deltaMasses™

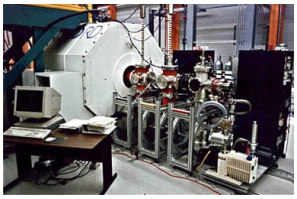
Differential PTM Detection
after High-Accuracy Mass Spectrometry

User Manual

version 5.2
build 787

20130127

Table of Contents


Flash Tutorials click to view (internet connection needed):

Installation

Upgrading to Discovery Edition

The deltaMassBase PTM fingerprint

5 minute tutorial on differential PTM detection

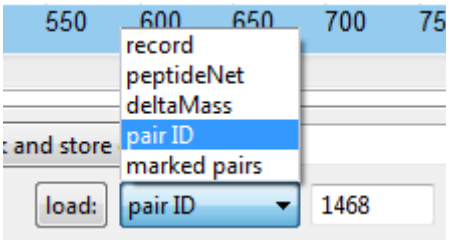
0 New in versions 5

Since Version 5.0, deltaMasses is only being distributed for Windows 64 bit architectures (both Windows 7 and Windows 8). Multicoreprocessing is now supported.

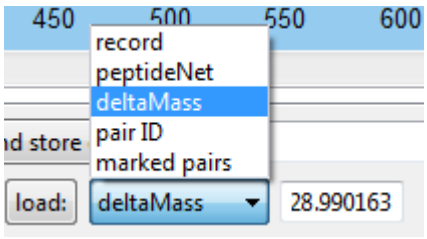
It is now possible to reload data easily into deltamasses. This assumes that you have the database installed. You can reload data based on record-id, deltaMass, pair-id, and based on your “favorites”. Commenting a pair is another new feature.

.

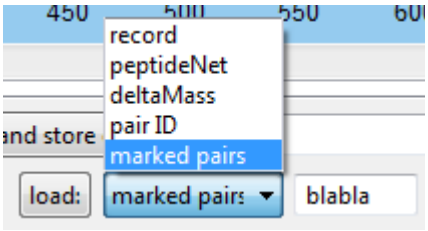
Example 1: to reload pair_id 1468, use the dialog at the „load“ dialog at the bottom of the application:



Example 2: to load all detected pairs with a deltaMass of 28.990163 [Dalton], corresponding to an S-Nitrosylation, use the loader as indicated below:



Example 3: to reload your „marked pairs“ or „favorite“pairs, use the „marked pairs loader. The text right to the load selector does not matter.



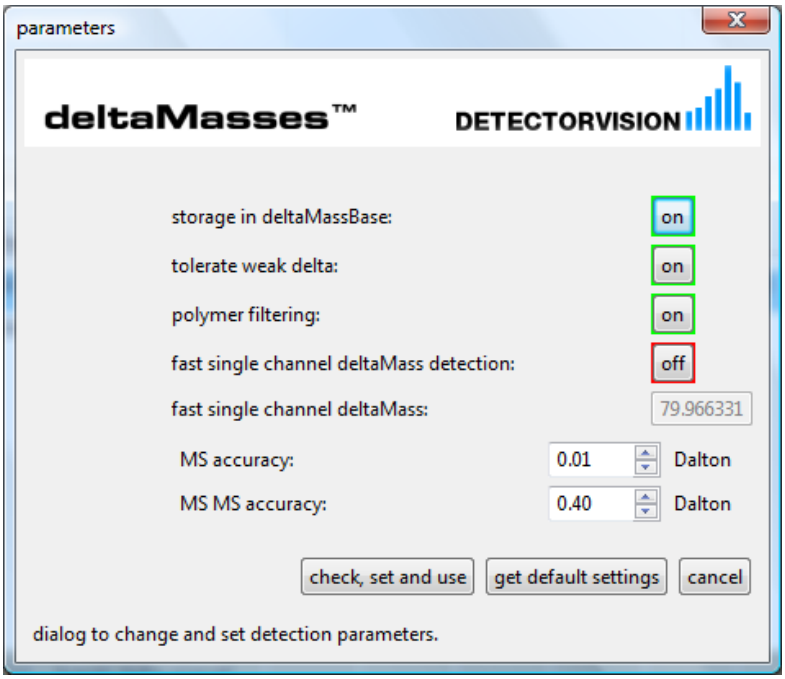
You can set detection parameters using the parameters dialog, see picture below.

Storage in deltaMassBase: If you switch this on, the results get stored automatically in the database. Default is off. If it is switdched off, you can manually store it using the new „store“ button

Tolerate weak delta: Default off. If you switch this on, pairs with a weak deltaMass signal are tolerated. This is especially interesting if you look for glycolysations and other unstable modifications. It is our first step in the direction of glycobiology. Use weak delta results with great care !

Polymer Filtering: Default off. Filters polymer-like spectra.

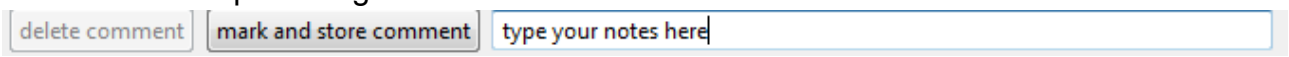
Fast single channel detection: Default off. If you switch this on, deltaMasses changes from „all channel detection“ to „single channel detection“; you will only detect signals of the fast single channel deltaMass. I.e. To detect phosphorylations only, switch it on, set the fast single deltaMass to 79.9663 [Dalton], and go ahead. Advantage: this is up to 2000% faster than „all channel detection mode”.



detection parameters dialog.

Favorites: mark and store comments on pairs in deltaMassBase:

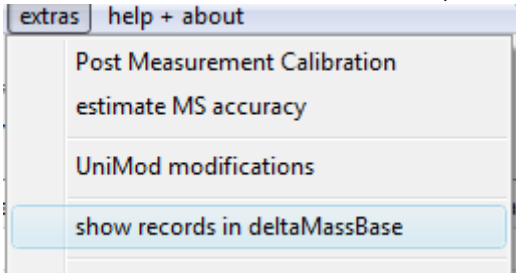
Just use the simple dialog in the interface:



You can reload your favorite by “reload marked pairs”, as described above. To delete and unmark a comment, use the delete comment button.

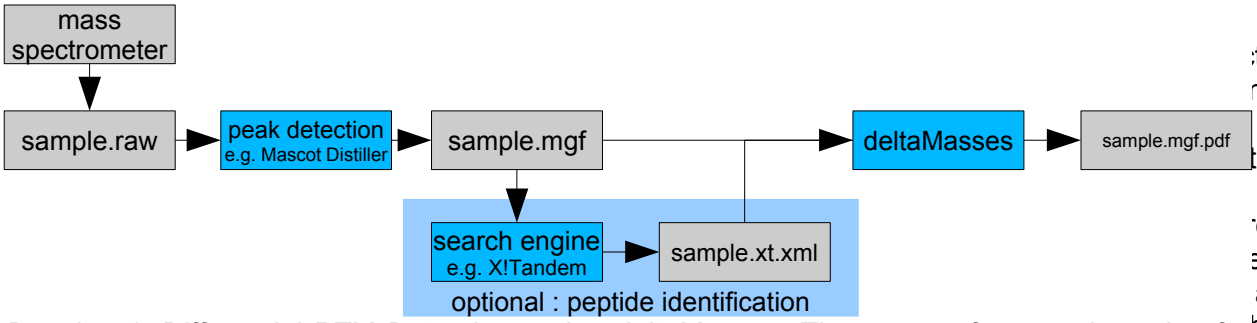
You can now list and delete records from deltaMassBase,

just go to extras->show records in deltaMassBase, and go ahead.



1 Differential PTM Detection

The analysis of proteins using collision induced mass spectrometry has been practiced for several years[i]. Recently, high-accuracy mass spectrometers with low-ppm mass accuracy became availble on the market[iiiiiiv] resulting in outstanding quality improvement. Peptide identification is often performed with classic search engines like Sequest [v], Protein Prospector, Mascot, X!Tandem etc.. These approaches rely on the comparison of synthetically calculated spectra with experimentally observed ones. They also depend on the usage of protein/DNA/mRNA sequence information. The classic search engines have well-known issues when used for the detection of post-translational modifications (PTM's). A new approach, ICATCHER[vi], had been described and implemented in March 2005, using **Differential PTM Detection (DPD)**. A similar approach was published under the name Modificomb in January 2006[vii]. DeltaMasses is the commercial strength implementation of Differential PTM Detection (DPD). Since the original March 2005 version, major functional extensions have been added; the most crucial being statistical scoring.



Drawing 1: Differential PTM Detection using deltaMasses. The usage of a search engine for peptide identification is optional but recommended and necessary if you want to localize the PTM. Input to deltaMasses is a peaklist in .mgf format (Mascot Generic Format). DPD can be performed on high-accuracy mass spectrometers only. MS accuracy should be below 10ppm.

Illustration 1: Differential PTM detection compares pairs of MSMS spectra against each other to detect and localize PTMs. Top: unmodified peptide. Below: modified peptide with a modification of weight 156.1[Da] located on T(8).

DPD is not an alternative to the classic search engines but a complementary new technology.

Table 1 compares the two approaches and how they supplement each other.

	Classic Search Engine (CSE)	***Differential PTM Detection (DPD)***
main purpose	peptide identification	PTM detection and localisation
depends on protein/DNA sequence information	yes	no
needs PTM knowledge to detect PTMs	yes	no
can detect unexpected modifications	no	yes
can detect a modification that is always present on a peptide	yes	no (unless unmodified peptide synthesized)
PTM-based biomarker discovery	limited	ideal
automation	limited	yes
useful information during measurement	yes	yes
makes assumptions about fragmentation patterns	yes	no
Localisation of PTM's	possible	better than the CSE
can identify a peptide	yes	no

Table 1: relation between DPD and the classic search engine – best used together.

As can be seen from above comparison, the approaches are complementary. The strength of the DPD is the usage for differential PTM detection, biomarker discovery or for functional studies. deltaMasses integrates with Mascot and X!tandem (Mascot recommended)

2 System Requirements

There are minimum requirements for the mass spectrometer and the computer on which you install deltaMasses:

Mass Spectrometer minimum requirements:

- Low-ppm-range MS accuracy
- Recommended minimum mass accuracy for MSMS is +/- 0.3 Dalton.

- Ability to export peaklist files in Mascot generic format (.mgf) with correctly determined precursor charge state. If you use Mascot as your search engine, it is sufficient to feed the mascot.xml output into deltaMasses, see section „Mascot Integration“ of this manual. We recommend Mascot Distiller for extraction of peak lists.
- Note: the statistical significance of results depends drastically on MSMS accuracy, see the statistics section
- Most measurements can be improved significantly by post-measurement callibration. In case you do cross-measurement PTM detection, this is strongly recommended. For PTM detection within a single measurement (run), deltaMasses is self-callibrating.
- Peak detection: aim for quality! When you transform your raw data to peaklists, it is an advantage if your software is able to merge scans in the time domain. This is not necessary yet recommend: your searches will run faster, your signals will get better and you avoid having multiple hits for the same peptide pair. We recommend Mascot Distiller for this task.
- MSMS signals need more than 10 signal peaks to yield DPD results.

