BHARATHIDASAN UNIVERSITY

Tiruchirappalli- 620024,
Tamil Nadu, India.

Programme : M.Sc., Biomedical Science

Course Title : Bioinformatics

Course Code : BM35S1BI

Unit-V

TOPIC: CHEMOINFORMATICS DRUG DATABASE

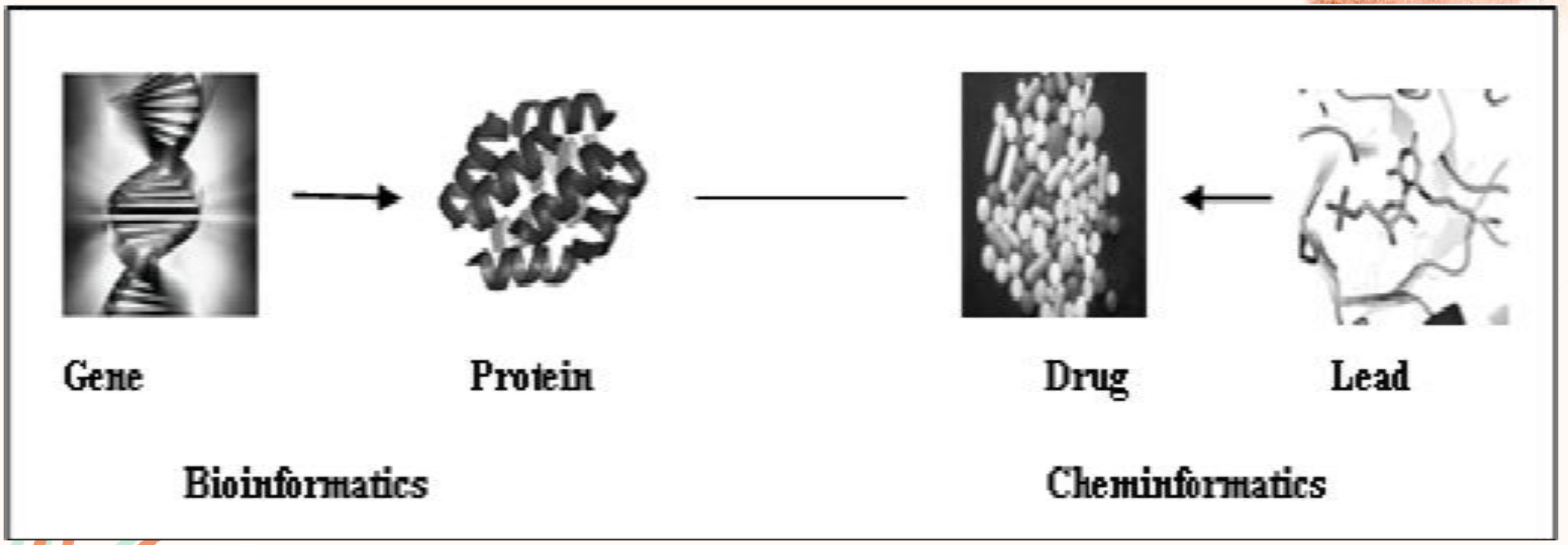
Dr. P. JEGANATHAN
Guest Lecturer
Department of Biomedical Science

