Structural Library of Natural Compounds

Subhash Chandra Bose. Kotte[1]*, Pavan Kumar K.V.T.S[2]*, Ravi Kumar Tumuluri[3]*, Shriram Raghavan[2], P.K. Dubey[1] and P.M. Murali[2]

[1] Jawaharlal Nehru Technological University Hyderabad, Kukatpally, Hyderabad, Andhra Pradesh, 500072, India.

[2] Dalmia Centre for Research & Development, 10/8 Bharathi Park Main Road, SAHSC Post, Saibaba Colony, Coimbatore,641043, India.

[3] Acharya Nagarjuna University- Dr. M.R. Appa Row Campus, NUZVID, 521201, India.

Abstract

Natural products with biological activity are of considerable interest to drug discovery community and a structural library of such compounds serves as data set for *insilico* experiments to predict the target specific activity before screening them in in-vitro experiments. This work leverages open source scientific tools to create a database of such compounds library in Structure Data Format with 3D coordinates which in turn may be used as an input file for various applications.

Keywords: *Natural Product library; SD file; ChemSketch; Open Babel*

1. Introduction

Colossal expansion in the field of computational biology and medicinal chemistry is providing thrust to the drug discovery community by opening new horizons. Improved algorithms for virtual screening of small molecules in batch mode has emphasized on the need for small molecule databases. To cater to this requirement, authenticated source is essential to mine structures of small molecules and N atural Products Updates (NPU) [1] acts as one such compendium that provides graphical abstracts of current developments in natural product chemistry, selected from dozens of key primary journals. NPU is published by Royal Society of Chemistry (RSC) [2] and its coverage includes isolation studies, biosynthesis, and new natural products, known compounds from new sources, structure determinations, new properties and biological activities. Around 200 graphical abstracts are contained in each monthly bulletin including structure diagrams, trivial and taxonomic names, molecular formulae and physical and biological properties. This is an ideal source to mine for natural compounds with or without specific biological activity from various natural sources of interest. B ut in order to conduct *insilico* experiments, it is important to convert these structures which are in image format into Structure Data Format [3]. Open source scientific tools of greater use for creating such database of these compounds and used in this effort.

2. Methodology

2.1 Structural data Collection

Electronic version of NPU journal was started in 1999 from Royal Society of Chemistry and each molecule will be identified with a unique reference number called NPU number. The current NPU issue is 5 - 2011 and a total number of ~ 23,000 molecules with structures were published till now. Graphical extracts “Fig. 1”, of these NPU molecules till issue 12-2008 were downloaded as Hyper Text Markup Language files (HTML) [4] and saved to draw structures.

