2022
2(2), 440

Journal of
Molecular Chemistry

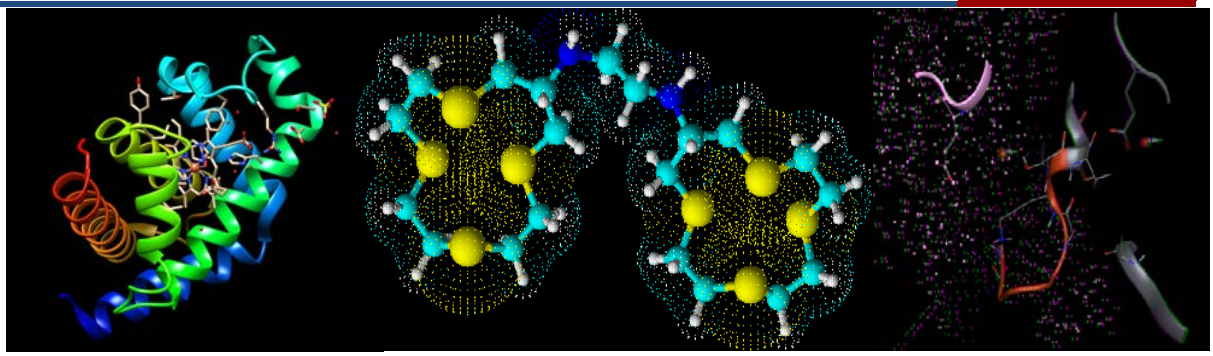
Recent advances and current strategies of cheminformatics with artificial intelligence for development of molecular chemistry simulations

Rajiv Kumar*[1], M. P. Chaudhary[2], Navneet Chauhan[3]

*[1]*University of Delhi, New Delhi. 110007.* [2]*International Scientific Research and Welfare Organization, New Delhi, India.* [3]*Department of Pharmaceutical Chemistry, Saraswati Institute of Pharmaceutical Sciences Dhanap, Gandhinagar 382355 Gujarat, India.*

Submitted on: 25-Jul-2022, Accepted and Published on: 01-Dec-2022

Review



ABSTRACT

The major focus of cheminformatic approaches for drug discovery thus far, notably in the medical field, has been on organic molecules. Cheminformatics has been used to analyze the characteristics of molecular compounds before chemical production and experimental assessment. Cheminformatics-inspired approaches employ the structural and chemical properties of molecules and pharmaceuticals to learn crucial information about the qualities of the molecules and materials being examined. The primary data mining methods utilized in cheminformatics intelligence include structural similarity matrices, descriptor computations, and classification algorithms, which are included in the property interpretations. Artificial chemical intelligence's core principles are focused on using it to find and create new drugs. This review investigates the underlying questions of this method by providing real-world case studies of molecules, medications, and practical uses of cheminformatics in drug design and discovery. In many areas of computer-aided drug discovery, including drug repurposing, metallodrugs, chemistry, material informatics, quantitative structure-activity relationship research, de novo drug design, and chemical space visualization, recent advances in cheminformatics and their use in current drug discovery processes have proven to be incredibly helpful.

Keywords: cheminformatics intelligence, molecular modeling, soft computing, bioinformatics.

INTRODUCTION

Cheminformatics is a branch of research that uses computer-based methods like machine learning and data science to solve various chemistry-related challenges and issues. Computational intelligence and cheminformatics look at a variety of metrics to show the characteristics of molecules, their structure, and forecasts that may be used in numerous medical claims and industries.[1] Signal processing, molecular chemistry simulations, manufacturing, sensor design, and predictive control are just a few of the computational disciplines where computational intelligence has found use outside of its initial purview. Additionally, computational intelligence is a crucial step in the evolution of computer science and has extensive applications in a wide range of scientific fields.[2] Current methods for managing enormous volumes of information include theoretical knowledge and computational intelligence-enabled research applications.[3] One of them, for instance, is molecular simulations, in which researchers anticipate different properties of molecules, such as their solubility, reactivity, and toxicity, using machine learning algorithms.[4] It has been used by researchers to find prospective medications. A few of the challenging problems involved with it are the dearth of data, determining which molecular characteristics are essential for predictions, and converting 3D molecule structures into inputs for the molecular model.[5] It is possible to think of molecules or

*Corresponding Author: Dr. Rajiv Kumar
Tel: +91-9810742944
Email: chemistry_rajiv@hotmail.com

URN:NBN:sciencein.jmc.2022.440

https://pubs.thesciencein.org/jmc

atoms as 3D quantum models when they have distinct positions inside molecules. The relative distance between each molecule, the atomic number, the structure of the cloud, the electron probability, and many other details may be learned from 3D models.[6] However, it is difficult to convert these outputs into a 3D quantum model **(Figure 1)**. Fortunately, a few molecular representations have been created to address this problem.[7]

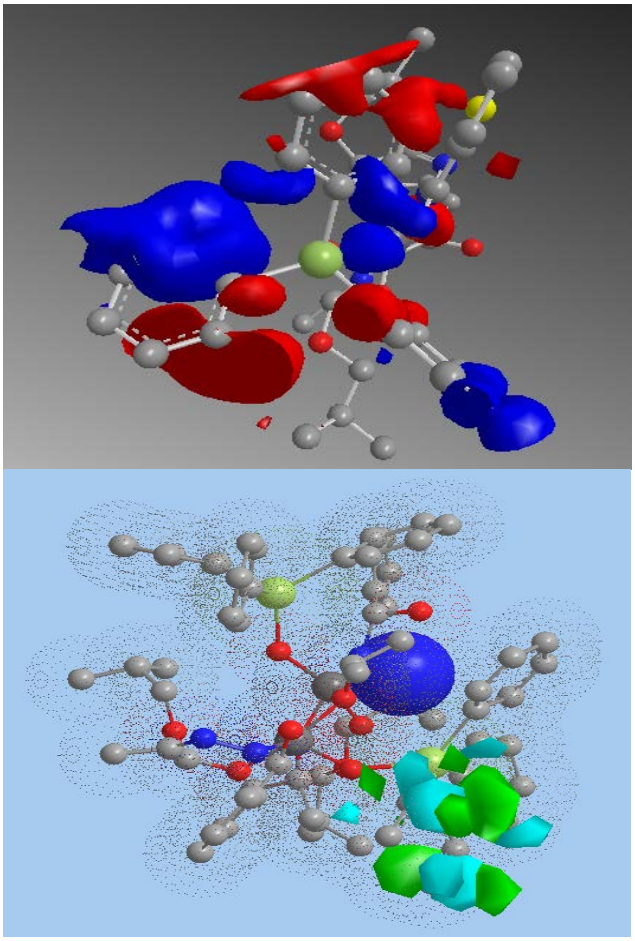


Figure 1. Transforming detected outputs into 3D quantum model that displayed molecular representations have been developed by resolving the existed challenges.

The topic of computational intelligence has lately seen a substantial expansion due to the development of soft computing and artificial intelligence methodologies, tools, and strategies utilized for conceptualizing the essence of intelligence contained in real-world observations. Researchers have been able to explain and comprehend real-world processes and practices that were previously thought to be nearly impossible to explore because they have a better understanding of the inherent imprecision, redundancies, and ambiguities as well as the lack of proper methods used to define the incompleteness and vagueness of the information.[8] This scientific and technological presentation of recent breakthroughs and current techniques of cheminformatics intelligence will be used to foresee emerging strategies for researching hidden regions by offering the most recent research findings.

Numerous computational algorithms have been created to satisfy every conceivable charge material, accuracy, and transport regime need due to the large variety of molecular materials. The straightforward Metabolomics platform has been integrated with bioinformatics tools based on machine learning and data mining methods. Applications for evaluating and comparing three-dimensional molecular objects are gathered under the umbrella of computational intelligence. It converts intricate three-dimensional data into comparable traits or models, enabling high-throughput comparisons and providing machine learning models with information to forecast critical properties like druggable cavities, interaction patterns, protein-protein interfaces, and interactive postures.[9] In order to gain a better knowledge of cheminformatics, the writers of this article study a wide variety of computational intelligence applications across many platforms and industries.

This review article also aims to bring together, clinical, academics and industry experts from the domains of science and engineering who are engaged in primary and transdisciplinary research in computational intelligence, soft computing, communication setups, and computer networks.

CHEMINFORMATICS AND COMPUTATIONAL INTELLIGENCE

The collection, storage, analysis, and manipulation of chemical data are the main concerns of cheminformatics, a relatively new field of information technology. The chemical data of interest (biological or engineering) typically includes information on microscopic molecules and their molecular formulas, spectra, activities, structures, and properties. More, the fluctuation in the microbial β-glucuronidase level in the human gut metagenomes was the explanation given by Ben Busby et al. (2021) for the prediction of drug-metagenome interactions.[10] Thus, cheminformatics has a big influence on many areas of biology, biochemistry, and chemistry even though it was first designed as a tool to help with the drug discovery and development process. The technological basis for spectrometry and spectroscopy, which resulted in the creation of cheminformatics and computational intelligence, required an early, significant revolution. Farahnaz R. Makhouri et al. (2018) talked about using computational methods for drug discovery and design to fight illness.[11]

Cheminformatics uses technology and software to communicate processes and quicken data processing. To define the functional characteristics of the molecular profiles present in the dataset, computational chemistry offers a contextual approach that incorporates ontologies and annotations. The main technological steps in processing, metabolomics, data collection, and statistical techniques, including machine learning, are all covered by different researchers in this work. Cheminformatics is primarily used for the storing and retrieval of information. It may be difficult to index, search for, and store the vast quantity of information that molecules and compounds convey. Thanks to advancements in computer science, particularly in data mining and artificial intelligence, chemists

may now access a massive library of 2D and 3D representations, including very accurate records of previous work. In chapter 9, titles “cheminformatics and computational techniques in metabolomics” of their book Computational Biology, Holger Husi et al. (2019) discuss it. This is important for research since it makes it possible for scientists to make better judgments in the lab because there is more information available.[12]

The chemical and structural characteristics of chemical compounds are used in modern cheminformatics-inspired tactics to give vital information on important molecular interactions. Reproducible computational drug discovery was described by Chanin Nantasenamat et al. in 2020.[13] This strategy could boost biological research and knowledge, notably in the quest for new medications and biomarkers. Its identification formerly dominated high-throughput screening, which was costly, time-consuming, and included extensive trials. Based on virtual screening, biological assessment, and molecular dynamics modeling, Mao, J. et al. (2022) discussed the finding of microtubule stabilizers with unique scaffold architectures.[14] The value of theoretical research is expanding, especially in relation to trend analysis of many elements, understanding of transport, and loss processes **(Figure 2)**.