Computer minimum system requirements:

- Windows 7 64 bit or Windows 8 64 bit.
- 8 Gigabyte RAM
- 2.5 GHz processor or faster.
- 200 Megabytes free diskspace on the C:\ drive.
- WSXGA display (1680 pixels × 1050 pixel)
- Ability to install to the directory C:\detectorvision\deltaMasses\
- .pdf reader installed
- Multicore processor is good to have.

As a general rule, one should not install more than the absolutely necessary software on an instrument PC. This is also true for deltaMasses. Never install deltaMasses onto your mass spec computer; deltaMasses consumes CPU power which might interfere with the functionality of the mass spectrometer.

3 Download, Installation, First Test

Usually, you do not have to use an administrator account to install deltaMasses. Try to install with your normal user account. An exception is when the C:\ drive is write-disabled for your normal user account. Please contact your system administrator in this case. Shortcuts to deltaMasses (Start->Programs->Detectorvision->deltaMasses, Desktop) are created only for the user account used during installation. deltMasses is a single-user desktop application, i.e. It is best installed on your „private“ desktop / laptop.

Registration and download:

1. request a free personal edition license from info@detectorvision.com
2. you will receive your license and download details by email (typically within 24 hours)
3. Download
4. For a Discovery Edition, you have to install a personal edition first. We request users to test the Personal Edition for at least a week, including deltaMassBase, before upgrading to Discovery Edition. Price 2011-09-01: 4000.- Euro subject to change at any time. Quotation required.

To install the database (deltaMassBase), please refer to Appendix A.

Uninstalling deltaMasses:

Start->Programs->DETECTORVISION->deltaMasses->Uninstall

First Test

A first test is always good to perform:

Start->Program Files->DETECTORVISION->deltaMasses

Click Open

Browse for the file

C:\detectorvision\deltaMasses\data\example.mgf

Click analyze.

The example.mgf file consists of 3'799 MSMS measurements; deltaMasses will thus compare 7'214'301 spectrum pairs. Analysis should take about 30 seconds on a 1.7 Ghz PC with 1361 spectrum pairs detected. On average, we expect 300'000 pair comparisons per second; the variation of this number is substantial.

For the example.mgf file, you should find the following modifications:

A screenshot of the example.mgf testrun is shown below showing an N-terminal carboxymethylation of a yeast protein.

If anything fails, please contact info@detectorvision.com +41 79 46 180 72

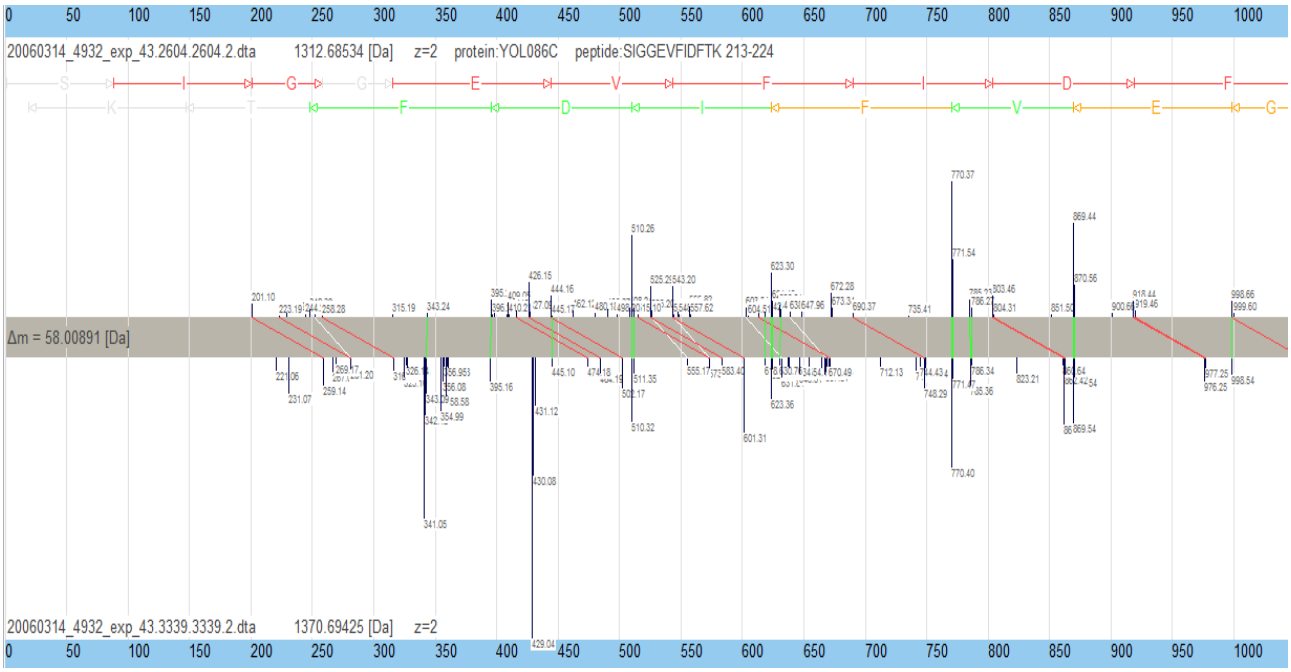
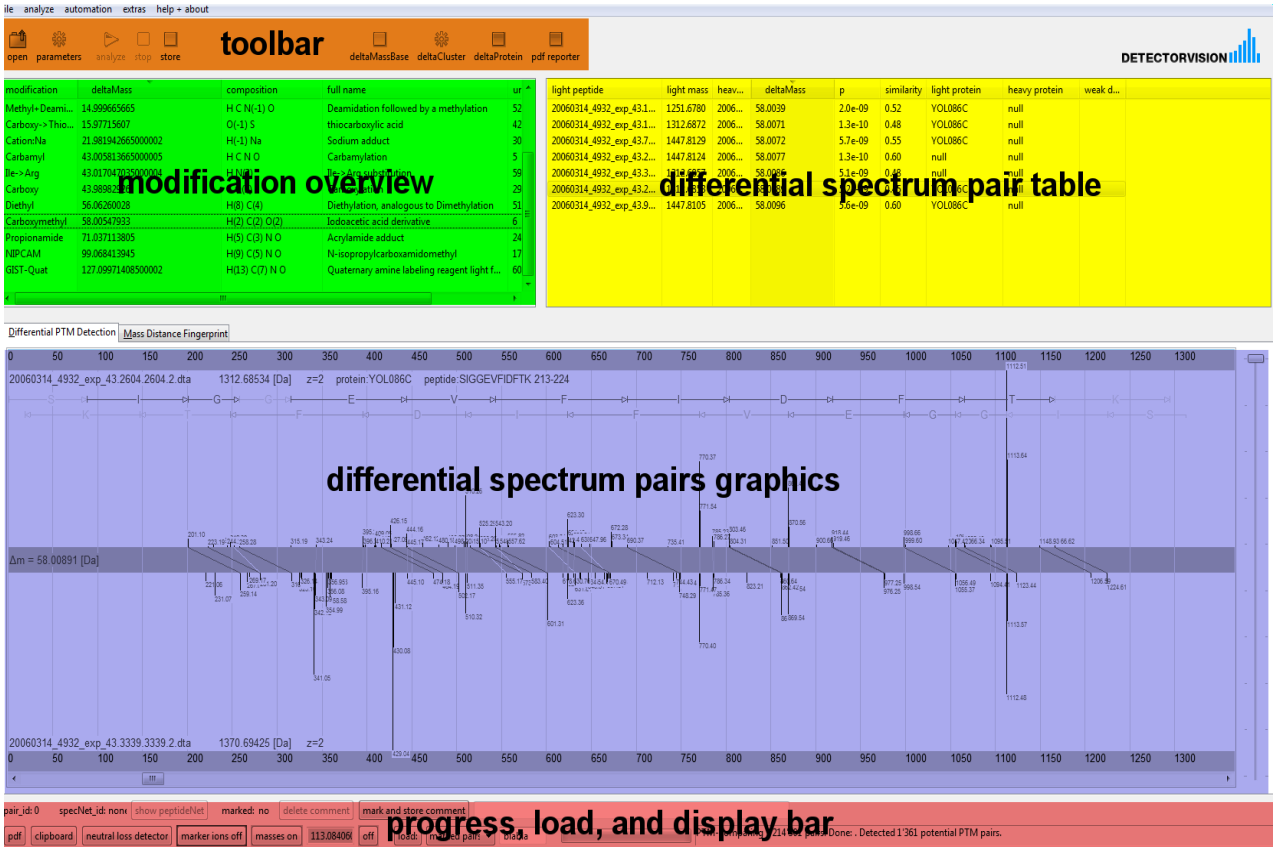


Illustration 2: Screenshot after running the c:\detectorvision\deltaMasses\data\example.mgf file. Shown is an N-terminal carboxymethylation of the peptide SIGGEVFIDFTK on the yeast protein YOL086C. Unmodified peptide: top, modified peptide below. Peptide is at position 213-224 of the protein.

4 Usage

Start->Programs->DETECTORVISION->deltaMasses



The **five main components of the interface** are:

1) ***toolbar*** to load and analyze peakfiles
2) ***modification overview*** table listing PTMs; both known and unknown.
3) ***differential spectrum pair table*** changes when you alter the selection in the modification overview. If peptide ID's have been loaded into deltaMasses, known peptides will be colored in blue in this table.
4) ***differential spectrum pairs graphics*** changes when you change your selection in the differential spectrum pair table.
5) ***Progress, load and display bar*** communicates messages from deltaMasses while loading, analyzing or printing. Used to load sets of spectra pairs and to set display characteristics.

deltaMasses is an easily operated **three stage** rocket:

1. open an .mgf peaklist / mascot.xml results file, adjust accuracy settings

2. Analyze by DPD : **D**ifferential **P**TM **D**etection
3. review DPD results

open a peaklist file

To load a peakfile (in .mgf format) click onto the *open* icon. Browse to the peakfile. Set both the MS and MSMS accuracy[1] in Dalton. Default settings (0.01/0.4 Dalton) are OK for a well-callibrated LTQ-FT/Orbitrap instruments. Given the accuracy of your massspec in ppm, you can approximate the Dalton setting using the formula

*(mass accuracy in Dalton) = (mass accuracy in ppm) * average_expected_mass / 1000000*

Example: for a 5 ppm instrument and an average expected precursor mass of about 4000 Dalton, the masss accuracy in Dalton would be 0.025. Some additional values are given in the table below. For a well-callibrated LTQ-FT[2] instrument, you would expect an MS accuracy of about 0.01[Da].

