

A practical guide to the MaxQuant computational platform for SILAC-based quantitative proteomics

Jürgen Cox¹, Ivan Matic¹, Maximiliane Hilger¹, Nagarjuna Nagaraj¹, Matthias Selbach², Jesper V Olsen¹ & Matthias Mann¹

¹Department for Proteomics and Signal Transduction, Max-Planck Institute for Biochemistry, Martinsried, Germany. ²Department for Cell Signalling and Mass Spectrometry, Max Delbrück Center for Molecular Medicine, Berlin, Germany. Correspondence should be addressed to J.C. (cox@biochem.mpg.de) or M.M. (mmann@biochem.mpg.de).

Published online 16 April 2009; doi:10.1038/nprot.2009.36

MaxQuant is a quantitative proteomics software package designed for analyzing large mass spectrometric data sets. It is specifically aimed at high-resolution mass spectrometry (MS) data. Currently, Thermo LTQ-Orbitrap and LTQ-FT-ICR instruments are supported and Mascot is used as a search engine. This protocol explains step by step how to use MaxQuant on stable isotope labeling by amino acids in cell culture (SILAC) data obtained with double or triple labeling. Complex experimental designs, such as time series and drug-response data, are supported. A standard desktop computer is sufficient to fulfill the computational requirements. The workflow has been stress tested with more than 1,000 liquid chromatography/mass spectrometry runs in a single project. In a typical SILAC proteome experiment, hundreds of thousands of peptides and thousands of proteins are automatically and reliably quantified. Additional information for identified proteins, such as Gene Ontology, domain composition and pathway membership, is provided in the output tables ready for further bioinformatics analysis. The software is freely available at the MaxQuant home page.

INTRODUCTION

Mass spectrometry-based proteomics¹ has become a very data-intensive science, especially since high-resolution Fourier transform mass spectrometers have become widespread^{2,3}. For instance, assume there would be 5,000 high-mass precision instruments, e.g., LTQ-Orbitraps^{4,5}, running worldwide without downtime, each performing 10 liquid chromatography/mass spectrometry (LC/MS) runs per day and generating 1 GB of raw data per run. Together, this would result in an annual data production rate of 18 PB (= 1.8×10^{16} bytes), which is more than the four experiments in the large hadron collider at CERN—the prime example for data-intensive experimental science—will produce per year (<http://lcg.web.cern.ch/LCG/>). Without any doubt, this situation calls for solid, efficient and standardized data-processing workflows that are widely applicable in quantitative proteomics. In ref. 6, we describe such a computational platform called MaxQuant, which is targeted at high-resolution quantitative data obtained with stable isotope labeling by amino acids in cell culture (SILAC)^{7,8} and which we have already successfully applied to a wide range of biological problems^{9–15}. Although the novel algorithmic concepts are explained in detail in ref. 6, here we give the step-by-step instructions how to analyze large-scale proteomic data with MaxQuant. Two example data sets are provided, which allow re-analysis of the data from two recent papers of our laboratory^{6,14,16}.

SILAC proteome experiments can be performed in many alternative ways, differing for instance in additional protein or peptide fractionation or in the kind of stable isotope labels used. Pre-fractionation of proteins, e.g., by gel electrophoresis, or a separation of digested peptides, e.g., with immobilized pI strips¹⁷, is compatible with the MaxQuant computational workflow. SILAC data from double- or triple-labeling¹⁸ experiments can be analyzed. Although a standard procedure is to label arginine and lysine when digesting proteins with trypsin, the kind of amino acids that are labeled as well as the atoms within the amino acids that are

replaced by stable heavy isotopes can be freely configured and may be adapted to any existing combination of labels. In addition to simple treatment/control SILAC comparisons, experiments with a more complex design can be analyzed, such as replicate measurements, SILAC time series with common reference points, label switches, multi-sample comparisons, and more. Furthermore, interaction data in the form of SILAC pulldowns¹⁹ can be analyzed. MaxQuant will in each case automatically assemble a matrix of SILAC ratios with rows corresponding to proteins and columns to different ‘samples’ or ‘conditions’ facilitating cross-experiment comparison of protein ratios. Thus, it is preferable in most situations to analyze data from different samples together in one MaxQuant project, so as to enable comparison of quantitative information directly. This is also a precondition for reliable control of protein false discovery rate (FDR).