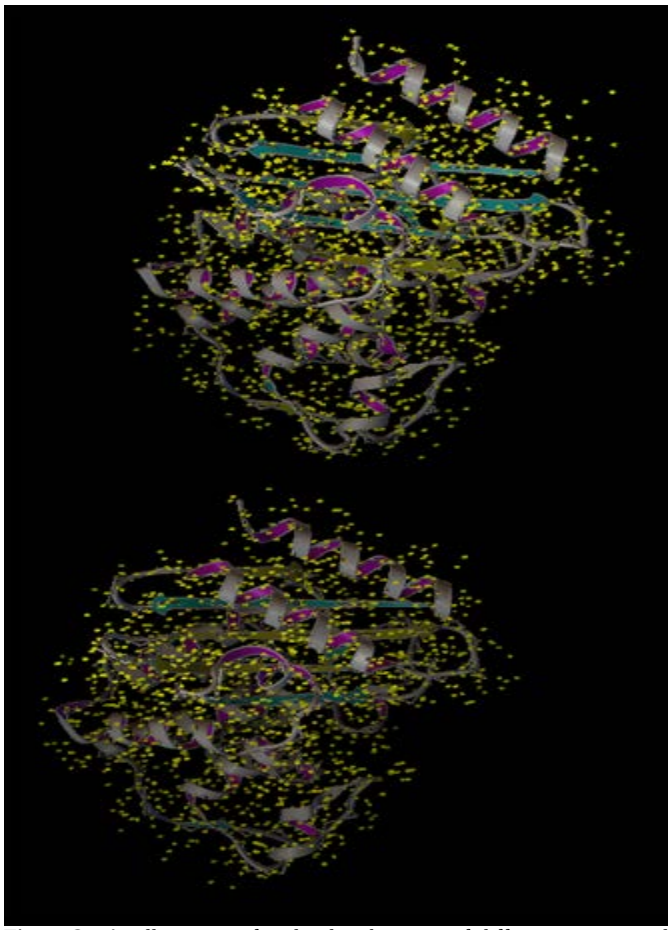


Figure 2. An illustration of molecular chemistry of different proteins and drugs resulting from cheminformatics intelligence to explore innumerable features.

For predicting the potential results of reactions, chemistry and medical breakthroughs have employed a number of computational intelligence methodologies and procedures.[12] Although it has demonstrated its value and importance in a number of medical and industrial claims, little is known about how well they hold up when it comes to the many molecular chemistry aspects. In a study of computational methodologies, Yorley Duarte et al. (2019) provided an explanation of the integration of target identification, drug development, and drug delivery.[15] The thriving field of molecular chemistry has sparked a rise in computational research and cheminformatics information on the chemical and physical characteristics of molecules and other materials, even at the nanoscale. The Bi(V) organic framework in an asymmetric system was described by Rajiv et al. (2008) using synthesis, spectroscopy, XRPD, and molecular modeling.[16] The pharmaceutical industry has begun using cheminformatics to create innovative treatments, but this would not be possible without the availability of virtual libraries.

Using information from both actual and hypothetical molecules, researchers may build virtual libraries of compounds that enable them to explore chemical environments and theories for the synthesis of novel compounds with a certain set of properties. Advanced machine learning-based algorithms and genuine classes of molecules can be used to create these unique substances. This method gives you a lot of time to estimate how a certain chemical will behave depending on its analysis. Analyzing the quantitative structure-activity link is essential as a result.[17] Chemical expert systems can be used by researchers to determine the physicochemical properties and biological potential of compounds. Multi-descriptor physical approaches to

computerized plant ecology, often known as multiSPAS, was de scribed by Oleg Gradov et al. in 2021.[18] Because it may add to a body of information that can support original ideas and direct new decisions in a variety of scenarios, this analytical approach goes beyond cost savings. Even if the most advanced and complicated computational tools are advanced and complicated, it is essential for researchers to keep collaborating with developers to make sure that digital platforms stay accurate and address the problems that the study itself raises. Therefore, the field of cheminformatics confronts major challenges.

Putative drug targets and prospective drug leads were discovered by David S. Wishart (2015) as beginning points for virtual screening and docking.[19] Creating sophisticated algorithms that can swiftly extract knowledge from big raw databases is the main and most notable feature of it. In the world of software development, the second problem is more important **(Figure 3)**. Current computational methodologies in chemistry (such as quantum mechanical methods, statistical machine learning, and molecular dynamics) are not appropriate for broad chemical research since they cannot be scaled up to large datasets. The main benefits and drawbacks of these methods may be contrasted in order to show how each strategy performs in various scenarios.

Drug discovery is the process of developing new compounds that have pharmacological effects on pathological disorders.[20] The entire process is costly and challenging. Despite the incredible technological advancements, it is not practical to create high-throughput screening experiments for all known compounds for a specified target(s). For instance, it is crucial to find interactions between drugs and protein binding sites while developing novel treatments.

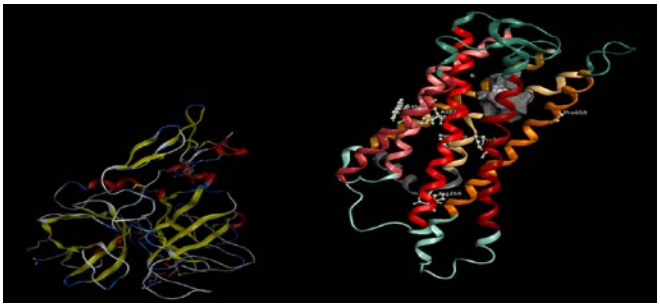


Figure 3. An illustration of cheminformatics intelligence applied for exploring different features of molecular chemistry.

The fact that only a tiny percentage of compounds in the chemical space may clinically retain medication-like qualities highlights the complexity of drug invention approach. Rajiv et al. in their (2015) expose architectural elements, molecular mechanics and computational models were used for studying organometallic assemblies, π-electron delocalization, lipophilic nature, bio-accessibility, μ-bridging spacers, flexibility, bioavailability, intracellular, antimicrobial assimilation and trafficking routes.[21] The process that kicks the whole thing off, leading to the identification of molecules, is likewise lengthy and challenging. Overall, chemoinformatics has a bright future, although it is complex.

Looking at the history of computer use in chemical research makes it simple to see how far the discipline has come.[12] As software developers and chemical professionals work together to advance computational abilities when applied to a variety of chemical-related difficulties, the foundation has been set for future advancement.

STRUCTURE- BASED DRUG SCREENING AND DECOMPOSING COMPOUNDS

Current cheminformatics intelligence methodologies have been employed for trend analysis of molecular chemistry simulations to the study points from the fragment's tolerance to tiny chemical changes without modifying its way of binding.[23] In light of that, it would seem appropriate to look at the structural data at the fragment level. Esther Kellenberger et al. (2018) emphasized structural insights on fragment binding mode conservation. It would be interesting to look at the molecular properties and binding mechanisms of molecular fragments from a structural viewpoint given the emphasis on the significance of molecular fragments in the drug development process **(Figure 4)**. For structure-based drug screening, Michael

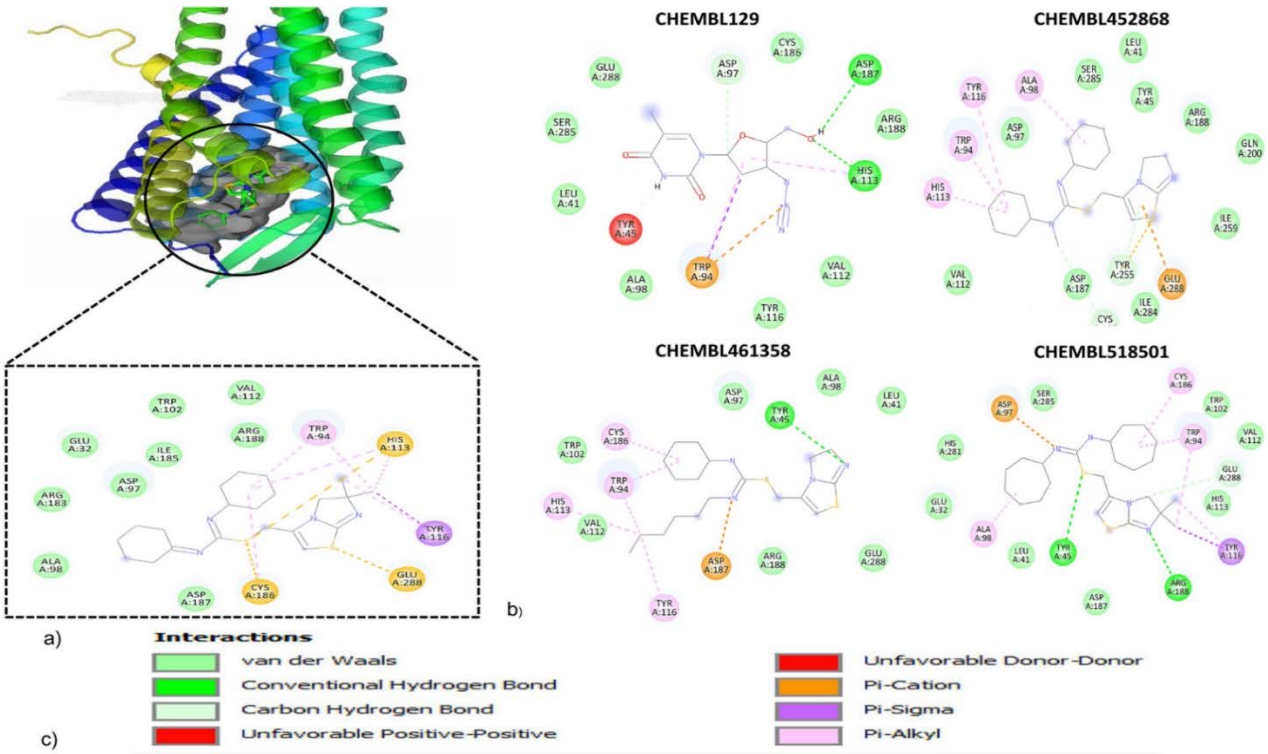


Figure 4. The PS module's representation of optimized hits. The true positives (TP) discovered by the PS-driven module of A-HIOT were converted into molecular IDs, combined with PLIP score (di), and ranked while adhering to the suggested CXCR4 binding interaction profile threshold. The shows the final ranking score for each ligand molecule that was subjected to optimization; a displays the CXCR4 interaction patterns and participating amino acid residues with its standard ligand (IT1t); b accumulates all four molecules (CHEMBL129, CHEMBL452868, CHEMBL461358, and CHEMBL518501) from an independent set administered for optimization coupling interaction patterns; and c describes the types of interaction and bond formation pattern. Reprinted (adapted) with permission from.[22] Copyright (2022) Vishal Acharya et al. Journal of Cheminformatics. BMC Springer Nature.

Schroeder et al. (2022) showed how decomposing compounds enabled the reconstruction of interaction fingerprints.[24] However, the majority of experiments were conducted in a constrained space.

