

NAME: SHALMON N. ANANDAS

CLASS: M.Sc PART – II

COURSE: BIOINFORMATICS

ACADEMIC YEAR: 2022-2023

ROLL NO: 92

PAPER CODE: GNKPSBI302 (PAPER 2)

**COURSE TITLE: CHEMISINFORMATICS &
DRUG DESIGN**

**GURU NANAK KHALSA COLLEGE
MATUNGA, MUMBAI 400019**

DEPARTMENT OF BIOINFORMATICS

CERTIFICATE

This is to certify that **Mr. Shalmon N Anandas (Roll no. 92)** of M.Sc. Part II Bioinformatics has satisfactorily completed the practical semester III course prescribed by the university of Mumbai during academic year 2022-2023

TEACHER IN CHARGE

HEAD OF DEPARTMENT

INDEX

Sr. No.	WEBLEM NAME	PAGE NO	DATE	Signature
1	Importance of 3D structures and method of generation from 1D and 2D representation for Drug designing and Chemoinformatics studies	1-5	08/10/22	
1a	To retrieve 2D/3D structure for Penicillin (Pubche Id- 22502) using PubChem Database and apply filters as: 1) To refine search using Substructure, Exact and Similar structure approach. 2) To screen structure based on chemical properties approach	6-15	08/10/22	
2	Introduction to Chemical Structure (1D, 2D and 3D) Drawing using BIOVIA DRAW Software and File conversion using Open Babel Tool	16-29	16/09/22	
2a	To perform drawing, editing & manipulation of 2D and 3D structures for query “Penicillin” using BIOVIA DRAW software	30-38	06/09/22	
2b	To retrieve, convert & store structure for “Penicillin” query and study various file formats using Open Babel Tool	39-45	06/09/22	
3	Introduction to conformational studies using Balloon software	46-50	07/10/22	
3a	To generate and analyze various structural conformation for Penicillin (Pubchem ID - 5904) molecule using Balloon software	51-56	07/10/22	
4	Introduction to Molecular Descriptors and PaDELPy software	57-70	09/10/22	
4a	To study molecular descriptors of Penicillin (Pubchem ID – 3672) using PaDELPy software and analyze the results.	71-76	09/10/22	
5	Introduction to Pharmacophore mapping	77-83	08/10/22	
5a	To generate and analyze Pharmacophore map for query Penicillin (Pubchem ID: 3672) using PharmMapper web server	84-89	08/10/22	
5b	To perform virtul screening based on pharmacophore features for Penicillin (Pubchem ID: 3672) using ZINCPharmer web server	90-98	08/10/22	
6	Introduction to SwissADME a free web server tool	99-109	08/10/22	
6a	To evaluate pharmacokinetics and drug likeness using SwissADME server for query penicillin (PubChem ID; 3672)	110-118	08/10/22	
7	Building a molecular database along with special emphasis on retrieval using structure input	119-122	04/10/22	

WEBLEM 1

Importance of 3D structures and method of generation from 1D and 2D representation for Drug designing and Chemoinformatics studies

The intensity and selectivity of therapies' interactions with proteins and other bioactive molecules are significantly influenced by the 3-dimensional (3D) structure of those molecules. Previous research has taken into account the strain introduced and tolerated by protein-induced conformational changes.

Ligand-protein interactions control how well new therapies are recognised by their intended targets. In order to create potent and targeted medications, structure-activity relationships (SAR) are often developed to examine noncovalent interactions, such as hydrogen and halogen bonding, salt bridges, and pi-pi stacking. The shape of inhibitors in the solvated vs. bound states can also influence the energetic favorability of inhibitor binding, in addition to many other protein-ligand interactions that can be adjusted. To produce effective medications, the variety of druggable protein targets needs structural and conformational flexibility in ligands. It has been investigated how much molecular strain and associated energy expenditures medications can withstand when they attach to proteins, and it was discovered that molecules binding to proteins might easily use 5–9 kcal/mol of strain energy. Although significant rearrangements are feasible, more diverse structural shapes have been linked to a spectrum of biological activities and positive therapeutic consequences. Modeling and docking of three-dimensional (3D) structures into protein active sites are now possible thanks to the accessibility and abundance of structures in the Protein Data Bank (PDB). In order to improve lead identification and therapeutic success, libraries with greater topological variety have been prepared for drug-discovery campaigns as a result of these observations and methods.

Method of Generation from 1D and 2D representatinog for drug designing and chemoinformatics studies

Mapping Chemical Graphs into 1D descriptors:

Chemical compound information is significantly reduced when mapping to a 0D representation, and 0D descriptors are typically too rough to encode molecules. As an alternative, 1D descriptors like:

- Systematic nomenclature
- Line Notations

Appeared to be enough for the efficient coding of molecules and for database searching.

The idea of a systematic nomenclature scheme (chemical name) by naming rules is human-oriented, hence it has been given top importance to make chemical names that are simple to pronounce and names that human chemists can easily read. But systematic nomenclature, like the IUPAC system, frequently results in names that are overly complex, reflecting the intricacy of chemical structures. A computer-oriented system does not always necessitate a nomenclature that is user-friendly. A preferred name programme (PNP) is currently being considered to address the difficulty of limiting the name generated for a particular structure to a single value, for instance. In 1986, Beilstein created the first system that permitted the entry of chemical structures in the form of their chemical names.

Another significant method for translating a 2D molecular graph into a linear list of letters and numbers is called a line notation. The first systems for this mapping were created even before computer applications, and their success peaked in the 1960s and 1970s when they gave scientists a quick way to code even big molecules.

Due of the small amount of space needed for data storage in computers at the time, this was particularly appropriate. Among crucial line notations throughout history:

- Wiswesser (WLN)
- ROSDAL
- SYBYL (SLN)
- SMILES

SMILES, the later of the two, obviously rules today. This is brought on by the system's intuitive design and its extremely basic semantic rules. We provide some sample cases in Fig. 3. Every chemical entry on Wikipedia is supported by SMILES codes, and practically every molecular editor can produce SMILES. In the other direction, SMILES can take the place of a molecular input into a number of databases, such as Reaxys and/or molecular editors. If we consider that a SMILES input can currently conduct a Google search without any intermediaries, there is a strong odour of success for this system.

A line notation accomplishes a mathematical reduction of data from a 2D into a 1D representation. The key to this success is dimensional reduction, which keeps 2D information intact. Data manipulation is simplified when the dimension is lower. The fact that this dimensionality reduction is completely reversible is a noteworthy observation. In order to recreate the original molecule, we can both obtain a SMILES for every molecule.

Mapping Chemical Graphs into 2D Descriptors:

For the needs of human scientists, molecular graphs were created to code molecules; yet, they first seemed unsuitable for providing computers with chemical data. The approaches created for mapping molecular graphs into computer-friendly data include matrix notations and linear codes. There are many different matrix systems (2D descriptors) that have been developed, and the numeric representation depends on the coding scheme for atoms and bonds. The following matrices are examples of popular notations:

- Adjacency
- Atom connectivity
- Distance (topological or geometrical)
- Bond
- Bond electrons
- Incidence

Although a matrix system works well, it is not practical for storing huge molecules or enormous datasets since the number of entries required to represent a molecule squares with the growth in the number of atoms in the molecule.

An alternative approach is the connection table (CT), which enables us to display molecular graphs as lists of the atoms and bonds within a molecule that grow linearly as the number of atoms increases. There are numerous CT variants available, and Gasteiger (2003) is recommended for a thorough introduction.

The most common method for depicting molecular structures today is a CT notation. In Fig. 6, we demonstrate a Molfile that codes the three-dimensional structure of salicylic acid while also providing the atoms' three-dimensional coordinates. Standard notations like Molfiles and SDFFiles have been established and are widely used for information communication. In essence, a globally standardised computer-readable form has been created from the global language of chemical graphs. (These representations can also be read by a human because they are coded in ASCII.) Bobach et al. provide an intriguing example of coding vitamin C, a chemical molecule containing all tautomers, via a CT (2012).

The molecular topology is defined by a 2D molecular representation (atomic constitution and connectivity). However, in its purest form, this does not encode stereochemistry, also known as atomic configuration—the spatial arrangement of atoms around an atomic centre. Chemical graphs can be enhanced with solid and dashed

wedged bonds to indicate above-the-plane or below-the-plane pointing in order to encode this. The configuration is a streamlined 3D molecular illustration that shows the stereoisomer's atomic arrangement but not its precise 3D atomic coordinates.

The development of stereochemical codes in chemistry aims to greatly simplify the viewing of 3D molecules in 2D for the human chemist. Information about stereochemistry lies between topology (2D) and topography (3D).

There are really just two representations for the actual 3D structure. The first step is to measure the real property of compounds (substances) using 3D atomic coordinates, such as by crystallography. As a result, databases contain such data (see section "Databases and Database Searching"). Second, there are 3D structural predictions made using molecular modelling. For the second scenario, read the section under "Representation of 3D Structures and Their Mapping into Lower Dimensionality" for an explanation of how this process maps a 2D (CT) or 1D (SMILES) molecular representation into 3D. Technically, there are numerous data formats in which 3D structures can be noted:

- SDfile (Structure Data) *.sdf
- Molfile (Connection Table Format) *.mol
- CIF (Crystallographic Information Fil) *.cif
- PDB (Protiens) *.pdf

These data files can represent both descriptor and property data and can be converted into each other.

PUBCHEM DATABASE:

Introduction and history:

PubChem is a database that the general public can access to learn more about chemical compounds and their biological effects. Since its debut in 2004 as a part of the US National Institutes of Health's (NIH) Molecular Libraries Roadmap Initiatives, PubChem has quickly developed into a crucial source of chemical information that supports scientific communities in a variety of fields, including cheminformatics, chemical biology, medicinal chemistry, and drug discovery.

One of the greatest collections of freely accessible chemical information may be found on PubChem. As of September 2015, it included more than 157 million descriptions of chemical substances contributed by depositors, 60 million descriptions of distinctive chemical structures, and 1 million descriptions of biological assays, encompassing roughly 10,000 different protein target sequences. These enormous amounts of data are organised by PubChem into three interconnected databases called Substance, Compound, and BioAssay. Information that was provided by depositors is kept in the Substance database. The Compound database stores unique chemical structures that are taken from the Substance database. Descriptions of biological tests on chemical compounds are kept in the BioAssay database. SID (SubstanceID), CID (CompoundID), and AID (AssayID), respectively, are the main identifiers for the Substance, Compound, and BioAssay databases.

Working:

Web interfaces for textual search:

The three main PubChem databases as well as other significant NCBI databases, including PubMed, Nucleotide and Protein Sequences, Protein Structures, Genome, Taxonomy, BioSystems, Gene Expression Omnibus (GEO), and many others, use the search and retrieval engine Entrez. By starting a search from the PubChem home page, which also offers launch points to other PubChem services, tools, assistance materials, and more, one can search the PubChem databases through Entrez. The search can also be started from the NCBI main page. Entrez searches all accessible Entrez databases by default if a specific database is not

selected in the search menu, and it reports the number of records in each database that are returned for this "global query."

Multiple records from an Entrez search are shown in a document summary (DocSum) report. There is some data-specific information and a link to the record's summary page for each record in the DocSum page. Controls are available on the DocSum page to alter the display type, sort the outcomes in different ways, or export the page to a file or printer. Additionally, the icons and buttons in the right column of the DocSum page give users the option to download the relevant records, alter or refine their search, get related records from other databases, perform additional analysis on the query result, and more.

Non-textual search using the chemical structure search tool:

Entrez cannot be used for searching that requires data types peculiar to PubChem, including chemical structures, since it is largely a text-based search system. A variety of chemical structure search methods, such as identity search, substructure/superstructure search, molecular formula search, and 2-D and 3-D similarity searches, can be used to query and subset the Compound database using the Chemical Structure Search tool. A number of query formats, including SMILES, SMARTS, InChI, CID, molecular formula, and SDF, are supported by the Chemical Structure Search tool. The PubChem Chemical Structure Sketcher can also be used to manually construct a query chemical structure. The user does not need to download or install any other software in order to use this JavaScript-based structure editor, which is cross-platform compatible and works with all major web browsers.

Users of the Chemical Structure Search tool have the option of limiting their search to a set of CIDs submitted in a file, the results of a prior Entrez or chemical structure search, or both. The search result can be restricted using optional filters depending on a variety of factors, including molecular weight, the number of heavy atoms, whether stereochemistry is present or not, the name or category of the depositor, and others. One can export a query to an XML file, import the query from the XML file, and repeat the search without having to fill out the search form once more. This XML file can also be used as a model for creating PUG interface questions.

Applications:

Numerous domains, including chemical biology, medicinal chemistry, and informatics research have utilised the vast amount of chemical structure and bioactivity data in PubChem and its online services. Drug discovery is supported by PubChem in various ways, including lead identification and optimization, compound-target profiling, polypharmacology research, and the clarification of unidentified chemical identities. Additionally useful for creating auxiliary databases, informatics tools, and online services, PubChem has grown into a platform. With its rising public accessibility, the PubChem resource provides assistance and fantastic opportunities for the investigation of pharmacological mechanisms and the genetic underpinnings of disorders, both of which are essential for drug development and repurposing.

A relatively recent concept in chemistry, chemoinformatics is based on the processing of information on chemical and molecular structures using computational analysis. These data can be analysed to examine the connection between molecular activity, chemical characteristics, and chemical structure. It is an in-silico technique, which is a type of scientific investigation carried out virtually on a computer using software and simulations. Choosing an illness to target is the first step in the typical drug discovery process. Next, prospective chemicals and molecules that could be used to lessen the disease's severity in some way are looked for.

Virtual screening can be used to filter out specific compounds early on that aren't compatible without the requirement for physical screening in order to lower expenses and accelerate drug development while screening for new prospective compounds that could be developed into medications. This strategy, which is

significantly more effective than traditional approaches, builds virtual screens and simulations using computer software to search for prospective compounds that could be turned into medications. In order to locate new prospective compounds to be utilised as medications, chemoinformatics and drug discovery depend on the 3D structures of compounds as well as their representation in 1D and 2D forms.

REFERENCES:

1. Kim, S., Thiessen, P., Bolton, E., Chen, J., Fu, G., & Gindulyte, A. et al. (2015). PubChem Substance and Compound databases. *Nucleic Acids Research*, 44(D1), D1202-D1213. doi: 10.1093/nar/gkv951
2. Cheng, T., Pan, Y., Hao, M., Wang, Y., & Bryant, S. (2014). PubChem applications in drug discovery: a bibliometric analysis. *Drug Discovery Today*, 19(11), 1751-1756. doi: 10.1016/j.drudis.2014.08.008
3. Prosser, K., Stokes, R., & Cohen, S. (2020). Evaluation of 3-Dimensionality in Approved and Experimental Drug Space. *ACS Medicinal Chemistry Letters*, 11(6), 1292-1298. doi: 10.1021/acsmedchemlett.0c00121
4. Polanski, J., & Gasteiger, J. (2017). Computer Representation of Chemical Compounds. *Handbook Of Computational Chemistry*, 1997-2039. doi: 10.1007/978-3-319-27282-5_50
5. Jack Davis, B. (2022). How is Chemoinformatics Used in Drug Discovery?. Retrieved 5 September2022, from <https://www.azolifesciences.com/article/How-is-Chemoinformatics-Used-in-Drug-Discovery.aspx#:~:text=Chemoinformatics%20can%20drastically%20enhance%20this%20processes%2C%20as%20one,calculate%20and%20visualize%20structures%20is%20crucial.%20Virtual%20screening>

WEBLEM 1a

Retrieve 2D/3D structures using Pubchem Database

(URL: <https://pubchem.ncbi.nlm.nih.gov/>)

AIM:

To Retrieve 2D/3D structure for Penicillin (Pubchem id - 22502) using PubChem Database and apply filters as:

1. To refine search using Substructure, Exact and Similar structure approach.
2. To screen structure based on chemical properties approach.

INTRODUCTION:

Penicillin is the procaine salt form of penicillin G, a broad-spectrum, beta-lactam, naturally occurring penicillin antibiotic with antibacterial activity. Penicillin G binds to and inactivates the penicillin binding proteins (PBPs) located inside the bacterial cell wall. Inactivation of PBPs interferes with the cross-linkage of peptidoglycan chains necessary for bacterial cell wall strength and rigidity. This interrupts bacterial cell wall synthesis and results in the weakening of the bacterial cell wall, eventually causing cell lysis.

PubChem is a database open to the public that contains details on chemical compounds and their biological effects. PubChem was first introduced in 2004 as a part of the Molecular Libraries Roadmap Initiatives of the US National Institutes of Health (NIH), and since then has grown quickly to become a major source of chemical information that supports scientific communities in a variety of fields, including cheminformatics, chemical biology, medicinal chemistry, and drug discovery.

One of the greatest collections of freely accessible chemical information may be found on PubChem. As of September 2015, it included more than 157 million descriptions of chemical substances contributed by depositors, 60 million descriptions of distinctive chemical structures, and 1 million descriptions of biological assays, encompassing roughly 10,000 different protein target sequences. These enormous amounts of data are organised by PubChem into three interconnected databases called Substance, Compound, and BioAssay. Information that was provided by depositors is kept in the Substance database. The Compound database stores unique chemical structures that are taken from the Substance database. Descriptions of biological tests on chemical compounds are kept in the BioAssay database. SID (SubstanceID), CID (CompoundID), and AID (AssayID), respectively, are the main identifiers for the Substance, Compound, and BioAssay databases.

Data filtering is crucial because redundant or unreliable pieces of information can make users confused or disoriented. Results can also be more effective by filtering the data. Data filters can also be used in other situations to limit access to private information.

METHODOLOGY:

- Open homepage for PubChem database
- Search for compound “Penicillin”
- Open result for best match
- Refine results obtained for Penicillin using Substructure, Exact and similar structure approach
- Screen results for penicillin using chemical properties approach
- Observe and interpret the results

OBSERVATIONS:

National Library of Medicine
National Center for Biotechnology Information

PubChem About Posts Submit Contact

Explore Chemistry

Quickly find chemical information from authoritative sources

Try covid-19 aspirin EGFR C9H₈O₄ 57-27-2 C₁=CC=C(C=C₁)C=O InChI=1S/C3H₆O/c1-3(2)4/h1-2H3

Use Entrez Compounds Substances BioAssays

FIG 1. Homepage of Pubchem Database

National Library of Medicine
National Center for Biotechnology Information

PubChem About Posts Submit Contact

SEARCH FOR ✖

Treating this as a text search.

BEST MATCH

Cillin; Benzopenicillin; Dropcillin; Gelacillin; Liquacillin; Pharmacillin; Cilopen; Galofak; ...

Compound CID: 2349
MF: C₁₆H₁₈N₂O₅S MW: 334.4g/mol
IUPAC Name: 3,3-dimethyl-7-oxo-6-[(2-phenylacetyl)amino]-4-thia-1-azabicyclo[3.2.0]heptane-2-carboxylic acid
Isomeric SMILES: CC1C(N2CC(C1)C(C(=O)N(C)C3=CC=CC=C3)C(=O)O)C
InChIKey: JGSARLDUGVTE-UHFFFAOYSA-N
InChI: InChI=1S/C16H18N2O4S/c1-16(2)12(15(2)2)18-13(20)11(14(18)23-16)17-10(19)8-9-6-4-3-5-7-9/h3-7,11-12,14H,8H2,1-2H3,(H,17,19)(H,21,22)
Create Date: 2005-03-25

[Summary](#) [Similar Structures Search](#) [Related Records](#)

Compounds (308)	Substances (1,620)	Genes (57)	Proteins (258)	Pathways (4)	BioAssays (6,433)	Literature (66,372)	Patents (5,126)
Searching chemical names and synonyms including IUPAC names and InChIKeys across the compound collection. Note that annotations text from compound summary pages is not searched. Read More...							
308 results	<input type="button" value="Filters"/>	SORT BY <input type="button" value="Relevance"/>	<input type="button" value="Download"/>	<input type="button" value="Search in Entrez"/>	ACTIONS ON RESULTS WITH ID TYPE: Compounds		
<p>Procaine Penicillin; 6130-64-9; Penicillin 100; Penicillin G Procaine Hydrate; Penicillin G Procaine [USP]; ...</p> <p>Compound CID: 22502 MF: C₂₉H₄₀N₂O₅S MW: 588.7g/mol</p>							

Fig 2. Hit page for compound “Penicillin”

COMPOUND SUMMARY

Penicillin G procaine

PubChem CID 22502

Structure

Find Similar Structures

Chemical Safety

Irritant Health Hazard

Laboratory Chemical Safety Summary (LCSS) Datasheet

Molecular Formula C₂₉H₄₀N₄O₇S

6130-64-9
Procaine penicillin
Penicillin 100
Penicillin G procaine hydrate
Penicillin G procaine [USP]

Synonyms

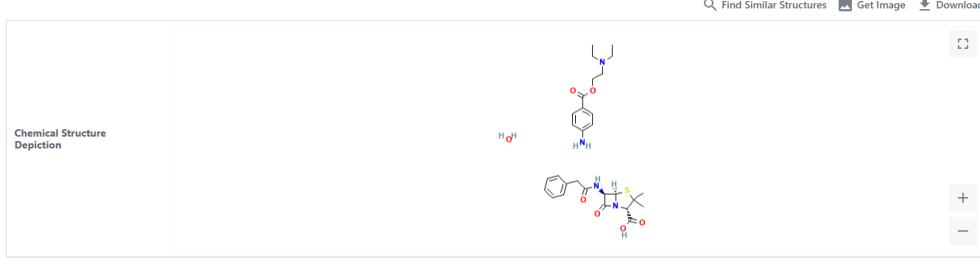
Molecular Weight 588.7

Cite	Download
CONTENTS	
Title and Summary	
1 Structures	▼
2 Names and Identifiers	▼
3 Chemical and Physical Properties	▼
4 Related Records	▼
5 Chemical Vendors	▼
6 Drug and Medication Information	▼
7 Food Additives and Ingredients	▼
8 Pharmacology and Biochemistry	▼
9 Use and Manufacturing	▼
10 Safety and Hazards	▼
11 Toxicity	▼
12 Associated Disorders and Diseases	▼
13 Literature	▼
14 Patents	▼
15 Classification	▼
16 Information Sources	▼

Fig 3. Result page with summary for Penicillin

1 Structures

1.1 2D Structure



1.2 3D Status

Conformer generation is disallowed since mixture or salt

▶ PubChem

2 Names and Identifiers

2.1 Computed Descriptors

2.1.1 IUPAC Name

Cite	Download
CONTENTS	
Title and Summary	
1 Structures	▼
2 Names and Identifiers	▼
3 Chemical and Physical Properties	▼
4 Related Records	▼
5 Chemical Vendors	▼
6 Drug and Medication Information	▼
7 Food Additives and Ingredients	▼
8 Pharmacology and Biochemistry	▼
9 Use and Manufacturing	▼
10 Safety and Hazards	▼
11 Toxicity	▼
12 Associated Disorders and Diseases	▼
13 Literature	▼
14 Patents	▼
15 Classification	▼
16 Information Sources	▼

Fig 3.1 Structure information for Penicillin

2 Names and Identifiers[?](#) [🔗](#)**2.1 Computed Descriptors**[?](#) [🔗](#)**2.1.1 IUPAC Name**[?](#) [🔗](#)

2-(diethylamino)ethyl 4-aminobenzoate;(2S,5R,6R)-3,3-dimethyl-7-oxo-6-[(2-phenylacetyl)amino]-4-thia-1-azabicyclo[3.2.0]heptane-2-carboxylic acid;hydrate
Computed by Lexichem TK 2.7.0 (PubChem release 2021.05.07)

▶ PubChem

2.1.2 InChI[?](#) [🔗](#)

InChI=1S/C16H18N2O4S.C13H20N2O2.H2O/c1-16(2)12(15(21)22)18-13(20)11(14(18)23-16)17-10(19)8-9-6-4-3-5-7-9;1-3-15(4-2)9-10-17-13(16)11-5-7-12(14)8-6-11/h3-7,11-12,14H,8H2,1-2H3,(H,21,22);5-8H,3-4-9-10,14H2,1-2H3;1H2/t11-,12+,14-;/m1..s1

Computed by InChI 1.0.6 (PubChem release 2021.05.07)

▶ PubChem

2.1.3 InChIKey[?](#) [🔗](#)

KZDCMKVLEYCGQX-UDPGNNSCCSA-N

Computed by InChI 1.0.6 (PubChem release 2021.05.07)

▶ PubChem

2.1.4 Canonical SMILES[?](#) [🔗](#)

CCN(CC(COC(=O)C1=CC=C(C=C1)NCC1(C(N2C(S1)C(C2=O)NC(=O)CC3=CC=CC=C3)C(=O)O)C)O

Computed by OEChem 2.3.0 (PubChem release 2021.05.07)

▶ PubChem

2.1.5 Isomeric SMILES[?](#) [🔗](#)**Fig 3.2. Names and Identifiers for Penicillin****3 Chemical and Physical Properties**[?](#) [🔗](#)**3.1 Computed Properties**[?](#) [🔗](#)

Property Name	Property Value	Reference
Molecular Weight	588.7	Computed by PubChem 2.1 (PubChem release 2021.05.07)
Hydrogen Bond Donor Count	4	Computed by Cactvs 3.4.8.18 (PubChem release 2021.05.07)
Hydrogen Bond Acceptor Count	10	Computed by Cactvs 3.4.8.18 (PubChem release 2021.05.07)
Rotatable Bond Count	11	Computed by Cactvs 3.4.8.18 (PubChem release 2021.05.07)
Exact Mass	588.26177080	Computed by PubChem 2.1 (PubChem release 2021.05.07)
Monoisotopic Mass	588.26177080	Computed by PubChem 2.1 (PubChem release 2021.05.07)
Topological Polar Surface Area	169 Å ²	Computed by Cactvs 3.4.8.18 (PubChem release 2021.05.07)
Heavy Atom Count	41	Computed by PubChem
Formal Charge	0	Computed by PubChem
Complexity	752	Computed by Cactvs 3.4.8.18 (PubChem release 2021.05.07)
Isotope Atom Count	0	Computed by PubChem
Defined Atom Stereocenter Count	3	Computed by PubChem
Undefined Atom Stereocenter Count	0	Computed by PubChem
Defined Bond Stereocenter Count	0	Computed by PubChem
Undefined Bond Stereocenter Count	0	Computed by PubChem
Covalently-Bonded Unit Count	3	Computed by PubChem
Compound Is Canonicalized	Yes	Computed by PubChem (release 2021.05.07)

▶ PubChem

4 Related Records[?](#) [🔗](#)**4.1 Related Compounds with Annotation**[?](#) [🔗](#)

Cite	Download
CONTENTS	
Title and Summary	
1 Structures	
2 Names and Identifiers	
3 Chemical and Physical Properties	
4 Related Records	
5 Chemical Vendors	
6 Drug and Medication Information	
7 Food Additives and Ingredients	
8 Pharmacology and Biochemistry	
9 Use and Manufacturing	
10 Safety and Hazards	
11 Toxicity	
12 Associated Disorders and Diseases	
13 Literature	
14 Patents	
15 Classification	
16 Information Sources	

Fig 3.3. Chemical and Physical properties of Penicillin

4 Related Records**4.1 Related Compounds with Annotation**

9 items View More Rows & Details



Structure	Compound CID	Name	Molecular Formula	Molecular Weight, g/mol
	5903	Procaine penicillin G	C ₂₉ H ₃₈ N ₄ O ₆ S	570.7
	22502	Penicillin G procaine	C ₂₉ H ₄₀ N ₄ O ₅ S	588.7
	23570	2-(Diethylamino)ethyl 4-amino-2-chlorobenzoate:3:3-dimethyl-7-oxo-6-[(2-phenylacetyl)amino]-4-thia-1-azabicyclo[3.2.0]heptane-2-carboxylic acid	C ₂₉ H ₃₇ CIN ₄ O ₆ S	605.1
	517319	2-(Diethylamino)ethyl 4-aminobenzoate:3:3-dimethyl-7-oxo-6-[(2-phenylacetyl)amino]-4-thia-1-azabicyclo[3.2.0]heptane-2-carboxylic acid	C ₂₉ H ₃₈ N ₄ O ₆ S	570.7
	3081515	Jenicillin A	C ₄₅ H ₅₈ N ₄ O ₁₀ S ₂	905.1

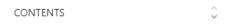
1 2 Next >

Fig 3.4. Related Records of Penicillin**5 Chemical Vendors**

Showing 1 Substance per Vendor View All

[View in Entrez](#)


3B Scientific (Wuhan) Corp	Purchasable Chemical: 3B3-059508
PubChem SID: 375108673	Purchasable Chemical: 3B3-059508
AA BLOCKS	Purchasable Chemical: AA01CC7G
PubChem SID: 381988336	Purchasable Chemical: AA01CC7G
Selleck Chemicals	Purchasable Chemical: 55515
PubChem SID: 404639703	Purchasable Chemical: 55515
BenchChem	Purchasable Chemical: B1679273
PubChem SID: 446397691	Purchasable Chemical: B1679273
Smolecule	Purchasable Chemical: 5538955
PubChem SID: 438485372	Purchasable Chemical: 5538955
THE BioTek	Purchasable Chemical: bt-278434
PubChem SID: 434303065	Purchasable Chemical: bt-278434
Ambinter	Purchasable Chemical: Amb22801185
PubChem SID: 364182307	Purchasable Chemical: Amb22801185
MedChemexpress MCE	Purchasable Chemical: HY-N7120
PubChem SID: 405010564	Purchasable Chemical: HY-N7120
LGC Standards	Purchasable Chemical: LGCFOR0302.00
PubChem SID: 340515512	Purchasable Chemical: LGCFOR0302.00
A2B Chem	Purchasable Chemical: AW55208
PubChem SID: 444240261	Purchasable Chemical: AW55208
3WAY PHARM INC	Purchasable Chemical: SWT-00733
PubChem SID: 438521790	Purchasable Chemical: SWT-00733
CymitQuimica	Purchasable Chemical: CO 6130-64-9
PubChem SID: 470666208	Purchasable Chemical: CO 6130-64-9



CONTENTS
Title and Summary
1 Structures
2 Names and Identifiers
3 Chemical and Physical Properties
4 Related Records
5 Chemical Vendors
6 Drug and Medication Information
7 Food Additives and Ingredients
8 Pharmacology and Biochemistry
9 Use and Manufacturing
10 Safety and Hazards
11 Toxicity
12 Associated Disorders and Diseases
13 Literature
14 Patents
15 Classification
16 Information Sources

Fig 3.5. Chemical vendors available for Penicillin

6 Drug and Medication Information

6.1 FDA Orange Book

1 item View More Details

Download

Trade Name	Marketing Status	Application Number	Applicant
DURACILLIN A.S.	Discontinued	A060093	ELI LILLY AND CO

► FDA Orange Book

6.2 Drug Labels for Ingredients

Label Information	Total 19 labels
Drug Ingredient	PENICILLIN G PROCAINE
NDC Code(s)	13985-036-02, 13985-036-04, 13985-036-05, 13985-522-02, 13985-522-04, 13985-522-05, 30798-213-10, 30798-213-13, 30798-213-17, 30798-236-10 ... total 48.
Packagers	A-S Medication Solutions; Aspen Veterinary; Aspen Veterinary Resources; Biomega, Inc.; Dispensing Solutions, Inc.; Durvet; MWI Veterinary Supply Co; Norbrook Laboratories Limited; Pfizer Laboratories Div Pfizer Inc; Physicians Total Care, Inc.; US Vet Inc; VetOne; Zoets Inc.

► DailyMed

6.3 Clinical Trials

6.3.1 EU Clinical Trials Register

1 item

Download

EudraCT	Title	Phase	Status	Date
2012-002836-97	Pharmacokinetics of penicillin, ampicillin and gentamicin in near- term and full-term neonates	Phase 4	Completed	2012-12-20

► EU Clinical Trials Register

Fig 3.6. Drug and Medication information for Penicillin

Fig 4. Substructure search for Penicillin

National Library of Medicine National Center for Biotechnology Information

PubChem About Posts Submit Contact

SEARCH FOR CID2349 structure

Treating this as a structure search for CID 2349. Search for CID2349 structure as text instead.

Identity (1) Similarity (>1,000) **Substructure (>1,000)** Superstructure (>1,000) 3D Similarity (>3)

Standard substructure search, finds structures in the database that contain the input structure as a part.

Percentage of the database searched: 15%. Search All

1,000 results (incomplete) SORT BY Relevance Download

Amoxicillin; Amoxycillin; 26787-78-0; Amoxicillin Anhydrous; Amoxicilline; ...

Compound CID: 33613
MF: C₁₄H₁₉N₃O₅ MW: 365.4g/mol
IUPAC Name: (2S,5R,6R)-6-[(2R)-2-amino-2-(4-hydroxyphenyl)acetyl]amino]-3,3-dimethyl-7-oxo-4-thia-1-azabicyclo[3.2.0]heptane-2-carboxylic acid
Isomeric SMILES: CC1(C(=O)N2[C@H](C1)[C@H](C=C2)O)C3=CC=C(C=C3)ON1C(=O)OC
InChIKey: LSQZLSEUYDQPKJ-NJDQSQKTSAN
InChI: InChI=1S/C16H19N3O5/c1-16(2)11(15(23)24)19-13(22)10(14(19)25-16)18-12(21)9(17)7-3-5-8(20)6-4-7/h3-6.9-11.14.20H,17H2.1-2H3.(H,18,21)(H,23,24)/t9-,10-,11+,14-/m1/s1
Create Date: 2005-06-24

Summary Similar Structures Search Related Records PubMed (MeSH Keyword)

69-53-4; Ampicillin; Aminobenzylpenicillin; Ampicillin Acid; Amcill; ...

ACTIONS ON RESULTS WITH ID TYPE: Compounds Save for Later

Fig 4.1 Hit page for substructure search for Penicillin

Structure search using exact search approach:

National Library of Medicine
National Center for Biotechnology Information

PubChem About Posts Submit Contact

SEARCH FOR
Penicillin
Treating this as a text search.

BEST MATCH

Cillin; Benzopenicillin; Dropicillin; Gelacillin; Liquacillin; Pharmacillin; Cilopen; Galofak; ...
Compound CID: 2349
MF: C9H14N2O5S MW: 334.4g/mol
IUPAC Name: 3,3-dimethyl-7-oxo-6-[(2-phenylacetyl)amino]-4-thia-1-azabicyclo[3.2.0]heptane-2-carboxylic acid
Isomeric SMILES: CCC1(C(N2C(C1)C(=O)NC(=O)CC3=CC=C3)C(=O)O)C
InChIKey: JGSARLDUGVTE-UHFFFAOYSA-N
InChI: InChI=1S/C16H18N2O4S/c1-16(2)12(15(21)22)18-13(20)11(14(18)23-16)17-10(19)8-9-6-4-3-5-7-9/h3-7,11-12,14H,8H2,1-2H3,(H,17,19)(H,21,22)
Create Date: 2005-03-25

Summary Similar Structures Search Related Records

Compounds (308) Substances (1,620) Genes (57) Proteins (258) Pathways (4) BioAssays (6,433) Literature (66,372) Patents (5,126)

Searching chemical names and synonyms including IUPAC names and InChIKeys across the compound collection. Note that annotations text from compound summary pages is not searched. [Read More...](#)

308 results Filters SORT BY Relevance DOWNLOAD Download

Procaine Penicillin; 6130-64-9; Penicillin 100; Penicillin G Procaine Hydrate; Penicillin G Procaine [USP]; ...
Compound CID: 22502
MF: C29H46N4O5S MW: 588.7g/mol

Search in Entrez ACTIONS ON RESULTS WITH ID TYPE: Compounds

Fig 5. Exact search for Penicillin

National Library of Medicine
National Center for Biotechnology Information

PubChem About Posts Submit Contact

SEARCH FOR
CID2349 structure
Treating this as a structure search for CID 2349. [Edit Structure](#) Search for [CID2349 structure as text](#) instead.

Identity (1) Similarity (>1,000) Substructure (>1,000) Superstructure (>1,000) 3D Similarity (>3) SETTINGS

Find structures very closely related to the input, comparing chemical connectivity, and optionally stereoisomers and isotopes.

1 result DOWNLOAD

Cillin; Benzopenicillin; Dropicillin; Gelacillin; Liquacillin; ...
Compound CID: 2349
MF: C9H14N2O5S MW: 334.4g/mol
IUPAC Name: 3,3-dimethyl-7-oxo-6-[(2-phenylacetyl)amino]-4-thia-1-azabicyclo[3.2.0]heptane-2-carboxylic acid
Isomeric SMILES: CCC1(C(N2C(C1)C(=O)NC(=O)CC3=CC=C3)C(=O)O)C
InChIKey: JGSARLDUGVTE-UHFFFAOYSA-N
InChI: InChI=1S/C16H18N2O4S/c1-16(2)12(15(21)22)18-13(20)11(14(18)23-16)17-10(19)8-9-6-4-3-5-7-9/h3-7,11-12,14H,8H2,1-2H3,(H,17,19)(H,21,22)
Create Date: 2005-03-25

Summary Similar Structures Search Related Records

ACTIONS ON RESULTS WITH ID TYPE: Compounds Push to Entrez Save for Later Linked Data Sets

Fig 5.1 Hit page for exact Search for Penicillin

Structure search using similarity search approach:

National Library of Medicine
National Center for Biotechnology Information

PubChem About Posts Submit Contact

SEARCH FOR
Penicillin
Treating this as a text search.

BEST MATCH

Cillin; Benzopenicillin; Dropicillin; Gelacillin; Liquacillin; Pharmacillin; Cilopen; Galofak; ...
Compound CID: 2349
MF: C9H14N2O5S MW: 334.4g/mol
IUPAC Name: 3,3-dimethyl-7-oxo-6-[(2-phenylacetyl)amino]-4-thia-1-azabicyclo[3.2.0]heptane-2-carboxylic acid
Isomeric SMILES: CC1(C(N2C(S(=O)(=O)CC3=CC=C(C=C3)C(=O)O)C2=O)C=C3)C=C(C=C3)N1C(=O)O
InChIKey: IGSARLDLUGVTF-UHFFFAOYSA-N
InChI: InChI=1S/C16H18N2O4S/c1-16(2)12(15(21)22)18-13(20)11(14(18)23-16)17-10(19)8-9-6-4-3-5-7-9/h3-7,11-12,14H,8H2,1-2H3,(H,17,19)(H,21,22)
Create Date: 2005-03-25

Summary Similar Structures Search Related Records

Compounds (308) Substances (1,620) Genes (57) Proteins (258) Pathways (4) BioAssays (6,433) Literature (66,372) Patents (5,126)

Searching chemical names and synonyms including IUPAC names and InChIKeys across the compound collection. Note that annotations text from compound summary pages is not searched. [Read More...](#)

308 results SORT BY Relevance ACTIONS ON RESULTS WITH ID TYPE: Compounds

Penicaine Penicillin; 6130-64-9; Penicillin 100; Penicillin G Procaine Hydrate; Penicillin G Procaine [USP]; ...
Compound CID: 22502
MF: C29H46N4O5S MW: 588.7g/mol

Fig 6. Similarity search for Penicillin

National Library of Medicine
National Center for Biotechnology Information

PubChem About Posts Submit Contact

SEARCH FOR
CID2349 structure
Treating this as a structure search for CID 2349. Search for [CID2349 structure as text](#) instead.

Identity (1) **Similarity (> 1,000)** Substructure (> 1,000) Superstructure (> 1,000) 3D Similarity (> 3)

Fingerprint Tanimoto-based 2-dimensional similarity search.

Percentage of the database searched: 64%.

1,000 results (incomplete) SORT BY Relevance ACTIONS ON RESULTS WITH ID TYPE: Compounds

69-53-4; Ampicillin; Aminobenzylpenicillin; Ampicillin Acid; Amcill; ...
Compound CID: 6249
MF: C9H14N2O5S MW: 349.4g/mol
IUPAC Name: (2S,5R,6R)-6-[[2(R)-2-amino-2-phenylacetyl]amino]-3,3-dimethyl-7-oxo-4-thia-1-azabicyclo[3.2.0]heptane-2-carboxylic acid
Isomeric SMILES: CC1(CC@H)N2C(H)(S1)C@@H(C2=O)NC(=O)C@@H(C3=CC=C(C=C3)N)C(=O)O
InChIKey: AVKJERGKGKZMTKX-NJDSDQKTS-A
InChI: InChI=1S/C16H19N3O4S/c1-16(2)11(15(22)23)19-13(21)10(14(19)24-16)18-12(20)9(17)8-6-4-3-5-7-8/h3-7,9-11,14H,17H2,1-2H3,(H,18,20)(H,22,23)/t9-,10-,11+,14-/m1/s1
Create Date: 2005-06-24

Summary Similar Structures Search Related Records PubMed (MeSH Keyword)

Penicillin G; Benzylpenicillin; 61-33-6; Benzylpenicillanic Acid; Free Penicillin II; ...

Fig 6.1. Hit page for similarity search for Penicillin

To screen structures based on chemical properties approach:

PubChem

[Compounds \(308\)](#) [Substances \(1,620\)](#) [Genes \(57\)](#) [Proteins \(258\)](#) [Pathways \(4\)](#) [BioAssays \(6,433\)](#) [Literature \(66,372\)](#) [Patents \(5,126\)](#)

Searching chemical names and synonyms including IUPAC names and InChIKeys across the compound collection. Note that annotations text from compound summary pages is not searched. [Read More...](#)

308 results SORT BY

Procaine Penicillin; 6130-64-9; Penicillin 100; Penicillin G Procaine Hydrate; Penicillin G Procaine [USP]; ...

Compound CID: 22502
MF: C25H40N4O5 MW: 588.7g/mol
IUPAC Name: 2-(diethylamino)ethyl 4-aminobenzoate(2S,5R,6R)-3,3-dimethyl-7-oxo-6-[(2-phenylacetyl)amino]-4-thia-1-azabicyclo[3.2.0]heptane-2-carboxylic acidhydrate
Isomeric SMILES: CCN(COC(=O)C1=CC=C(C=C1)N2C@H(S1)C@H(C2@H)C2=O)NC(=O)CC3=CC=C(C=C3)C(=O)O)CO
InChIKey: KZDCMKVLEYCGQX-UDPGNSCSA-N
InChI: InChI=1S/C16H18N2O4S.C13H20N2O2.H2O/c1-16(2)12(15(21)22)18-13(20)11(14(18)23-16)17-10(19)8-9-6-4-3-5-7-9;1-3-15(4-2)9-10-17-13(16)11-5-7-12(14)8-6-11;/h3-7,11-12,14H,8H2,1-2H3,(H.17,19)(H.21,22);5-BH3;4-9-10,14H2,1-2H3;1H2/t11-12+,14-:/m1./s1
Create Date: 2005-08-08

[Summary](#) [Similar Structures Search](#) [Related Records](#) [PubMed \(MeSH Keyword\)](#)

Penicillin G Potassium; 113-98-4; Penicillin G Potassium Salt; Benzylpenicillin Potassium; Potassium Benzylpenicillin; ...

Compound CID: 23664709
MF: C10H17KN3O5 MW: 372.5g/mol
IUPAC Name: potassium(2S,5R,6R)-3,3-dimethyl-7-oxo-6-[(2-phenylacetyl)amino]-4-thia-1-azabicyclo[3.2.0]heptane-2-carboxylate
Isomeric SMILES: CCl((C@H)N2C(=O)C(C@H)C2=O)NC(=O)CC3=CC=C(C=C3)C(=O)O)K+
InChIKey: IVNDOLXRXUOGIU-LQDWDTQKMSA-M
InChI: InChI=1S/C16H18N2O4S.K/c1-16(2)12(15(21)22)18-13(20)11(14(18)23-16)17-10(19)8-9-6-4-3-5-7-9;/h3-7,11-12,14H,8H2,1-2H3,(H.17,19)(H.21,22);/q+1/p-1/t11-12+,14-:/m1./s1
Create Date: 2008-02-05

[Summary](#) [Similar Structures Search](#) [Related Records](#) [PubMed \(MeSH Keyword\)](#)

Penicillin G Sodium; Penicillin G Sodium Salt; 69-57-8; Benzylpenicillin Sodium; Crystapen; ...

Compound CID: 23668834

Fig 7. Applying filters to screen structure based on chemical properties

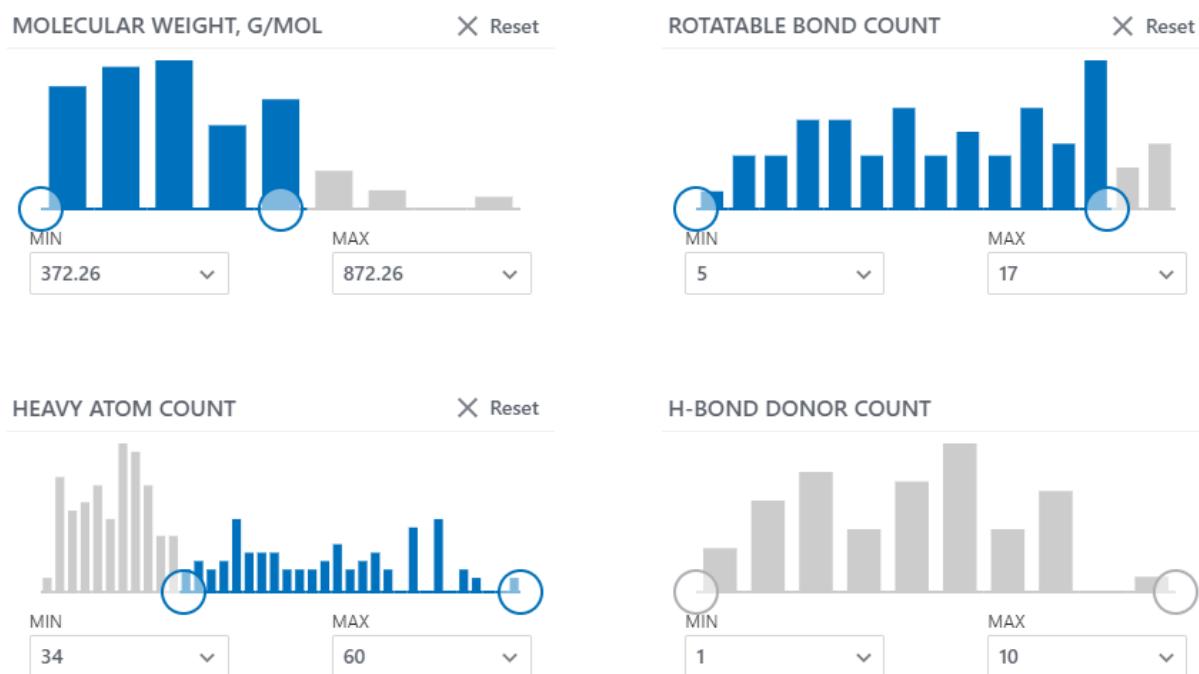


Fig 7.1 Filters applied to refinement based on chemical properties

PubChem Penicillin

Compounds (308)	Substances (1,620)	Genes (57)	Proteins (258)	Pathways (4)	BioAssays (6,433)	Literature (66,372)	Patents (5,126)
---------------------------	------------------------------	----------------------	--------------------------	------------------------	-----------------------------	-------------------------------	---------------------------

Searching chemical names and synonyms including IUPAC names and InChIKeys across the compound collection. Note that annotations text from compound summary pages is not searched. [Read More...](#)

74 results Filters (4) SORT BY Relevance Download Search in Entrez ACTIONS ON RESULTS WITH ID TYPE: Compounds Push to Entrez Save for Later Linked Data Sets

Penicaine Penicillin; 6130-64-9; Penicillin 100; Penicillin G Procaine Hydrate; Penicillin G Procaine [USP]; ...

Compound CID: 22502
MF: C23H40N4O5S MW: 588.7/g/mol
IUPAC Name: 2-(diethylamino)ethyl 4-aminobenzoate:(25.5R,6R)-3,3-dimethyl-7-oxo-6-[(2-phenylacetyl)amino]-4-thia-1-azabicyclo[3.2.0]heptane-2-carboxylic acidhydrate
Isomeric SMILES: CCN(CC(COC(=O)C1=CC=C(C=C1)N.CC1([C@H](N2[C@H](S1)[C@H](C2=O)NC(=O)C3=CC=C3)C(=O)O)C2=O)C3=CC=C3)C(=O)O)C2=O
InChIKey: KZDCMKVLEVGQX-UDPGNSCCSA-N
InChI: InChI=1S/16H18N2O4S.C13H20N2O2.H2O/c1-16(2)(12(15(2)22)18-13(20)11(14(18)23-16)17-10(19)8-9-6-4-3-5-7-9-1-3-15(4-2)9-10-17-13(16)11-5-7-12(14)8-6-11/h3-7,11-12,14H,8H2,1-2H3,(H,17,19)[H,21,22]5-8H,3-4,9-10,14H2,1-2H3/t11,-12+,4-11-/m1,-/s1
Create Date: 2005-08-06

[Summary](#) [Similar Structures Search](#) [Related Records](#) [PubMed \(MeSH Keyword\)](#)

Penicillin G Dibenzylamine; UNII-AF5F3C86Z1; DIBENZYLAMINE PENICILLIN G; AF5F3C86Z1; 7179-52-4; ...

Compound CID: 197868
MF: C30H33N3O5 MW: 531.7/g/mol
IUPAC Name: N-benzyl-1-phenylmethanamine:(25.5R,6R)-3,3-dimethyl-7-oxo-6-[(2-phenylacetyl)amino]-4-thia-1-azabicyclo[3.2.0]heptane-2-carboxylic acid
Isomeric SMILES: CCl([C@H](N1[C@H](C2=O)NC(=O)C3=CC=C3)C(=O)O)C1=CC=C(C=C1)CNCC2=CC=C2
InChIKey: UPYLDHVPOFTZGP-LQDWTKMSA-N
InChI: InChI=1S/15H16N18N2O4S.C14H15N/c1-16(2)(12(15(2)22)18-13(20)11(14(18)23-16)17-10(19)8-9-6-4-3-5-7-9-1-3-7-13(8-4-1)11-15-12-14-9-5-2-6-10-14/h3-7,11-12,14H,8H2,1-2H3,(H,17,19)[H,21,22]2-10,15H,11-12H2/t11,-12+,4-11-/m1,-/s1
Create Date: 2005-08-09

[Summary](#) [Similar Structures Search](#) [Related Records](#)

Procaine Penicillin G; 54-35-3; Duphenap; Hostacillin; Hydracillin; ...

Compound CID: 5903

Fig 7.2. Hit page for results obtained after applying filters

Results:

2D/3D structure for Penicillin (Pubchem id - 22502) were retrieved using Pubchem Database. The results showed one best match and 308 similar compounds. The best match was used to refine the search based on substructure and similarity structure approach wherein both results showed <1000 compounds. Search refinement using exact search approach gave one identical compound. The 308 compounds retrieved earlier were filtered using chemical properties approach. Filters were applied for molecular weight, rotatable bond count and heavy atom count and the search was narrowed down to 74 compounds.

Compounds:

PubChem is a public repository for information on chemical substances and their biological activities which can be used to retrieve 2D/3D structures of chemical compounds. The database also provides various search options using different ways of representing molecular structures such as SMILES, InChI key, molecular formula, Pubchem CID or even by drawing the structure. It also provides various options to refine the search to retrieve structures that have substructure of query compounds, structures that are similar or the exact match. Filter can also be applied using chemical properties to narrow down the search results. Thus, Pubchem serves as a useful resource many areas such as cheminformatics, chemical biology, medicinal chemistry and drug discovery where there is requirement for information on chemical compounds.

References:

1. What is Data Filtering? - Definition from Techopedia. (2022). Retrieved 5 September 2022, from <https://www.techopedia.com/definition/26202/data-filtering>
2. Penicillin- an overview / ScienceDirect Topics. (n.d.). [Www.sciencedirect.com](http://www.sciencedirect.com). Retrieved 5September 2022, from <https://www.sciencedirect.com/topics/neuroscience/penicillin>
3. PubChem. (2022). *The PubChem Project*. Nih.gov; National Library of Medicine. Retrieved 5September 2022, from <https://pubchem.ncbi.nlm.nih.gov/>
4. PubChem. (2022). *The PubChem Project*. Nih.gov; National Library of Medicine. Retrieved 5September 2022, from <https://pubchem.ncbi.nlm.nih.gov/#query=penicillin>
5. PubChem. (2022). *The PubChem Project*. Nih.gov; National Library of Medicine. Retrieved 5September 2022, from <https://pubchem.ncbi.nlm.nih.gov/compound/22502>

WEBLEM 2

BIOVIA Draw software & Open Babel Tool

Aim:

Introduction to Chemical Structure (1D, 2D and 3D), Drawing using BIOVIA DRAW Software and File conversion using Open Babel Tool

Introduction:

Chemical Structure:

A drug's chemical composition impacts its physicochemical characteristics, which in turn influence its ADME/Tox characteristics, which in turn influence its pharmacological action. By changing the shape of drug molecules, medicinal chemists can control the pharmacological activity of a given substance. A drug's ring systems and functional groups are crucial parts. The percentage of heavy atoms in a medicine is determined by the ratio of non-hydrogen to non-hydrocarbon atoms. The three variables have a lot of potential for determining whether organic compounds have drug-like characteristics. To the best of our knowledge, however, no studies have been done to systematically examine the simultaneous impacts of the quantity of aromatic and non-aromatic rings, the quantity of specific functional groups, and the fraction of heavy atoms on the drug-like qualities of an organic molecule.

1. BIOVIA Draw Software:

History:

Accelrys was formed in 2001 as a wholly owned subsidiary of Pharmacopeia, Inc. from the fusion of five companies: Molecular Simulations Inc., Synopsys Scientific Systems, Oxford Molecular, the Genetics Computer Group (GCG), and Synomics Ltd. MSI, itself a result of the combination of Biodesign, Cambridge Molecular Design, Polygen and, later, Biocad and BiosymTechnologies.

Introduction:

BIOVIA is a software company headquartered in the United States, with representation in Europe and Asia. It provides software for chemical, materials and bioscience research for the pharmaceutical, biotechnology, consumer packaged goods, aerospace, energy and chemical industries. Previously named Accelrys, it is a wholly owned subsidiary of Dassault Systèmes after an April 2014 acquisition and has been renamed BIOVIA.

