Available online at www.ijpcr.com
International Journal of Pharmaceutical and Clinical Research 2015; 7(6): 462-467

ISSN- 0975 1556

Research Article

E-Analogs: A Web-Based Cheminformatics Algorithm for Automated Ligand Analog Library Generation and In-Silico Drug Design

Renjith P[1*], Balakrishnan G[1], Suganya S[2]

[1]Department of Boinformatics, Liatris Biosciences LLP, Kakkanad, Kochi, Kerala – 682037, India
[2]Department of Computer Science & Engineering, St. Michael College of Engineering & Technology, Sivaganga, Tamil Nadu – 630551, India

Available Online:31st October, 2015

ABSTRACT
Making use of computational approaches like virtual screening of compound libraries are essential steps in drug discovery mission. Several thousands of ligands and their analogs need to be computationally modeled and tested during early stages of drug discovery in order to identify the best molecule with desired characteristics against the target protein which causing the disease. Manual sketching of ligands and their analogs is a time consuming approach with high degree of human errors involved. eAnalogs is a novel computational algorithm to override the manual methods. eAnalogs is a fast, reliable web-accessible software for construction of automated ligand analog libraries. The algorithm is based on simple random sample statistical approach to create each analog of pre-defined atoms or functional groups with equal selection probability. The de novo ligand design tool helps medicinal chemists to rapidly and accurately generate hundreds of analogs from basic 2-dimensional chemical scaffold. The algorithm is capable of generating multiple analogs from single input chemical structure by adding user-defined atoms, functional groups and ring structures to the selected positions in a combinatorial manner. Software allows users to download result library in PDF and Structure Data Formats (SDF) that can be further used for processing virtual High Throughput Screening (vHTS), Quantitative Structure-Activity Relationship (QSAR), Lead Identification and Optimization techniques. The eAnalogs software is accessible at: http://insilico.liatrisbio.com/.

INTRODUCTION

Computational approaches have an increasingly important role to play in drug discovery as can be seen by the increasing adoption of Cheminformatics and Bioinformatics methods[1]. In fact many Pharmaceutical companies and drug discovery research groups have already adopted computational virtual screening methods as an alternate or to complement conventional high throughput screening experiements[2]. Virtual screening libraries are typically collections of very large numbers of small molecules which are more drug like substances[3]. These large compound libraries are screened by advanced computational tools in order to identify new compounds that have desired activities against the drug target[4,5]. Novel compounds from these computational screening methods are most likely to be binding with the disease causing agent usually a receptor or an enzyme and activates or inhibits their biological functions[6,7]. Quality of such screening libraries is a crucial factor for the identification of best compounds that have the maximum probability of successfully making through the drug discovery pipeline[8]. Knowledge-based smarter computational methods not only help the pharmaceutical industry to cut down their research costs, but also to speedup introducing more drugs into the market for human welfare[9]. Computer Aided Drug Design (CADD) approaches often process multiple ligand scaffolds and their analogs for virtual bioactivity screening. Medicinal chemists search for the best analogue with optimum biological activity towards the target protein with desired absorption, distribution, metabolism, excretion, toxicity (ADMET) features. Several thousands of analogs are usually designed and screened for this purpose. Medicinal chemists frequently need to experiment with random functional groups and/or atoms and their combinations at multiple positions on scaffold in order to discover novel ligands. Most of the virtual screening libraries are enriched with several structural analogs of chemical scaffolds[10,11].

Structural analogs are compounds having similar chemical scaffold but differ in certain components like atoms, functional groups, or substructures[12]. Structural analogs have a chemical similarity but may vary in physical, chemical, biological, or pharmacological properties[13]. Manual methods of compound sketching are time consuming; require added man power and it indirectly affect the cost of drugs. Efficient computational methods can replace these manual methods and can speedup lead identification and optimization processes.

Present research explains and demonstrates all the aspects of the software eAnalogs, which is a web-accessible tool to automate ligand analog design and compound library construction. The aim for development of new software for drug design is to replace or complement manual curation efforts of medicinal chemists to construct chemical analog libraries for their drug discovery researches. Further this study is extended to explore the contributions of this tool in the ligand design and optimization studies.