The constrained set is biased toward the study interests of crystallographers, as seen in the work of Drwal et al. (2018), which was restricted to fragments crystallized as small molecule ligands in PDB structure. In this study, the authors describe the binding of molecular fragments in all drugs from the parallel database (PDB), create a structural metric to evaluate the conservation of the fragments' binding modes, and then use the results to reconstruct the whole drugs' binding modes in the absence of structural data.[25] The authors evaluate whether it is feasible to transfer structural knowledge from a fragment level to a drug level in order to expand the applicability of structure-based drug repositioning and other techniques dependent on the structural characterization of pharmaceuticals. In 2016, Steffen Lindert et al. explored computational approaches to drug development.[26] Drug repositioning gives already-available medications new applications. This approach has the benefits of a decreased risk of failure, a quicker medication development process, and cheaper costs. As a result, drug positioning offers a strong substitute for conventional drug research and development.

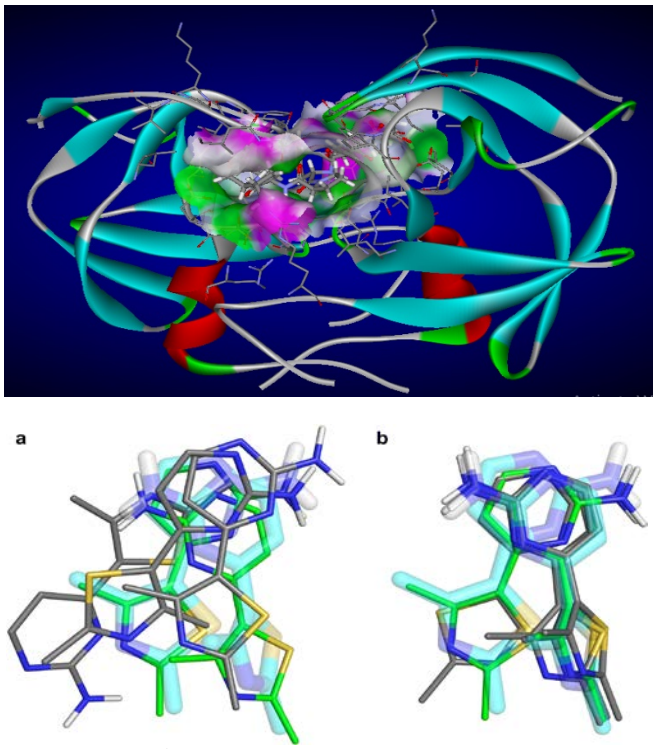


Figure 5. CK2 is found in several positions within cyclin-dependent kinase 2. Two binding mechanisms were identified by crystallographic structures in the protein site (Uniprot: P24941). (a) Transparent blue sticks are used to illustrate the crystallographic poses (PDB 1PXJ and 2C5O). The top ranked accurate poses (RMSD to the native pose 1.0) are displayed in green. The worst stances are shown in grey. The top six poses as determined by GRIM rating. (b) The top seven poses as determined by ROCS ranking. Reprinted (adapted) with permission from.[30] Copyright (2019) Esther Kellenberger et al. Journal of Cheminformatics. BMC Springer Nature.

Structure-based drug repositioning employs the 3D model of proteins to specify the binding strategy of drugs to their protein objects from an energetic/geometrical perspective **(Figure 5).**[27] The modeling of three-dimensional protein structures for use in drug creation was described by Torsten Schwede et al. (2014).[28] For a complete understanding at the quantum level, Rajiv (2021) discussed computational studies and interpretations of bonding in metallopharmaceuticals.[29] As a result, this knowledge makes it possible to screen for and identify new drug-target associations that have the potential for novel applications.

The structural concept encompasses pharmacophore-based screening, docking[31], interaction similarity screening, binding site prediction, and other techniques. The improvement of fragment docking with binding mode information was researched by Esther Kellenberger et al. in 2019.[30] Later on, Bi(V)-MCs produced from streptomycin derivatives were molecularly modeled by Rajiv et al. (2015) through synthesis and spectroscopic research.[32] Researchers can identify new targets for repositioning candidates by contrasting the binding characteristics of different drugs in in silico screening based on 3D interaction data.[33] The concept of potential interactions based on medication placement has been employed in several research in the past. Ibrutinib was found to be a new inhibitor of the autoimmune-related target VEGFR2, but more recently, a fingerprinting technique based on the protein-ligand interactions profiler tool has been successfully used for many tasks, including the identification of repurposed medications for treatments.[24] The repositioning of amodiaquine as a cancer medication and the forecasting of new LRRK2 inhibitors are further tasks. The foundation of any pipeline for structure-based medication repositioning is the collection of structural data that describes the geometric conformations of medicinal substances.

Recent advancements in Parkinson's disease LRRK2-targeted treatment were studied by Nicolas Dzamko et al. in 2019.[34] Furthermore, it is hard to identify them structurally given that the vast majority of the millions of pharmaceutical compounds discovered in chemical libraries like Pubchem or ChEMBL lack structures for crystalline targets. The availability of structural data remains a significant obstacle to structure-based medication repositioning.[35] The concept of "building blocks" offers a different perspective on the problem. A drug compound is a group of specific chemical building components that give the medicine the intended properties. Theoretically, a large number of functional groups involved in protein binding are present in fragments, and many of these groups exactly match the target subpockets. Moreover, because to their diminished size and complexity, fragments enable efficient exploration of protein binding sites.[36] Recent research by Kozakov et al. (2017) showed that conserved binding modes are frequently present in the segments corresponding to low-energy hot spots.[37] Later, Drwal et al. (2019) carried out a comprehensive analysis of PDB on a large scale in order to get a full knowledge of fragment binding to ligand-able targets.[38] Fragments and drug-like superstructures that have binding sites connect to the same protein. Giordanetto et al. (2019)[39] investigated the

deposited protein structures with bound fragment hits in-depth, and their results are consistent with the notion that the majority of the fragment-hit complexes are stabilized by attractive contacts like hydrogen bonds, water bridges, and coordination bonds to catalytic metal ions.

PYTHON- BASED INFORMATICS

The identification of novel natural and synthetic chemicals is becoming more and more crucial in today's data-driven world, and silicochemical processing has become an essential part of biological and chemical research. PIKAChU is a Python-based informatics tool for assessing chemical units, according to Barbara R. Terlouw et al. (2022).[40] Bioactivities and biological pathways can predict the structures of molecules. COCONUT online, collection of open natural products database was utilized by Maria Sorokina et al. in 2021.[41] The chemical structure predicts medicinal qualities. Chemical databases like PubChem, NP Atlas, ChEBI, and COCONUT are linked to novel metabolites.ChEBI in 2016: improved services and an expandin g collection of metabolites was discussed by Janna Hastings et a l. (2016).[42] More, support vector machines were used by Jonathan Alvarsson et al. (2016) to explain large-scale ligand-based predictive modeling.[43] These studies depend on trustworthy cheminformatics tools that carry out basic chemical operations, such substructure matching, fingerprint-based similarity searches, molecular visualization, and chemical visualization for machine learning.

Chemical data formats, such as one-dimensional to three-dimensional molecular representations, may be read by cheminformatics kits to begin the molecule processing process. One of these forms is the simplified molecular input line entry system (SMILES). Smiles drawer, parsing and drawing SMILES-encoded molecular structures using client-side javascript was described by Jean-Louis Reymond et al. (2018).[44] ScottiSistematX is a web-based cheminformatics application developed by Marcus Tullius et al. (2019) for managing secondary metabolite data.[45] It comprises details on a molecule's stereochemistry, connectivity, composition, and atomic charge and represents a molecule as a one-dimensional string. The coordinates of an atom in three dimensions as well as the aforementioned attributes are stored in text files using more complicated formats like PDB and MOL.[46] Depending on the application, several forms and follow-up processing are necessary.

Comprehensive cheminformatics kits have gathered into sizable software libraries because to the large number of potential chemical research; yet, these libraries can be difficult to browse and rely on a great number of dependencies, making it challenging to use them in software packages. For users who should perform basic in silico investigations like reading in SMILES, sketching a molecule, or visualizing a substructure, the cost-benefit analysis of the time required to acquire and integrate these cheminformatics tools into a codebase performing real analyses is consequently disproportionate. A well-known open-source cheminformatics toolkit with these problems is RDKit.[47] The fact that RDKit is written in both Python and C^{++} and has so many dependencies that it frequently makes it difficult to integrate RDKit into other programs disproportionately increases the number of libraries that must be installed despite the fact that it is a lightning-fast and powerful library that supports a huge range of potential chemical operations. Due to this, RDKit is great for in-depth in silico research like computing a molecule's 3D conformers or producing electron density maps, but a bit heavyweight for the fundamental operations required by the majority of scientists in cheminformatics and bioinformatics. Another common cheminformatics toolbox is CDK. The Cytoscape claim chemViz2, the COCONUT database[41], and the scientific workflow software KNIME have all successfully employed it for molecular processing. Workflow-driven cheminformatics was described by Stephan Beisken et al. as KNIME-CDK.[48] (Konstanz Information Miner) It was created in Java and is highly suited for use in online applications. However, considering that Python is becoming the programming language of choice for researchers, particularly those working in the developing field of (deep) neural networks, CDK is not necessarily the ideal choice.[49]

To make fundamental cheminformatics processing more accessible to Python programmers, the authors provide PIKAChU, a Python-based informatics tool for analyzing chemical units. PecanPy, a quick, effective, and parallelized Python version of node2vec, was introduced by Renming Liu et al. in 2021.[50] Thus, PIKAChU is a flexible cheminformatics toolkit with some dependencies. The Python implementation of the subtype and stage inference methods, pySuStaIn, was explained by Leon M. Aksman.[51] It can visualize chemical structures, carry out extended connectivity finger-printing and infrastructures in Matplotlib, parse compounds from SMILES, and carry out Tanimoto similarity searches, with an emphasis on fundamental chemical processes with a focus on natural product chemistry. Because PIKAChU just needs the most elementary chemical processing, researchers believe it will be a useful substitute for many Python-based bio- and cheminformatics applications and databases.

The algorithm

To discuss the algorithm[52] applied for our purpose, we need to follow eight steps, as described below:

Step-1: Necessary program is launched from the__init.py script from a local RDKit environment. This script accesses SMILES strings of the cycloadducts saved in the SMILES.txt file, reads user input from the main configuration file (da. ini), and contains flags that must be set according to the type of jobs the user will like to carry out.

Step-2: SMILES strings accessed are converted into *mol* objects, which are then used to locate and keep track of the reactive sites (RS). As such, an ordered list of atomic indices is returned, in which four atoms originate from diene and two others from the dienophile. In case, cycloadducts have more than one cyclohexene substructure, a $2 \times n$-shaped list is returned, where n is the number of cyclohexene substructures; and its outcome is dealing within competing paths that are treated separately.

