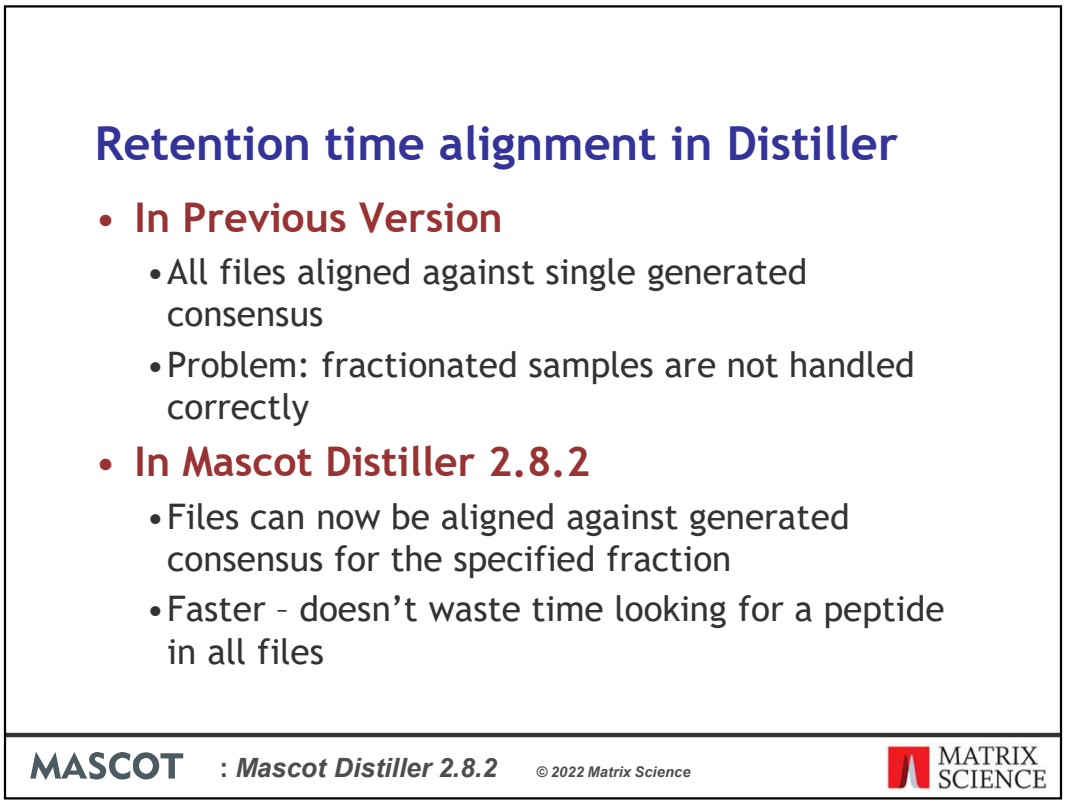
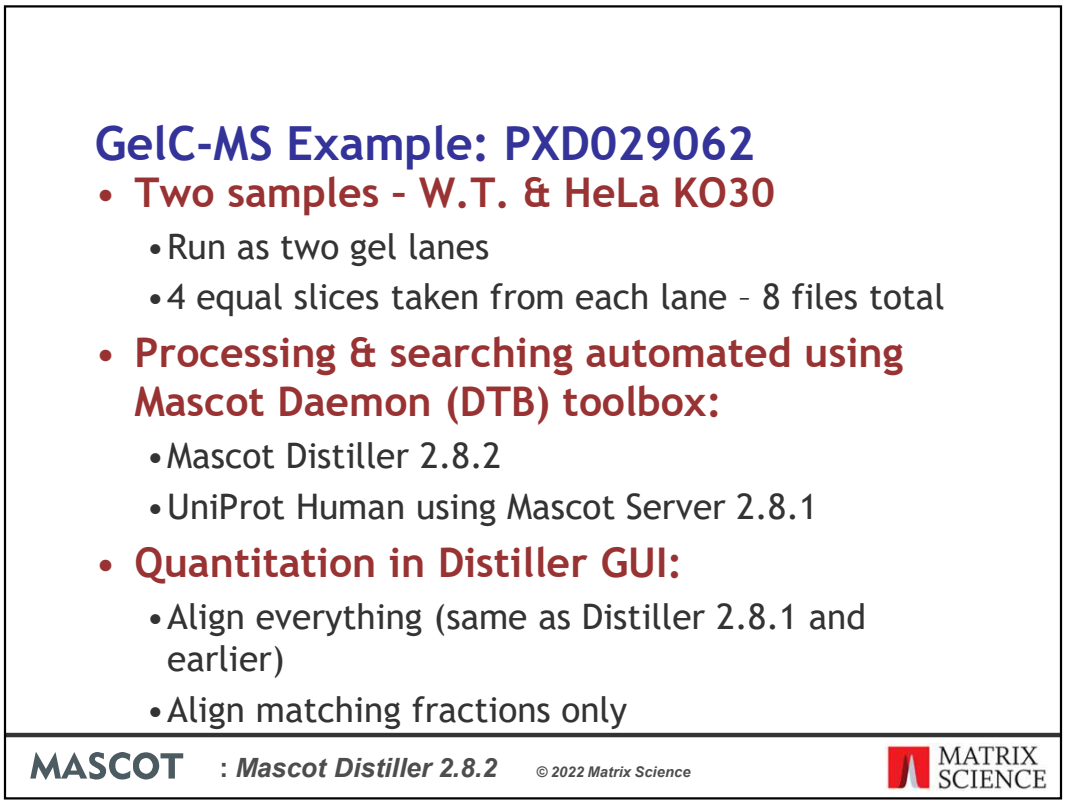
Fractionated LFQ in Mascot Distiller 2.8.2