*Author for Correspondence

Table 1: Fragments as choices for R1 (labeled “U”) and R2 (labeled “V”) positions.

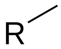
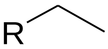
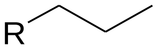
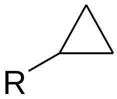
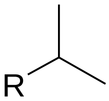
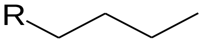
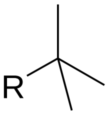
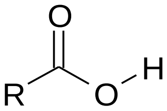
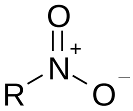
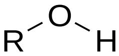
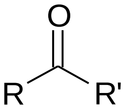
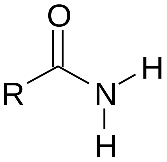
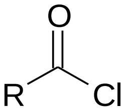
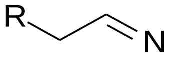
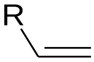
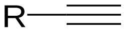
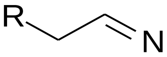
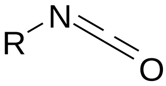
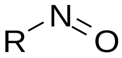
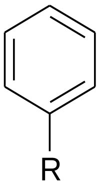
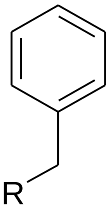
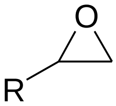
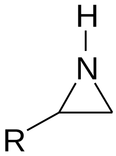
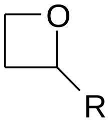
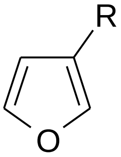
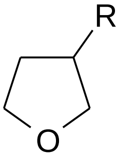
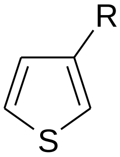
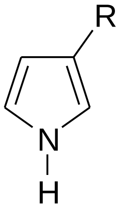
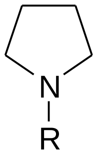
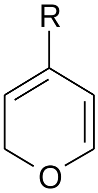
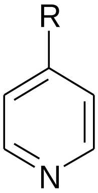
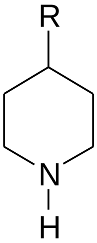
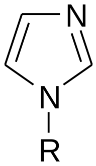
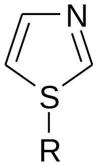
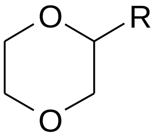
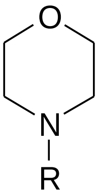
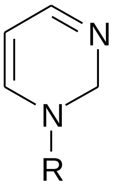
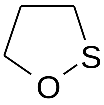
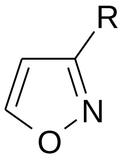
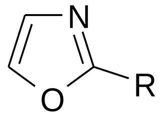
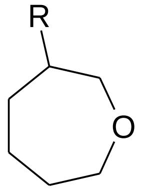
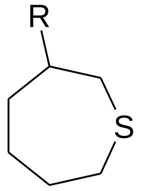
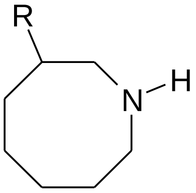
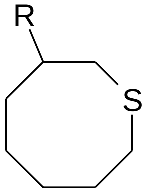
Atoms	R-N Nitrogen	R-O Oxygen	R-P Phosphorus	R-S Sulphur	
	R-F Fluorine	R-Cl Chlorine	R-Br Bromine	R-I Iodine	
Functional Groups	 Methane	 Ethane	 Propane	 Cyclopropane	 Isopropane
	 Butane	 Tertiary butane	 Carboxyl group	 Nitro group	 Alcohol group
	 Ketone group	 Amide group	 Acid chloride	 Carboxylate ion	 Alkene
	 Alkyne	 Ethanimine	 Isocynate	 Nitroso	
Heterocyclic Rings	 Phenyl group	 Benzyl group	 Ethylene oxide	 Ethylenimine	 Trimethylene oxide
	 Furan	 Tetrahydrofuran	 Thiophene	 Pyrrole	 Pyrrolidine
	 Pyran	 Pyridine	 Piperidine	 Imidazole	 Thiazole
	 Dioxane	 Morpholine	 Pyrimidine	 1,2-Oxathiolane	 Isoxazole

Table 1: Fragments as choices for R1 (labeled “U”) and R2 (labeled “V”) positions.