Step-3: 3-dimensional geometry of cycloadducts in Cartesian coordinates are obtained using a sequential procedure, including the embedding of its *mol* object and optimization of the returned conformer using UFF force-field.[53] Kindly note that, UFF is a broadly applicable force field that contains parameters for almost all the atoms of the periodic table. This guarantees no error is returned when studying a system with such uncommon atoms like actinides due to inexistent force field parameters. However, since UFF is a non-reactive force field, the topology of the system under investigation is kept intact during this conformation search, preventing any bond cleavage or formation.[54] Additionally in case of failure, the procedure is repeated, this time looking for more conformers (up to 60) and increasing the number of runs (up to 2000). Selected conformer is used as input in constrained optimization towards the pseudo-guess of the TS as briefly describes in Step-4.

Step-4: A constrained optimization in internal coordinates coerces the cycloadducts to adopt a symmetric two-fragment configuration where two pairs of terminal C atoms from the diene and dienophile are separated by 2.15 Å. Please note that the positions of these C atoms are retrieved from the List Atoms Int obtained at the procedure described in Step-2. The default separation distance of 2.15 Å can be modified by the user by simultaneously editing the ini.da configuration file and the Gen_gjf_file_ts() method of the Geom_3D.py module. Using default settings, this optimization returns 16 successive configurations of the system, of which the highest energy structure (a 2-fragment structure for intermolecular DA reactions) corresponds to the pseudo-guess TS. The latter is isolated and cleaned up at the same level of theory using the TS single-ended Berny algorithm.[55] This gives rise to the guess-TS. It is observed that the PM6 semi-empirical method has been found to perform well at this step.

Step-5: For every system, a new TS calculation is performed to refine the previous guess structure at a user-defined quantum mechanics level of theory. This step is followed by a vibrational check to make sure that the predicted stationary point is a real TS. This check is meant to assure that the returned TS has only one imaginary frequency. For the same, we examine (extract and count) normal vibrational modes of the system. Only structures with a unique negative (imaginary) frequency are retained as actual TSs. Rejected stationary points are automatically copied to an appropriate folder named ERROR_FILES. Steps 3–5 are revealed in for a small set of three randomly selected systems. Additional S1 depicts (optimized) geometries of the associated reactants (diene and dienophile), whose SMILES strings were first generated by applying the process retro Diels Alder function of the retro_DA() module to the cycloadducts SMILES, before being sequentially optimized at the PM6 and B3LYP/6-31G(d) levels respectively.

Step-6: If the TS has been located, we can determine the path for IRC. Details about the theory level or the number of IRC points are defined by the user in the da.ini file. We have observed that at least 60 points per IRC direction from the TS are sufficient to obtain a good IRC path (for mid-size systems with heteroatoms) linking the reactants to cycloadducts through the TS. Geometries of the reactants and cycloadducts can be optimized alongside that of the TS (during steps 2–4), if the RC_FLAG in the da.ini is set to 1.

After the IRC calculations, a separate script named myIRCAnalyzer.py can be executed to perform reaction force analyses (RFA, step7), and atomic decompositions of the reaction force and reaction force constant (as discussed in Step-8 below) for specific reactions. Details about the system to analyze or the atoms to consider in the decomposition must be given in a separate configuration file (analysis.ini). Description and usage of all keywords (and sections) found in the analysis.ini configuration files are discussed.

Step 7 A special module (namely RFA, see [56] for details) has been integrated to the package for executing all the calculations related to the reaction force analysis. Two important quantities of theory are the reaction force F and reaction force constant κ, which are defined by Equations (1) and (2), as given below, where E is the system's energy along the IRC path ξ. Torro-Labé and his co-workers (2009) have provided strong evidence showing that F and κ can help gain insight into the mechanism of several reactions.[57]

$$F_\xi = -\frac{dE}{d\xi}, \quad (1)$$

$$k_\xi = \frac{d^2E}{d\xi} = -\frac{dE}{d\xi}. \quad (2)$$

where all symbols are having their usual meanings. Also, F and κ are numerically calculated at each point of the IRC path using the finite difference approach. Technically, an average value of forwarding and backward derivatives at each given point is used as a good estimation of the exact derivative, except for the first and the last points of the IRC path. Any attempt to run this analysis will be ignored if the RFA_FLAG in the da.ini file has not been set to 1.

Step 8 Komorowski, et al. (2016) described atomic resolution for the energy derivatives on the reaction Path.[58]. In the reaction force F and force constant κ can be decomposed into atomic contributions by introducing the Hellman–Feynman[59] forces acting on each nucleus in the standard definition of F and κ [see Equations (3)–(4)]

$$F_\xi = -\frac{dE}{d\xi} = -\sum_{A\in M}\frac{\partial E}{\partial R_A}\frac{\partial R_A}{d\xi} = \sum_{A\in M} F_A\frac{\partial R_A}{d\xi} = \sum_A F_A(\xi), \quad (3)$$

$$k_\xi = -\frac{dE}{d\xi} = -\sum_{A\in M}\frac{\partial}{d\xi}\left[F_A\frac{\partial R_A}{d\xi}\right] = \sum_A k_A(\xi). \quad (4)$$

where all symbols are having their usual meanings. Moreover, κ_ξ can be split into two components originating from the atoms and bonds of the molecule as described in the following equation (5).

$$k_\xi = \sum_A^N k_{AA}(\xi) + 2\sum_A^N\sum_{B<A}^N k_{AA}(\xi) = k_\xi^{atoms} + k_\xi^{bonds}. \quad (5)$$

where all symbols are having their usual meanings, and the sums run over all the atoms in the molecule. We have also incorporated in the AMADAR package a module called RFD, which implements in above Equations (3)–(5) in case of DA reactions. To perform the series of decomposition analyses available in the module, the RFD_FLAG in the da.ini file must be set to 1 before running the myIRCAnalyzer.py script.

Calculation of a single reference PADIF

After the individual references, PADIFs are extracted from the GOLD rescoring files, usually, they are combined into a single, median reference PADIF. Hence, for each PADIF element, the median of all respective values in the reference PADIFs is calculated for negative values so that only favourable interactions are considered.[60] In case, all the reference values for a certain element are zero or positive, the respective element value is set to zero. Additionally, a weighting matrix is generated which assigns weighting factors to the elements depending on how often respective interaction occurs in the reference PADIFs (for instance, if it occurs in four of ten complexes, the weighting factor is 0.4).

Mathematical Computation for Scoring

The PADIF based scoring follows are being computed by applying the following steps: (I). Compute the R elements (m, n) whose value is less than zero in the reference PADIF (in case favourable reference interactions). (II). Compute the P elements (m, n) whose value is less than zero in the pose PADIF (favourable interactions in the pose fingerprint).[61] (III). Compute the maximum possible Overlap O_{max} (= P/R) between reference and pose PADIF, but at maximum. (IV). For the R elements (m, n) check the respective values in the pose PADIF and compute the individual elements score S(m,n) by following small steps;

(i) S(m, n) = w(m, n) if pose PADIF(m, n) <0.
(ii) S(m, n) = 0 if pose PADIF(m, n) = 0.
(iii) S(m, n) = − w(m,n) if pose PADIF (m, n) > 0.

1. Compute the value for actual Overlap
   O_{real} = (P∩ R)/R.
2. Compute the value for relative Overlap
   O_{rel}= O_{real} /O_{max}

Compute the total score S_{tot} (= ΣS(m, n)−(1.0 − Orel)·|ΣS(m, n)|) by summing up the individual scores of all elements (for many unfavourable interactions, it might be a negative value) and decrease the total score depending on the deviation to a perfect overlap of 1.0.

As per our observation, authors have a general prediction that, for the combination with ChemPLP, ranking first contains only the best three percent of poses by ChemPLP followed by PADIF based ranking of the rest. Purpose of this was to combine the strength of both methods to yield desired outputs.

Cheminformatics, metabolites, and chemical structures

The names of structurally annotated metabolites were provided by the Metabolomics Workbench. We automatically tried the chemical structures for all 130 metabolites using the PubChem API.[62] All the structures were standardized according to the already published chemical curation protocols[63]. Ten, metabolite chemical structures were characterized using MACCS fingerprints[63] computed using the RDKit[47] in Knim [64]. A Pearson correlation coefficient cut-off of 0.9 was used to filter out the highly correlated bits in the fingerprints. Hierarchical clustering of metabolites based on their chemical structures encoded as MACCS fingerprints was performed according to Soergel distances (also known as Tanimoto distances) and the average linkage. The ggtree package[65] was used to create circular dendrograms, and clustering of the significant metabolites for the construction of multi-metabolite models proceeded in the same way. As a part of the hierarchical clustering procedure, the number of clusters (k) was selected to achieve a reasonable partitioning of the metabolites. We selected the k value that resulted in the highest average silhouette width (ASW) for cluster assignments. By maximizing the ASW, we aimed to find the most "natural" number of clusters in the data, in which cluster members are most like each other, and distant from members belonging to other clusters. In general sense, let a(i) be the average distance between metabolite i and all the other members of its cluster, and b(i) be the smallest average distance between metabolite i and the members of any other cluster.

Then, the silhouette for metabolite i, s(i), is expressed by the following mathematical equation:

$$s(i) = \frac{a(i) - b(i)}{\max\{a(i), b(i)\}}$$

Where all the symbols are having their usual meaning. However, the ASW is the average of the silhouette values for all the i metabolites.

CHEMINFORMATICS INTELLIGENCE: METABOLITE PROFILES

Metabolites research is advancing swiftly and appears to be catching up to proteomics and genome-based approaches, which have been more frequently applied in the search for disease biomarkers. The consequences of the imputation of missing values using statistical analysis across analytical metabolomics platforms and by the kind of biological matrix must be assessed, however, in order to deal with missing data. Cheminformatics technique was investigated by Hiroshi Tsugawa et al. (2019) to characterize metabolomes in stable-isotope-labeled organisms.[66] Building foundations, SOPs, and protocols are done in this manner. By including and integrating more contextual biological analogs like genomes and proteomics, we may gain a complete picture of the system under study. Currently, matching experimental spectrum data requires scanning several independent database resources in order to maximize chemical identification and have the optimum coverage. As well as spectrum and compound chemical properties, these databases should ideally provide data on biological activity obtained from multiple sources. By combining mass spectrometry cheminformatics and metabolome databases, Zijuan Lai et al. (2018) discovered metabolites.[67] For the highest degree of accuracy, it is recommended to personally annotate and amend records.