If your peakfile was allready searched with X!Tandem[viii], you can read the corresponding X!tandem xml file to read peptide identifications into deltaMasses. See separate section in this manual for instructions.

When ready, click „OK“ to load the file(s) into deltaMasses. deltaMasses supports loading of a single .mgf file. Current limitation for the peakfile are: maximum number of MSMS spectra: 30000; minimum number: 2; exact charge state of the precursor peptide must be known; positive ion mode, maximal 5000 MSMS signals/spectrum.

DPD : differential PTM detection

To perform the DPD analysis, just click onto the „Analyze“ Icon. Assuming that the file you loaded into deltaMasses contains *N* MSMS spectra, deltaMasses will now differentially detect PTM's in all *N(N-1)/2* spectrum pairs.

5 License personal edition

1 deltaMasses supports only Dalton.

2 LTQ FT and Orbitrap are trademarks of Thermo Scientific Corporation

exception are external packages as listed in the deltaMasses user manual.

This software is provided "as-is," without any expressed or implied warranty. In no event shall Detectorvision AG be held liable for any damages arising from the use of this software.

The deltaMasses package may not be redistributed. The only source is http://www.detectorvision.com/deltaMasses i.e. every single installation has to be done by download from this adress.

The licensee must register with correct name, email, city, and country. Not doing so does not void this agreement and is a legal offense against Detectorvision.

Detectorvision AG is not allowed to publish above registration data except city and country.

This license is valid for the registered person only and is not transferable.

In any kind of publication, including all public or nonpublic presentations, of deltaMasses results, the authors are legally bound to cite www.detectorvision.com/deltaMasses

It is forbidden to re-engineer deltaMasses by e.g. reverse compilation or other.

By clicking the "accept" button, or by installing, downloading, or using the software, or by opening the package, you are consenting to be bound by and are becoming a party to this agreement with Detectorvision AG. If you do not agree to all of the terms of this agreement, do not use, install or download the software.

Detectorvision AG will not grant you a license to install, download, or use the software or accept any order for the software unless you agree to all of the terms of this license agreement. Copyright laws and international copyright treaties, as well as other intellectual property laws and treaties, protect this software.

You are not allowed to sublicense deltaMasses.

This license is valid for a single installation on a single computer for personal use only.

You are receiving a non-transferable, personal, and non-exclusive license to use deltaMasses.

The software and all intellectual property rights relating to the Software, including code, operation, architecture, implementation, look and feel remain at all times the exclusive property of Detectorvision AG.

You are not receiving any right, title, or interest in the software, except to the extent of this license. You may not modify, translate, reverse engineer, decompile, disassemble, or create derivative works or emulators of the software. Decompiling or disassembling the software may violate the software's copyright.

You may not delete or alter any copyright, trademark, or other proprietary rights notices of Detectorvision AG and any licensors appearing on the software, all of which must be reproduced on all copies made of the software.

You may not distribute the software in any form to any third party, and you may not sublicense, rent, lease, sell, transfer, or grant any rights for deltaMasses in any form to any third party without the written consent of Detectorvision AG. If you violate these terms, Detectorvision AG has the right to seek injunctive or other equitable relief.

Detectorvision AG reserves the right to terminate this agreement at any time. Upon termination, your right to use deltaMasses will immediately cease. You agree that you destroy all copies of deltaMasses after notification.

Please submit feedback to info@detectorvision.com including eventual problems with installation or usage, and suggestions to improve the software. You agree that Detectorvision AG may use your submissions, whether orally or written, and you grant Detectorvision AG a perpetual, unrestricted, irrevocable, and royalty-free license to incorporate your comments, suggestions into Detectorvision AG products or services.

You understand and agree that the software is offered "as is" without any warranty whatsoever, expressed or implied or statutory, including, without limitation, any implied warranty of merchantability or fitness for a particular purpose. Detectorvision AG does not warrant that the software is error free or that errors in the software will be corrected.

You may not assign this agreement, by operation of law or otherwise, without the prior written consent of Detectorvision AG.

This Agreement shall be construed under and be governed in all respects by the laws of Switzerland without any reference to the principles of conflict of laws thereof. All disputes out of or in connection with this Agreement shall be settled exclusively by the competent courts of Zurich.

If any provision of this agreement is invalid or unenforceable, the remainder of this agreement will be interpreted so as best to reasonably effect the intent of the parties.

This agreement constitutes the entire understanding and agreement of the parties hereto with respect to its subject matter and supersedes any prior and contemporaneous agreements, purchase orders or understandings between the parties. Any term or provision of this agreement may be amended, and the observance of any term of this agreement may be waived, only by ammendments in writing signed by the party or parties to be bound thereby.

End of deltaMasses Personal Edition license text.

6 Support and Development

No support is granted under the free personal edition license. Contact info@detectorvision.com for support contracts, further development or lab automation requests.

7 FAQ

- Q: For the Differential PTM Detection (DPD), isn't it true that MSMS accuracy is more important than MS accuracy? Why is ppm accuracy in the precursor mass required ?

  A: Correct, the MSMS accuracy is the single most important factor and the quality of the DPD increases exponentially with MSMS mass accuracy.

- Q: In the spectrum pair list, I get many repeats of essentially the same peptide with the same modification on and on again. Why is that?

  A: Most likely, this is because your raw->peaklist transformation software is not doing scan merging in the time domain. So when you have the same peptide fragmented several times, you will have the „same“ spectrum repeated several times in your peaklist. Imagine you have the unmodified peptide 10 times and the modified 10 times as well. In this case, DPD would give you 100 detected pairs for one „real world“ pair. If you use deisotoping software which is capable of merging scans in the time domain probably, this problem will almost fade away. Try e.g. Mascot Distiller to get better peaklists.

- Q: Does DPD tolerate dominating neutral losses in MSMS?

  A: Yes for phosphorylations. Currently no for all other modifications. Neutral loss information is not displayed in the graphical comparison except for phosphorylations.

- Q: If there is a dual modification on a peptide, can this be handeled by deltaMasses?

  A: Not really when both modifications occur always together. Yes if they are variable. An „extreme“ example of this was a detection of a 4-fold phosphorylated peptide; the detection was possible because all 4 stages of phosphorylation had been measured in that case.

- Q: If a peptide is always modified in a certain way, can DPD detect it?

  A: No – unless you synthesize the unmodified peptide and measure it. This approach might be used for e.g. virus coat protein analysis.

- Q: So both the modified and the unmodified peptide must be MSMS-measured, otherwise DPD cannot detect the modification?

  A: Correct. DPD is a differential method. This is both the biggest advantage and disdvantage of the method. It is an advantage because DPD can be used for biomarker discovery and regulational investigations. The disadvantage is obvious. If you run DPD after a search engine, you combine the strengths of both approaches.

- Q: deltaMasses does not display all signals of my MSMS spectrum

  A: This is because the DPD may use just some information of your MSMS spectrum and only displays those peaks actually used. Many if not most classic search engines use the same approach. Also, when you calibrate a peakfile using „post measurement calibration“, deltaMasses exports only the most intense MSMS signals.

- Q: What is the sensitivity of DPD?

  A: Depends on MSMS accuracy and many other factors. With MSMS accuracy worse than 0.4 Dalton, you will loose sensitivity. deltaMasses in its current form is not detecting neutral loss prone modifications (exception: phosphorylations).

- Q: What is the „*Similarity*“ ?

  A: The *Similiarity* is between 0 and 1 and indicates the overlap between the spectra in the pair. A similarity of 1 does not necessarily imply that the two spectra are equal, but they will be very similar.

- Q: When using X!Tandem, I need two files as an input: a file somename.mgf and somename.xt.xml . This is irritating – why do you do this ?

  A: This is indeed irritating. Reason: The X!Tanem results file (.xt.xml) contains MSMS information only for the MSMS spectra having a peptide identification. So in a typical proteomics experiment, you would loose up to 90% of your MSMS spectra before doing DPD. This is why you need to have the .mgf file as well. For the Mascot case, this is easier because you can export Mascot results including ALL MSMS information.

- Q: Exporting mascot.xml files takes me 2 minutes for each file – this is a waste of time.

  A: In Discovery Edition, this export can be done automatically.

8 Statistics of Differential PTM Detection

In this section, a definition to the statistical filtering of the differential PTM detection process is given. The deltaMasses DPD statistics uses both MSMS m/z values and intensity values. Here, only the m/z value statistics is discussed.

Before using deltaMasses, the raw spectrum has to be converted into a correctly deisotoped peaklist[3]. This list is read in Mascot Generic Format (.mgf) or mascot.xml format. The file must contain two or more MSMS measurements.

If a single spectrum is compared to N other spectra, the probability of randomly hitting at least one other spectrum is N*P; P being defined as the probability for a random hit for a single comparison. P dramatically depends on the MSMS mass accuracy of the instrument approximately following an exponential relation; P being roughly „proportional to δ^{m}“ where δ denotes the mass accuracy and *m* the number of peaks matching between two spectra. If you assume that a typical deltaMasses match has about 16 matching MSMS peaks, *m=16;* an improvement of the mass accuracy δ by a factor of two would decrease the random hit probability P by more than six million percent (!). **MSMS mass accuracy is exponentialy significant.**

DeltaMasses compares all N MSMS spectra with each other, resulting in a total of N*(N-1)/2 comparisons. The probability to randomly match at least one spectrum pair is limited from above by P_{random}= P*N*(N-1)/2. For a spectrum pair to be „significant“, deltaMasses uses a threshold of P_{random}< 0.05 (5%). The P_{random}< 0.05 cutoff is to be understood on the level of the whole dataset. This is clarified by the following example:

3 A number of tools are available for this. Distiller from www.matrixscience.com is a good option..

N=5000 MSMS spectra resulting in 12,497,500 comparisons. P_{random}<5% implies

$$P_{random} = P*N*(N-1)/2 < 0.05$$

$$P < 0.05 * 2 / N*(N-1)$$

$$P < 4.001 * 10^{-9}$$

Therefore, the probability P_{random} is below 5% for the whole set of 12,497,500 comparisons. On the level of the single comparison, the probability would be below $4.001 * 10^{-9}$. In this way, P_{random} depends on the size of the dataset; comparable to the E-value of a Blast search. Statistically speaking, one would expect one false positive pair in 20 experiments of N MSMS measurements each; i.e. one false pair for 20*N*(N-1)/2 comparisons.

The above statistical cutoff is a minimum requirement; deltaMasses uses additional filters in the matching process. Above criteria are always fulfilled. Spectrum pairs with P higher than 10^{-6} are not reported.