Atoms	R-N	R-O	R-P	R-S	
	Nitrogen	Oxygen	Phosphorus	Sulphur	
	R-F	R-Cl	R-Br	R-I	
	Fluorine	Chlorine	Bromine	Iodine	
	 Oxazole	 Oxepane	 Thiepine	 Azocane	 Thiocane

MATERIALS AND METHODS

Web Server

eAnalogs software is configured on a web server connected to the internet for online access. Users from worldwide can access the software using their personal computers connected to the internet by visiting the website, http://insilico.liatrisbio.com/. Internet browsers like Internet Explorer, Google Chrome or Mozilla can be used for accessing the web-based software. Users are requested to allow pop-ups or add-ons on their internet browsers in order to load the page successfully. The software is using a Java molecule sketcher applet for input molecule submission and is also required by the user to add the web server link to the Java security exception site lists.

Software Design

eAnalogs software consists of molecule input interface, algorithm processing phase and result output phase. Input interface enables user to provide input molecule structure in any of the three methods of manual sketching, inserting SMILE format, or importing molecule structure file in CML[17], MOL[18], or SDF[19] formats. As a next step, user needs to define maximum of two variable positions on the input molecule where fragments are to be added in a combinatorial manner and construct the analog library. These two positions are to be labeled as ‘U’ and ‘V’ using the atom label option on the sketcher applet. A list of atoms, functional groups and heterocyclic rings are listed on the input page in which user can select the desired fragments to be added to the ‘U’ and ‘V’ positions. A job name must be provided in order to submit the job for algorithm processing. Input molecule structure and parameters have been taken to the next level of processing phase from the input interface. Algorithm used by the software analyzes the input structure and attach chosen atoms, functional groups or heterocyclic rings to the user defined positions in a combinatorial manner. Once the process is finished, algorithm passes the newly generated analog structures to the output result page automatically. Result page displayed to the user lists the resulted analogs as individual structure images. Users can download the results in PDF or SDF formats using the download options provided on the output page.

Algorithm

eAnalogs software uses a proprietary algorithm in order to generate structural analogs in a fast and efficient manner. Every structural analog is created by attaching choosen atoms or functional groups to the pre-defined positions on the input scaffold based on a simple random sampling statistical approach[14]. User defined atoms and chemical groups are sampled with equal probability in order to create analogs with equal structural distribution. Several iterations of samplings are conducted to attain all possible combinations of chemical analogs. Algorithm is terminated after attaining the maximum possible analog combinations.

Input Methods

On the input page, user is permitted to supply the scaffold structure into which selected fragments are to be added and generate the analog library. JChemPaint[15] molecule editing applet incorporated into the eAnalogs interface allows the user to submit molecule in three methods, 1) manual sketching, 2) SMILE[15] format, and 3) structure file in CML[17], MOL[18], or SDF[19] formats.

Sketch Molecule

The molecule editing applet featured with sketching components permits user to manually sketch and design the input structure (Fig. 1a). The applet also provides various prebuilt structural fragment templates to complement the manual structure drawing efforts.

Import SMILE Notation

The simplified molecular-input line-entry system (SMILE format)[15] is a common method of structure representation. SMILE notation of the input molecule can be inserted into the molecule editing applet and convert it to structure.