Structure-based elucidation of new compounds is a challenging, drawn-out process. Although this load can be lessened and the in-line combined analysis of higher-dimensional NMR studies with high-resolution MS can be improved using computationally aided tools and algorithms, precise identifications can still be made. In the approaching years, there will surely be advancements in the creation of comprehensive bioinformatics tools for accurate identification, metabolite spectrum libraries, algorithms, and biological elucidation of metabolite outlines. J. R. Ullmann et al. (1976)

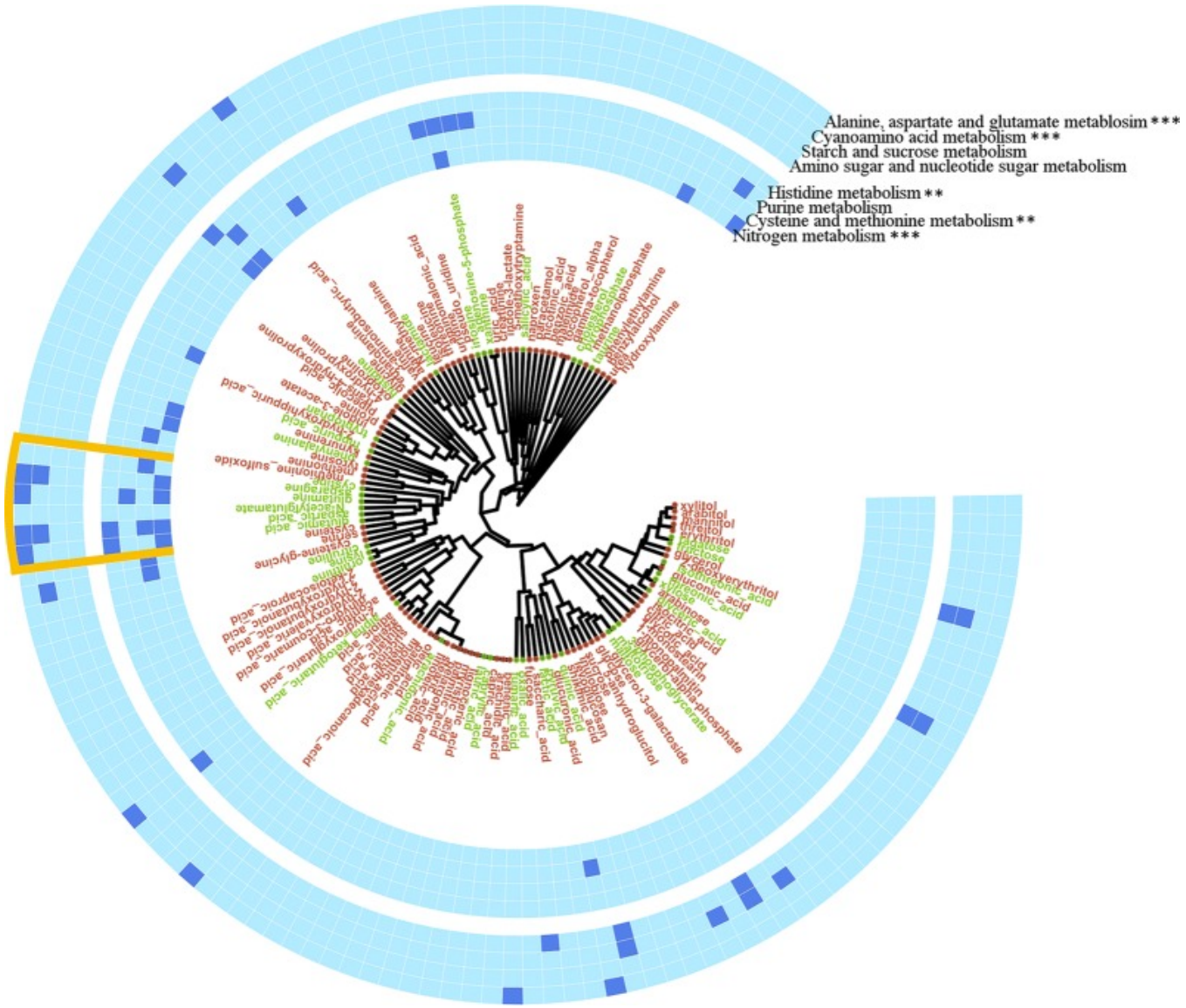


Figure 6. An illustration of metabolites identified in the differential analysis as significant, cells next to the names of those metabolites are shaded dark blue. Exploring and modelling trait-associated metabolite patterns using cheminformatics. If a metabolite was significant in at least one data set, named in green. Reprinted (adapted) with permission from.[70] Copyright (2019) Denis Fourches al. Journal of Cheminformatics. BMC Springer Nature.

demonstrated a subgraph isomorphism method.[68] These bioinformatics technologies will affect the development of new drugs and individualized medical treatments. Additionally, the current understanding of biology and disease causation is being illuminated by these cutting-edge approaches. An individual's metabolome serves as a representation of their molecular phenotype. Metabolomics and cheminformatics analysis of the antifungal function of plant metabolites were investigated by Miroslava Cuperlovic-Culfet al. in 2016.[69] Profiling metabolites, or small molecules with a molecular weight of 1500 Da, present in a given sample (such as serum, urine, and plasma) can be used to conduct in-depth research into various biochemical perturbations with internal (such as disease, microbiome, and drug metabolites) and external (such as exposome and drugs) origins. Metabolomic profiling stands out from other omics techniques due to its exceptional sensitivity to alterations in biological pathways with a mechanistic role in these biochemical activities. The idea that a small number of metabolites may be identified as disease biomarkers has given birth to a rapidly expanding collection of metabolomics investigations.

Cheminformatics technique was recommended by Jeremy R. Ashet al. (2019) for investigating and modeling trait-associated metabolite profiles.[70] Metabolomics has been used, for instance, to hunt for biomarkers for multiple sclerosis, colon cancer, and Alzheimer's disease.[60] Investigations on the efficacy, pharmacokinetic and pharmacodynamic properties, and toxicity of drug candidates and their metabolites typically employ metabolomics.[71] In the subject of pharmacometabolomics, the investigation of the function of metabolites in medicine response has also been useful.

Medicinal chemists employ metabolic profiling to more effectively assess a compound's potential for unfavorable side effects. They were able to investigate lead compound screening

in vivo as a result. In every metabolomics inquiry, chemical structure is the key to understanding the structure of the metabolites; nevertheless, it is typically underutilized in the trait association analysis that follows.[72] Recent research has shown that carefully considering the chemical makeup of metabolites can greatly improve our capacity to conduct in-depth analyses of metabolomics data.

In order to quantitatively and systematically characterize the structural properties of substances through the consistent computation of molecular descriptors, cheminformatics approaches were developed. Therefore, one may predict additional depiction of metabolites by calculating quantitative molecular descriptors to characterize their chemical structures **(Figure 6)**. As a result, a recent research that compared the chemical structures of medicines with endogenous human metabolites found that 90% of commercially available medications had a medium-to-high similarity (Tanimoto > 0.5) with their structurally most comparable human metabolite. The fact that metabolites in connected metabolic pathways share a common chemical structure has also recently been demonstrated by new algorithms that swiftly scan metabolic networks using chemical fingerprints.[73] The MetamapR network visualization tool has demonstrated that it is feasible to create hypotheses about the biological mechanisms behind an observed phenotype by grouping metabolites into chemical categories. The same research team has now released ChemRICH, a tool for identifying metabolites for enrichment analysis based on chemical similarities rather than biological annotation.

To the best of our knowledge, these strategies have not yet been applied with a predictive modeling methodology. Overall, the information included in metabolite chemical structures is vast, but they have not yet been used as the primary analytical and modeling framework for metabolomics datasets to create more comprehensible trait-metabolite linkages. The inability to annotate metabolic pathways substantially limits the use of different techniques (such as route or reaction pair databases) for figuring out the enzymatic links between metabolites, especially for understudied species.

Compared to multi-metabolite models, which take into account a variety of metabolite concentrations, single-metabolite models can predict an interest trait more accurately. Single-metabolite models are still often used in biomarker detection since multi-metabolite models typically lack interpretability. Small-molecule metabolites have been examined in order to identify biomarkers and potential treatment targets.[74] Metabolites that are biochemically related and that have the same trait-metabolite linkages are likely to be grouped if metabolites are bound for multi-metabolite models based on their structural similarity.

In biological processes, enzymes make it easier for molecules with comparable chemical properties to interact. It makes perfect sense given that the biological impact of interest typically acts at the level of biochemical pathways.[75] This technique may boost the output of single-metabolite models while keeping their intended interpretability because the resulting models may still reveal pathways mechanistically associated to the characteristic of interest.[70] Following this procedure, further methods (targeted metabolomics and isotope labeling) may be used to further investigate the compounds discovered and the pathways leading to them. It could be able to interpret metabolomics data in novel ways and find trait-metabolite linkages that would have been overlooked with earlier approaches since the metabolites in these models are coupled biochemically.

As a result, scientists have developed a cheminformatics method that combines chemically-informed clustering with a multi-metabolite modeling strategy.[76] In a case study involving lung cancer with adenocarcinoma, this technique was applied. The primary goal must be to identify clusters of structurally similar metabolites linked to pathways that have mechanistic and/or important involvement in lung cancer. The scientists hypothesized that structure-based clustering of metabolites might assist in creating multi-metabolite classifiers that are more accurate, reproducible, and understandable for patients' cancer status than existing approaches.

CHEMINFORMATICS INTELLIGENCE: DRUG DISCOVERY

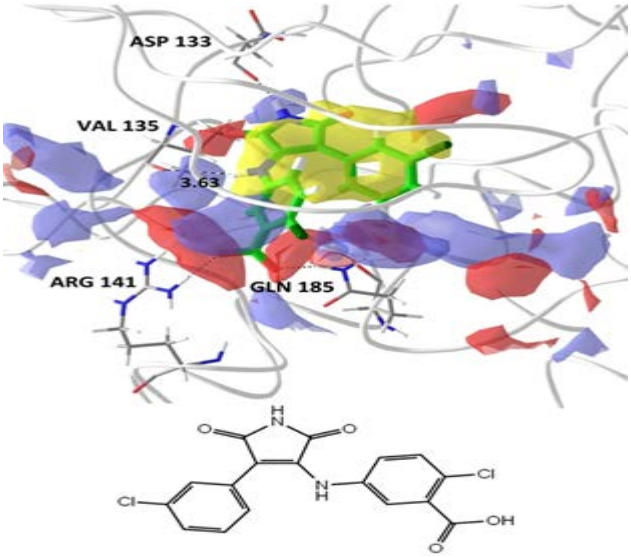
In the realm of drug development, the use of machine learning algorithms has grown increasingly widespread. Artificial neural networks with several hidden processing layers provide the foundation for machine abilities to automatically extract features from input data, capture nonlinear input-output correlations, and demonstrate deep learning capabilities. It has been demonstrated that deep learning algorithms offer a number of benefits over traditional machine learning techniques that rely on manually created chemical descriptors. Neural networks, an overview of early research, contemporary frameworks, and fresh difficulties were all explained by Ignacio Rojaset al. (2016).[77] Due to the relatively late rise in interest in deep learning, there has already been an unmatched explosion in the use of innovative modeling approaches and applications in the pharmaceutical industry.

Numerous areas of the chemical sciences have already benefited from the deep learning industry's ongoing improvements. This opinion post discusses key elements that have enabled deep learning techniques to flourish and, in some cases, outperform the industry's accepted practices for chemoinformatics. The use of artificial intelligence in drug discovery: current advancements and future possibilities was discussed by José Jiménez-Luna et al. in 2021.[78] Structure-based modeling, de novo molecule design, synthesis prediction, and property-relation/quantitative structure-activity based on ligands are specifically covered.