Note: probabilities are reported in „statistical units“, i.e. P=0.05 means P=5%.

Almost all CPU time is spent on comparing MSMS spectra. The number of MSMS spectra in a peaklist is the main factor influencing runtime. The table below shows some runtime examples on a rather usual laptop[4]. A 1-hour MSMS measurement with 5000 MSMS spectra, resulting in more than 12 million comparisons, has an average analysis time of about three minutes. As can be seen from the table, the runtime also depends on other factors than just the number of MSMS spectra; the 7454 MSMS measurement runs faster than the 4512 measurement.

number MSMS spectra	comparisons	Time [m:s]	pairs detected	P
2902	4408965	1:11	1821	< 1.2e-8
4512	10850811	2:45	1767	< 4.7e-9
7454	27628461	1:05	1358	< 1.9e-9
24931	310640275	19	2695	< 1.7e-10
14842	111728826	32	7898	< 4.5e-10

Table 2: deltaMasses v 1.1 runtime examples, all with 0.01 Dalton MS accuracy and 0.4 Dalton MSMS accuracy. HP Pavilion dv4000 Laptop, 1 Gbyte RAM, Pentium 1.73 Ghz. P is the estimated upper limit of the probability for a reported pair to be a random hit (i.e. the actual probability might be even lower). Note that deltaMasses v 1.1 compares all spectra against all spectra; neither retention time nor precursor charge information is used to filter pairs.

4 HP Pavilion dv4000, 1 Gbyte RAM, Pentium 1.73 GHz

9 Automation: DeltaMasses Discovery Edition

To DPD-analyze hundreds of samples, you might want to automate deltaMasses and store the results in a structured way. Personal Edition, the free version of deltaMasses does not provide automation capabilities nor database export functionality. Both features are available in the *deltaMasses Discovery Edition*. With Discovery Edition, you can run DPD automatically on all peaklists contained within a directory. Discovery Edition connects to a relational database called deltaMassBase. Additional features are: .pdf reporting, .csv reporting.

Before installing Discovery Edition,we request users to test the personal edition for at least two weeks. Connectivity to the postgreSQL driven deltaMassBase database engine can be tested in the personal edition as well, and we encourage users to do so before upgrading to Discovery Edition.

10 Automatic pdf analysis report

Discovery Edition only

For each .mgf file analzyed a pdf analysis report is written to the file system automatically. To give an example, the input file

D:\terraStorage\G353\QTOF6510\20060905\HIV_gp120\20060905_HIV_gp120_34.mgf

will automatically yield a deltaMasses output file

D:\terraStorage\G353\QTOF6510\20060905\HIV_gp120\20060905_HIV_gp120_34.mgf.deltaMasses.pdf

This file is written if the current user has write acess to this location and no equally named file exists. This feature cannot be switched off.

Structure of the pdf report

On the top of the page, the name of the peaklist file is printed in barcode 128 format [ix] . This can be used to track experimental information in professionally organised laboratories. The Barcode 128

Illustration 3: Barcode 128 representation of a filename

system can only encode the 128 characters contained in the 7-bit ASCII characterset[x]. If your peaklist filename contains other characters, these will be converted to undersores in the barcode presentation.

满足无线数据库管理的……
速度需求。

deltaMasses™ DETECTORVISION

example.mgf

C:\detectorvision\deltaMasses\data\example.mgf
deltaMasses analysis time: Mon Feb 19 10:35:25 CET 2007 by frank
number of MSMS spectra: 3799
spectra with peptide id: 252 = 6.63%
number of charge 1 spectra: 1413 = 37.19%
number of charge 2 spectra: 1219 = 32.09%
number of charge 3 spectra: 913 = 24.03%
number of charge 4 spectra: 166 = 4.37%
number of charge 5 spectra: 54 = 1.42%
number of spectra with higher charge: 170 = 4.47%
maximal accepted random probability: 6.93e-09
detected pairs: 1336 originating from 282 different spectra
spectra with peptide ID including ID by pairing: 331 = 8.71%
spectra ID'd by pairing: 79 improvement: 31.35%
paired spectra without pair having any peptide ID: 192

potential modification signals and attempted explanations			
delta mass	composition	short name	number of pairs
-	-	unknown	48
43.00581	H C N O	Carbamyl	1122
58.00548	H(2) C(2) O(2)	Carboxymethyl	7
99.06841	H(9) C(5) N O	NIPCAM	56
-18.01056	H(-2) O(-1)	Dehydrated	19
71.03711	H(5) C(3) N O	Propionamide	1
-17.02655	H(-3) N(-1)	Gln->pyro-Glu	1
21.98194	H(-1) Na	Cation:Na	6
127.09971	H(13) C(7) N O	GIST-Quat	1
-30.01056	H(-2) C(-1) O(-1)	Pro->Pyrrolidinone	1
14.99967	H C N(-1) O	Methyl+Deamidated	17
43.98983	C O(2)	Carboxy	3
15.97716	O(-1) S	Carboxy->Thiocarboxy	1
56.06260	H(8) C(4)	Diethyl	22
-15.95853	H(4) C O(-2)	Asp->Val	11
43.01705	H N(3)	Ile->Arg	20
Listed are the nearest modifications in terms of delta mass. If you find e.g. 100 pairs, they might result from 10 unmodified spectra and 10 modified spectra from the same peptide.			

Illustration 4: example of a .pdf report

Configuring the laboratory logo on the top left of the pdf reports

To put your laboratory logo onto the top left corner of the pdf report, name it laboratory_logo.png and put it into the directory

C:\detectorvision\deltaMasses\config\laboratory_logo.png

It will be automatically scaled, with image ratios conserved, into a rectangle of size 196*44 points (width*height). 1 point = 0.352mm = 1/72 inch. An example is given below (the customer_logo being the chinese text):

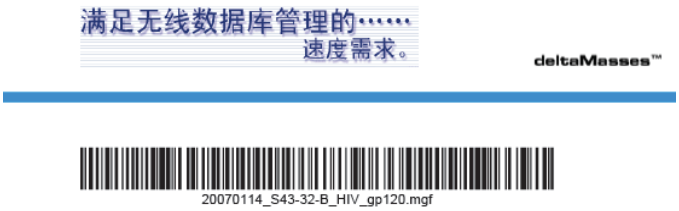


Illustration 5: laboratory logo (top left) is read from the file .../config/laboratory_logo.png

The laboratory_logo must be strored in .png format (portable network graphics).

Configuring the footer of the pdf reports

The information at the bottom of the pdf reports can be configured in the file

C:\detectorvision\deltaMasses\config\deltaMassBase.config.txt

The four footer fields can be set as shown below. Observe that there MUST be one space between the equality sign and the value of the footer field.

```
#
FOOTER_FIELD_1= Proteomics Core Laboratory
FOOTER_FIELD_2= St. Moritz
FOOTER_FIELD_3= Switzerland
FOOTER_FIELD_4= www.myswitzerland.com
#
```

Above example results in a footer as illustrated in the picture below:

11 deltaProtein

DeltaMasses stores information about related spectra in the deltaMassBase database. DeltaProtein is an analysis tool collecting these relations into so-called peptide networks. The principle is best illustrated, see below.

DeltaProtein is started from the deltaMasses interface by clicking onto the deltaProtein button, the status of this tool is still pre-release.

12 Technologies used

Technologies used for the development and implementation of deltaMasses are listed below including licensing details

- deltaMasses runs with Java® 6. Java is a trademark or registered trademark of Sun Microsystems, Inc[xi].
- deltaMasses is developped using Eclipse[xii] 3.2
- The jar files are packed using the fatjar eclipse plugin[xiii]
- To create .pdf files, deltaMasses uses the itext package[xiv] under the Mozilla Public License[xv]
- The installer is from Innosetup[xvi] version 5.1.6 by Jordan Russell, see This installer is for Windows Vista / XP / 2000 only.
- deltaMasses uses SWT (Standard Widget Toolkit), an open source toolkit for Java designed to provide efficient, portable access to the user-interface facilities of the operating systems on which it is implemented. DeltaMasses runs on platforms supporting Java 5 and SWT, the installer is for Windows XP/2000.
- The Graphical user interface (GUI) is defined using XSWT[xvii] . XSWT is licensed under the Common Public License[xviii]
- The postgresql database system is used under the BSD license[xix]. BSD is a trademark of the University of California, Berkeley.
- DeltaMasses connects to the deltaMassBase database using postgreSQL JDBC under the BSD license[xx]. Copyright (c) 1997-2005, PostgreSQL Global Development Group
- http://www.xmlpull.org/ xmlpull_1_1_3_4b.jar under free license[xxi]. XML is a trademark of MIT and a product of the World Wide Web Consortium.

- http://www.jdom.org/ jdom is licensed under Apache license[xxii].
- http://kxml.sourceforge.net/ kxml2-2.4.0.jar is licensed under BSD license[xxiii]. BSD is a trademark of the University of California, Berkeley.
- Some statistical calculations are done using the Jakarta mathematics library commons-math-1.1.jar from http://jakarta.apache.org/commons/math/ distributed under the Apache 2.0 license[xxiv]. Apache is a trademark of The Apache Software Foundation
- The database of the Discovery Edition ist postgresql distributed under the BSD license. The BSD license can be found below.
- deltaMasses was developped on Linux and Windows operating systems. Linux is a trademark registered to Linus Torvalds. Windows is a trademark of Microsoft[xxv] Corporation.
- Documents and flowcharts authored with OpenOffice.org version 2.0
- Graphics with open source GIMP version 2.2.11
- base64 encoding is done with http://iharder.net/base64

13 Quality Control

This section lists some quality control methods for Differential PTM Detection. If you have suggestions for additional quality tests, please let us know.

Does DPD detect expected PTM's ?

Are the mass differences within your mass spectrometer's accuracy?

Random spectra control: DPD should detect close to no pairs when you throw random spectra at it.
Since version 1.1.0.2, you can do this directly from deltaMasses:
after OPEN, choose .random.params for the file selector (default: .mgf)
deltaMasses will now generate and analyze the random spectra.
you can influence the random spectra control by changing the paramters in your *.random.params file, a sample is shown below. You can find it at
C:/detectorvision/deltaMasses/data/sample.random.params

```
MIN_AMINOS=10
MAX_AMINOS=25
MS_ACCURACY_DALTON=0.01
MSMS_ACCURACY_DALTON=0.01
NUM_SPECTRA=299
MIN_MSMS_MASS=400
MAX_MSMS_MASS=4000
CALIBRATION_A=0.01
CALIBRATION_B=0.0001
```

To use the random test feature, you need to feed deltaMasses with a *random.params file; you can easily create one yourself from above example.
Currently, the precursor masses correspond to randomly generated peptides, while the MSMS signals are randomly distributed.