CHEMOINFORMATICS DRUG DATABASE

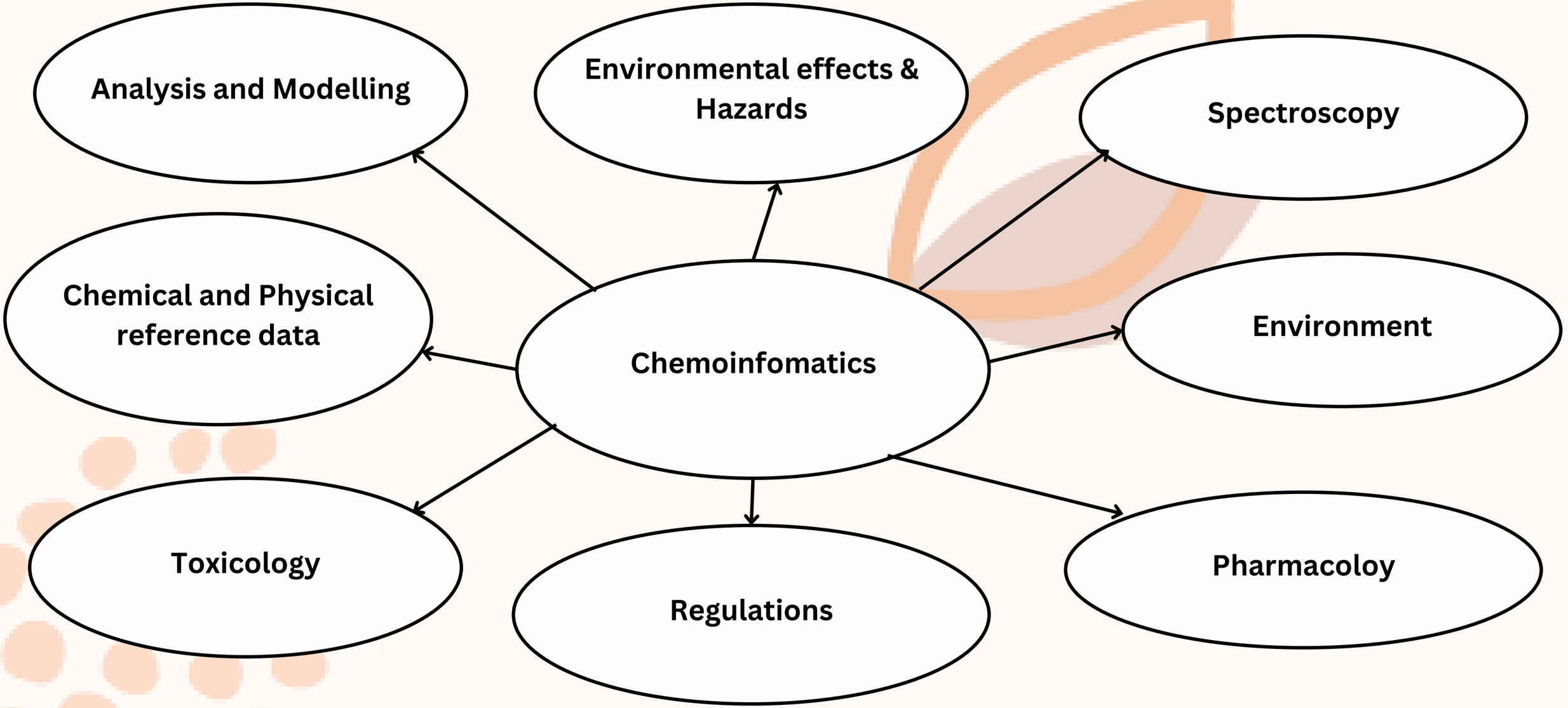
INTRODUCTION

CHEMOINFORMATICS

- It is the combination of chemical synthesis, biological screening and data mining approaches used to guide drug discovery and development .
- **CI** is the use of computer software to assist in the acquisition, analysis and management of data and information relating to chemical compounds and their properties.
- **CI** applies **IT** to chemical data and includes topics such as chemical databases,combinatorial library design, structure-activity relationships and structure based on drug design.


Gene
Protein
Drug
Lead
Bioinformatics
Cheminformatics

NEEDS OF CHEMOINFORMATICS



THREE MAJOR ASPECTS OF CHEMOINFORMATICS

- INFORMATION ACQUISITION,is the process of generating and collecting data empirically (experimentation) or from theory (molecular simulation).
- INFORMATION MANAGEMENT deals with storage and retrieval of information.
- INFORMATION USE, which includes data analysis, correct And application to problems in the chemical and biological sciences.

