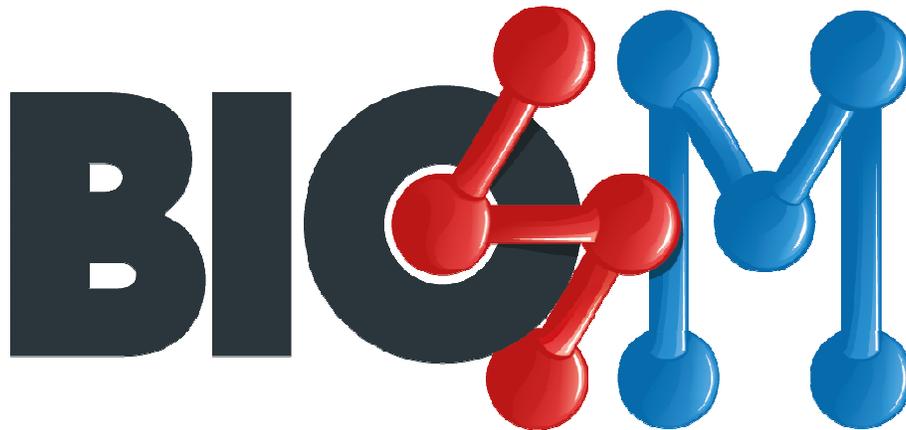# User Guide

11/1/2012

Mai Hamdalla

Computer Science and Engineering Department

University of Connecticut

Availability and Requirements

Project home page: http://metabolomics.pharm.uconn.edu/

Operating System(s): Platform independent. BioSM should work on any operating system (OS) including Windows, Linux/Unix Mac OSX or Solaris.

Other requirements: Java 1.6 or higher. Java standard edition (SE) run time (jre 1.6 / jre 1.7) can be freely downloaded from

http://www.oracle.com/technetwork/java/javase/downloads/index.html

Memory: We recommend at least 2 GB of RAM for java virtual machine. However, more RAM may be required depending on the number and size of structures being classified. The bundled program execution scripts (**BioSM.bat** and **BioSM.sh** for Windows and Linux respectively) found in the BioSM folder will allocate 2GB of RAM for the java virtual machine. Please note that the RAM allocated for the java virtual machine can be easily changed. (See FAQ)

Processor: There are no processor requirements. However, faster (multi-core) processors will certainly help since BioSM is designed to take advantage of modern multi-core architectures.

Restrictions to use by non-academics: NONE

Installation Instructions

No installation is required. Simply extract the downloaded zip file into a folder.

To start BioSM on

- Windows OS, double click on **BioSM.bat**

- Linux/Mac OSX/Solaris, double click on BioSM.sh or type `./BioSM.sh` in the command

  prompt terminal.

Contents of the BioSM Folder

Folders:

- *config*: contains configuration files.

- *input*: contains input files with candidate structures to be tested.

- *output*: contains individual output folders with results for each BioSM run.

Files:

- *BioSM.bat*: program startup script for Windows OS.

- *BioSM.jar*: BioSM platform independent executable file.

- *BioSM.sh*: program startup script for Linux.

Frequently Asked Questions (FAQ)

1. **Nothing happens when I double click BioSM.bat.**

    This happens when operating system cannot locate the java run time (jre) Add

    `C:\Program Files (x86)\Java\jre7\bin` or `C:\Program Files\Java\jre7\bin`; to

    windows path. Please note that jre might be found at a different location. So search for

    `jre6\bin` or `jre7\bin` directories on your system.

2. **How do I know java runtime is correctly installed and java path is correctly set?**

    Type `java –version` at the command prompt (dos prompt in Windows or terminal in

    Linux). This command should return the version number of java runtime.

3. **Why should I double click on the startup script (not the jar file) to start the program?**

    Configuration files location and memory allocation information are passed on to BioSM via

    startup scripts.

4. **How do I change the amount of memory allocated to java virtual machine?**

    Use a text editor to open the startup script BioSM.bat or BioSM.sh. Change the highlighted

    number to desired amount. (Note: 1024m = 1 GB; m = mega bytes)

    `java –Xmx1024m –jar BioSM.jar candidates.smiles`

    Recommended text editors:

    Windows OS: Notepad++, freely available at http://notepad-plus-plus.org/.

    Mac OSX: TextEdit or TextWrangler.

    Linux: gedit

5. **How can I provide BioSM with some candidate structures to classify?**

   First, copy the input file (with candidate structures to examine) into the input folder in the

   BioSM folder. Then use a text editor to open the startup script (BioSM.bat or BioSM.sh) and

   replace the highlighted text with the new file name.

   ```
   java –Xmx1024m –jar BioSM.jar candidates.smiles
   ```

   Please note that BioSM is a computationally intensive program. Providing an input file with

   a large number of candidate structures might result in delayed responses.

6. **I can't find the results generated by BioSM? The output folder has a lot of folders that I**

   don't recognize.

   Every time BioSM is run, a new folder (labeled with run date_time: `yyyyMMdd_HHmmss`) is

   created in the output folder in the BioSM folder. That is where you would find the results.

7. **What is the content of output folder generated by running BioSM?**

   ▪ *CpdsNotClassifiable.txt*: contains structures that were filtered out as they do not
      comply with BioSM's input requirements. The file specifies the reason for the
      elimination of each structure.
   ▪ *Detailed_Substructure_Results_Batch 1.xls*: contains the details of which scaffold
      matched with which candidate structure (as a substructure) and the similarity score.
   ▪ *Detailed_Superstructure_Results_Batch 1.xls*: contains the details of which scaffold
      matched with which candidate structure (as a superstructure) and the similarity
      score.
   ▪ **Final_Results_Batch 1.xls**: this is the main output file containing BioSM predictions per
      candidate structure.

8. **I have a question that was not addressed in this user guide.**

    Please contact the authors with your question(s) for further assistance

    mai@engr.uconn.edu or david.grant@uconn.edu