Welcome to ProteinGoggle 2.0.2, a top-down intact protein database search engine implemented with isotopic mass-to-charge ratio and envelope fingerprinting (iMEF) search algorithm. ProteinGoggle is developed to be run on PCs with Windows operation systems.

ProteinGoggle User Manual

Contents

1. Installation

1.1 MySQL

Please acknowledge separately (other than ProteinGoggle 2.0) that "All ProteinGoggle customized databases are managed by MySQL Server 5.0".

Download MySQL Server 5.0.96 (Microsoft Windows 32. (ZIP format)) from http://downloads.mysql.com/archives.php?p=mysql-5.0&o=-win.

Run "MySQL Server Instance Configuration Wizard" following the wizard with default setting. Please make sure the dedicated port to be "3306" and set "root" as the password for the "root" account.

Restart the computer.

1.2 MySQL Tools

Download MySQLQueryBrowser 1.1.20 (Microsoft Windows (ZIP format) **(7 Feb 2006, 4.9M)**) from

http://downloads.mysql.com/archives.php?p=mysql-query-browser&o=-win .

Download MySQLAdministrator 1.1.9 (Microsoft Windows (ZIP format) **(16 Feb 2006, 4.9M)**) from

http://downloads.mysql.com/archives.php?p=mysql-administrator-suite&o=-win. Just place the folder in some place of your computer; no installation is needed.

Connect My SQL with ProteinGoggle.

Run "MySQLQueryBrowser.exe" to get the GUI shown in **Figure 1**;

Click the square button with "…" to see the GUI shown in **Figure 2**;

Click "New Connection" to see the GUI shown in **Figure 3** and do the input: Connection=ProteinGoggle, Username=root, Password=root, Hostname=localhost, Schema=test. After input, click "Apply" and then "Close". Now you get the GUI shown in **Figure 4**. Click "OK" to finish connecting ProteinGoggle with My SQL.

Database deletion and sharing

All databases created through ProteinGoggle are shown In "MySQLQueryBrowser".

A database could be deleted through right-click "Drop Schema".

Databases could be shared among computers by using "Restore" and "Backup" buttons in "MySQLAdministrator", which could be opened by clicking "MySQLAdministrator.exe" and connected with inputs shown in **Figure 5**. The password here is also "root".

Change a default setting of My SQL

To be safe, follow "Startup Variables→Advanced Networking→Data/Memory size→Max.packet size" and set "max_allowed_packet" as 100M or much larger as shown in **Figure 6**.
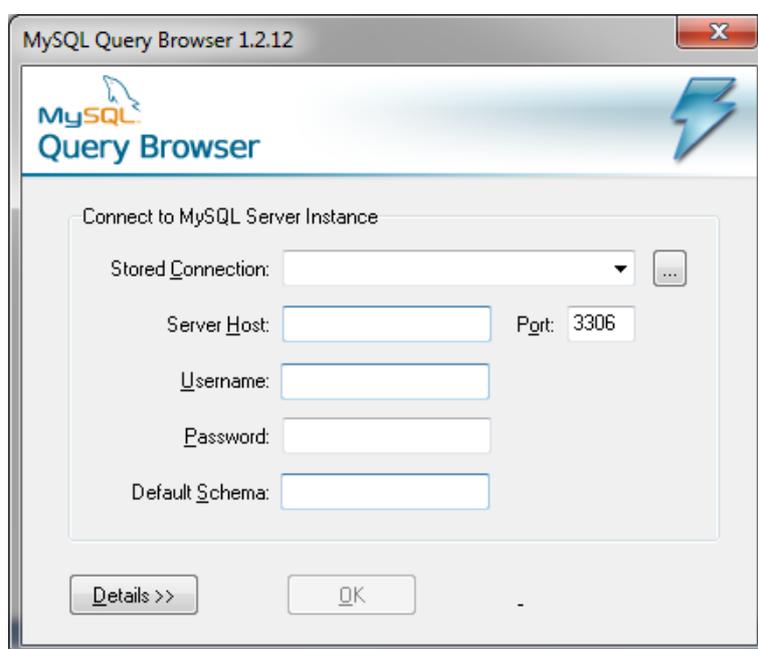