BIOVIA Draw has the same look-and-feel as ISIS/ Draw, but brings additional speed and efficiency to chemical structure drawing:

- Continuously draw bonds, pull out rings and add atoms using all-purpose drawing tool
- Drag-and-drop commonly-used structures and chemical abbreviations onto the toolbar for reuse
- Right-click for atom, bond, fragment properties and query options
- Hover over atoms and edit them without right-click
- Quickly retrace steps using Multiple Undo/Redo
- Easily create structures with R groups for queries or enumerations
- Annotate reaction schemes with text, colour and a variety of arrow styles
- Easily create publication quality structures for inclusion in Microsoft Office Documents and presentations.

Applications:

Simulate, visualize and analyse chemical and biological systems and to communicate the results to other scientists.

Installation Steps:

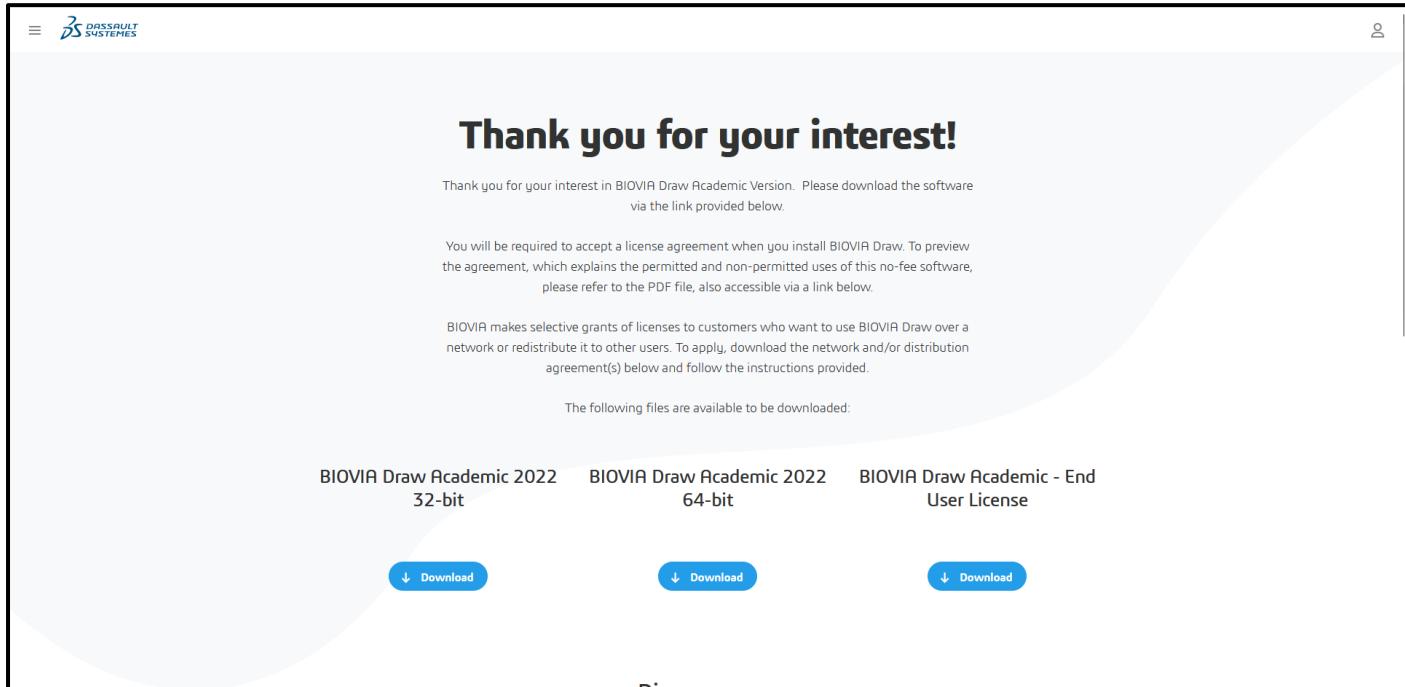
1. Open Homepage of BIOVIA DRAW software website and then click on BIOVIA Draw for Academics (URL: <https://www.3ds.com/products-services/biovia/products/scientific-informatics/biovia-draw/>).

The screenshot shows the BIOVIA DRAW software homepage. At the top, there's a navigation bar with links: PRODUCTS & SERVICES, BIOVIA, PRODUCTS, SCIENTIFIC INFORMATICS, and BIOVIA DRAW. Below the navigation is the BIOVIA logo and a yellow header bar with social media icons (Twitter, LinkedIn, Facebook, YouTube, Email) and a 'CONTACT US' button. The main content area features a large image of a molecular structure with binary code (10100101010101010) overlaid. Below this image is a text box stating: 'BIOVIA Draw enables scientists to draw and edit complex molecules, chemical reactions and biological sequences with ease, facilitating the collaborative searching, viewing, communicating, and archiving of scientific information.' Another text box below it says: 'BIOVIA Draw offers scientists unique capabilities for managing complex biological entities including the ability to register and retrieve peptides, oligonucleotides, and oligosaccharides. Scientists have access to many features including a biological sequence editor that allows the definition of custom residues and linkers, Markush structure tools, and haptic and hydrogen bond tools.' To the right, there are sections for 'Learn More' (with a 'BIOVIA Draw Datasheet' thumbnail showing two people working on a computer), 'BIOVIA Draw for Academics' (with a thumbnail of a molecular structure), and an 'Explore the Products' section with a 'CONTACT US' button.

2. Fill the registration details and click on submit

The screenshot shows the 'BIOVIA Draw for Academics' registration page. It features a large background image of a molecular structure. The title 'BIOVIA Draw for Academics' is prominently displayed in the center. Below the title, it says 'Free access for academics and non-commercial users'. A section titled 'Draw and edit complex molecules' explains that students, teachers, and researchers can download the software for free. A 'Register to download' button is provided. On the right side, there's a 'Register now' form with fields for First Name*, Last Name*, Company*, Country (India selected), City*, and Phone Number (+91 + 91234 56789). There's also a checkbox for privacy policy and a 'Submit' button.

3. Download BIOVIA DRAW Academic 2022

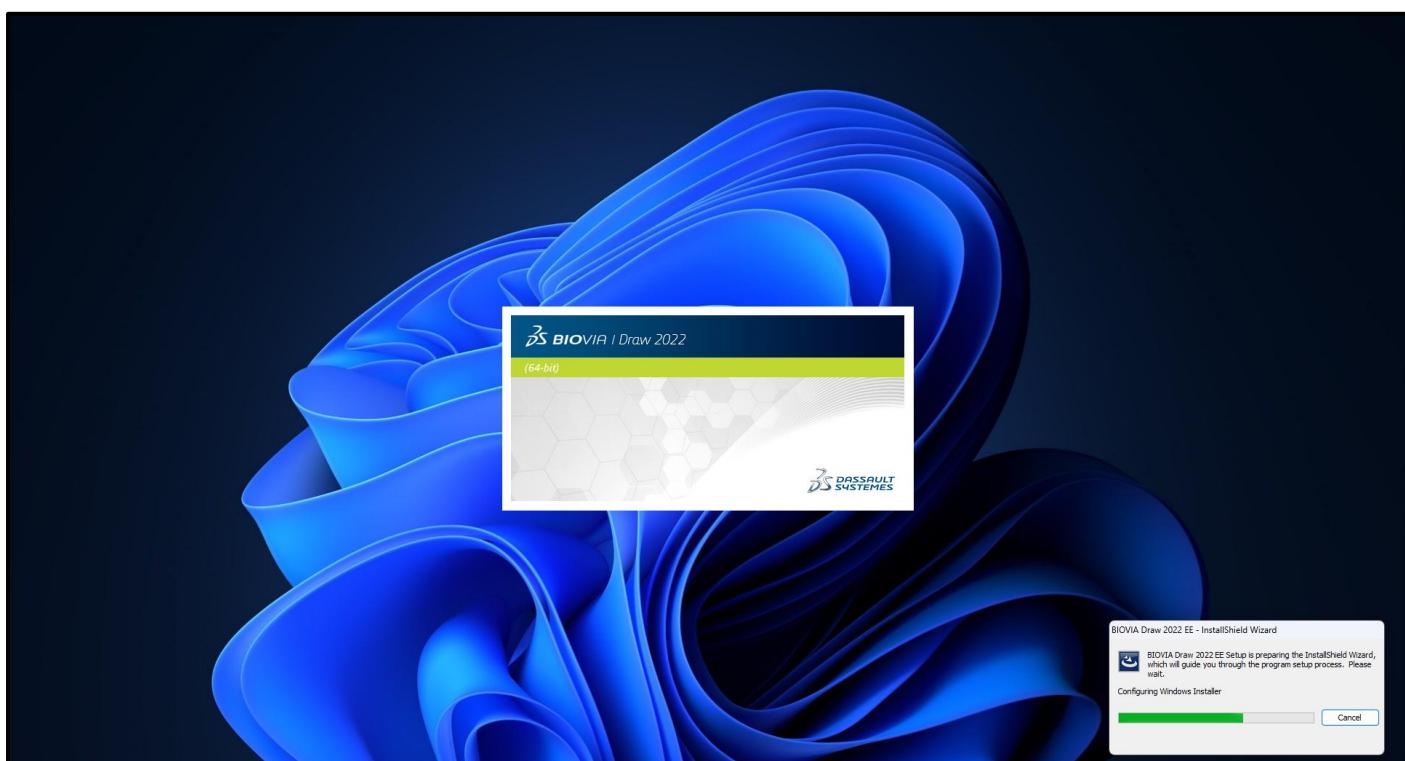


The screenshot shows the download page for BIOVIA Draw Academic 2022. At the top left is the Dassault Systèmes logo. On the right is a user icon. The main heading is "Thank you for your interest!". Below it, a message says: "Thank you for your interest in BIOVIA Draw Academic Version. Please download the software via the link provided below." A note follows: "You will be required to accept a license agreement when you install BIOVIA Draw. To preview the agreement, which explains the permitted and non-permitted uses of this no-fee software, please refer to the PDF file, also accessible via a link below." Another note states: "BIOVIA makes selective grants of licenses to customers who want to use BIOVIA Draw over a network or redistribute it to other users. To apply, download the network and/or distribution agreement(s) below and follow the instructions provided." A link to "Discover more" is at the bottom.

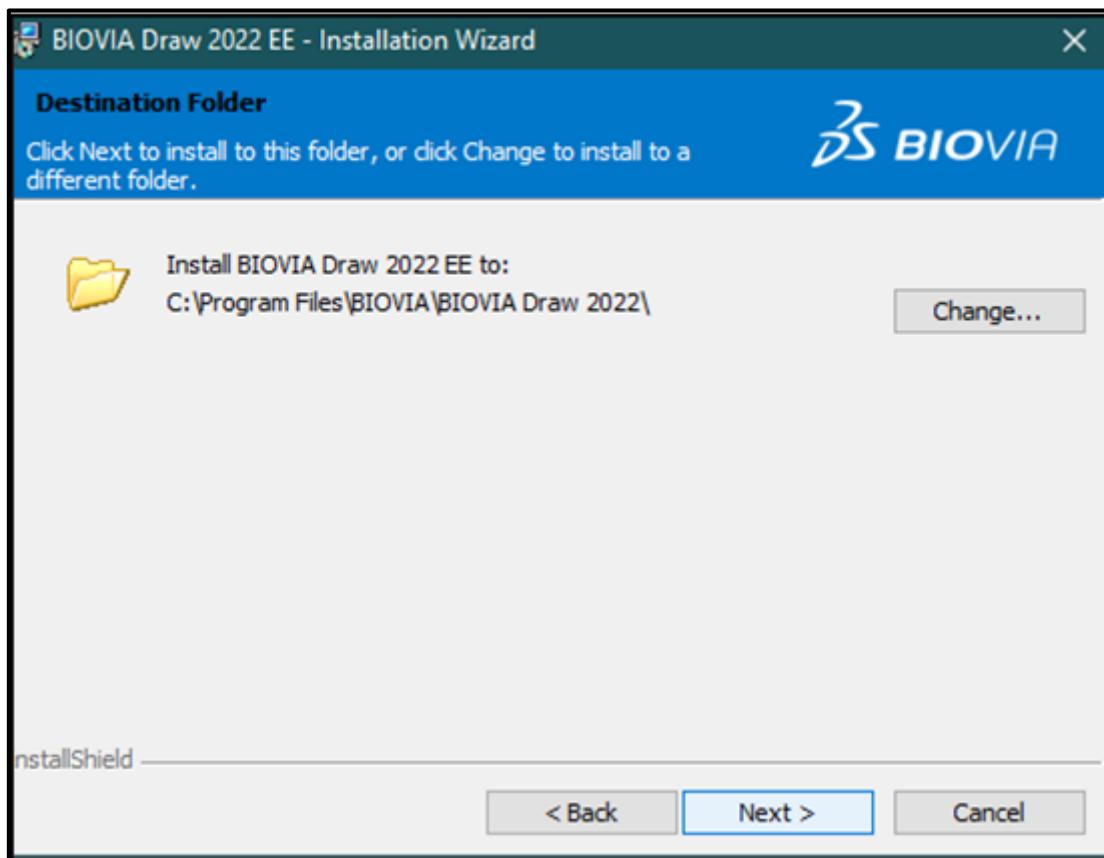
BIOVIA Draw Academic 2022
32-bit BIOVIA Draw Academic 2022
64-bit BIOVIA Draw Academic - End
User License

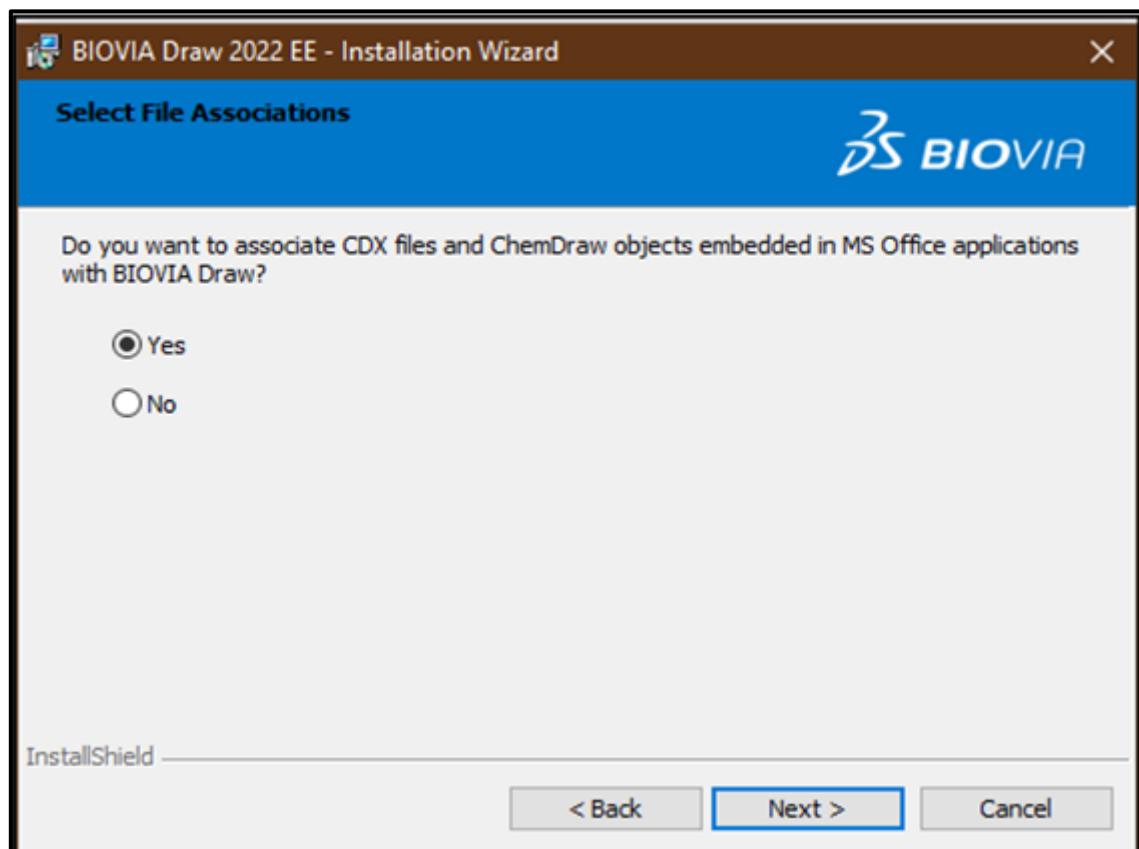
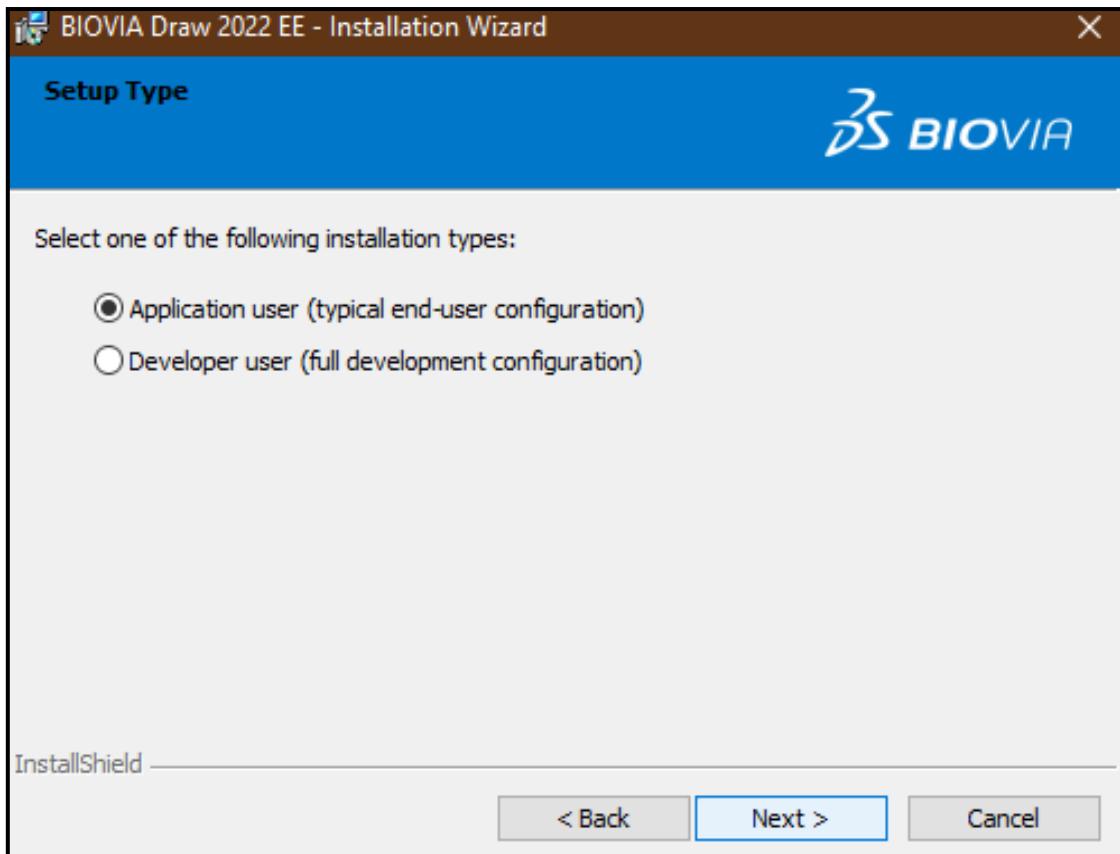
[Download](#) [Download](#) [Download](#)

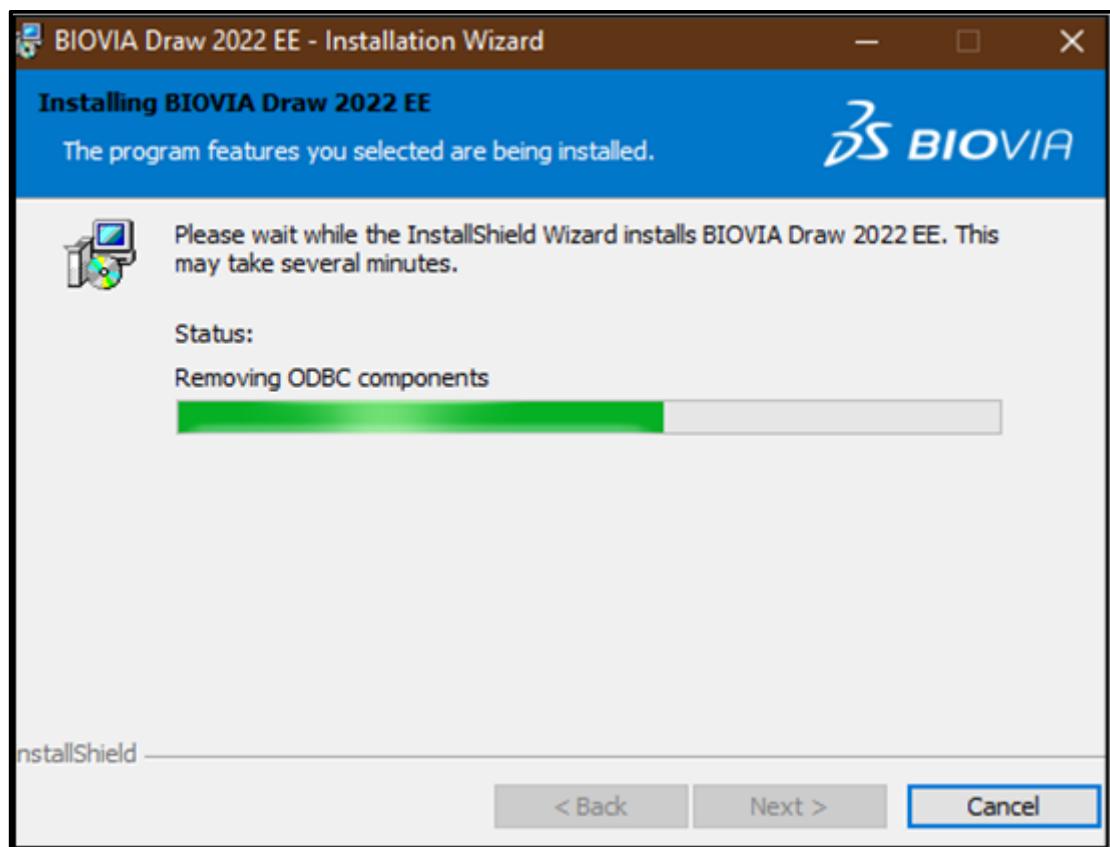
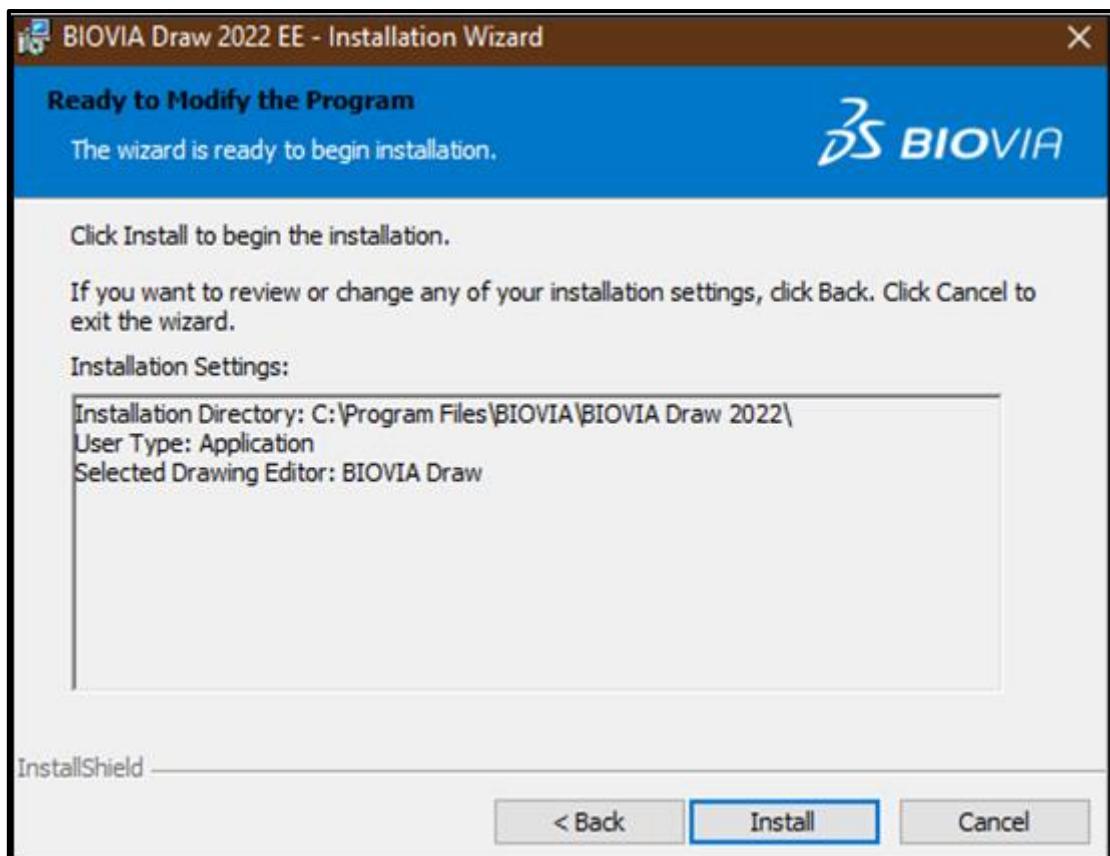
4. After Downloading, open it on your PC or Laptop



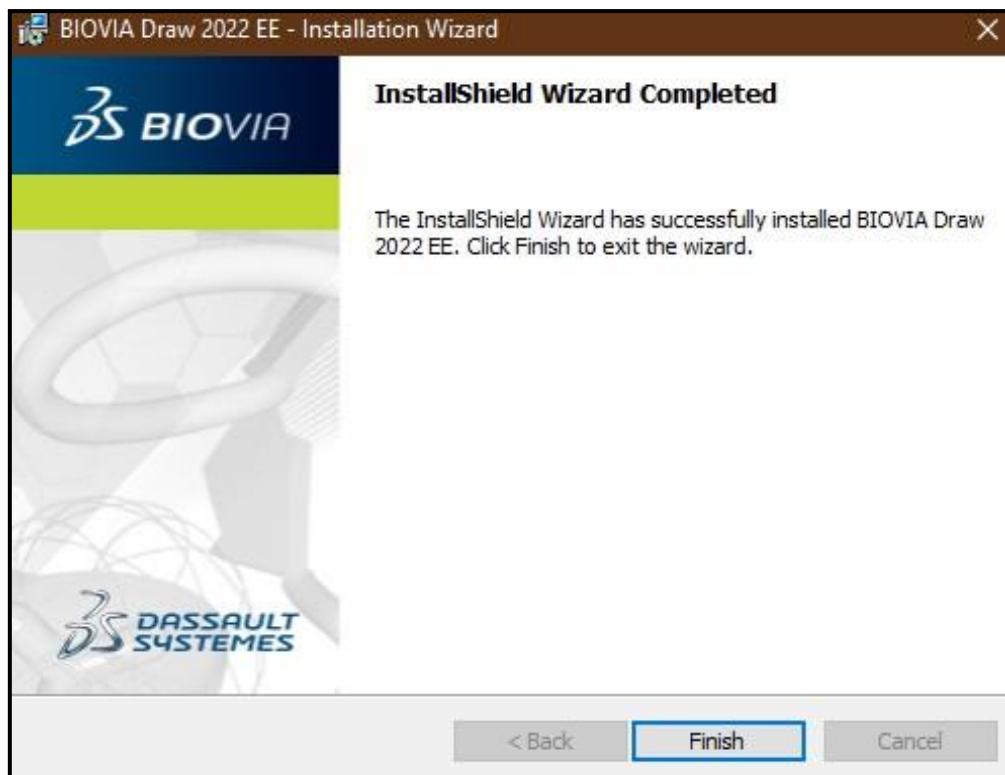
- Check and select the optional components, and then click on 'Next'



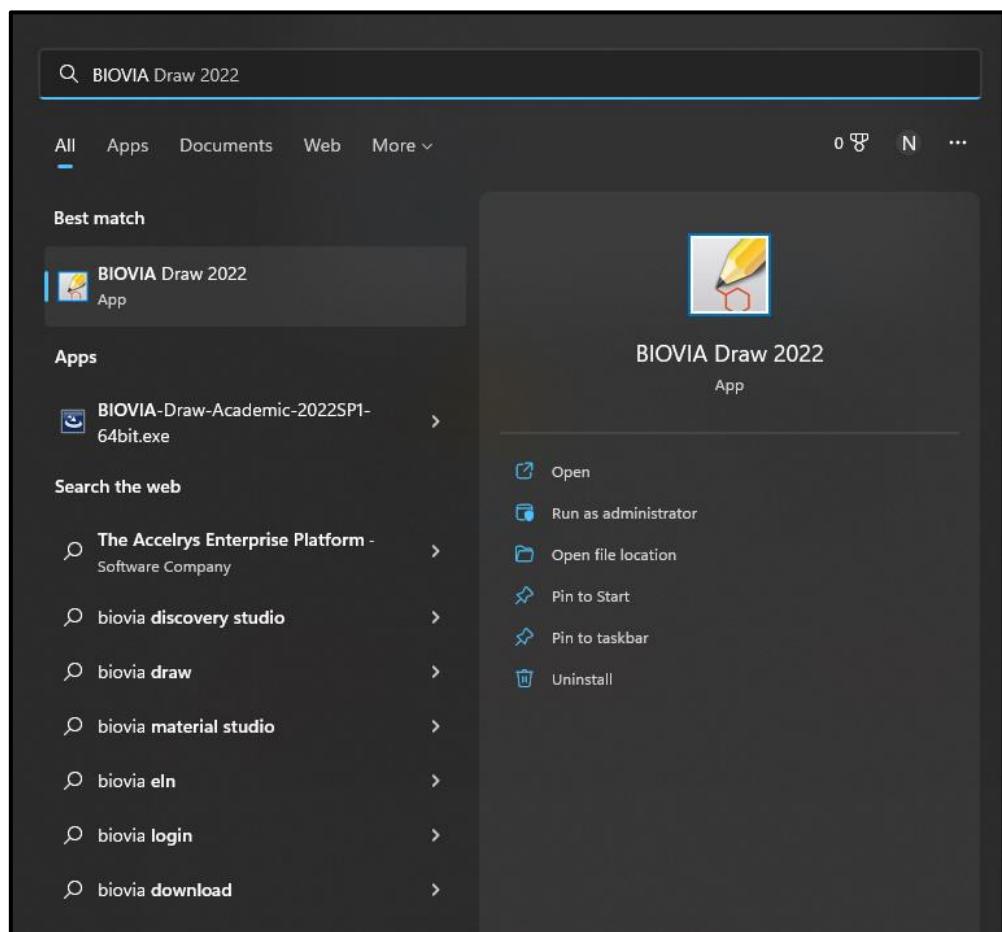




6. The Installation is completed. Click finish to exit the wizard



7. BIOVIA after installation



II. OPEN BABEL Tool

HISTORY:

Open Babel and JOELib were derived from the OELib cheminformatics library. In turn, OELib was based on ideas in the original chemistry program Babel and an unreleased object-oriented programming library called OBabel.

INTRODUCTION:

The development and use of the Open Babel project, a full-featured open chemical toolbox, designed to "speak" the many different representations of chemical data. It allows anyone to search, convert, analyze, or store data from molecular modeling, chemistry, solid-state materials, biochemistry, or related areas. It provides both ready-to-use programs as well as a complete, extensible programmer's toolkit for developing cheminformatics software. It can handle reading, writing, and interconverting over 110 chemical file formats, supports filtering and searching molecule files using Daylight SMARTS pattern matching and other methods, and provides extensible fingerprinting and molecular mechanics frameworks. We will discuss the frameworks for file format interconversion, fingerprinting, fast molecular searching, bond perception and atom typing.

Open Babel has its origin in a version of OELib released as open-source software by OpenEye Scientific under the GPL (GNU Public License). In 2001, OpenEye decided to rewrite OELib in-house as the proprietary OEChem library, so the existing code from OELib was spun out into the new Open Babel project. Since 2001, Open Babel has been developed and substantially extended as an international collaborative project using an open-source development model. It has over 160,000 downloads, over 400 citations, is used by over 40 software projects, and is freely available from the Open Babel website.

FEATURES:

- 1) File Format Support.
- 2) Fingerprints and Fast Searching.
- 3) Bond Perception and Atom Typing.
- 4) Canonical Representation of Molecules.
- 5) Coordinate Generation in 2D and 3D.

Installation Steps:

1. Open Homepage of Open Babel website and then click on download (URL: http://openbabel.org/wiki/Main_Page).

[Create account](#) [Log in](#)



Page [Discussion](#)

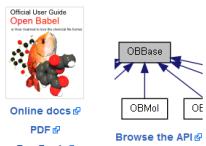
Read [View source](#) [View history](#) [Search](#)

Open Babel: The Open Source Chemistry Toolbox

Open Babel is a chemical toolbox designed to speak the many languages of chemical data. It's an open, collaborative project allowing anyone to search, convert, analyze, or store data from molecular modeling, chemistry, solid-state materials, biochemistry, or related areas.

- **Ready-to-use programs, and complete programmer's toolkit**
 - Read, write and convert over 110 chemical file formats [🔗](#)
 - Filter and search molecular files using SMARTS and other methods
 - Supports molecular modeling, cheminformatics, bioinformatics
 - Organic chemistry, inorganic chemistry, solid-state materials, nuclear chemistry
 - Downloaded over 325,000 times [🔗](#) and used by over 40 related projects
 - More about Open Babel
 - Open Babel on SourceForge [🔗](#)

To support Open Babel, please cite J. Cheminf. 2011, 3:30 [🔗](#)

Office User Guide
Open Babel
Open Babel is an open source chemical toolbox

Online docs [🔗](#) PDF [🔗](#) Buy Book [🔗](#)

Browse the API [🔗](#)

Read Paper [🔗](#)

Navigation

- [Get Open Babel](#) (downloads, installation, release notes)
- [Need Help?](#) (getting support, reporting bugs, making suggestions...)
- [Capabilities](#) (features, file formats, related projects, academic papers...)
- [Using Open Babel Programs](#) (manuals, tutorials, support lists...)
- [Develop with Open Babel](#) (how does it work, examples, developer documentation...)
- [Get Involved](#) (tell others, report bugs, suggest features, citing Open Babel...)

News

- 2016-09-21 Open Babel 2.4.0 Released [🔗](#)
- 2012-10-11 Open Babel 2.3.2 Released [🔗](#)
- 2011-10-17 Open Babel 2.3.1 Released [🔗](#)
- 2011-10-07 Open Babel Paper Published [🔗](#)
- 2011-03-24 Pre-registration open for 2nd Open Babbie [🔗](#)
- 2010-10-26 Open Babel 2.3.0 Released [🔗](#)
- 2010-07-06 Silicos NV contributes Spectrophore code [🔗](#)
- 2009-07-31 Open Babel 2.2.3 Released [🔗](#)
- 2009-07-10 Open Babel 2.2.2 Released [🔗](#)
- 2009-02-03 Open Babel 2.2.1 Released [🔗](#)
- 2008-07-04 Open Babel 2.2.0 Released [🔗](#)
- 2008-03-20 Paper on Pybel published [🔗](#)
- 2008-02-26 New OpenBabel.org Website [🔗](#)
- 2007-07-07 Open Babel 2.1.1 Released [🔗](#)
- 2007-04-07 Open Babel 2.1.0 Released [🔗](#)



InChiTRUST
SUPPORTER
2011

2. Open Babel is available on Windows, Linux and Mac. For Windows install the 64bit version.

[Create account](#) [Log in](#)



Category [Discussion](#)

Read [View source](#) [View history](#) [Search](#)

Category:Installation

(Redirected from [Get Open Babel](#))

Open Babel is available for Windows, Linux and Mac OSX.

Windows

- [Open Babel GUI](#) Provides a graphical user interface for Open Babel, as well as a command-line interface.
Get the latest installer [🔗](#) for 64-bit (recommended) or 32-bit (has "x86" in name).
[Documentation](#) [🔗](#)
- [Python module](#) [🔗](#) (requires OpenBabelGUI above) Provides access to the Open Babel libraries from Python.
- [Java library](#) [🔗](#) (requires OpenBabelGUI above) Provides access to the Open Babel libraries from Java.
- [OBDotNet assembly](#) [🔗](#) (requires OpenBabelGUI above) Provides access to the Open Babel libraries from .NET languages.

The following options are only recommended for experienced developers.

- [Compile from source](#)
 - Download the source for the latest release [🔗](#)
 - or Get latest development code [🔗](#)
 - Compile instructions [🔗](#)

Open Babel can be compiled using any of MSVC++, Cygwin or MinGW

Linux

- [Compile from source](#) Compile Open Babel:
Download the latest release [🔗](#)
or Get latest development code (today)
[How to compile](#) [🔗](#)
[How to use obabel](#) [🔗](#)
[How to develop with Open Babel](#) [🔗](#)
Scripting language modules:
[Perl](#) [🔗](#), [Python](#) [🔗](#), [Ruby](#) [🔗](#), [Java](#) [🔗](#), [Mono](#) [🔗](#)
or
- [Install a binary package](#) Several Linux distributions provide binary packages. For scripting languages, the package may be named like 'openbabel-perl' or 'python-openbabel'
Alternatively, with Conda: `conda install openbabel -c conda-forge`
Also available as a snap package: `snap install openbabel`

MacOSX

There are several ways to install Open Babel on MacOSX:

- With Conda, `conda install -c conda-forge openbabel`
- With HomeBrew, `brew install open-babel`
- [Compile the source code](#) Compile Open Babel using clang or gcc

Once installed, you may wish to try [IBabel](#) [🔗](#), a graphical interface to Open Babel.

This category has the following 3 subcategories, out of 3 total.

M

- [Macintosh](#)

R

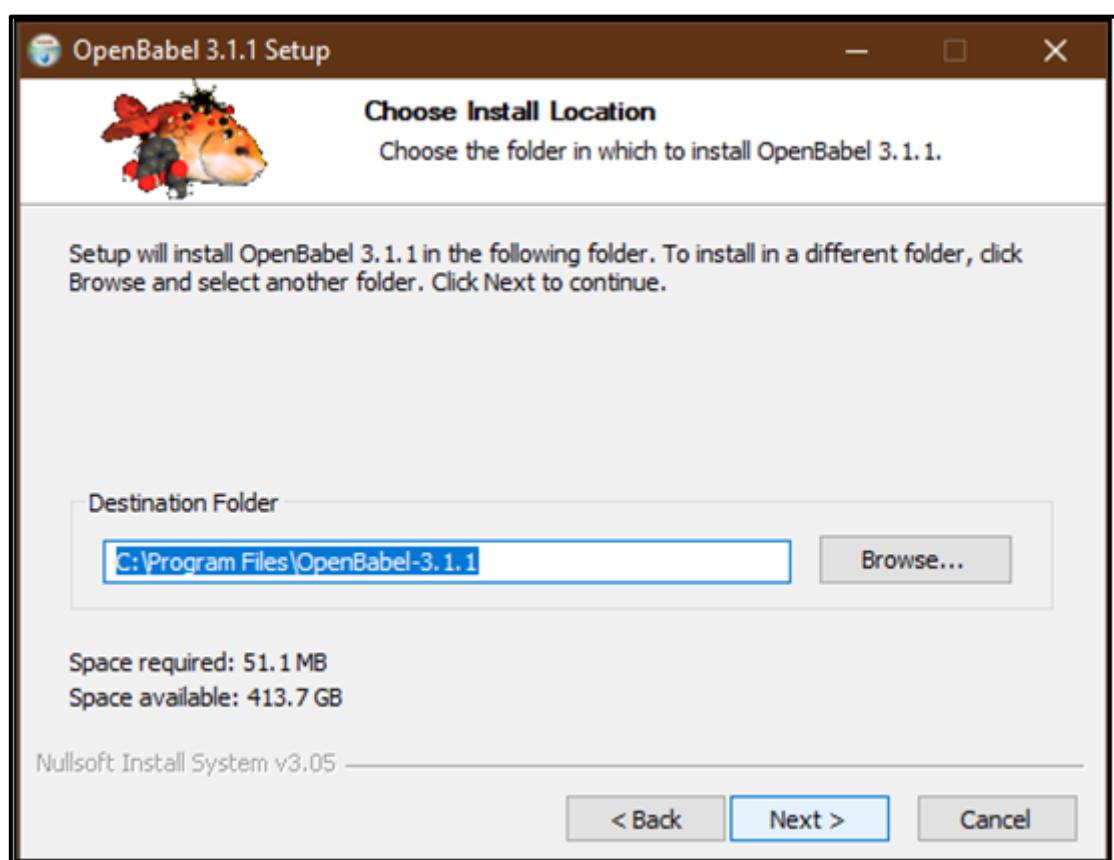
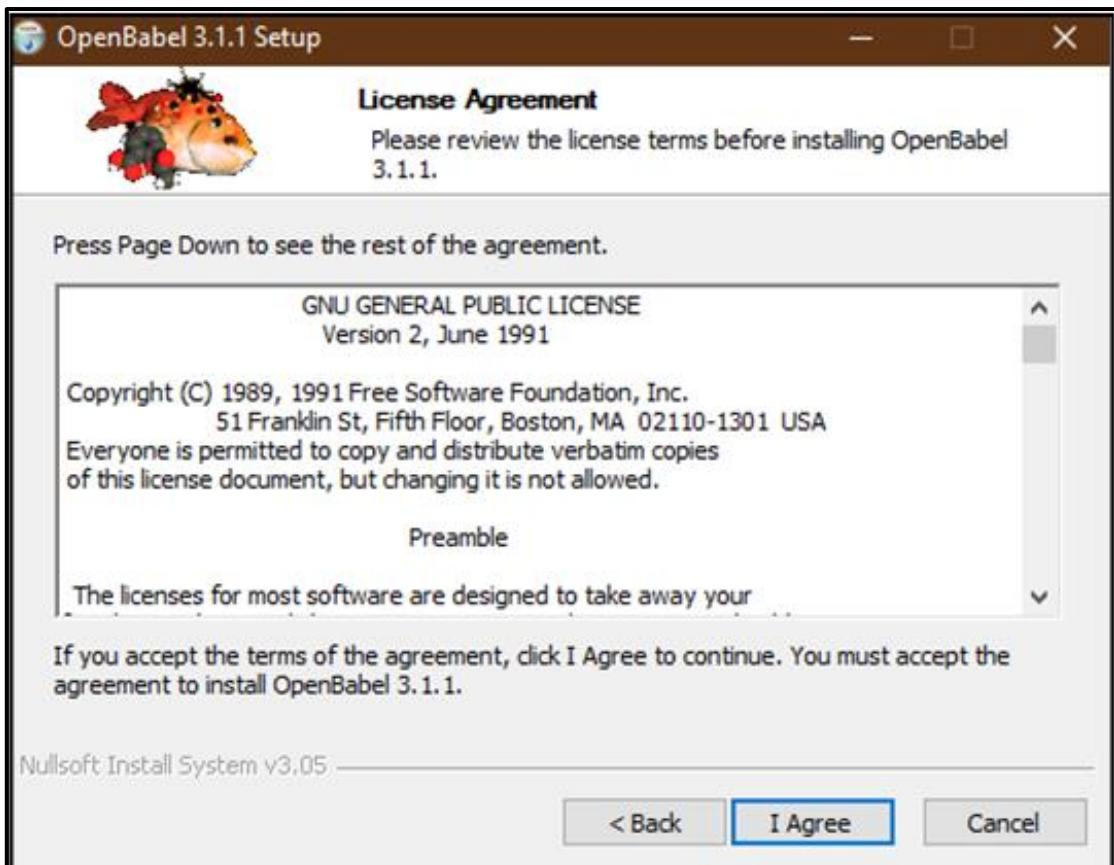
- [Releases](#)

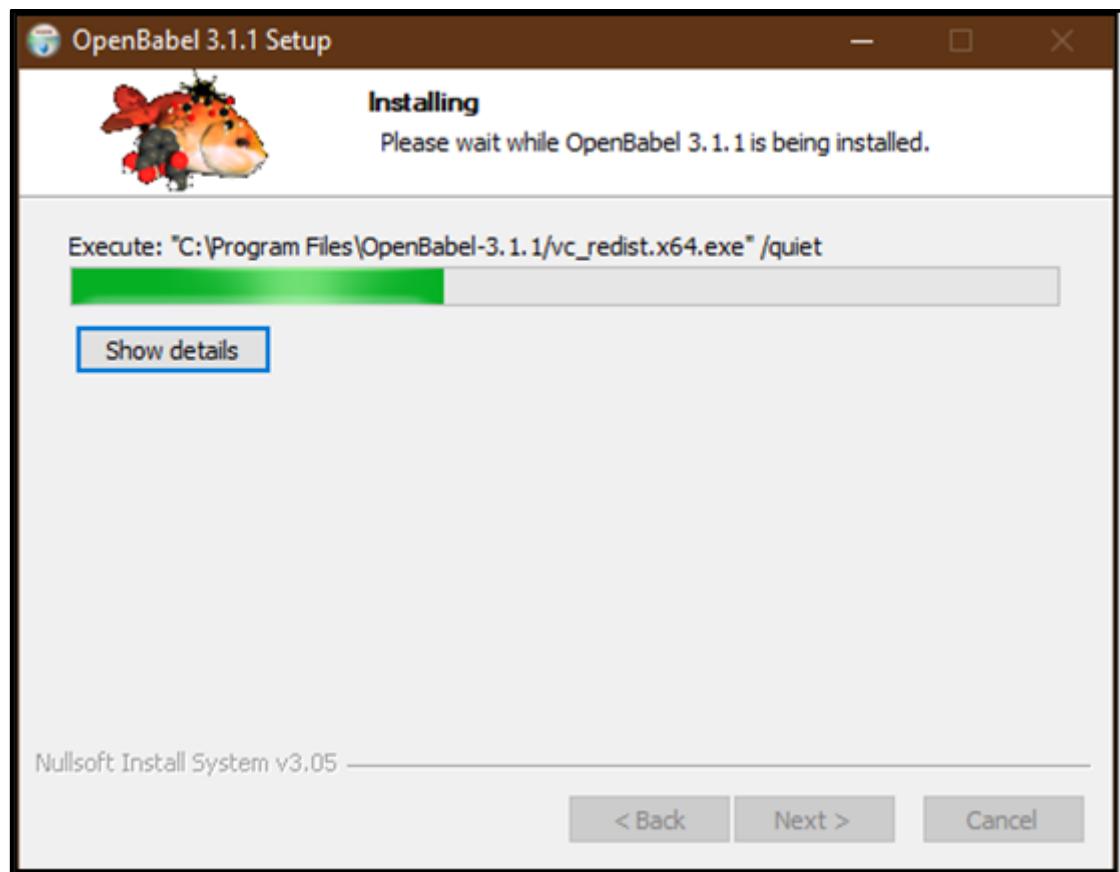
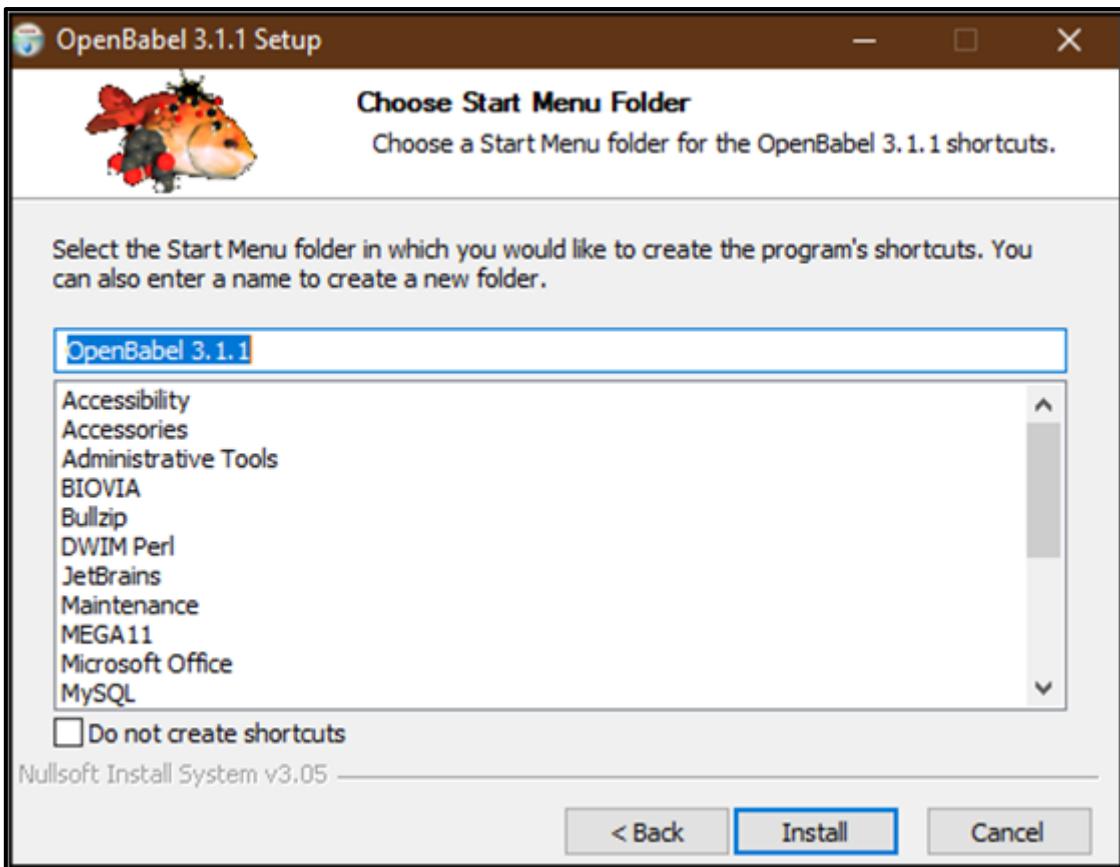
3. Click on OpenBabel-3.1.1.-64.exe

The screenshot shows the GitHub release page for the 'openbabel/openbabel' repository. The page displays the 'Open Babel 3.1.1' release, which is the latest version. It includes a summary of the release, a list of assets (source code in various formats), and a note about fixes for packaging on Linux and FreeBSD. The GitHub interface at the top shows the repository's statistics: 528 issues, 22 pull requests, 1 project, 1 wiki, security information, and insights.

4. GO through the installation procedure

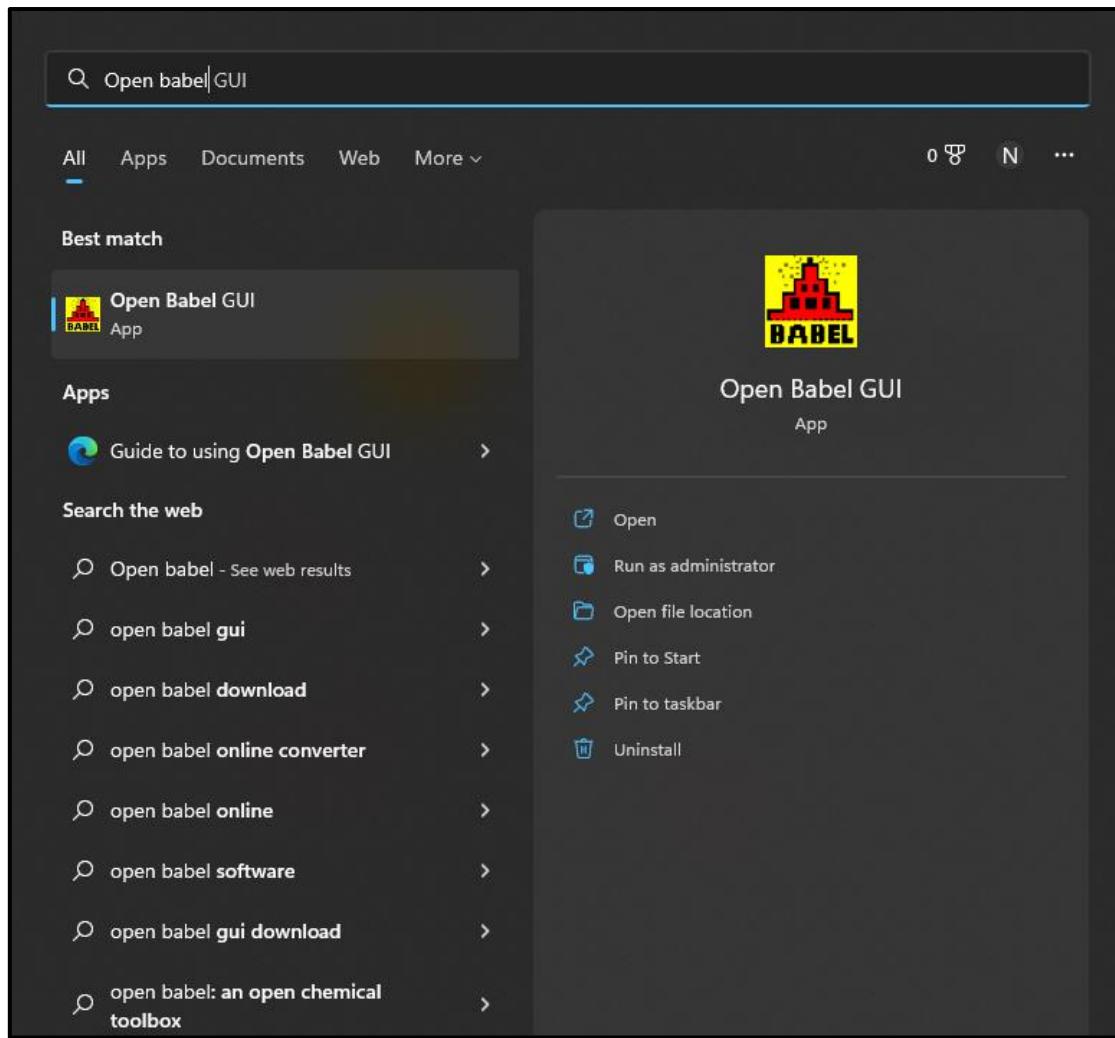








5. Icon of Open Babel GUI after installation



References:

1. Biovia draw. (n.d.). BIOVIA Draw. Retrieved September 7, 2022, from https://lib.cnu.edu.tw/3_2_eresource/doc/biovia-draw.pdf
2. Price Waterhouse Coopers (PWC) From Vision to Decision. Pharma 2020 Report. PWC; London, UK: 2012. [(accessed on 16 November 2015)].
3. Weininger, D. (1988). SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. J Chem Inf Comput Sci. 28, 31-36.

