

## An Approach for Computer-Based Chemical Compound Design

Amena Mahmoud<sup>A</sup>, Mazdi Zakaria<sup>B</sup> and Taher Hamza<sup>C</sup>

<sup>A</sup>Computer Science Department, Mansoura University, Mansoura, Egypt

<sup>B</sup>Computer Science Department, Mansoura University, Mansoura, Egypt

<sup>C</sup>Computer Science Department, Mansoura University, Mansoura, Egypt

---

### Abstract

To aid the design and development of new drugs, automated computational methods can be used to generate molecules that could be used as lead compounds for possible new drugs. Lead compounds are chemical compounds that have pharmacological or biological activity of which the chemical structures can be used as a starting point for chemical modifications in order to improve potency, selectivity, or pharmacokinetic parameters. The proposed approach has the ability to design chemical compound structures plus the availability to transform these structures into their computer representation SMILES in order to be used for classification and testing purposes.

**Keywords:** Thermal Analysis, Pre-conditioner, In-mold, Inoculant's efficiency.

---

### I. INTRODUCTION

Conventional drug design techniques are based on trial-and-error testing using cells or animals. High-throughput screening for chemicals with desired bioactivities requires specialized labs that make the process costly. With a growing number of known experimental structures of target molecules, computational methods have been used successfully to supplement and speed up drug discovery.

Computer-based molecular design combines methods of informatics, medicine, and biophysics. This cross-disciplinary field has accelerated drug research by predicting the potential therapeutic effectiveness of designed molecules prior to laborious experiments and costly preclinical trials. In addition, computational modeling has led to discoveries of structures of novel small molecules.

To have a successful drug molecule, two procedures are needed. The first one is developing a structure design application that aids graphical molecule editing with the availability of using different tools (palette of fonts, arrows, lines and curves, arcs and other shapes or graphic primitives, etc.) and then converting them to computer readable format. The second procedure is the testing of the designed molecule using computational algorithms to decide its availability to be a specific drug.

The first procedure leads to the development of the current proposed chemical designing software which can be integrated with databases, other bundled program or be embedded within a web site.

### II. CRITICAL REVIEW

Every software has its unique features, advantages/disadvantages, but all have the a common thing that they will generate chemical structure in as less time as possible, which is not possible with the ordinary softwares and are freely accessible. In this aspect, the most

popular high quality chemical structure drawing applications are reviewed.

Accelrys Draw [1], an alternative input module to its own with a brand new and more user-friendly interface. It supports some features like CrossFire database use and many drawing tools are used. Chem(Bio)Draw [2], an integrated application with the Chem(Bio)Office and E-Notebook, available as Windows or Mac version. The drawing of biological objects have been developed lately. SoftShell's old ChemWindow has melted into the KnowItAll application suite and has been split to DrawIt [3] and ReportIt – separate molecular editor and general editor applications, the latter is to import objects such as structures and spectra, arrange them on the page, and add text, annotations, reaction arrows, orbitals or other graphical objects. ChemSketch [4] is the input module of ChemFolder and several other applications of the ACD suite. It has several modules, extensions, and add-ins concerning the calculation of physical-chemical properties, import, processing and database using of spectra and chromatograms, naming of molecules, prediction of NMR spectra, etc. A new player in the field of chemical editors is ChemDoodle [6] for Windows, Mac and Linux. It has everything necessary for creating publication quality chemical documents. It is integrated with ChemDoodle Web Components and ChemDoodle Mobile (iOS and Android phones). In addition to structure and document editing it contains several tools (some of them called "widgets") such as constitutional and physicochemical descriptors and properties, embedding, editing and simulation of spectra, etc. The main drawback of the current version that round-trip editing is not possible.

All of the mentioned software packages are stand-alone applications, they are members of integrated suites as well, and they serve as the input modules of other database and modeling software applications or full information management systems. They are commercially released and not open source of free-ware applications.

The current proposed approach deals with the previous approaches disadvantages and aims to develop an application which is open-source, open platform, free-ware with an easy navigate interface.

### III. CHEMICAL MOLECULE DESIGN

The structural formula of a chemical compound is a graphical representation of the molecular structure, showing how the atoms are arranged. The chemical bonding within the molecule is also shown, either explicitly or implicitly. Also several other formats are used, as in chemical databases, such as SMILES, InChI and CML.

Unlike chemical formulas or chemical names, structural formulas as in fig.1, provide a representation of the molecular structure. Chemists describe a chemical reaction or synthesis using structural formulas rather than chemical names, because the structural formulas allow the chemist to visualize the molecules and the changes that occur.

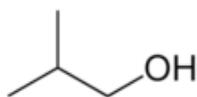


Fig.1. Example of chemical formula of isobutanol

Two atoms can be bonded by sharing more than one pair of electrons. The common bonds to carbon are single, double and triple bonds as in fig.2. Single bonds are most common and are represented by a single, solid line between two atoms in a skeletal formula. Double bonds are denoted by two parallel lines, and triple bonds are shown by three parallel lines.