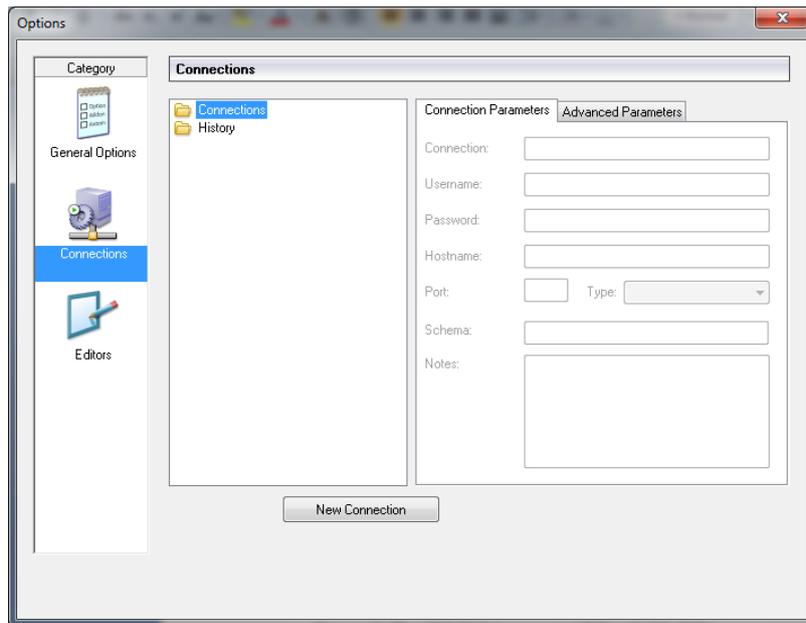**Figure 1.** The 1st GUI to connect ProteinGoggle with My SQL.

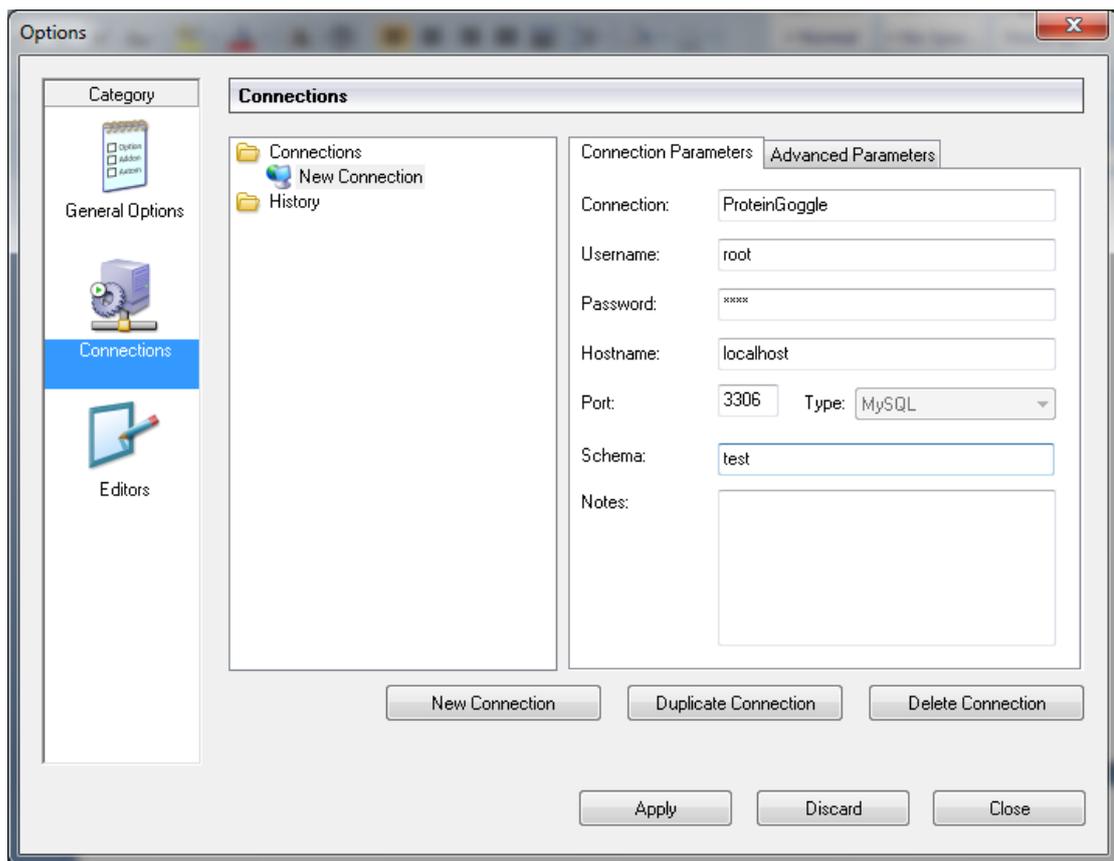**Figure 2.** The 2nd GUI to connect ProteinGoggle with My SQL.