WEBLEM: 2A**BIOVIA DRAW SOFTWARE**

([URL:https://www.3ds.com/products-services/biovia/products/scientific-informatics/biovia-draw/](https://www.3ds.com/products-services/biovia/products/scientific-informatics/biovia-draw/)).

Aim:

To perform drawing, editing & manipulation of 2D & 3D structures for query “Penicillin” using BIOVIA DRAW software.

Introduction:

Penicillin V is a member of the penicillin family exhibiting broad-spectrum antibiotic property. Penicillin V binds to penicillin binding proteins (PBP), the enzymes that catalyze the synthesis of peptidoglycan, which is a critical component of the bacterial cell wall. This leads to the interruption of cell wall synthesis, consequently leading to bacterial cell growth inhibition and cell lysis.

Phenoxycephalothin is a narrow spectrum antibiotic also commonly referred to as Penicillin V or Penicillin VK. It is a phenoxycephalothin analog of Penicillin G, or [benzylpenicillin]. An orally active naturally penicillin, phenoxycephalothin is used to treat mild to moderate infections in the respiratory tract, skin, and soft tissues caused by penicillin G-sensitive microorganisms. Phenoxycephalothin has also been used in some cases as prophylaxis against susceptible organisms. While there have been no controlled clinical efficacy studies that were conducted, phenoxycephalothin has been suggested by the American Heart Association and the American Dental Association for use as an oral regimen for prophylaxis against bacterial endocarditis in patients with congenital heart disease or rheumatic or other acquired valvular heart disease when they undergo dental procedures and surgical procedures of the upper respiratory tract, except for those who are at an elevated risk for endocarditis.

With the use of BIOVIA Draw, scientists can quickly and easily sketch and edit complex compounds, chemical reactions, and biological sequences, easing group searches, viewing, sharing, and archiving of scientific data. With the ability to save and retrieve peptides, oligonucleotides, and oligosaccharides, among other complicated biological things, BIOVIA Draw provides scientists with specialised management skills. A biological sequence editor that enables the definition of unique residues and linkers, Markush structure tools, and haptic and hydrogen bond tools are just a few of the features available to scientists. Developers can build unique add-ins and drop them in as tools, buttons, or menu items using BIOVIA Draw's published API. Tools for enumeration, bioavailability, isotopomer distribution, stoichiometry calculations, and many other tasks are available as add-ins.

Methodology:

1. Open BIOVIA Draw software.
2. Draw the Chemical Structure using various tools.
3. In the chemistry option use Clean the Structure.
4. Apply different filter options such as Calculator, Atom Number, Generate Text from Structure (IUPAC, SMILE), Show sequence view, etc.
5. Observe and save the results.

Observations:

Shortcut

Header

Template

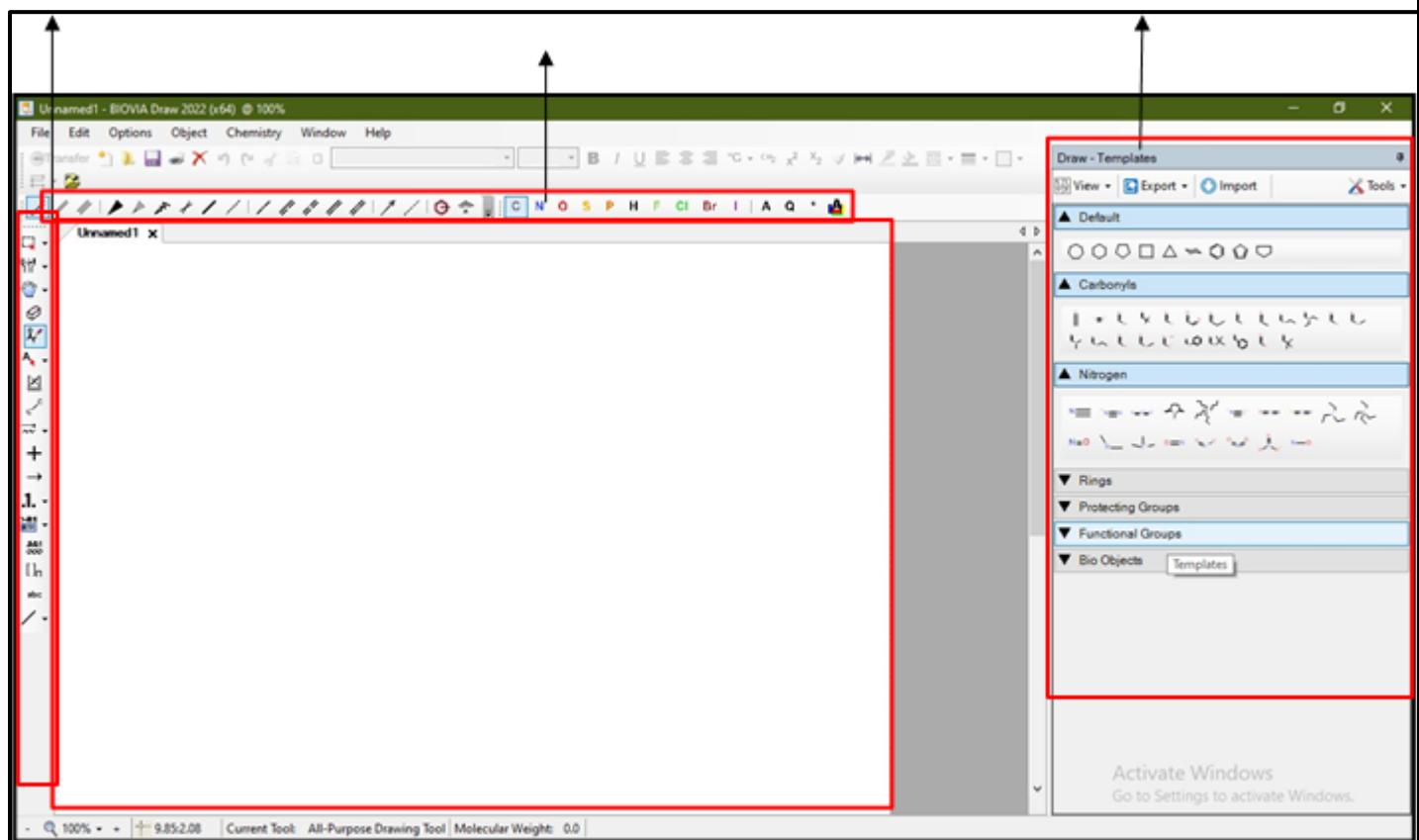


Fig 1. Homepage of BIOVIA DRAW SOFTWARE

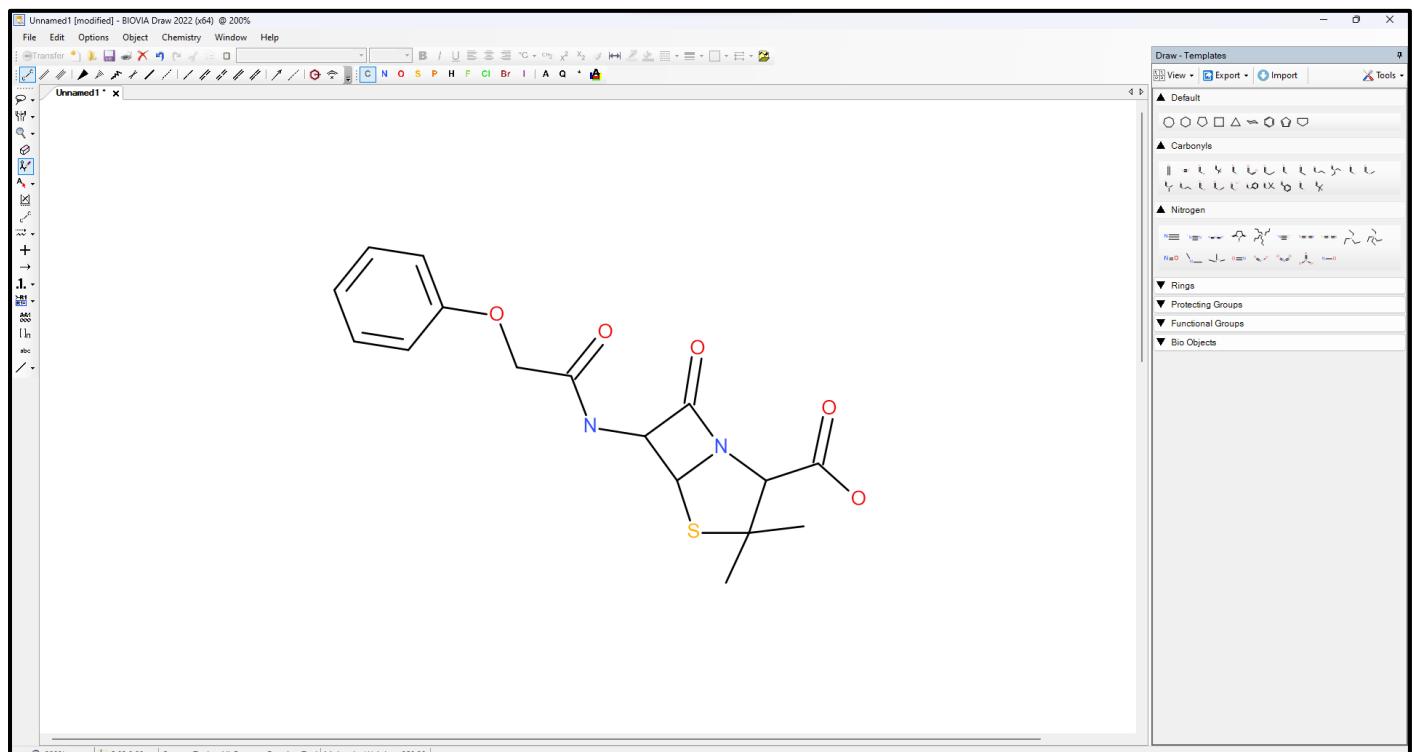


Fig 2. Structure of Penicillin chemical compound

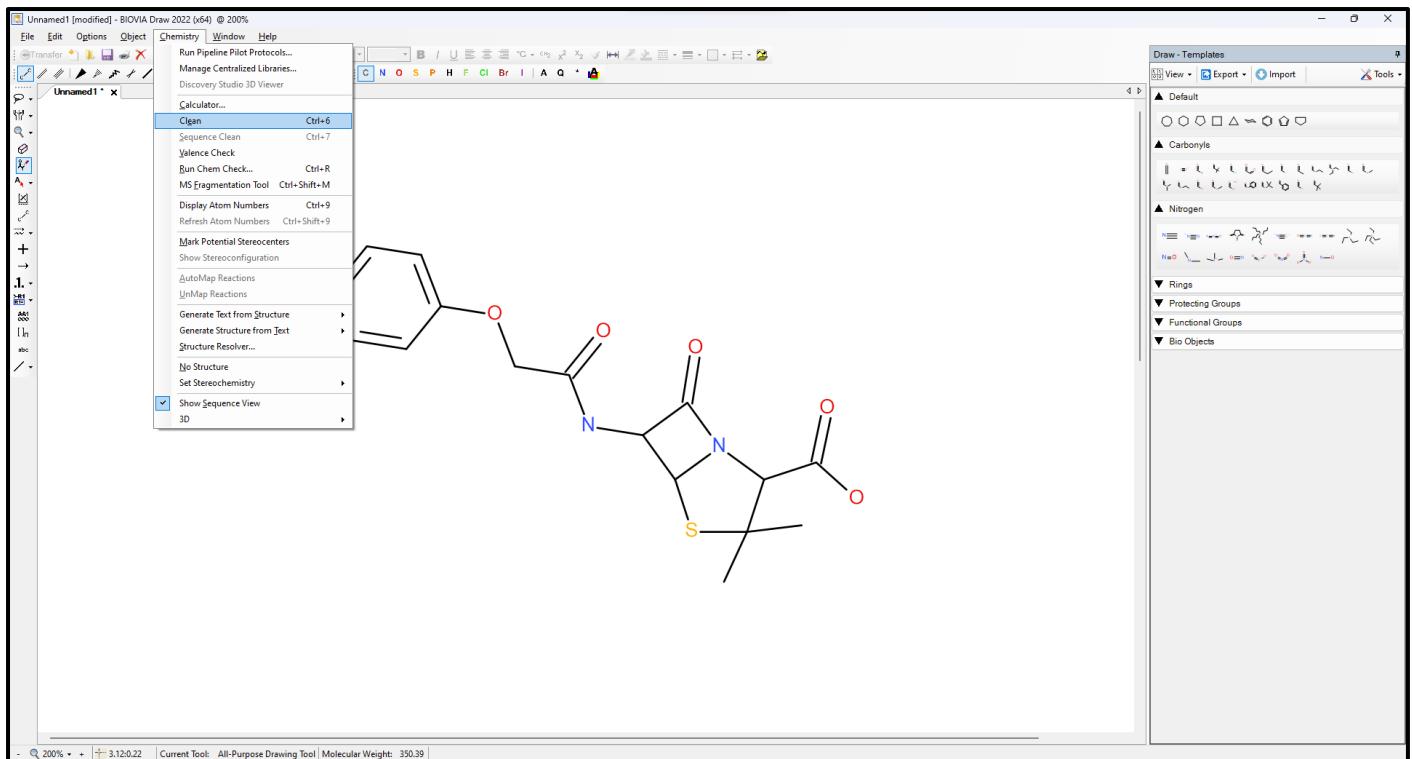


Fig 3. Option applied “Clean structure”

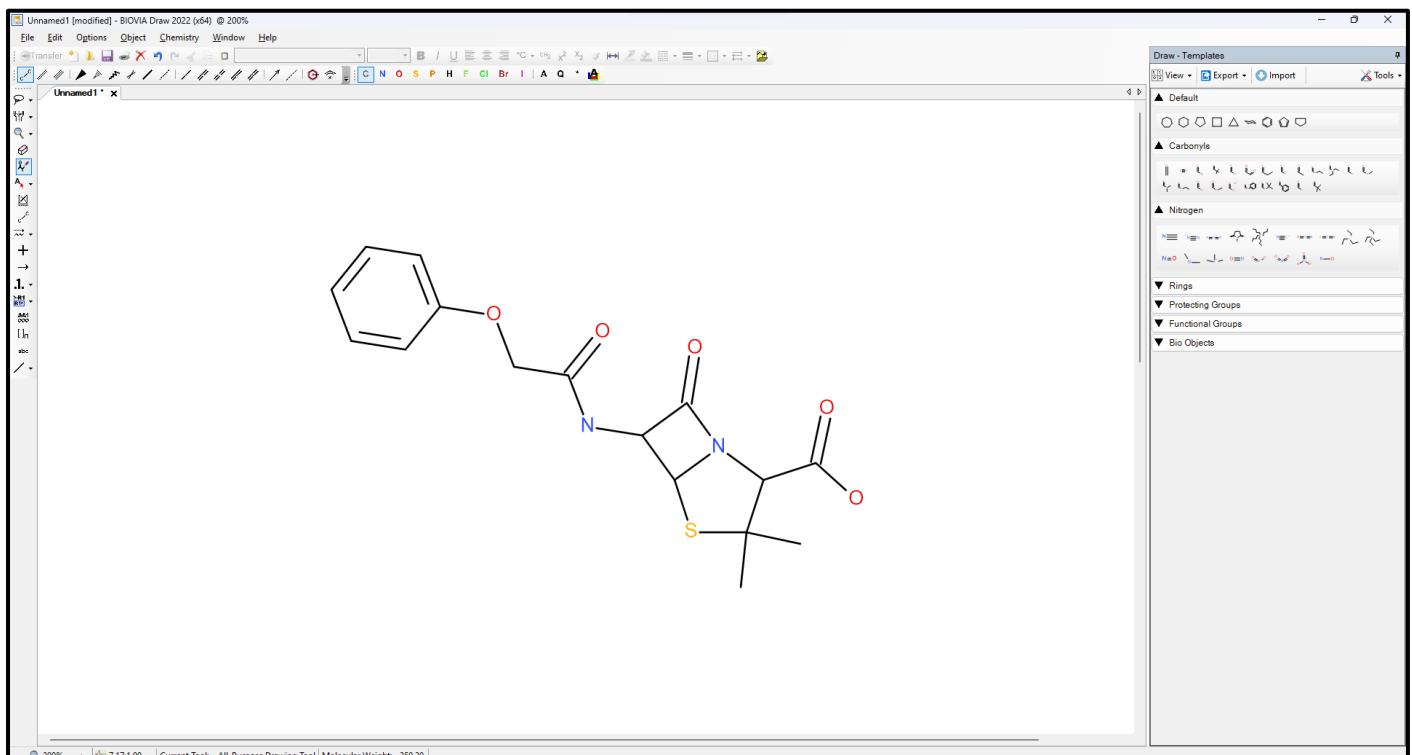


Fig 4. Clean structure of Penicillin

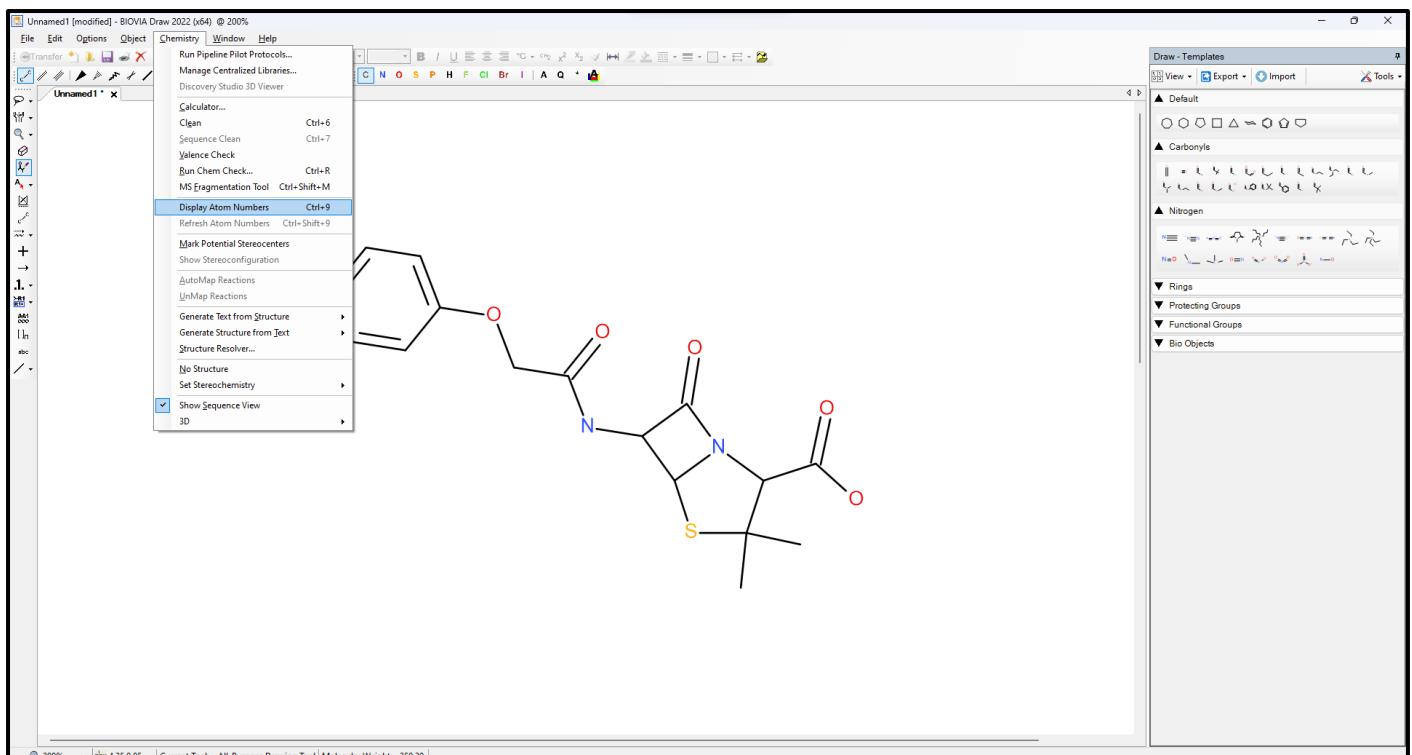


Fig 5. Option applied “Display Atom Numbers”

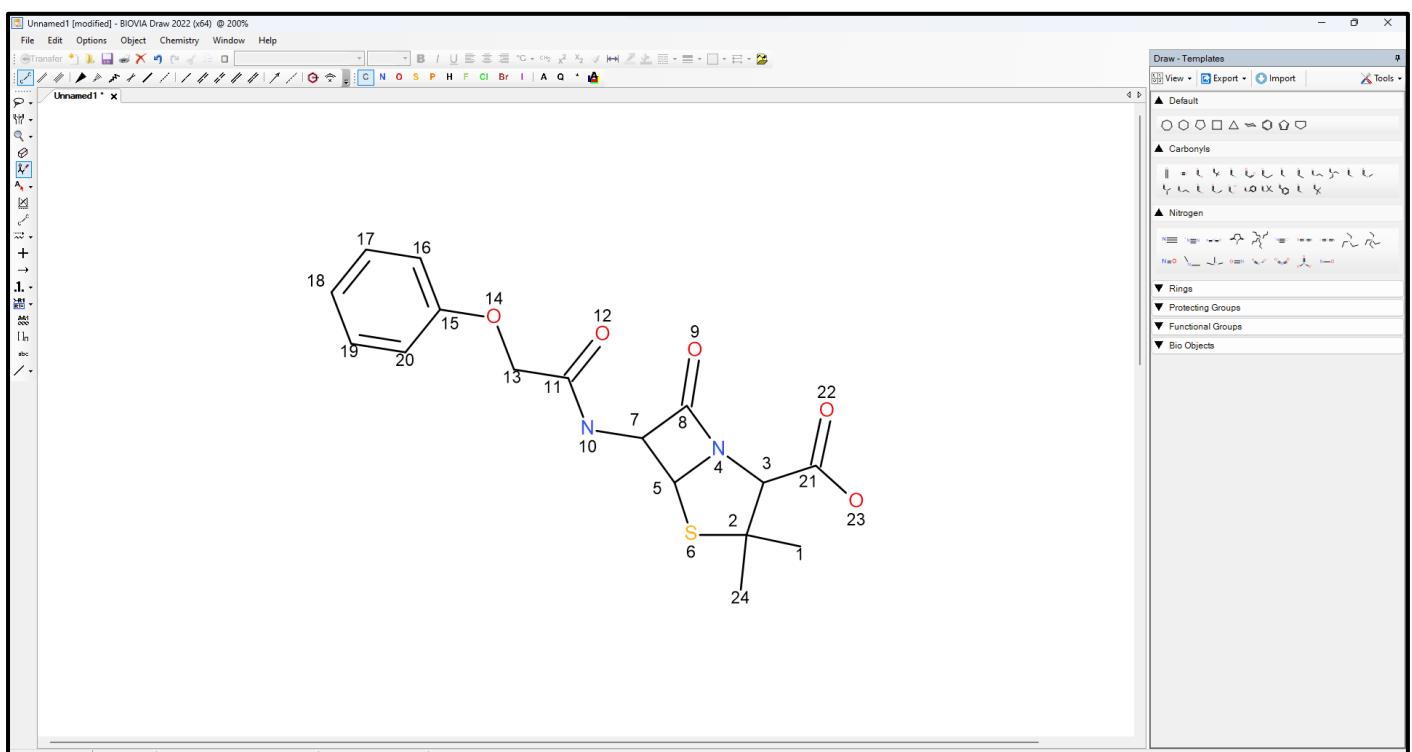


Fig 6. Atom numbering of Penicillin structure

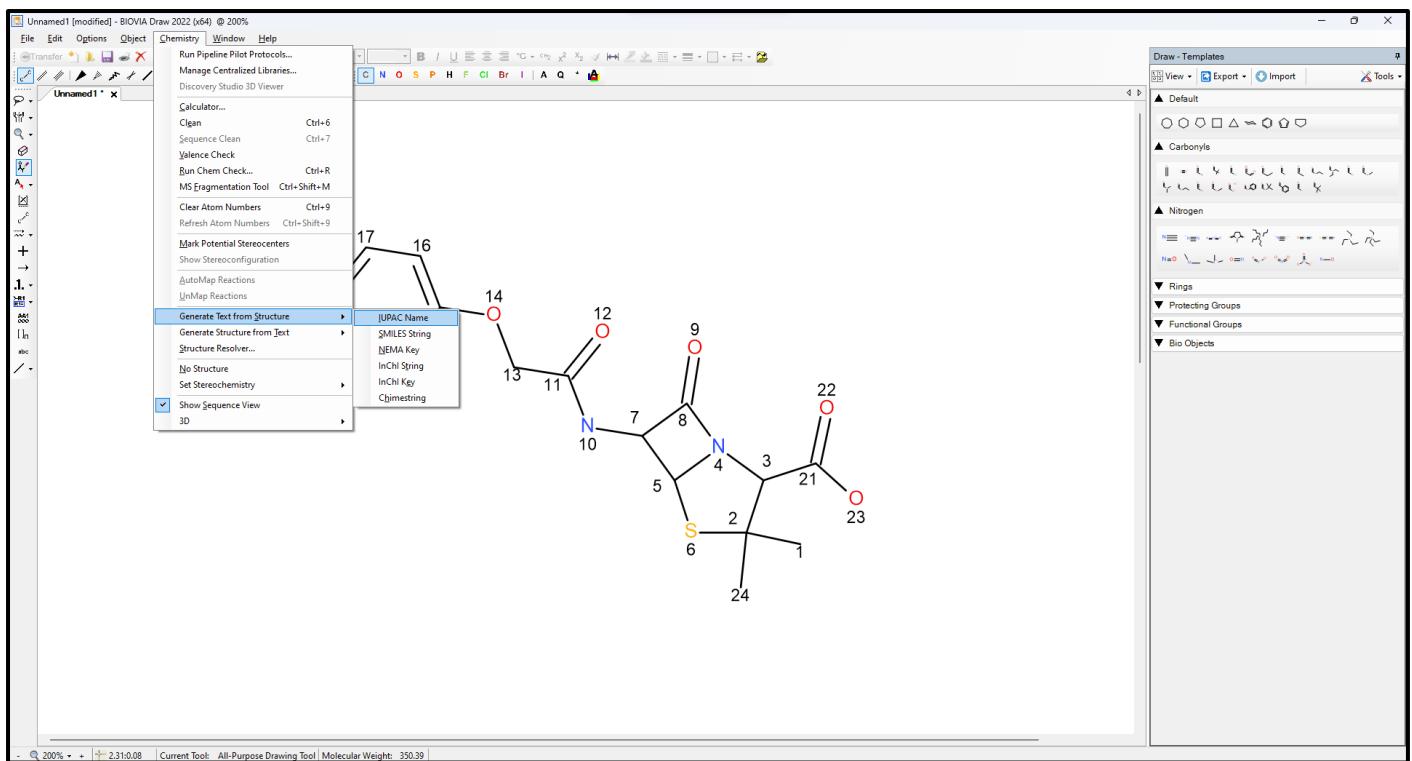


Fig 7. Option applied: To generate IUPAC name

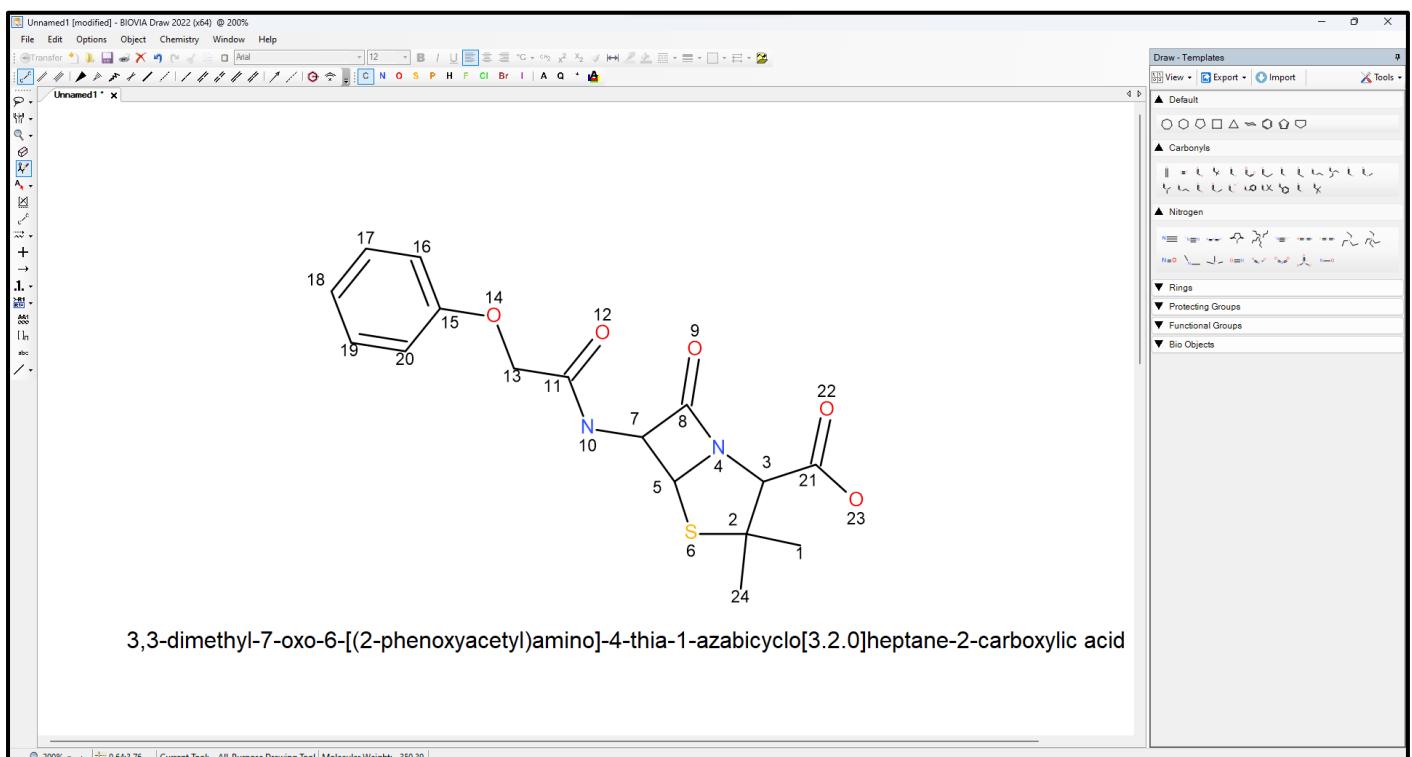


Fig 8. IUPAC name of Penicillin

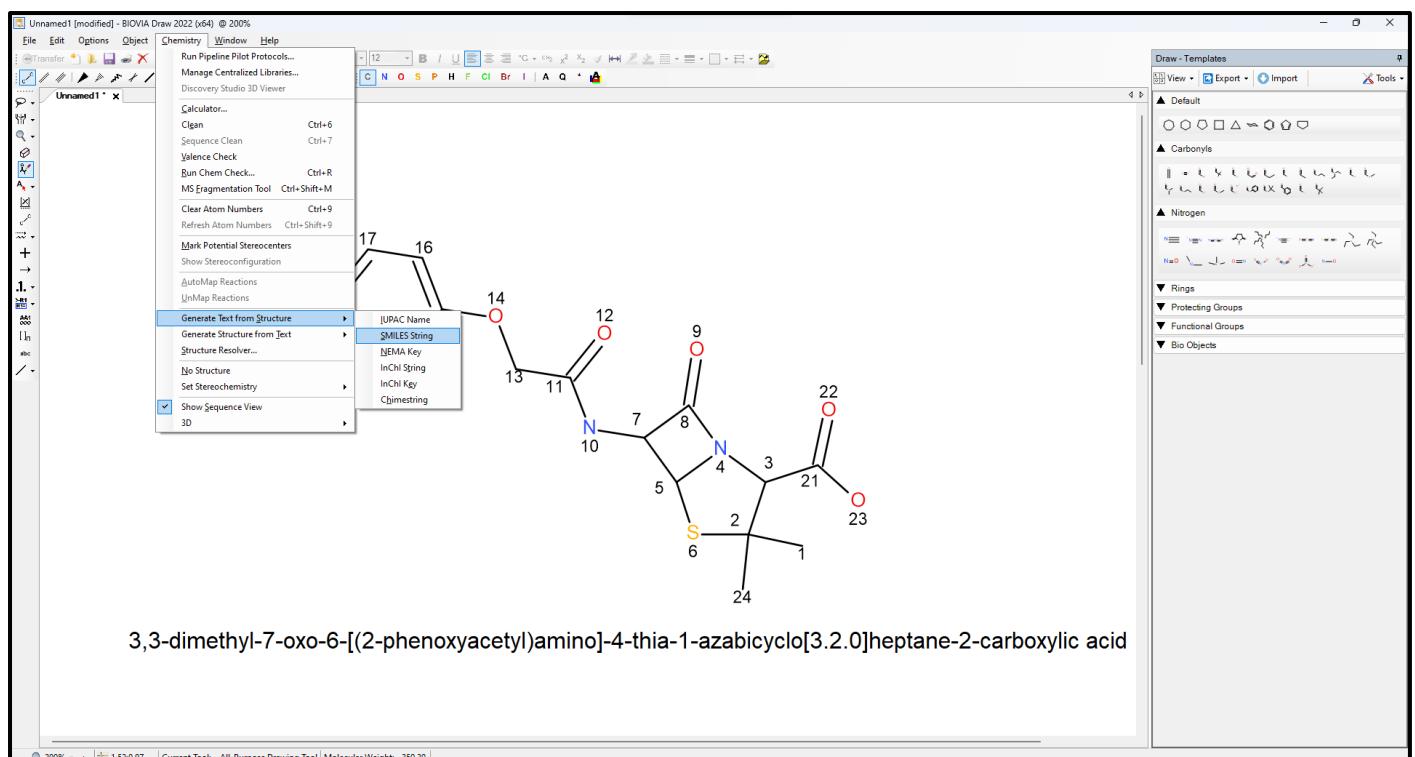


Fig 9. Option applied: To generate SMILES string

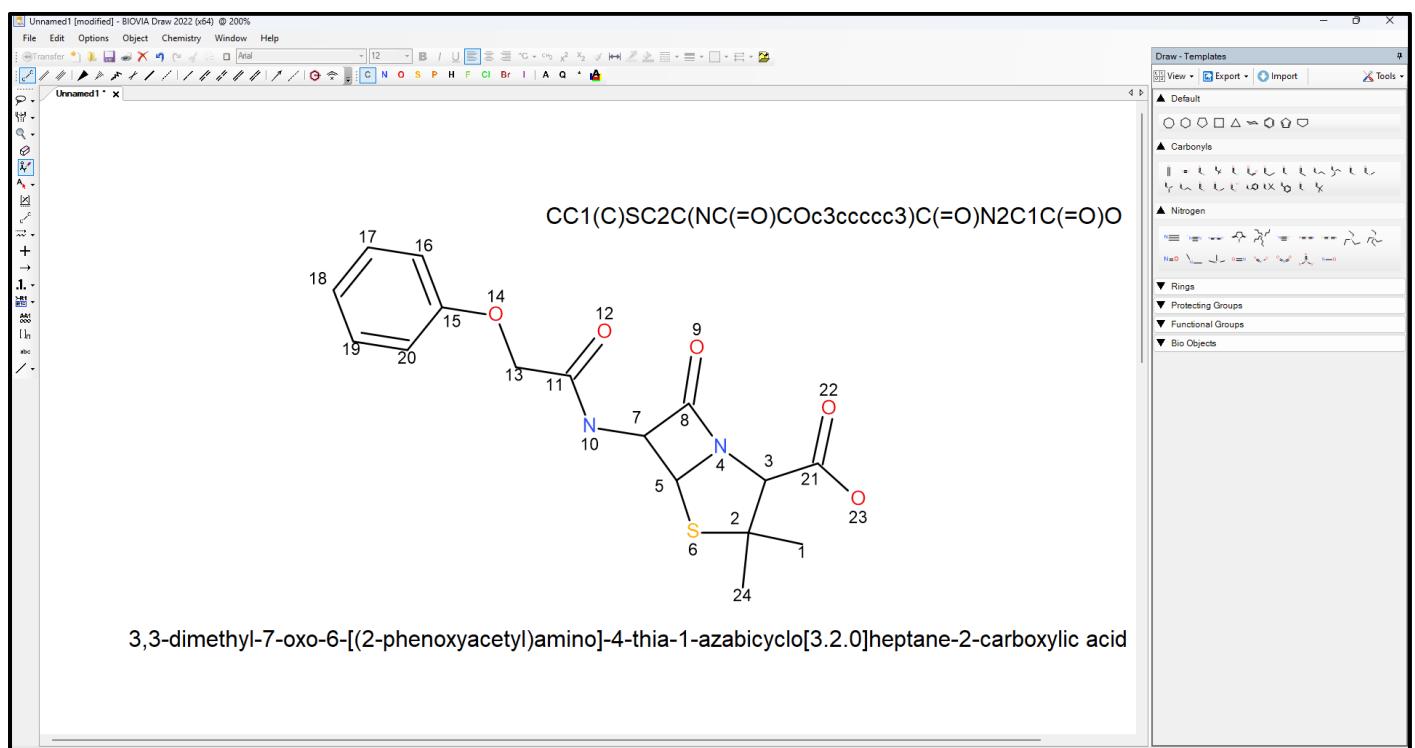


Fig 10. SMILES string of Penicillin

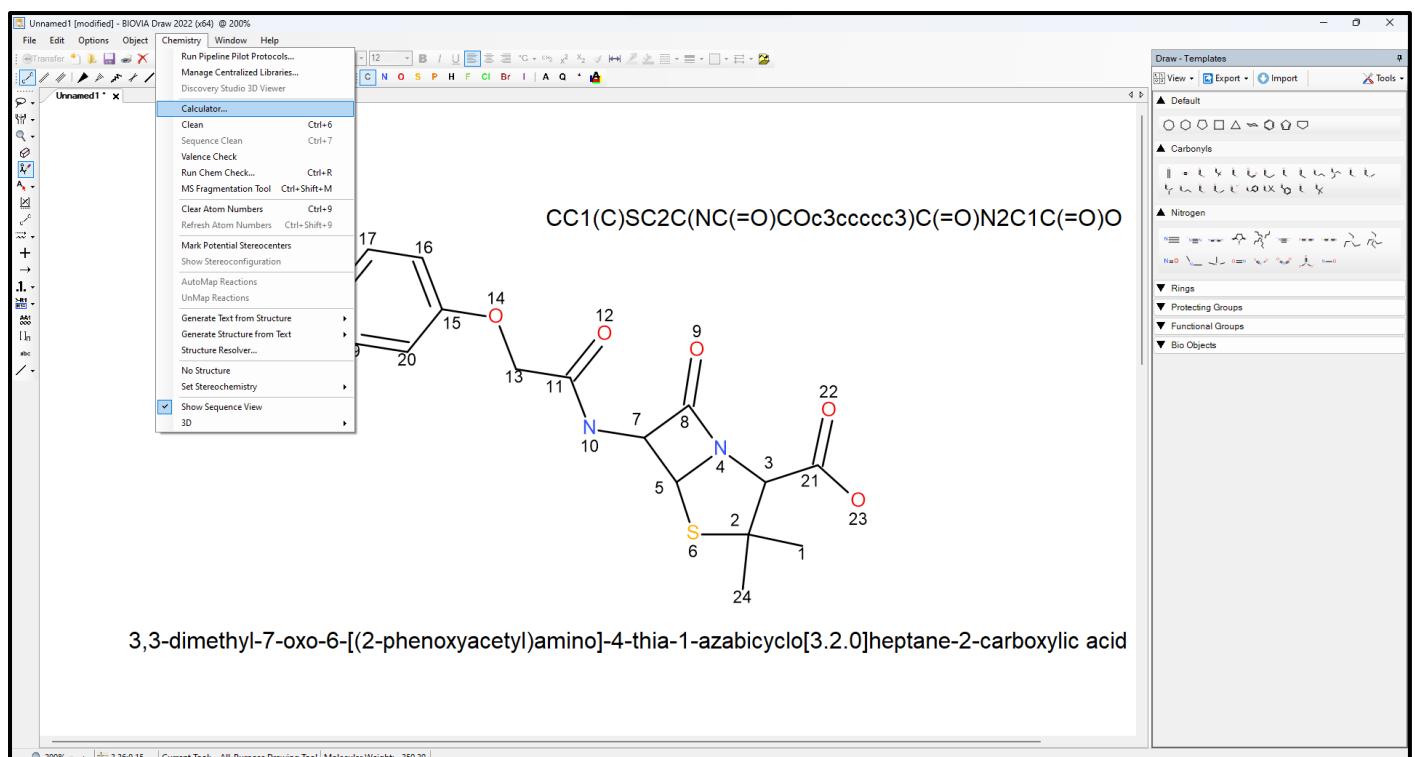


Fig 11. Option applied: To calculate Physicochemical properties of Penicillin

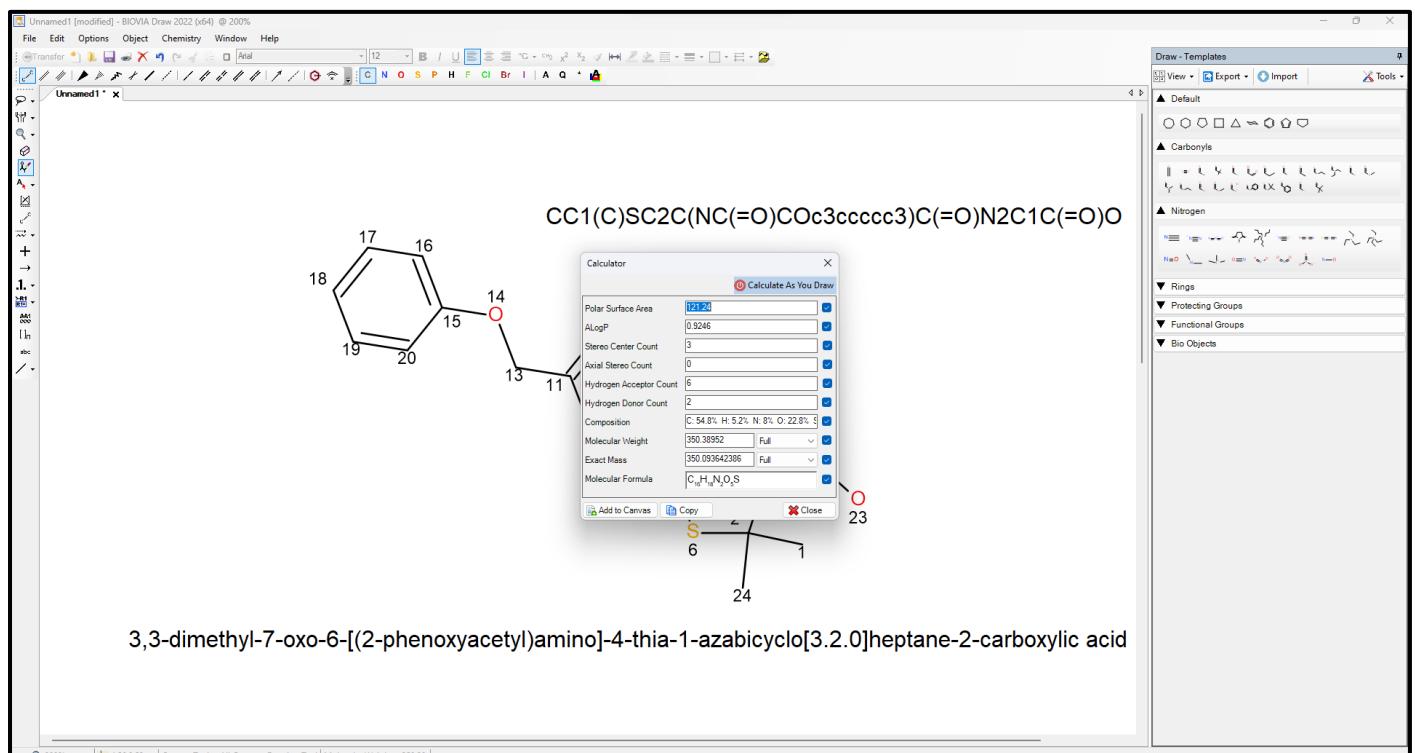


Fig 12. Different Physicochemical properties available for Penicillin

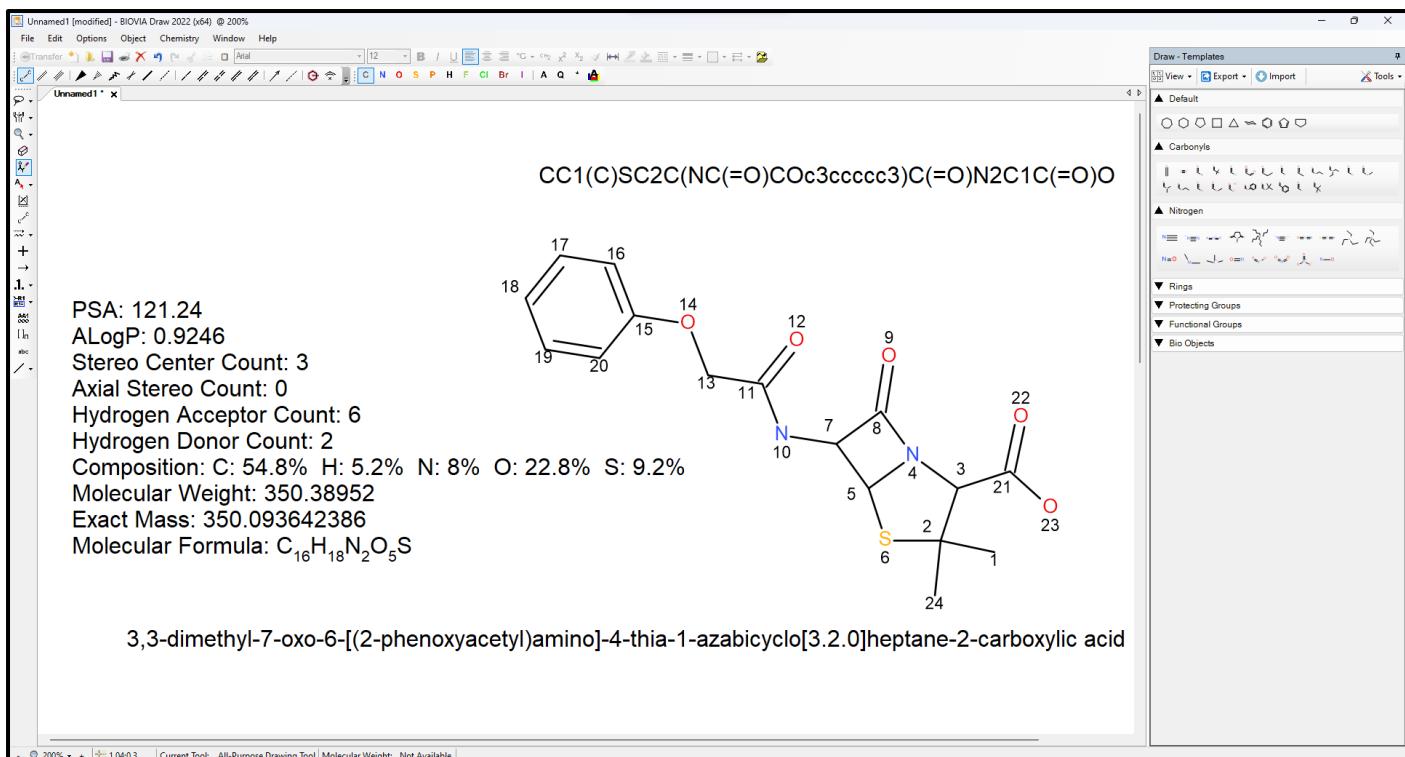


Fig 13. Calculated Physicochemical properties of Penicillin

Results:

The structure of Penicillin was drawn using the BIOVIA DRAW structure. Following are the information retrieved for the structure using software in tabulation format

SR NO	Name of Option	Results
1.	Physicochemical Properties	PSA: 121.24 ALogP: 0.9246 Stereo Center Count: 3 Axial Stereo Count: 0 Hydrogen Acceptor Count: 6 Hydrogen Donor Count: 2 Composition: C: 54.8% H: 5.2% N: 8% O: 22.8% S: 9.2% Molecular Weight: 350.38952 Exact Mass: 350.093642386 Molecular Formula: C ₁₆ H ₁₈ N ₂ O ₅ S
2.	IUPAC name	3,3-dimethyl-7-oxo-6-[(2-phenoxyacetyl)amino]-4-thia-1-azabicyclo[3.2.0]heptane-2-carboxylic acid
3.	SMILES string	CC1(C)SC2C(NC(=O)COc3ccccc3)C(=O)N2C1C(=O)O

Conclusion:

BIOVIA Draw enables the drawing and editing of complex biologics, molecules and chemical reactions. It provides structure and query drawing, allows registering, searching and reporting on chemically modified peptide or nucleotide sequences and to add structure drawing and display to user's applications, and customizing according to their organizational workflows. The software can facilitate the collaborative searching, viewing, communicating and archiving of scientific information as well.

References:

1. Klein, F., Mouquet, H., Dosenovic, P., Scheid, JF., Scharf, L., Nussenzweig, MC. (2013). Antibodies in HIV-1 vaccine development and therapy. *Science*,341(6151), 1199-204. doi: 10.1126/science.1241144
2. Pirard, D., Vereecken, P., Mélot, C., Heenen, M. (2005). Three percent Thalidomide in 2.5% hyaluronan gel in the treatment of actinic keratoses: a meta-analysis of the recent studies. *Arch DermatolRes*, 297(5), 185-9.

WEBLEM 2B

Open Babel Tool

Aim:

To retrieve, convert & store structure for “Penicillin” query and study various file formats using Open Babel Tool

Introduction:

Penicillin V is a member of the penicillin family exhibiting broad-spectrum antibiotic property. Penicillin V binds to penicillin binding proteins (PBP), the enzymes that catalyze the synthesis of peptidoglycan, which is a critical component of the bacterial cell wall. This leads to the interruption of cell wall synthesis, consequently leading to bacterial cell growth inhibition and cell lysis.

Phenoxycephalothin is a narrow spectrum antibiotic also commonly referred to as Penicillin V or Penicillin VK. It is a phenoxycephalothin analog of Penicillin G, or [benzylpenicillin]. An orally active naturally occurring penicillin, phenoxycephalothin is used to treat mild to moderate infections in the respiratory tract, skin, and soft tissues caused by penicillin G-sensitive microorganisms. Phenoxycephalothin has also been used in some cases as prophylaxis against susceptible organisms. While there have been no controlled clinical efficacy studies that were conducted, phenoxycephalothin has been suggested by the American Heart Association and the American Dental Association for use as an oral regimen for prophylaxis against bacterial endocarditis in patients with congenital heart disease or rheumatic or other acquired valvular heart disease when they undergo dental procedures and surgical procedures of the upper respiratory tract, except for those who are at an elevated risk for endocarditis.

Open Babel is a free, open-source version of the Babel chemistry file translation program. Open Babel is a project designed to pick up where Babel left off, as a cross-platform program and library designed to interconvert between many file formats used in molecular modeling, computational chemistry, and many related areas. Open Babel includes two components, a command-line utility and a C++ library. The command-line utility is intended to be used as a replacement for the original babel program, to translate between various chemical file formats. The C++ library includes all of the file-translation code as well as a wide variety of utilities to foster development of other open source scientific software.