Import Structure File

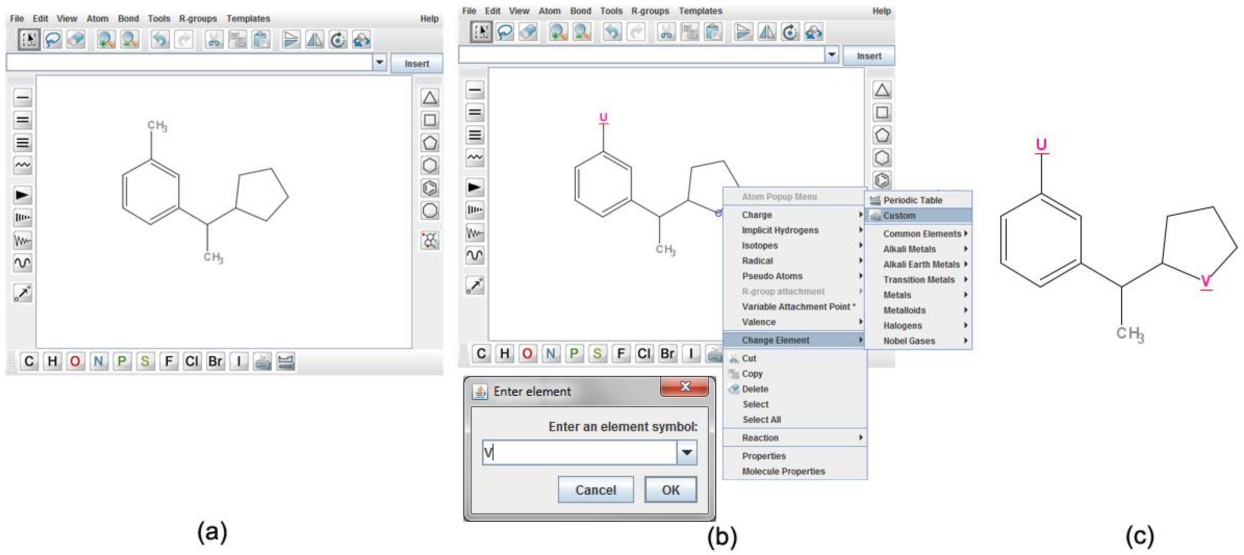
As third option, input molecule can be opened from structure files. Molecule editing applet supports most commonly used structure files such as Chemical Markup Language (CML)[17], MDL MOL file[18], and Structure Data File (SDF)[19].

Define Variable Structural Positions (R Positions)

Current version of eAnalogs allows user to select up to two positions on the scaffold structure as variable positions for analog library construction[20,21]. User needs to define first and second variable positions using the select and custom change element tools on the sketcher applet and label positions as “U” and “V” consecutively (Fig. 1b and 1c).

Fragments Selection

Input interface contains a list of fragments for user to select as R position variables and to be inserted into the core



Enter element

Enter an element symbol:

V

Figure 1 (a): Manual sketching of input structure (b) labeling two variable positions as “U” and “V” using the custom change element tool on the applet (c) input scaffold structure with two variable positions successfully labeled and defined.

compound. Current version of eAnalogs is featured with 8 atoms (Nitrogen, Oxygen, Phosphorous, Sulphur, Fluorine, Chlorine, Bromine, and Iodine), 20 functional groups and 20 heterocyclic rings as choices of additions at the R-Positions (Table 1).

Job Submission

Once the input molecule, R-positions, and the fragments are defined, the job to be submitted for processing with a job name mandatorily provided.

Functionality

Automated Ligand Analog Library Generation

eAnalogs software functions in a systematic combinatorial manner[22-24] in order to generate unique structural analogs. The structure of the core compound, submitted by the user is analyzed as a primary step and variable positions are mapped for the fragment replacements to be made. Newly generated structures are stored to form a library for future reference and to avoid repeated construction of same compounds. Software is featured with downloading results in structure data formats which offers research scientists an easier way to process and characterize multiple ligands for virtual High Throughput Screening. Processing page leads to the results page with analog structures displayed as images and option to view and download the results in PDF as well as Structure Data Formats (SDF) (Fig. 2).

In existing systems, manual molecule sketching and full filling positions with fragments one by one is a time consuming process and high degree of human errors involved. Proposed software tool is designed to dominate on the existing systems with increased accuracy and speed in combinatorial library generation.

RESULTS AND DISCUSSION

Displaying Analogs as Images

Once the eAnalogs algorithm is processed and analog library is generated, results are displayed as two dimensional images of the newly designed structures by replacing the ligands in the core scaffold compound given as input. In the existing system, three dimensional structural views are not facilitated. Current version of the software allows two dimensional views of structures which again blocks the information against molecular conformational changes.