Patrick Emery



In Mascot Distiller 2.8 we changed the way time alignment worked for label-free quantitation so that all the individual files in a project were aligned against a consensus generated by roughly aligning and the combining the Total Ion Chromatograms of each raw file. The method as implemented in Distiller 2.8 and 2.8.1 has the drawback that fractionated samples are not handled correctly, as the consensus generated from the different fractions will be incorrect due to there being little or no overlap in the peptides between fractions – at least if the fractionation has been efficient.

In Mascot Distiller 2.8.2 we have changed this so that multiple consensuses are generated, one for each fraction. Files are then aligned to the consensus for their assigned fraction. In addition to supporting fractionation, this makes quantitation of these types of sample faster as the system doesn't waste time looking for a peptide in all the files – only in the files for the fraction(s) it was identified in

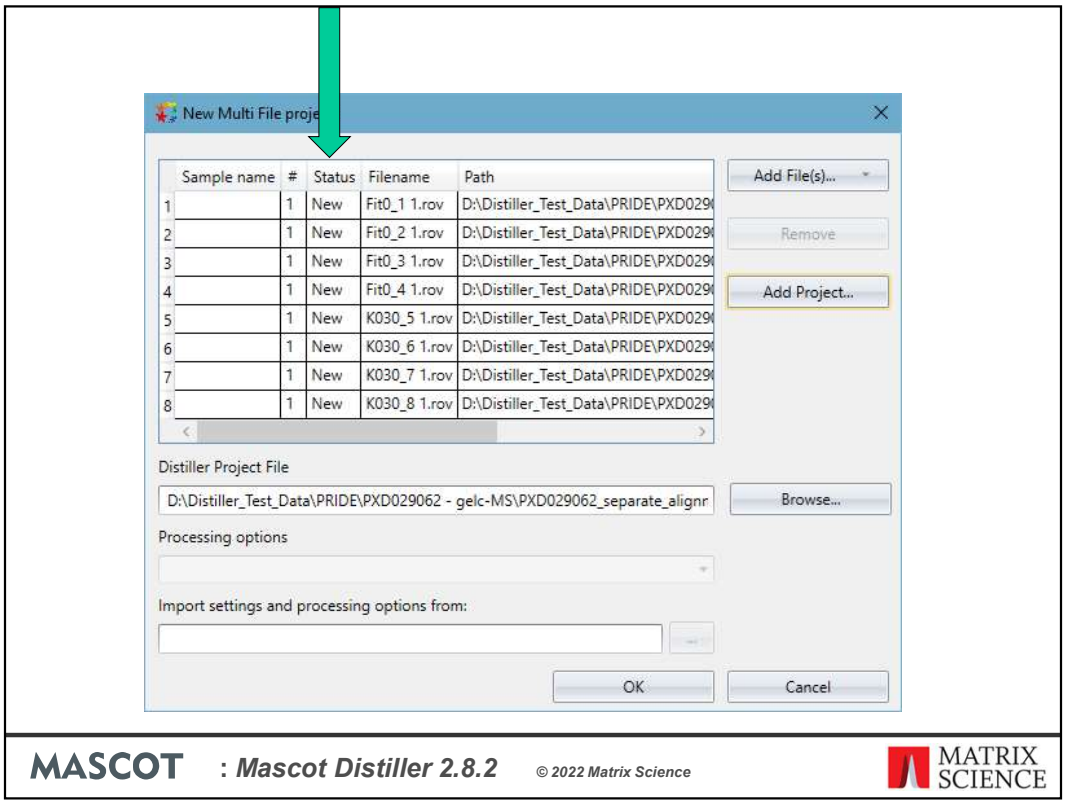