Parameters:

```
MIN_AMINOS                      minimal number of amino acids in a peptide
MAX_AMINOS                      maximal ....
MS_ACCURACY_DALTON              MS accuracy in Dalton (Std. Deviation)
MSMS_ACCURACY_DALTON            MSMS....
NUM_SPECTRA                     number of spectra to be generated
MIN_MSMS_MASS=400               minimal MSMS signal
MAX_MSMS_MASS=4000              maximal ...
CALIBRATION_A=0.01              calibration model of the data is    a+b*mass
CALIBRATION_B=0.0001
PEPTIDE=LKEVMDSLK               generate MSMS of peptide LKEVMDSLK
PEPTIDE=LKEVMDSLK=5=15.994915   as above, put a modification on the M
```

This framework makes it easy to produce synthetic .mgf's. An example is a simulation of SMC4_HUMAN protein, as illustrated by the files:
C:/detectorvision/deltaMasses/data/SMC4_HUMAN.random.params
C:/detectorvision/deltaMasses/data/SMC4_HUMAN.random.mgf
C:/detectorvision/deltaMasses/data/SMC4_HUMAN.random.params.mascot.xml

X!Tandem and Mascot results:
http://human.thegpm.org/gpm/archive/GPM32100016712.xml
http://www.matrixscience.com/cgi/master_results.pl?file=../data/20070112/FLEcieHw.dat

If you run deltaMasses in this mode, it will automatically generate an mgf file which you can use on your search engine:
input:
C:/detectorvision/deltaMasses/data/sample.random.params
output:
C:/detectorvision/deltaMasses/data/sample.random.params.mgf

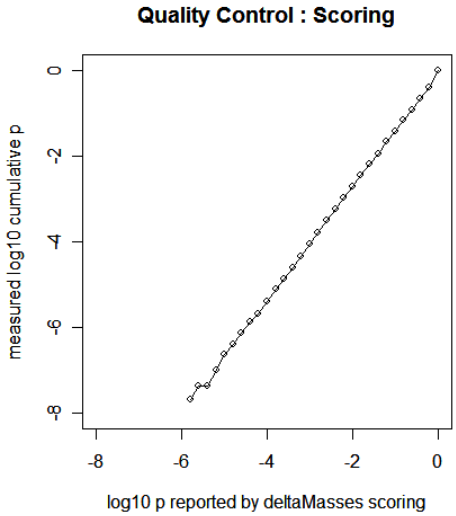
WARNING: allready existing *.random.params.mgf files will be overwritten

Quality control example

A file with random signals only
www.detectorvision.com/deltaMasses/exampledata/random_10000_75_01_4.zip
was used as a **random test for deltaMasses**:

At www.detectorvision.com/deltaMasses/exampledata/random_10000_75_01_4.dat is the result from comparing the 10.000 MSMS spectra i.e. 49,995,000 pairs. DeltaMasses cutoff probability for 49,995,000 pairs is p = 0.05 / 49,995,000 i.e. 2e-8 which is two orders of magnitude below the minimum seen random probability of e-6. These results illustrate that the deltaMasses probability is

conservative. Below, we plot the measured cumulated log10 probability against the log10 probability reported by deltaMasses. As can be seen from the figure, p(reported)>p(measured) =p(true). Furthermore, the relation is linear indicating a true probabilistic scoring.



14 Acknowledgements

The concept of Differential PTM Detection (DPD) originates from research at the Functional Genomics Center Zurich[xxvi] during 2003-2005 as described in 42. First proof-of concept prototype in Perl dates back to May 2003 from Seattle, USA. The C-implementation was done by Jiri Ocenasek and Martin Pelikan at the Computational Laboratory of ETH Zurich in 2004. A first Java prototype of deltaMasses (v 0.5) was implemented by Detectorvision AG during January-July 2006. deltaMasses v 1.0 / 1.1 was implemented by Raphael Bosshard, Michael Lehmann, and Frank Potthast in cooperation with Zürcher Hochschule Winterthur[xxvii] under supervision of Andreas Meier and Frank Potthast. The statistical scoring has been developped by Detectorvision AG. Jari Häkkinen established the synchronized version infrastructure, the trac system, and the major reorganisation of the program code in 2008.

15 Known Issues

No guarantee that this list is complete.

1. Precursor mass accuracy must be better than MSMS accuracy (at least a factor of

two is recommended)

2. Maximal number of detected pairs limited to 250000
3. Maximal number of MSMS spectra 30000
4. Maximal number of signals within one MSMS spectrum is 5000
5. Supports positive ion mode only
6. Sometimes, polymers get erraneously detected as modified peptides
7. No installer for Linux / Macintosh
8. Peptide ID integration only with Mascot and X!Tandem
9. The DPD results are annotated with known PTMs as listed in www.unimod.org. In case there are several modifications within mass accuracy, deltaMasses has no way to determine which it would be. More exactly, it can only detect a potential mass difference and thereby a potential elemental composition difference. There is, by construction, no way to circumvent this issue except MS3 measurements which deltaMasses does not adress.
10. If you have a 4700 TofTof and use the Peaks 2 Mascot tool, you will typically get too few MSMS signals to run DPD succesfully.

Please refer to www.detectorvision.com/deltaMasses.html for updates of this list.

16 Version History

Detailed version history starts with version 1.1.0.0 released on 20061012.

20030509	0.1		Perl prototype developped by Frank Potthast in Seattle, USA on behalf of Functional Genomics Center Zurich, ETH/Uni Zurich
20040604	0.5		„Icatcher“, a C command-line tool for the detection of ICAT labelled peptides. Code written by Jiri Ocenasek and Martin Pelikan of Computational Laboratory, ETH Zurich.
20050310			Publication of Icatcher in Journal of Chromatography B
20060529	1.0	#1	Java prototype by Detectorvision AG in Carcassonne, France including the DPD statistical framework
20061012	1.1.0.0	#57	New structure in cooperation with Zürcher Hochschule Winterthur by Michael Lehmann, Raphael Bosshard, and Frank Potthast
20061015	1.1.0.1	#58	Added newest unimod.xml file to install package. Transparent installer icon. Transparent desktop and shortcut icons. Fixed Q-Tof .mgf PEPMASS issue (intensity after mass).
20061017	1.1.0.2	#59	Random spectrum feature added, see section „quality testing“. Can now process .mgf's made with mzxml2other. Added 10e3 separator for big numbers in statusbar.

20030509	0.1	Perl prototype developped by Frank Potthast in Seattle, USA on behalf of Functional Genomics Center Zurich, ETH/Uni Zurich
20061022	1.1.0.3 #60	X!Tandem .xml integration finished. First version of automatic pdf report implemented. Can now read .mgf's produced with Mascot Distiller.
20061027	1.1.0.4 #61	Trivial b- and y-ion series labelling for the light spectrum (if the peptide ID has been loaded from X!tandem). Fixed deltaMass scaletick problem.
20061101	1.1.0.5 #62	Fix of a fileread bug for .mgf files. (in this case, no results reported). Occured on: Orbitrap.
20061112	1.1.0.6 #63	Mascot integration. XML export (Discovery Edition only)
20070112	1.2 #64	Post Measurement Calibration, XML export for all editions.
20070128	2.0 #68	Speed increased by about 500%-1000% . Java JRE delivered together with deltaMasses for easiest installation procedure. Prototype of ICPL-Quant. ICPL is a trademark of Toplab GmbH Germany, www.toplab.de. Focussed comparisons (Discovery Edition).
20070228	2.1 #74	Installation now in two parts: Infrastructure + deltaMasses. PostgreSQL database support.
20070330	2.2 #81	DeltaMassBase report, neutral loss detector, deltaMassBase PTM fingerprint, copy to clipboard
20080127	3.0 #209	Peptide Net functionality integrated (also referred to as deltaProtein)
20080815	4.0 #307	Reload of records, pairs, marked pairs. Annotation of pairs. Complete restructuring of the software (transparent to the user). About 40 minor improvements and crashfixes
20081231	4.2 #440	Introduction of the weak delta detection (glycoanalysis). Parameter settings included. Deletion of records. 28 minor improvements and crashfixes.
20091231	4.5 #565	Diverse fixes and enhancements.
20101231	4.6 #621	Smaller fixes.
20111010	5.0 #700	Windows 64 Multicore Edition, upgrade to Windows 7 / 8

17 X!Tandem integration

deltaMasses supports X!Tandem. Instructions:

Assume you have you .mgf file at

`D:\terraStorage\G353\QTOF6510\20060905\HIV_gp120\20060905_HIV_gp120_34.mgf`

Copy the X!Tandem .xml results file (NOT PRIDE FORMAT) to

`D:\terraStorage\G353\QTOF6510\20060905\HIV_gp120\20060905_HIV_gp120_34.xt.xml`

Blue marking above for the important issues:

The X!Tandem .xml file must

1. be located in the same directory as the .mgf file
2. must have the same name as the .mgf file with file ending .xt.xml

deltaMasses supports X!Tandem results files version 2.0. Below is a self-explaining screenshot of deltaMasses used in conjunction with X!Tandem.

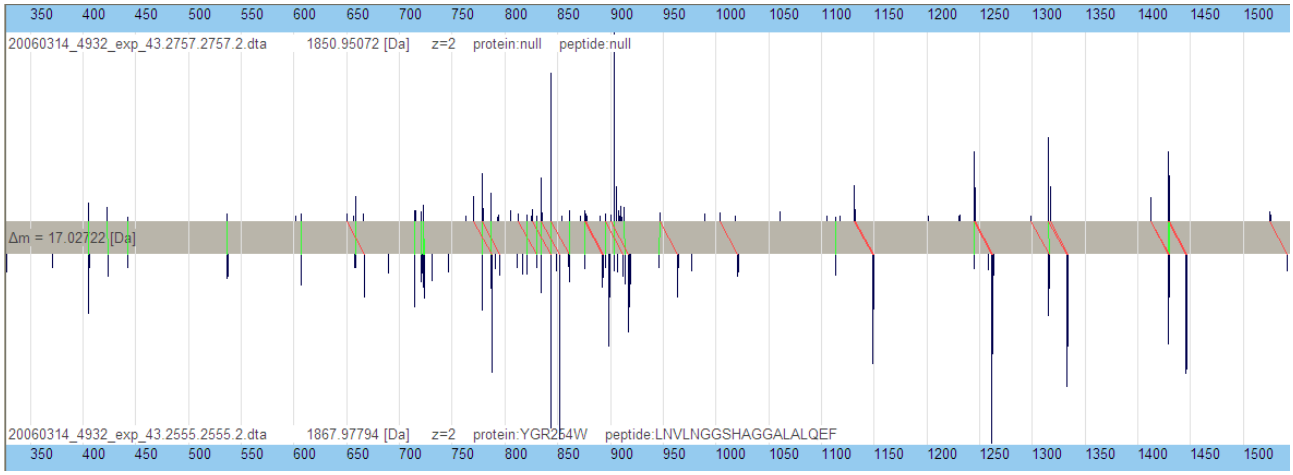


Illustration 6: X!Tandem integration example. Shown is a spectrum pair at deltaMass 17.02722 [Da]. The heavy peptide has been identified by X!Tandem to be YGR254W, i.e. Enolase 1. It is immediately clear from this plot that the modification lives on either the N- or the C-terminus of the peptide with a likely elemental composition difference of H(-3)N(-1).