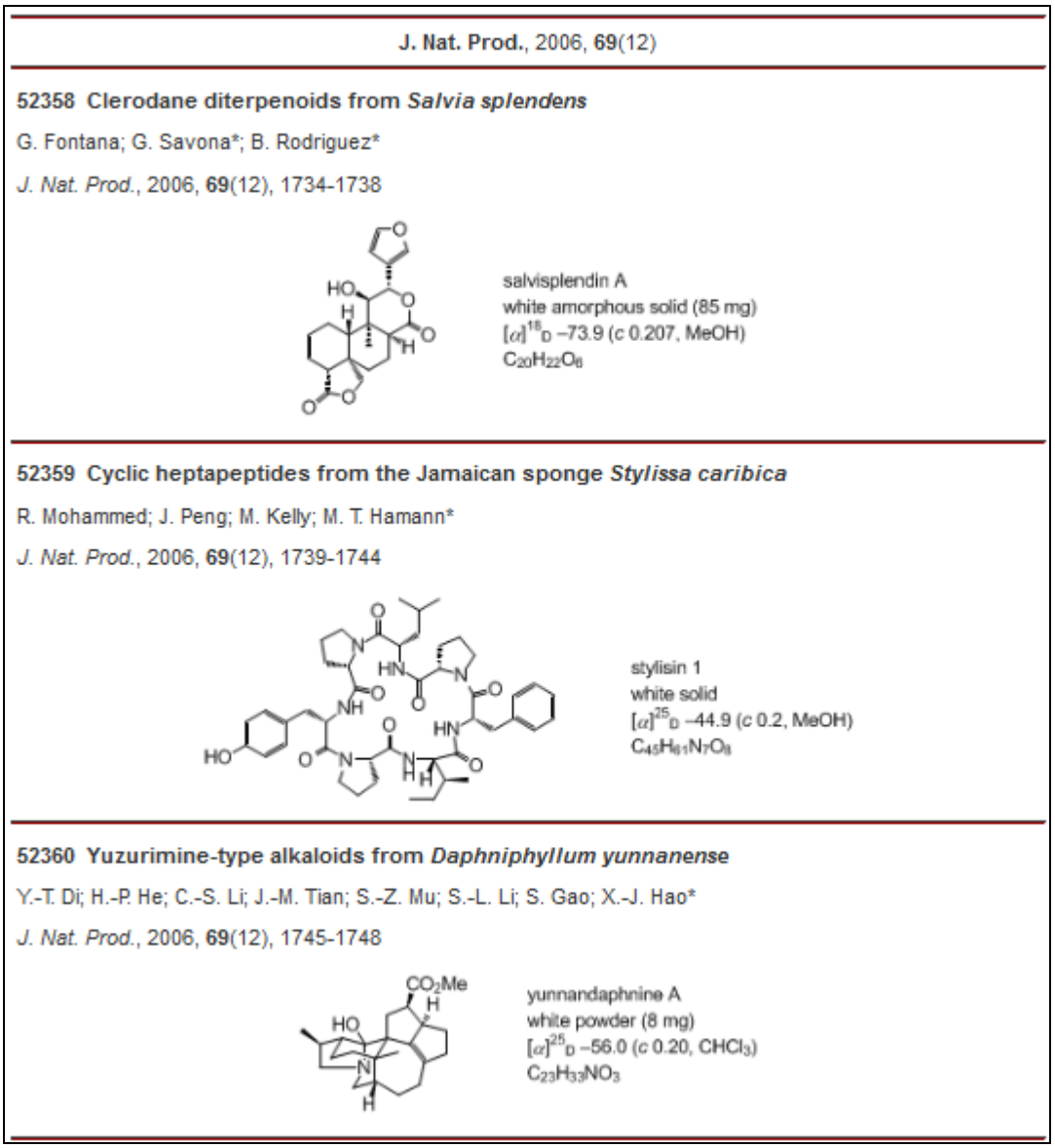

Fig 1. Natural Product Updates Graphical structure information.

2.2 Drawing of structures in 2D format

Since one of the objectives of this effort is to deploy open source tools to create structural database, a freeware "ChemSketch" "Fig. 2", from ACD (Advanced Chemistry Development Inc.) [5] Was used to draw structures. The saved Joint Photographic Experts Group (JPEG) [6] files of the NPU molecules were meticulously drawn in to 2D, indexed with respective NPU number and converted in to a single Structure Data file, the most convenient chemistry structures holding file format for most of the *insilico* applications, comprises of 18, 464 structures.