To demonstrate the new functionality we took a GelC-MS dataset from the PRIDE repository. This is a small, two sample dataset consisting of Wild Type and HeLa KO30 human cell lines. These were run as two separate lanes on a gel, and 4 equal slices were taken from each lane, digested with Trypsin and the run on the Mass Spec for a total of 8 raw files; 4 from each sample.

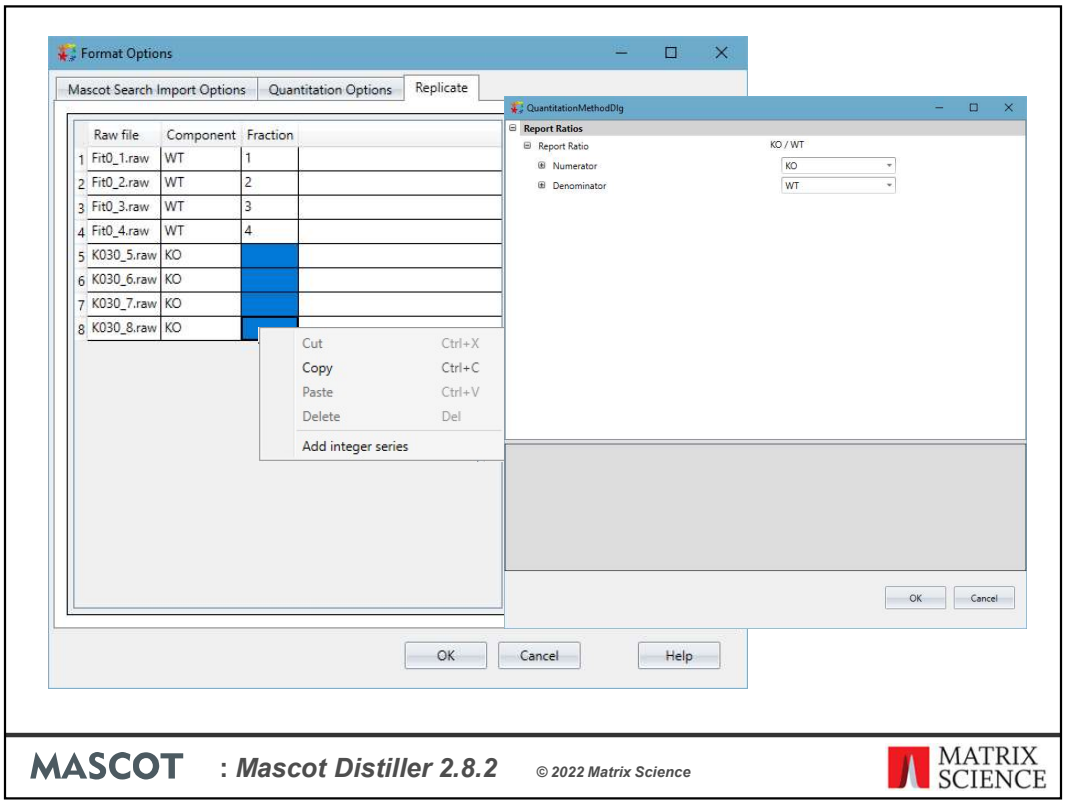
The raw files were processed and searched using Mascot Daemon with the Mascot Distiller Daemon Toolbox. Mascot Distiller 2.8.2 was used by Daemon for peak-picking, and the generated peak lists files searched with the "Label-free" quantitation method selected against the UniProt Human reference proteome using Mascot Server 2.8.1

The .rov project files generated by Daemon were then used to create a multifile project in the Mascot Distiller GUI and the data quantified in two different ways:

1. Align all files against each other, giving the same results as you would get in Distiller 2.8 or 2.8.1
2. Align only matching fractions using the new settings in Distiller 2.8.2



To create a multifile project from existing .rov project files in Distiller, go to File->New Multifile Project. Then click on “Add Project…” in the dialog that opens and select the .rov project files. When you do this, remember to check the “Status” column in the “New Multi File project” dialog for any errors.