Methodology:

1. Open homepage for Pubchem database
2. Search for compound “Penicillin”
3. Download the compound in various file formats such as SMILES, sdf and mol file formats
4. Open the OPEN BABEL Tool and upload the structure file derived from pubchem
5. Convert the file format into the various structural file formats
6. Observe the penicillin in different file formats and interpret the results

Observations:

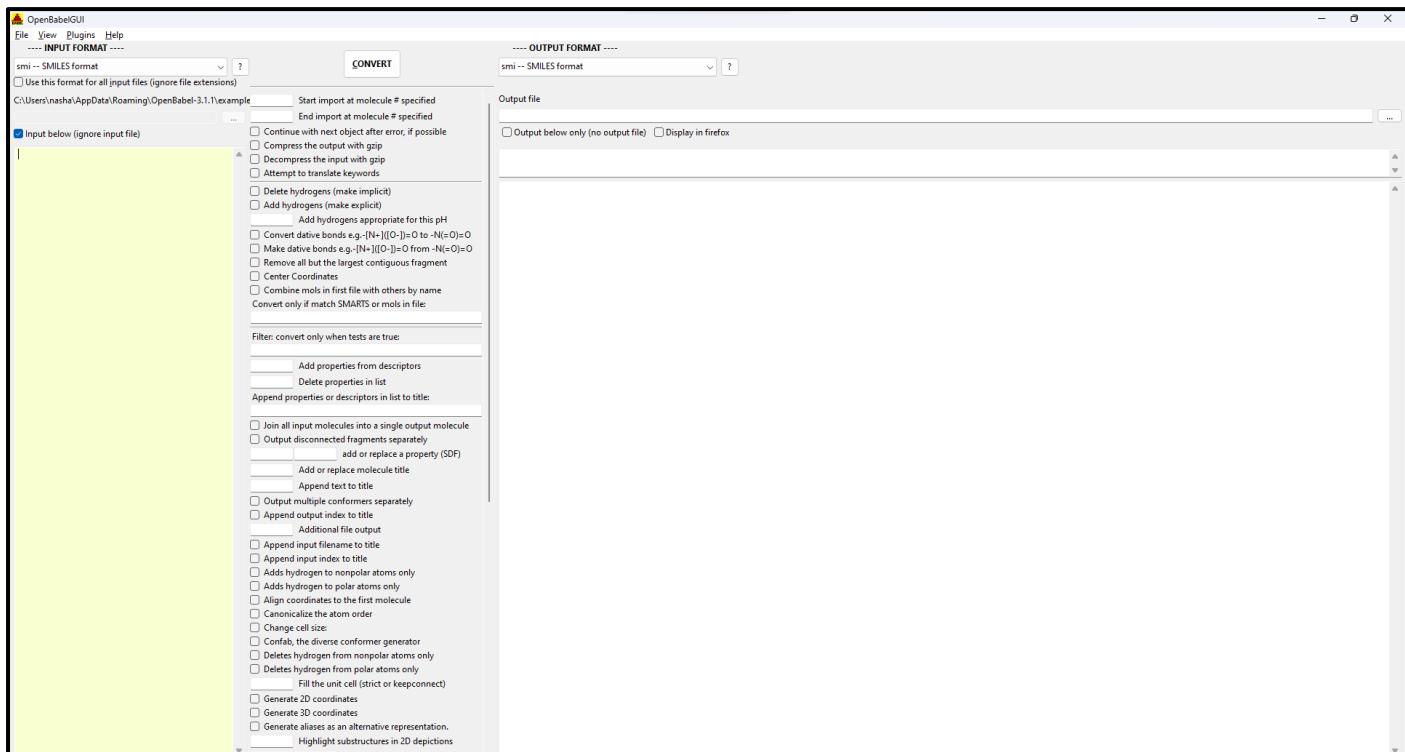


Fig 1. Homepage of OpenBabel tool

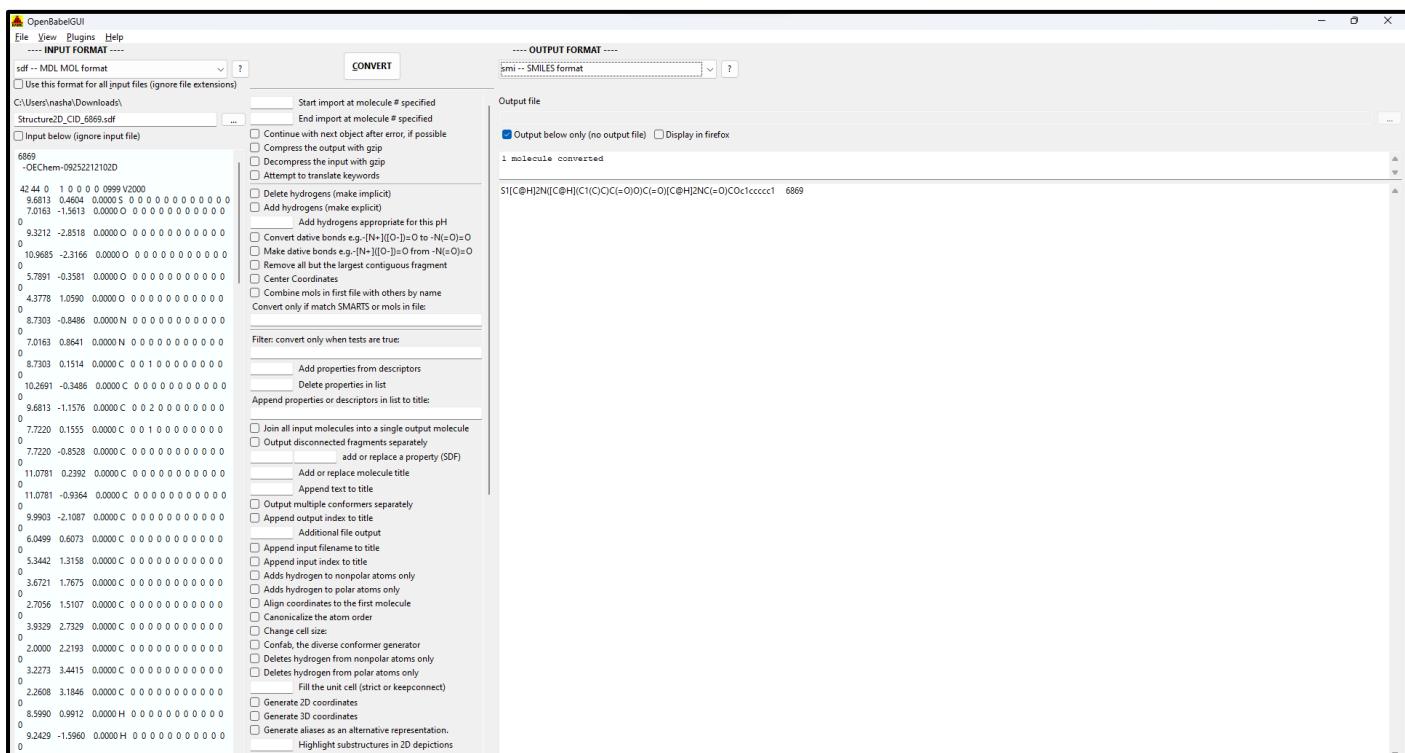


Fig 2. Convert sdf file into smiles for Penicillin

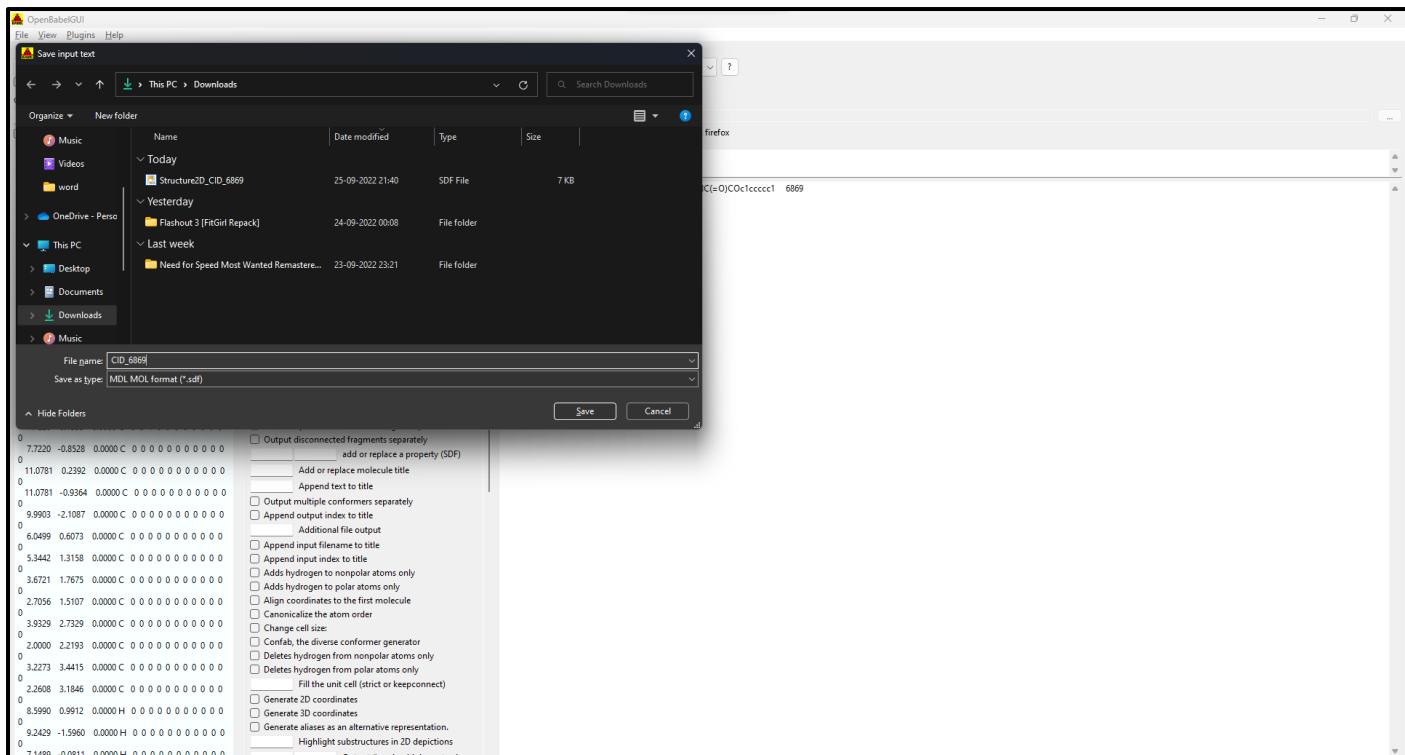


Fig 3. Storage Option for SMILES under Open Babel

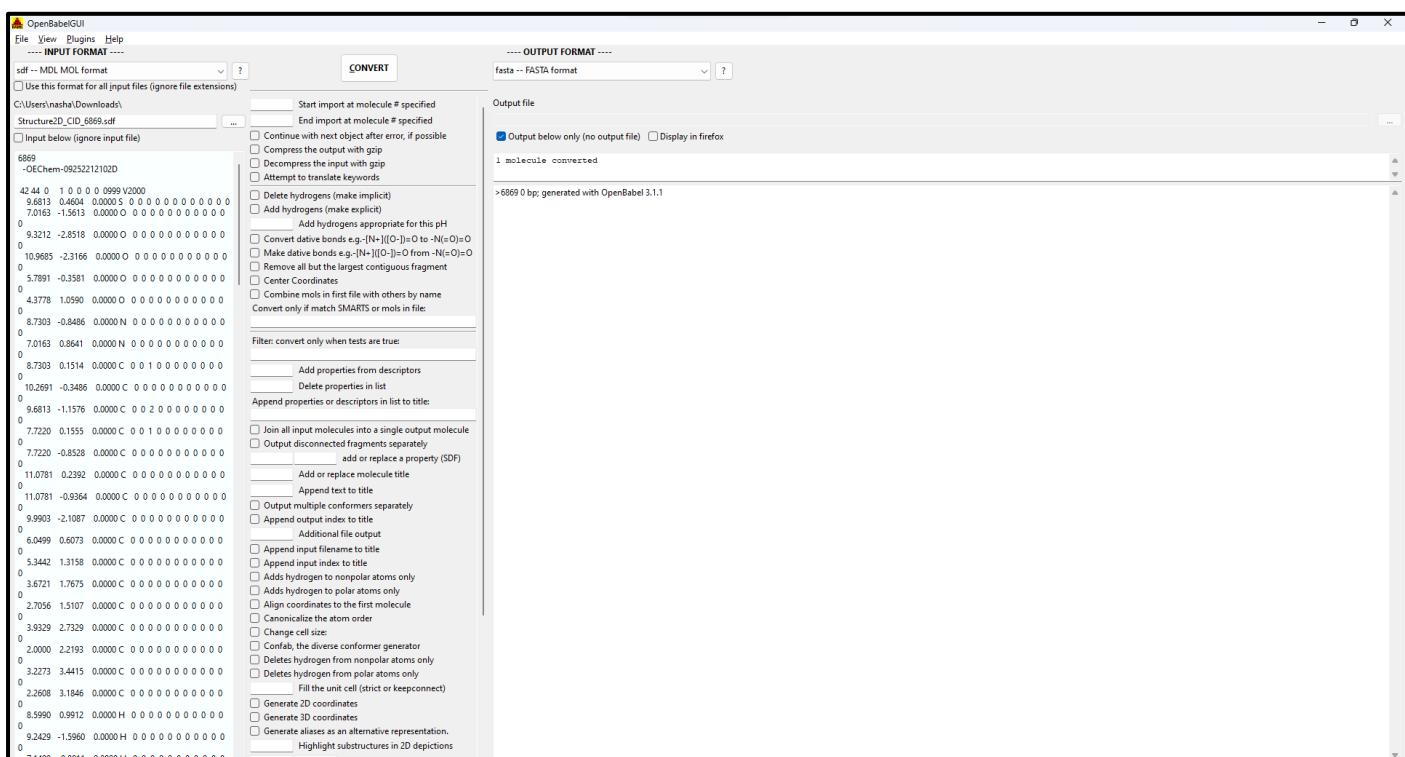


Fig 4. Convert sdf file into fasta file for Penicillin

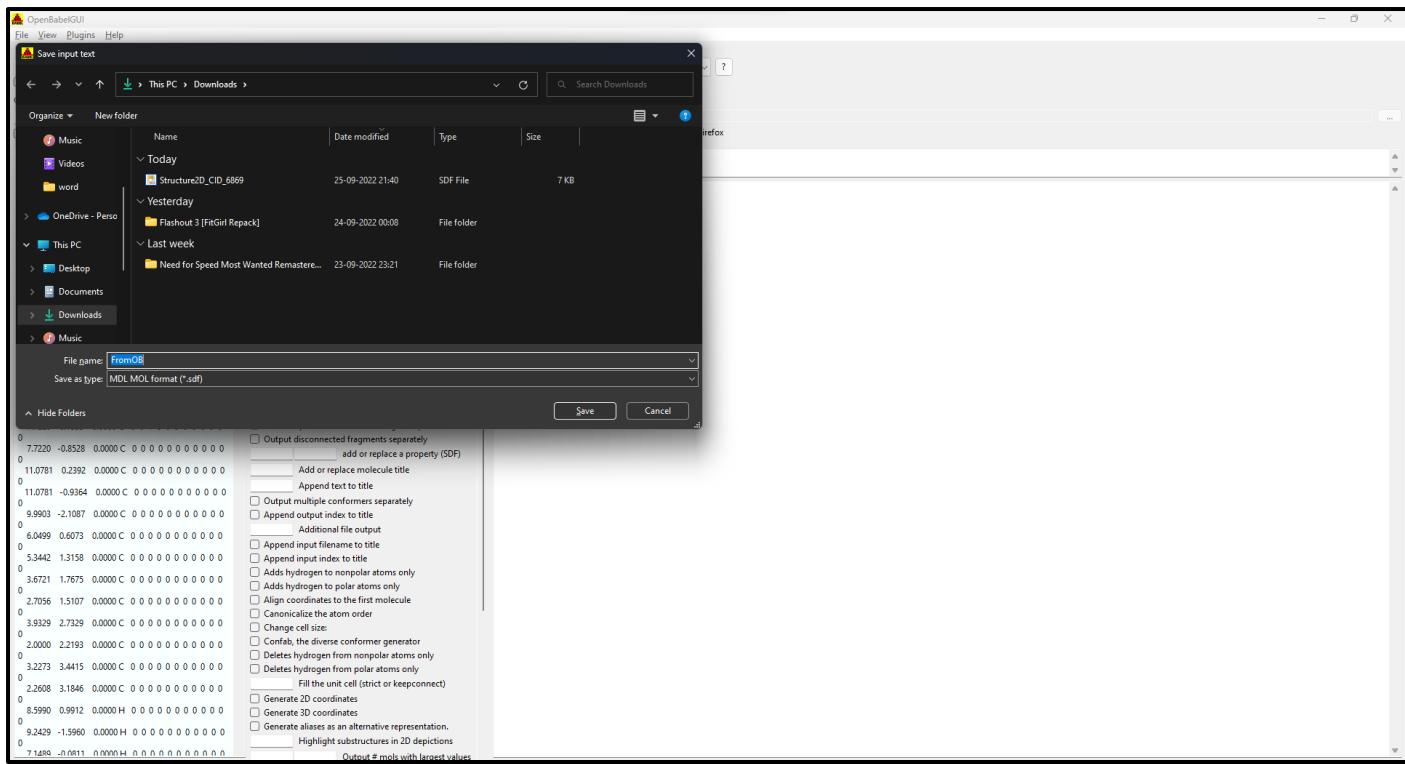


Fig 5. Storage option for fasta under Open Babel

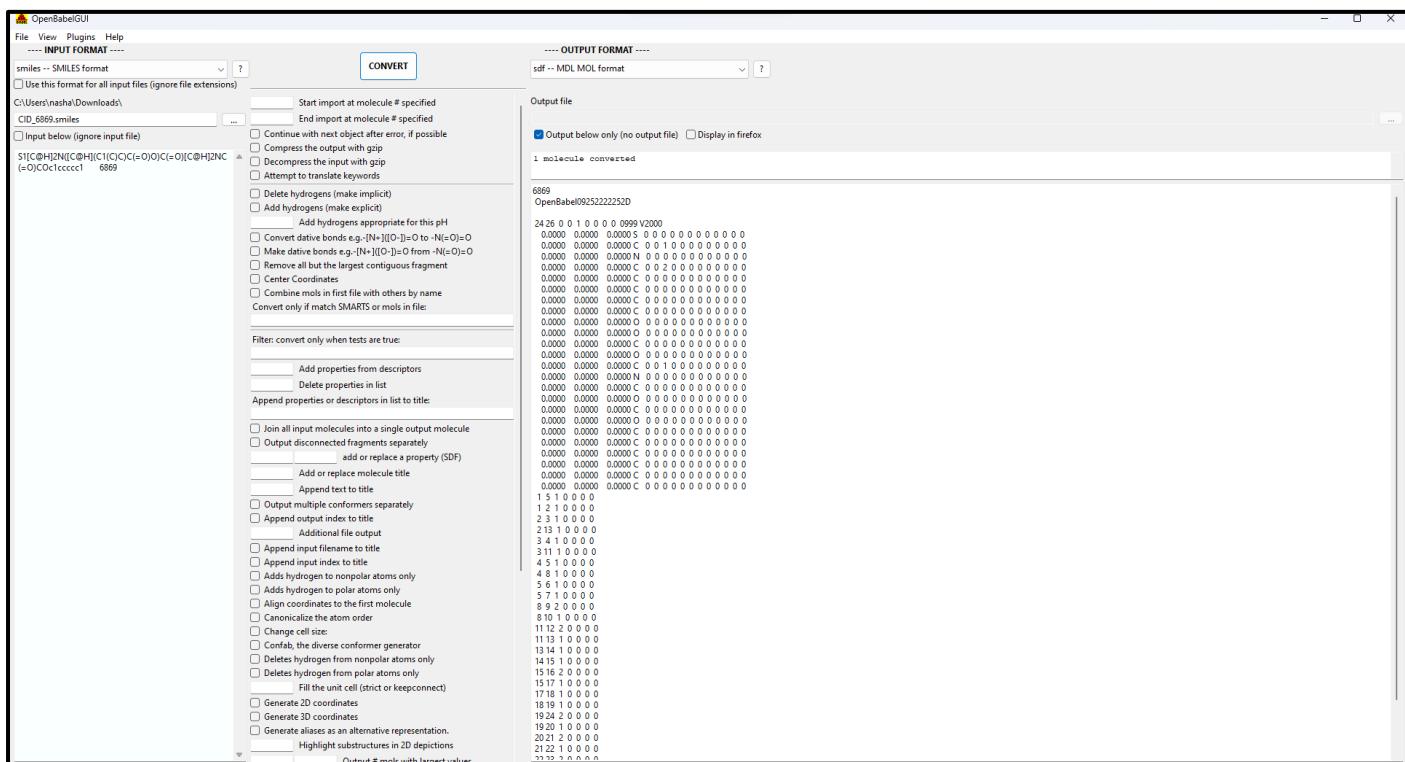


Fig 6. Convert smiles file into SDF file for Penicillin

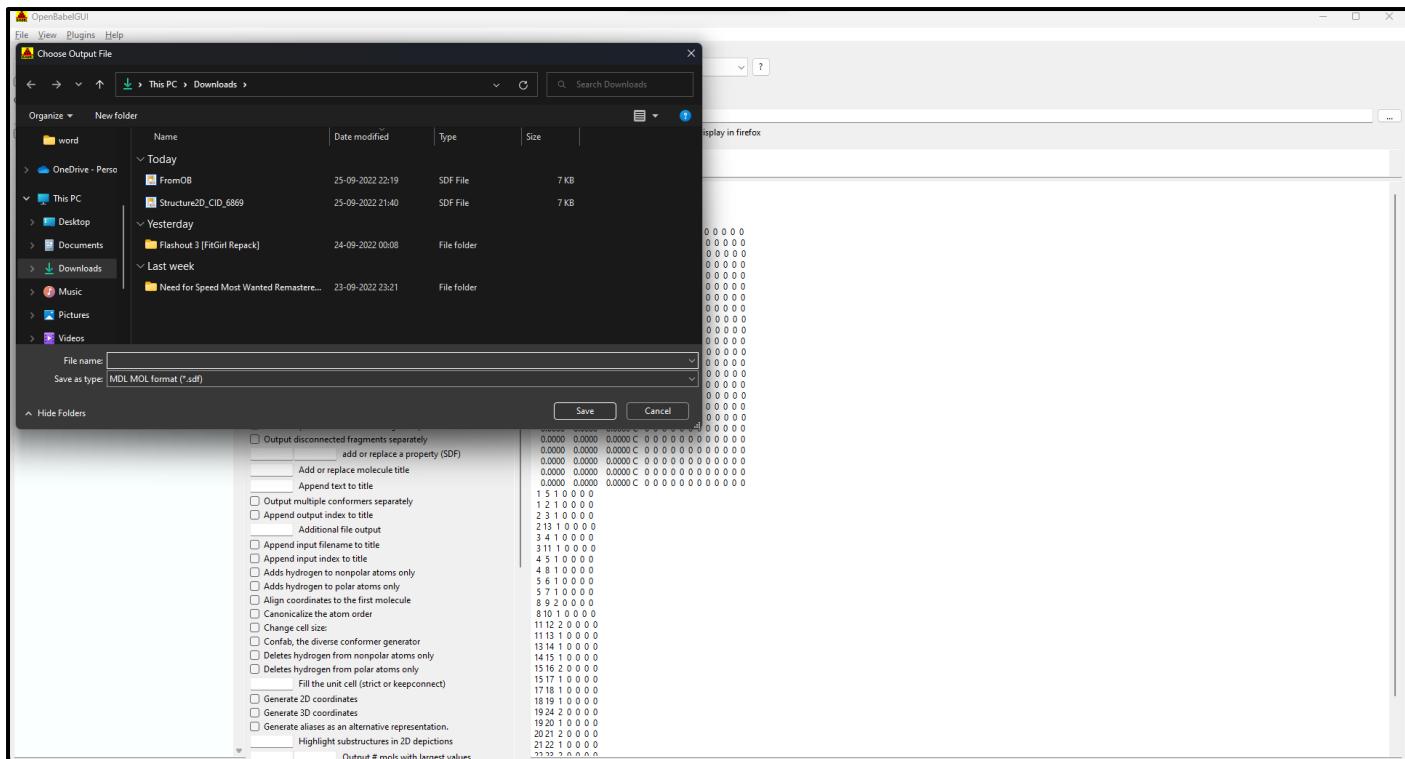


Fig 7. Storage options for sdf under Open Babel

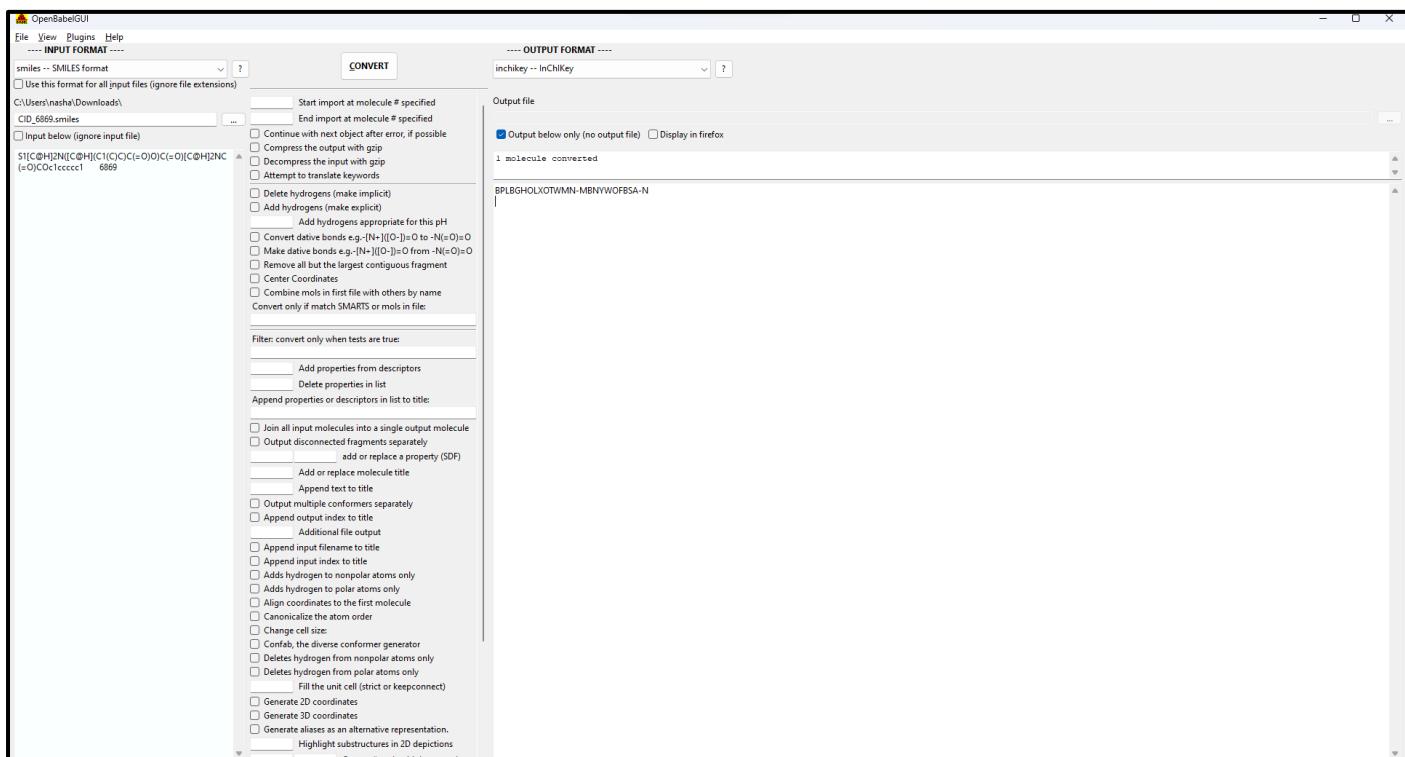


Fig 8. Convert smiles file into InChIKey file for penicillin

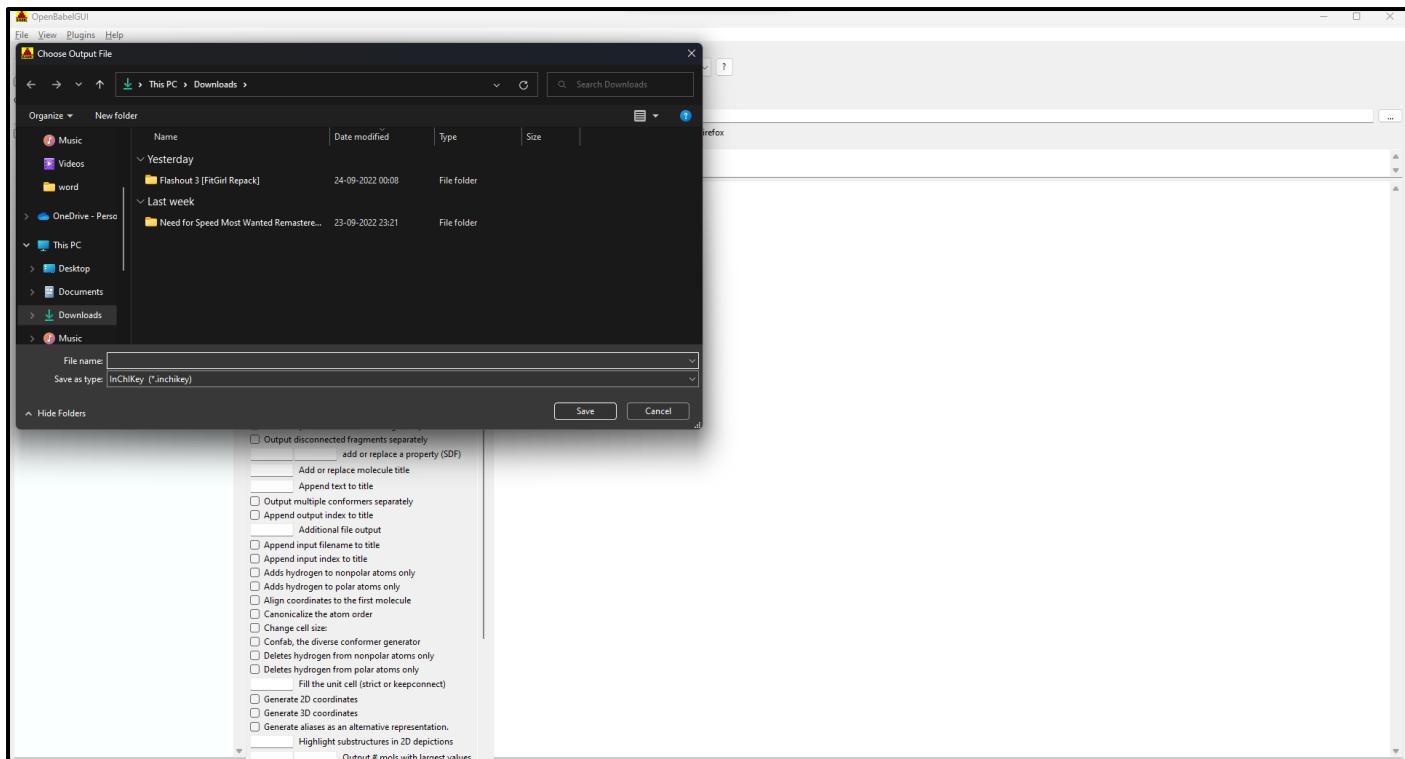


Fig 9. Storage options for InChIKey under open babel

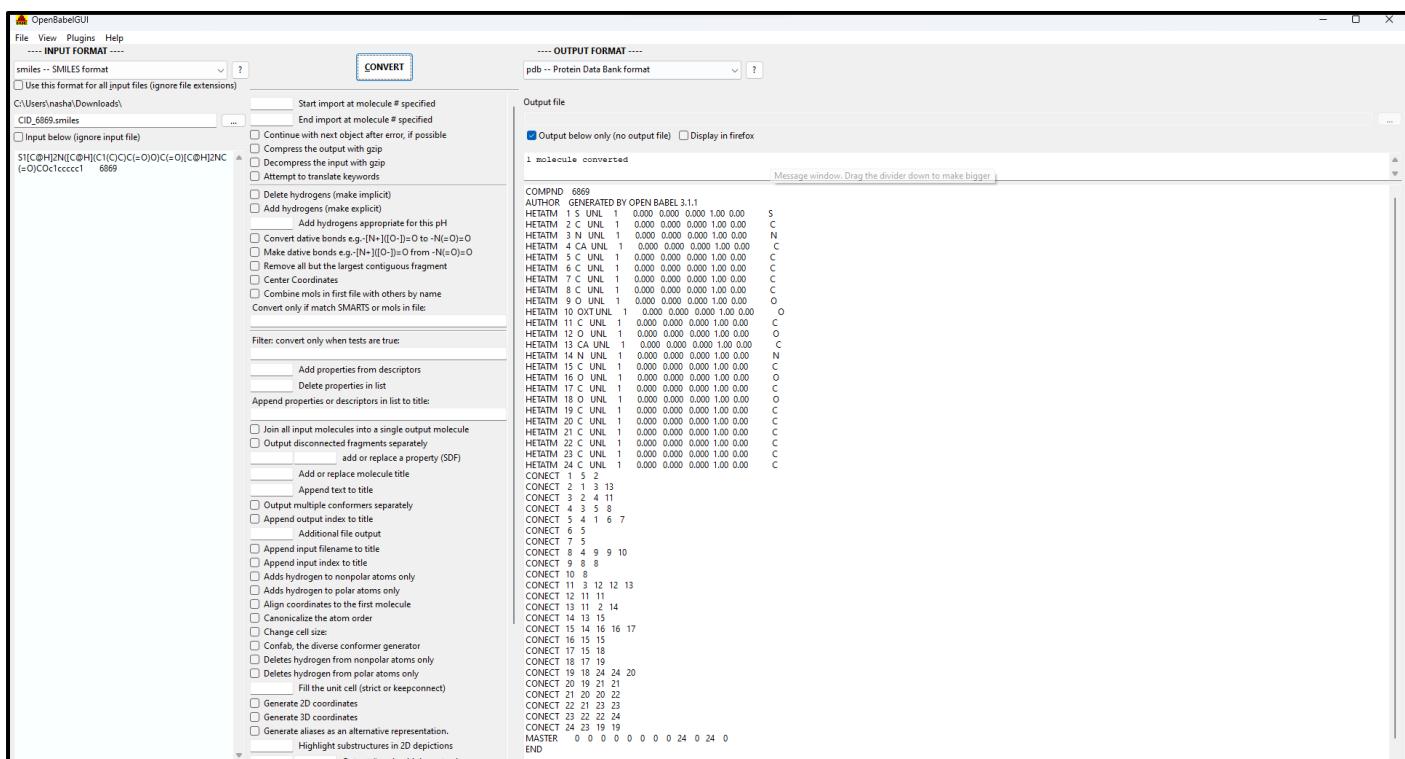


Fig 10. Convert smiles file into pdb file for penicillin

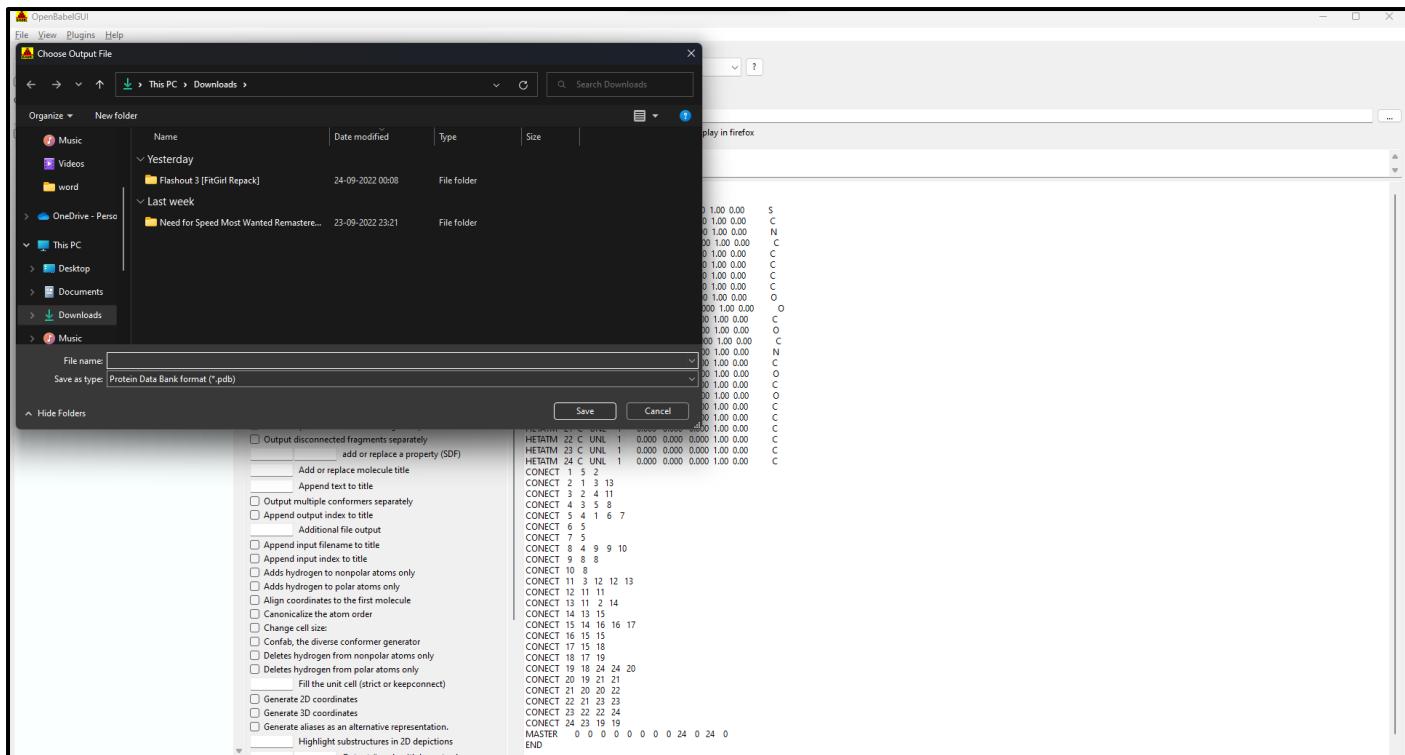


Fig 11. Storage Option for pdb under Open babel

Results:

The structure for Penicillin was retrieved from PUBCHEM in sdf file format. By using Open Babel tool, it was possible to convert various structural file formats into various other formats such as SMILES, MDL MOL, fasta, pdb, InChI key and the output is saved in the desired format for further studies

Conclusion:

The data of the chemical compound “Penicillin” was studied using different file formats. Therefore, the Open Babel tool presents a solution to the increase of multiple chemical file formats. It gains by way of its users, contributors, developers, related projects, and the general chemical community. It allows searching, converting, analyzing, or storing data from molecular modeling, chemistry, solid-state materials, biochemistry, or related areas. It provides both ready-to-use programs as well as a complete, extensible programmer's toolkit for developing cheminformatics software.

References:

1. Tawfeek, N., Mahmoud, MF., Hamdan, DI., Sobeh, M., Farrag, N., Wink, M., El-Shazly, AM. (2021).
2. Phytochemistry, Pharmacology and Medicinal Uses of Plants of the Genus Salix: An Updated Review. Front Pharmacol, 12(593856).
3. Weininger, D. (1988). SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. J Chem Inf Comput Sci. 28, 31-36.

WEBLEM 3:**Introduction to Conformational search studies using BALLOON software**

(<http://users.abo.fi/mivainio/balloon/index.php>)

AIM:

Introduction to conformational search and Balloon software.

INTRODUCTION TO CONFORMATIONAL SEARCH:

3-D ligand conformations are required for most ligand-based drug-design methods, such as pharmacophore modeling, shape-based screening, and 3-D QSAR model building. Many studies of conformational search methods have focused on the reproduction of crystal structures; however, for ligand-based modeling, the key question is how to generate a ligand alignment that produces the best results for a given query molecule. In general, we find that virtual screening results are relatively insensitive to the conformational search protocol; hence, a conformational search method that generates fewer conformations could be considered “better” because it is more computationally efficient for screening.

A number of methods have been described to generate ligand conformations, such as:

1. Random torsional angle changes
2. Random coordinate changes
3. Distance geometry
4. Rule-based methods
5. Knowledge-based methods
6. Low mode search

Each mode has its strength and weaknesses and the performance depends on multiple factors. All the methods are “unbiased”, in that they do no explicitly consider information about the query molecule being used in the virtual screen or QSAR alignments. While generating more conformations will necessarily lead to an increased probability of finding a bioactive conformation, there are potential of finding a bioactive conformation, there are potential drawbacks to generating conformations.

NEED OF CONFORMATIONAL SEARCH:

- Conformational analysis is an important step in molecular modeling as it is necessary to reduce time spent in screening of compounds for activity.
- Most drugs are flexible molecules with the ability to adopt different conformations by means of rotation about single bonds.
- Conformations play an important role in prediction of not just physico-chemical properties but also the biological activity of the drug.
- The major objective of conformational analysis is to gain insight on conformational characteristics of drugs and also to identify the relation between the role of conformational flexibility and their activity.
- Therefore, it plays a significant role in computer aided design as well.
- The significance of conformational analysis not just extends to computational docking and screening but also for lead optimization.
- The analysis of the conformational collection that was sampled and optimized is essential so as to ascertain the conformational properties of the molecule that is being studied.
- This helps to underline the global properties and to exemplify features of overall flexibility and to recognize common inclination in the conformation set.
- Alternatively, it may be used to identify a smaller subset of characteristic low energy conformations, which may be used to direct future drug development efforts.

INTRODUCTION TO BALLOON:

Balloon creates 3D atomic coordinates from molecular connectivity via distance geometry and conformer ensembles using a multi-objective genetic algorithm. The input can be SMILES, SDF or MOL2 format. Output is SDF or MOL2. Flexibility of aliphatic rings and stereochemistry about double bonds and tetrahedral chiral atoms is handled.

The software has been ported to Linux, Mac OS X, and Microsoft Windows platforms.

Balloon was introduced to the scientific community during a poster session of the The 16th European Symposium on Quantitative Structure-Activity Relationships & Molecular Modelling 10 - 17 September 2006 held in MSC Opera on the Mediterranean Sea. There were a couple of other studies about conformational analysis presented as posters as well, which indicates that despite the problem has been studied for decades people still do not consider the job done.

PROCEDURE (INSTALLATION OF BALLOON):

1. Navigate to the official Balloon website (<http://users.abo.fi/mivainio/balloon/>)
2. Go to “Downloads” section
3. Scroll down to the bottom and selected your platform for which you want to download the software.
4. A zip file will be downloaded.
5. Create a folder or directly extract the contents of the zip file into this folder or at the desired location.
6. The installation is completed.

Balloon

about news features download change log faq

Latest version 1.8.2 (March 13 2022)

About

Balloon creates 3D atomic coordinates from molecular connectivity via distance geometry and conformer ensembles using a multi-objective genetic algorithm. The input can be SMILES, SDF or MOL2 format. Output is SDF or MOL2. Flexibility of aliphatic rings and stereochemistry about double bonds and tetrahedral chiral atoms is handled.

The software has been ported to Linux, Mac OS X, and Microsoft Windows platforms.

Contact information

Mikko Vainio
email: mikko.vainio@abo.fi

Bibliographic references

Mikko J. Vainio and Mark S. Johnson (2007) *Generating Conformer Ensembles Using a Multiobjective Genetic Algorithm*. *Journal of Chemical Information and Modeling*, **47**, 2462 - 2474.
The structures used for the test runs are available for download.

Figure 1: Homepage of Balloon Software

Download

BALLOON is distributed "as is", free of charge, and without warranty of any kind. The use of the program is not restricted, but we appreciate if you acknowledge use of BALLOON in any reports or publications of results obtained with BALLOON.

BY ACCESSING THE PROGRAM, YOU ACKNOWLEDGE THAT YOU HAVE READ THE TERMS OF THE END USER LICENSE AGREEMENT ABOVE AND AGREE TO BE BOUND BY ITS TERMS.

Current version is 1.8.2 (March 13 2022)

Please select your platform:

Linux 2.6.18, 64-bit x86_64 ▾

[Download](#)

System Requirements

BALLOON is a command-line program and will most likely run on any platform that can run any of the operating systems for which a binary executable is provided.

Related material

Other downloads related to BALLOON:

Reference structures used in the publication ▾

[Download](#)

Figure 2: Bottom of Downloads page where the platform selection is there

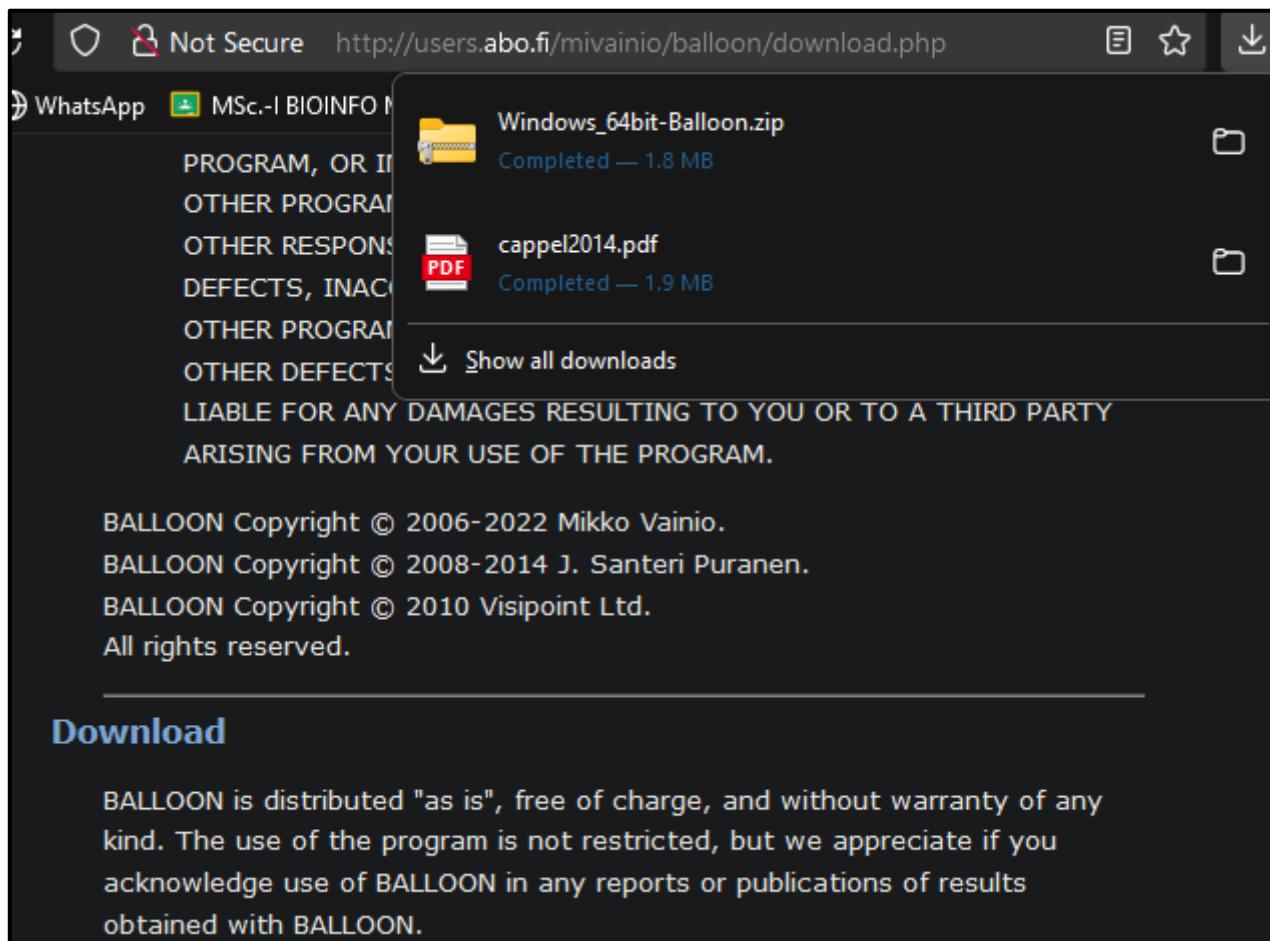


Figure 3: The downloaded ZIP file

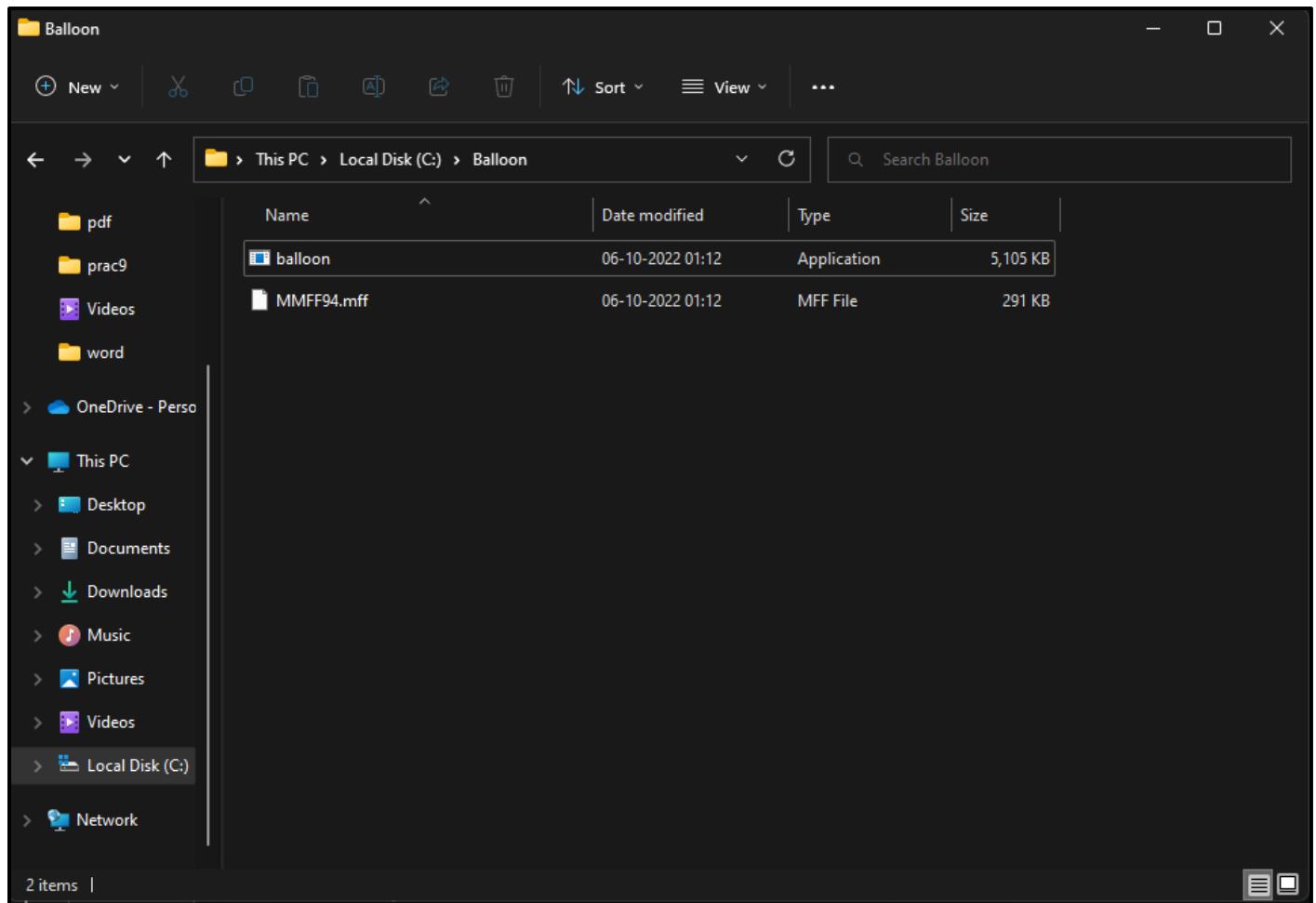


Figure 4: Folder with the contents of the ZIP file extracted

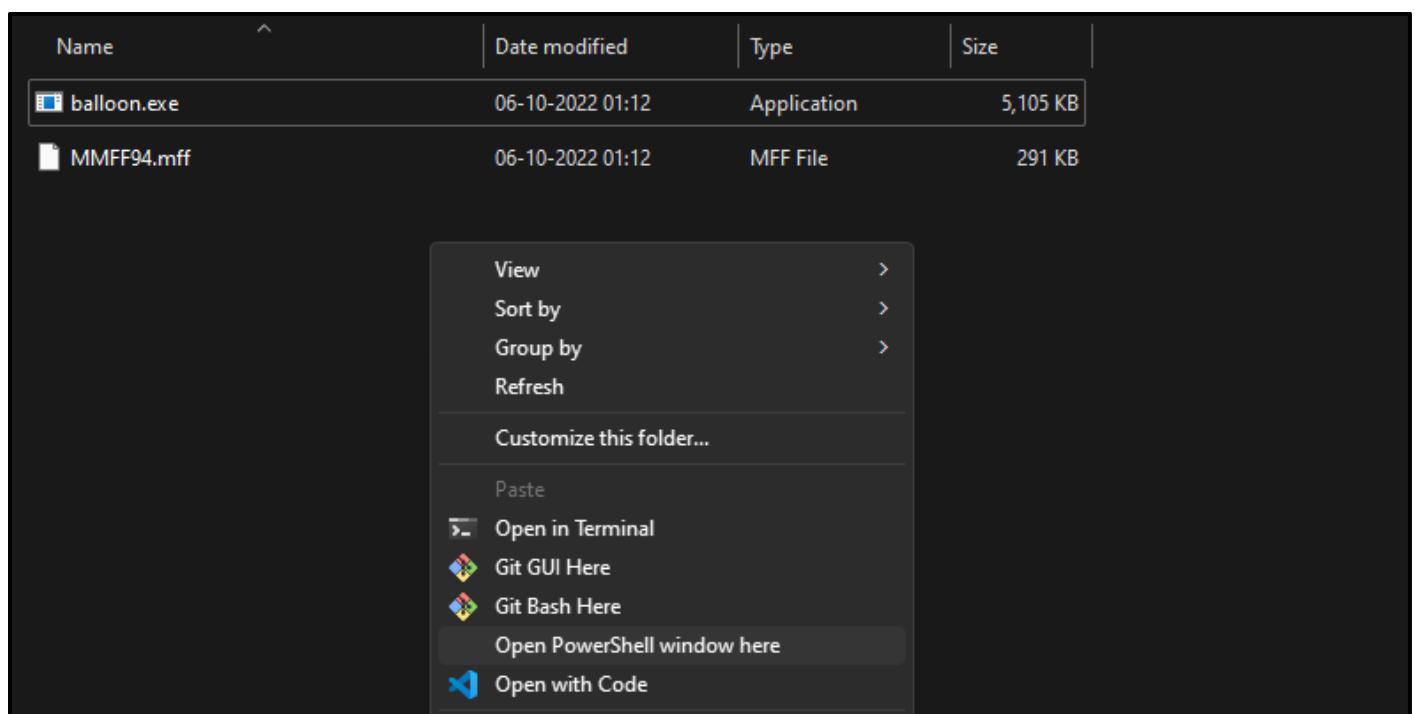


Figure 5: Open PowerSHell Option after Pressing Shift+Right click for BALLOON Software

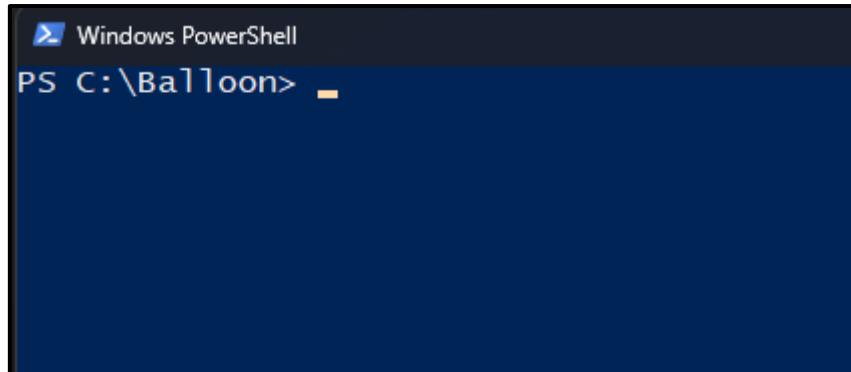


Figure 6: Windows PowerShell for BALLOON Software

REFERENCES:

1. Cappel, D., Dixon, S. L., Sherman, W., & Duan, J. (2014, November 19). Exploring conformational search protocols for ligand-based virtual screening and 3-D QSAR modeling. *Journal of Computer-Aided Molecular Design*, 29(2), 165–182. <https://doi.org/10.1007/s10822-014-9813-4>
 2. 8.2: Conformational Analysis. (2020, May 12). Chemistry LibreTexts. [https://chem.libretexts.org/Bookshelves/Organic_Chemistry/Organic_Chemistry_I_\(Cortes\)/08%3A_Conformational_Analysis](https://chem.libretexts.org/Bookshelves/Organic_Chemistry/Organic_Chemistry_I_(Cortes)/08%3A_Conformational_Analysis)
-

Weblem 3a:**Introduction to Conformational search studies using BALLOON Software**

(<http://users.abo.fi/mivainio/balloon/index.php>)

AIM:

To generate and analyze various structural conformation for Penicillin (Pubchem ID-5904) molecule using Balloon software.