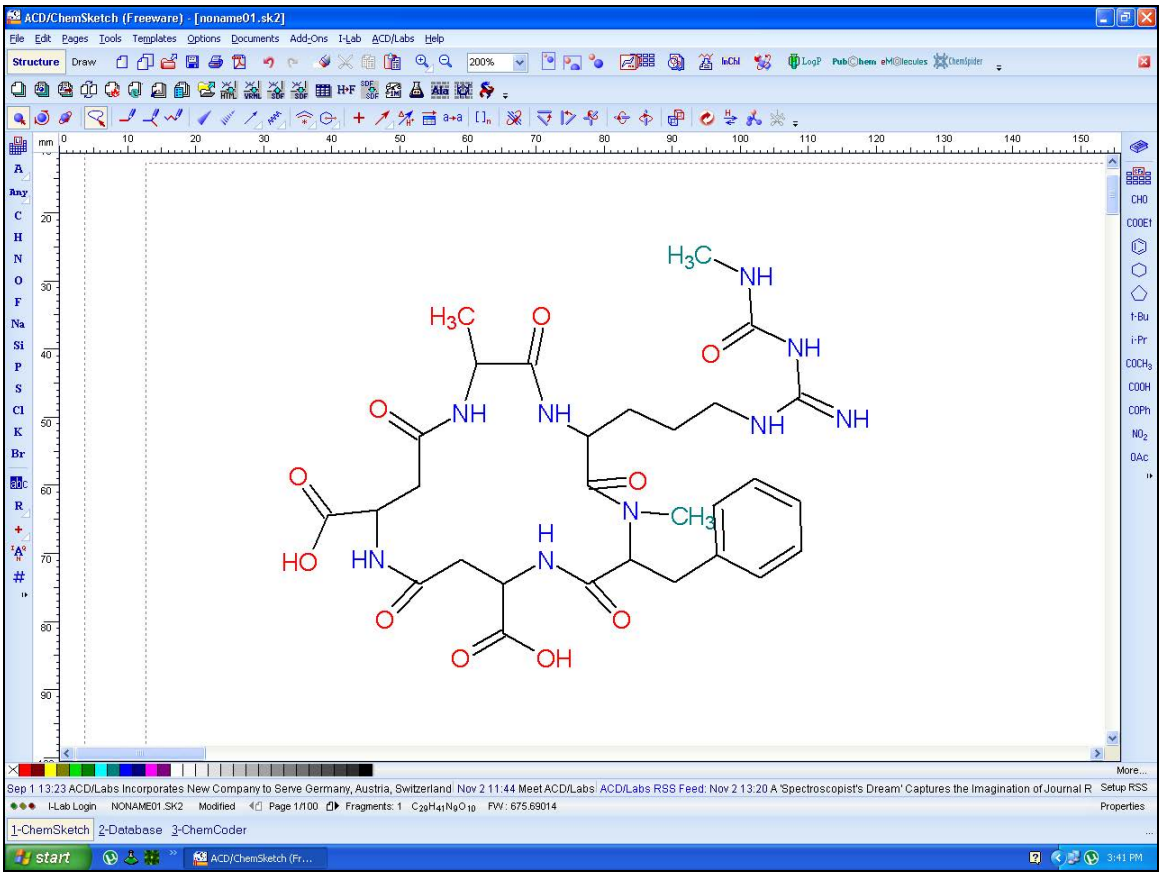

Fig 2. ACD Chemsketch Structure drawing tool.

2.3 Conversion of 2D chemical structures into 3D format

To understand small molecule interaction with macro molecules such as enzymes, Receptors etc., one of the important properties is stereochemistry of these molecules. Considering this the library of these 18,464 molecules were converted in to 3D "Fig. 3", format by using another freeware called Open Babel [7], a chemical toolbox designed to speak the many languages of chemical data for interconverting, searching, modifying, and analyzing chemical files or store data for *insilico* experiments.

The entire process flow of the library creation methodology was described in "Fig.4"

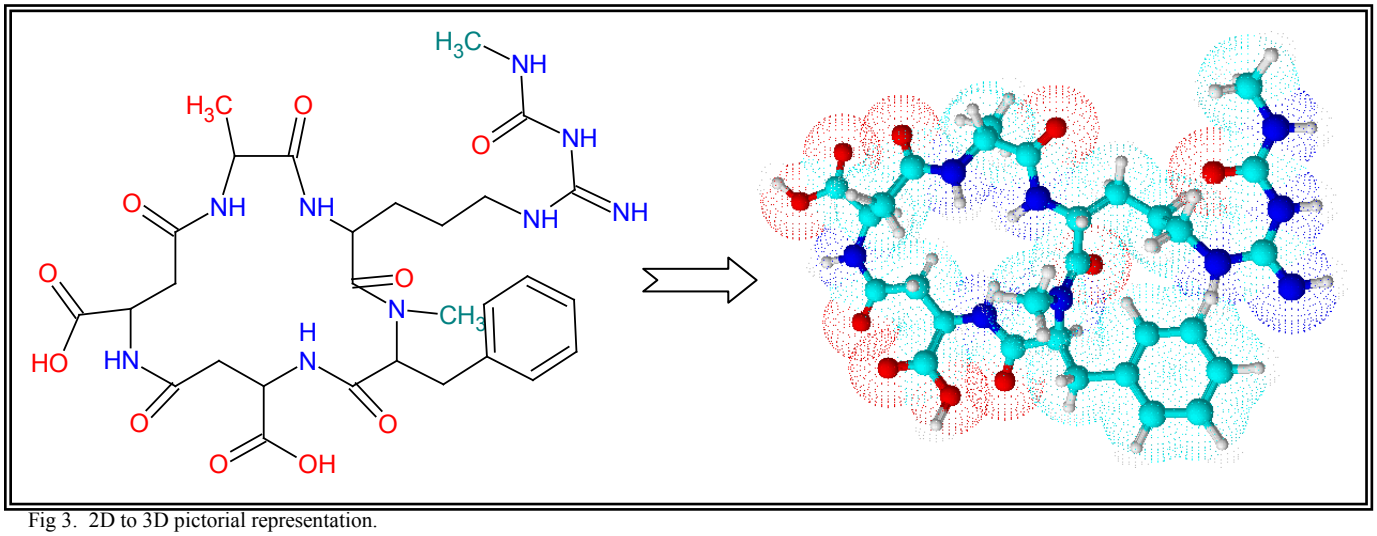
Fig 3. 2D to 3D pictorial representation.

3. Advantages

This NPU library is quite helpful in hastening the throughput of screening prospective leads for drug targets of interest by way of automation. This will also be of use for building Structure Activity Relationship (SAR) studies.