APPLICATION OF CHEMOINFORMATICS

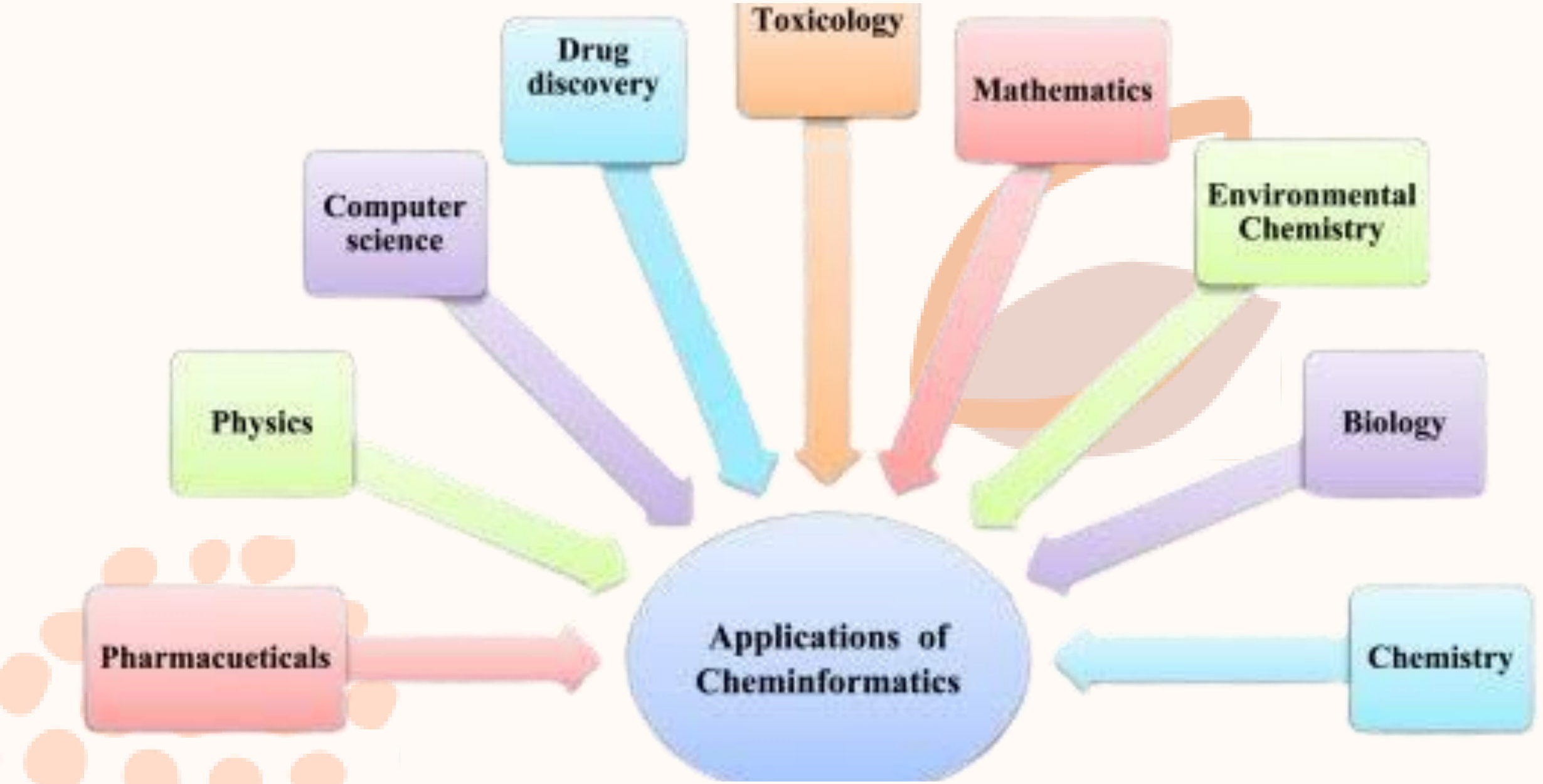
1. Storing data generated through experiments or from molecular simulation retrieval of chemical database (software libraries).
2. Prediction of physical, chemical and biological properties of chemical compounds.
3. Elucidation of the structure of a compound based on spectroscopic data.
4. Structure, substructure, similarly and diversity searching from chemical database.