Figure 3. The 3rd GUI to connect ProteinGoggle with My SQL.
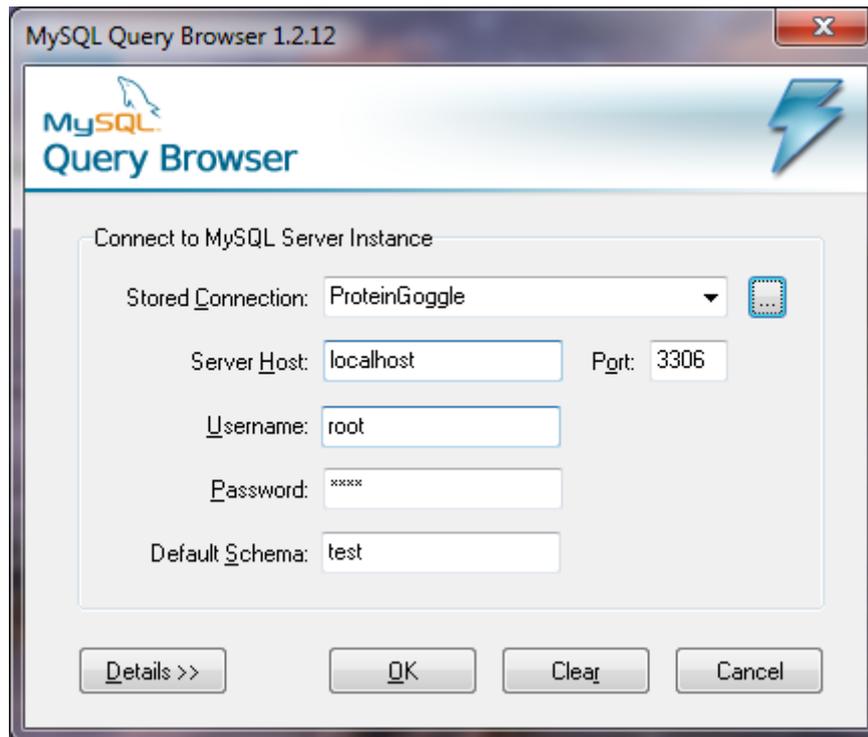
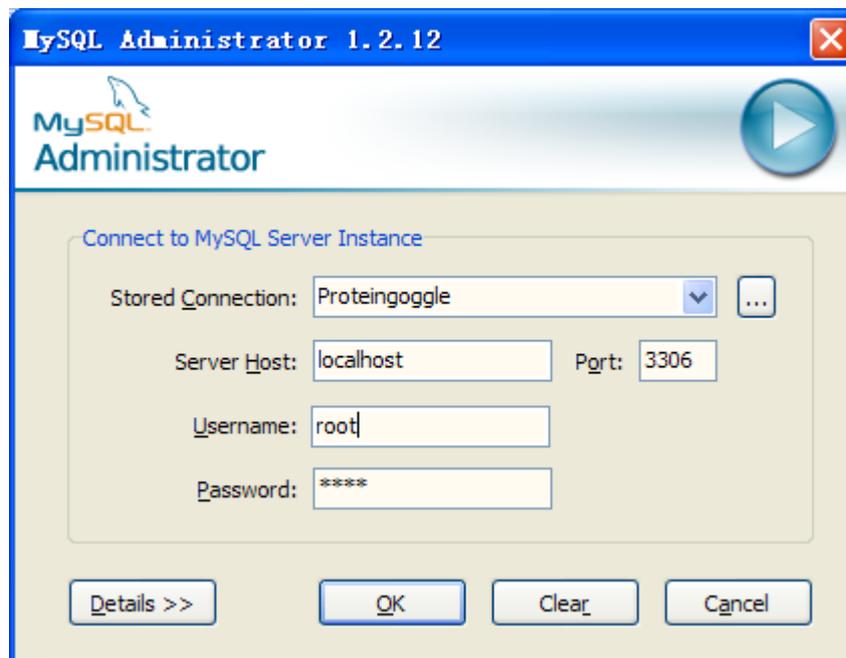**Figure 4.** The 4th GUI to connect ProteinGoggle with My SQL.