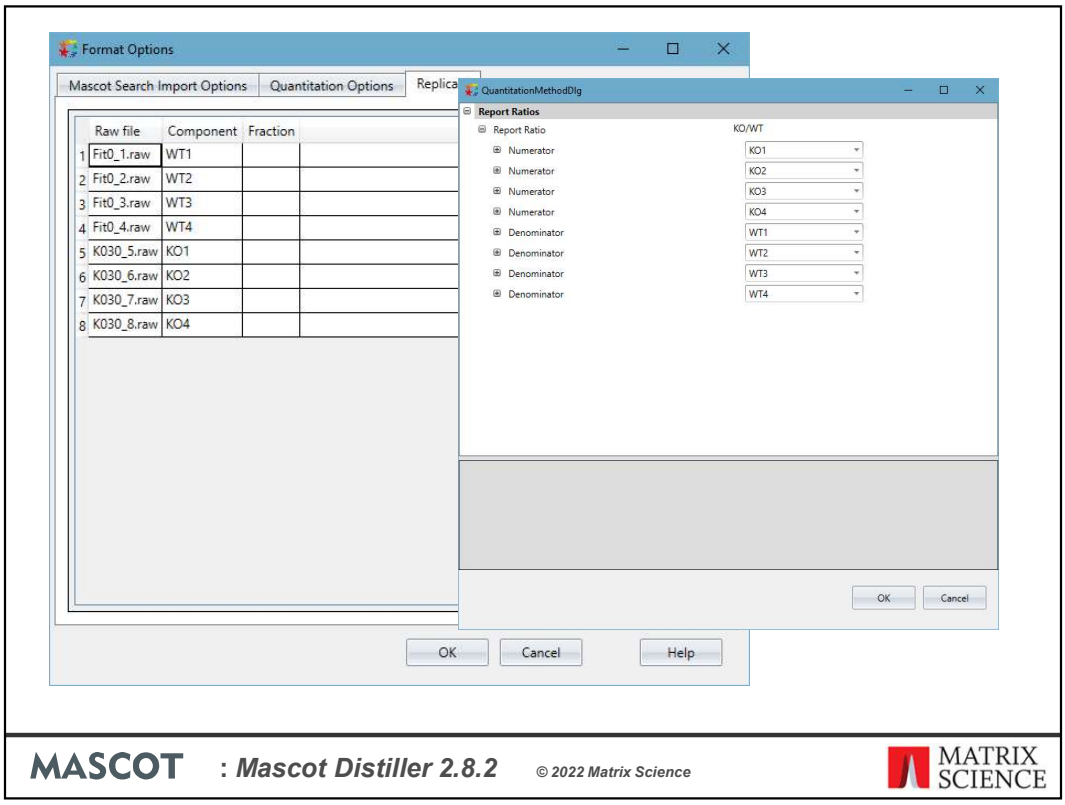


Once we've generated the mutlfile project and loaded the merged search, we need to setup the sample components and fractions in the "Format Options" dialog. The "Replicate" tab on this dialog has changed in Mascot Distiller 2.8.2 to allow you to do this. It now shows a table with the project raw files in the first column, and the columns for "Component" and "Fraction" which are user editable. Component is the sample ID – for this dataset we have two samples – W.T. and KO – so we've entered that into the "Component" column cells for the files as shown. If you have multiple technical or biological replicates of a sample, you should encode that information into the component names.

Each raw file is from a separate gel-slice or fraction, with each gel lane having been cut into four slices or fractions – we enter this information in the "Fraction" column. In this case we just have the four fractions per sample, so manually typing in the fraction numbers would be easy enough. However, if you have many more samples comprised of a larger number of fractions, it could get tedious to manually enter this information so you can instead select a range of cells, then right click to bring up the context menu and click "Add integer series". This will fill the cells with an integer series from 1 to the number of rows selected.