5. High Throughout Screening (HTS) is the integration of technologies (laboratory automation, asssay technology, microplate based instrumentation, etc.) to quickly screen chemical compunds in search of a desired activity.

6. DOCKING - Interaction between two macromolecule.

7.Drug discovery .

8.Molecular Science , Materials Science , Food Science (nutraceuticals), Atmospheric Chemistry, Polymer Chemistry,Textile Industry, Combinational Organic Synthesis (COS).


Toxicology
Drug discovery
Mathematics
Computer science
Environmental Chemistry
Physics
Biology
Pharmacueticals
Applications of Cheminformatics
Chemistry

RECENT DEVELOPMENT

- Computational chemistry.
- Morden combinational chemistry.
- Drug design and discovery.
- Data sequence, mining and visualisation.
- Chemical database design and their management.
- Chemical information sources.
- Medicinal chemistry etc. has resulted in the emergence of the discipline of CI, which involves the creation, retrieval, organisation, dissemination and processing of chemical information.

SCOPE FOR CHEMOINFORMATICS

- **CI** is the latest area is now becoming a reality India too.
- Till data advance countries like US, UK, Japan and few European countries were working on this.
- In India CI is making inroad in Indian software, Department of Bio-tech Govt of India, R & D organisations, pharmaceuticals and other industries too.
- CI companies are in great need of people with knowledge of chemistry and computer skills to handle the data generated by chemical researchers.

TOOLS USED FOR CHEMOINFORMATICS

1. **ISIS-DRAW** is a chemical structure drawing program for Windows, published by MDL Information Systems. ZIt is the interfacial software to ISIS/ Base database.
2. **ChemDoodle -** chemical structure environment with a main focus on 2D graphics and publisher to create media for structures, reactions and spectra.
3. **ChemOffice** suit of programs and is available for Macintshand Microsoft Windows.
4. **ChemDraw** is a molecular editor developed by the chemoinformatics company CambridgeSoft, ChemDraw is along with Chem3D .

5.LogCHEM, an Inductive Logic Programming (ILP) based tool for discriminative interactive mining of chemical fragments.

6.AmberTools is used Bio molecular Simulation and analysis of polymers, nucleotides, and synthethic Organic structures.

7.PubChem the worlds largest collection of freely available chemical information. It contains close to 100M compounds and more than 236M substances. The dataset seems to be well-maintained including documentation and tutorials. Definitely worth checking out' the link is https://pubchem.ncbi.nlm.nih.gov/

8. ZINC a free database of commercially - available compunds for virtual screening. The link is https://zinc.docking.org/

PubChem

WHAT IS PUBCHEM?

- PubChem is the world's largest collection of freely accessible chemical information. Search chemicals by name, molecular formula, structure, and other identifiers. Find chemical and physical properties, biological activities, safety and toxicity information, patents, literature citations and more.

- PubChem is a database of chemical molecules and their activities against biological assays. The system is maintained by the National Center for Biotechnology Information, a component of the National Library of Medicine, which is part of the United States National Institutes of Health.

PubChem CONTAINS :

- **>243.9** million substance description,
- **>97.6** million unique chemical structure,
- **>264.8** million biological activity test results,
- **>1.3** million biolog assays, covering **>10,000** unique protein sequence target.