18 Neutral loss detection

Neutral loss detection was introduced in v 1.1.0.3 for phosphorylations. Since v 1.1.0.4 this information is also displayed as shown below.

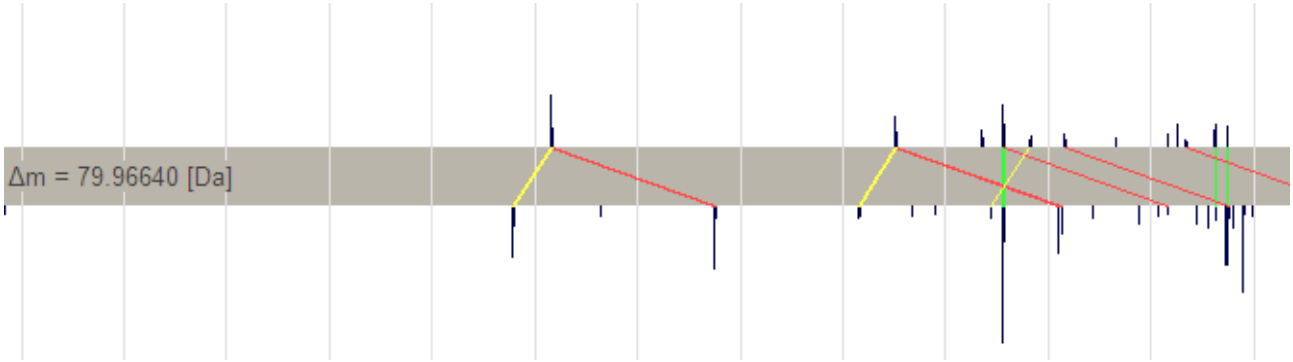


Illustration 7: Neutral loss detection. Upper spectrum is the lighter one. The phosphorylation triangle is a strong indicator for a phosphorylation. The yellow line denotes the neutral loss of H3PO4, 97.976896 Dalton, after adding HPO3, 79.966331 Dalton (red line).

19 Mascot Integration

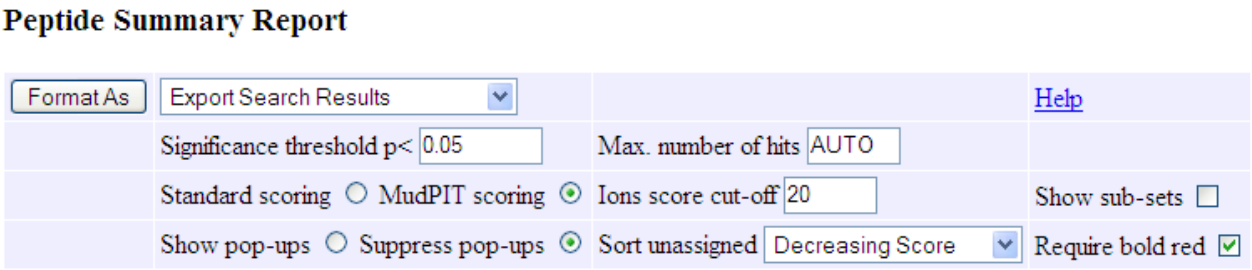
deltaMasses can read output from the industry standard search engine Mascot (see www.matrixscience.com). Mascot is a trademark of Matrixscience Ltd. London.

Flight instructions (follow carefully):

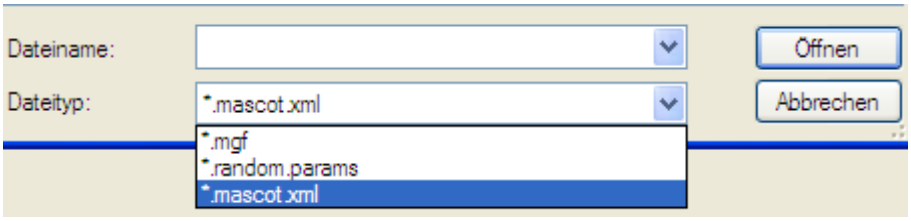
In Discovery Edition, you can automatically perform the following steps; instructions below are for Personal Edition.

After having done the following five times you'll do it in 20 seconds.

1. On the *Peptide Summary Report* page, check the „**Require bold red**“ checkbox
2. On the *Peptide Summary Report* page, set **Ions score cut-off** to 0.1
3. On the *Peptide Summary Report* page, set „*Format as*“ to „***Export Search Results***“
4. Hit The „***Format As***“ Button



5. On the next page, check all checkboxes in the two sections „*Optional Search Information*“ and „*Optional Peptide Match Information*“. Please dont change anything else unless you know exactly what you do. **In Mascot 2.2,** check all checkboxes in the three sections „Search Information“, „Peptide Match Information“ and „Query Level Information“; leave the rest as-is.
6. Click the „**Export search results**“ on the bottom of the page.
7. **Save the resulting file with a file name ending in .mascot.xml** . As an example, you might store the result for HIV_gp120_exp_43.mgf as HIV_gp_120_exp_43.mgf.mascot.xml .
8. Start deltaMasses, open, set the file selector to *.mascot.xml and load the file



example from a Swiss version of Windows XP

9. Done

Going this way, you achieve two things at once: 1) loading the peakfile into deltaMasses 2) loading peptide identifications from Mascot into deltaMasses. The .mascot.xml export is automated in Discovery Edition.

20 Post Measurement Calibration

Post Measurement Calibration (PMC) is provided only in conjunction with mascot.xml input files. Below is an example. Exporting calibrated data in .mgf format is available in the Discovery Edition only. For each MSMS spectrum, the 75 most intense MSMS peaks get exported. The calibration effects MS masses only. At least ten peptides need to have been significantly identified; otherwise, post measurement calibration is not performed. Analysis of the calibration is available in both Personal and Discovery Edition.

Exporting calibrated .mgf: (Discovery Edition only)

extras->Post Measurement Calibration->export calibrated .mgf

beeps upon completion. The exported file is written to the same directory as the original datafile, with the file ending .cal.mgf. Warning: Existing .cal.mgf files are overwritten without warning.

Example:

original datafile: C:\somewhere\20070211_p443_i44.mascot.xml

calirated mgf goes to: C:\somewhere\20070211_p443_i44.mascot.xml.cal.mgf

Copying calibration data to the clipboard: (Discovery Edition only)

extras->Post Measurement Calibration->data to clipboard

beeps upon completion

Export calibration graphics to the clipboard: (Discovery edition only)

extras->Post Measurement Calibration->calibration graphics to clipboard

beeps upon completion

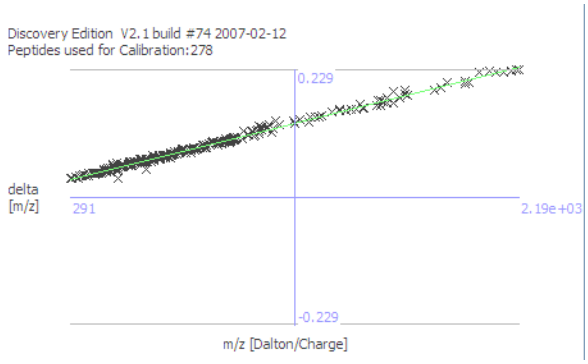


Illustration 8: before calibration

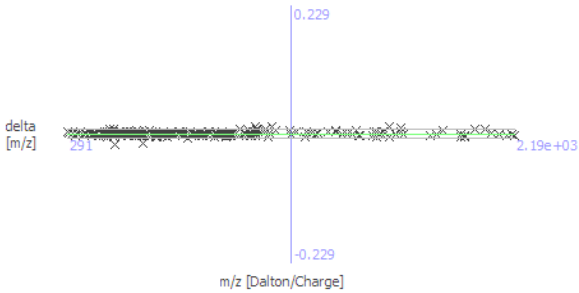


Illustration 9: after calibration

21 Directory Structure

On Windows, deltaMasses installs to

C:\detectorvision\deltaMasses

No changes in this directory and subdirectories, please.

Starting from C:\detectorvision\deltaMasses, the structure is as follows:

bin\deltaMasses.jar	Executable. Cannot be moved!
deltaProtein	All files belonging to deltaProtein
deltaProtein\dist\lib	Executable files and packages needed for deltaProtein
bin\swt-win32-3139.dll	SWT dll. Exact number (3139) may vary
bin\lib*.jar	Program packages to which deltaMasses links.
config\	Configuration files preferences.xml, preferences.xsd
data\	Some example data for first tests
documentation\	Manual
images\	...
license\	License files
license\license.txt	License text
license\registration.txt	The actual registration file
log\	Log files
automation\	Automation configuration and log files
log\automation.log.txt	automation log file
automation\mascot\data\	the mascot.xml files are stored in this directory. No subdirectories.
automation\mascot\mascot.server.txt	Holds the http://www.yourmascotserver.org mascot http address pointing to your mascot server. Can be changed with the menu automation->set Mascot server
automation\mgf\data	.mgf's to be processed by the automation robot. No subdirecctories.
automation\cluster	import/export directory for deltaMassCluster
unins000.exe	Uninstaller executable.
unins000.dat	Data needed for de-installation.
config\deltaMassBase.config.txt	database connectivity parameters Discovery Edition only
tmp\	Temporary files (for example deltaMassBaseReport.pdf). **This directory might be cleaned by the application once in a while without warning.**

Automation: Mascot (Discovery Edition only)

Put the searches.log file[5] (containing only those lines you want to process) into the directory

`C:\detectorvision\deltaMasses\automation\mascot\searches.log`

The automation robot will look for all files having the format

`"..\data\[0-9]{8}\[A-Za-z0-9]+.dat"`

i.e. they begin with „`..\data\`“ followed by 8 digits, followed by any number of alphabetic characters/digits and ending with „`.dat`“

In case you do not want to process all .dat files living on your mascot server, you have to delete the non-wanted lines from this list [6]. Only searches of type „MIS“ can be processed.