Once we've setup the components and fractions, click on the "Ratios" button to set up

the ratio(s) you want to calculate – selecting the required numerators and denomitors.



If you want all the files to be aligned against each other, this is how you setup the components on the "Replicates" tab – leave the fraction column blank. Each component name now needs to be unique, so we call the WT1, WT2 etc. When we set up the ratios, we now need to include each KO and WT component in the definition, as shown here.

GelC-MS Example: PXD029062

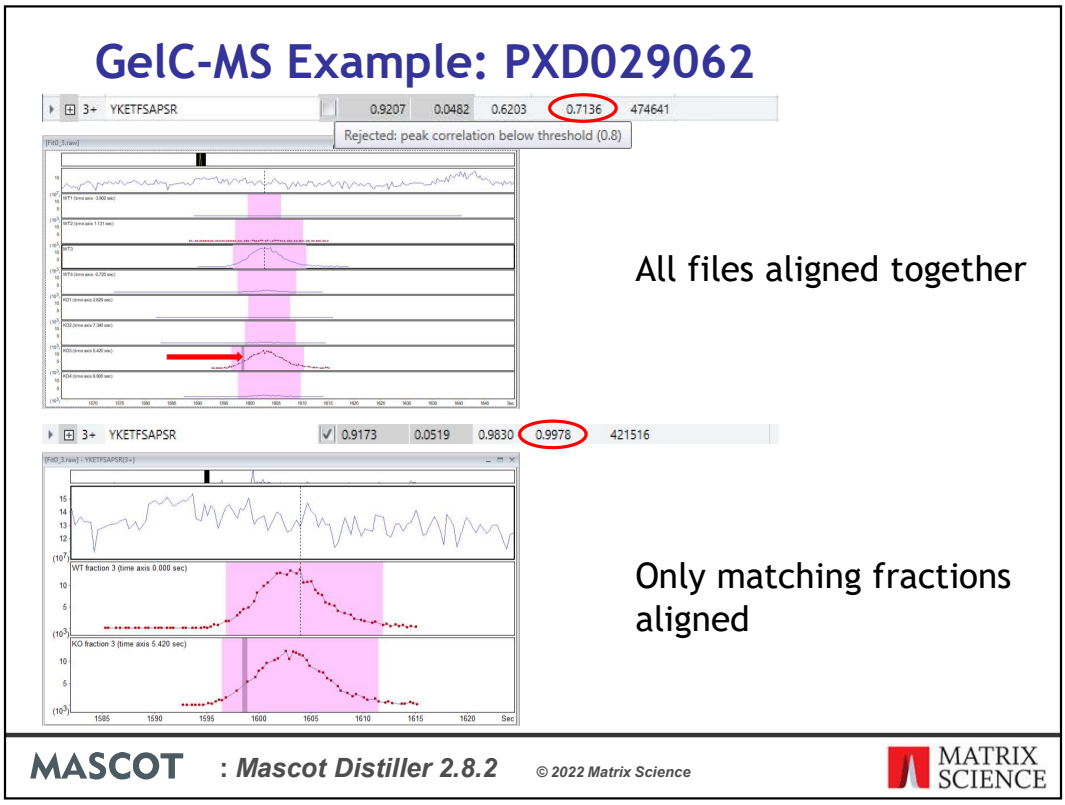
	Align Everything	Align Fractions
No. peptide ratios	1417	1417
Time taken (min)	26	11
No. valid peptide ratios	569	1137

MASCOT : *Mascot Distiller 2.8.2* MATRIX SCIENCE

Here are the overall results of running quantitation on the dataset with either aligning all files together, or just aligning matching fractions. In both cases we've quantified a total of 1417 peptides. However, quantitation where we've aligned just the fractions took only 11 minutes to complete compared with 26 minutes if we align all files – this is because Distiller doesn't waste time looking for peptides which were only identified in, say, fraction 3 in the files for fractions 1,2 and 4.