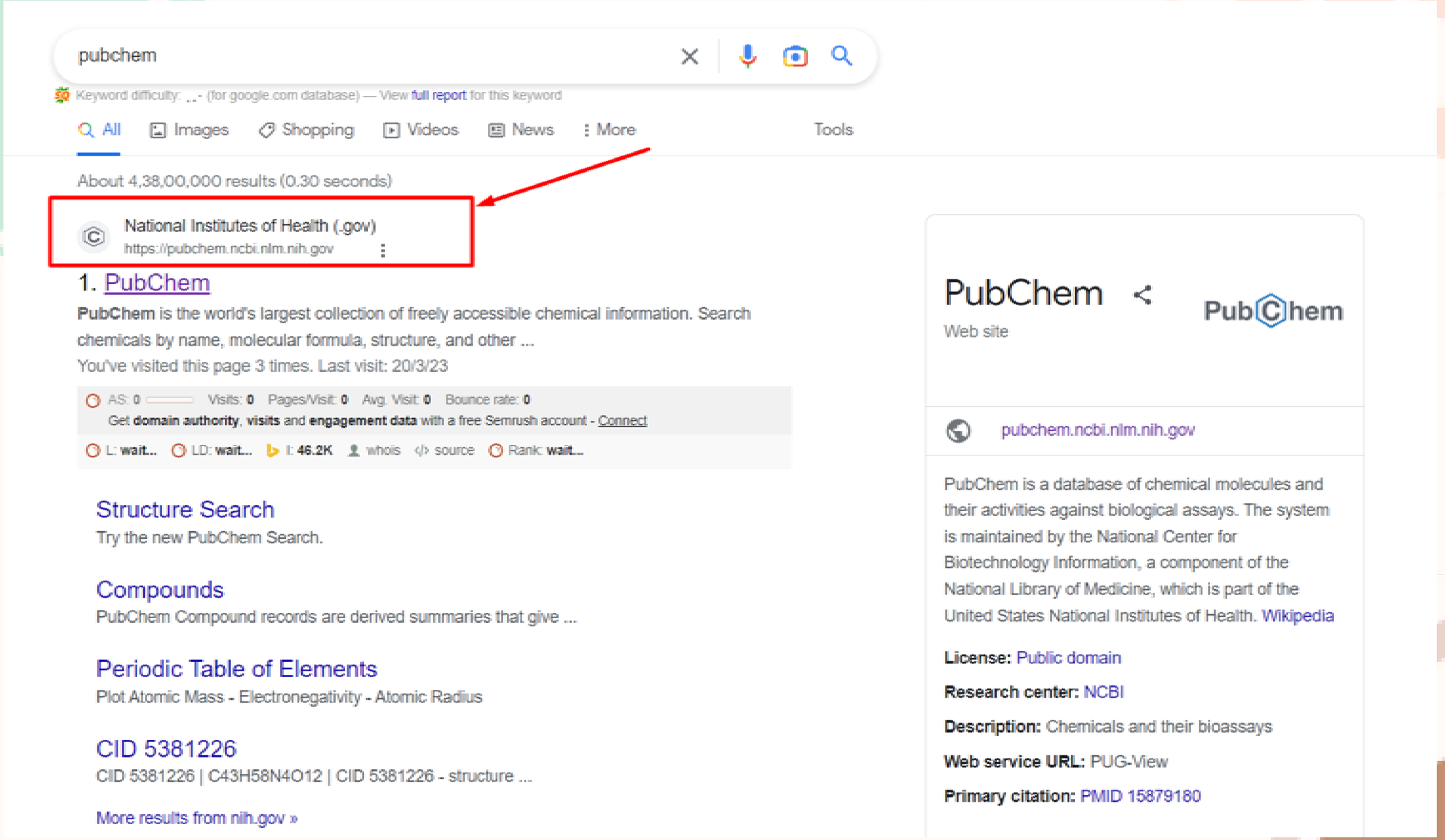
USES OF PUBCHEM :

