TTCS: Three-Dimensional and Two-Dimensional Compound Structure Search Online Tools

Hongzhi Li, Haiqing Li, Edward Lee and Yate-Ching Yuan
Bioinformatics Core Facility, Department of Molecular Medicine, Beckman Research Institute,
City of Hope Medical Center, Duarte, CA 91010, USA

Abstract - *TTCS is a web based online tool to perform Two- and Three-Dimensional Compound Similarity structure search. It performs compound similarity search from several in-house compound libraries integrated with millions of public available compound libraries such as NCI DTP database, NCI FDA-approved Drug database, and DrugBank, etc. TTCS offers user friendly web interface for 2D and 3D structure query with features such as tracing user's search history, job queuing, and visualization. The website is specially designed for researches' need in bioactive chemical discovery and lead optimization for drug discovery.*

Keywords: similarity search, drug discovery, ligand-based drug discovery, three-dimensional similarity, cheminformatics

1 Introduction

Computer-assisted drug design can reduce the number of potential drug candidates from millions of compounds to tens for experimental synthesis and validation. There are mainly two types of approaches, structure-based and ligand-based drug design. Compound structure similarity search is widely used cheminformatics tool to identify clusters of structural similarity compounds for lead optimization in ligand-based drug discovery. TTCS offers the compound similarity searching measured by calculating the Tanimoto coefficient of the lead compound two-dimensional (2D) fingerprints[1], as well as comparing 3D pharmacophore fingerprints[2] and comparing with integrated large bioactive compound libraries with purchase information.

Structure similarity search by using 2D molecular fingerprint is one of the common methods utilized for comparing large compound libraries in *silico*. TTCS integrated a suite of compound structure similarity search tools which are effectively parallelized running on a scalable ScaleMP computation server to search millions compound libraries with reasonable accuracy. The similarity can be determined by calculating the Tanimoto coefficient of the 2D fingerprints and 3D pharmacophore patterns. For example, DiverseSolution BCUT descriptors from Tripos capture the properties of the molecule in chemical space, including topological or distance information, such as hydrogen bonds between atoms in a molecule. Many BCUT descriptors are calculated for a compound library/database, but only the most important 3-6 descriptors that are most useful for the diverse chemical space of the library. The pre-selected chemical space determined as the nearest neighbor representing a clusters of structural similarity compounds for lead optimization in ligand-based drug discovery.

Several public available compound databases, such as NCBI PubChem[3], UCSF ZINC[4] or NCI DTP integrates 2D fingerprint similarity search engine on their online tools. However, none of them provides both 2D and 3D pharmacophore fingerprints similarity search tools due to computational limitation. In this study, we present an online comprehensive similarity searching tools Two- and Three-Dimensional Compound Similarity structure search (TTCS) to provide users with more searching functionality to efficiently narrow down clusters of potential leads from several in-house commercial compound libraries integrated with large public available compound libraries such as NCI DTP database, NCI FDA-approved Drug database, and DrugBank database for further validation in ligand based drug discovery.