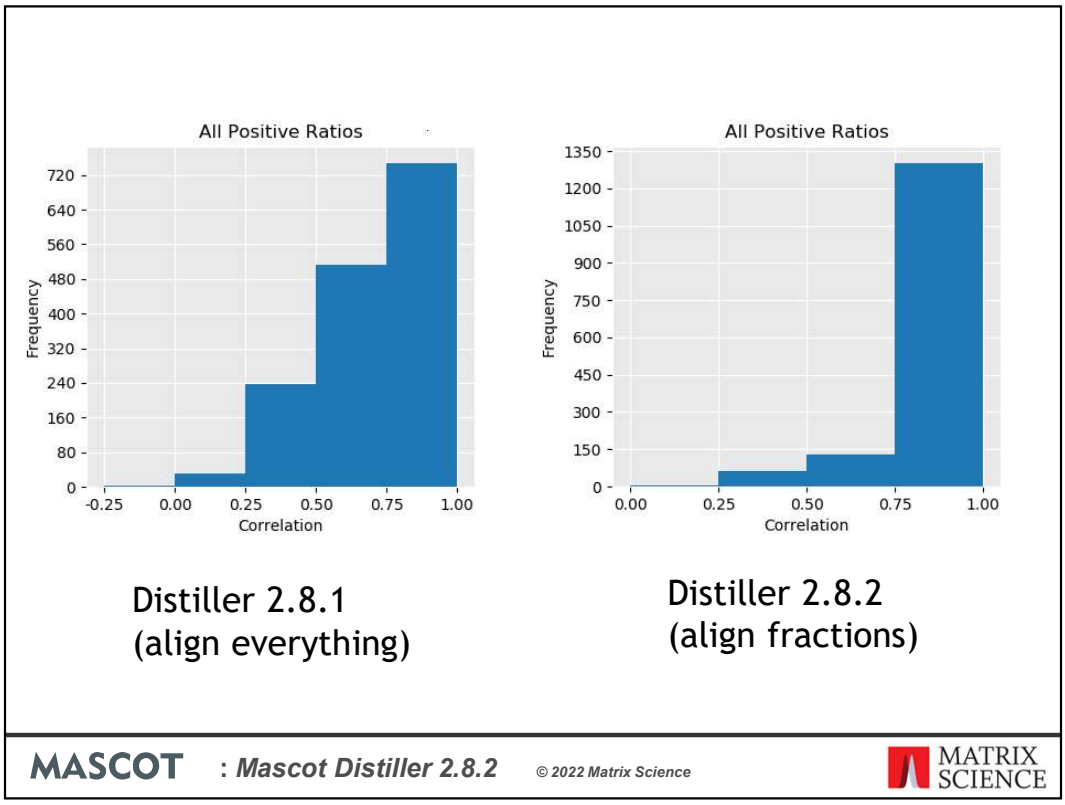
In addition to that, we're getting much better quality results back if we align just the individual fractions with 1137 quantified peptide passing our quality thresholds as opposed to just 569 if we align everything.

Therefore, aligning everything together – as you have to do in previous versions – is clearly the wrong way to handle these types of data.



Here we have an example of how only aligning matching fractions is improving our results. This is a quantified peptides from the dataset where we've aligned all the files together. It's only been identified by a single MS/MS match in the KO sample, fraction 3. But here, Distiller has looked for peptide signal in all the files in the project – as you can see, we have strong XICs from the WT and KO fraction 3 files, and essentially nothing from the other files – searching in which wasted time. The overall correlation for the observed and expected precursor isotope distributions for this peptide is 0.7136 – because the junk from fractions 1,2 and 4 has dragged the correlation down. This means that the peptide has been rejected because the correlation is below our default threshold of 0.8 – despite actually having good XICs from the correct fraction

Here we have the same match from the dataset where we've only aligned files within their matching fractions. You can immediately see that we don't have XICs for the other fractions where the peptide wasn't identified. Our precursor isotope distribution correlation between observed and expected signal is now 0.9978 because we don't have any junk matching from the other fractions, and so the peptide has passed our correlation threshold quality filter.



We can see the effect of this for the dataset as a whole using the "Quality" report in Distiller, which now includes frequency histograms of various metrics. Here we can see that the number of peptides with a correlation of between 0.75 and 1.0 is increased from ~720 to ~1300 if we align within matching fractions only.

A few other changes

- Option to export Volcano Plot report as CSV file
- Exposed Log2 ratio threshold on Volcano Plot report
- Separate component correlation, fraction columns on table-peptides-int report
- Frequency histograms added to Quality reports
- *…and many bug fixes*

And finally, a few other changes we've made in Distiller 2.8.2. If you already have a copy of Mascot Distiller, this is a free update.