LiSIs: a Galaxy-based platform for Life Sciences Informatics Research

http://lisis.cs.ucy.ac.cy

http://www.granatum.org

Christos C. Kannas, Ioanna Kalvari, Ioannis Kirmitzoglou, Zinonas Antoniou, Kleo G. Achilleos, Christiana M. Neophytou, Christiana G. Savva, Christos A. Nicolaou, Vasilis J. Promponas, Andreas I. Constantinou and Costantinos S. Pattichis

Department of Computer Science, University of Cyprus

Department of Biological Sciences, University of Cyprus

1. Introduction

Life **S**ciences **I**nformatic**s** (**LiSIs**) platform, a new, open Scientific Workflow Management Systems (SWMSs), with several unique features designed to enhance user experience and facilitate user adoption. **LiSIs** is an online system based on the widely popular Galaxy SWMS. **LiSIs** provides five tool categories dedicated to small molecule virtual screening and, a selection of native Galaxy tools.

2. The LiSIs Platform

LiSIs aims to provide a set of tools to create, update, store and share Scientific Workflows (SWs) for the discovery of active compounds for biomedical researchers. The system is available via a web interface through a password protected, tiered login process. Specifically, the login process provides different level access to platform functionalities based on the user profile. The user is able to assemble SWs utilizing available in silico models and tools loaded into the platform. Depending on the user profile and associated permissions, users may also construct new models and tools through the development of custom workflows made available by the system for this purpose. Workflows execute on the system server. The execution results can also be stored on the user's **GRANATUM** workspace, where the user is able to access, manipulate or share them with other users.

LiSIs: Input Layer

UPLOAD DATA FILES
- Upload File from your computer
- GRANATUM File Loader

PARSE DATA FILES
- SDF File Reader
- SMI File Reader
- Property File Reader

Fig 1. Illustrates **LiSIs** Input Layer Tools. These tools are responsible for uploading and reading chemical and biological data files.

LiSIs: Pre-processing Layer

COMPOUND PRE-PROCESSING TOOLS
- Compound Fragmenter
- Descriptor Calculator
- Fingerprint Calculator

DOCKING PRE-PROCESSING TOOLS
- Coord Calculator
- Protein Cleaner

Fig 2. Illustrates **LiSIs** Pre-Processing Layer Tools. These tools provide essential chemoinformatics functionality for chemical descriptor calculation, 2D fingerprint generation, compound fragmentation, 3D coordinate calculation and protein cleaning.

LiSIs: USER processing Layer

FILTERS
- Chemical Properties Filter
- GRANATUM Ro5 Filter
- Lipinski Ro5 Filter
- Similarity Filter
- Diversity Filter
- Substructure Filter

USE AN EXISTING PREDICTIVE MODEL
- Property Predictor Predict a biological or chemical property of compounds using one of the existing Predictive Models

DOCKING MODEL PREDICTORS
- Vina Predictor

Fig 3. Illustrates **LiSIs** USER Processing Layer Tools. The tools are indented for use from novice users. They provide mainly filtering functionalities based on chemical, biological and structural characteristics of the compounds.

LiSIs: EXPERT processing Layer

BUILD AND EVALUATE A PREDICTIVE MODEL USING ANY OF THE FOLLOWING ALGORITHMS
- Linear SVM
- Decision Trees
- Random Forest
- k-Nearest Neighbors

DOCKING MODEL BUILDERS
- Dock Vina Builder

Fig 4. Illustrates **LiSIs** EXPERT Processing Layer Tools. The tools are indented for use from expert users, who have knowledge of classification algorithms and/or docking modelling. They provide functionalities for building Biological Property Prediction Models and Docking Prediction Models.

LiSIs: Post-processing Layer
- Output Reformater
- Binary File Merger

Fig 5. Illustrates **LiSIs** Post-Processing Layer Tools. The tools provide functionalities for manipulating intermediate and final results.

LiSIs: Output Layer

WRITE DATA FILES
- SMI Writer
- SDF Writer
- CSV Writer
- TAB Writer

SENT DATA FILES
- GRANATUM File Writer

Fig 6. Illustrates **LiSIs** Output Layer Tools. The tools provide functionalities for saving results in human readable format.

