. E. Three-Dimensional Models. In: VERLI, H. (Ed.). Bioinformatics: from Biology to Molecular Flexibility. 1st Edition ed. São Paulo: SBBq, 2014. p. 148-171.

GONZALEZ, M. S. P.; CHEOHEN, C.; ANDRIOLO, B. V.; DA SILVA, M. L. “Database of Peptides with Potential for Pharmacological Intervention in Human Pathogen Molecular Targets”, Mendeley Data, V1, 2023. DOI: 10.17632/2zhgy9ggdv.1.

LOPEZ M. J.; Mohiuddin, S. S. Biochemistry, Essential Amino Acids. StatPearls, Treasure Island (FL): StatPearls published in January of 2022 -. Available in: <https://www.ncbi.nlm.nih.gov/books/NBK557845>.

MORRIS, G.; RUTH, H.; LINDSTROM. W. Software news and updates AutoDock4 and AutoDockTools4: Automated docking with selective receiver flexibility. Journal of Computational Chemistry, (2009), 30(16).

NELSON, D. L; COX, M. M. Lehninger's Principles of Biochemistry. 6th ed. Porto Alegre: Artmed, 2014. 85-90 p.

O'BOYLE, N. M. et al. Open Babel: An open chemical toolbox. Journal of Cheminformatics, v. 3, no. 33, p. 1-14, 2011.

ROSE, A. J. Amino Acid Nutrition and Metabolism in Health and Disease. Nutrients. 2019;11(11):2623. Published 2019 Nov 1. DOI: 10.3390/nu11112623.

WEININGER, D. SMILES, a chemical language and information system. 1. Introduction and rules of encoding. J.Chem. Inf. Computer science 1988, 28, 31.