Instead of the searches.log file, you can also put a file containing a list of „.dat“ files. The „.dat“ files in this list have to be separated by any amount of whitespace. The following example illustrates this. You have to name this file „searches.log“ as well. Note the direction of slashes „/“.

```
../data/20070104/F000324.dat
../data/20070104/F000324.dat
../data/20070104/F000325.dat
```

For each file in the searches.log file, the deltaMasses automation does the following:

1. fetch the corresponding data in .mascot.xml format to the directory C:\detectorvision\deltaMasses\automation\mascot\data
2. processes it and stores the result in the deltaBase database
3. deletes the mascot.xml file fetched in step 1 above.
4. Writes a short report to the log file log/MascotAutomation.log.txt

`In deltaMasses version 2.2, step 2 is performed with the settings MS accuracy=0.01 Dalton and MSMS accuracy =0.4 Dalton. To be improved in version 2.3.`

PostgreSQL License / Copyright (valid for deltaMasses Discovery Edition)

[more at: www.postgresql.org]

5 The searches.log file can be found in the config directory of your Mascot Server.
6 Please never change the searches.log file on the Mascot server.

deltaMassBase

deltaMasses is best used in conjunction with a professional database engine. DeltaMasses supports postgreSQL which is an open source database. Please note that future releases of deltaMasses are likely to effect the database structure.

Setting up deltaMassBase

Detailed instructions on how to set up the postgreSQL database are found in the .zip package delivered to you. Typical time do get this done: about 20 minutes.

postgreSQL database facts:

Maximum Database Size	Unlimited
Maximum Table Size	32000 Gigabyte
Maximum Row Size	16000 Gigabyte
Maximum Field Size	1 GB
Maximum Rows per Table	Unlimited
Maximum Columns per Table	250 - 1600 depending on column types
Maximum Indexes per Table	Unlimited
License	Free BSD – it's free for all purposes !
Transactions	Yes

The deltaMassBase configuration file is at

C:\detectorvision\deltaMasses\config\deltaMassBase.config.txt

which you have to configure to the details of your postgreSQL installation. There must be whitespace between variable names and variables. No variable might contain any whitespace. (i.e. The password „good evening“ would fail)

DELTAMASSBASE_NAME=	deltaMassBase
DELTAMASSBASE_USER=	postgres
DELTAMASSBASE_PASSWORD=	4.3.jjMM
DELTAMASSBASE_PORT=	5432
DELTAMASSBASE_HOST=	localhost

deltaMassBase Database structure

deltaMassBase consists of the ten tables *deltacomp, deltamass, elements, multimod, peptide, record, spectrum, specnet, meta_db, experiment.*

For each table[7] <*tableName*> , there is a row called <*tableName*>_id ; being the unique primary key to the table; the primary key is always an integer. If a table <*tableName*> refers to a primary key of <*anotherTableName*>, the referring column is called fk_<*anotherTableName*>_id.

The most important relation is between the 6 tables *experiment,record, spectrum, deltamass, specnet,* and *peptide:*

each detected pair is stored as one row in the table *deltamass* with two references (fk_l_spectrum_id, fk_h_spectrum_id) to the *spectrum* table. Each spectrum has a unique reference (fk_record_id) to the table *record*. A spectrum might have a peptide associated (table peptide). Each *record* belongs to one and only one *experiment.*

table record

record_id	int	unique primary key
filename	string	name of the original file
pepfilename	string	name of file holding peptide ID's if applicable
username	string	windows username of operator
deltadate	timestamp	indicates when deltaMass was run
url	string	http refererence to data if applicable
msprecision	double	
msmsprecision	double	
hasretention	boolean	true if retention information known on this record
originmethod	string	mgf/mascot/xtandem
fk_experiment_id	int	reference to an experiment. Is set to 0 by default.

7 Exception: table meta_db

table spectrum

spectrum_id	int	unique primary key
fk_record_id	int	reference to the record
fk_specnet_id	int	reference to the spectrum net. If NULL->no reference
precursormass	double	
charge	int	
title	string	title as reported in an .mgf file
retention	double	(often not present)– the retention time
queryid	int	ID as e.g. Used in a mascot search result
mzbase64	text	base64 encoded ArrayList<Double>
signalbase64	text	base64 encoded ArrayList<Double>

table deltamass

Holds information about detected pairs and references to the involved spectra.

deltamass_id	int	unique primary key
dm	double	mass difference between the two precursors (redundant)
sim	double	similarity between the two spectra, within [0,1]
p	double	probability that this is a random pair [0,1]
fk_l_spectrum_id	int	reference to the light spectrum
fk_h_spectrum_id	int	reference to the heavy spectrum
mass_light	double	mass of the light peptide in Dalton (redundant)
lfk_specnet_id	int	lazy foreign key to the specnet.

table peptide

peptide_id	int	unique primary key
pepmass	double	theoretical mass of the peptide
proteinasc	string	accession of the respective protein
pepsequence	string	peptide sequence
pepmz	double	m/z value (an experimental value)
peperror	double	error of the measured mass

pepstart	int	where this peptide starts within the protein
pepend	int	as above
fk_spectrum_id	int	reference to the spectrum

table specnet

Holds information about „spectrum nets“, presenting a set of connected pairs.

specnet_id	int	unique primary key
numspecs	int	number of specs belonging to this specNet (number of nodes of the net)
numpairs	int	number of pairs belonging to this specNet (number of edges of the net)
minmass	double	minimal mass in the net
maxmass	double	maximum mass in the net
numphospho	int	number of pairs having a deltaMass of 79.96...+/- 0.01 Dalton
fk_experiment_id	int	to be removed, i.e. not to be used anymore

table experiment

experiment_id	int	unique primary key.
name	string	name of this experiment.

In case you have several records belonging to one experiment, these can be grouped into experiments using the experiment table. By default, each record loaded has the experiment_id set to 0 (default experiment). To change the record-experiment relation, the user has to change the fk_experiment_id in the record table. The experiment table is currently only used in deltaProtein.

Appart from above tables, there are three additional tables describing the influence of modifications:

table multimod is derived from www.unimod.org

table deltacomp lists possible differential elemental compositions.

table elements lists masses of elements being relevant to biomolecules.

The table meta_db holds information about the database itself:

table meta_db

db_Schema_version	int	version of the database schema.
createDate	timestamp	when this database was created
lastModificationDate	timestamp	last change of DB relevant to the specnet table
lastSpecnetDate	timestamp	last update of the specnet table

Purpose: If a service is using the specnet table, it is important that the last update of that table happened after the last change of the database. In case this is not the case, the service has to make sure that the specnet table is updated first. (The specnet table is „redundant“ because it is completely derived from other data in the database).

Table spectrum: How the columns mzbase64 and signalbase64 are encoded

These two columns are of type text and hold the base64 encoded Java datatype Arraylist<Double>. Technically, the base64 package from http://iharder.net/base64 is used. For those interested, here is an example program of how to use the base64 encoding/decoding to read the mzbase64 and signalbase64 columns.

```
import java.util.ArrayList;
public class testEncodeDecode {
	public static void main(String[] args) {
		ArrayList<Double> list = new ArrayList<Double>();
		for(int i=0;i<10;i++){//produce some nonsense data
			list.add(((double)i) * 1.0/3.0);
		}
		//serialize Arraylist<Double >list to String encodedString
		String encodedString   = Base64.encodeObject(list,Base64.DONT_BREAK_LINES);
		System.out.println("encodedString:"+encodedString.length()+" "+encodedString);

		ArrayList<Double> decodedList = new ArrayList<Double>();
		//decode encodedString to object decodedList
		decodedList = (ArrayList<Double>)Base64.decodeToObject(encodedString);
		for(int i=0;i<decodedList.size();i++){
			System.out.println("i:"+i+"  value:"+decodedList.get(i));
		}
	}
}
```

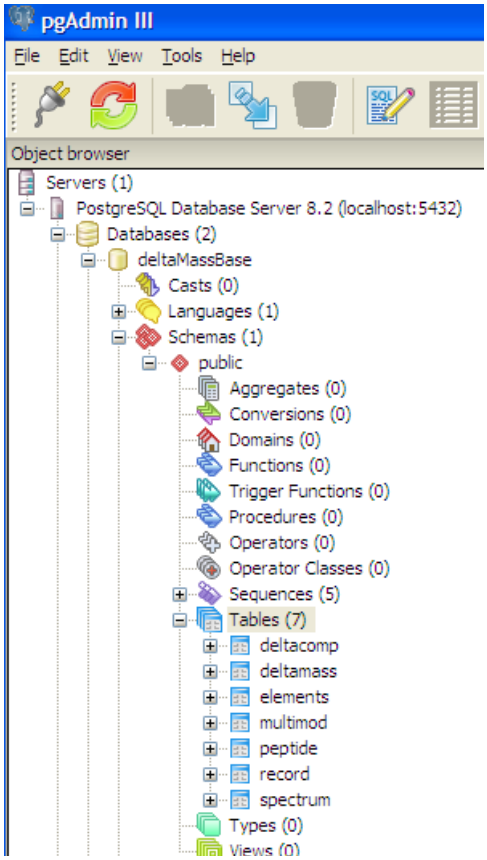


Illustration 10: browsing the database definition within pgAdmin III (part of postgreSQL)

deltaMassCluster

Status: fully functioning prototype („use on your own risk“)

You need: deltaMasses Discovery Edition with the deltaMassBase database installed.

Operators: To use deltaMassCluster, at least minor experience with command-line computing is necessary.

License: Shipped with Discovery Edition

Motivation: deltaMasses compares MSMS spectra within „records“, i.e. Within „runs“ only. DeltamassCluster enables scientists to do cross-record PTM detection. This step is computationally intensive and suited for high-performance computing.

Support Currently no support.

Remark: You do not need to have a „supercomputer“; if you want, you can run deltaMassCluster on your laptop.

OS: Windows, Linux, Unix, Apple, mobile phones (everything which can run Java 6)

deltaMassCluster is an extension to deltaMasses for high-throughput applications. A typical workflow consists of four steps:

step 0: remove ALL files in the directory C:\detectorvision\deltaMasses\automation\cluster

step1: **export** data from the deltaMassBase database to the file system

start->programs->DETECTORVISION->deltaMasses->cluster->export

step2: run the deltaMasses algorithm on the exported data („**cluster**“). You need extra software to do this. Contact us by email.

step3: **import** results back to deltaMassBase

Currently, the export exports ALL spectra of deltaMassBase and the cluster compares ALL spectra against ALL spectra.

deltaMassCluster Example

109 mgf files with 157352 MSMS spectra in total. Processed by deltamasses using automation in 48 minutes with 8034 pairs detected. With deltaMassCluster, the number of pairs goes up to 46848 pairs. Computing time on a dual core CPU was 14 hours.

Thus, the number of detected PTM pairs increased by almost 500 percent.

Installing the deltaMassBase database

You might want to contact your system administrator before proceeding.