INTRODUCTION:

Balloon creates 3D atomic coordinates from molecular connectivity via distance geometry and conformer ensembles using a multi-objective genetic algorithm. The input can be SMILES, SDF or MOL2 format. Output is SDF or MOL2. Flexibility of aliphatic rings and stereochemistry about double bonds and tetrahedral chiral atoms is handled.

Penicillin G is a broad-spectrum, beta-lactam naturally occurring penicillin antibiotic with antibacterial activity. Penicillin G binds to and inactivates the penicillin binding proteins (PBPs) located inside the bacterial cell wall. Inactivation of PBPs interferes with the cross-linkage of peptidoglycan chains necessary for bacterial cell wall strength and rigidity. This interrupts bacterial cell wall synthesis and results in the weakening of the bacterial cell wall and eventually causing cell lysis.

METHODOLOGY:

1. Download the sdf file of the query Penicillin G from PubChem Database.
<https://pubchem.ncbi.nlm.nih.gov/>
2. Paste the structure in the folder where BALLOON Software installed.
3. Open BALLOON Software in PowerShell by pressing Shift+Right key.
4. Type the command “.\balloon.exe -f .\MMFF94.mff --nconfs 20 --nGenerations 5 --input- file .\Structure2D_CID_6167.sdf --output-file .\Structure2D_CID_6167_OPT.sdf”
5. Press the “Enter” button.
6. Command will run within a few seconds.
7. Observe the results and from output file extract the energy data for each conformations.

NOTE: DESCRIPTION OF COMMAND:

- a. ./balloon.exe (specifies which exe to execute in our case it's Balloon)
- b. -f MMFF94.mff (symbolizes and specifies which forcefield file should be used)
- c. --nconfs 20 (specifies the number of conformations)
- d. --nGenerations 5 (specifies the number of generations)
- e. --input-file Structure2D_CID_5904.sdf (specifies the input file)
- f. --output-file Structure2D_CID_5904_out.sdf (specifies the output file)

OBSERVATIONS:

National Library of Medicine
National Center for Biotechnology Information

PubChem About Posts Submit Contact Search PubChem

COMPOUND SUMMARY

Penicillin g

PubChem CID 5904

Structure 2D 3D Find Similar Structures

Chemical Safety Irritant Health Hazard Laboratory Chemical Safety Summary (LCSS) Datasheet

Molecular Formula C₁₆H₁₉N₂O₄S

Synonyms penicillin g Benzylpenicillin 61-33-6 Benzylpenicillanic acid Free penicillin II More...

Molecular Weight 334.4

CONTENTS

- Title and Summary
- 1 Structures
- 2 Names and Identifiers
- 3 Chemical and Physical Properties
- 4 Spectral Information
- 5 Related Records
- 6 Chemical Vendors
- 7 Drug and Medication Information
- 8 Food Additives and Ingredients
- 9 Pharmacology and Biochemistry
- 10 Use and Manufacturing
- 11 Identification
- 12 Safety and Hazards
- 13 Toxicity
- 14 Associated Disorders and Diseases
- 15 Literature
- 16 Patents
- 17 Biomolecular Interactions and Pathways
- 18 Biological Test Results

Figure 1: PubChem page for my query Penicillin

PubChem Penicillin g (Compound)

Corynebacterium diphtheriae, and *Erysipelothrix rhusiopathiae*. Natural penicillins have limited activity against gram negative organisms; however, they may be used in some cases to treat infections caused by *Neisseria meningitidis* and *Pasteurella*. They are not generally used to treat anaerobic infections. Resistance patterns, susceptibility and treatment guidelines vary across regions.

DrugBank FDA Pharm Classes

1 Structures

1.1 2D Structure

Chemical Structure Depiction

Find Similar Structure Download SDF Save Display JSON Save Display XML Save Display ASNT Save Display

1.2 3D Conformer

Figure 2: Download the 2-D structure of the query in .sdf format

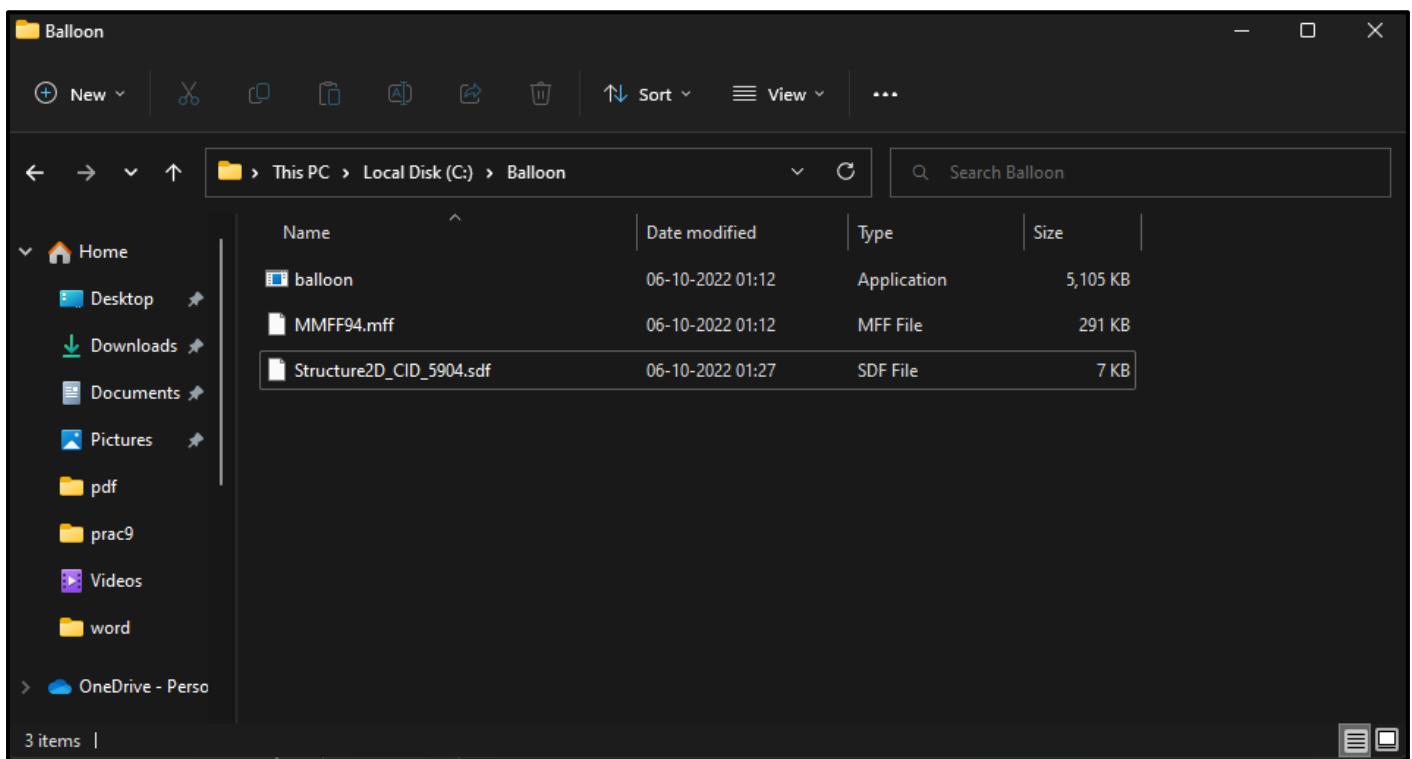


Figure 3: Place the downloaded structure in the same folder as the Balloon Installation

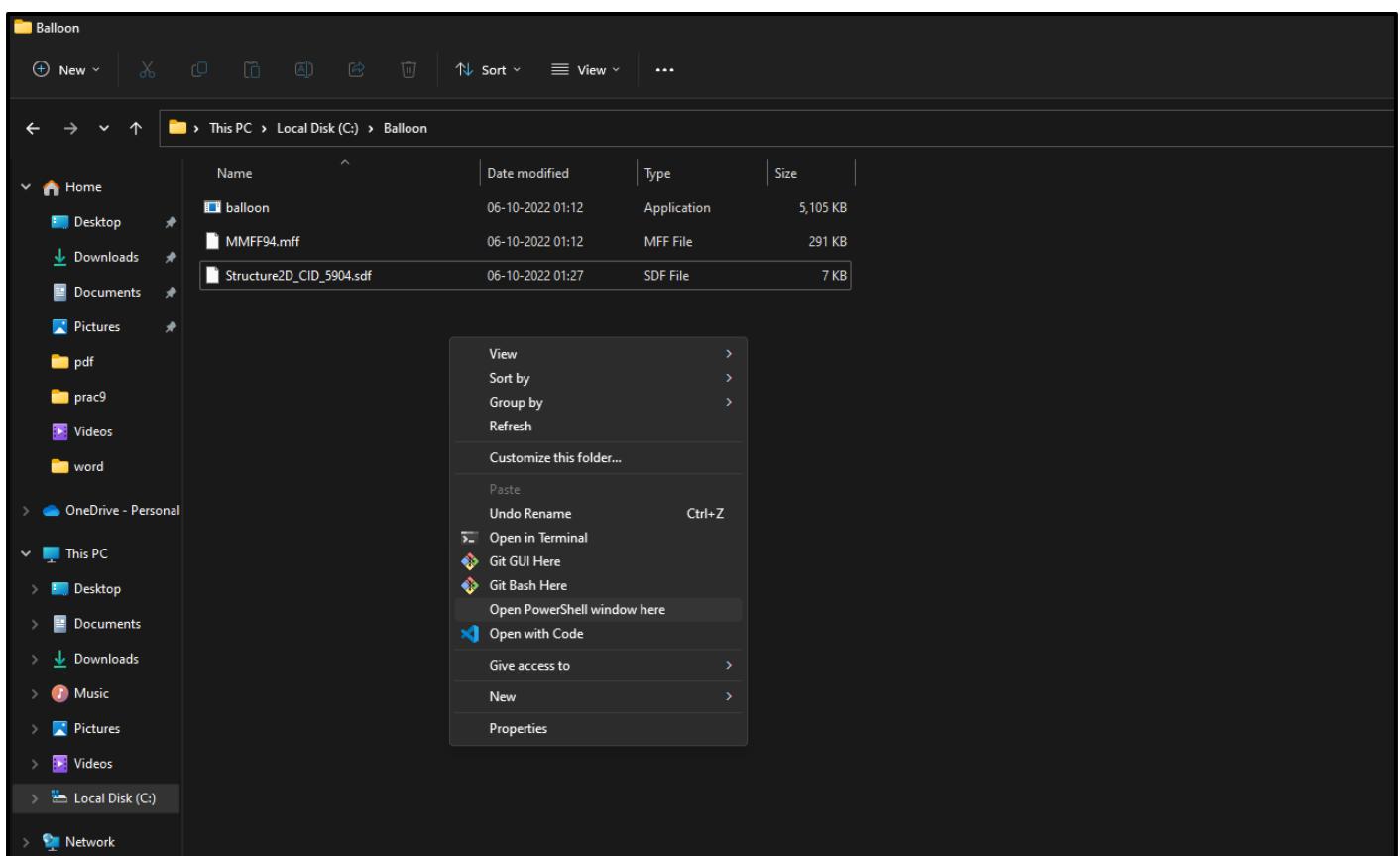
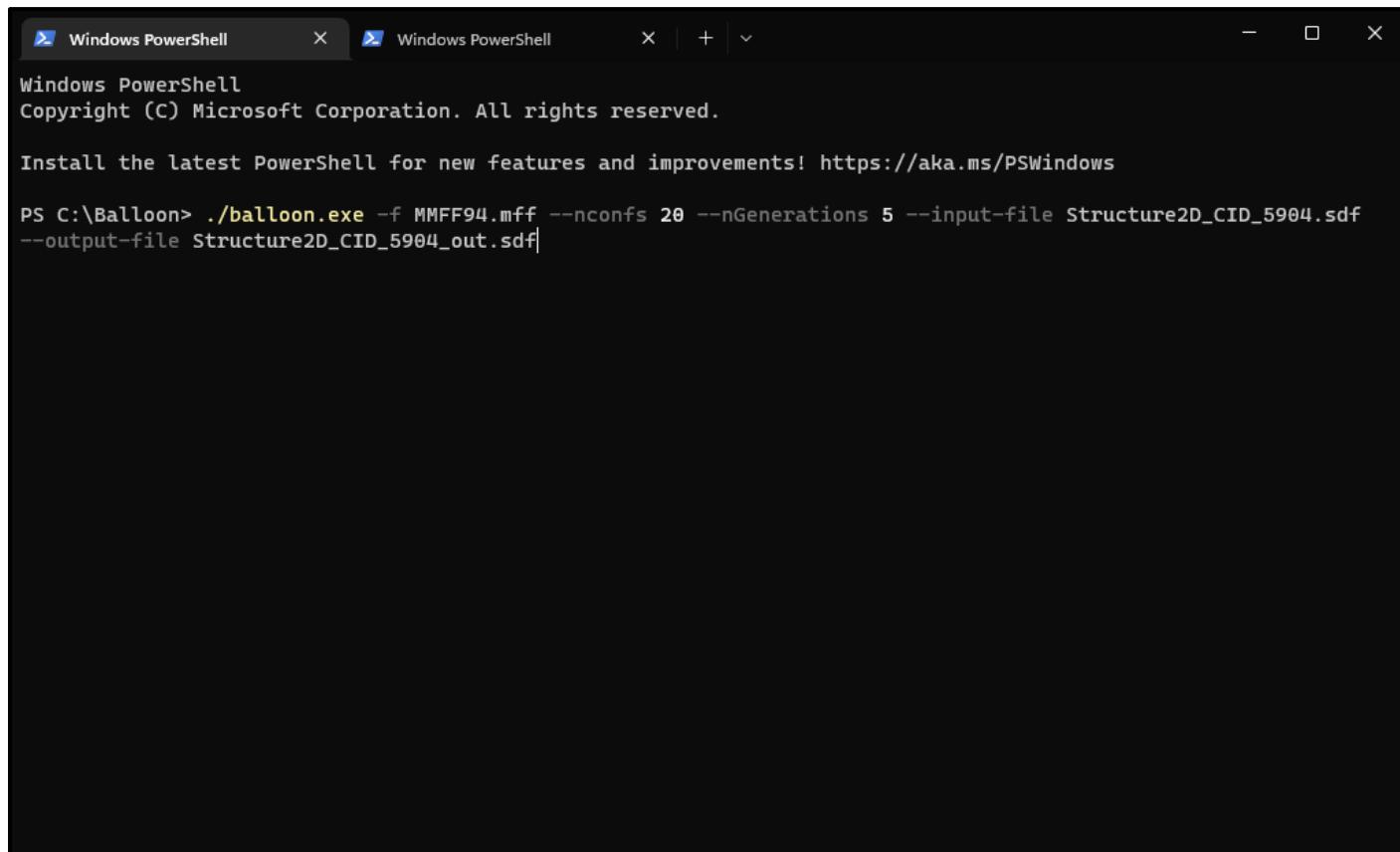


Figure 4: Right-click in the blank area of the folder and click “Open PowerShell window here”

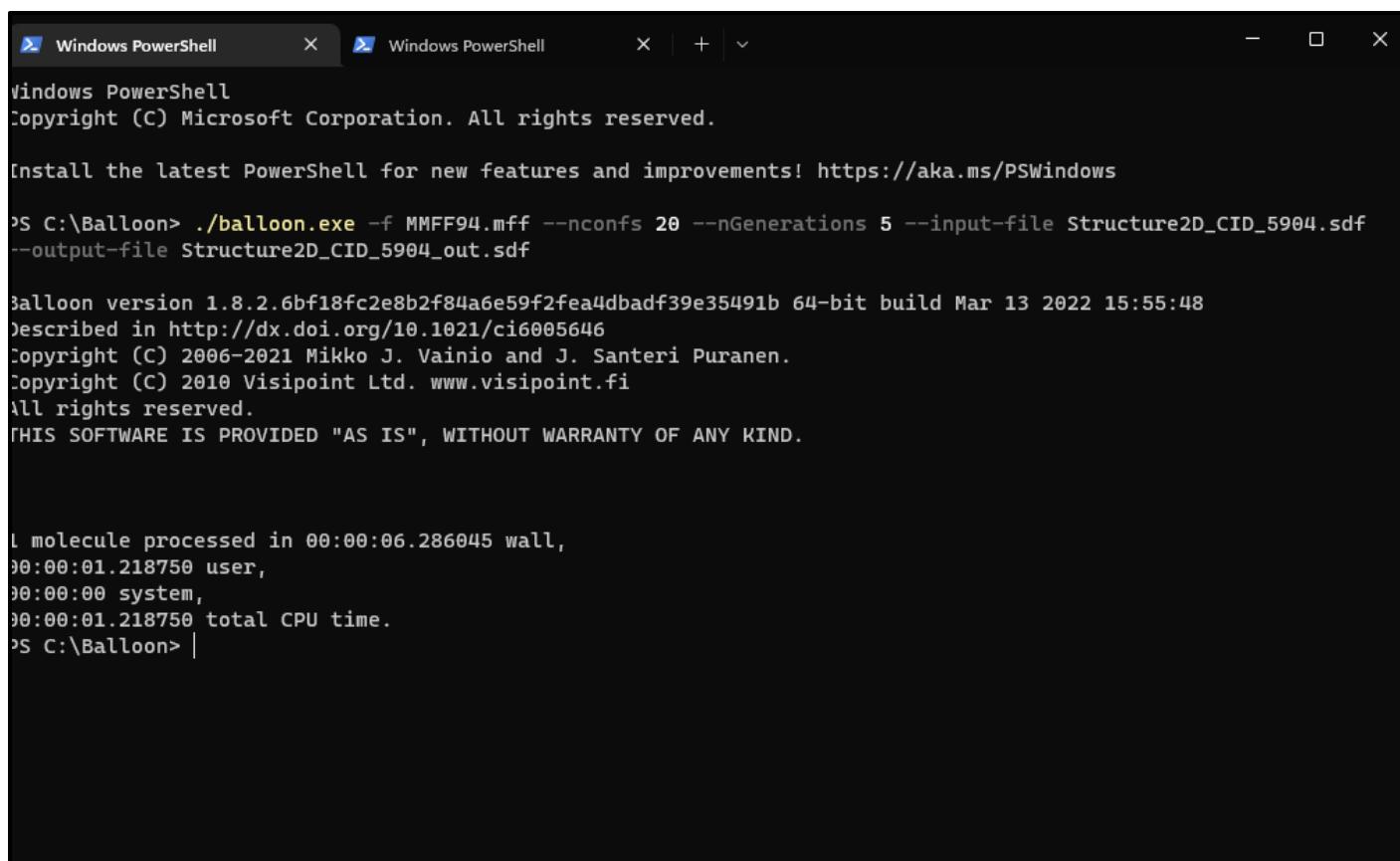


```
Windows PowerShell | Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.

Install the latest PowerShell for new features and improvements! https://aka.ms/PSWindows

PS C:\Balloon> ./balloon.exe -f MMFF94.mff --nconfs 20 --nGenerations 5 --input-file Structure2D_CID_5904.sdf
--output-file Structure2D_CID_5904_out.sdf
```

Figure 5: Enter Command to Generate Conformations for Penicillin Structure



```
Windows PowerShell | Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.

Install the latest PowerShell for new features and improvements! https://aka.ms/PSWindows

PS C:\Balloon> ./balloon.exe -f MMFF94.mff --nconfs 20 --nGenerations 5 --input-file Structure2D_CID_5904.sdf
--output-file Structure2D_CID_5904_out.sdf

Balloon version 1.8.2.6bf18fc2e8b2f84a6e59f2fea4dbadf39e35491b 64-bit build Mar 13 2022 15:55:48
Described in http://dx.doi.org/10.1021/ci6005646
Copyright (C) 2006-2021 Mikko J. Vainio and J. Santeri Puranen.
Copyright (C) 2010 Visipoint Ltd. www.visipoint.fi
All rights reserved.
THIS SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND.

1 molecule processed in 00:00:06.286045 wall,
00:00:01.218750 user,
00:00:00 system,
00:00:01.218750 total CPU time.
PS C:\Balloon> |
```

Figure 7: Execution of command for conformation generations

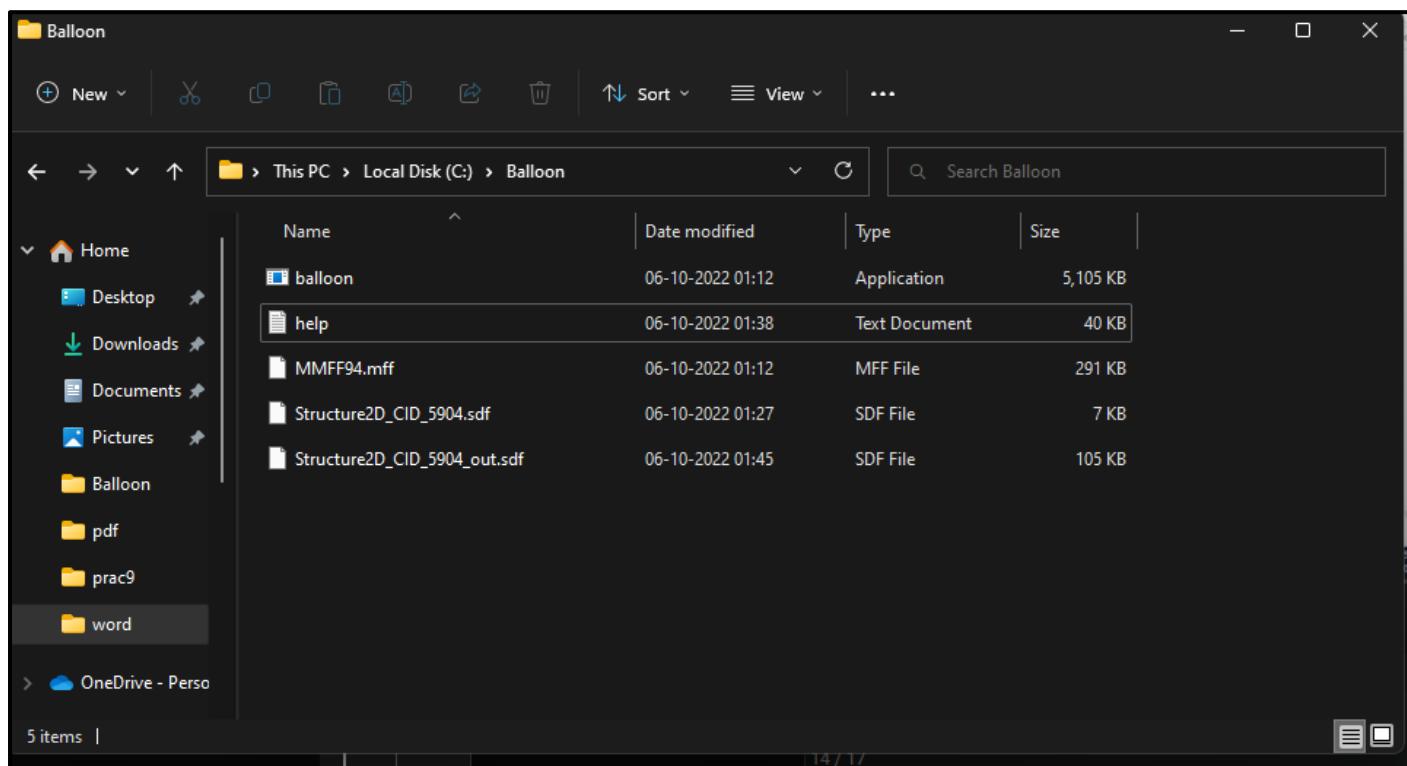


Figure 8: An output file has now been generated in folder

```
41 43 0 0 0 0 0 0 0999 V2000
-2.6097 -0.8563 1.5175 S 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0.3539 -0.4402 -1.5239 O 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
-3.8979 0.3274 -2.3484 O 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
-4.1108 2.2522 -1.2013 O 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
1.5241 -3.7221 -0.1554 O 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
-1.9191 -0.6088 -1.0295 N 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
-0.0995 -2.5314 0.9299 N 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
-2.3156 -1.6723 -0.0896 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
-2.2578 0.8544 0.8974 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
-2.2311 0.7795 -0.6498 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
> <PUBCHEM_OPENEYE_ISO_SMILES>
CC1([C@H](N2[C@H](S1)[C@H](C2=O)NC(=O)CC3=CC=CC=C3)C(=O)O)C

> <PUBCHEM_TOTAL_CHARGE>
0

> <PUBCHEM_XLOGP3>
1.8

> <energy>
48.422748811322933
```

Figure 9: Output File Opening in Notepad Containing Information of Conformations

	A	B	C	D
1	5904_1	48.42274881		
2	5904_2	51.6148302		
3	5904_3	53.73922719		
4	5904_4	54.20091676		
5	5904_5	54.85140344		
6	5904_6	55.32364057		
7	5904_7	55.97578657		
8	5904_8	56.15576191		
9	5904_9	56.17307521		
10	5904_10	56.45572829		
11	5904_11	56.46135098		
12	5904_12	56.85778572		
13	5904_13	57.67564736		
14	5904_14	58.56860916		
15	5904_15	60.11844718		
16	5904_16	60.12870487		
17	5904_17	61.04395685		

Figure 10: Excel Sheet showing Generated Conformations with lowest Energies by BALLOON Software

RESULTS:

To generate and analyze various structural conformations for Penicillin molecule , BALLOON software was used. It had provided 17 different conformations for query structure with their lowest energies. Also, other information like monoisotopic weight, smiles, LogP values etc. were interpreted by BALLOON Software.

CONCLUSION:

Balloon creates 3D atomic coordinates from molecular connectivity via distance geometry and conformer ensembles using a multi-objective genetic algorithm. It is distributed "as is", free of charge, and without warranty of any kind. Its a command-line program and will most likely run on any platform that can run any of the operating systems for which a binary executable is provided. The performance of Balloon is dependent on the performance of the used force field, both time-wise and with regard to the quality of produced geometries.

REFERENCES:

1. Huhtala, M. V. A. M. (n.d.). Balloon. Retrieved October 6, 2022, from <http://users.abo.fi/mivainio/balloon/index.php>
2. NCBI - WWW Error Blocked Diagnostic. (n.d.). Retrieved October 6, 2022, from <https://pubchem.ncbi.nlm.nih.gov/compound/5904>

WEBLEM 4

Introduction to Molecular Descriptors and PaDELPy software

Molecular descriptors can be defined as mathematical representations of molecules' properties that are generated by algorithms. The numerical values of molecular descriptors are used to quantitatively describe the physical and chemical information of the molecules. An example of molecular descriptors is the LogP which is a quantitative representation of the lipophilicity of the molecules, it is obtained by measuring the partitioning of the molecule between an aqueous phase and a lipophilic phase which consists usually of water/n-octanol. Molecular descriptors can be useful in performing similarity searches in molecular libraries, as they can find molecules with similar physical or chemical properties based on their similarity in the descriptors' values. The molecular descriptors are used in ADMET prediction models to correlate the structure–property relationship to help in predicting the ADMET properties of molecules based on their descriptors values (Khan and sylte, 2007).

The molecular descriptors that are used in ADMET models can be classified as being one-dimensional (1D), two-dimensional (2D), or three-dimensional (3D) descriptors based on the level of molecular representation required for calculating the descriptor. The 1D descriptors are the simplest type of molecular descriptors, these represent information that are calculated from the molecular formula of the molecule, which includes the count and type of atoms in the molecule and the molecular weight.

The 2D descriptors are more complex than the 1D descriptors, usually, they represent molecular information regarding the size, shape, and electronic distribution in the molecule. Calculating the 2D descriptors depends mainly on the database size, and the calculation of parts of a molecule in which the data is missing could largely result in a false result.

The 3D descriptors describe mainly properties that are related to the 3D conformation of the molecule, such as the intramolecular hydrogen bonding. Examples of descriptors obtained from calculations involving the 3D structure of the molecules are the polar and nonpolar surface area (PSA and NPSA, respectively). More advanced calculation like quantum mechanics calculations can be used to obtain 3D descriptors that describe the valence electron distribution in the molecules (Bergström, 2005).

Common Molecular Descriptors:

Constitutional

- Functional groups
- Molecular weight
- Simple counts e.g., number of atoms, bonds, rings

Topological

- Atom-pairs⁵
- Balaban index⁶
- BCUT⁷
- Information content indices⁸
- Kappa shape indices⁹
- Kier and Hall connectivity indices¹⁰
- Kier flexibility index¹¹
- Kier shape indices¹¹
- Molecular walk counts¹²
- Randic indices¹³
- Wiener index¹⁴

Geometric

- Gravitation index20
- Molecular surface area
- Molecular volume21
- Shadow indices22
- Solvent accessible molecular surface area

Electrostatic

- Charged polar surface area33
- Galvez topological charge indices34
- Hydrogen bonding capacities
- Maximum and minimum partial charges35
- Molecular polarizabilities36

Fingerprints

- Daylight37
- MDL keys38
- UNITY39

Hydrophobic

- Aromaticity indices3
- Hansch substituent constant4
- Log D
- Log P

Steric

- Charton steric parameter15
- Molar refractivity16
- Parachor17
- Taft steric parameter18

Quantum chemical19

- Charges
- HOMO and LUMO energies
- Orbital electron densities
- Superdelocalizabilities
- Atom-atom polarizabilities
- Molecular polarizabilites
- Dipole moments and polarity indices
- Energies

Combination

- 3D-MoRSE23
- Electrotopological state indices24
- GETAWAY
- LSER26
- MolSurf27
- Moreau-Broto topological autocorrelation28

- Randic molecular profiles²⁹
- RDF³⁰
- VolSurf³¹
- WHIM³²

PaDELPy: A Python wrapper for PaDEL-Descriptor software

PaDELPy provides a Python wrapper for the PaDEL-Descriptor molecular descriptor calculation software. It was created to allow direct access to the PaDEL-Descriptor command-line interface via Python.

Installation

Installation via pip:

```
$ pip install padelpy
```

PaDEL-Descriptor is bundled into PaDELPy, therefore an external installation/download of PaDEL-Descriptor is not necessary. There are currently no additional Python dependencies for PaDELPy, however it requires an installation of the Java JRE version 6+.

Basic Usage

In addition to providing a complete interface between Python and PaDEL-Descriptor's command line tool, PaDELPy offers two functions to acquire descriptors/fingerprints within Python - obtaining descriptors/fingerprints from a SMILES string, and obtaining descriptors/fingerprints from an MDL MolFile

STEPS FOR INSTALLATION OF PADELPy:

Installation of Java JDK and JRE:

The screenshot shows the Oracle Java website. At the top, there's a dark header with the ORACLE logo, a search icon, 'View Accounts' button, and a 'Contact Sales' button. Below the header, a banner for 'JavaOne is back!' is displayed, featuring a Java coffee cup logo and text about the conference. The main content area has a blue background with wavy patterns. It includes links for 'Java downloads', 'Tools and resources', and 'Java archive'. At the bottom, there's a footer bar with links for 'Looking for other Java downloads?', 'OpenJDK Early Access Builds', and 'JRE for Consumers'. The URL 'https://www.oracle.com/java/technologies/downloads/#jdk19-windows' is visible at the bottom left.

STEP 1: Go to homepage of Oracle

(URL: <https://www.oracle.com/java/technologies/downloads/>)

Java SE Development Kit 8u341

Java SE subscribers will receive JDK 8 updates until at least December of 2030.

The Oracle JDK 8 license changed in April 2019

The [Oracle Technology Network License Agreement for Oracle Java SE](#) is substantially different from prior Oracle JDK 8 licenses. This license permits certain uses, such as personal use and development use, at no cost -- but other uses authorized under prior Oracle JDK licenses may no longer be available. Please review the terms carefully before downloading and using this product. FAQs are available [here](#).

Commercial license and support are available for a low cost with [Java SE Subscription](#).

JDK 8 software is licensed under the [Oracle Technology Network License Agreement for Oracle Java SE](#).

[JDK 8u341 checksum](#)

[Linux](#) [macOS](#) [Solaris](#) [Windows](#)

Product/file description	File size	Download
x86 Installer	159.66 MB	jdk-8u341-windows-i586.exe
x64 Installer	173.16 MB	jdk-8u341-windows-x64.exe

STEP 2: Scroll down to Java 8 for windows and download .exe file

You must accept the [Oracle Technology Network License Agreement for Oracle Java SE](#) to download this software.



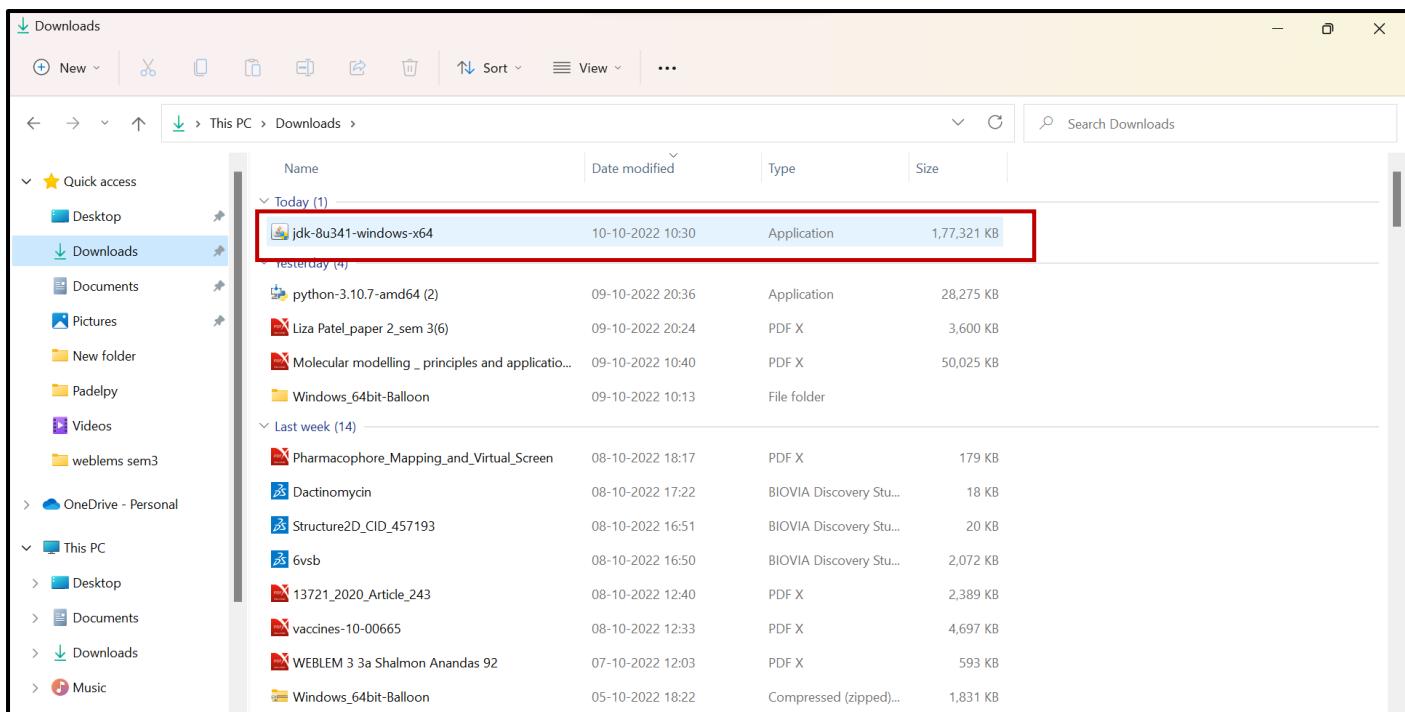
I reviewed and accept the Oracle Technology Network License Agreement for Oracle Java SE

Required

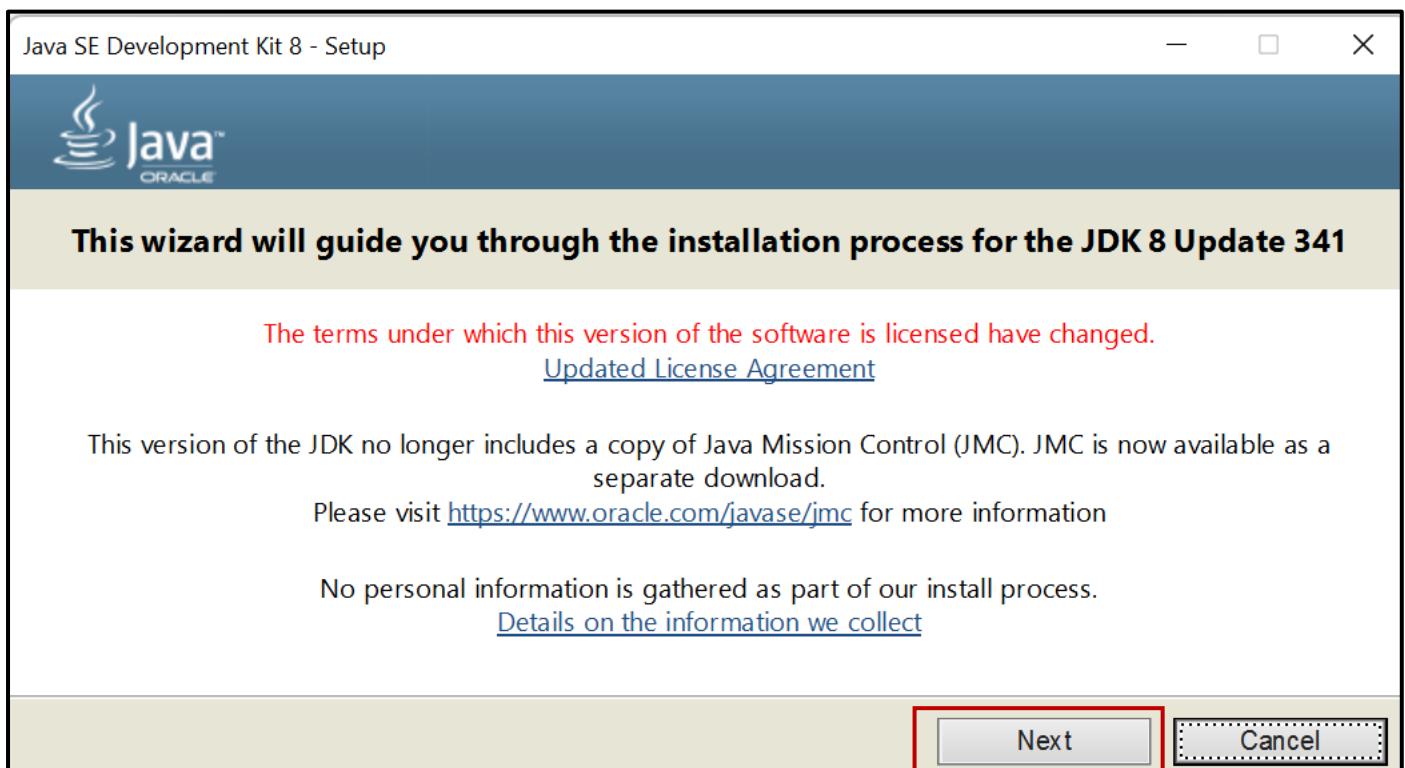
You will be redirected to the login screen in order to download the file.

[Download jdk-8u341-windows-x64.exe](#)

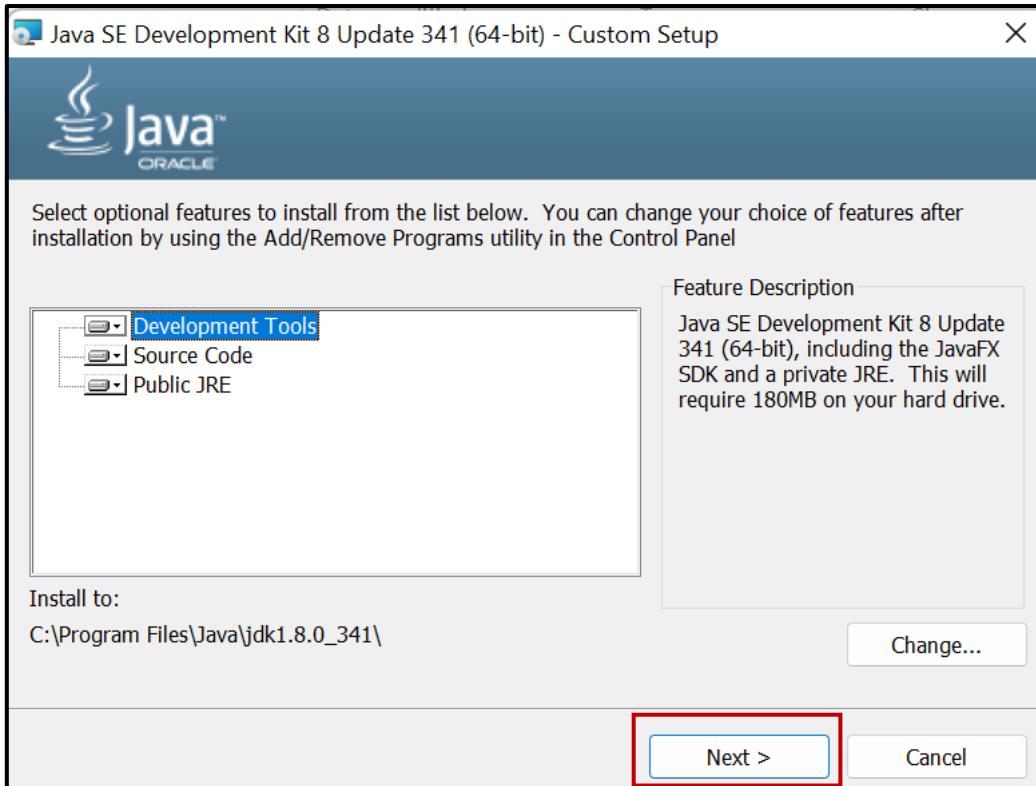
STEP 3: Accept the license agreement and click on download



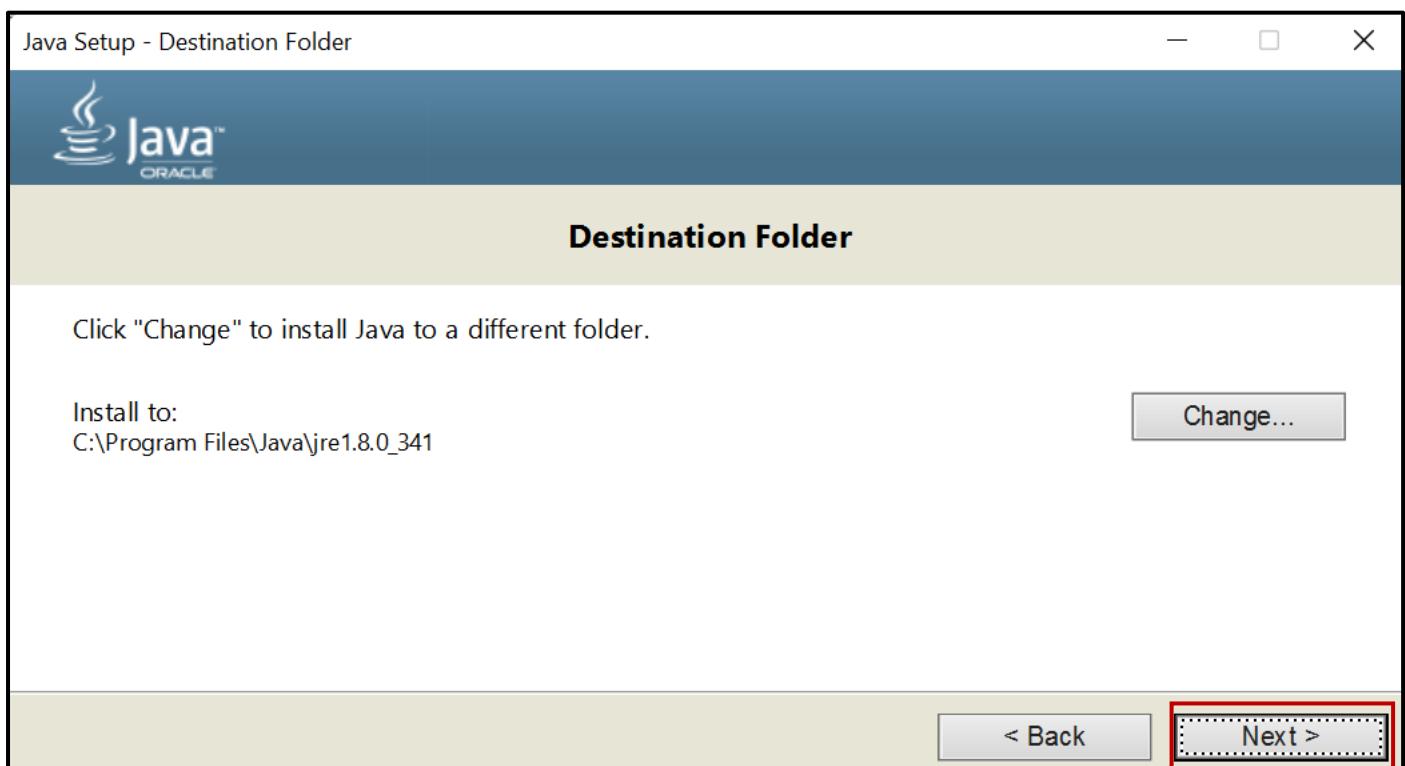
STEP 4: Click on the downloaded .exe file from downloads



STEP 5: Click on next option for setup



STEP 6: Click on next for custom setup



STEP 7: Click on next for selection of destination folder



STEP 8: After installation is complete click on close option

The screenshot shows the Java download page. The "JRE 8" section is highlighted with a red box. It includes links for "Documentation Download" and "Server JRE 8". The "Server JRE (Java SE Runtime Environment) 8u341" link is also highlighted with a red box. A note states: "Server JRE 8 software is licensed under the Oracle Technology Network License Agreement for Oracle Java SE". Below this, there are download options for "Linux", "Solaris", and "Windows". The "Windows" option is highlighted with a red box. The "x64 Installer" download link is also highlighted with a red box.

STEP 9: Scroll down to JRE 8 on oracle page and download .gz file for windows

You must accept the [Oracle Technology Network License Agreement](#) for Oracle Java SE to download this software.



I reviewed and accept the Oracle Technology Network License Agreement for Oracle Java SE

Required

[Download server-jre-8u341-windows-x64.tar.gz](#) 

STEP 10: Accept license agreement and click on download

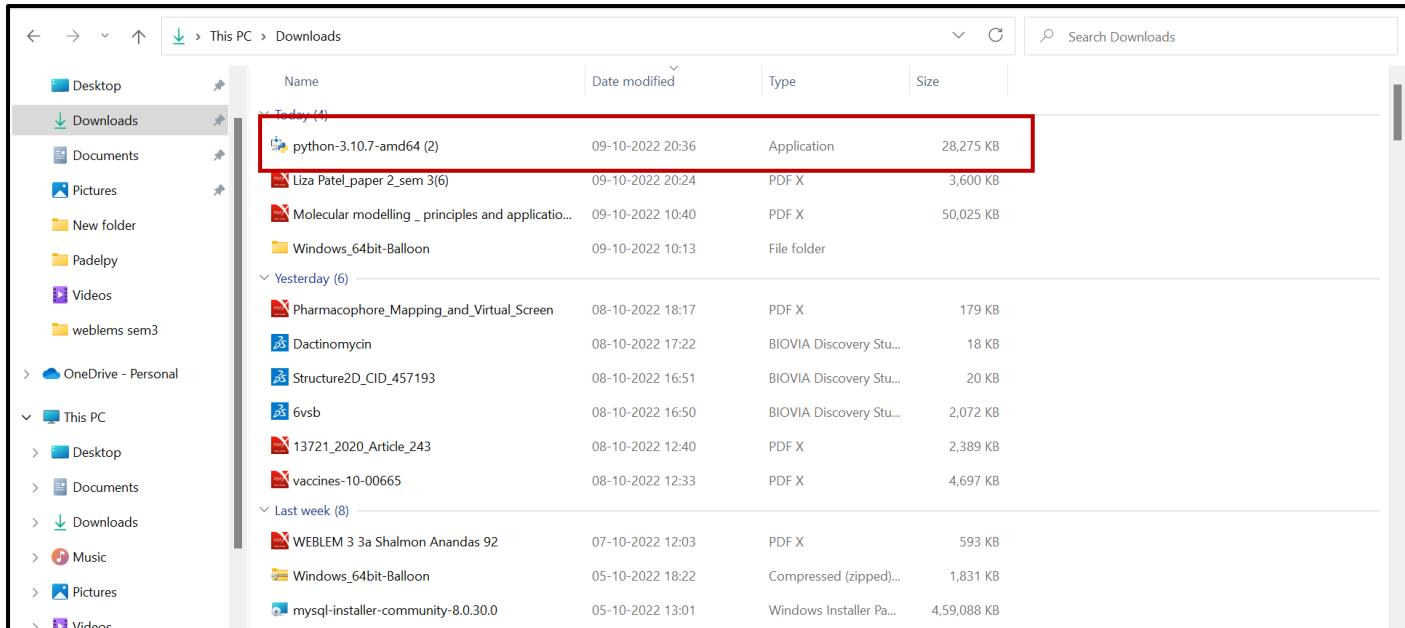
Installation of Python:



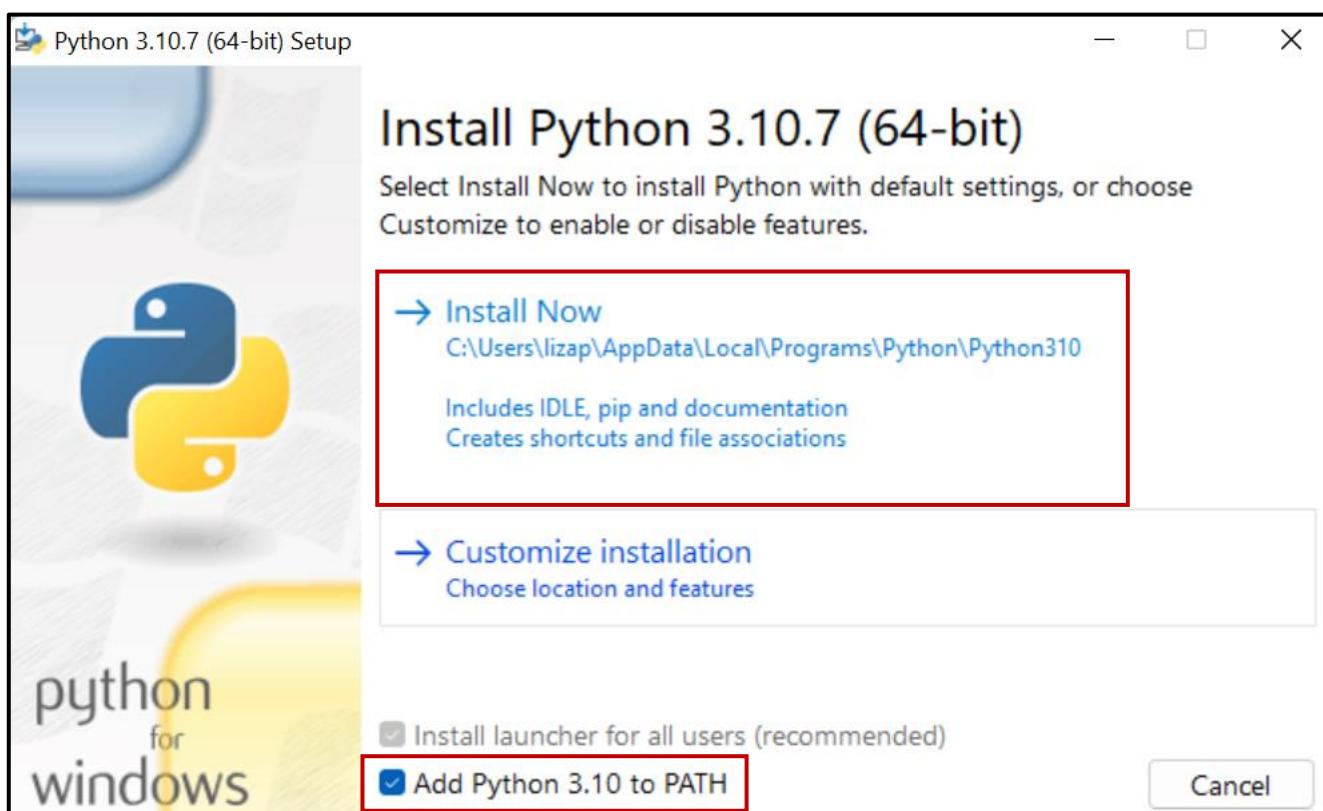
The screenshot shows the official Python website homepage. At the top, there's a navigation bar with tabs for Python, PSF, Docs, PyPI, Jobs, and Community. Below the navigation bar is the Python logo and a search bar with a 'GO' button. A prominent yellow button labeled 'Download the latest version for Windows' is centered on the page. This button is highlighted with a red rectangular box. Below it, there's text for other operating systems like Linux/UNIX, macOS, and Other, along with links for Prereleases and Docker images. To the right of the text, there's a cartoon illustration of two boxes with parachutes falling from the sky. At the bottom left, the URL 'https://docs.python.org' is visible.