Results in PDF Form

Users are permitted to download the results in the form of PDF report on successful completion of the process.

Analog Library in Structure Data Format (SDF)

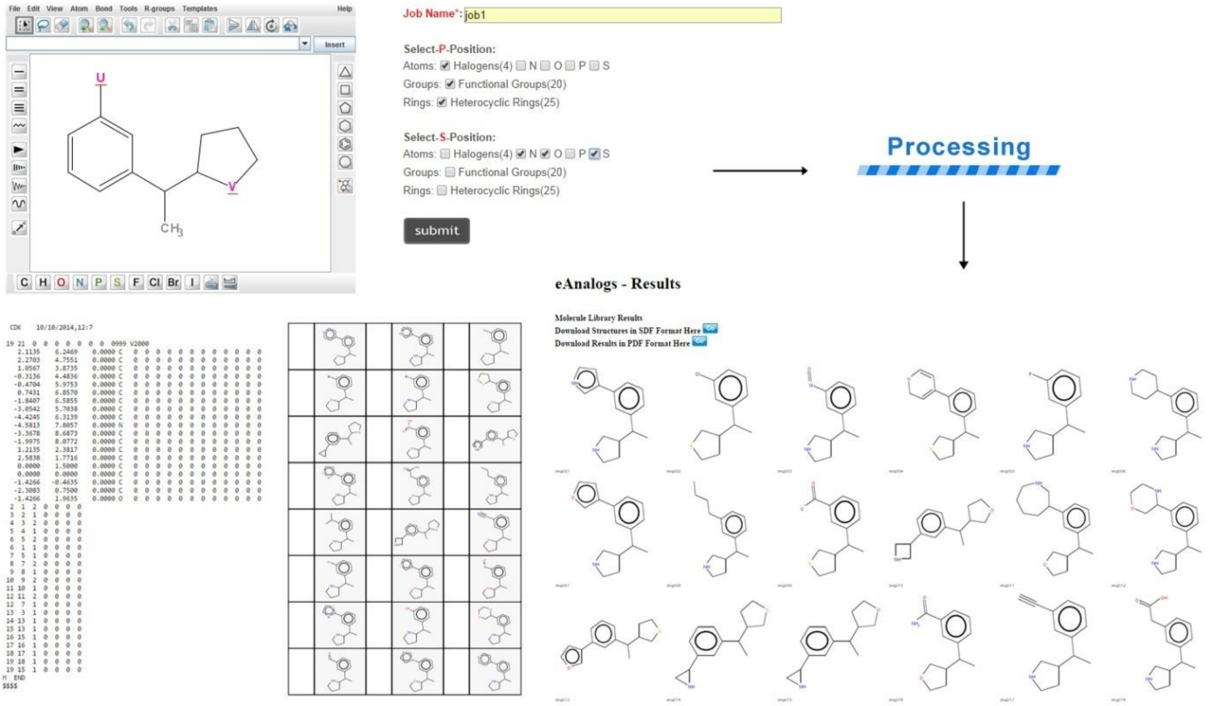
Newly generated analog library can be downloaded in the form for Structure Data Format (SDF) on successful completion of the process. These libraries can be used as inputs for computer aided drug design processes.

Automated structure enumerators[21] play a key role in lead identification and optimization drug discovery steps. It enables fast discovery of new lead compounds and thereby leads to efficient drug discovery process. It also helps in creating a compound database with minimal time investment. The algorithm also supports in near future with multiple options of atoms, functional groups and fragments for position replacements. The calculated ADMET properties of the new analogs created will also be added as a new feature into eAnalogs software.

Newly generated analog libraries have their own advantages of comparing the structures for different fragment combinations and finding the best one having high conformational stability. This high conformational stability has a key role in Drug design, as more stable compounds can be most widely used in the production of new drug.

A better idea about the ADMET properties also plays a key role in choosing the ligand for production of drugs, as these ADMET properties give much more idea about the solubility of the drug and also about the stability of the newly designed compound.