Fig. 2. Example of carbon-carbon double bond

Benzene [8] rings are common in organic compounds. To represent the delocalization of electrons over the six carbon atoms in the ring, a circle is drawn inside the hexagon of single bonds. This style is very common in introductory organic chemistry. Although it could be considered inaccurate as it implies three single bonds and three double bonds (benzene would therefore be cyclohexa-1,3,5-triene), all qualified chemists are fully aware of the delocalization in benzene as in fig.3.

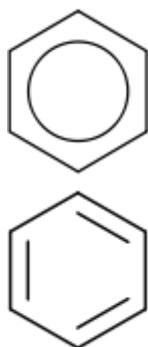


Fig. 3. Example of Benzene rings

Rules of representing atoms and bonds in the chemical formula:

- Hydrogen bonds are sometimes denoted by dotted or dashed lines.
- Solid lines represent bonds in the plane of the paper or screen.
- Wedges represent bonds that point out of the plane of the paper or screen, towards the observer.
- Dashed lines represent bonds that point into the plane of the paper or screen, away from the observer.
- Wavy lines represent either unknown stereochemistry or a mixture of the two possible stereoisomers at that point, fig.4.

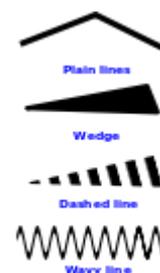


Fig. 4. Types of bonds

### IV. THE PROPOSED APPROACH

#### A. Motivation

2D chemical structure editors are central tools in fields like chemoinformatics, computational chemistry and synthetic chemistry. No matter if one wants to submit a structure query to a database, prepare a starting structure for molecular modeling, draw a set of structures for good lists and bad lists in Computer Assisted Structure Elucidation (CASE) or just sketch a structure or reaction scheme for a publication - in all of these cases the starting point is opening a structure editor.

Nevertheless, there are no state-of-the-art programs available in source code with a free licensing scheme, thus allowing researchers to adapt and embed them into their own programs without paying license fees. Such an open source structure editor would be of interest for many reasons, of which just not having to pay for it is certainly the weakest argument.

Firstly, it would ease the work of all those developers who need to be able to change and adapt the source code of a module they use in order to integrate it into their projects. The authors of programs that calculate NMR shifts or generate the IUPAC name for a given structure would not have to rewrite this standard piece of software again and again. Secondly, open source projects are known to solve problems with faulty software in a short time since bugs are much more easily found and improvements are much more easily made if everyone can have a look at the source code.

## B. System Design

Taking into account the fairly large number of programs available in this field, there would seem to be no great challenges in designing such a system. However, some aspects are tricky to program and pose interesting problems, for example, written in an object oriented programming language (JAVA), with its clear and modular design, and its source code available to everyone, seems to be the ideal playground for trying new ideas and optimizing existing solutions.

The program already offers some amount of functionality. Fig. 5, shows a screen shot for the program drawing a chemical compound with the use of ring templates.

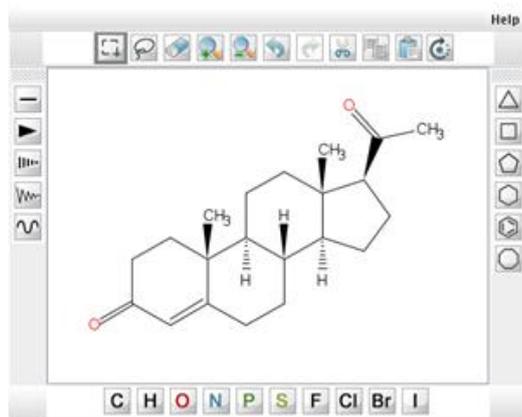


Fig. 5. Screen shot of the program

The interface of the designed applet supports the following needs for the Chemical Structure:

- Single, double and triple bonds
- Stereo descriptors
- Atoms ( elements names )
- Benzene rings ( size 3 to 8 )
- Flipping and rotating selected parts of the molecule

## C. Graph-based definition

In terms of a graph-based computational procedure, SMILES [10] is a string obtained by printing the symbol nodes encountered in the program of a chemical graph. The chemical graph is first trimmed to remove hydrogen atoms and cycles are broken to turn it into a spanning tree. Where cycles have been broken, numeric suffix labels are included to indicate the connected nodes. Parentheses are used to indicate points of branching on the tree.

Typically, a number of equally valid SMILES can be written for a molecule. For example, CCO, OCC and C(O)C all specify the structure of ethanol. Algorithms have been developed to ensure the same SMILES are generated for a molecule regardless of the order of atoms in the structure. This SMILES is unique for each structure, although dependent on the canonicalization algorithm used to generate it, and is termed the Canonical SMILES. These algorithms first convert the SMILES to an internal

representation of the molecular structure and do not simply manipulate strings as is sometimes thought.

## D. Examples

### Atoms

Atoms are represented by the standard abbreviation of the chemical elements, in square brackets, such as [Au] for gold. Brackets can be omitted for the "organic subset" of B, C, N, O, P, S, F, Cl, Br, and I. All other elements must be enclosed in brackets. If the brackets are omitted, the proper number of implicit hydrogen atoms is assumed; for instance the SMILES for water is simply O.