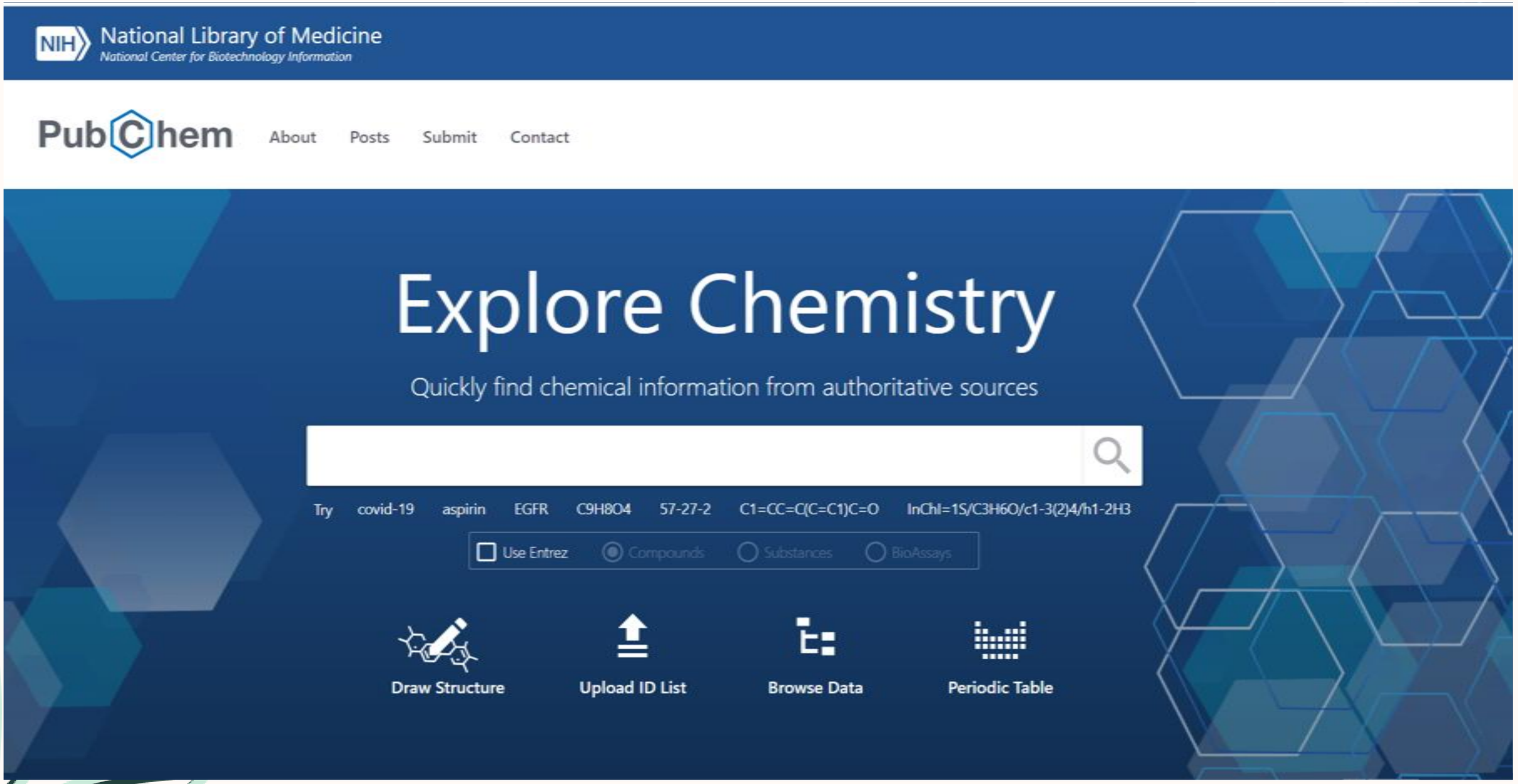
- PubChem supports drug discovery in many aspects such as lead identification and optimization, compound-target profiling, polypharmacology studies, and unknown chemical identity elucidation.

- PubChem archives and organizes information about the biological activities of chemical compounds into a comprehensive database and is the informatics backbone for the Molecular Libraries and Imaging Initiative, which is part of the NIH Roadmap.