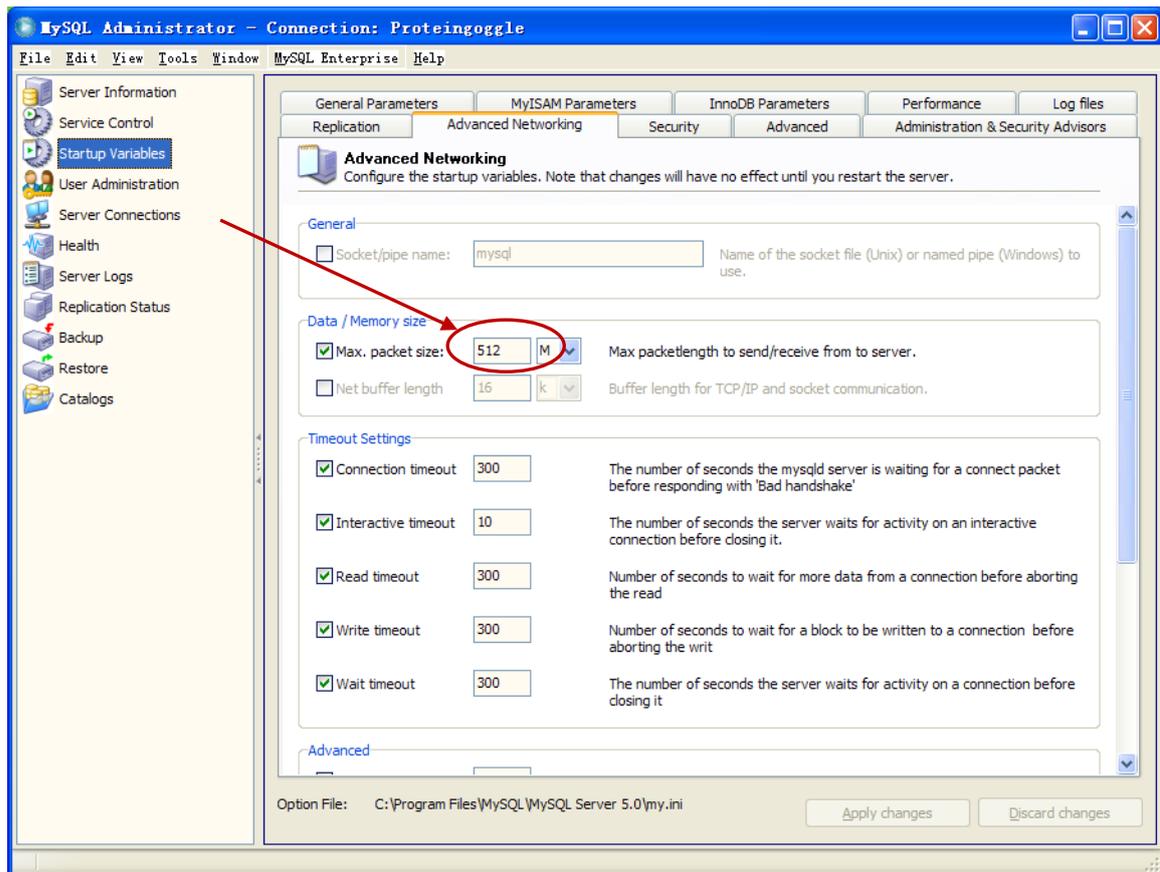**Figure 5.** The connection of MySQLAdministrator and MySQL Server