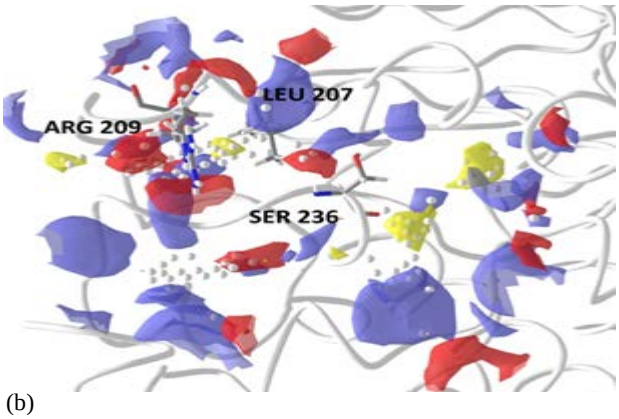
The authors also discuss the drawbacks of current artificial intelligence (AI) in each area under consideration and forecast how it might change the field of computer-aided drug discovery in the future. There is evidence that AI applications are beginning to be widely used in the drug research and design process. With notable developments in QSAR modeling, de novo molecular design, synthesis planning, and others, these techniques are gradually living up to some of the community's expectations. Direct steering of de novo molecule creation with

descriptor-conditional recurrent neural networks was debated by Esben Jannik Bjerrumet al. in 2020.[79] It still must be proven whether these methods can help scientists create and synthesize "better drug candidates faster and ultimately be helpful.

In the context of ligand-based property prediction, approaches depending on more "raw" chemical representations (such as graph neural networks and SMILES-based recurrent neural networks) can be anticipated to perform at least as well as traditional descriptor-based models. Deep learning approaches also make it possible to use data more efficiently, for example, through multitasking and online learning.[44] These methodologies are also easily adaptable to a larger range of chemical entities and modeling tasks. Contrarily, conformation-aware deep learning is still in its early stages, particularly when considering the methods that take three-dimensional symmetries into consideration while designing their algorithms **(Figure 7)**. Nevertheless, it is reasonable to expect that their use in drug discovery and related disciplines like quantum physics and material science will advance quickly, especially if they serve as a substitute for the computationally more demanding first-



(a)



(b)

Figure 7. Hydrophobic map–Yellow map; Hydrogen bond donor map–Blue map; Hydrogen bond acceptor map – Red map. Figure Comparison of the surface maps of pockets of GSK-3β identified by SiteMap. (A) Druggable pocket (ATP-binding site) (B) Difficult pocket (Allosteric site)

principle computations.[80] Rule-based and rule-free methods to de novo drug creation have gradually gained in favor over the past few years. These later compounds have limitations, such as a restricted capacity for synthesis, yet they exhibit promise for exploring hitherto unknown chemical space. A suitable solution could be provided by combining rule-free and rule-based techniques (sometimes referred to as "hybrid").

A specific emphasis will be given on generative techniques that may make use of fresh data sources, such as the groundbreaking research involving gene expression, conformational space, and ligand binding site data. On the basis of gene expression patterns relevant to various illness stages, Tareq B. Malas et al. (2020) discussed new ADPKD treatment candidates.[81] Automated synthesis planning and reaction prediction will continue to be inspired by and propelled by advances in natural language processing. There will be much-needed focus on issues like yield estimate, predicting favorable reaction conditions, and side-product formation.[82]

Robotics and reinforcement learning advancements will provide the groundwork for completely automated synthesis during the next few years. With the rebirth of interest in explainable AI, which encompasses techniques like feature attribution, instance-based molecular counterfactual explanation, and uncertainty estimation, the use of AI-supported drug development will rise. To create and validate these solutions, further multidisciplinary research is required.[83] Additionally, techniques that may employ data in low-data regimes, such as transfer learning, multitasking, and meta-learning, will be given significant attention. The challenges to learning about and then putting into practice deep learning techniques have greatly diminished in recent years for practitioners who are interested. This trend suggests that these approaches will soon become more widely available due to the ongoing creation of extensive, high-level research, the distribution of software packages, and comprehensive documentation.

BINDING SITE CHARACTERIZATION

In structure-based drug design, analyzing a target protein's potential binding sites is the first step. The target binding site's druggability is a vital step that may be determined by several in-silico pocket detection programs. The SiteMap program from Schrödinger, Inc. is the most effective instrument for determining druggability. With the use of tiny chemical probes typically molecules of methane or water that are docked on a specific target protein, SiteMap is an energy-based technique that identifies binding sites and determines whether a certain protein region interacts with its environment positively. SiteMap can discriminate between a target protein's "druggable," "difficult," and "undruggable" locations.[25] An analysis of the druggability of the glycogen synthase kinase enzyme was recently conducted using a dataset of more than 24 distinct X-ray crystal structures that were retrieved from the protein data bank. Based on the druggability score (Dscore), SiteMap revealed two potential binding sites for GSK-3.

A protein's binding sites can be divided into "druggable" (Dscore > 0.98), "undruggable" (Dscore 0.83), and "medium druggable/difficult" (Dscore between 0.83 and 0.98) sites based on the Dscore criterion. The allosteric pocket of GSK-3 simply didn't pass the druggability test, while the ATP-binding site was shown to be a druggable location. The maps of both sites in GSK-3 provide a visual representation of the capacity of SiteMap to discern between a druggable and problematic location.[25] The ATP-binding site may be recognized from the allosteric site's dispersed hydrophobic area by a clearly defined hydrophobic region (yellow map). Both in terms of site identification and druggability prediction, these approaches are more accurate.

SIMILARITY SCREENING

Shape-based approaches via virtual screening have been successful for lead generation strategies. Computational approaches with combined shape-based and electrostatic similarity can bind to a given protein and are thus valuable methods for lead identification.[84] A recent progress in this regard was the identification of a potent fibrinolysis inhibitor.[85]

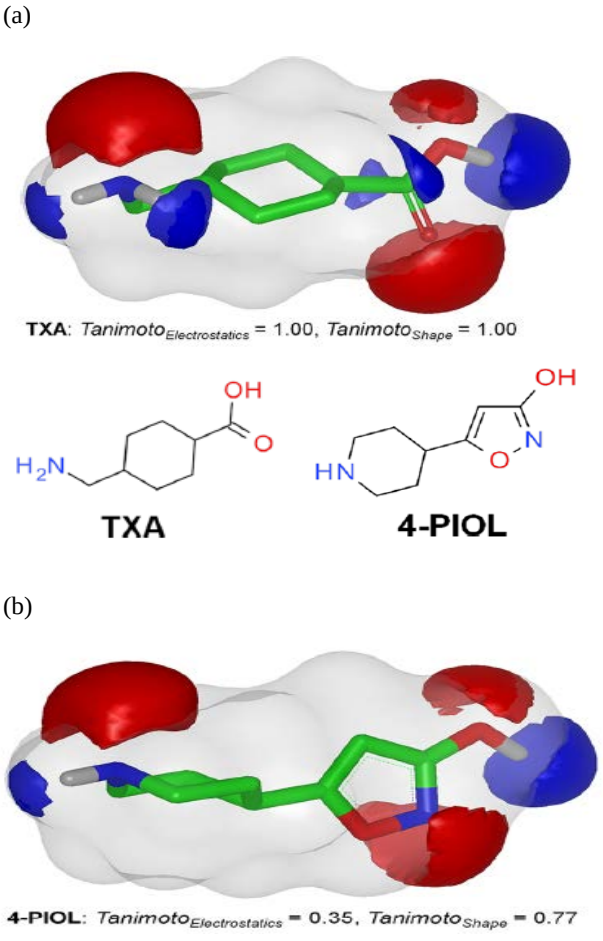


Figure 8. A representation showing the shapes and the electrostatic potentials for TXA and 4-PIOL. Red color denotes electronegative areas, whereas blue color shows electropositive areas.

The applied computational technique with a high-quality low-throughput screen identified 5-(4-piperidyl)-3-isoxazolol (4-PIOL) as a potent plasminogen binding inhibitor with the potential for the treatment of various bleeding disorders. Remarkably, 4-PIOL was found to be more than four times as potent as the drug TXA **(Figure 8)**. The calculated Tanimoto values show that 4-PIOL is electrostatically very similar to TXA. The similarity screening strategy identified 4-PIOL as a potent compound in the Clot-Lysis buffer assay with an IC50 of 2.8 μM. Similarity screening methods with combined shape-based and electrostatic similarity can generate potent lead compounds.

CONCLUSION

Cheminformatics intelligence has produced useful knowledge and predictions that have been used to shorten the time needed for screening compounds. In order to analyze complex molecular models, computational techniques have become a powerful source of extra traditional experimental procedures. By showcasing how current developments in computational techniques have created sophisticated computational models and how these cheminformatics intelligence procedures have been used to the desired outcomes These initiatives have showed a lot of promise, reduced human work by removing prejudice, and expedited the use of computational models at the same time.

The authors provide a view on the current use of reported methodology and how these approaches may alter future methodologies in this article. Additionally, contributors express their views on how issues developed, persisted, and eventually offered solutions for each of them. Therefore, various techniques for computational and cheminformatics intelligence include (I) reaction mechanism-based themes, which compute all necessary elementary steps; (II) identify the steps that determine rate and selectivity and then use kinetic analysis to predict performance; (III) a descriptor-based approach, which uses physical and chemical considerations to identify molecular properties; and (IV) a data-driven approach, which applies via statistical analysis and machine learning. As said earlier, remedies based on artificial intelligence and ideal networking in several industries might be amazing ways to remedy the ongoing issue. These computational-based methods must be improved, tested again, and trained to function fast and efficiently.

Numerous domains, including robotics, algorithms, high computing, system and process optimization, molecular modeling and simulation, and cheminformatics intelligence, have adopted recent developments and contemporary cheminformatics intelligence tactics.[86] Additionally, these computational techniques were used in biomedical imaging, protein structure prediction, modeling of biosets, and intelligent manufacturing systems. These computational methods go deeper into the study of optical communication systems, swarm robotics, intelligent sensor design, environmental restoration, ecological engineering, industrial system optimization, and hazardous material detection methods. In expert system design, environmental control and monitoring, interaction paradigms, environmental sustainability, bioinformatics, and biomedical engineering, several technologies generated via

cheminformatics intelligence are used. In order to have a more sophisticated perspective and flawless outcomes, distributed artificial intelligence, image processing, and machine intelligence all utilize greater computational intelligence.