- PubChem has been widely used as a ‘big data’ source in machine learning and data science studies for virtual screening, drug repurposing, chemical toxicity prediction, drug side effect prediction and metabolite identification, and so on.

https://pubchem.ncbi.nlm.nih.gov/




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
Draw Structure
Upload ID List
Browse Data
Periodic Table

Search Documents

About
Tutorials
COVID-19/SARS-CoV-2
News
Publications
Citation Guidelines
Data Organization
Data Sources
Search and Tools
Submissions
Downloads
Programmatic Access
Widgets
RDF
Glossary
Data Specification
PubChem3D
Statistics

DOCS

About PubChem

For latest announcements, please visit the PubChem News page.

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.

Where does the data in PubChem come from? 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.

New Web Interface

Multiple updates to the PubChem web interface were released in March 2019. The updates include a new look and feel for

ON THIS PAGE

ADVANTAGES OF PUBCHEM

- Popularity
- Sustainability
- zero-cost to students

DISADVANTAGES OF PUBCHEM

- The **context** of the data is frequently ignored.
- Difficult to standardise chemical structure representation.
- Ambigous / incorrect name- structure association.

ZINC

WHAT IS ZINC?

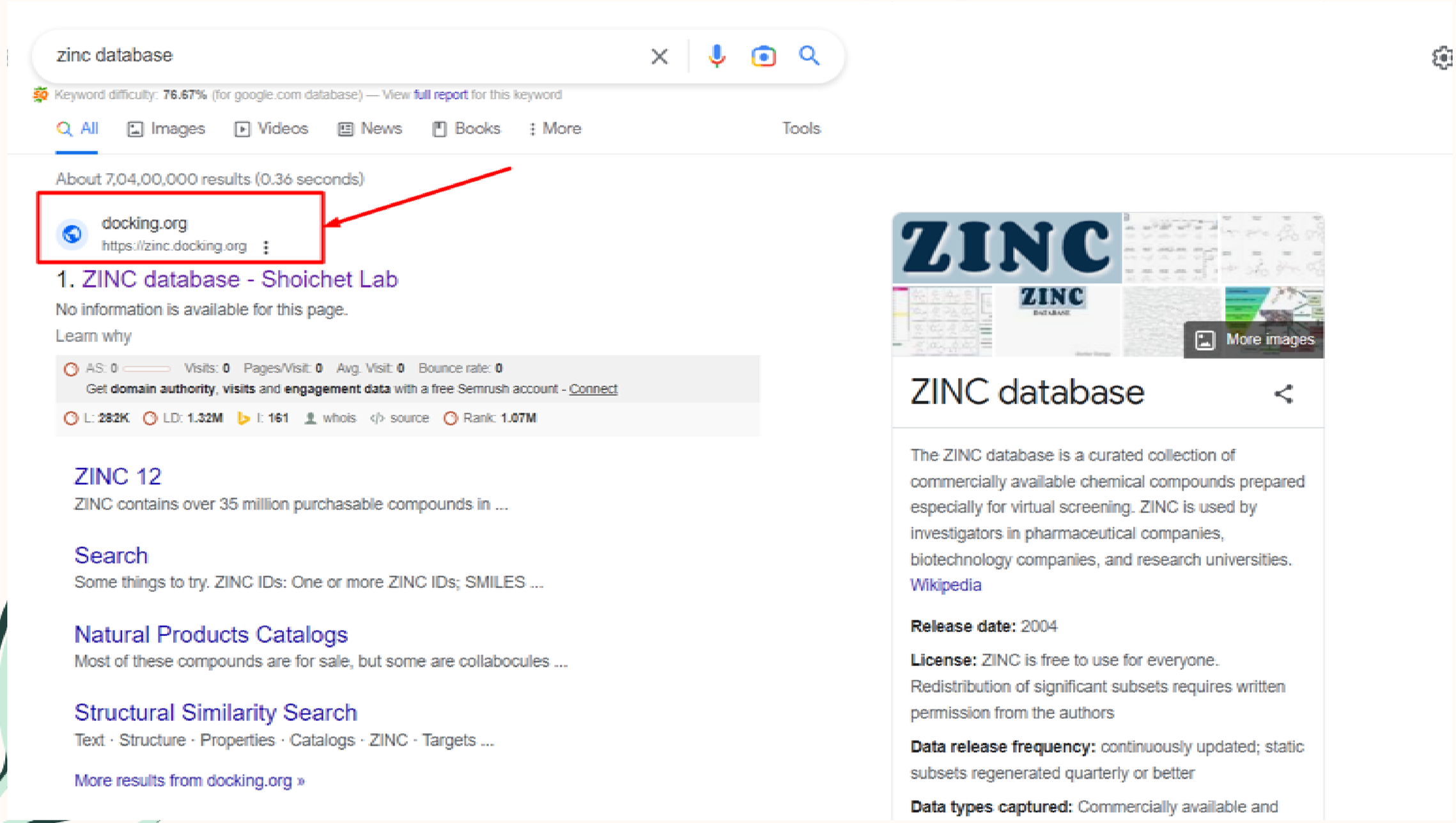
- ZINC is a research tool for investigators seeking chemical matter for their biological targets. It incorporates purchasable compounds from over one hundred vendors and annotated compounds from over twenty databases.