Existing systems are limited with features like automated multiple replacement or insertions of fragments to the scaffold. Even if present, those are either time consuming



Job Name*: job1

Select-P-Position:
Atoms: ☑ Halogens(4) □ N □ O □ P □ S
Groups: ☑ Functional Groups(20)
Rings: ☑ Heterocyclic Rings(25)

Select-S-Position:
Atoms: □ Halogens(4) ☑ N ☑ O □ P ☑ S
Groups: □ Functional Groups(20)
Rings: □ Heterocyclic Rings(25)

submit

```
 CDK    10/10/2014,12:7

19 21  0  0  0  0  0  0  0  0999 V2000
    2.1135    6.2469    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    2.2703    4.7551    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    1.0567    3.8735    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -0.3136    4.4836    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -0.4704    5.9753    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    0.7431    6.8570    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -1.8407    6.5855    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -3.0542    5.7038    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -4.4245    6.3139    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -4.5813    7.8057    0.0000 N   0  0  0  0  0  0  0  0  0  0  0  0
   -3.3678    8.6873    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -1.9975    8.0772    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    1.2135    2.3817    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    2.5838    1.7716    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    0.0000    1.5000    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    0.0000    0.0000    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -1.4266   -0.4635    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -2.3083    0.7500    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -1.4266    1.9635    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
  2  1  2  0  0  0  0
  3  2  1  0  0  0  0
  4  3  2  0  0  0  0
  5  4  1  0  0  0  0
  6  5  2  0  0  0  0
  6  1  1  0  0  0  0
  7  5  1  0  0  0  0
  8  7  2  0  0  0  0
  9  8  1  0  0  0  0
 10  9  2  0  0  0  0
 11 10  1  0  0  0  0
 12 11  2  0  0  0  0
 12  7  1  0  0  0  0
 13  3  1  0  0  0  0
 14 13  1  0  0  0  0
 15 13  1  0  0  0  0
 16 15  1  0  0  0  0
 17 16  1  0  0  0  0
 18 17  1  0  0  0  0
 19 18  1  0  0  0  0
 19 15  1  0  0  0  0
M  END
$$$$
```

eAnalogs - Results

Molecule Library Results
Download Structures in SDF Format Here
Download Results in PDF Format Here

Figure 2: eAnalogs process outline

or non-user friendly. Manual structure drawing tools consume more man power to get the structures of the newly synthesized compounds. Users need to draw the core compound manually and attach individual atoms or fragments to make individual analog. Information related to the structure stability or the conformational changes made by the functional group added to the structure are also not available. This made the structure an unstable one in proceeding further in the drug design. Manual errors also made major issue while drawing the same structures repeatedly, ultimately leading to a wrong core structure.

Proposed algorithm is trying to overcome the issues faced with existing systems. In eAnalogs, 'n' number of positions can be added or replaced with atoms or fragments at the same time. This reduces the time consumption, which is very vital in early stage drug discovery process. Since the software itself automatically generates the multiple combinations of structures, it reduces the man power needed exponentially. Also it is possible to save newly designed structure images and structure data files for future studies.

CONCLUSION

The present investigation introduced and demonstrated eAnalogs, a web-based software intending to automate ligand analogs sketching and combinatorial compound library construction. Software is featured with processing user sketched or imported single chemical scaffold with defined positions where structural changes to be reflected. Desired atoms, functional groups or chemical rings to be selected from a pre-defined list provided by the software. Software generates multiple analogs from a single core chemical scaffold by adding and replacing user-defined atoms, functional groups and chemical rings of interests and their combinations. The software then combines these input data and then goes on multiplying the structure of the new compounds with replacements. The application of the developed methodology to dynamic image database environment (i.e., support insertion and deletion of images in the database) is highly appreciable as it is used to insert and delete bonds in the scaffold structure obtained for enumeration. eAnalogs is presently incorporated to the web-based software platform Cheminformatica workbench, which will be added with new modules like ADMET screening, Bioactivity prediction etc., in the near future.