Figure 1 shows an overview of the computational pipeline and the types of files that are exchanged. Raw files are generated by the instrumentation software and transferred to the local computer where they are loaded into the ‘Quant’ module. It performs all tasks that can be done before knowing the identity of peptides. In particular, the assembly of isotope patterns into SILAC pairs is already done here, before the submission of data to a tandem mass spectrometry (MS/MS) search engine. An advanced three-dimensional peak and isotope pattern detection is also carried out in this module. Output files are generated containing processed MS/MS spectra (‘msm’ files) bundled together from all LC/MS runs analyzed together, ready for submission to the Mascot search engine²⁰. A parameter file containing the search engine parameters is also created (‘par’ file), facilitating the submission of MS/MS spectra with Mascot Daemon (see below). The ‘Identify’ module takes the search engine results, the raw files (as well as intermediate results from the ‘Quant’ module), performs integration and statistical validation, assembles peptides into proteins, quantifies

proteins and writes out several tables containing the results as tab-separated text files (.txt). These can be uploaded for browsing and downstream bioinformatics analysis into programs such as Microsoft Office Excel, R in conjunction with Bioconductor, Spotfire, Matlab and the like.

The MaxQuant website will be expanded as a repository of tools and documents supporting the use of MaxQuant. Presently, we provide protein sequence databases for the most common organisms in MaxQuant-compatible formats ready for upload to Mascot. In addition, here we describe and provide a program ('Sequence Reverser') that allows creating a MaxQuant-compatible FASTA file given the user's own organism-specific protein list. Furthermore, a list of common contaminants—which can be expanded or configured—is included in Sequence Reverser.

MATERIALS

EQUIPMENT

Hardware requirements

- A personal computer (PC) with at least 2 GB of RAM ▲ **CRITICAL** At least a dual-core processor is recommended. Most computational parts scale with the number of available computing cores because of parallelization.
- Local storage is used for all raw files belonging to a project, and about half of this size for intermediate results. An external disc connected through USB 2.0 would be sufficient

Software requirements

- 32 bit versions of Windows XP or Windows Vista operating systems
- ▲ **CRITICAL** 'Regional and Language Options' have to be set to English.
- .NET Framework 2.0 is to be installed
- Thermo Fisher Scientific Xcalibur software ▲ **CRITICAL** Version must be compatible with your .raw files.
- Access to a Mascot (Matrix Science) server (currently version 2.2)
- Mascot Daemon installed on your local computer. Currently, MaxQuant uses Mascot (version 2.2, Matrix Science) as MS/MS search engine. For convenient and automatic submission of .msm and .par files generated by Quant, we recommend Mascot Daemon, a client application included in the Mascot purchase
- Microsoft Office Excel 2007. Recommended for browsing the result files
- ▲ **CRITICAL** Older versions of Excel will likely lead to problems due to limitations in allowed numbers of rows and columns.
- Sequence database (see EQUIPMENT SETUP)
- MaxQuant (see EQUIPMENT SETUP)

EQUIPMENT SETUP

MaxQuant software installation Go to MaxQuant home page and navigate to the 'Downloads' section. Please read the software license agreement carefully and stop at this point if you do not agree to its terms and conditions. When downloading MaxQuant you will receive a zipped file containing all necessary binaries and configuration files. Unzip this file (e.g., with WinZip) and store the resulting folder named 'MaxQuant' anywhere on the computer that one is going to use for the computations. No installation script needs to be executed. After saving the 'MaxQuant' folder, one needs to adapt the files in the folder 'MaxQuant\conf' to your local environment. The files 'enzymes,' 'mod_file,' 'mascot.dat' and 'unimod.xml' are copies of the corresponding files in the configuration folder of Mascot installation. These files have to be identical in the MaxQuant installation and on the Mascot server. Whenever anything is changed in the configuration of the Mascot server, e.g., modifications or sequence databases are added, one has to copy those files from the Mascot server into the MaxQuant installation, because local editing will lead to malfunction. SILAC labels can be configured in the file 'labels.txt,' which is a tab-separated text file and may be edited with Microsoft Office Excel.

Table 1 contains some examples of definitions for some standard SILAC labels. In the 'Composition' column, the labeled form of the amino acid is specified by its empirical formula. For this purpose, ^{13}C , ^2H and ^{15}N are represented by 'Cx,' 'Hx' and 'Nx,' respectively. The 'Mascot Name' column contains the modification as it appears in the 'mod_file.' Note that in the 'mod_file' the amino acid type (e.g., 'R') is automatically appended at the end of the name in brackets. 'Short Name' defines how the label is denoted in the graphical user interface of Quant.