STEP 11: Go to homepage of python and click on download python 3.10.7

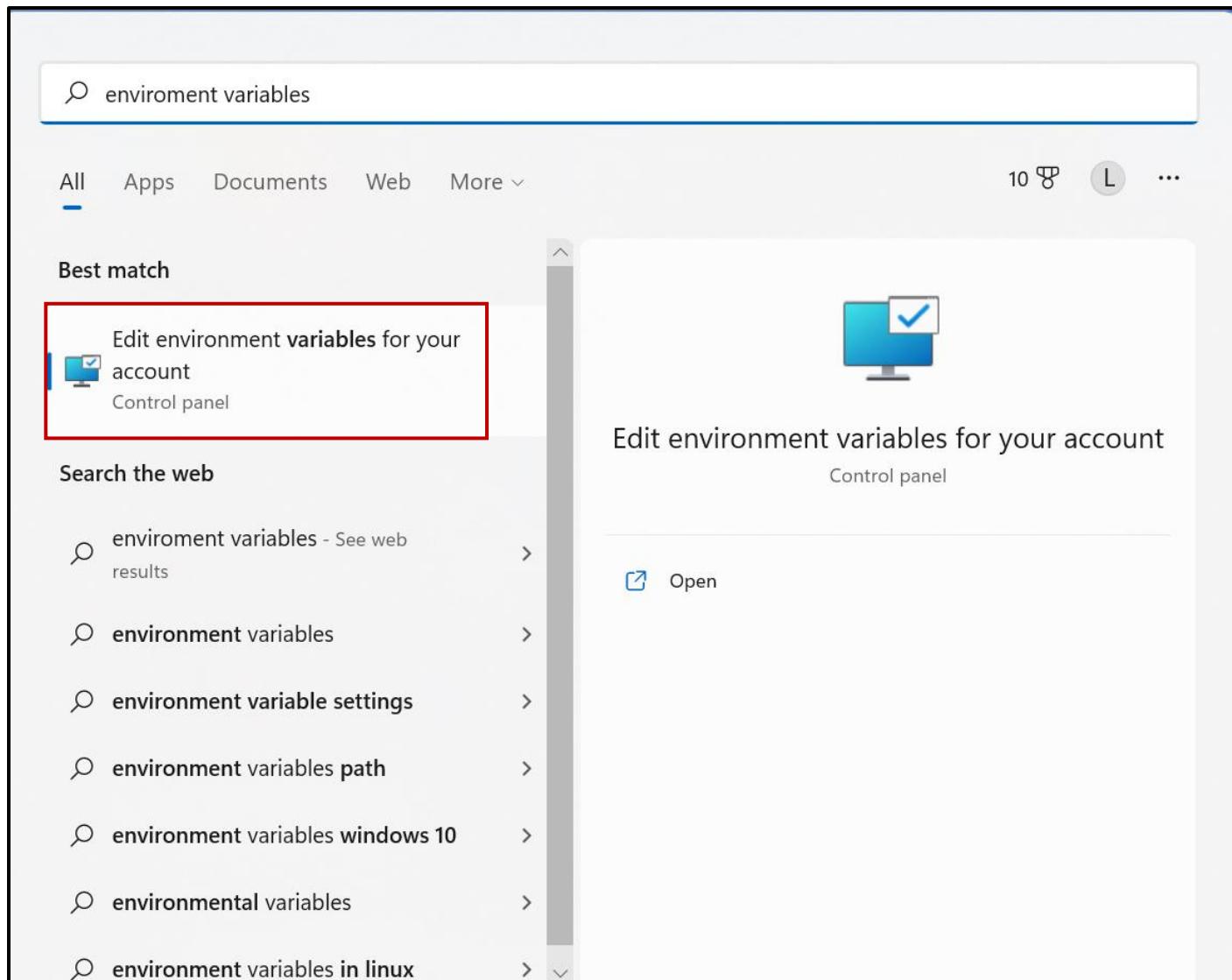
(URL: <https://www.python.org/downloads/>)



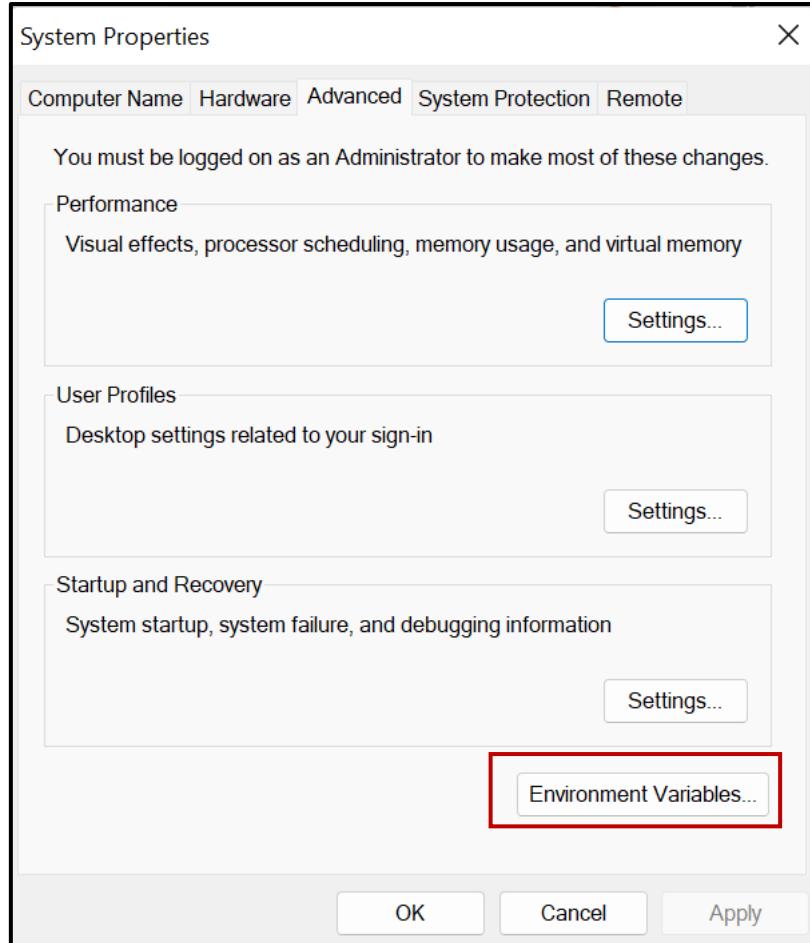
STEP 12: Go to downloads and open the .exe file



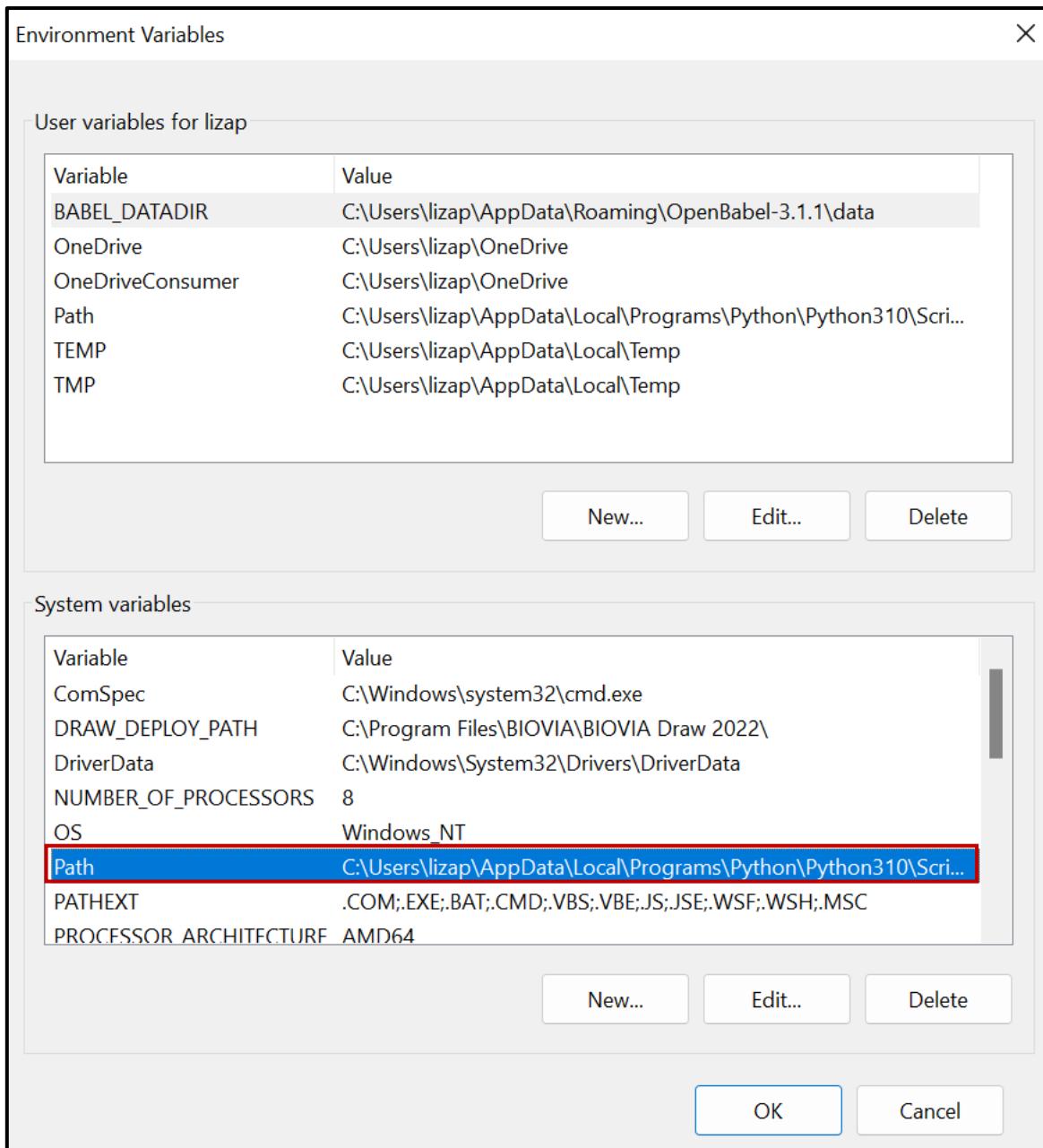
STEP 13: Select Add Python 3.10 to PATH and click on Install Now



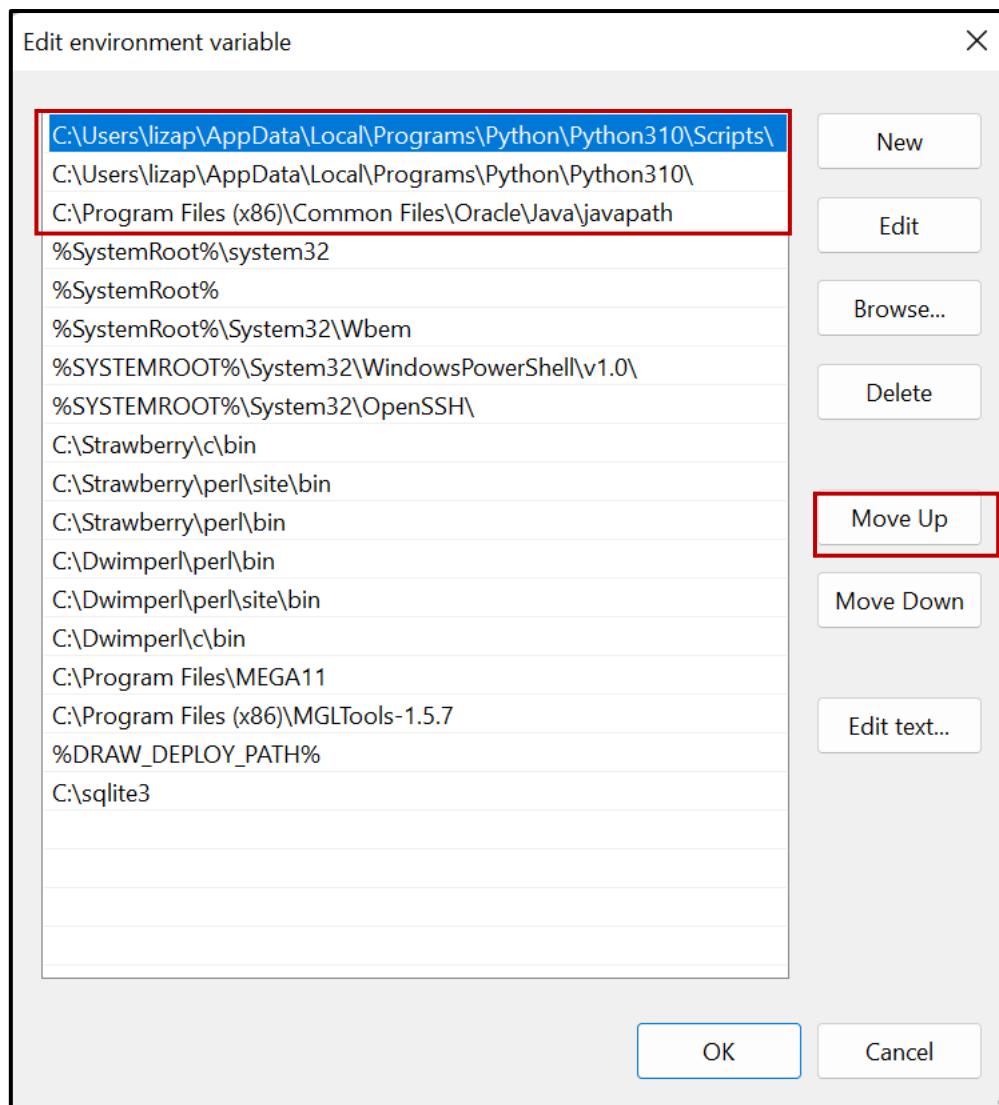
STEP 14: Search for environment variables option on your system and open it



STEP 15: Click on environment variables option



STEP 16: Double click on path



STEP 17: Pull path for python and java to the top using move up key and click on ok

```
C:\> Command Prompt
Microsoft Windows [Version 10.0.22000.978]
(c) Microsoft Corporation. All rights reserved.

C:\Users\lizap>pip install padelpy
Requirement already satisfied: padelpy in c:\users\lizap\appdata\local\programs\python\python310\lib\site-packages (0.1.12)

C:\Users\lizap>
```

STEP 18: Open command prompt and type “pip install padelpy” and press enter

Application:

Molecular descriptors play important roles in the fields of quantitative structure–activity relationship studies (QSAR) as well as quantitative structure–property relationship studies (QSPRs). The path breaking progress in the field of chemoinformatics has showed us new paths for identifying key links between the molecular structure and their biological properties. Molecular descriptors can be useful in performing similarity searches in molecular libraries, as they can find molecules with similar physical or chemical properties based on their similarity in the descriptors’ values.

REFERENCES:

1. *Molecular Descriptor - an overview / ScienceDirect Topics.* (n.d.). [Www.sciencedirect.com](http://www.sciencedirect.com). Retrieved October 9, 2022, from <https://www.sciencedirect.com/topics/medicine-and-dentistry/molecular-descriptor#:~:text=A%20molecular%20descriptor%20is%20a>
2. Kamath, V., & Pai, A. (2017). Application of Molecular Descriptors in Modern Computational Drug Design-An Overview. *Research Journal of Pharmacy and Technology*, 10(9), 3237. <https://doi.org/10.5958/0974-360x.2017.00574.1>
3. *PaDELPy: A Python wrapper for PaDEL-Descriptor software.* (2022, September 28). GitHub. Retrieved October 9, 2022, from <https://github.com/ecrl/padelpy#readme>
4. *Download Python.* (2019). Python.org; Python.org. Retrieved October 9, 2022, from <https://www.python.org/downloads/>
5. *Download the Latest Java LTS Free.* (2021). Oracle.com. Retrieved October 9, 2022, from <https://www.oracle.com/java/technologies/downloads/>

WEBLEM 4a

PaDELPy

(URL: <https://github.com/ecrl/padelpy#readme>)

AIM:

To study molecular descriptors for Ibuprofen (Pubchem Id- 3672) using PaDELPy software and analyze the results.

INTRODUCTION:

Ibuprofen is (2RS)-1[4-(2-methyl propyl) phenyl] propionic acid (BP. 2004). Ibuprofen was the first member of propionic acid derivatives to be introduced in 1969 as a better alternative to Aspirin. Gastric discomfort, nausea and vomiting, though less than aspirin or indomethacin, are still the most common side effects.

Ibuprofen is the most commonly used and most frequently prescribed NSAID.^{2,3} It is a non-selective inhibitor of cyclo-oxygenase-1 (COX-1) and Cyclooxygenase-2 (COX-2). Although, its anti-inflammatory properties may be weaker than those of some other NSAIDs, it has a prominent analgesic and antipyretic role. Its effects are due to the inhibitory actions on cyclo-oxygenases, which are involved in the synthesis of prostaglandins. Prostaglandins have an important role in the production of pain, inflammation and fever.

PaDELPy provides a Python wrapper for the PaDEL-Descriptor molecular descriptor calculation software. It was created to allow direct access to the PaDEL-Descriptor command-line interface via Python.

PaDEL-Descriptor is bundled into PaDELPy, therefore an external installation/download of PaDEL-Descriptor is not necessary. There are currently no additional Python dependencies for PaDELPy, however it requires an installation of the Java JRE version 6+.

In addition to providing a complete interface between Python and PaDEL-Descriptor's command line tool, PaDELPy offers two functions to acquire descriptors/fingerprints within Python - obtaining descriptors/fingerprints from a SMILES string, and obtaining descriptors/fingerprints from an MDL MolFile

METHODOLOGY:

- Retrieve canonical SMILES for Ibuprofen from Pubchem database (URL: <https://pubchem.ncbi.nlm.nih.gov/>)
- Open github page for PaDELPy: A Python wrapper for PaDEL-Descriptor software (URL: <https://github.com/ecrl/padelpy#readme>)
- Copy script for SMILES to Descriptors/Fingerprints to notepad
- In the script for function “from_smiles” pass canonical SMILES of Ibuprofen as parameter.
- Save it as script1.py to a folder named Padelpy
- Open command prompt
- Give path for Padelpy folder
- Run script1
- Observe and interpret the results

OBSERVATIONS:



```
script1 - Notepad
File Edit View
from padelpy import from_smiles
# calculate molecular descriptors for ibuprofen
descriptors = from_smiles('CC(C)CC1=CC=C(C=C1)C(C)C(=O)O')
# calculate molecular descriptors for ibuprofen and butane
descriptors = from_smiles(['CC(C)CC1=CC=C(C=C1)C(C)C(=O)O', 'CCCC'])
# in addition to descriptors, calculate PubChem fingerprints
desc_fp = from_smiles('CC(C)CC1=CC=C(C=C1)C(C)C(=O)O', fingerprints=True)
# only calculate fingerprints
fingerprints = from_smiles('CC(C)CC1=CC=C(C=C1)C(C)C(=O)O', fingerprints=True, descriptors=False)
# setting the number of threads, this uses one cpu thread to compute descriptors
descriptors = from_smiles(['CC(C)CC1=CC=C(C=C1)C(C)C(=O)O', 'CCCC'], threads = 1)
# save descriptors to a CSV file
_ = from_smiles('CC(C)CC1=CC=C(C=C1)C(C)C(=O)O', output_csv='descriptors.csv')
```

FIG 1. Python script for SMILES to Descriptor/Fingerprints

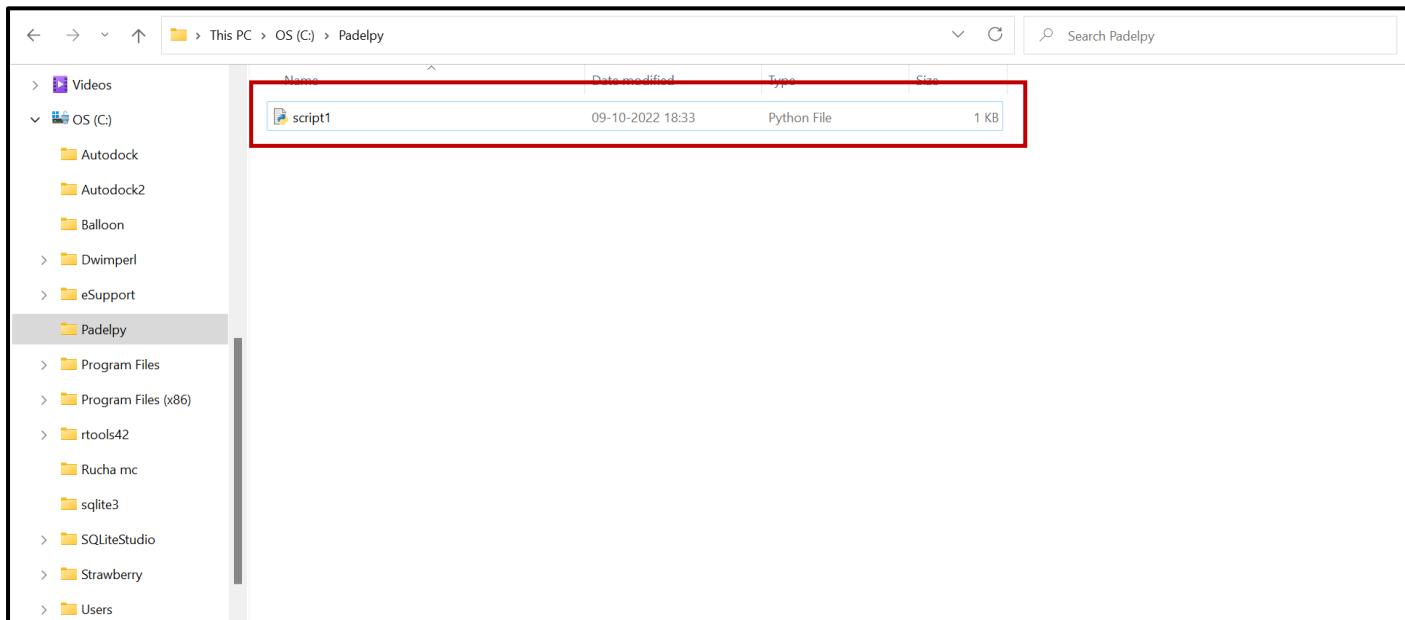


FIG 2. Script saved as script1.py in Padelpy folder

```
Command Prompt
Microsoft Windows [Version 10.0.22000.978]
(c) Microsoft Corporation. All rights reserved.

C:\Users\lizap>cd..

C:\Users>cd..

C:\>cd Padelpy

C:\Padelpy>
```

FIG 3. Path to Padelpy folder in command prompt

```
Command Prompt
Microsoft Windows [Version 10.0.22000.978]
(c) Microsoft Corporation. All rights reserved.

C:\Users\lizap>cd..

C:\Users>cd..

C:\>cd Padelpy

C:\Padelpy> pip install padelpy
Requirement already satisfied: padelpy in c:\users\lizap\appdata\local\programs\python\python310\lib\site-packages (0.1.12)

C:\Padelpy>script1.py
```

FIG 4. Running script1.py

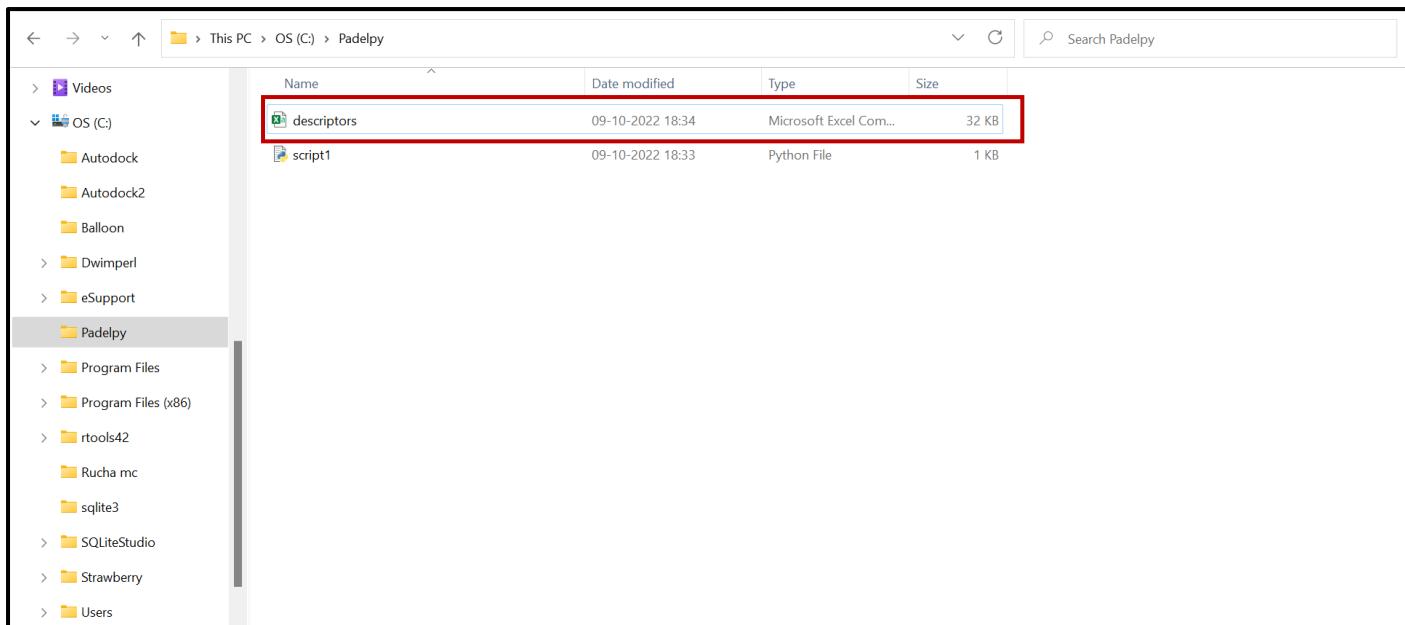


FIG 5. Descriptors excel file as .csv generated in Padelpy folder

A screenshot of Microsoft Excel. The ribbon is visible at the top with tabs like Home, Insert, Page Layout, etc. The main area shows a table titled 'descriptors1'. Row 1 contains column headers: Name, nAcid, ALogP, ALogP2, AMR, apol, naAromAt, nAromBor, nAtom, nHeavyAt, nH, nB, nC, nN, nO, nS, nP, nF, nR, nCl, nB. Row 2 contains data for Ibuprofen: AUTOPEN, 1, 1.9613, 3.846698, 64.1114, 36.48627, 0, 0, 33, 15, 18, 0, 13, 0, 2, 0, 0, 0, 0, 0. The status bar at the bottom says 'descriptors1'.

Name	nAcid	ALogP	ALogP2	AMR	apol	naAromAt	nAromBor	nAtom	nHeavyAt	nH	nB	nC	nN	nO	nS	nP	nF	nR	nCl	nB
AUTOPEN	1	1.9613	3.846698	64.1114	36.48627	0	0	33	15	18	0	13	0	2	0	0	0	0	0	
3																				
4																				
5																				
6																				
7																				
8																				
9																				
10																				
11																				
12																				
13																				
14																				
15																				
16																				
17																				
18																				
19																				

FIG 6. Descriptors calculation for Ibuprofen molecule

RESULTS:

PaDELPy which is a Python wrapper for the PaDEL-Descriptor molecular descriptor calculation software was used to calculate molecular descriptors of Ibuprofen.

CONCLUSION:

PaDELPy provides a Python wrapper for the PaDEL-Descriptor molecular descriptor calculation software. It allows direct access to the PaDEL-Descriptor command-line interface via Python. Thus, PaDELPy can be used for quantitative structure–activity relationship studies (QSAR) as well as quantitative structure–property relationship studies (QSPRs) as molecular descriptors play an important role here. Molecular descriptors can be useful in performing similarity searches in molecular libraries, as they can find molecules with similar physical or chemical properties based on their similarity in the descriptors’ values.

REFERENCES:

1. *Molecular Descriptor - an overview / ScienceDirect Topics.* (n.d.). [Www.sciencedirect.com](http://www.sciencedirect.com). Retrieved October 9, 2022, from <https://www.sciencedirect.com/topics/medicine-and-dentistry/molecular-descriptor#:~:text=A%20molecular%20descriptor%20is%20a>
2. Kamath, V., & Pai, A. (2017). Application of Molecular Descriptors in Modern Computational Drug Design-An Overview. *Research Journal of Pharmacy and Technology*, 10(9), 3237. <https://doi.org/10.5958/0974-360x.2017.00574.1>
3. *PaDELPy: A Python wrapper for PaDEL-Descriptor software.* (2022, September 28). GitHub. Retrieved October 9, 2022, from <https://github.com/ecrl/padelpy#readme>
4. Bushra, R., & Aslam, N. (2010). An overview of clinical pharmacology of Ibuprofen. *Oman Medical Journal*, 25(3), 155–1661. <https://doi.org/10.5001/omj.2010.49>

WEBLEM 5

Introduction to Pharmacophore mapping

When the 3D structure of the protein target has not been characterized, and/or when a certain number of ligands (with or without associated binding affinity) are available, pharmacophore models can be developed and used as search queries for virtual screening of databases. Pharmacophore models may range from sub-structural type pharmacophores to feature-based pharmacophores, in the latter the pharmacophoric points are represented by chemical features like hydrogen - bond acceptors/donors, hydrophobic points, acidic or basic features etc. Moreover, when the necessity occurs to move to the 3D level, virtual screening has to deal with enhanced complexity with regard to functionality and flexibility of molecules, which requires more sophisticated tools for analyzing this type of data. For implementation of this concept into virtual screening, the chemical function-based approach is the most generic one. The originality of this type of pharmacophores mostly resides in the fact that their definition is general and represents different types of interactions between organic molecules and proteins. The utility of such models as queries for 3D database search has been recently reviewed. Such pharmacophores can be generated indifferently from ligand sets or from an active site structure. At the end of virtual screening filtering procedure, a reliable method for ranking the hits obtained according to their expected bioactivity is required.

Methods for Pharmacophore Generation There are two ways to deduce a pharmacophore: direct- and indirect-methods. The former uses both the ligand and the receptor information, while the latter employs only a collection of ligands that have been experimentally observed to interact with a given receptor. Indirect methods can be used even in the absence of structure of the receptor and hence are more advantageous in the present scenario where the crystal structures of less than 10 % of drug targets are available. However, direct methods are becoming extremely important with the rapidly increasing number of known protein structures, which is the outcome of the Structural Genomics project. Once identified, a pharmacophore model is a versatile tool for the discovery and development of new lead compounds. **Steps in Identifying a Pharmacophore** In general, all the algorithms for pharmacophore identification utilize the following six steps:

- 1) Input
- 2) Conformational Search
- 3) Feature extraction
- 4) Structure Representation
- 5) Pattern Identification
- 6) Scoring

PharmMapper:

PharmMapper server is a freely accessed web server designed to identify potential target candidates for the given small molecules (drugs, natural products or other newly discovered compounds with unidentified binding targets) using pharmacophore mapping approach. PharmMapper hosts a large, in-house repertoire of pharmacophore database (namely PharmTargetDB) annotated from all the targets information in TargetBank, BindingDB, DrugBank and potential drug target database, including over 7000 receptor-based pharmacophore models (covering over 1500 drug targets information). PharmMapper automatically finds the best mapping poses of the query molecule against all the pharmacophore models in PharmTargetDB and lists the top N best-fitted hits with appropriate target annotations, as well as respective molecule's aligned poses are presented. Benefited from the highly efficient and robust triangle hashing mapping method, PharmMapper bears high throughput ability and only costs 1 h averagely to screen the whole PharmTargetDB.

WORKING:

Construction of potential targets pharmacophore databases

PharmMapper requires a sufficient number of available pharmacophore models describing the binding modes of known ligands at the binding sites of protein targets. The target protein structures co-complexed with small molecules were carefully selected from DrugBank, BindingDB, PDBBind and our PDTD databases. DrugBank hosts a complete list of known targets with appropriate annotations, while BindingDB and PDBBind provide public, web-accessible databases of measured binding affinities, focusing chiefly on the interactions of those proteins considered to be drug targets with small or drug-like molecules. Only those proteins with available 3D crystal structures were selected and used for pharmacophore model extraction.

LigandScout, which is a software tool that allows rapid extraction of 3D pharmacophores from structural data of macromolecule–ligand complexes in a fully automated and convenient way, was used in the process of pharmacophore model derivation. Six primary types of pharmacophore features were adopted in this process: hydrophobic center (H), positive-charged center (P), negative-charged center (N), hydrogen bond acceptor vector (HBA), hydrogen bond donor vector (HBD) and aromatic plane (AR) and one optional feature [metal interaction center (M)]. Each ligand binding site was manually analyzed after generation of corresponding pharmacophore model and the corresponding shape was characterized by several excluded volumes centered at each residue of the binding pocket. All the small ligands with molecular weight lower than 100, such as solvents, buffers and metal cations, and all the cofactors with molecular weight over 600, such as CoAs, polypeptides and nucleic acids were regarded as ‘environment atoms’ instead of binding ligands. In this context, the corresponding pharmacophore models were not generated. For the proteins existing as homopolymers, only one monomer was reserved for analysis. For the proteins determined by NMR with multiple structure models, only the first model was selected for pharmacophore generation. As a result, we generated 7302 pharmacophore models (2241 entries are annotated as ‘Human protein targets’) and deposited them in PharmTargetDB. The target annotations were extracted from DrugBank, PDBSum, UniProt and in-house TargetBank (our unpublished data) and were categorized as follows: UniProt access ID, target name, target function and indication/disease involved.

Reverse pharmacophore mapping procedure using PharmMapper

PharmMapper consists of two parts: a front-end web interface written in both PHP and HTML, with MySQL as database system, and a back-end tool for reverse pharmacophore mapping. The reverse pharmacophore mapping procedure is as follows: (i) PharmMapper flexibly aligns the given small molecule onto each pharmacophore model of proteins in the target list, and the fit values between the small molecule and the pharmacophores are calculated and recorded; (ii) PharmMapper presents the aligned pose with the corresponding pharmacophore model and prioritizes candidate targets based on the fit values to analyze the reverse mapping result. In general, PharmMapper outputs the top N hits of the ranking list, from which the user may select protein candidates for further bioassay validation.

Generally, the algorithm suggests to solve the molecule pharmacophore best fitting task in a strategy of sequential combination of triangle hashing (TriHash) and genetic algorithm (GA) optimization, which consists of following major steps: (i) ligand initialization and preparation; (ii) ligand as well as target pharmacophore model features triangulation; (iii) pairwise alignment and GA post optimization; and (iv) solution filtering, ranking and output. The readers can refer to the Supplementary Data for more details about the pharmacophore mapping algorithm used by PharmMapper.

PharmMapper server is open-accessed and free of charge. Users are expected to upload the mol2 file of the test molecule, customize the mapping parameters and submit a job. A job identity number, namely the JOB ID, is assigned to each job by the web server, and the number is appended to a job queue in the back-end server. The user may use the JOB ID to check the status of the submitted job.

Input

PharmMapper's interface is very simple. Its input form has only one mandatory field: a file with single drug-like molecule or natural product stored in Mol2 format. The user must make sure the uploaded molecule has appropriate 3D structural information. Multiple commercial or open source toolkits are recommended to complete this task, including CORINA, CONCORD and ChemAxon's Standardizer. The user can choose or not to leave an email address in order to receive a notification when the job is finished. After uploading the file, the user is encouraged to set some optional parameters in the following pop-up form instead of accepting corresponding default values to reduce the computational cost or achieve more accurate result. Since PharmMapper uses semi-flexible alignment strategy, a conformer ensemble has to be generated prior to mapping. For single 3D conformer provided by the user, an in-house program Cyndi is used by default to generate multiple conformations. Of course, the user can skip this step by uploading pre-generated conformation ensemble with other programs, such as CAESAR, MacroModel and Omega. Additionally, the user can specify the minimum number of each pharmacophore feature type to skip those target pharmacophore models, of which the number of corresponding pharmacophore features are less than the threshold values. Moreover, the scoring weights assigned to each type of pharmacophore feature can be adjusted according to the user's judgment towards the structural, physicochemical features presented by the molecule (e.g. if the molecule bears dominantly hydrophobic features, the scoring weight assigned to the Hydrophobic Score can be moderately increased to favor the hydrophobic interaction with the pharmacophore models). Detailed explanations for each field can be displayed in the pop-up windows when the mouse is lifted on the corresponding field and are also available in the Help page.

Output

A typical run of PharmMapper task takes 1–2 h, depending on the flexibility of the input molecule and filter parameters assigned by the user. To ensure successful job submission, the user is prompted to activate a self-refreshed alert page to monitor the job status. The user can bookmark this alert page so as to check the status of corresponding job at any time in the feature. Once the job completes, the user is automatically redirected to the computational results via the self-refreshed page or expected to input the JOB ID in the 'Get Result' page to access the computational results. The hyperlink to the result page is also contained in the notification after the job is finished, if the users have left their email address during job submission. The result will be kept on the server for up to 3 months so that the user may access the result at any time later via the same JOB ID.

The output of a PharmMapper run is demonstrated in the form of a ranked list of hit target pharmacophore models that are sorted by fit score in descending order. User can also re-rank the result list by normalized fit score or number of pharmacophore features in descending order via clicking the arrow icons in the corresponding columns. The 3D structural information can be accessed via the hyperlinks in the 'PDB ID' column to the Protein Database Bank (PDB) website. The hotlink to UniProt database as well as functional and therapeutic annotations of each target will be presented in the pop-up window by mouse lifting over the corresponding PDB IDs. As Figure 1B shows, a pull-down window will appear by clicking the '+' mark at the starting of each line of the result table, which illustrates the details of each pharmacophore model candidate, including the numbers of each pharmacophore feature (rendered in different colors scheme), a 3D interactive visualization of molecule-pharmacophore alignment poses displayed via a modified version of Jmol applet (<http://www.jmol.org>), and the download links of the aligned pose of molecule as well as the corresponding pharmacophore model (in hypoedit format). The radio buttons in the pull-down window allow the users to show/hide either the pharmacophore model, query molecular conformation or the features from the query molecule in display, which may provide better visual assessment for the matching quality between the input probe molecule and the identified potential target pharmacophore models. All the text-based targets information is downloadable in comma separated values (CSV) format via the hotlink at the bottom of the result page.

ZINCPharmer:

A pharmacophore describes the structural arrangement of the essential molecular features of an interaction between a ligand and its receptor. Searching chemical libraries for compounds that match a specific pharmacophore is an established method of virtual screening. The two main challenges of pharmacophore-based virtual screening are identifying a representative pharmacophore for an interaction and then identifying the compounds within a relevant chemical library that match the pharmacophore. ZINCPharmer is a pharmacophore search engine for purchasable chemical space that addresses both these challenges.

An interaction pharmacophore may be elucidated from a set of known active ligands by identifying a consensus pharmacophore that is conformationally accessible to all these ligands. These techniques do not require a ligand-bound structure, but may be computationally demanding if the input set contains many flexible ligands. PharmaGist is a free web server that can identify a consensus pharmacophore of a set of up to 32 ligands in a few minutes. Alternatively, structure-based approaches require a ligand-bound structure and identify a potential pharmacophore by analyzing the interaction site. ZINCPharmer provides a mechanism for deriving an initial pharmacophore hypothesis directly from structures within the PDB (Protein Data Bank), and also supports importing pharmacophore definitions developed using more computationally demanding approaches implemented in third-party tools.

Given a library of explicit compound conformations, conformers that match a 3D pharmacophore can be found using either fingerprint-based or alignment-based approaches. Fingerprints are well suited for similarity metrics, but, since they discretize the pharmacophore representation, provide inexact results. The EDULISS online database provides fingerprint-based screening of a single-conformer library of a few million compounds, but the query fingerprint must be manually constructed from pairwise distance constraints. Alignment-based approaches produce more accurate and interpretable results, at the expense of more computation. For example, a library of fewer than a million conformers may take minutes or hours to screen. However, since there are substantially fewer protein targets than there are possible ligands, alignment-based pharmacophore screening can be used effectively when performing a reverse screen that identifies matching protein targets instead of ligands. PharmMapper takes as input a single ligand and screens a database of over 7000 receptors for potential targets.

Both fingerprint and alignment-based approaches typically evaluate every conformer in the library, resulting in search times that scale with the size of the database. Newer methods, such as Pharmer and Recore use indexing approaches so that search times scale with the complexity and breadth of the query, not the size of the library. ZINCPharmer uses the open-source Pharmer software to enable the interactive search of more than 176 million conformations in just a few minutes, if not seconds.

WORKING:

ZINCPharmer searches a database of conformations calculated from the purchasable compounds of the ZINC database. ZINC is a comprehensive collection of commercially available, biologically relevant compounds suitable for screening. Purchasable compounds have an expected availability of <10 weeks and are either available from vendor stock or are make-on-demand. The ZINCPharmer library is synchronized with the ZINC library on a monthly basis. Compounds are both added and removed to maintain consistency and ensure that only currently purchasable compounds are retained. ZINC compounds are converted into 3D conformations using omega2 from OpenEye Scientific Software. Conformers are generated using the default settings and -rms.7, which improves the sampling of conformational space compared to the default setting of .5. The 10 best conformers are saved.

The generated conformers are converted into an efficient search format using the Pharmer open-source software. Pharmer identifies hydrophobic, hydrogen bond donor/acceptor, positive/negative ions and aromatic pharmacophore features using the SMARTS matching functionality of the OpenBabel toolkit. Currently, the default set of SMARTS definitions is used, but these are subject to refinement based on user input. These features are stored in an efficient spatial index to support the rapid search of large chemical libraries.

The graphical user interface for defining, refining and visualizing pharmacophore queries and their results is implemented using JavaScript and the Java-based Jmol molecular viewer. A modern, standards compliant browser with a recent Java plugin is required. Session state, which includes the pharmacophore definitions, can be saved in a human-readable JSON (JavaScript Object Notation) format and the aligned search results can be saved in the sdf molecular format. An internet forum hosts a user guide and provides technical support.

DEFINING A PHARMACOPHORE QUERY

Using the Pharmer software, ZINCPharmer can automatically extract a set of pharmacophore features from molecular structure. Each feature consists of the feature type (hydrophobic, hydrogen bond donor/acceptor, positive/negative ion or aromatic), a position, and a search radius. Features may be derived from a single ligand structure, a protein–ligand structure, a protein–protein structure or from the output of third-party software.

REFINING A QUERY

Although ZINCPharmer is capable of automatically extracting a pharmacophore from an interaction, it is expected that the user will further refine the query to enhance its specificity and applicability. This can be done by editing the properties of the query or by applying filters to the results.

PHARMACOPHORE SEARCH

Having defined a pharmacophore, searching for matching purchasable compounds is as simple as clicking the ‘Submit Query’ button. Searches take anywhere from a few seconds to a few minutes. Queries with more features, queries with many hydrophobic features (which are the most common features), queries with large search tolerances and symmetric queries (which require the processing of many orientations per a matching conformer) will have longer search times. Results are returned and displayed in the results browser as they are found. An orientation of a conformer is only returned as a hit if all the matching features are within the specified search tolerances of the query when the conformer is aligned to minimize the weighted RMSD.

RESULTS VISUALIZATION

Each hit represents a unique orientation of a conformation to the query. For each hit, the ZINC identifier, RMSD to the query, molecular weight ('Mass'), and number of rotatable bonds ('RBnds') is shown. The ZINC identifier is a hyperlink that points to the corresponding compound web page in the ZINC database where purchasing information may be found. The results may be sorted by any of the numerical properties by clicking on the property heading in the results table. The complete set of oriented hits may be saved to an sdf file through the ‘Save Results’ button. The hits in this file are unordered and include the RMSD to the query as extra data attached to each molecule. This file is immediately useful as input to a secondary screening protocol such as ranking by energy minimization.

Individual hits are visualized with the query and a receptor (if present) by clicking on the corresponding row in the results browser. The viewer tab contains a wide assortment of colors and styles (wireframe, stick, spheres, etc.) for visualizing the results, the query ligand, the receptor residues and the receptor surface.

Different ways to view the receptor, ligand and result in the viewer panel:

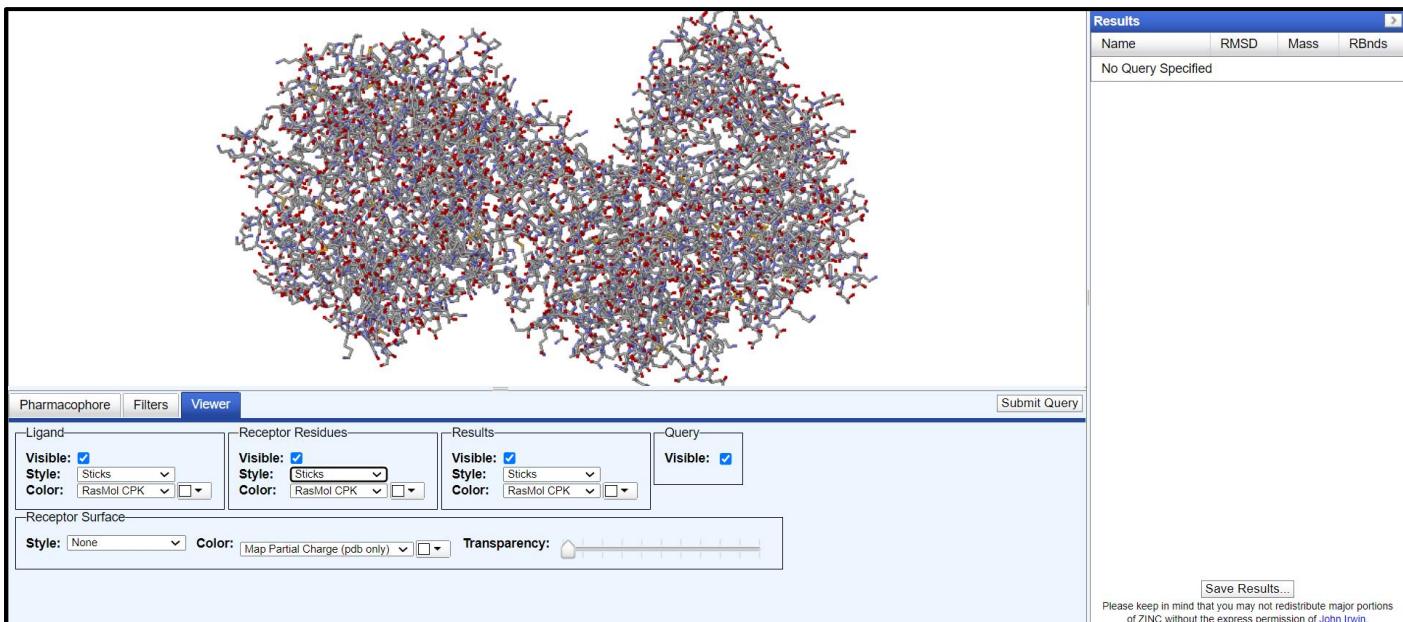


FIG 1. Visualisation in sticks style

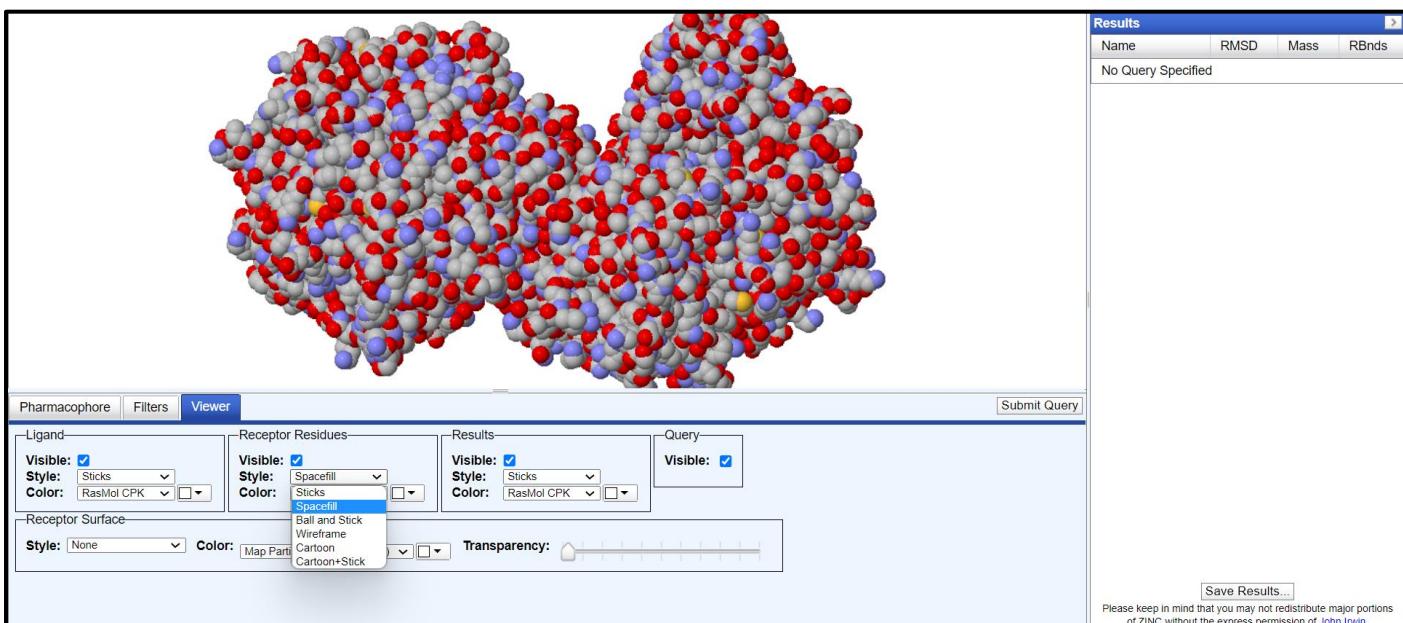


FIG 2. Visualisation in spacefill style

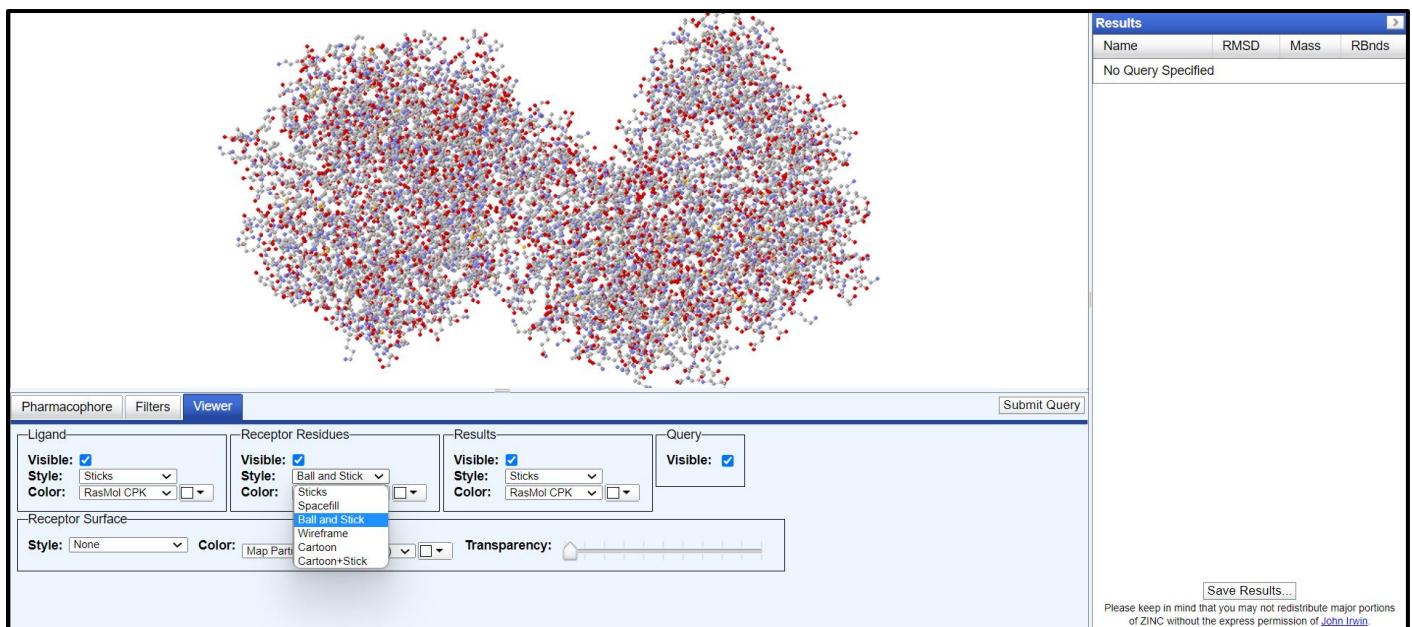


FIG 3. Visualisation in Ball and Stick style

REFERENCES:

1. Ijsrch, I. J. of S. R. in C. (2020). Pharmacophore Mapping and Virtual Screening. *International Journal of Scientific Research in Chemistry*. https://www.academia.edu/45098809/Pharmacophore_Mapping_and_Virtual_Screening
2. Liu, X., Ouyang, S., Yu, B., Liu, Y., Huang, K., Gong, J., Zheng, S., Li, Z., Li, H., & Jiang, H. (2010). PharmMapper server: a web server for potential drug target identification using pharmacophore mapping approach. *Nucleic Acids Research*, 38(suppl_2), W609–W614. <https://doi.org/10.1093/nar/gkq300>
3. Koes, D. R., & Camacho, C. J. (2012). ZINCPharmer: pharmacophore search of the ZINC database. *Nucleic Acids Research*, 40(W1), W409–W414. <https://doi.org/10.1093/nar/gks378>

WEBLEM 5a**PharmMapper**(URL: <http://www.lilab-ecust.cn/pharmmapper/>)**AIM:**

To generate and analyze Pharmacophore map for query Ampicillin (PubChem id: 6249) using PharmMapper web server.

INTRODUCTION:

Ampicillin is a semisynthetic penicillin derivative. A beta-lactam antibiotic, ampicillin is active against gram-positive cocci, including nonpenicillin resistant streptococcal, staphylococcal, and enterococcal species. It displays activity against some gram-negative organisms, gram-positive anaerobic organisms, and gram-negative anaerobic organisms. Ampicillin also has activity against certain spirochetes. Sulbactam, a beta-lactamase inhibitor, is administered with ampicillin to extend its spectrum of activity against penicillinase producing bacteria. Ampicillin is used for the treatment of upper and lower respiratory infections, skin and skin structure infections, urinary tract infections, and otitis media.

PharmMapper Server is a freely accessed web-server designed to identify potential target candidates for the given probe small molecules (drugs, natural products, or other newly discovered compounds with binding targets unidentified) using pharmacophore mapping approach. Benefited from the highly efficient and robust mapping method, PharmMapper bears high throughput ability and can identify the potential target candidates from the database within a few hours.

PharmMapper is backed up by a large, in-house repertoire of pharmacophore database extracted from all the targets in TargetBank, DrugBank, BindingDB and PDTD. Over 7,000 receptor-based pharmacophore models (covering 1,627 drug targets information, 459 of which are human protein targets) are stored and accessed by PharmMapper. PharmMapper finds the best mapping poses of the user uploaded molecules (in Tripos Mol2 or MDL SDF format) against all the targets in PharmTargetDB and top N potential drug targets as well as respective molecule's aligned poses are outputted.