Figure 6. Change of default "max_allowed_packet" setting.

1.3 Microsoft .NET Framework 4.0

Download from
http://www.microsoft.com/en-us/download/details.aspx?id=17718 and choose
"dotNetFx40_Full_x86_x64.exe".

1.4 ProteinGoggle

Download from http://proteingoggle.tongji.edu.cn/download/pg and
decompress the file. Just place the unzipped folder anywhere in your computer,
no installation is needed;

2. Database search of an intact protein MS/MS spectrum using ProteinGoggle

Run "ProteinGoggle.UI.exe".

Follow "menu→database→create a new database" to open the "Create a Database" GUI. Open the downloaded flat_txt file (www.uniprot.org, e.g., input "P62975" to search for ubiquitin), and then set the parameters of static or dynamic PTMs, etc. Check the "MS2" box on lower right corner and click "Execute".

To run the search on the main GUI, open the MS/MS spectrum and choose the corresponding database. The default search parameters are optimized at the proteome level with FDR control. Note: click the "refresh" button to use the database you just created without restarting ProteinGoggle.

It should be noted that ProteinGoggle reads .mzXML file directly; if you want it read Thermo .raw file directly, you need to install the newest Xcalibur on your computer.

3. Database search of an intact proteome DDA-LC-MS/MS dataset

Run "ProteinGoggle.UI.exe".

Follow "menu→database→create a new database" to open the "Create a Database" GUI. Open the downloaded flat_txt file (www.uniprot.org, e.g., use the criteria of 'Organism [OS]' = escherichia coli, 'Sequence_Fragment' = No, 'Sequence_Sequence length' = 1–200, and 'Reviewed' = Yes to search for proteins of *E. coli*), and then set the parameters of static or dynamic PTMs, etc. Click "Execute". Do NOT check the "MS2" box on lower right corner.

To run the search on the main GUI, open the MS/MS dataset and choose the corresponding database. The default search parameters are optimized at the proteome level with FDR control. Note: click the "refresh" button to use the database you just created without restarting ProteinGoggle.

Please note: 1) if you datasets contains more than one dissociation methods, just choose all of them when you create the corresponding database; ProteinGoggle recognizes them automatically and uses appropriate fragments; 2) currently only Thermo Orbitraps (including Exactive Series) DDA spectra or datasets have been tested in ProteinGoggle, if you have other type of sepctra or datasets, we are happy to work with you as well; 3) to search a dataset with thousands of spectra, it is good to specify the "Spectrum_Range" having useful MS/MS spectra to save the search time; 4) For FDR control, decoy search with a "random" database normally gives you more proteoform IDs than a "reverse"

database.

4. Notice

The use of ProteinGoggle is currently free. All publications that result from the use of this software should cite the research article(s) from the developer on the "References" page at .

ALL RIGHTS IN THIS COMPUTER SOFTWARE ARE RESERVED BY THE DEVELOPER. YOU MAY NOT MODIFY OR REDISTRIBUTE THIS SOFTWARE. YOU TAKE FULL LIABILITY FOR THE USE OF THIS SOFTWARE.