3. Third Party Tools powering LiSIs

Galaxy	Open, web-based platform for data intensive biomedical research, used for the customized SWMS platform.	http://galaxyproject.org/
RDKit	Open source cheminformatics toolkit, used to support all the cheminformatics functionalities.	http://www.rdkit.org/
The R Project	Statistical environment used to support data mining, machine learning and statistics related functionalities; caret (Classification and Regression Training) package is used for the generation of Predictive Models.	http://www.r-project.org/
AutoDock Vina	Docking prediction application used to support docking experiments functionalities.	http://vina.scripps.edu/

7. Acknowledgements

This work has been partially supported through the EU-FP7 GRANATUM project, "A Social Collaborative Working Space Semantically Interlinking Biomedical Researchers, Knowledge and data for the design and execution of In Silico Models and Experiments in Cancer Chemoprevention", contract number ICT-2009.5.3.

4. Showcase and Results

LiSIs has been used for the implementation of a VS experiment in order to identify molecules able to bind to Estrogen Receptor-α (ER-α) and/or Estrogen Receptor-β (ER-β).

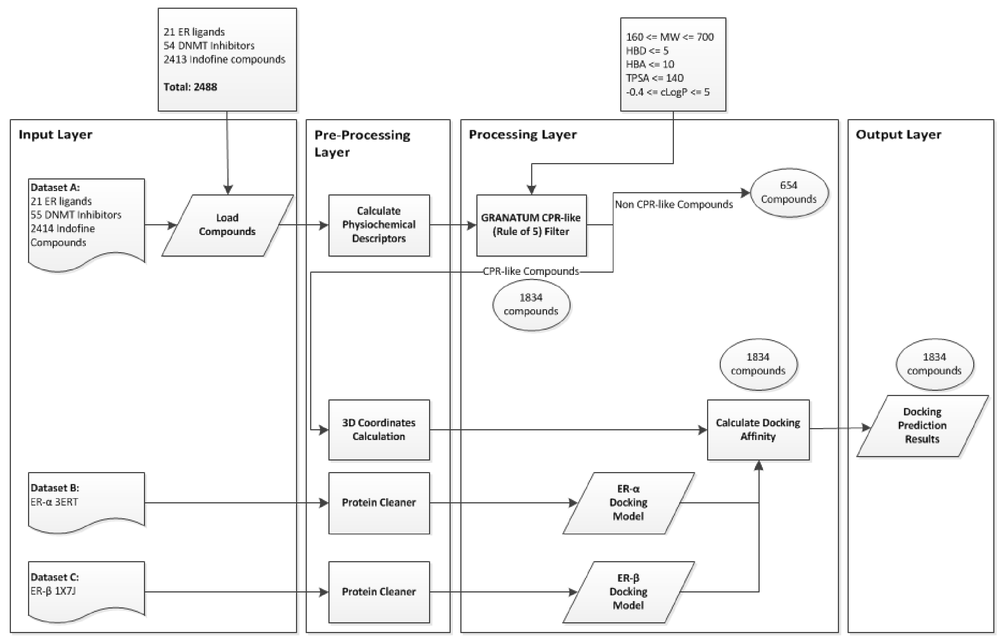

Fig. 7 illustrates the schematic of the workflow used for the VS experiment.