METHODOLOGY:

- Retrieve 2D structure of Ampicillin in .sdf format from Pubchem database (URL: <https://pubchem.ncbi.nlm.nih.gov/>)
- Open homepage of PharmMapper. (URL: <http://www.lilab-ecust.cn/pharmmapper/>)
- Go to the submit job tab.
- Submit the Ampicillin structure retrieved from Pubchem database.
- Copy the job id and paste in the get result tab.
- Observe and interpret the results.

OBSERVATIONS:

National Library of Medicine
National Center for Biotechnology Information

PubChem About Posts Submit Contact Search PubChem

COMPOUND SUMMARY

Ampicillin

PubChem CID 6249

Structure 2D 3D

Find Similar Structures

Cite Download

CONTENTS

Title and Summary

1 Structures

2 Names and Identifiers

3 Chemical and Physical Properties

4 Spectral Information

5 Related Records

6 Chemical Vendors

7 Drug and Medication

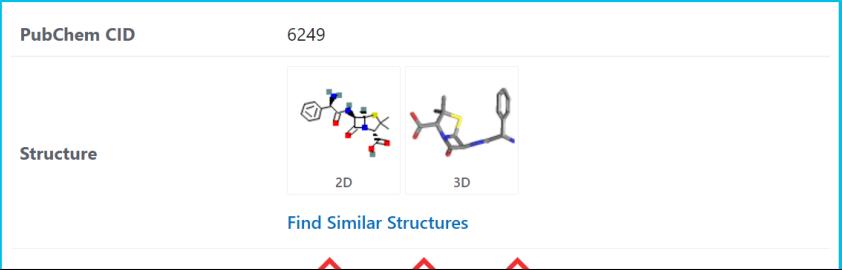


Fig1. Pubchem database page for Ampicillin

PharmMapper



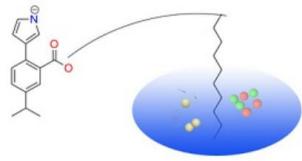
PharmMapper Introduction Submit Job Check Job Get Result Help Doc

PharmMapper Server: An updated integrated pharmacophore matching platform with statistical method for potential target identification.

Drug target identification, which includes many distinct algorithms for finding genes and proteins, is the first step in drug discovery. When 3D structures of the targets are available, the problem of target identification is usually converted to finding the best interaction mode between the potential target candidates and probe small molecules. Pharmacophore, which is the spatial arrangement of features that is essential for a molecule to interact with a specific target receptor, is an alternative method despite of molecular docking for achieving this goal.

PharmMapper Server is a freely accessed web-server designed to identify potential target candidates for the given probe

FIG 2. Homepage of PharmMapper server



PharmMapper Introduction **Submit Job** Check Job Get Result Help Doc

Step 1: Specify molecule file to perform calculation

Upload Query File

Ampicillin.sdf

Browse

We DO NOT support sdf V3000 format file.

Email Address

lizapatel80@gmail.com

We will send you an email when your job finished.

Job Description

[- Optional -]

FIG 3. Submission of Ampicillin structure on server

PharmMapper Introduction **Submit Job** Check Job Get Result Help Doc

Step 2: Parameter set

A. Conformation Generation

Generate Conformers

Yes No

Maximum Generated Conformations

300

Advanced Options



B. Pharmacophore Mapping

Human Protein Targets Only (v2010, 2241) [?](#)

All Targets (v2010, 7302) [?](#)

Druggable Pharmacophore Models (v2017, 16159) [?](#)

Pharmacophore Models whose $pK_d \geq 6.0$ (v2017, 52431) [?](#)

Select Targets Set

300

Number of Reserved Matched Targets (Max 1,000)

FIG 4. Parameters for the query submission

PharmMapper Introduction Submit Job Check Job Get Result Help Doc

Advanced Options	
- Perform GA Match	<input checked="" type="radio"/> Yes <input type="radio"/> No
- Include Exclude Volume	<input checked="" type="radio"/> Yes <input type="radio"/> No
- Fit Value Cutoff	2.0
- Vector Angle Cutoff	90
- Weight of Hydrophobic Score	1.0
- Weight of HBA Score	1.0
- Weight of HBD Score	1.0
- Weight of Positive Score	1.0
- Weight of Negative Score	1.0
- Weight of Aromatic Ring Score	1.0

FIG 5. Advanced options to set parameters

PharmMapper



PharmMapper Introduction Submit Job Check Job **Get Result** Help Doc

Get Job Result

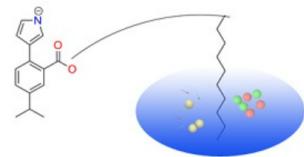
Job ID

Please input the job ID given after you submitted the job.

Get Result

FIG 6. Submission of job ID to get the results

PharmMapper



[PharmMapper](#) [Introduction](#) [Submit Job](#) [Check Job](#) [Get Result](#) [Help Doc](#)

Result of 221003150003

Top 300 targets ranked by normalized fit score in descending order

Ampicillin.sdf - Pharmacophore map for Ampicillin

Ligand: 6249

Rank	PDB ID	Target Name	Number of Features ↑	Fit Score ↑	Normalized Fit Score ↑
+ 1	3IH9	NONE	5	3.873	0.7745

Waiting for www.google-analytics.com...

FIG 7. Targets obtained for the query hit

PharmMapper [Introduction](#) [Submit Job](#) [Check Job](#) [Get Result](#) [Help Doc](#)

Rank	PDB ID	Target Name	Number of Features ↑	Fit Score ↑	Normalized Fit Score ↑
+ 1	3IH9	NONE	5	3.873	0.7745
- 2	2D9H	Zinc finger protein 692	5	3.742	0.7485

Pharmacophore Color Scheme

Hydrophobic	Positive	Negative	Donor	Acceptor	Aromatic
3	0	0	1	1	0

Molecule and Pharmacophore Model [Download Pharmacophore](#) [Download Aligned Ligand](#)

Molecule and Molecular Features [Pharmacophore Model and Molecular Features](#)

Reset to original view

* The visualization of hypoedit format pharmacophore models is implemented with [NGL](#).

FIG 8. Result for Zinc finger protein 692 protein target

RESULTS:

To generate and analyze Pharmacophore map for query Ampicillin (PubChem id: 6249) PharmMapper web server was used. 300 targets ranked with normalized fit score in descending order were retrieved. Molecule and pharmacophore model was observed for Zinc finger protein 962 with fit score of 3.742 and 5 matching features with query was observed.

CONCLUSION:

PharmMapper Server can be used identify potential target candidates for the given probe small molecules (drugs, natural products, or other newly discovered compounds with binding targets unidentified) using pharmacophore mapping approach. Benefited from the highly efficient and robust mapping method, PharmMapper bears high throughput ability and can identify the potential target candidates from the database within a few hours.

REFERENCES:

1. *Ampicillin - an overview / ScienceDirect Topics.* (n.d.). Www.sciencedirect.com. Retrieved October 8, 2022, from <https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/ampicillin>
2. Liu, X., Ouyang, S., Yu, B., Liu, Y., Huang, K., Gong, J., Zheng, S., Li, Z., Li, H., & Jiang, H. (2010). PharmMapper server: a web server for potential drug target identification using pharmacophore mapping approach. *Nucleic Acids Research*, 38(suppl_2), W609–W614. <https://doi.org/10.1093/nar/gkq300>
3. *PharmMapper.* (n.d.). Www.lilab-Ecust.cn. Retrieved October 8, 2022, from <http://www.lilab-ecust.cn/pharmmapper/>

WEBLEM 5b**ZINCPharmer**(URL: <http://zincpharmer.csb.pitt.edu/>)**AIM:**

To perform virtual screening based on pharmacophore features for Dactinomycin (PubChem Id- 457193) using ZincPharmer web server.

INTRODUCTION:

Dactinomycin is used in combination with other medications, surgery, and/or radiation therapy to treat Wilms' tumor (a type of kidney cancer that occurs in children) and rhabdomyosarcoma (cancer that forms in muscles) in children. Dactinomycin is also used in combination with other medications to treat certain types of testicular cancer and Ewing's sarcoma (a type of cancer in bones or muscles). Dactinomycin is also used alone or in combination with other medications to treat gestational trophoblastic tumors (a type of tumor that forms inside a woman's uterus while she is pregnant). Dactinomycin may also be used to treat certain types of cancerous tumors that are located in a specific area of the body. Dactinomycin is a type of antibiotic that is only used in cancer chemotherapy. It works by slowing or stopping the growth of cancer cells in your body.

ZINCPharmer is free pharmacophore search software for screening the purchasable subset of the ZINC database (updates occur monthly). ZINCPharmer can import LigandScout and MOE pharmacophore definitions, as well as identify pharmacophore features directly from structure.

METHODOLOGY:

- Retrieve structure of protein 6VSB from PDB database in .pdb format (URL: <https://www.rcsb.org/>)
- Retrieve 2D of Dactinomycin in .sdf format from Pubchem database (URL: <https://pubchem.ncbi.nlm.nih.gov/>)
- Open homepage of ZINCPharmer. (URL: <http://zincpharmer.csb.pitt.edu/>)
- Click on “Search ZINC”
- Load the receptor i.e the protein structure.
- Load the features i.e the Dactinomycin structure
- Enable the features and apply filters
- Click on “Submit Query”
- Observe and interpret the results.

OBSERVATIONS:

The screenshot shows the RCSB PDB homepage. At the top, there are links for Deposit, Search, Visualize, Analyze, Download, Learn, More, Documentation, and Careers. On the right, there are buttons for MyPDB and Contact us. The main header features the RSCB PDB logo and statistics: 196,108 Structures from the PDB and 1,000,361 Computed Structure Models (CSM). A search bar allows users to enter search terms, entry IDs, or sequences, with an option to include CSM. Below the search bar are links for Advanced Search and Browse Annotations, along with a Help button. The navigation bar includes links for PDB-101, wwPDB, EMDDataResource, NUCLEIC ACID DATABASE, and www.PDB Foundation. The main content area displays the structure of Biological Assembly 1 (6VSB), which is a Prefusion 2019-nCoV spike glycoprotein with a single receptor-binding domain up. The structure is shown as a 3D ribbon model. To the right of the structure, there is a summary box containing the PDB ID (6VSB), classification (VIRAL PROTEIN), organism (Severe acute respiratory syndrome coronavirus 2), expression system (Homo sapiens), and mutation information. It also lists the deposition date (2020-02-10), release date (2020-02-26), and deposition author(s) (Wrapp, D., Wang, N., Corbett, K.S., Goldsmith, J.A., Hsieh, C., Abiona, O., Graham, B.S.).

FIG 1. Structure of 6VSB protein from PDB database

The screenshot shows the PubChem website. The top navigation bar includes the NIH National Library of Medicine logo, the PubChem logo, and links for About, Posts, Submit, and Contact. A search bar on the right is labeled "Search PubChem". The main content area is titled "COMPOUND SUMMARY" and features the compound "Dactinomycin". The summary card for Dactinomycin shows its PubChem CID (457193) and a 2D chemical structure. Below the structure is a link to "Find Similar Structures". To the right of the summary card is a "CONTENTS" sidebar with sections for Title and Summary, 1 Structures, 2 Names and Identifiers, 3 Chemical and Physical Properties, 4 Spectral Information, 5 Related Records, 6 Chemical Vendors, and 7 Publications.

FIG 2. Structure of Dactinomycin from PubChem database

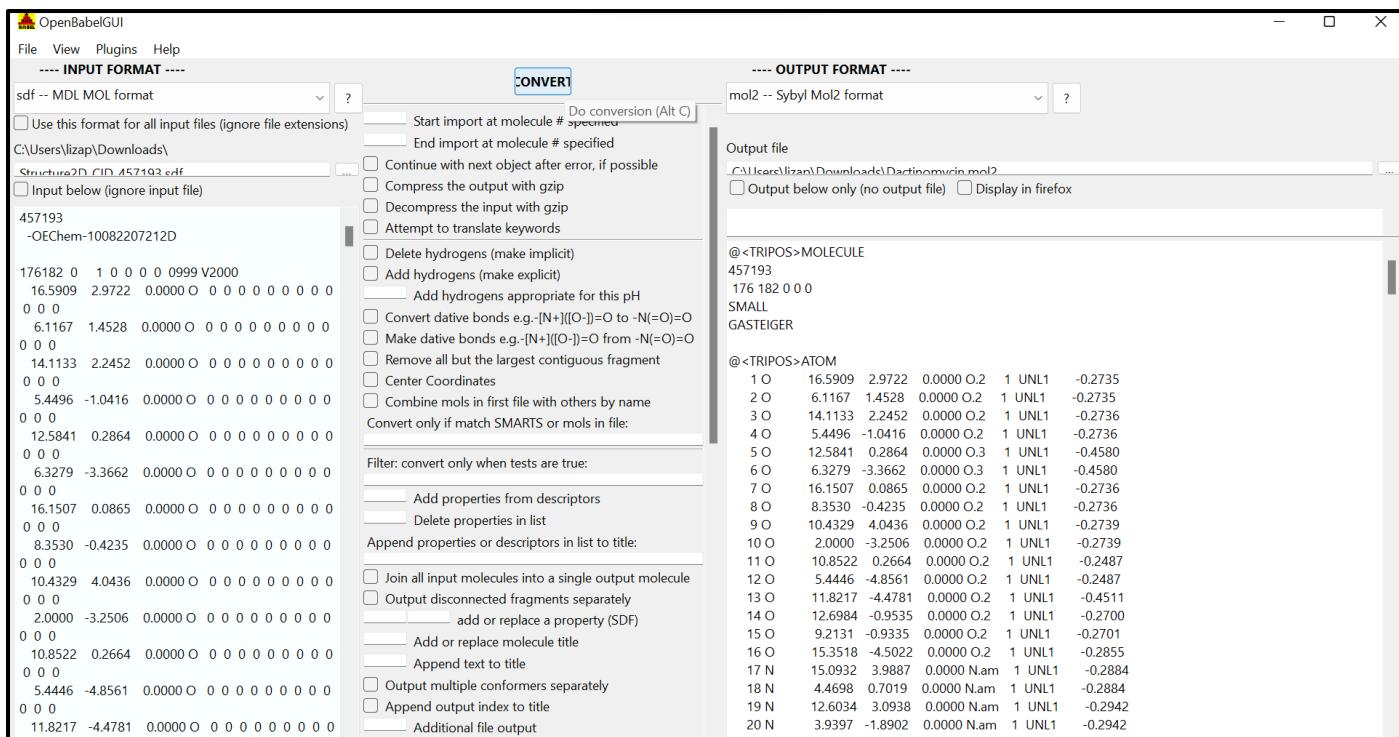


FIG 3. Conversion of Dactinomycin structure from .sdf file format to .mol2 file format using open babel tool

The screenshot shows the ZINCPharmer homepage. At the top, it says 'ZINCPharmer'. Below that, a text block explains: 'ZINCPharmer is free [pharmacophore](#) search software for screening the purchasable subset of the [ZINC](#) database (updates occur monthly). ZINCPharmer can import [LigandScout](#) and [MOE](#) pharmacophore definitions, as well as [identify pharmacophore features](#) directly from structure.' There are two main search buttons: 'Search ZINC' and 'Search Molport'. To the right, there's a 'From PDB structure' input field with 'PDB' and 'Ligand' dropdowns, and a 'Start' button. Below the search buttons, it says 'Java Problems? Try [HTML5](#)'. Further down, there are links for 'User Guide and Forum', 'Video Tutorials: [Defining Pharmacophores Searching and Filtering](#)', and a box for 'Targeting a protein-protein interaction? Design a pharmacophore from the PPI with [PocketQuery](#)'. At the bottom, there are 'Like 27' and 'Share' buttons, and a link for 'Interactive Examples'.

FIG 4. Homepage of ZINCPharmer server

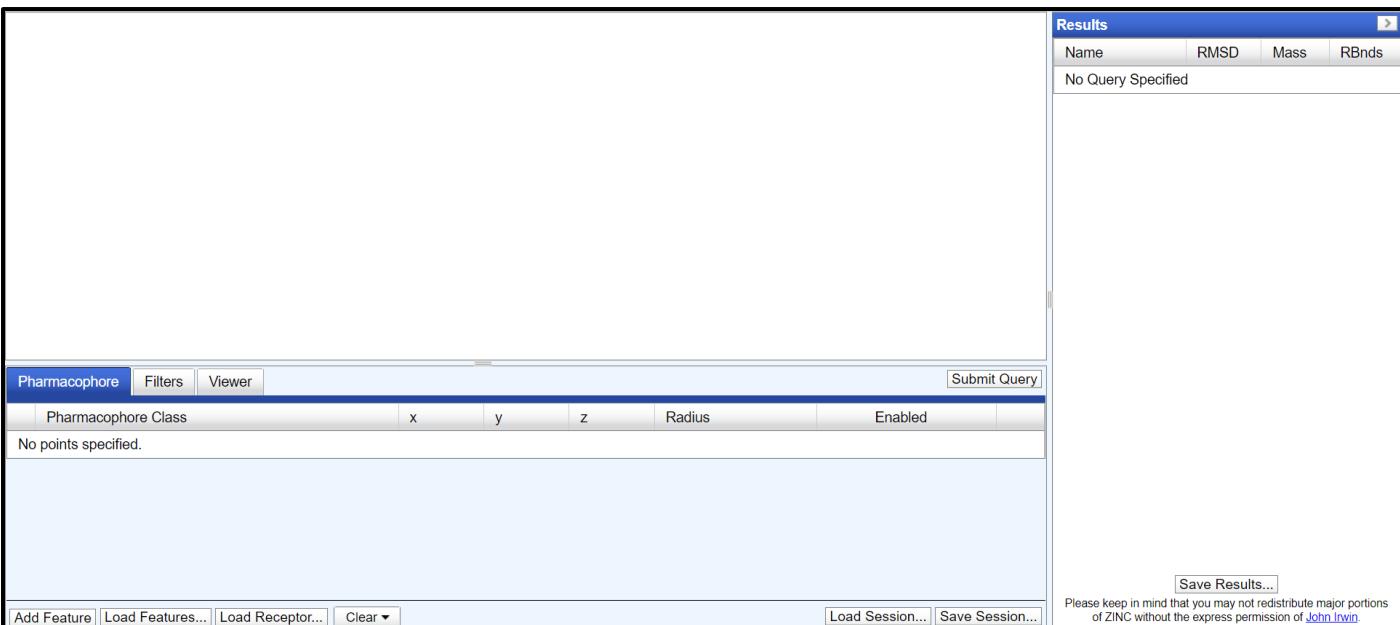


FIG 5. Panel for ZINC Search

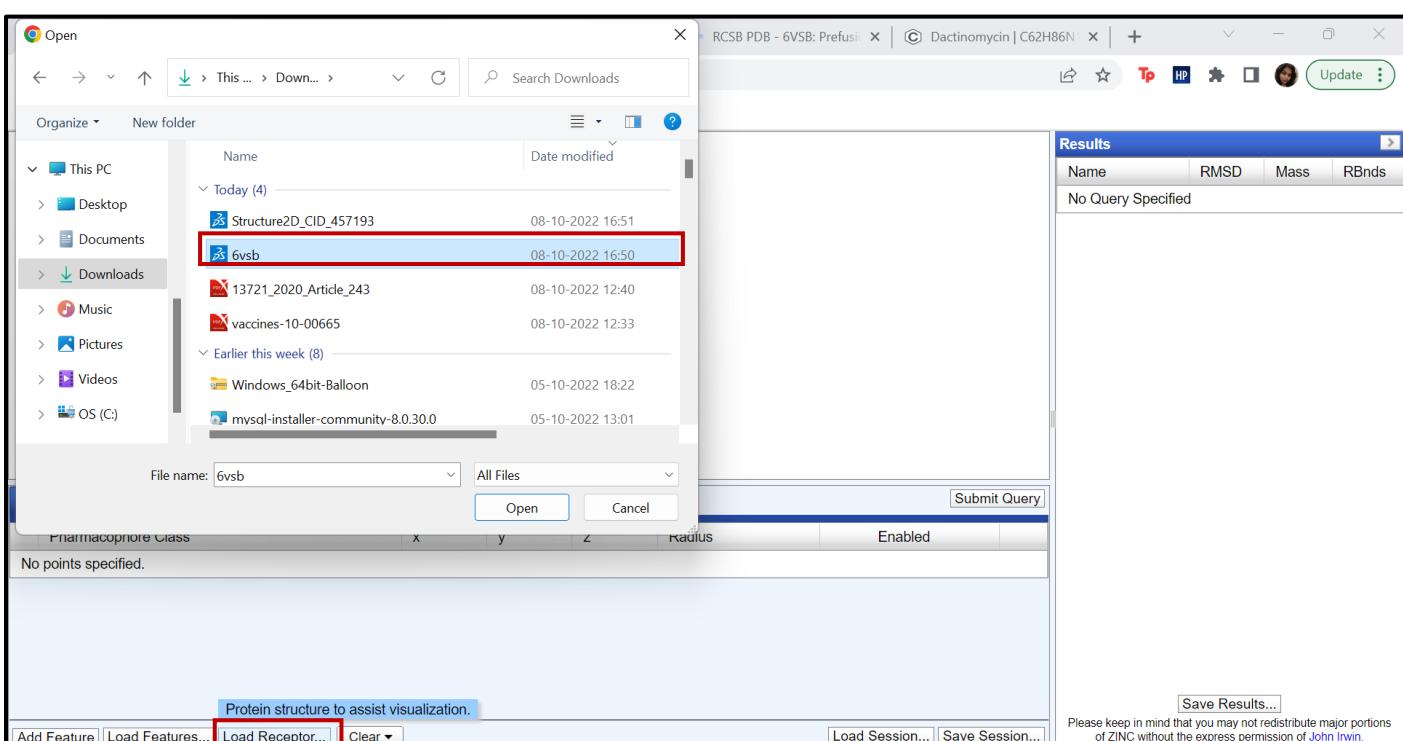


FIG 6. Loading receptor

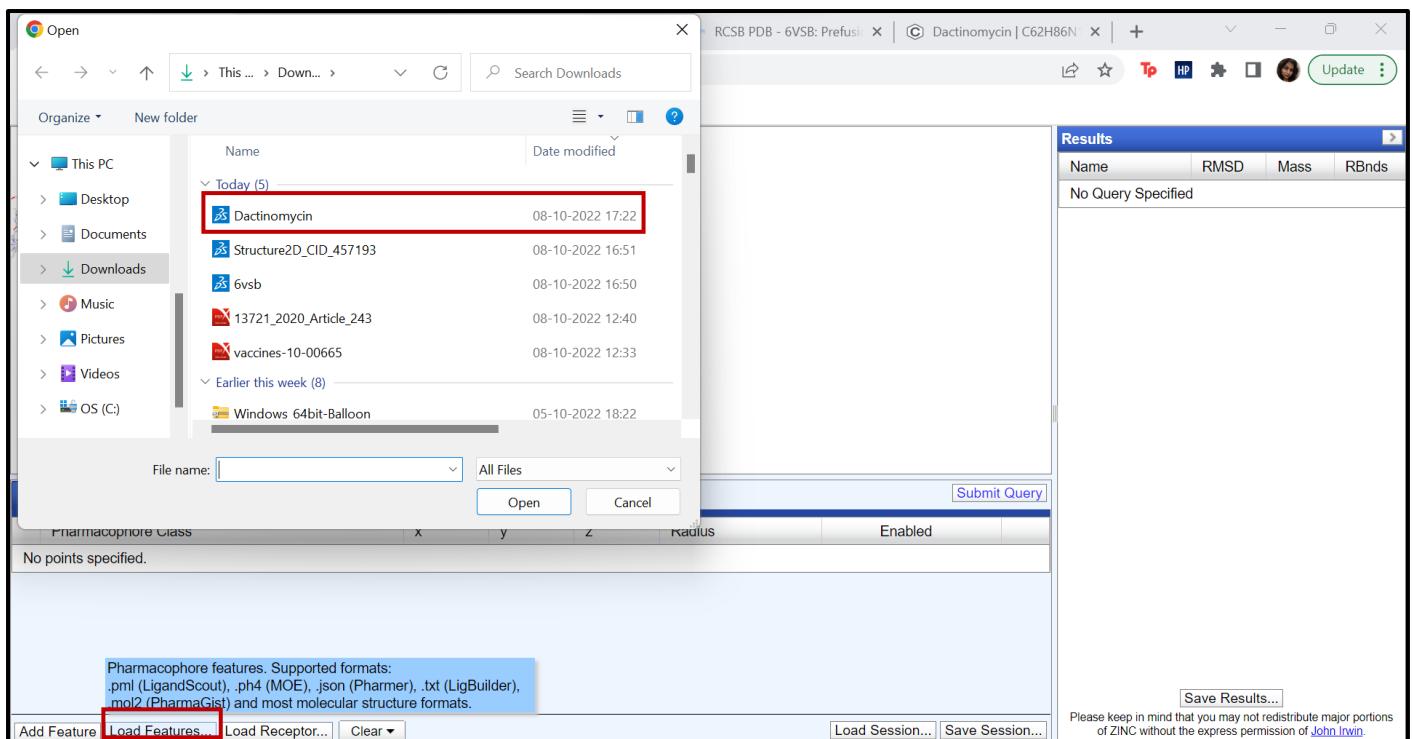


FIG 7. Loading features for query

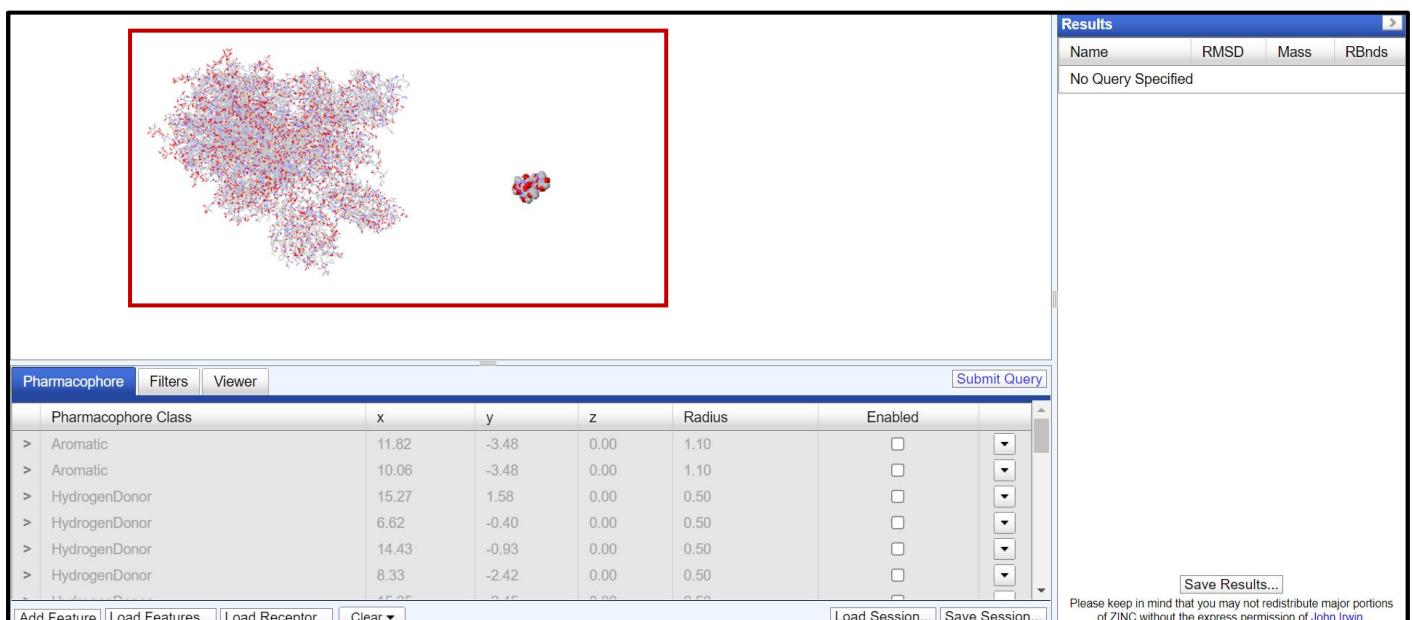


FIG 8. Loaded receptor and features for analysis

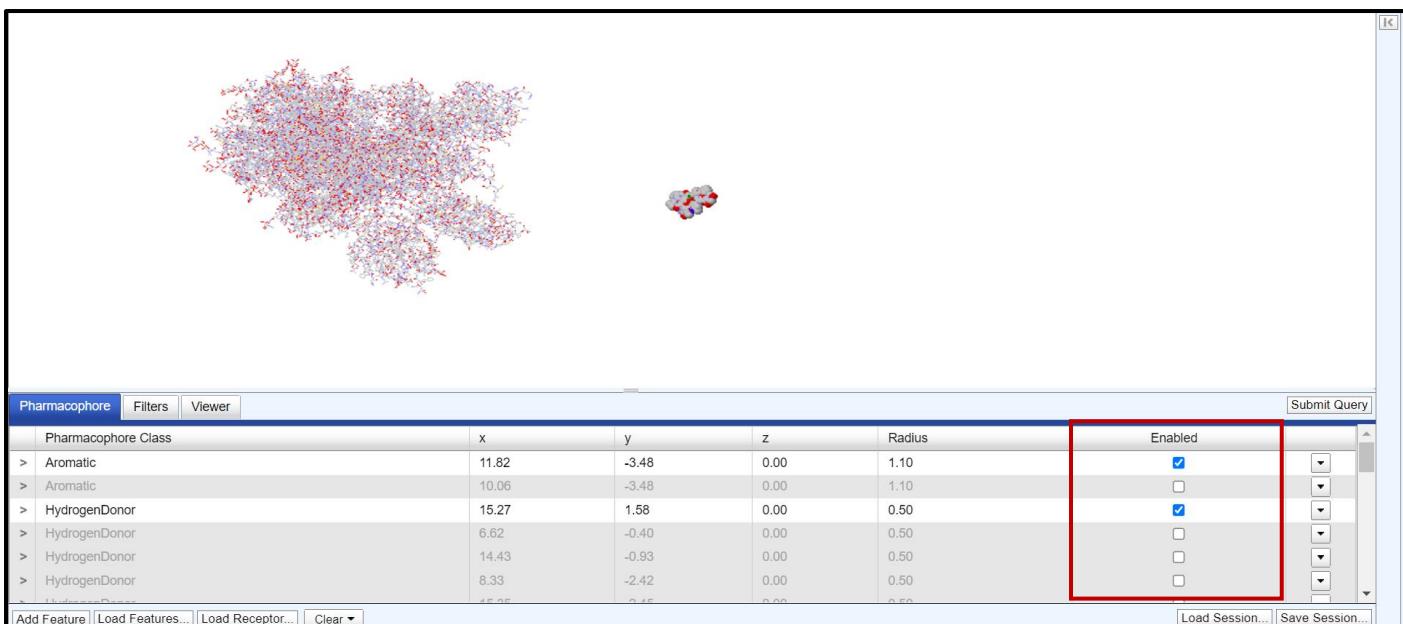


FIG 9. Enabling features for the search

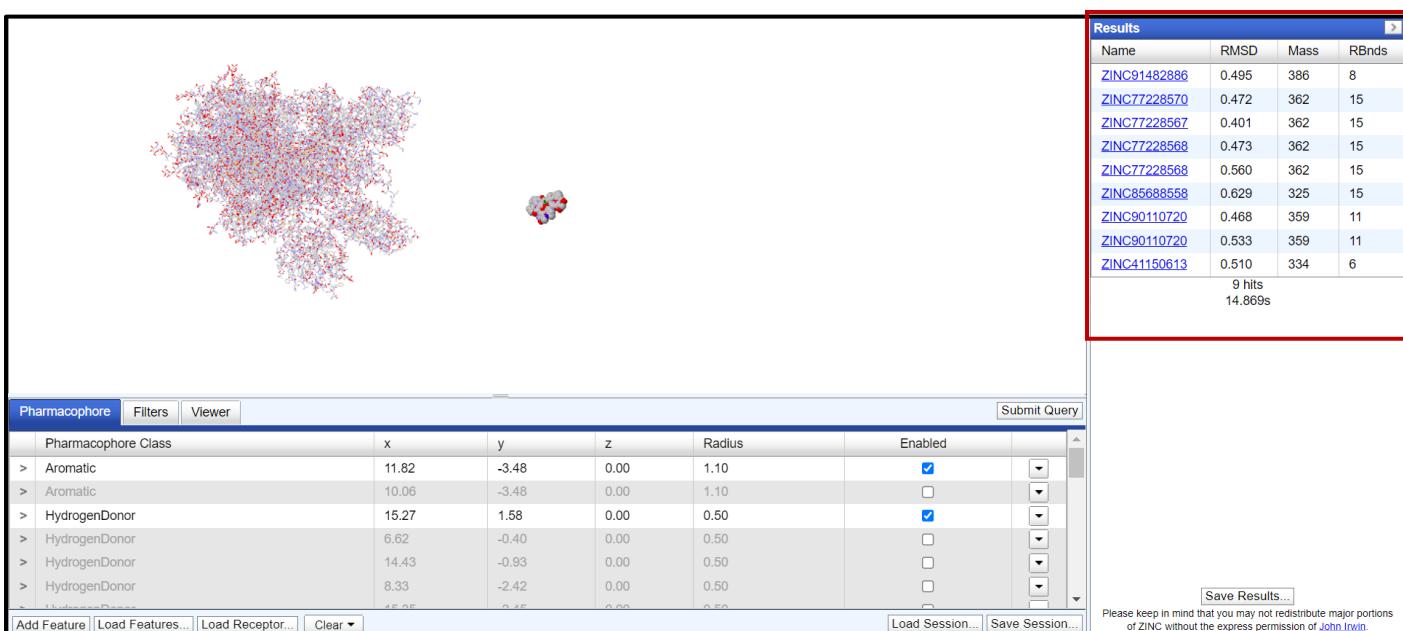


FIG 10. Search results

Results

Name	RMSD	Mass	RBnds
ZINC91482886	0.495	386	8
ZINC77228570	0.472	362	15
ZINC77228567	0.401	362	15
ZINC77228568	0.473	362	15
ZINC77228569	0.560	362	15
ZINC85688558	0.629	325	15
ZINC90110720	0.468	359	11
ZINC90110720	0.533	359	11
ZINC41150613	0.510	334	6

9 hits
14.869s

Pharmacophore **Filters** **Viewer** **Submit Query**

Hit Reduction

Hit Screening

Subset Selection

Save Results...

Please keep in mind that you may not redistribute major portions of ZINC without the express permission of [John Irwin](#)

FIG 11. Applying filters to refine the search

Results

Name	RMSD	Mass	RBnds
ZINC90110720	0.468	359	11
ZINC90110720	0.533	359	11
ZINC41150613	0.510	334	6
ZINC91482886	0.495	386	8

4 hits
5.132s

Pharmacophore **Filters** **Viewer** **Submit Query**

Hit Reduction

Hit Screening

Subset Selection

Save Results...

Please keep in mind that you may not redistribute major portions of ZINC without the express permission of [John Irwin](#)

FIG 12. Search results after refinement

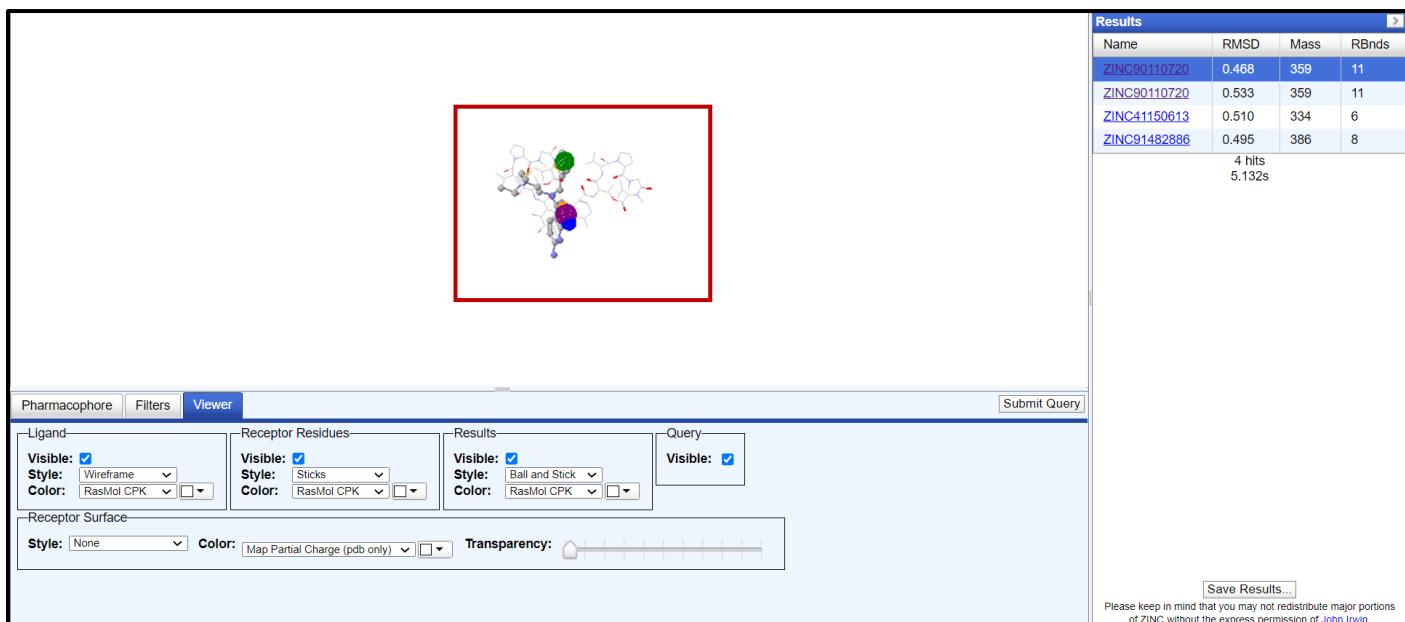


FIG 13. ZINC conformer mapped over ligand “Dactinomycin”

RESULTS:

Virtual screening based on pharmacophore features for Dactinomycin (PubChem Id- 457193) was performed using ZincPharmer web server. After submitting the query 9 hits were obtained and after applying filters for molecular weight and rotatable bonds 4 hits were obtained based on provided pharmacophore features.

CONCLUSION:

A pharmacophore describes the structural arrangement of the essential molecular features of an interaction between a ligand and its receptor. Searching chemical libraries for compounds that match a specific pharmacophore is an established method of virtual screening. The two main challenges of pharmacophore-based virtual screening are identifying a representative pharmacophore for an interaction and then identifying the compounds within a relevant chemical library that match the pharmacophore. ZINCPharmer is a pharmacophore search engine for purchasable chemical space that addresses both these challenges. The ultimate goal here is the discovery of novel compounds which exhibit a set of desired pharmacophoric features that are considered crucial for biological activity towards a particular target of interest. Hit-lists made up of molecules belonging to different structural classes can serve as a valuable source of “ideas” for the development and optimization of novel lead compounds that might not have been discovered by a traditional rational drug design processes alone.

REFERENCES:

1. Welcome to ZINCPharmer. (n.d.). Zincpharmer.csb.pitt.edu. <http://zincpharmer.csb.pitt.edu/>
2. Seidel, T., Wieder, O., Garon, A., & Langer, T. (2020). Applications of the Pharmacophore Concept in Natural Product inspired Drug Design. *Molecular Informatics*, 39(11), 2000059. <https://doi.org/10.1002/minf.202000059>
3. Koes, D. R., & Camacho, C. J. (2012). ZINCPharmer: pharmacophore search of the ZINC database. *Nucleic Acids Research*, 40(W1), W409–W414. <https://doi.org/10.1093/nar/gks378>

Weblem 6: SwissADME

(<http://www.swissadme.ch>)

Aim:

Introduction to SwissADME, a free web server tool

Introduction:

A large number of molecular structures are evaluated based on a wide range of criteria during the time- and resource-intensive processes of drug discovery and development in order to guide the choice of which chemicals to synthesise, test, and promote, with the ultimate goal of identifying those with the best potential to become a medicine that works for patients. The compounds must exhibit both a high level of biological activity and low toxicity. Equally crucial is the availability to and concentration at the therapeutic target in the body. The conventional approach to thinking about pharmacokinetics (i.e., what happens to a medicinal chemical in the body) is to separate the numerous effects that affect the target's access into separate parameters.

It has been shown that early ADME calculation during the discovery phase significantly lowers the percentage of clinical failures attributable to pharmacokinetics. In the early stages, when there are many researched chemical structures but few available compounds, computer models have been promoted as a viable alternative to experimental approaches for the prediction of ADME. The goal of many in-silico techniques is to predict ADME parameters from molecular structure. It is noteworthy that Lipinski et al ground-breaking research looked at orally active chemicals to identify physicochemical ranges with a high likelihood of being an oral medication. The link between pharmacokinetic and physicochemical characteristics was defined by the so-called Rule-of-Five.

Substructure searches can directly describe molecules whereas physicochemical characteristics provide a more general description of the structure. Structural Alert, PAINS, and Lilly MedChem filters, which are used to rid chemical libraries of compounds that are expected to be unstable, reactive, poisonous, or likely to interfere with biological tests due to generalised frequent hits, dyes, or aggregators, are all based on these approaches. The SwissADME web tool described here is freely available at <http://www.swissadme.ch> and designed for easy submission and result analysis, even for those unfamiliar with CADD. SwissADME's strong points include, non-exhaustively: many input ways, computation for multiple compounds, and unique access to proficient methodologies when compared to the state-of-the-art of free web-based ADME and pharmacokinetics applications.

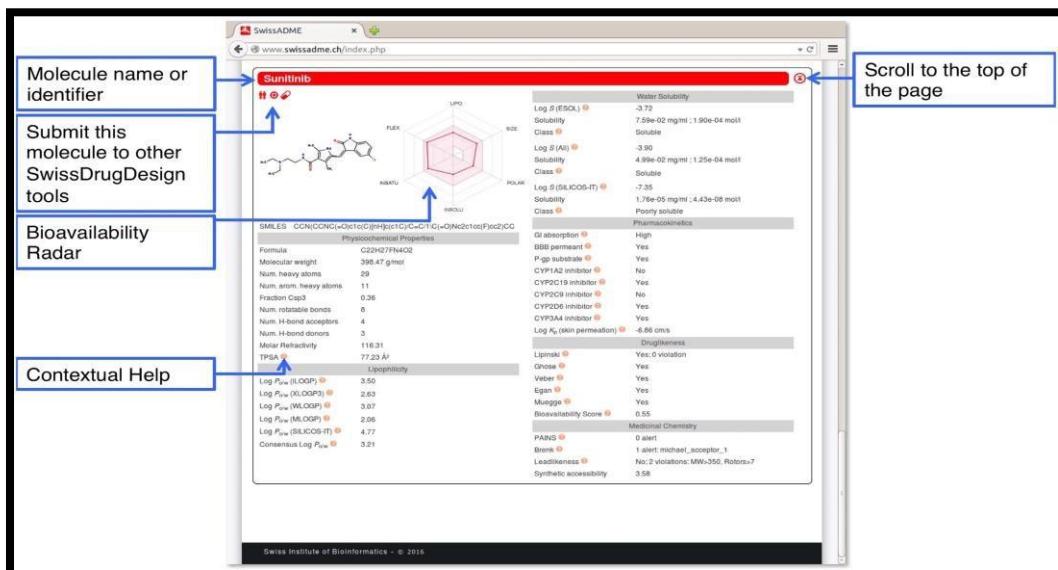
SIB Swiss Institute of Bioinformatics, including molecular docking (SwissDock), bioisosteric design (SwissBioisostere), ligand-based virtual screening (SwissSimilarity), and molecular mechanics.

Submission Web Page:

- Accessing <http://www.swissadme.ch> in a web browser displays directly the submission page of SwissADME, where molecules to be estimated for ADME, physicochemistry, drug-likeness, pharmacokinetics and medicinal chemistry friendliness properties can be input.
- A black toolbar at the top of the Webpage allows the user to navigate within the different SwissDrugDesign tools.
- A second bar gives access to different information regarding SwissADME, among which the FAQ and Help pages as well as legal disclaimer and contacts.
- The input zone itself comprises a molecular sketcher based on ChemAxon's Marvin JS that enables the user to import (from a file or an external database), draw, and edit a 2D chemical

structure, and to transfer it to a list of molecules.

- This list, on the right-hand side of the submission page, is the actual input for computation.
- It can be edited as a standard text, allowing for typing or pasting SMILES. The list is made to contain one input molecule per line, defined by SMILES and optionally a name separated by a space.
- If name is omitted, SwissADME will automatically provide an identifier.
- Noteworthy, both buttons for transferring the sketch to SMILES list and for running the computation are dynamic, in the sense that they are active only if the action is possible.
- At the time of writing, one can expect a result in 1 to 5 seconds for a drug-like molecule.
- Examples can be loaded in the SMILES list by clicking on the “Fill with an example” button.



One-panel-per-molecule Output

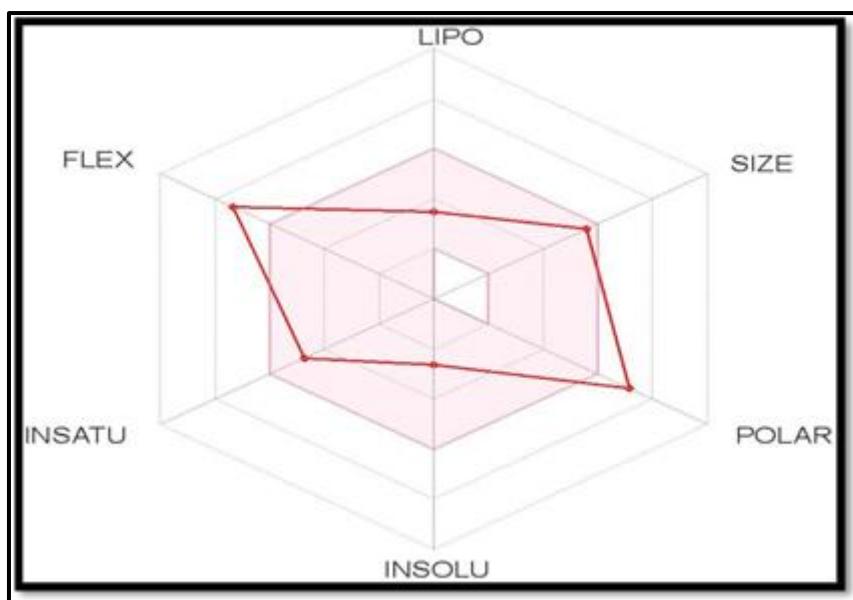
- The output panels are loaded in the same Web page.
- There is one panel compiling all values for each molecule. It is filled immediately after calculation completion, one molecule after the other.
- This way it is possible to inspect the results for the first compounds without waiting for the whole list to be treated.
- This one-panel-per-molecule is headed by the molecule name and divided into different sections.

Chemical Structure and Bioavailability Radar

- The first section, including two-dimensional chemical structure and canonical SMILES, is located below the title.
- It shows on which chemical form the predictions were calculated (refer to Computational Methods).
- Moreover, our Bioavailability Radar is displayed for a rapid appraisal of drug-likeness.
- Six physicochemical properties are taken into account: lipophilicity, size, polarity, solubility, flexibility and saturation
- A physicochemical range on each axis was defined by descriptors and depicted as a pink area in which the radar plot of the molecule has to fall entirely to be considered drug-like.
- Leaving the mouse over the radar gives further information about the descriptors (see also Physicochemical Properties and Computational Methods).

Physicochemical Properties:

This section compiles straightforward molecular and physical descriptors such molecular weight (MW), molecular refractivity (MR), count of particular atom kinds, and polar surface area (PSA). OpenBabel, version 2.3.0, is used to calculate the values. By treating sulphur and phosphorus as polar atoms, the topological polar surface area (TPSA), a fragmental method, is used to determine the PSA. This has demonstrated to be an effective descriptor in many models and rules to quickly estimate some ADME properties, particularly with regards to biological barrier crossing like absorption and brain access.



Lipophilicity:

- The partition coefficient between n-octanol and water ($\log P_{o/w}$) is the classical descriptor for Lipophilicity.
- It has a dedicated section in SwissADME due to the critical importance of this physicochemical property for pharmacokinetics drug discovery.
- Many computational methods for $\log P_{o/w}$ estimation were developed with diverse performance on various chemical sets.
- Common practice is to use multiple predictors either to select the most accurate methods for a given chemical series or to generate consensus estimation.
- The models behind the predictors should be as diverse as possible to increase the prediction accuracy through consensus $\log P_{o/w}$.
- The consensus $\log P_{o/w}$ is the arithmetic mean of the values predicted by the five proposed methods.

Water Solubility:

- Having a soluble molecule greatly facilitates many drug development activities, primarily the ease of handling and formulation.
- Moreover, for discovery projects targeting oral administration, solubility is one major property influencing absorption.
- As well, a drug meant for parental usage has to be highly soluble in water to deliver a sufficient quantity of active ingredient in the small volume of such pharmaceutical dosage.
- Two topological methods to predict Water Solubility are included in SwissADME.
- The first one is an implementation of the ESOL model³⁶ and the second one is adapted from Ali et al.

- Both differ from the seminal general solubility equation since they avoid the melting point parameter; the latter being challenging to predict.
- Moreover they demonstrate strong linear correlation between predicted and experimental values ($R^2 = 0.69$ and 0.81 , respectively).
- SwissADME third predictor for solubility was developed by SILICOS-IT.
- SwissADME also provides solubility in mol/l and mg/ml along with qualitative solubility classes.

Pharmacokinetics:

- Specialized models, whose predictions are compiled in the Pharmacokinetics section, evaluate individual ADME behaviors of the molecule under investigation.
- One model is a multiple linear regression, which aims at predicting the skin permeability coefficient (K_p).
- It is adapted from Potts and Guy, who found K_p linearly correlated with molecular size and lipophilicity ($R^2 = 0.67$).
- The more negative the log K_p (with K_p in cm/s), the less skin permeant is the molecule.
- The predictions for passive human gastrointestinal absorption (HIA) and blood-brain barrier (BBB) permeation both consist in the readout of the BOILED-Egg model, an intuitive graphical classification model, which can be displayed in the SwissADME result page by clicking the red button appearing below the sketcher when all input molecules have been processed.
- Other binary classification models are included, which focus on the propensity for a given small molecule to be substrate or inhibitor of proteins governing important pharmacokinetic behaviors.
- The knowledge about compounds being substrate or non-substrate of the permeability glycoprotein is key to appraise active efflux through biological membranes, for instance from the gastrointestinal wall to the lumen or from the brain.
- One major role of P-gp is to protect the central nervous system (CNS) from xenobiotics. Importantly as well, P-gp is over expressed in some tumor cells and leads to multidrug-resistant cancers.
- Also essential is the knowledge about interaction of molecules with cytochromes P450 (CYP).
- This super family of isoenzymes is a key player in drug elimination through metabolic biotransformation.
- It has been suggested that CYP and P-gp can process small molecules synergistically to improve protection of tissues and organisms.
- Inhibition of these isoenzymes is certainly one major cause of pharmacokinetics- related drug-drug interactions leading to toxic or other unwanted adverse effects due to the lower clearance and accumulation of the drug or its metabolites.
- SwissADME enables the estimation for a chemical to be substrate of P-gp or inhibitor of the most important CYP isoenzymes.
- We applied the support vector machine algorithm (SVM) on meticulously cleansed large datasets of known substrates/non-substrates or inhibitors/non-inhibitors.
- In similar contexts, SVM was found to perform better than other machine-learning algorithms for binary classification.
- The models return “Yes” or “No” if the molecule under investigation has higher probability to be substrate or non-substrate of P-gp (respectively inhibitor or non- inhibitor of a given CYP).
- SVM models rely merely on molecular and physicochemical descriptors generated by SwissADME.
- We believe that this improves robustness and sustainability of the underlying methodologies.

- In particular, not using molecular fingerprints, molecular graphs or other structural descriptions can be an handicap to generate high statistical values but should also limit over fitting biases and yield more generalist predictive models, not necessarily influenced by specific chemical scaffolds or moieties.
- In our practice, these well-performing models able to estimate important ADME behaviors are of great support for pharmacokinetics optimization and evaluation of small molecules.

Drug Likeness:

- As defined earlier, “drug-likeness” assesses qualitatively the chance for a molecule to become an oral drug with respect to bioavailability.
- Drug-likeness was established from structural or physicochemical inspections of development compounds advanced enough to be considered oral drug-candidates.
- This notion is routinely employed to perform filtering of chemical libraries to exclude molecules with properties most probably incompatible with an acceptable pharmacokinetics profile.
- This SwissADME section gives access to five different rule-based filters, with diverse ranges of properties inside of which the molecule is defined as drug-like.
- These filters often originate from analyses by major pharmaceutical companies aiming to improve the quality of their proprietary chemical collections.
- The Lipinski (Pfizer) filter is the pioneer rule-of-five implemented.
- The Ghose (Amgen), Veber (GSK), Egan (Pharmacia) and Muegge (Bayer) methods were adapted.
- Any violation of any rule described here appears explicitly in the output panel.
- This semi-quantitative rule-based score relying on total charge, TPSA, and violation to the Lipinski filter defines four classes of compounds with probabilities of 11%, 17%, 56% or 85%.4

Medicinal Chemistry:

- The purpose of this section is to support medicinal chemists in their daily drug discovery endeavors.
- Two complementary pattern recognition methods allow for identification of potentially problematic fragments.
- PAINS (for pan assay interference compounds, a.k.a. frequent hitters or promiscuous compounds) are molecules containing substructures showing potent response in assays irrespective of the protein target.
- Such fragments, yielding false positive biological output, have been identified by Baell et al. in analyzing six orthogonal assays and breaking down the molecules active on 2 or more assays into 481 recurrent fragments, considered as potentially leading to promiscuous compounds.
- SwissADME returns warnings if such moieties are found in the molecule under evaluation.
- In SwissADME, it is possible to have a chemical description of the problematic fragments found in a given molecule by flying over the “question mark” icon appearing after the fragment list.
- This is implemented for both PAINS and Brenk filters.
- By applying these and other physicochemical filters to design screening libraries, Brenk et al.5 observed that most of the remaining compounds satisfy criteria for “leadlikeness”.
- This concept is similar to drug-likeness, yet focusing on physicochemical boundaries defining a good lead, i.e. a molecular entity suitable for optimization.