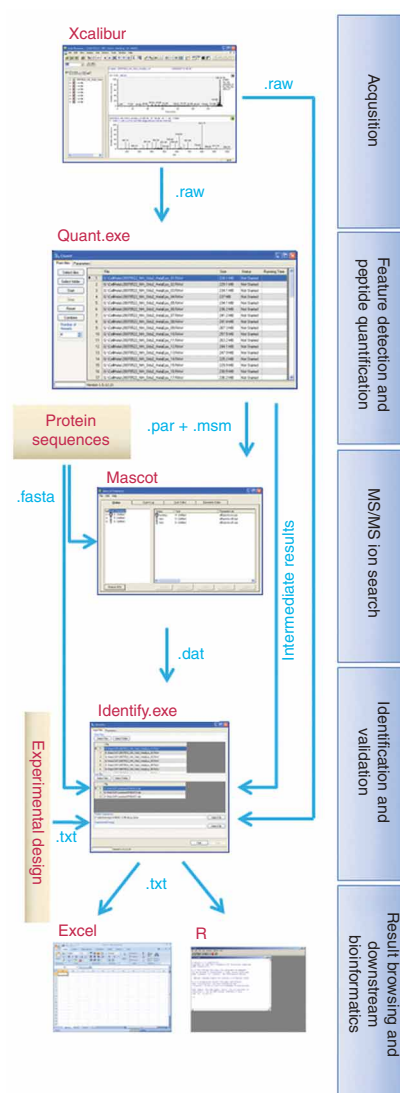


Figure 1 | Overview of the computational workflow. It consists of five steps: the first step—data acquisition—is performed by the vendor software of the mass spectrometer used; in the second step, the 'Quant.exe' module of MaxQuant detects peak features and quantifies peptides; in the third step, a search engine (here Mascot) associates fragment spectra with amino acid sequences; in the fourth step, the 'Identify.exe' module of MaxQuant validates and scores peptide identifications, assembles them to protein identifications and determines protein ratios; and in the fifth step, downstream bioinformatic analysis is performed by general purpose software (spreadsheets), statistical packages or bioinformatic packages.

Sequence database We recommend using a species-specific protein sequence database that includes all predicted proteins from an organism with fully sequenced genome. For instance, when analyzing human data, the human International Protein Index database²¹ may be used, or, alternatively, all human protein entries contained in Uniprot²², including the TrEMBL part, or all ENSEMBL²³ proteins. Redundancy of protein sequences does not matter at this point, as it will be dealt with at a later stage by the software, when assembling the identified peptides to proteins. MaxQuant validates scoring statistics on the basis of the hits to reversed protein entries in a target-decoy database²⁴. Therefore, it is necessary to include a reversed version for each original entry in the protein database FASTA file. Reversed entries have to be indicated by a recognizable prefix to the protein ID, e.g., 'REV_'. In addition, a set of common contaminant proteins can be included, also having a specific prefix, e.g., 'CON_'. These may, for instance, contain different forms of



PROTOCOL

TABLE 1 | The default content of the SILAC label configuration file ‘labels.txt’.

Composition	Mascot name	Short name	Amino acid
Cx6H12N4O	Arginine-13C6 (R-13C6) (R)	Arg6	R
Cx6H12Nx4O	Arginine-13C615N4 (R-full) (R)	Arg10	R
C6H8Hx4N2O	Lysine (D4) (K)	Lys4	K
Cx6H12N2O	Lysine-13C6 (K-13C6) (K)	Lys6	K
Cx6H12Nx2O	Lysine-13C615N2 (K-full) (K)	Lys8	K
C6H8Hx3NO	Leucine (D3) (L)	Leu3	L

It may need adaptation depending on the environment.

keratins or of abundant proteins from bovine serum if applicable. Visit the accompanying web page for pre-built MaxQuant-compatible FASTA files for the most common organisms and for downloading the program Sequence Reverser for reversing each entry and adding contaminants of

arbitrary protein sequence collections. The FASTA file should be uploaded to the Mascot server in the normal way as described in the Mascot manual. A copy of exactly the same file needs to be available locally on the computer where MaxQuant is running.

PROCEDURE

Preparation of data files

1 | Copy all Xcalibur .raw files belonging to a project to a single folder on a local computer. The files may reside on an external hard drive.

▲ CRITICAL STEP MaxQuant currently supports only files produced by LTQ-FT-ICR and LTQ-Orbitrap. Additionally, the current version of the Xcalibur software can only open ‘raw’ files smaller than 2 GB. Therefore, we suggest recording the MS/MS spectra in the centroid mode, which will keep file size sufficiently small for normal length gradients.

2 | Make sure that there is sufficient storage space available on the hard disc containing the raw data. In addition to the raw files, there should be at least half of this space available for intermediate results generated by MaxQuant during the calculations.

Quant.exe: feature detection and peptide quantification

3 | Start Quant.exe by double clicking on it. The program is in the ‘MaxQuant’ folder on your local PC.

4 | Go to the ‘Raw files’ tab. The location of this tab and those of the tabs described in later steps are shown in **Figure 2**.

5 | Load the Xcalibur .raw files to be analyzed by clicking the ‘Select files’ button. Alternately, select all the .raw files in a folder with the ‘Select folder’ button. The raw files will appear in the main table.