In different areas such as nanoscience and advanced computing, knowledge-based simulation, computer games, intelligent photonics, multidimensional signal processing, intelligent controllers, information theory and coding, and mobile databases, computational intelligence has been applied to better outputs. Further, during toxicity assessment, epidemiological studies, developing antennas, sensor networking, adaptive intelligent systems, engineering intelligent networks, cryptography, and pattern recognition[87] discussed computational analysis methods and tools smeared. In the fields of knowledge discovery, optical engineering, photonics, gene expression analysis, bio-imaging, signaling, computation, mechanics, and computer-assisted medical diagnostic systems, computational intelligence has explored various dimensions. In the areas of environmental pollution, remediation, quality guidelines, wastewater, sludge treatment, industrial wastewater treatment, solid waste management, and recycling, the cheminformatics tools and intelligence have been applied to yield perfect results. Innovative intelligent systems, image and information processing, and medical innovative technologies have been developed via cheminformatics intelligence and used for trend analysis of innumerable features of molecular chemistry.

Author Contributions

Dr. Rajiv Kumar drafted the work and figures. Prof. M. P. Chaudhary and Dr. Navneet Chauhan contributed to the writing and revised it critically for important intellectual content and approved the final draft.

Acknowledgments

Lead author (Rajiv Kumar) gratefully acknowledges his younger brother Bitto for motivation.

Availability of data and materials

Wherever necessary, relevant citations included in the reference section.

Competing interests

The author has declared that no competing interest exists.

References and Notes

1. Wishart, D. S. Introduction to cheminformatics. *Curr. Protoc. Bioinforma.* **2016**, (2016).
2. Rajiv Kumar. Molecular Entities of Antimicrobial Drugsand Resistance Mechanism Underlying: Bioaccessibility, Bioavailability, and Bioaccumulation. *Int. J. Bioorganic Med. Chem.* **1**, 17–19 (2021).
3. Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O. & Walsh, A. Machine learning for molecular and materials science. *Nature* vol. 559 (2018).
4. Lu, J. *et al.* Transfer learning using computational intelligence: A survey. *Knowledge-Based Syst.* **80**, (2015).
5. Ul-Haq, Z. & Madura, J. D. *Frontiers in Computational Chemistry*. *Frontiers in Computational Chemistry* vol. 1 (Elsevier Inc., 2015).
6. Turner, K. Reasons to craft your own molecular models. *Education in Chemistry* vol. 56 (2019).
7. David, L., Thakkar, A., Mercado, R. & Engkvist, O. Molecular representations in AI-driven drug discovery: a review and practical guide. *Journal of Cheminformatics* vol. 12 (2020).
8. Ma, Q. & Wang, Y. Application of embedded system and artificial intelligence platform in Taekwondo image feature recognition. *J. Ambient Intell. Humaniz. Comput.* (2021) doi:10.1007/s12652-021-03222-9.
9. Rajiv, K. & Rajni, J. Computational approach on architecture and tailoring of organic metal complexes derived from streptomycin and Zn, Cd and Pb: Antimicrobial effectiveness. *Appl. Organomet. Chem.* **25**, (2011).
10. Elmassry, M. M., Kim, S. & Busby, B. Predicting drug-metagenome interactions: Variation in the microbial β-glucuronidase level in the human gut metagenomes. *PLoS One* **16**, (2021).
11. Makhouri, F. R. & Ghasemi, J. B. Combating Diseases with Computational Strategies Used for Drug Design and Discovery. *Curr. Top. Med. Chem.* **18**, 2743–2773 (2019).
12. Fernandes, M., Sanches, B. & Husi, H. Cheminformatics and Computational Approaches in Metabolomics. in *Computational Biology* (2019). doi:10.15586/computationalbiology.2019.ch9.
13. Schaduangrat, N. *et al.* Towards reproducible computational drug discovery. *Journal of Cheminformatics* vol. 12 (2020).
14. Mao, J. *et al.* Discovery of microtubule stabilizers with novel scaffold structures based on virtual screening, biological evaluation, and molecular dynamics simulation. *Chem. Biol. Interact.* **352**, (2022).
15. Duarte, Y., Márquez-Miranda, V., Miossec, M. J. & González-Nilo, F. Integration of target discovery, drug discovery and drug delivery: A review on computational strategies. *Wiley Interdisciplinary Reviews: Nanomedicine and Nanobiotechnology* vol. 11 (2019).
16. Kumar, R. & Mishra, P. Bi(V) organic framework in an asymmetric system: Synthesis, spectroscopic, XRPD and molecular modeling. *Main Gr. Chem.* **6**, (2007).
17. Rajiv Kumar, Tanya Chauhan, Anil Kumar, P. M. Ecology of supramolecular entities of antibiotics (SMEAs), on-demand controlled guest capture and release systems: tuning spacer and polarities in abiotic and biotic factors. *Integr. Res. Adv.* **4**, 61–68 (2017).
18. Orekhov, F. & Gradov, O. MultiSPAS: Multi-Descriptor Physical Approaches to Computerized Plant Ecology. in *Smart Innovation, Systems and Technologies* vol. 247 295–333 (Springer Science and Business Media Deutschland GmbH, 2022).
19. Wishart, D. S. Identifying putative drug targets and potential drug leads: Starting points for virtual screening and docking. *Methods Mol. Biol.* **1215**, 425–444 (2015).
20. Kumar, R. & Mishra, P. Spectroscopic, thermal and X-ray powder diffraction patterns of Bi(V) complexes with [2S-[2α, 5α, 6β(S*)]-6-[amino(4-hydroxyphenyl) acetylamino]-3, 3-dimethyl-7-oxo-4-thia-1-azabicyclo[3,2,0] hepatane-2-carboxylic acid and (6R)-6-(α-phenyl-D-glycyamino) p. *Main Gr. Chem.* **7**, (2008).
21. Kumar, R. & Chhikara, B. S. Organometallic assemblies: π-electron delocalization, μ-bridging spacers, flexibility, lipophilic nature, bio-accessibility, bioavailability, intracellular trafficking pathways and antimicrobial assimilation. *J. Organomet. Chem.* **776**, (2015).
22. Kumar, N. & Acharya, V. Machine intelligence-driven framework for optimized hit selection in virtual screening. *J. Cheminform.* **14**, 48 (2022).
23. Rajiv Kumar, Kiran K Gulia, Mina M Chandra, Anil K Aggarwal, Anil Kumar, Sandeep Mittan, P. M. Biomolecular interaction simulation of supramolecular topologies of organometallic assemblies of Bi(V) with antibiotic Tetracycline Amoxicillin drugs and their experimental activities evaluation. *J. Biomed. Ther. Sci.* **6**, 61–72 (2019).
24. Adasme, M. F., Bolz, S. N., Al-Fatlawi, A. & Schroeder, M. Decomposing compounds enables reconstruction of interaction fingerprints for structure-based drug screening. *J. Cheminform.* **14**, (2022).

25. Halgren, T. New method for fast and accurate binding-site identification and analysis. *Chem. Biol. Drug Des.* **69**, (2007).
26. Leelananda, S. P. & Lindert, S. Computational methods in drug discovery. *Beilstein Journal of Organic Chemistry* vol. 12 2694–2718 (2016).
27. Halgren, T. A. Identifying and characterizing binding sites and assessing druggability. *J. Chem. Inf. Model.* **49**, (2009).
28. Schmidt, T., Bergner, A. & Schwede, T. Modelling three-dimensional protein structures for applications in drug design. *Drug Discovery Today* vol. 19 (2014).
29. Kumar, R. Computational Explorations and Interpretation of Bonding in Metallopharmaceutical for a Thorough Understanding at the Quantum Level. *Sustain. Chem. Eng.* (2021) doi:10.37256/sce.212021828.
30. Jacquemard, C., Drwal, M. N., Desaphy, J. & Kellenberger, E. Binding mode information improves fragment docking. *J. Cheminform.* **11**, (2019).
31. Martin, S. J., Chen, I. J., Chan, A. W. E. & Foloppe, N. Modelling the binding mode of macrocycles: Docking and conformational sampling. *Bioorganic Med. Chem.* **28**, (2020).
32. Rajni Johar, Rajiv Kumar, P. M. Molecular modeling of Bi (V)-MCs derived from streptomycin derivatives: synthesis and spectroscopic studies. *J Integr Sci Technol* **3**, 18–21 (2015).
33. Kumar, R. Metal organic and covalent organic frameworks as antiviral drugs for intervention in viral infection and future prospective. *Front. Infect. Dis. Microbiol.* **1**, 1–2 (2021).
34. Zhao, Y. & Dzamko, N. Recent Developments in LRRK2-Targeted Therapy for Parkinson's Disease. *Drugs* **79**, (2019).
35. Rajiv, K. & Rajni, J. Spectral, structural elucidation and coordination abilities of Co(II) and Mn(II) coordination entities of 2,6,11,15-tetraoxa-9,17-diaza-1,7,10,16-(1,2)- tetrabenzenacyclooctadecaphan-8,17-diene. *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.* **79**, (2011).
36. Volkamer, A., Kuhn, D., Rippmann, F. & Rarey, M. Dogsitescorer: A web server for automatic binding site prediction, analysis and druggability assessment. *Bioinformatics* **28**, (2012).
37. Kozakov, D. *et al.* The ClusPro web server for protein-protein docking. *Nat. Protoc.* **12**, (2017).
38. Drwal, M. N. *et al.* Structural Insights on Fragment Binding Mode Conservation. *J. Med. Chem.* **61**, (2018).
39. Giordanetto, F., Jin, C., Willmore, L., Feher, M. & Shaw, D. E. Fragment Hits: What do They Look Like and How do They Bind? *J. Med. Chem.* **62**, (2019).
40. Terlouw, B. R., Vromans, S. P. J. M. & Medema, M. H. PIKAChU: a Python-based informatics kit for analysing chemical units. *J. Cheminform.* **14**, 34 (2022).
41. Sorokina, M., Merseburger, P., Rajan, K., Yirik, M. A. & Steinbeck, C. COCONUT online: Collection of Open Natural Products database. *J. Cheminform.* **13**, (2021).
42. Hastings, J. *et al.* ChEBI in 2016: Improved services and an expanding collection of metabolites. *Nucleic Acids Res.* **44**, (2016).
43. Alvarsson, J. *et al.* Large-scale ligand-based predictive modelling using support vector machines. *J. Cheminform.* **8**, (2016).
44. Probst, D. & Reymond, J. L. SmilesDrawer: Parsing and Drawing SMILES-Encoded Molecular Structures Using Client-Side JavaScript. *J. Chem. Inf. Model.* **58**, (2018).
45. Scotti, M. T. *et al.* Sistematx, an online web-based cheminformatics tool for data management of secondary metabolites. *Molecules* **23**, (2018).
46. Engel, T. Representation of Chemical Compounds. in *Chemoinformatics* (2003). doi:10.1002/3527601643.ch2.
47. Landrum, G. RDKit: Open-Source Cheminformatics Software. *Http://Www.Rdkit.Org/* (2021).
48. Beisken, S. *et al.* KNIME-CDK: Workflow-driven cheminformatics. *BMC Bioinformatics* **14**, (2013).
49. Zamil, A. M. A., Nusairat, N. M., Vasista, T. G., Shammot, M. M. & Yousef, A. PREDICTION OF SALES BASED ON AN EFFECTIVE ADVERTISING MEDIA SALE DATA: A PYTHON IMPLEMENTATION APPROACH. *Acad. Strateg. Manag. J.* **20**, (2021).
50. Liu, R. & Krishnan, A. PecanPy: a fast, efficient and parallelized Python implementation of node2vec . *Bioinformatics* **37**, (2021).
51. Aksman, L. M. *et al.* pySuStaIn: A Python implementation of the Subtype and Stage Inference algorithm. *SoftwareX* **16**, (2021).
52. Kamada, T. & Kawai, S. An algorithm for drawing general undirected graphs. *Inf. Process. Lett.* **31**, (1989).
53. Casewit, C. J., Colwell, K. S. & Rappé, A. K. Application of a Universal Force Field to Organic Molecules. *J. Am. Chem. Soc.* **114**, (1992).
54. Kumar, R. & Mishra, P. Spectral analysis, XRPD, quantum structures and evaluation of antimicrobial activity of supramolecular topologies of Bi(V) and Co(II) metal based antibiotic drugs complexes. *J. Integr. Sci. Technol. Vol 4, No 2* (2016).
55. Frisch, M. J. *et al. Gaussian 09, Revision E. 01, Gaussian Inc.: Wallingford CT. Gaussian Inc.* (2016).
56. Toro-Labbé, A. Characterization of chemical reactions from the profiles of energy, chemical potential, and hardness. *J. Phys. Chem. A* **103**, (1999).
57. Toro-Labbé, A., Gutiérrez-Oliva, S., Murray, J. S. & Politzer, P. The reaction force and the transition region of a reaction. in *Journal of Molecular Modeling* vol. 15 (2009).
58. Jedrzejewski, M., Ordon, P. & Komorowski, L. Atomic Resolution for the Energy Derivatives on the Reaction Path. *J. Phys. Chem. A* **120**, (2016).
59. Politzer, P. & Murray, J. S. The Hellmann-Feynman theorem: a perspective. *J. Mol. Model.* **24**, (2018).
60. Marechal, E. Editorial [Hot Topic: Chemogenomics: A Discipline at the Crossroad of High Throughput Technologies, Biomarker Research, Combinatorial Chemistry, Genomics, Cheminformatics, Bioinformatics and Artificial Intelligence Chemogenomics (Guest Editor: Eric Marech. *Comb. Chem. High Throughput Screen.* **11**, 582–582 (2008).
61. Bouysset, C. & Fiorucci, S. ProLIF: a library to encode molecular interactions as fingerprints. *J. Cheminform.* **13**, (2021).
62. Kim, S. *et al.* PubChem substance and compound databases. *Nucleic Acids Res.* **44**, (2016).
63. Fourches, D., Muratov, E. & Tropsha, A. Trust, but Verify II: A Practical Guide to Chemogenomics Data Curation. *Journal of Chemical Information and Modeling* vol. 56 (2016).
64. Durant, J. L., Leland, B. A., Henry, D. R. & Nourse, J. G. Reoptimization of MDL keys for use in drug discovery. *J. Chem. Inf. Comput. Sci.* **42**, (2002).
65. Yu, G., Smith, D. K., Zhu, H., Guan, Y. & Lam, T. T. Y. ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* **8**, (2017).
66. Tsugawa, H. *et al.* A cheminformatics approach to characterize metabolomes in stable-isotope-labeled organisms. *Nat. Methods* **16**, (2019).
67. Lai, Z. *et al.* Identifying metabolites by integrating metabolome databases with mass spectrometry cheminformatics. *Nat. Methods* **15**, (2018).
68. Ullmann, J. R. An Algorithm for Subgraph Isomorphism. *J. ACM* **23**, (1976).
69. Cuperlovic-Culf, M., Rajagopalan, N., Tulpan, D. & Loewen, M. C. Metabolomics and cheminformatics analysis of antifungal function of plant metabolites. *Metabolites* **6**, (2016).
70. Ash, J. R., Kuenemann, M. A., Rotroff, D., Motsinger-Reif, A. & Fourches, D. Cheminformatics approach to exploring and modeling trait-associated metabolite profiles. *J. Cheminform.* **11**, (2019).
71. Liang, G. *et al.* Inhibition of LPS-induced production of inflammatory factors in the macrophages by mono-carbonyl analogues of curcumin. *J. Cell. Mol. Med.* **13**, (2009).