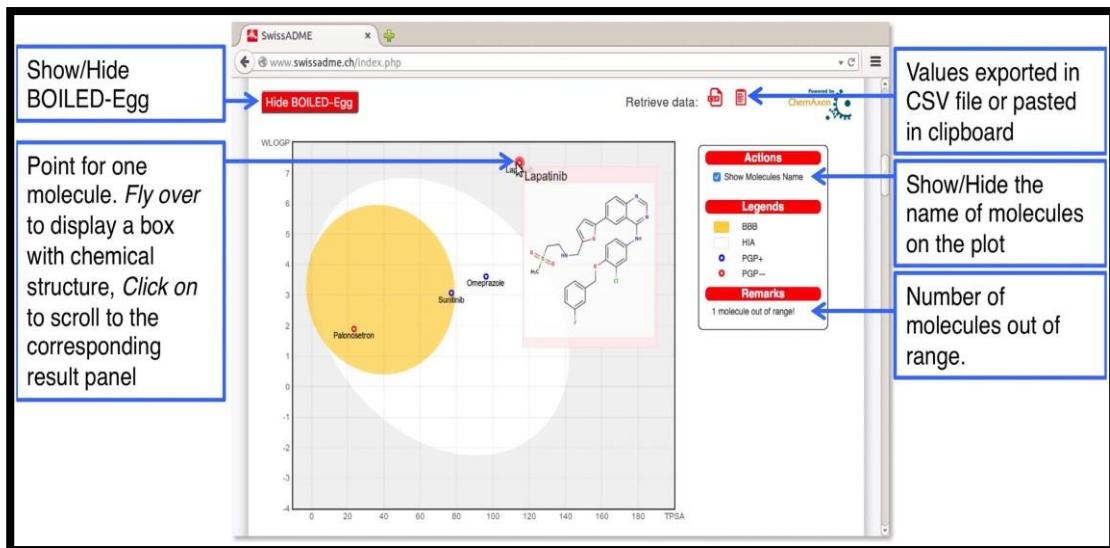
- By definition, leads are subjected to chemical modifications that will most likely increase size and lipophilicity.
- As a consequence, leads are required to be smaller and less hydrophobic than drug-like molecules.
- Synthetic accessibility (SA) is a major factor to consider in this selection process. Obviously, for a reasonable number of molecules, medicinal chemists are the best able to determine SA.
- However, when too many molecular structures prevent an expert evaluation, in silico estimation can be used for pre-filtering.
- Ertl & Schuffenhauer proposed a fingerprint-based approach for SA estimation but including closed-source information about fingerprint definition that prevents a straightforward implementation in our tool open to the scientific community.
- For a given molecule, the fragmental contributions to SA are summed and corrected by the terms describing size and complexity, such as macrocycles, chiral centers, or Spiro functions as defined by Ertl & Schuffenhauer.
- After normalization, the SA Score ranges from 1 (very easy) to 10 (very difficult).
- Human evaluation of synthetic complexity is undeniably subjective and relies on individual chemist's training and experience.
- However, significant linear correlation and small errors especially with SwissADME SA Score that outperformed the reference methods on both sets with smaller errors, and equal or higher linear correlation coefficients demonstrate how this simple and fast methodology can help prioritizing molecules to synthesize.

Graphical Output:

- After all calculations completed, the “Show BOILED-Egg” red button appears below the sketcher to display the graphical output on the same page.
- This consists primarily in the BOILED-Egg, an intuitive method to predict simultaneously two key ADME parameters, i.e. the passive gastrointestinal absorption (HIA) and brain access (BBB).
- Although conceptually very simple as it relies on two physicochemical descriptors only (WLOGP and TPSA, for lipophilicity and apparent polarity), this classification model was built with extreme care regarding statistical significance and robustness.
- The egg-shaped classification plot includes the yolk (i.e. the physicochemical space for highly probable BBB permeation) and the white (i.e. the physicochemical space for highly probable HIA absorption).
- Both compartments are not mutually exclusive and the outside grey region stands for molecules with properties implying predicted low absorption and limited brain penetration.
- In practice, the BOILED-Egg has proven straightforward interpretation and efficient translation to molecular design in a variety of drug discovery settings.
- Contrary to the one-panel-per-molecule concept for other parameters, the graphical output includes prediction for all molecules submitted to SwissADME, thus additional capacities were implemented to enable interactive navigation and easy evaluation.
- Flying over a specific point makes a semi-transparent box appear including the name and structure of the molecule.
- Clicking on a specific point makes the page scroll to the corresponding output panel including all predictions for the molecule.
- Getting back to the graphical output is achieved by clicking on the red up-arrow on the top-right corner of the panel.
- Moreover, on the right of the plot are displayed possible actions (at the moment, to show the name of

the molecules on the graph, only), legends and possible remarks (the number of molecules outside the range of the plot).

- The user may want to hide the plot by clicking the “Hide BOILED-Egg” red button.



Methodology:

- Using Google search engine search Swiss ADME.
- In Molecular Sketcher Add, Import or Draw a structure.
- Transfer the sketched structure to SMILES list.
- Run the calculations and observe the result page.
- Browse through various parts of this result page.

Observations:

Figure 1: SwissADME Homepage

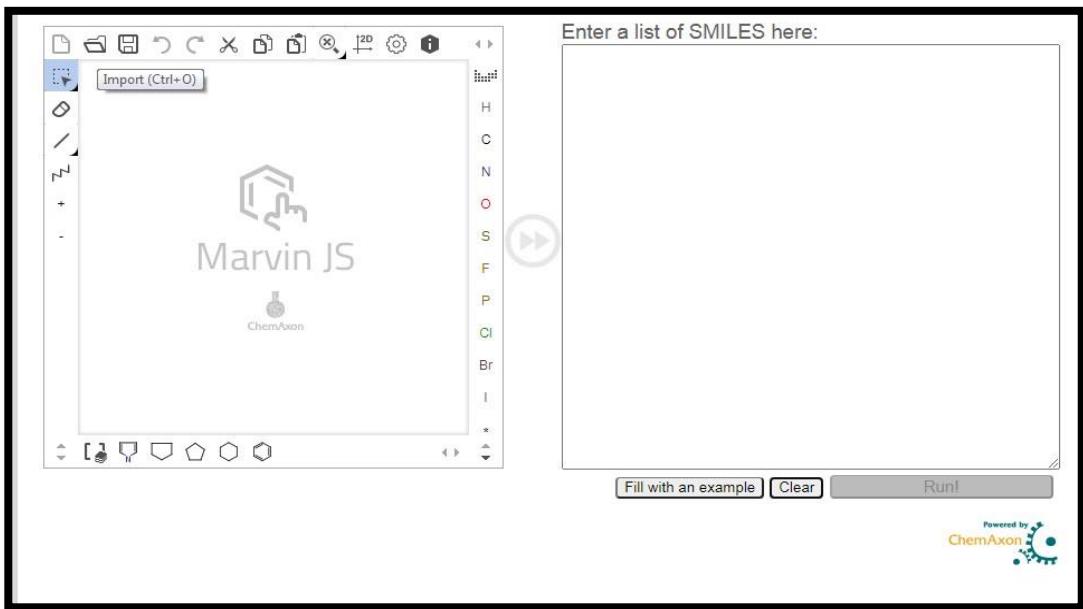


Figure 2: Molecular Sketcher and Smiles List



Figure 3: Smiles List filled with an example from the given option below

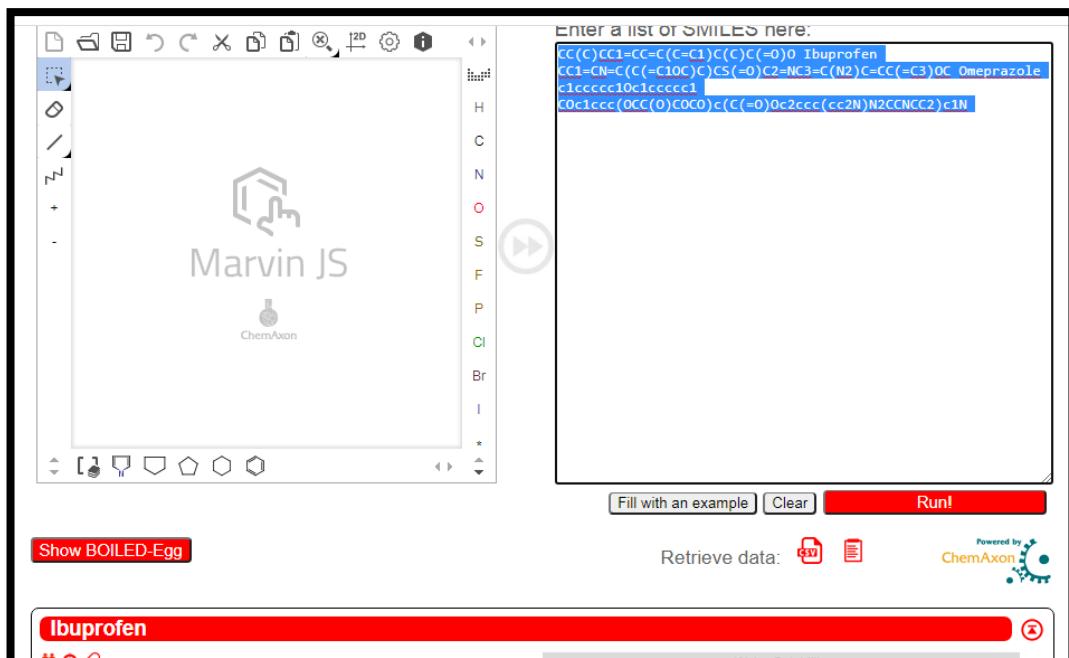


Figure 4: Run the example shown in Smiles Format

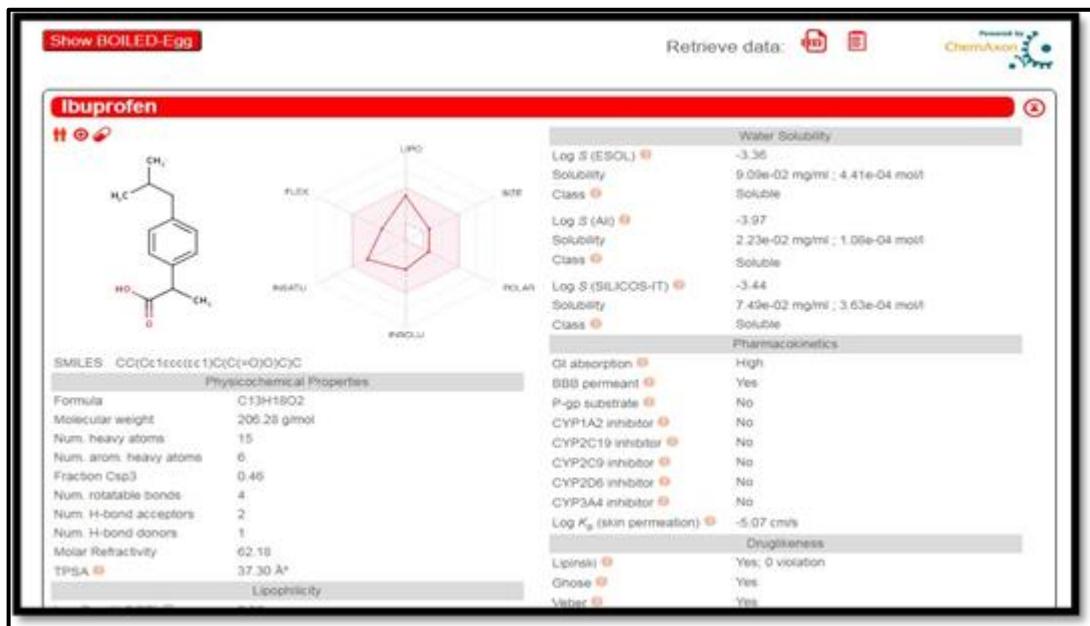


Figure 5: Result page of the given example

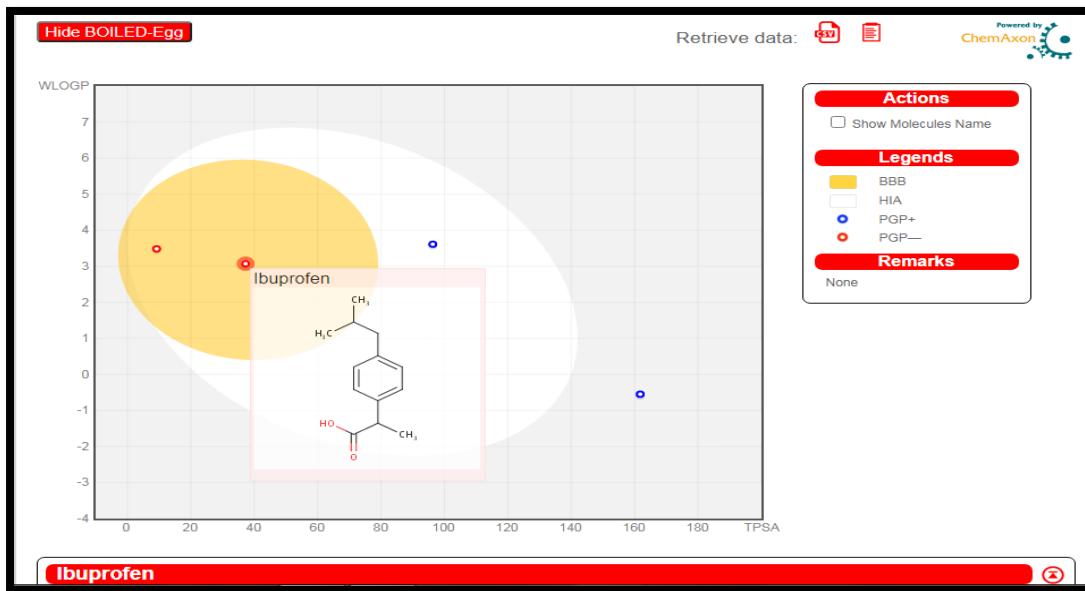


Figure 6: Graphical output of the example

The screenshot shows the SwissADME web interface. At the top, there is a navigation bar with links to Click2Drug, SwissDock, SwissParam, SwissSidechain, SwissBioscience, SwissTargetPrediction, SwissADME (highlighted in blue), SwissSimilarity, and About us. Below the navigation bar is the SIB logo and the text "Swiss Institute of Bioinformatics". The main content area is titled "SwissADME". It includes a "Help" section with information about the toolbar and molecular sketcher, and a "Input" section with instructions for entering molecules. A screenshot of the molecular sketcher interface is shown, with arrows pointing to the "SwissDrugDesign toolbar" and the "SwissADME toolbar: Home, FAQ, Help and Disclaimer".

Figure 7: Help page on SwissADME

Result:

SwissADME server is used for studying properties of chemical compounds or drugs their permeability in blood, gastrointestinal tract, etc. Results stated physicochemical characteristics lipophilicity, drug likeliness, pharmacokinetics and water solubility also it showed a graphical view of the compound.

Conclusion:

For one or more molecules, the SwissADME Web tool enables the computation of physicochemical, pharmacokinetic, drug-like, and associated characteristics. Free open-access models were integrated that demonstrated statistical significance, predictive capability, and simple molecular design application. These models were modified from well-known published methods. Through interactive capabilities, simple input and effective output analysis are made possible. Additionally, direct access was provided to additional SwissDrugDesign web tools including SwissSimilarity (virtual screening), SwissBioisostere (ligand-based design), and SwissTargetPrediction (prediction of biotargets). SwissADME was developed as a result to aid the entire community in the development of new drugs.

References:

1. SwissADME Homepage (<http://www.swissadme.ch>)
2. SwissADME help manual (<http://www.swissadme.ch/help.php>)
3. SwissADME literature (<https://www.nature.com/srep>)

Weblem 6a SwissADME

(<http://www.swissadme.ch>)

Aim:

To evaluate pharmacokinetics and drug likeness using SwissADME server for query (PubChem ID- 5904)

Introduction:

Acetylsalicylic acid (ASA), popularly known as aspirin, is a drug used to treat inflammation, fever, and pain. Aspirin is used to treat a variety of inflammatory disorders, including Kawasaki disease, pericarditis, and rheumatic fever. When taken soon after a heart attack, aspirin reduces the risk of dying. Long-term usage of aspirin is also used to help those at high risk avoid further heart attacks, ischemic strokes, and blood clots. Additionally, it might make some cancers less likely, especially colorectal cancer. Effects often start within 30 minutes for pain or fever. Aspirin is a nonsteroidal anti-inflammatory medication (NSAID) that functions in a manner similar to other NSAIDs while also inhibiting platelet function. An uncomfortable stomach is a typical side effect. The exacerbation of asthma as well as stomach bleeding and ulcers are more serious adverse effects. People who are older, drink alcohol, use other NSAIDs, or are taking other blood thinners have a higher risk of bleeding. Taking aspirin during the final trimester of pregnancy is not advised. Due to the possibility of developing Reye syndrome, it is generally not advised for children who have infections. Ringing in the ears has been linked to high doses. One of the most often used drugs in the world is aspirin, which is taken in between 50 and 120 billion pills annually at an estimated 40,000 tonnes (44,000 tonnes) of weight. It is listed as one of the Essential Medicines by the World Health Organization.

Lipinski's rule of five:

- Lipinski's rule of five is a rule of thumb that describes the drugability of a determinate molecule.
- This rule helps to determine if a biologically active chemical is likely to have the chemical and physical properties to be orally bioavailable.
- The Lipinski rule bases pharmacokinetic drug properties such as absorption, distribution, metabolism and excretion on specific molecular properties such as:
 - No more than 5 hydrogen bond donors
 - No more than 10 hydrogen bond acceptors
 - Molecular mass less than 500 Da
 - Partition coefficient not greater than 5
- The violation of 2 or more of these conditions predicts a molecule as a non- orally available drug.

PubChem

PubChem is an open chemistry database at the National Institutes of Health (NIH). “Open” means that you can put your scientific data in PubChem and that others may use it. Since the launch in 2004, PubChem has become a key chemical information resource for scientists, students, and the general public. Each month our website and programmatic services provide data to several million users worldwide. PubChem mostly contains small molecules, but also larger molecules such as nucleotides, carbohydrates, lipids, peptides, and chemically-modified macromolecules. We collect information on chemical structures, identifiers, chemical and physical properties, biological activities, patents, health, safety, toxicity data, and many others. PubChem records are contributed by hundreds of data sources. Examples include:

government agencies, chemical vendors, journal publishers, and more. The amount of data in PubChem is ever-growing, please visit the PubChem Statistics page to find out what the latest data counts are.

SwissADME:

During the time and resource-consuming processes of drug discovery and development, a large number of molecular structures are evaluated according to very diverse parameters in order to steer the selection of which chemicals to synthesize, test and promote, with the final goal to identify those with the best chance to become an effective medicine for the patients. The molecules must show high biological activity together with low toxicity. Equally important is the access to and concentration at the therapeutic target in the organism. The traditional way to consider pharmacokinetics (i.e. the fate of a therapeutic compound in the organism) is to break down the various effects that impact the access to the target into individual parameters. It has been demonstrated that early estimation of ADME in the discovery phase reduces drastically the fraction of pharmacokinetics-related failure in the clinical phases. Computer models have been fostered as a valid alternative to experimental procedures for prediction of ADME, especially at initial steps, when investigated chemical structures are numerous but the availability of compounds is scarce. A large variety of in silico methods share the objective of predicting ADME parameters from molecular structure. Noteworthy, the pioneer work of Lipinski et al. examined orally active compounds to define physicochemical ranges for high probability to be an oral drug. This so-called Rule-of-five delineated the relationship between pharmacokinetic and physicochemical parameters. Whereas physicochemical parameters give a global description of the structure, molecules can be directly described by substructure searches. These techniques are at the root of Structural Alert, the PAINS or the Lilly MedChem filters applied to cleanse chemical libraries from compounds most likely unstable, reactive, toxic, or prone to interfere with biological assays because unspecific frequent hitters, dyes or aggregators. The SwissADME web tool presented here is freely accessible at <http://www.swissadme.ch> and meant for user-friendly submission and easy analysis of the results, also for non-expert in CADD. Compared to the state-of-the art of free web-based tools for ADME and pharmacokinetics and apart from unique access to proficient methods, SwissADME strong points are, non-exhaustively: different input methods, computation for multiple molecules, and the possibility to display, save and share results per individual molecule or through global intuitive and interactive graphs.

Methodology:

1. Using Google search engine search Swiss ADME.
2. In Molecular Sketcher Add, Import a structure.
3. The Structure will be from PubChem (ID- 2244).
4. Download the structure and save is SDF format.
5. Import the structure to the Molecular Sketcher in SwissADME.
6. Transfer the structure to SMILES list.
7. Run the calculations and observe the result page.
8. Browse through various parts of this result page.

Observations:

This website allows you to compute physicochemical descriptors as well as to predict ADME parameters, pharmacokinetic properties, druglike nature and medicinal chemistry friendliness of one or multiple small molecules to support drug discovery.

The main article describing the web service and its underlying methodologies is **SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules**. *Sci. Rep.* (2017) 7:42717.

For details about development and validation of iLOGP, please refer to this article: **iLOGP: a simple, robust, and efficient description of n-octanol/water partition coefficient for drug design using the GB/SA approach**. *J. Chem. Inf. Model.* (2014) 54(12):3284-3301.

For details about development and validation of the BOILED-Egg, please refer to this article: **A BOILED-Egg to predict gastrointestinal absorption and brain penetration of small molecules**. *ChemMedChem* (2016) 11(11):1117-1121.

Developed and maintained by the **Molecular Modeling Group** of the SIB | Swiss Institute of Bioinformatics.

Figure 1: SwissADME Homepage

National Library of Medicine
National Center for Biotechnology Information

PubChem

About Posts Submit Contact

Search PubChem

COMPOUND SUMMARY

Penicillin g

PubChem CID 5904

Structure

Chemical Safety

Molecular Formula C₁₆H₁₈N₂O₄S

Synonyms penicillin g, Benzylpenicillin, 61-33-6, Benzylpenicillanic acid, Free penicillin II

Molecular Weight 334.4

Dates Modify 2022-10-01 Create 2004-09-16

Penicillin G is a broad-spectrum, beta-lactam naturally occurring penicillin antibiotic with antibacterial activity. Penicillin G binds to and inactivates the penicillin binding proteins (PBPs) located inside the bacterial cell wall. Inactivation of PBPs interferes with the cross-linkage of peptidoglycan chains necessary for bacterial cell wall strength and rigidity. This interrupts bacterial cell wall synthesis and results in the weakening of the bacterial cell wall and eventually causing cell lysis.

CITE DOWNLOAD

CONTENTS

Title and Summary

1 Structures

2 Names and Identifiers

3 Chemical and Physical Properties

4 Spectral Information

5 Related Records

6 Chemical Vendors

7 Drug and Medication Information

8 Food Additives and Ingredients

9 Pharmacology and Biochemistry

10 Use and Manufacturing

11 Identification

12 Safety and Hazards

13 Toxicity

14 Associated Disorders and Diseases

15 Literature

16 Patents

17 Biomolecular Interactions and Pathways

18 Biological Test Results

19 Taxonomy

20 Classification

Figure 2: PubChem Compound ID

The screenshot shows the PubChem compound summary page for Penicillin g (CID 5904). A modal window titled "DOWNLOAD" is open, listing various formats for download: JSON, XML, ASNT, and SDF. Below this, sections for "2D Structure" and "3D Conformer" also provide download links for SDF, JSON, XML, and ASNT. The main page displays the chemical structure of Penicillin g, its molecular formula ($C_{16}H_{18}N_2O_4S$), and its synonyms (penicillin g, Benzylpenicillin, 61-33-6, Benzylpenicillanic acid, Free penicillin II).

Figure 3: Compound structure, download in SDF format

The screenshot shows the SwissADME website. At the top, there is a navigation bar with links to various services: SwissDrugDesign, SwissDock, SwissParam, SwissSidechain, SwissEloisostere, SwissTargetPrediction, SwissADME, SwissSimilarity, and About us. The main content area features the SwissADME logo and a brief description of the service's purpose: "This website allows you to compute physicochemical descriptors as well as to predict ADME parameters, pharmacokinetic properties, druglike nature and medicinal chemistry friendliness of one or multiple small molecules to support drug discovery." Below this, a note cites a paper: "The main article describing the web service and its underlying methodologies is SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. Rep.* (2017) 7:42717." Further down, it mentions two other articles: "For details about development and validation of iLOG, please refer to this article: iLOGP: a simple, robust, and efficient description of *n*-octanol/water partition coefficient for drug design using the GB/SA approach. *J. Chem. Inf. Model.* (2014) 54(12):3284-3301." and "For details about development and validation of the BOILED-Egg, please refer to this article: A BOILED-Egg to predict gastrointestinal absorption and brain penetration of small molecules. *ChemMedChem* (2016) 11(11):1117-1121." At the bottom, it states: "Developed and maintained by the Molecular Modeling Group of the SIB | Swiss Institute of Bioinformatics." On the left, there is a Marvin JS interface for drawing chemical structures, and on the right, a text input field labeled "Enter a list of SMILES here:".

Figure 4: Import The structure in molecular structure

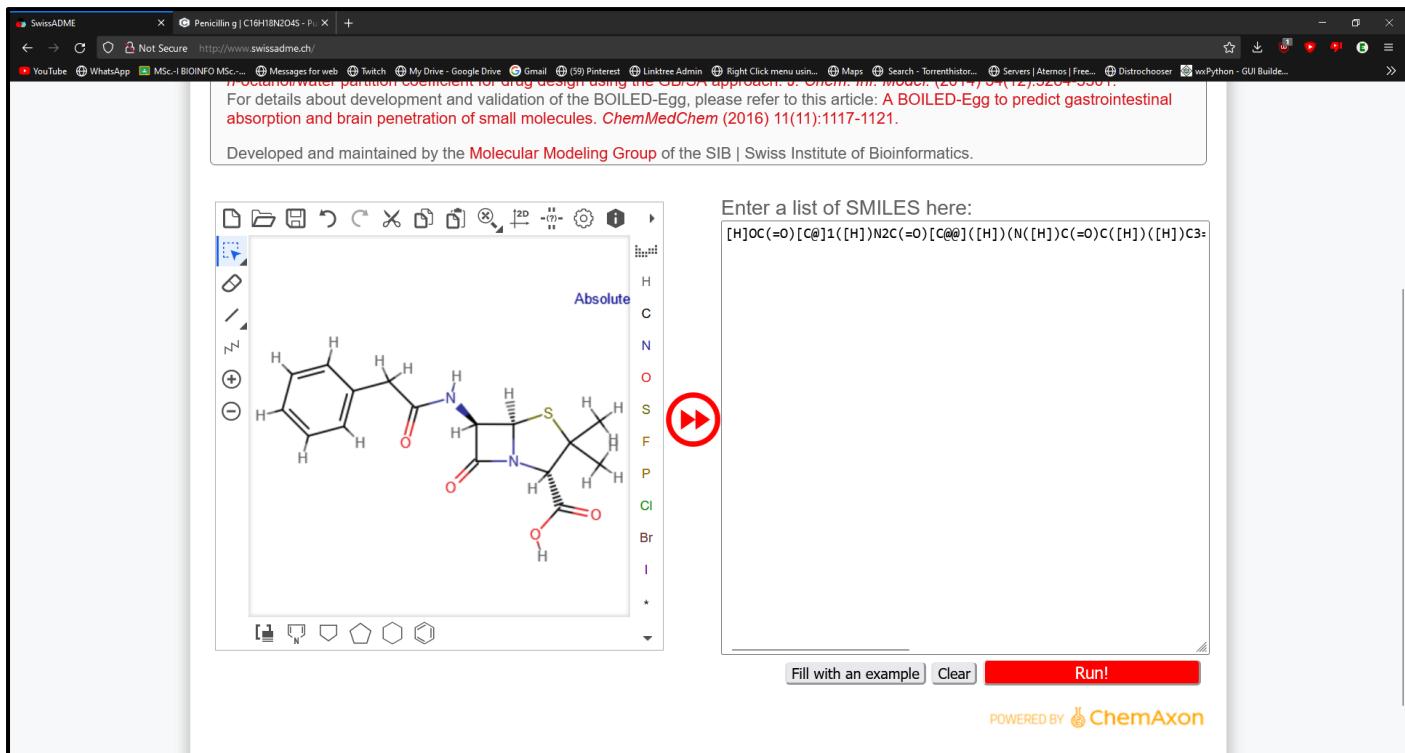


Figure 5: Transfer in to SMILES list

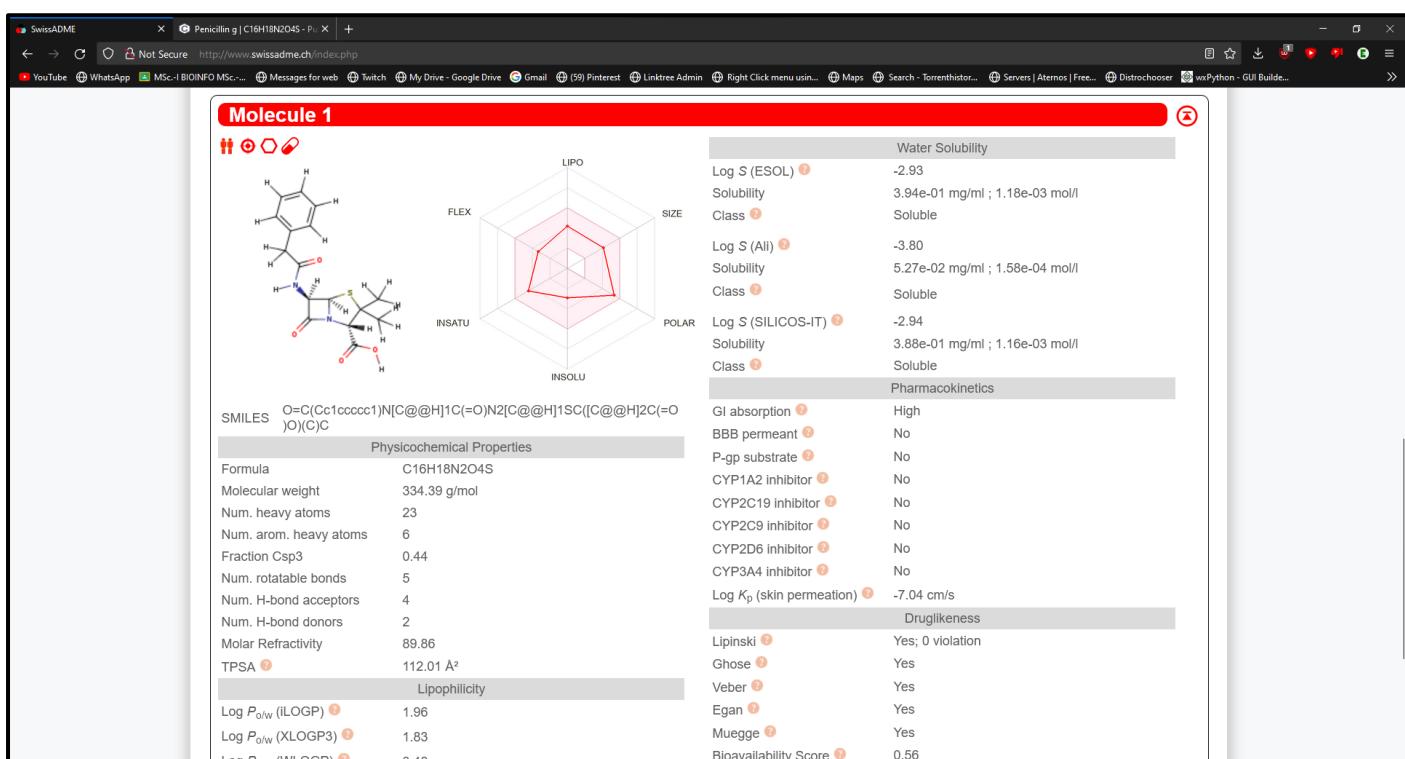


Figure 6: After running the calculations, Result page of the molecule

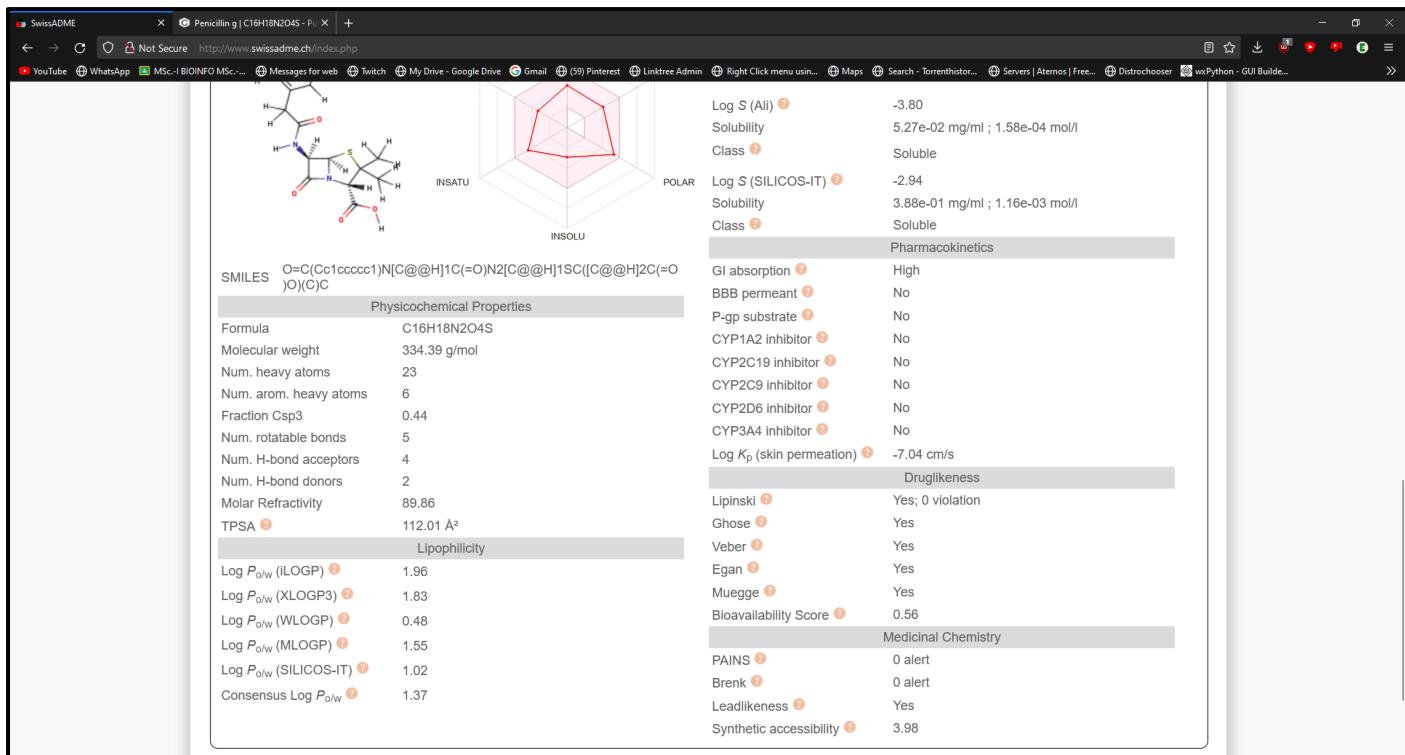


Figure 7: Properties, drug likeness, Pharmacokinetics, etc of the compound in the result

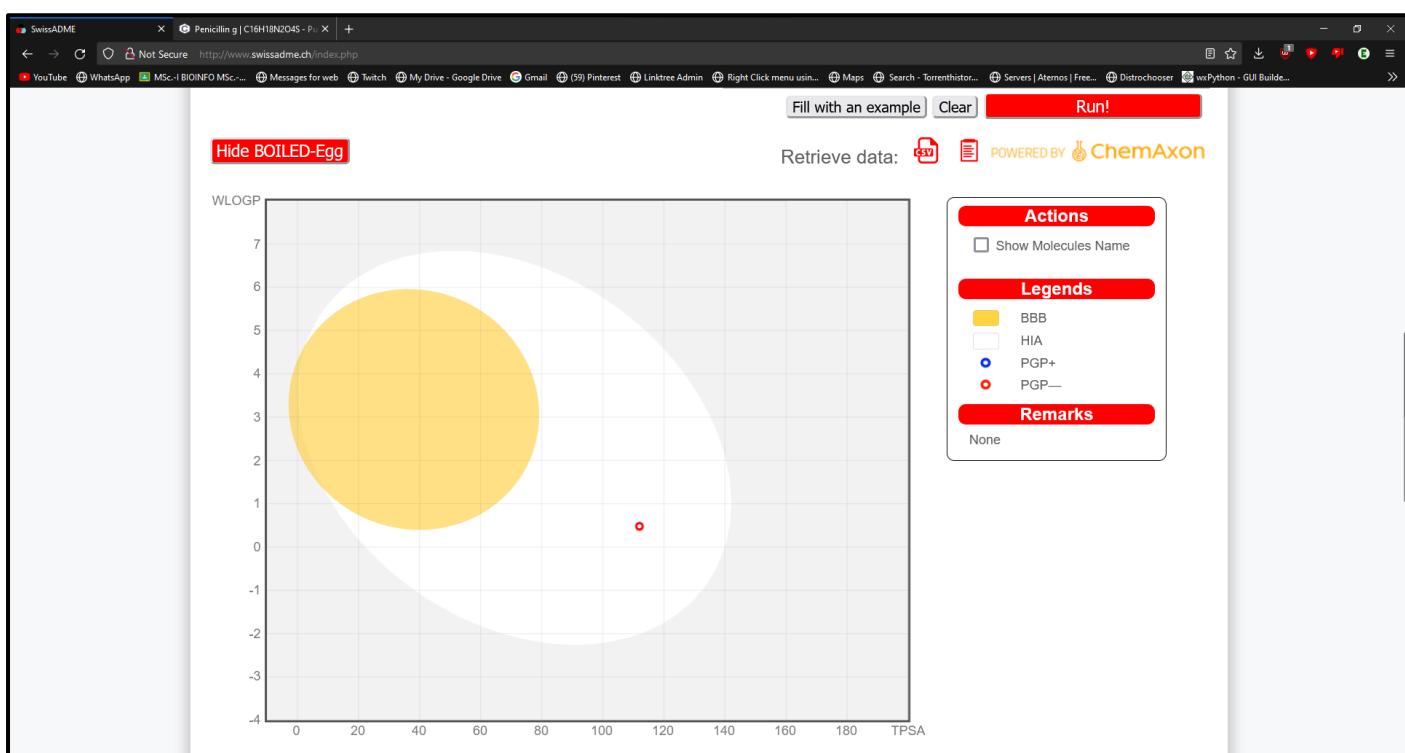


Figure 8: Boiled Egg view/ Graphical view/ Pharmacokinetic result of the compound

This website allows you to perform ligand-based virtual screening of several libraries of small molecules, using different approaches.

If you use this web tool, please, cite the following paper: Zoete, V., Daina, A., Bovigny, C., & Michielin, O. SwissSimilarity: A Web Tool for Low to Ultra High Throughput Ligand-Based Virtual Screening., *J. Chem. Inf. Model.*, 2016, 56(8), 1399-1404.

Choose a reference small molecule

Paste a SMILES in this box, or draw the reference molecule

[H]C1=CC([H])=C([H])=C1[H]C@C([H])[C@@H]2[C([H])SC1(C([H])([H])[H])C([H])[H])C([H])[H]

Examples:

Choose a method and a library to screen

Choose a library of small molecules to screen and the screening methods in the list below.

Perform the screening

(Provide a SMILES before submitting)

Figure 9: Similarities searched through SwissSimilarity

SwissDrugDesign | SwissDock | SwissParam | SwissSidechain | SwissEloisostere | **SwissTargetPrediction** | SwissADME | SwissSimilarity | About us

Query Molecule

Target Classes

Top 15
Top 25
Top 50
All

Target Class	Percentage
Membrane receptor	26.7%
Protease	26.7%
Family A G protein-coupled receptor	13.3%
Nuclear receptor	6.7%
Enzyme	6.7%
Undclassified protein	6.7%
Phosphatase	6.7%

Export results:

Show 15 entries Search:

Target	Common	Uniprot	ChEMBL ID	Target Class	Probability*	Known actives
--------	--------	---------	-----------	--------------	--------------	---------------

Figure 10: Target prediction of the compound in SwissTargetPrediction

Result:

In Boiled egg section molecule-1 is present in yellow section i.e. Boiled eggs yolk where molecules are predicted to passively permeate through the blood brain barrier and according to its physiochemical properties and lipophilicity the molecule also obeys the Lipinski 5 rule.

Lipinski 5 Rule:

1. Molecular weight less than 500 Dalton - Molecular weight (180.16g /mol).
2. High lipophilicity expressed as LogP less than 5 – LogP (1.30)
3. Less than 5 hydrogen bond donors - No. of H bonds donors (1)
4. Less than 10 hydrogen bond acceptors - No. of H bonds acceptors (4)
5. Molar refractivity should be between 40-130 – Molecular Refractivity (44.90)

Pharmacokinetics:

Gastrointestinal absorption – High

BBB Permeant – Yes

CYP1A2 inhibitor – NO

CYP2C9 inhibitor - NO

CYP2C19 inhibitor – NO

Drug likeness:

Lipinski – Yes; 0 Violation

Ghose - Yes

Veber - Yes

Egan - Yes

Muegge No; - 1 violation: MW<200

Bioavailability Score - 0.85

Conclusion:

The Pubchem ID of the query was taken & imported in SwissADME database and SMILES were generated. After generating SMILES and running calculations various results were obtained, where in Boiled egg section the molecule-1 was present in the yellow section which means that it can enter in the BBB i.e. Blood brain barrier but it cannot cross the (GI) Gastro intestinal track. In Drug likeness section, filters like Lipinski, Ghose, Veber, Egan was observed with 0 violations with bioavailability score 0.85.

References:

1. SwissADME Homepage (<http://www.swissadme.ch>)
2. SwissADME literature (<https://www.nature.com/srep>)
3. PubChem Compound (<https://pubchem.ncbi.nlm.nih.gov/>)
4. Compound literature (<https://www.medicalnewstoday.com/articles/161255>)
5. SwissSimilarity (<http://www.swisssimilarity.ch/>)
6. SwissTargetPrediction (<http://www.swisstargetprediction.ch/>)

WEBLEM 7: CombiGlide

Aim: CombiGlide: A building molecular database along with special emphasis on retrieval using structure input.

Introduction:

The number of drug-discovery initiatives that have a high-resolution crystal structure of the receptor available in recent years has increased, and this trend is projected to continue, as a result of the human genome project and high-throughput crystallography studies. A common computational strategy in this case is to dock molecules into the receptor from a physical or virtual database and then apply an appropriate scoring function to calculate the binding affinity. Numerous dockings programmes are widely used by the pharmaceutical and biotechnology sectors; the most common ones seem to be GOLD, FlexX, and DOCK. Over the past few years, these programmes have shown significant success with virtual screening applications.

Grid-based ligand docking with energetics is the new docking technique used by the Glide software suite from FirstDiscovery. Glide was created to screen large libraries quickly enough while doing the most thorough possible search of the positional, orientational, and structural space. This was accomplished by using several hierarchical filters, as will be described in more detail below. The current performance characteristics of Glide are as follows.

- 1) On an AMD Athelon MP 1800+ CPU running Linux, docking times for data sets with 0–10 rotatable bonds typically take less than one minute.
- 2) In comparison to what is published in the present literature for widely used docking algorithms, robustness in binding mode prediction is qualitatively greater. For instance, a comparison with the results obtained by the creators of GOLD shows that for the 72 noncovalently bound cocrystallized ligands of the GOLD test set¹⁵ that have 10 or fewer rotatable bonds, Glide's average rmsd is 1.46 Å while GOLD's is 2.56. Even more favourably, FlexX is compared. Similar conclusions can be drawn from comparisons for ligands with up to 20 rotatable bonds.
- 3) The predicted binding affinities for cocrystallized complexes are reasonable (2.3 kcal/mol rmsd), although they could certainly be better.
- 4) The results of the library screening, which are given in the paper that follows, 16 are quite positive. Additionally, compared to earlier Glide versions, Glide 2.5 yields database enrichment factors that are noticeably greater.

Overview: Docking Methodology

Glide employs a system of filters to search for probable ligand locations in the receptor's active-site region. The shape and properties of the receptor are described by different sets of fields on a grid that enable progressively more accurate scoring of the ligand posture. These fields are generated during the preprocessing steps of the calculation, thus they only need to be computed once for each receptor.

The following procedure results in a number of initial ligand conformations. These conformations are given in a compact combinatorial form and are chosen from an extensive list of the minima in the ligand torsion-angle space. Given these ligand conformations, preliminary searches are made across the ligand's full phase space to identify viable ligand poses.

While avoiding the use of stochastic methods, which can miss important phase-space regions occasionally and prevent the development of a truly robust algorithm, his prescreening significantly reduces the region of phase space over which computationally expensive energy and gradient evaluations will later be carried out. While Glide is unique in that it relies on comprehensive systematic search techniques, reasonable computational speed necessitates approximations and truncations.

Using a typical molecular mechanics energy function along with a distance-dependent dielectric model, the ligand is minimised in the field of the receptor starting from the poses chosen by the initial screening.

Finally, on the three to six lowest-energy sites discovered in this manner, a Monte Carlo method is employed to search for close torsional minima. Such a method may occasionally be required to accurately orient peripheral groups and potentially alter internal torsion angles. The standard molecular mechanics energy function offers a trustworthy model for predicting binding modes even in the absence of a solvent.

However, it is sufficient for ranking different ligands, such as ligands with different net charges. A modified and improved version of the ChemScore! scoring tool, GlideScore, is created to forecast binding affinity and rank-order ligands in database screens. Combining the GlideScore, the ligand-receptor molecular mechanics interaction energy, and the ligand may determine the optimal docked configuration.

It is necessary to modify the scoring method, particularly the molecular mechanics component, to take into account the fact that the protein structure utilised for docking is frequently not optimised to fit a particular ligand. This is the key last issue. Most actives cannot fit because the protein cavity is too small when docking a library of ligands into a single stiff receptor shape. Van der Waals radii of often chosen (e.g., nonpolar) protein and/or ligand atoms are scaled down to expand the binding pocket in order to prevent this. According to studies, this tactic is effective here. The default values are sufficient, even if altering the scale parameters for a specific receptor typically leads to superior enrichment.

Scoring Function:

The starting point for Glide scoring is the empirically based ChemScore function of Eldridge et al., which can be written as:

$$\Delta G_{\text{bind}} = C_0 + C_{\text{lipophilic}} \sum f(r_{lr}) + C_{\text{hydrogen bond}} \sum g(\Delta r) h(\Delta \alpha) + C_{\text{metal}} \sum f(r_{lm}) + C_{\text{rotational}} H_{\text{rot}}$$

The summation include all ligand-atom/receptor-atom pairs classified as lipophilic by ChemScore, whereas the third term encompasses interactions involving ligand-receptor hydrogen bonds. For lengths or angles that are within nominal bounds and fall inside greater threshold values, f, g, and h are functions that provide a complete score (1.00) and a half score (1.00-0.00), respectively. For instance, g(r) is 1.00 if the H—X hydrogen bond distance is less than or equal to 0.25 of the nominal value of 1.85, but it linearly decreases to zero if it is between 2.10 and 2.50. In a similar manner, h(R) equals 1.00 if the Z-H— X angle is between 150° and 120° and zero between 180° and 30°.

GlideScore 2.5 modifies and extend the CHemScore function as follows:

$$\begin{aligned} \Delta G_{\text{bind}} = & C_{\text{lipophilic-lipophilic}} \sum f(r_{lr}) + C_{\text{hydrogen-bond-neutral-neutral}} \sum g(\Delta r) h(\Delta \alpha) + C_{\text{hydrogen-bond-neutral-charged}} \sum g(\Delta r) h(\Delta \alpha) \\ & + C_{\text{hydrogen-bond-charged-charged}} \sum g(\Delta r) h(\Delta \alpha) + C_{\text{max-metal-ion}} \sum f(r_{lm}) + C_{\text{rotational}} H_{\text{rot}} + C_{\text{polar-phobic}} V_{\text{polar-phobic}} + C_{\text{coulombic}} E_{\text{coul}} + C_{\text{vdW}} E_{\text{vdW}} + \text{solvation terms} \end{aligned}$$

According to ChemScore, the phrase "lipophilic-lipophilic" means. The ChemScore form is used for the hydrogen-bonding term as well, but it is divided into variously weighted components depending on whether the donor and acceptor are both neutral, one is neutral and the other is charged, or both are charged. The first of these contributions is discovered to be the most stabilising and the final, the charged-charged term, to be the least significant in the optimised scoring system. While using the same functional form as ChemScore, the metal-ligand interaction term differs in three key aspects.

First, only interactions with anionic acceptor atoms are taken into account (such as either of the two oxygens of a carboxylate group). With this adjustment, Glide is now able to distinguish metalloproteases' significant propensity for coordinating the functionality of anionic ligands to metal centres. In addition, when two or more metal ligations are discovered, Glide 2.5 only counts the best interaction. We chose a decent value for the coefficient—2.0 kcal/mol—even though the parameter refinement would have favoured a value that was even more drastically negative. Third, we evaluate the metal ion's net charge in the unligated apo protein (generally straightforward via examination of the directly coordinated protein side chains). An anionic ligand is preferred if the net charge is positive.

Docking Accuracy:

The capacity of Glide to mimic the cocrystallized ligand geometries of a sizable group of 282 publicly accessible PDB24 complexes is assessed in this section. This set contains the majority of the well-known GOLD and FlexX test set members, 50 PDB complexes that were used to evaluate Glide for potential customers, and an additional 50 complexes whose experimental binding affinities were used to create one or more of the empirical scoring functions mentioned in the literature (e.g., ChemScore). These complexes were used to calibrate the GlideScore algorithm along with others discovered in the FlexX and GOLD test sets. Our coverage of the GOLD and FlexX sets is incomplete since Glide cannot handle ligands with more than 35 rotatable bonds or ligands that are covalently linked (seven cases: 1aec, 1ase, 1blh, 1tpp, and 1lmp) (one case: 2er6). Furthermore, one complex (6rsa) was ignored since it has a vanadium atomic species, which is missing parameters from Glide's OPLS-AA force field.

CombiGlide includes the following capabilities and features:

- Library enumeration:
1. Enumerate complete combinatorial libraries.
 2. Untangles and minimizes structures.

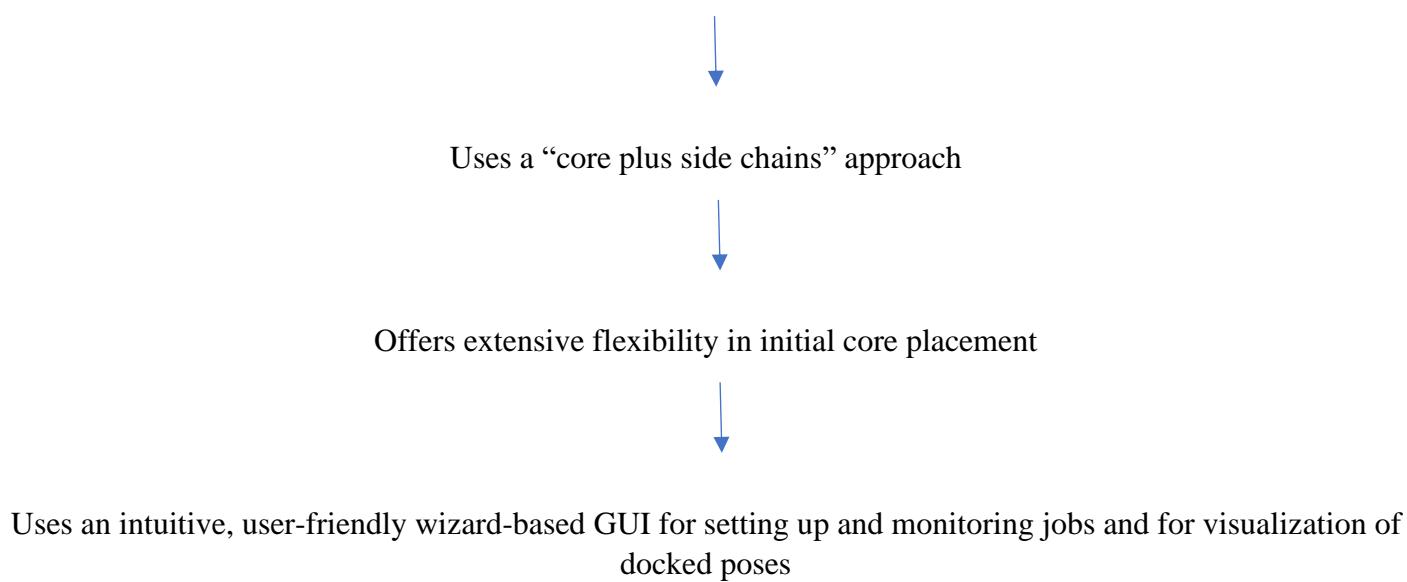
- Interactive enumeration:
1. Define and manage collections of fragments (R groups).
 2. Enumerate complete combinatorial libraries.

Virtual Combinatorial screening:

1. Performs rapid screening of large virtual combinatorial libraries against 3D targets
2. Are orders of magnitude faster than docking the entire library
3. Performs flexible docking using the standard and extra precision (XP) modes of Glide
4. Provides multiple post-docking library selection strategies and options
5. Allows for incorporation of predicted ADME properties into selection process
6. Analyzes selected libraries for enrichment of actives and chemical features

All workflows:

Provides automated reagent file preparation: 2D to 3D conversion, generation of reasonable ionization and tautomeric states, stereoexpansion, assignment of attachment points



References:

1. Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., Repasky, M. P., Knoll, E. H., Shelley, M., Perry, J. K., Shaw, D. E., Francis, P., & Shenkin, P. S. (2004). Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *Journal of medicinal chemistry*, 47(7), 1739–1749. <https://doi.org/10.1021/jm0306430>
2. CombiGlide 2.8: User Manual. (n.d.). Retrieved October 3, 2022, from http://gohom.win/ManualHom/Schrodinger/Schrodinger_2012_docs/combiglide/combiglide_user_manual.pdf