- Welcome to ZINC, a free database of commercially-available compounds for virtual screening. ZINC contains over 230 million purchasable compounds in ready-to-dock, 3D formats. ZINC also contains over 750 million purchasable compounds you can search for analogs in under a minute.

USES OF ZINC DATABASE

- **ZINC** is used by investigators (generally people with training as biologists or chemist)In pharmaceutical companies, biotechnology companies, and research universities.
- It is different from other chemical databases because it aims to represent the biologically relevant, three dimensional form of the molecule.

https://zinc.docking.org/



ZINC Substances Catalogs Tranches Biological More About

ZINC20

Welcome to ZINC, a free database of commercially-available compounds for virtual screening. ZINC contains over 230 million purchasable compounds in ready-to-dock, 3D formats. ZINC also contains over 750 million purchasable compounds you can search for analogs in under a minute.

ZINC is provided by the Irwin and Shoichet Laboratories in the Department of Pharmaceutical Chemistry at the University of California, San Francisco (UCSF). We thank NIGMS for financial support (GM71896).

To cite ZINC, please reference: Irwin, Tang, Young, Dandarchuluun, Wong, Khurelbaatar, Moroz, Mayfield, Sayle, *J. Chem. Inf. Model 2020*, in press. https://pubs.acs.org/doi/10.1021/acs.jcim.0c00675. You may also wish to cite our previous papers: Sterling and Irwin, *J. Chem. Inf. Model, 2015 http://pubs.acs.org/doi/abs/10.1021/acs.jcim.5b00559*. Irwin, Sterling, Mysinger, Bolstad and Coleman, *J. Chem. Inf. Model, 2012* DOI: 10.1021/ci3001277 or Irwin and Shoichet, *J. Chem. Inf. Model. 2005;45(1):177-82* PDF, DOI.

Getting Started

- Getting Started
- What's New
- About ZINC 20 Resources
- Current Status / In Progress
- Why are ZINC results "estimates"?

Explore Resources

Chemistry
Tranches, Substances, 3D Representations, Rings, Patterns
And More
Catalogs, Genes, ATC Codes

Ask Questions

You can use ZINC for **general** questions such as

- How many substances in current clinical trials have PAINS patterns? (150)
- How many natural products have names in ZINC and are not for sale? (9296) get them as SMILES, names and calculated logP
- How many endogenous human metabolites are there? (47319) and how many of these can I buy? (8271) How many are FDA approved drugs? (94)
- How many compounds known to aggregate are in current clinical trials? (60)
- How many epigenetic targets have compounds known? (53) and Which of these substances can I buy? (278)
- How many ligands are there for the NMDA 1 ion channel GRIN1? (662) and How many of these are for sale? (60)
- More...

ZINC20 News

- ZINC20 has been released

Caveat Emptor: We do not guarantee the quality of any molecule for any purpose and take no responsibility for errors arising from the use of this database. ZINC is provided in the hope that it will be useful, but you must use it at your own risk.

REFERENCE

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC102409/

THANK YOU