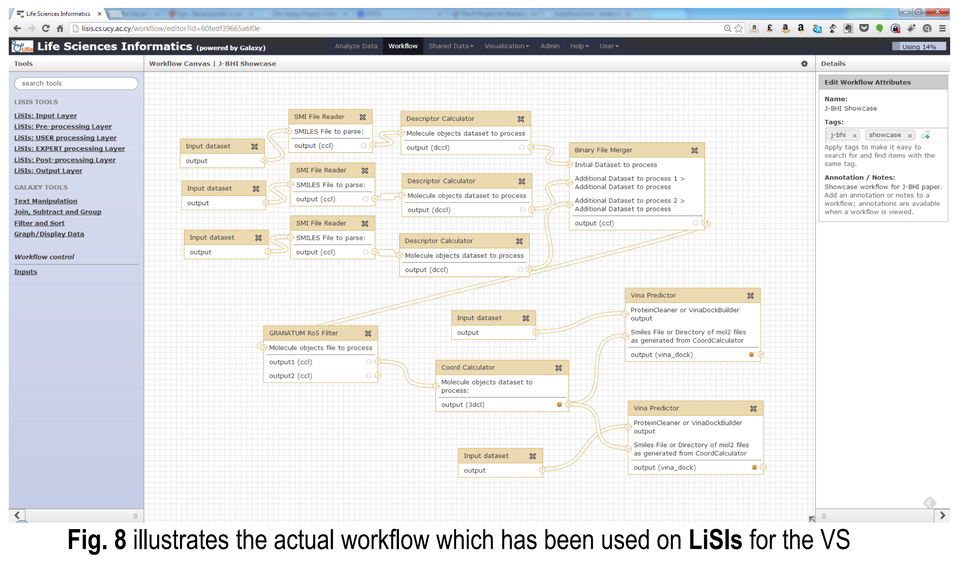

Fig. 8 illustrates the actual workflow which has been used on **LiSIs** for the VS experiment.

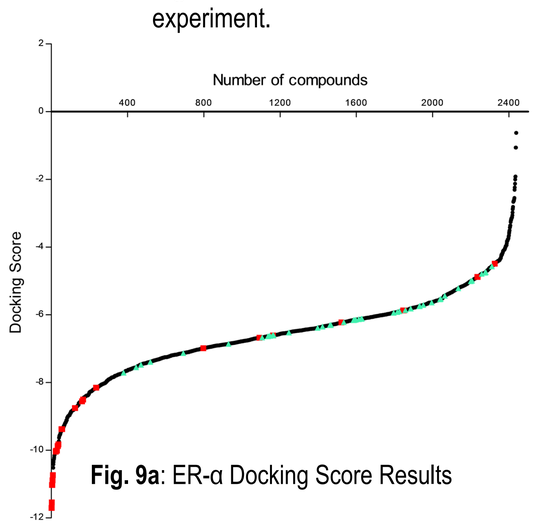

Fig. 9a: ER-α Docking Score Results

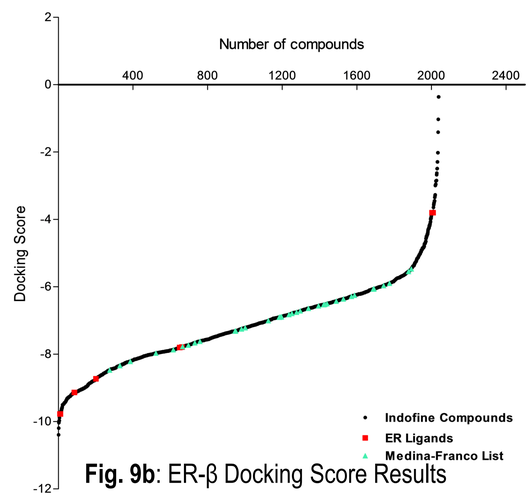

Fig. 9b: ER-β Docking Score Results

Finally a selection of molecules highly ranked was hand-picked. These molecules have undergone in vitro investigation to provide feedback for the calibration of the tools used by **LiSIs** platform and also to select a small set for further research.

5. Discussion

- In the present study, we used a Virtual Screening Workflow implemented using the **LiSIs** platform to screen the Indofine database of 2413 compounds.
- Based on their drug-like criteria and docking results we selected 18 potential ER ligands.
- Further in vitro investigation showed that:
  - Five (5) agents displayed strong affinity for ER-α,
  - Three (3) showed selectivity for ER-β and
  - Seven (7) were able to bind to both receptors with similar affinity.
- **In total 15 out of 18 compounds (83.3%) were experimentally confirmed active**.
- Therefore, the use of **LiSIs** system may allow researchers to execute complex biomedical studies and in silico experiments on largely available and high quality data repositories in order to facilitate the selection and prioritize the investigation of novel chemopreventive compounds in vitro.

6. Conclusion

- The **LiSIs** platform aims to fill the current void in the application of advanced cheminformatics and computational chemistry technology in determining efficacy and predicting possible mechanism of action or identifying a possible receptor for a chemopreventive agent in life sciences research.
- Its successful deployment may have a substantial impact on enabling biomedical researchers to utilize state of the art computational techniques to search for promising chemical compounds that may lead to the discovery of novel agents with chemopreventive properties.
- Utilizing the **LiSIs** platform in conjunction to a widely used docking program we identified compounds that can bind to ER-α and/or ER-β with a high degree of success rate.
- This in silico approach is expected to facilitate the process of identification of lead compounds with estrogenic or anti-estrogenic activity and to enhance considerably the discovery process for new therapeutic agents.