REFERENCES

1. Li XJ. Kong DX, Zhang HY. Chemoinformatics approaches for traditional chinese medicine research and case application in anticancer drug discovery. Curr Drug Discov Technol. 2010; 7(1): 22-31.
2. Handen JS. High-throughput screening - challenges for the future. Drug Discov. World 2002: 47–50.
3. Walters WP, Stahl MT, Murcko MA. Virtual screening – an overview. Drug Discov. Today 1998; 3(4): 160–178.
4. Bajorath J. Selected concepts and investigations in compound classification, molecular descriptor analysis, and virtual screening. J. Chem. Inf. Comput. Sci. 2001; 41(2): 233–245.
5. Bajorath J. Virtual screening: methods, expectations, and reality. Curr. Drug Discov. 2002; 2(3): 24–28.
6. Rester U. From virtuality to reality - Virtual screening in lead discovery and lead optimization: A medicinal chemistry perspective. Curr Opin Drug Discov Devel. 2008; 11(4): 559–68.

7. Rollinger JM, Stuppner H, Langer T. Virtual screening for the discovery of bioactive natural products. Progress in Drug Research 2008; 65: 211, 213–49.
8. Smith A. Screening for drug discovery: the leading question. Nature 2002; 418(6896): 453–459.
9. Drews J. Drug discovery: a historical perspective. Science 2000; 287(5460): 1960–1964.
10. Willett P, Barnard JM, Downs GM. Chemical Similarity Searching. Journal of Chemical Information and Computer Science 1998; 38(6): 983−996.
11. Johnson MA, Maggiora GM. Concepts and Applications of Molecular Similarity. John Willey & Sons, New York 1990.
12. Nikolova N, Jaworska J. Approaches to Measure Chemical Similarity - a Review. QSAR & Combinatorial Science 2003, 22(9-10): 1006–1026.
13. Martin, YC, Kofron JL, Traphagen LM. Do Structurally Similar Molecules Have Similar Biological Activity?. Journal of Medicinal Chemistry 2002; 45(19): 4350–4358.
14. Sunter AB. List Sequential Sampling with Equal or Unequal Probabilities without Replacement. Journal of the Royal Statistical Society. Series C (Applied Statistics) 1977; 26(3): 261-268.
15. Stefan K, Egon W, Christoph S. JChemPaint - Using the Collaborative Forces of the Internet to Develop a Free Editor for 2D Chemical Structures. Molecules 2000; 5(1): 93-98.
16. Anderson E, Veith GD, Weininger D. SMILES: A line notation and computerized interpreter for chemical structures. Duluth, MN: U.S. EPA, Environmental Research Laboratory-Duluth 1987.
17. Murray-Rust P, Rzepa HS. Chemical Markup, XML and the World Wide Web. 4. CML Schema. J. Chem. Inf. Comput. Sci. 2003; 43(3): 757–772.
18. Dalby A; Nourse JG, Hounshell WD, Gushurst AKI, Grier DL, Leland BA, Laufer J. Description of several chemical structure file formats used by computer programs developed at Molecular Design Limited. Journal of Chemical Information and Modeling 1992; 32(3): 244.
19. De Martino G, Silverstri R. et al. Arylthioindoles, Potent Inhibitors of Tubulin Polymerization. J. Med. Chem. 2004; 47(25): 6120-6123.
20. Barnard JM, Downs GM. Use of Markush Structure Techniques to avoid Enumeratrion in Diversity Analysis of Large Combinatorial Libraries, Barnard Chemical Information 1997.
21. Kvasnicka V, Pospichal J. Constructive Enumeration of Acyclic Molecules. Collect. Czech. Chem. Commun. 1991; 56: 1777-1802.
22. Martinez WL. Graphical user interfaces. WIREs Comp Stat. 2011; 3(2): 119–133.
23. Roberto A, Pierre H, Federico M. Chemical Trees Enumeration Algorithms. Quarterly Journal of the Belgian, French and Italian Operations Research Societies 2003; 1(1): 67-83.
24. Trinajstic N, Nicolic S, Knop JV, Mauller WR, Szymanski K. Computational Chemical Graph Theory. Ellis Horwood: New York 1991.