Download Win-x86-64 version 9.1.1-1 from www.postgresql.org

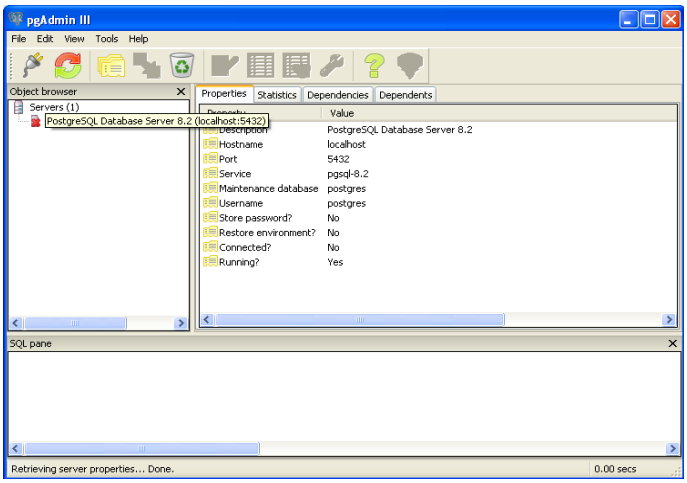
IMPORTANT: Make sure to set the encoding to **UTF-8**
Standard password for the postgres user: **4.3.jjMM**

Advanced setup (if you want to spend more time installing ...)
You may change the Port number, superuser name, password, this can be adjusted in detaMasses. Leaving it like above is the “simple road”. If you choose the more advanced road, you can configure deltaMasses in the file C:\detectorvision\deltaMasses\config\deltaMassBase.config.txt: just change
DELTAMASSBASE_NAME= deltaMassBase
DELTAMASSBASE_USER= postgres
DELTAMASSBASE_PASSWORD= 4.3.jjMM
DELTAMASSBASE_PORT= 5432
DELTAMASSBASE_HOST= localhost
DELTAMASSBASE_NAME is the name of the database which we will create below. You might change this name as well, if you want to. By default, *DELTAMASSBASE_NAME* is *deltaMassBase*
(end of advanced setup)

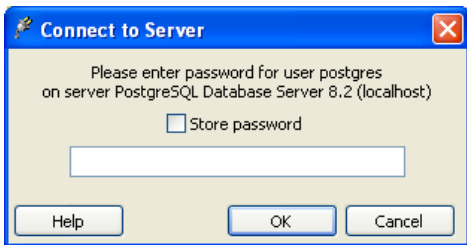
-> Finished with installing postgreSQL.

Now, to create the database “deltaMassBase”, start the postgreSQL admin tool, pgAdmin III:

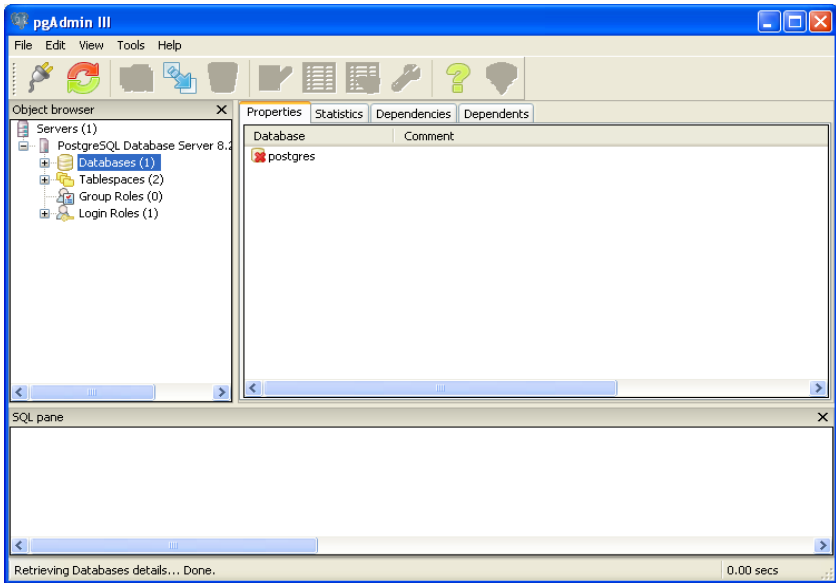
Start->Programs->postgreSQL 8.2->pgAdmin III



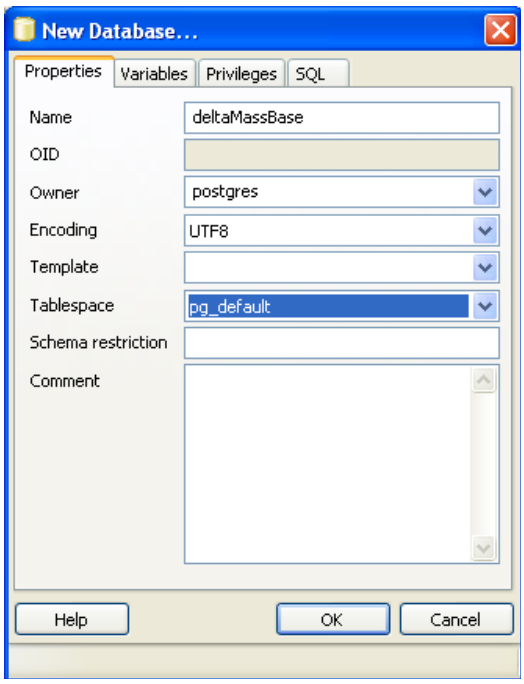
double-click onto PostgreSQL Database Server



enter the password, 4.3.jjMM (if you took the simple road). You might want to check the “store password” option.

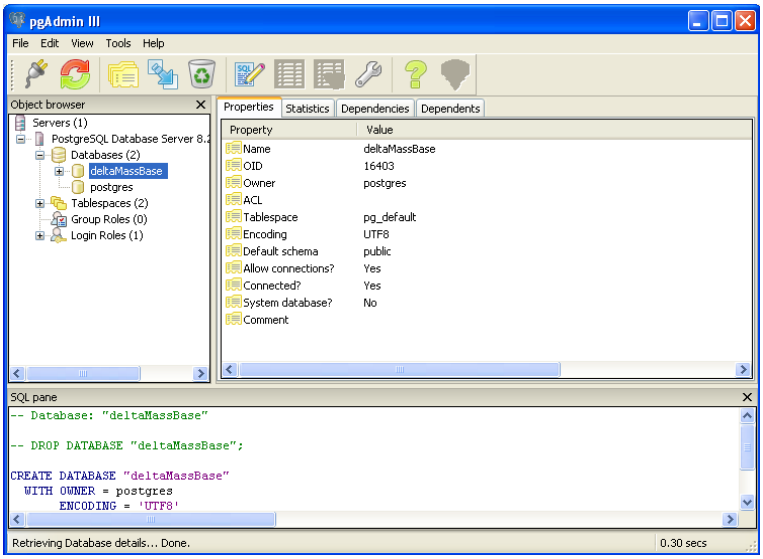


Right-click onto “Databases(1)” and choose “New Database…” Make sure to choose UTF-8 as database encoding.



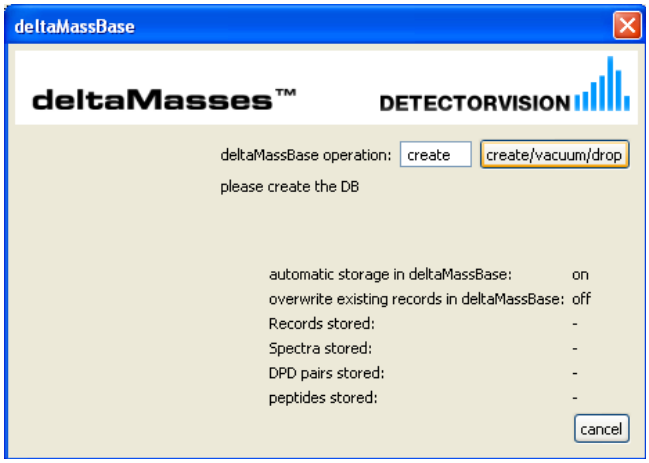
Make sure your settings are as above, and click OK.

You can now see the deltaMassBase at



Now, when you first run deltaMasses,

Click onto “deltaMassBase” to create (the schema of) the database:



This is also the place where you can drop the database (i.e. delete all data). The vacuum option should be run once in a while – it cleans the database; vacuuming is NOT removing data. You do not have to run “create” after having done a “drop”. If you choose “create” for a populated database, this will delete your database (!).

Important note on the database: we reserve the right to change the database schema with each release.

i Papayannopoulos, I.A. (1995) The interpretation of collision-induced dissociation tandem mass spectra of peptides. *Mass Spectrom. Rev.*, **14**, 49–73
ii http://www.thermo.com/ltqft
iii Qizhi Hu. et. Al. J Mass Spectrom. 2005 Apr ;40:430-43. The Orbitrap: a new mass spectrometer.
iv Parts per Million Mass Accuracy on an Orbitrap Mass Spectrometer via Lock Mass Injection into a C-trap. Matthias Mann et. Al. Molecular & Cellular Proteomics 4:2010-2021, 2005.
v SEQUEST is a registered trademark of the University of Washington.
vi Database independent detection of isotopically labelled MS/MS spectrum peptide pairs. Potthast F, Ocenasek J, Rutishauser D, Pelikan M, Schlapbach R. J Chromatogr B. 2005 Mar 25;817(2):225-30.
vii ModifiComb, a new proteomic tool for mapping substoichiometric post-translational modifications, finding novel types of modifications, and fingerprinting complex protein mixtures. Savitski MM, Nielsen ML, Zubarev RA. Mol Cell Proteomics. 2006 May;5(5):935-48.
viii Craig, R. and Beavis, R.C. (2004) TANDEM: matching proteins with tandem mass spectra. *Bioinformatics*, **20**, 1466–1467 http://www.thegpm.org
ix http://en.wikipedia.org/wiki/Barcode
x http://en.wikipedia.org/wiki/ASCII
xi http://www.sun.com/
xii http://www.eclipse.org
xiii http://fjep.sourceforge.net/
xiv http://www.lowagie.com/iText/
xv http://www.lowagie.com/iText/MPL-1.1.txt.
xvi http://www.jrsoftware.org/
xvii http://xswt.sourceforge.net/
xviii http://en.wikipedia.org/wiki/Common_Public_License
http://www-128.ibm.com/developerworks/library/os-cpl.html
xix http://www.postgresql.org/about/licence
xx http://jdbc.postgresql.org/license.html
xxi http://www.xmlpull.org/v1/download/unpacked/LICENSE.txt
xxii http://www.jdom.org/docs/faq.html#a0030
xxiii http://en.wikipedia.org/wiki/BSD_license
xxiv http://jakarta.apache.org/commons/license.html
xxv http://www.microsoft.com
xxvi http://www.fgcz.ethz.ch
xxvii http://www.init.zhwin.ch