Using machine learning in Mascot Server 3.1 with Mascot Distiller

Version: 2025-02-10
Author: Matrix Science Ltd

Introduction

Mascot Distiller communicates with Mascot Server via HTTP/HTTPS and is compatible with Mascot Server 2.0 and later, including the latest versions, Mascot Server 3.0 and 3.1.

With Mascot Server 3.1, you can automatically refine results with machine learning and import the refined results into Mascot Distiller. Machine learning is optional and disabled by default. If you do not wish to enable it, then you do not need to follow the instructions below, as Mascot Server 3.1 works with Mascot Distiller without extra steps.

If you enable machine learning, the table below illustrates the expected improvement, using a subset of Thermo Orbitrap QE HF-X raw files from PRIDE project PXD028735.

Mascot Server	**Protein Hits**	**Peptide Sequences (1% FDR)**
3.0 (no refining)	5089	27080
3.1 (with MS2PIP:HCD2021 instrument)	5961	34222

Requirements

Mascot Server: version 3.1

Mascot Distiller: version 2.8.5.1

Three steps are needed:

1. Configure a new instrument definition in Mascot
2. Update Mascot Distiller 2.8.5.1
3. Increase the HTTP Timeout setting in Mascot Distiller
4. Create a new single or multifile project in Mascot Distiller using the new instrument definition and with automatic decoy searching enabled

MATRIX SCIENCE

1. Configuring a new instrument definition in Mascot

Mascot Distiller 2.8 and earlier does not have a user interface for selecting an MS²PIP model. Instead, the model can be configured as part of the instrument definition.

Go to your local Mascot home page and select Configuration Editor. Go to the Instruments panel.

Mascot Configuration

Amino Acids	Amino Acid Data
Modifications	Modification definitions
Symbols	Symbols used in chemical formulae
Linkers	Linker definitions
Enzymes	Enzyme definitions
Instruments	Fragmentation Rules
Quantitation	Quantitation Methods
Crosslinking	Crosslinking Methods
Configuration Options	Global Options in mascot.dat
Database Manager	Sequence databases, Parse Rules and automated downloads

Create a new instrument definition. Select the fragmentation series based on an existing instrument. For example, if you regularly use the ESI-TRAP instrument when submitting searches, select the same ion series:

Instruments

Ion series	New	Default	ESI QUAD TOF	MALDI TOF PSD	ESI TRAP
1+	☑	X	X	X	X
2+ (precursor>2+)	☑	X	X		X
2+ (precursor>3+)	☐				
immonium	☐			X	
a	☐	X		X	
a*	☐	X		X	
a0	☐			X	
b	☑	X	X	X	X
b*	☑	X	X	X	X
b0	☑		X	X	X
c	☐				
x	☐				
y	☑	X	X	X	X
y*	☑	X	X		X
y0	☑		X		X
z	☐				

Enable refining with machine learning and choose the desired model. Give it a name that includes the model's name, so that it is easy to differentiate between instruments.

If you want to use several different MS²PIP models, for example for different instrument or experiment types, add a new instrument definition for each one.

For Thermo instruments, Mascot Server ships with four suitable models:

- **HCD2021**: HCD fragmentation. Use for qualitative studies as well as label-free quantitation.
- **CID**: CID fragmentation. Use for qualitative studies as well as label-free quantitation.
- **TMT**: Experiments using TMT or TMTpro labels.
- **iTRAQ**: Experiments using iTRAQ labels.

If you have a Sciex instrument, the **TTOF5600** model was trained on qtof data from a Sciex Triple-TOF 5600. Use this for qualitative studies as well as label-free quantitation.

There are two models for Bruker timsTOF data**, timsTOF2023** and **timsTOF2024**. Use these for qualitative studies as well as label-free quantitation.

2. Update Mascot Distiller

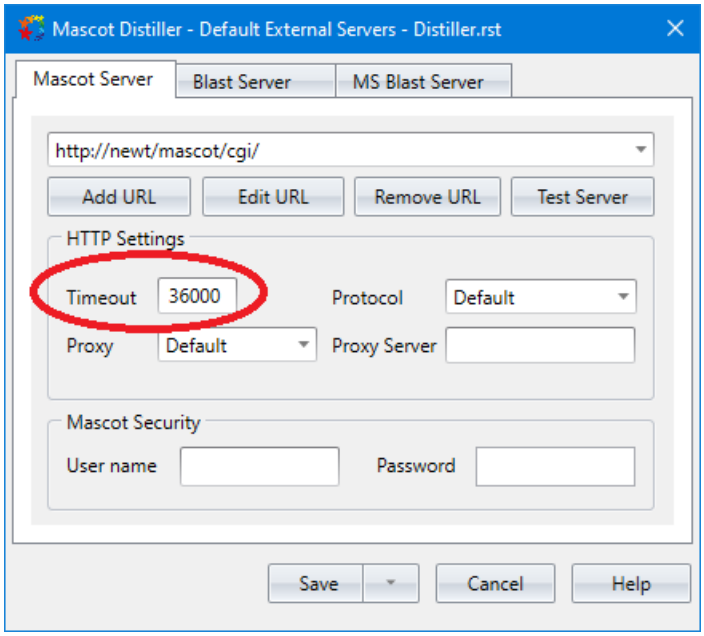
Download Mascot Distiller 2.8.5.1 and install the update:

https://www.matrixscience.com/distiller_download.html

This point release is a free update that fixes an issue with HTTP/S timeout handling. It is required for the machine learning integration to work.

3. Mascot Distiller configuration

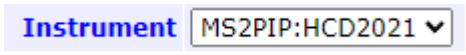
Once you have updated to Mascot Distiller 2.8.5.1, open the software and without creating a project, open the main Mascot Distiller workstation GUI and open the Tools->External Servers dialog. On the Mascot Server tab, under HTTP Settings, increase the Timeout value from the default of 60 seconds to a much higher number such as 36000 (10 hours). This is required because the connection between Mascot Distiller and the Mascot Server will have to be held open while the results are refined on the Mascot Server. During this period, no data will be sent from the Mascot Server to Mascot Distiller, and without increasing the timeout setting the connection will be terminated by Mascot Distiller before refining has completed.



Note that changing this setting will change the default HTTP timeout for all future projects created. For any pre-existing projects, the setting will be the old default value and you will need to change the timeout on a project by project basis if you wish to resubmit searches from an existing dataset.

4. Submitting a Mascot Search

To enable the machine learning, submit the search from either Mascot Distiller or Mascot Daemon with the instrument with machine learning refinement enabled selected on either the search form in Mascot Distiller, or in the parameter set in Mascot Daemon.



In Mascot Daemon you should also deselect any machine learning settings from the parameter set you are using to avoid confusion between the settings in the parameter set and from the instrument definition.

Important: machine learning in Mascot 3.1 requires that an automatic decoy search is carried out.

If you're submitting a search from Mascot Daemon, make sure that the "Decoy" checkbox is checked in the parameter set:

If you're submitting from within the Mascot Distiller, make sure the "Decoy" checkbox is checked on the search form:

There is currently no progress reported from Mascot Server 3.1 to Mascot Distiller during the machine learning refinement step so during this stage Mascot Distiller will show 0% progress for an

extended period. This will be fixed in the next release of Mascot Distiller. For further details, please see Troubleshooting ML integration with Mascot client software