### Bonds

Bonds between aliphatic atoms are assumed to be single unless specified otherwise and are implied by adjacency in the SMILES string. For example the SMILES for ethanol can be written as CCO. Ring closure labels are used to indicate connectivity between non-adjacent atoms in the SMILES string, which for cyclohexane and dioxane can be written as C1CCCCC1 and O1CCOCC1 respectively.

## E. Chemical Graph Construction Algorithm

The auxiliary graph  $G_u$  is constructed from the graph  $G$  and the binary vector  $u$  as follows: for each vertex  $v$  in  $G$ , there are two vertices  $v$  and  $v_0$  in the  $G_u$ . By adding both edges, we obtain a closed path in  $G$ . This might not be a cycle; however, it always contains an odd parity simple cycle as a subpath.

Input: Graph  $G$ , kernel vector  $u$

Output: Auxiliary graph  $G_u$

```

1: for all vertices  $v$  of  $G$  do
2:   Add two vertices  $v$  and  $v_0$  to  $G_u$ 
3: end for
4: for all edges  $e = (x, y)$  of  $G$  do
5:   if  $u_e = 1$  then
6:     Add the two edges  $(x, y_0), (x_0, y)$  to  $G_u$ 
7:   else
8:     Add the two edges  $(x, y), (x_0, y_0)$  to  $G_u$ 
9:   end if
10: end for

```

## F. Results and Discussion

In some aspects the designed system is different from other 2D editors:

- It's an open source and free software. We believe that scientific software should be free and open for use and development from others.
- It's written in JAVA, so it can run on any computing platform and operating system for which a JVM is implemented.

- It contains a help tutorial for how to use.

The supported features are:

1. A subset of the regular drawing features of commercial programs, such as
  - drawing of single, double and triple bonds
  - stereo descriptors
  - deletion of bonds and atoms
  - templates for rings of size 3 to 8
  - one click attachment of templates to an atom or a bond
  - flipping and rotating selected parts of the molecule
2. Loading and saving structures in SMILES .
3. Allowing keyboard entry of new symbols.

## V. CONCLUSION

The proposed program, a 2D molecular structure editor, can be considered as a development model and a software that might attract the attention of a potentially large group of users and of some highly welcome new co-developers. We have also shown the powerful distributed development model that is open sourcing which can help the development of this software.

A great number of improvements and new features still await implementation. Professional quality output is, for example, not yet possible nor is adaptation to different types of layout. Another needed capability to be added to the program is converting the painted graph into saved images. An obvious new feature is the implementation of an open source algorithm for conversion to and from structural names conforming to the IUPAC nomenclature rules. A lot of work also remains to be done on the interface of the program. The experiences from other projects show that a critical mass of working features must be implemented in order to attract new contributors.

## REFERENCES

- [1] Accelrys (Symyx) Draw 4.0 (Symyx Technologies, Inc) URL: <http://accelrys.com/products/informatics/cheminformatics/draw/>.
- [2] ChemDraw 12.0.2; CambridgeSoft Corporation, Cambridge, MA, USA; URL: <http://www.cambridgesoft.com>.
- [3] DrawIt 9.0 (KnowItAll Academic Edition); Bio-Rad Laboratories Informatics Division, Philadelphia PA, USA; URL: <http://www.chemwindow.com> or <http://www.bio-rad.com>.
- [4] ACD/ChemSketch 12.0; Advanced Chemistry Development, Inc, Toronto, Canada; URL: <http://www.acdlabs.com>.
- [5] Chemistry 4-D Draw 8.4.6; ChemInnovation Software Inc., San Diego, USA; URL: <http://www.cheminnovation.com>.
- [6] a) ChemDoodle 4.1.1, iChemLabs LLC, Piscataway, NJ, USA; URL: <http://www.chemdoodle.com> , b) ChemDoodle Review. URL: [http://homepage.mac.com/swain/Macinchem/Reviews/chemdoodle\\_review/chemdoodle\\_review.html](http://homepage.mac.com/swain/Macinchem/Reviews/chemdoodle_review/chemdoodle_review.html) , c) Exploring ChemDoodle Web Components. URL: [http://homepage.mac.com/swain/Macinchem/Reviews/Chemdoodle/chemdoodle\\_web\\_components.htm](http://homepage.mac.com/swain/Macinchem/Reviews/Chemdoodle/chemdoodle_web_components.htm) (accessed 28.12.2011)
- [7] Skeletal formula; URL: [http://en.wikipedia.org/wiki/Skeletal\\_formula](http://en.wikipedia.org/wiki/Skeletal_formula)
- [8] Benzene rings ; URL: <http://en.wikipedia.org/wiki/Benzene>
- [9] Stereochemistry ; URL: <http://en.wikipedia.org/wiki/Stereochemistry>
- [10] SMILES; URL: [www.daylight.com](http://www.daylight.com)