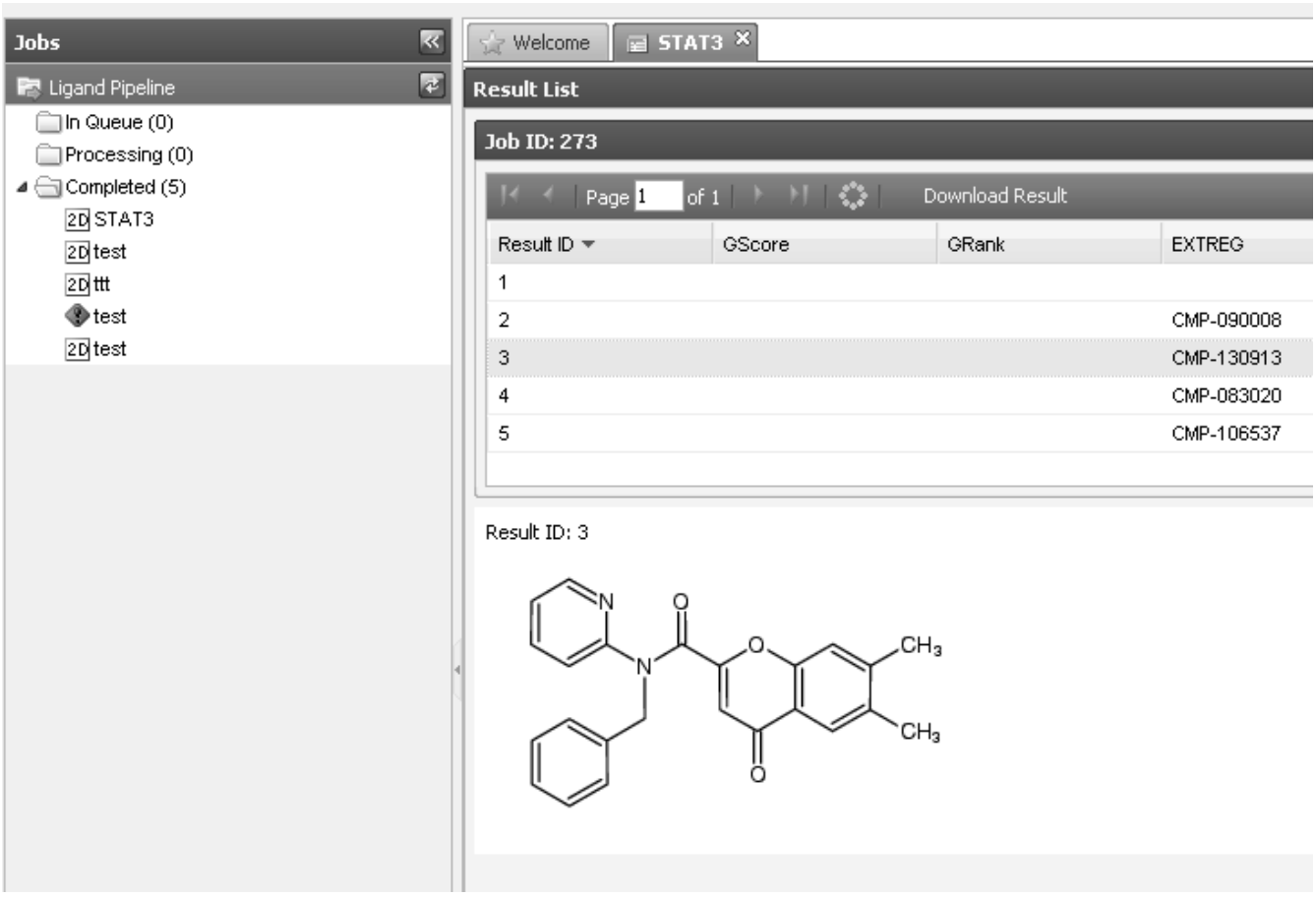
2 Methods and Design

2.1 Methodology

The 2D fingerprint similarity of two compounds is determined by the Tanimoto coefficient:

$$Tc = \frac{Nab}{Na + Nb - Nab}$$

where *Na* is the number of the bits in the 2D fingerprint of the first molecule, *Nb* is that of the second molecule, and *Nab* is the number of common bits for both molecules. The chemistry space for 3D pharmacophore fingerprint searching is selected from the best 4 BCUT descriptors that provide the best separation of the compound library. The low-dimensional BCUT chemistry space is divided into cells and the compounds in the same cell of the target compound are selected as the similar compounds.



2.2 Web Design

The TTCS website is specially designed with a user friendly web interface for researchers to perform effective 2D and 3D structure similarity query to explore large chemical libraries. The parallelization of the 2D and 3D similarity algorithms runs as back-end on a scalable ScaleMP HP LINUX servers. As shown in the figure, the web application can trace user's search history, summary status for the batch job and display the output compounds in image. It also provide results downloading in different format, email results to the user, and many other features.

3 Conclusions

TTCS, a user-friendly online compound similarity tool that combines the 2D fingerprint searching and 3D pharmacophore fingerprint searching for lead optimization in ligand based drug discovery. It will be available to public soon for users' feedbacks.

4 References

[1] Peter Willett. "Similarity-based virtual screening using 2D fingerprints", Drug Discovery Today, Volume 11, Issues 23-24, 1046-1053, December 2006

[2] Brett R. Beno and Jonathan S. Mason**. "**The design of combinatorial libraries using properties and 3D pharmacophore fingerprints", Drug Discovery Today, Volume 6, Issue 5, 251-258, March 2001

[3] Bolton E, Wang Y, Thiessen PA, Bryant SH. "PubChem: Integrated Platform of Small Molecules and Biological Activities". Chapter 12 in *Annual Reports in Computational Chemistry*, Volume 4, American Chemical Society, Washington, DC, April 2008

[4] Irwin JJ and Shoichet BK. "ZINC—a free database of commercially available compounds for virtual screening", J Chem Inf Model, Volume 45, Issues 1, 177-182, Jan-Feb, 2005