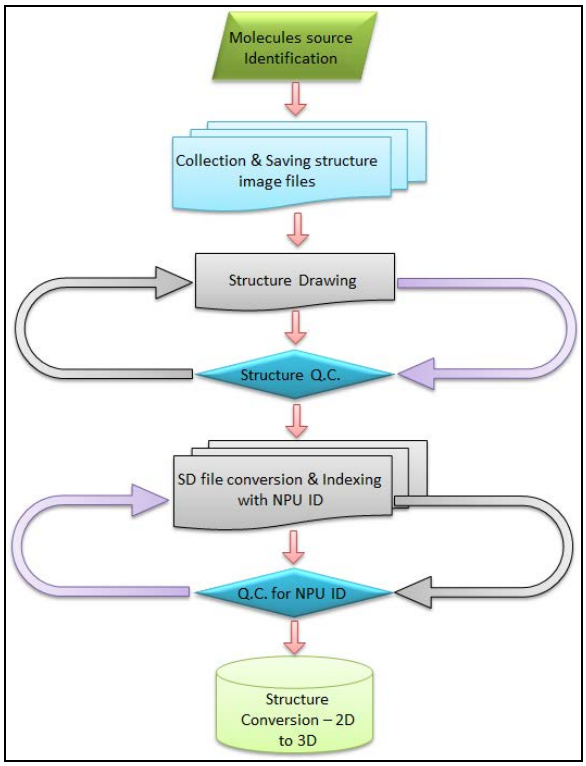

Fig 4. Flow diagram of Database creation.

4. Future Development

As RSC continuous to publish natural products information in NPU Journal and only 18,464 structures out of 23,000 were drawn till now, the efforts will be continued to update this structural database and we plan to develop provision to search the database on a web based platform to the scientific community.

5. Acknowledgments

We would like to thank for constant support from Dalmia Centre for Research & Development and Indian Institute of Chemical Technology for literature support.

6. References

[1] http://www.rsc.org/Publishing/CurrentAwareness/npu/index.asp
[2] http://www.rsc.org/Publishing
[3] Arthur Dalby, James G. Nourse, W. Douglas Hounshell, Ann K. I. Gushurst, David L. Grier, Burton A. Leland, John Laufer "Description of several chemical structure file formats used by computer programs developed at Molecular Design Limited," Journal of Chemical Information and Computer Sciences. vol. 32, pp. 244-255, may 1992.
[4] http://en.wikipedia.org/wiki/HTML
[5] Allen D. Hunter "ACD/ChemSketch 1.0 (freeware); ACD/ChemSketch 2.0 and its Tautomers, Dictionary, and 3D Plug-ins; ACD/HNMR 2.0; ACD/CNMR 2.0," J. Chem. Educ., vol 74, p 905, August 1997.
[6] http://en.wikipedia.org/wiki/JPEG
[7] Rajarshi Guha, Michael T. Howard, Geoffrey R. Hutchison, Peter Murray-Rust, Henry Rzepa, Christoph Steinbeck, Joerg Kurt Wegner, Egon Willighagen. "The Blue Obelisk - Interoperability in Chemical Informatics." J. Chem. Inf. Model., vol 46, pp. 991-998, February 2006.

First Author:
1. Subhash Chandra Bose. Kotte obtained his Masters in Chemistry from Jawaharlal Nehru Technological University at Hyderabad. He specializes in Chemical synthesis, Analytics and presently a researcher in this field with Jawaharlal Nehru Technological University Hyderabad.
2. K.V.T.S.Pavan Kumar obtained his Masters in biochemistry. He specializes in Bio-informatics and is presently a researcher in the field of Plant biotechnology with Dalmia Centre for Research and Development.
3. T. Ravi Kumar obtained his Masters in biochemistry. He is presently a researcher in the field of biochemistry with Acharya Nagarjuna University

Second Author: Shriram Raghavan obtained his Masters in biochemical engineering and biotechnology from HP University at Shimla. He specializes in Bio-informatics and is presently a researcher in this field with Dalmia Centre for Research and Development.

Third Author:
1. P.K. Dubey is M.Sc.(OU), Ph.D (O.U), Specialized in Organic Chemistry. His research interests include Synthetic organic chemistry. He guided over 20 Ph.D. Students and presently guiding seven students for Ph.D. Degrees. He published over 140 research papers. He has 31 years of teaching and r esearch experience.

2. P.M. Murali received a Ph.D. in Microbiology and M icrobial technology from Madurai Kamaraj University, having over 20 years of experience in Pharmaceutical & Healthcare R&D, including management of more than 10 clinical trials, in particular in respiratory diseases. He has remained the Founder and Director of Dalmia Centre for Research and Development for 16 years, developing and launching natural product based therapeutics, and Founder and C hairman of MLC & Netpeople group of IT & Telecom companies (Networking solutions, Banking Security and communication services). Dr. Murali is a former Indo-US scientist at Battelle-Kettering, Ohio and fellow of Unilever India. He is currently General Secretary of the Association of Biotechnology Led enterprises in India.