72. Shen, F., Cheek, C. & Chandrasekaran, S. Dynamic network modeling of stem cell metabolism. in *Methods in Molecular Biology* vol. 1975 (2019).
73. Gasteiger J., Engel, T. *Applied Chemoinformatics: Achievements and Future Opportunities. Applied Chemoinformatics* (2018).
74. Bishai, J. D., & Palm, N. W. Small molecule metabolites at the host–microbiota interface. *The Journal of Immunology*, (2021). 207(7), 1725-1733.
75. Martínez-Zamudio, R. I., Robinson, L., Roux, P. F. & Bischof, O. SnapShot: Cellular Senescence Pathways. *Cell* (2017) doi:10.1016/j.cell.2017.07.049.
76. Nigam, M. Phytomedicine: Scope and current highlights. in *Preparation of Phytopharmaceuticals for the Management of Disorders: The Development of Nutraceuticals and Traditional Medicine* 39–54 (Elsevier, 2020). doi:10.1016/B978-0-12-820284-5.00013-7.
77. Prieto, A. *et al.* Neural networks: An overview of early research, current frameworks and new challenges. *Neurocomputing* **214**, (2016).
78. Jiménez-Luna, J., Grisoni, F., Weskamp, N. & Schneider, G. Artificial intelligence in drug discovery: recent advances and future perspectives. *Expert Opin. Drug Discov.* **16**, (2021).
79. Kotsias, P. C. *et al.* Direct steering of de novo molecular generation with descriptor conditional recurrent neural networks. *Nat. Mach. Intell.* **2**, (2020).
80. Dasgupta, A. Computational challenges for reliability assessment of next-generation micro & nano systems. in *7th International Conference on Thermal, Mechanical and Multiphysics Simulation and Experiments in Micro-Electronics and Micro-Systems, EuroSimE 2006* vol. 2006 (2006).
81. Malas, T. B. *et al.* Prioritization of novel ADPKD drug candidates from disease-stage specific gene expression profiles. *EBioMedicine* **51**, (2020).
82. Cheng, A. C. Chapter 2 Predicting Selectivity and Druggability in Drug Discovery. in *Annual Reports in Computational Chemistry* vol. 4 23–37 (Elsevier BV, 2008).
83. Kitano, H. Biological Complexity and the Need for Computational Approaches. in *History, Philosophy and Theory of the Life Sciences* vol. 20 (2017).
84. Kumar, R., Mishra, P., Singh, R. P. & Singh, R. P. Synthesis, spectroscopic, XRPD and computational study of transition metal-organic frameworks (TMOFs) derived from (6R)-6-(α-phenyl-D-glycylamino) penicillanic acid. *Russ. J. Inorg. Chem.* **54**, (2009).
85. Boström, J., Grant, J. A., Fjellström, O., Thelin, A. & Gustafsson, D. Potent fibrinolysis inhibitor discovered by shape and electrostatic complementarity to the drug tranexamic acid. *J. Med. Chem.* **56**, (2013).
86. Kumar, R. & Mishra, P. Main group metal-organic frameworks of Bi(V) and Pb(II) derived from 2-[4,6-diamino-3-[3-amino-6-(1-methyamino-ethyl) tetrahydropyran-2-yl] oxy-2-hydroxy-cyclohexoxy]-5-methyl-4-methylaminotetrahydropyran-3,5-diol and their molecular modeling. *Main Gr. Chem.* **6**, (2007).
87. Kumar, R. & Singh, S. Coordination patterns: Spectroscopic and electrochemical investigation on Mn(II) complexes with macrocyclic ligands derived from two different organic skeletons. *Russ. J. Coord. Chem. Khimiya* **33**, (2007).

Authors Biographies

Dr. Rajiv Kumar received a M.Sc. (Institute Topper) post-graduate with honors in chemistry from CCS University (1999) and a Ph.D. in macrocycles (nature molecules) from the University of Delhi (2004). His group is specialized in chemistry, nanoscience, and nanotechnology, computational chemistry, molecular modeling, computational nanotechnology, computational models of biomolecules, macromolecular and apoptotic pathways, atmospheric chemistry, spectroscopy, and biomarkers. He has published 112 research articles, 12 book chapters or review articles, and 2 theses. His book entitled *"Nanotechnology and Nanomaterials in the Treatment of Life-threatening Diseases" by the* author(s), N. Kumar and Rajiv Kumar, published by Elsevier Publication in 2013, covers the contents for researchers working in the areas of nanoscience and medicine, pharmaceutical industries, and practicing physicians as well as surgeons. Dr. Kumar has worked as a c*o-investigator* in the research project sponsored by DRDO, Ministry of Defense, India.
One of the articles by Dr. Rajiv Kumar et al. was featured on the cover page of MOLECULAR OMICS, ROYAL SOCIETY OF CHEMISTRY, 2021. Dr. Rajiv Kumar is also contributing as a reviewer and editor of scientific journals of international repute.

Prof. M. P. Chaudhary, Ph.D., is an Indian mathematician, who is currently working in number theory, combinatorics and special functions especially in the fields of interest to Srinivasa Ramanujan. Currently, he is serving as Albert Einstein Chair Professor of Mathematical Sciences at International Scientific Research and Welfare Organization (ISRWO), India. He had served as professor of mathematics at International Institution (in abroad) (supported by United Nations Development Programme-UNDP) during academic years 2012 to 2018-19. Recently, he had been offered for Advisor position at the National Institute of Educational Planning and Administration (NIEPA), (Deemed to be University), Ministry of Human Resource Development, Government of India, New Delhi, India. He initiated his research career at the department of mathematics, University of Delhi (1999-2005), also conducted his research works at Jamia Millia Islamia, New Delhi (2008-2011); and worked as Research Scientist at an external research centre of the Department of Sciences and Technology (DST), Government of India, New Delhi (2011). He was visiting researcher at the department of mathematics, the Pennsylvania State University, USA(2007); academic visitor at OCCAM / Mathematical Institute, University of Oxford, England, U.K. (between 2013-2018). He has more than 300 publications to his credits at both national and international level which includes more than 225 scientific research paper, 40 book(s) / monograph(s).

Dr. Navneet Chauhan received M. Pharm. in Medicinal Chemistry (2008) and Ph.D. (2018) from Institute of Institute of Pharmacy, Nirma University, Ahmedabad. His group specialized in Computational Chemistry and Medicinal Chemistry. He has been granted one Indian Patent and published many research and review articles in various reputed journals and guided five M. Pharm students. He has recently received a grant sponsored by Gujarat Technological University, Ahmedabad. His areas of interest include anti-tubercular drug discovery, anti-cancer drug discovery, microwave assisted organic synthesis and neglected disease drug discovery.