▲ CRITICAL STEP All raw files to be analyzed should be in the same folder.

▲ CRITICAL STEP MaxQuant’s strength is in the analysis of large numbers of LC/MS runs. When analyzing only very few LC/MS runs, the statistical evaluation of peptide identifications may suffer, as a sufficiently high number of identified MS/MS spectra are required to determine histograms used in intermediate steps of the calculation. Although it is possible to analyze even single runs separately with practically usable results, it is recommended to analyze together at least 10 LC/MS runs of reasonably high peptide complexity.

6 | Select the number of threads that will be run in parallel by MaxQuant. Each file will be analyzed by one process and the use of multiple processes will considerably shorten the time of analysis. If the number of threads selected is the same or higher than the number of available computing cores, the computer will be very busy and it will hardly be possible to use it for other purposes during the processing time.

7 | Go to the ‘Parameters’ tab.

8 | Choose the type of instrument that has produced the files. In the current MaxQuant version, ‘FT’ and ‘Orbitrap/FT Ultra’ can be chosen.

9 | Select the following types of SILAC experiment:

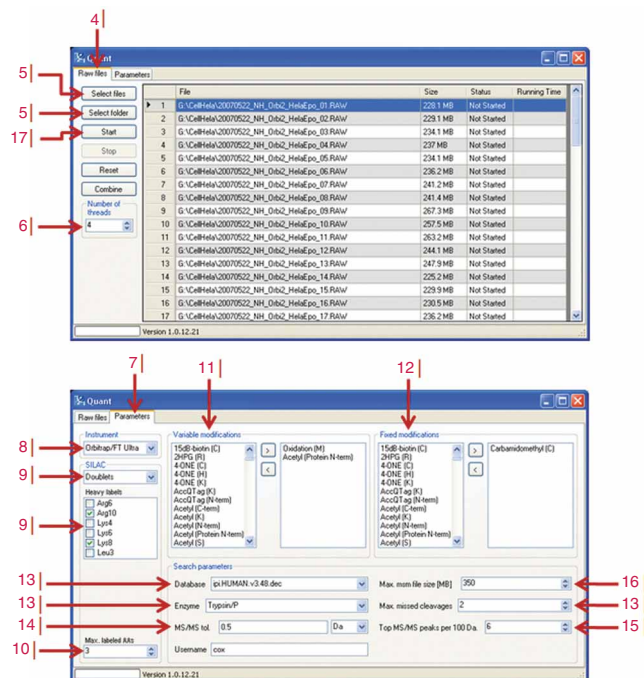


Figure 2 | The graphical user interface of the Quant module. The upper panel shows the ‘Raw files’ tab, whereas in the lower panel, the ‘Parameters’ tab is shown. The positions in the user interface of Quant.exe that correspond to numbered steps in the procedure are indicated by the step number.

- (a) 'Singlets' if no isotopic labeling was used; peptides and protein are identified and no quantification will be provided.
- (b) 'Doublets' (default selection) in case of a double SILAC labeling; a 'Heavy labels' panel will appear.
- (c) 'Triplets' for a triple SILAC labeling; a 'Medium labels' and a 'Heavy labels' panel will appear.

Select the appropriate labeled amino acids used in the SILAC experiment in the appropriate panels. It is assumed that in the 'Light' SILAC state, all amino acids have a natural isotopic composition. Note that the available labels are read from the labels.txt file in the 'conf' folder mentioned above.

10| Specify the maximum number of labeled amino acids ('Max. labeled AAs') a peptide can have to be detected by Quant.

11| Add variable modifications by selecting the desired modification on the left of the 'Variable modifications' panel and click on the right arrow button. The modification will then appear in the right panel. Per default oxidation of methionine and N-terminal protein acetylation are used. To remove a modification, select the modification to be removed in the right panel and click on the left arrow button. Variable modifications may or may not be present on a specific residue or a terminus.

▲ CRITICAL STEP Modifications related to SILAC labeling must not be specified here, as they are already automatically taken care of by defining the SILAC experiment type in Step 9.

▲ CRITICAL STEP Mascot, and consequently MaxQuant, only allows up to nine variable modifications. It is important to note that SILAC modifications (e.g., Arg10 and Lys8) as selected in the SILAC panels are *de facto* variable modifications when the database search is performed on the unpaired isotope patterns. In general, large numbers of variable modifications should be avoided because if they do not occur sufficiently often, they only tend to decrease the number of identifications at fixed FDR, and they will increase search times considerably (combinatorial explosion)²⁵.

12| Select or deselect, as desired, any 'Fixed modifications' in the same way as described for variable modifications.

Fixed modifications are applied to every occurrence of the specified residue or the terminus. For example, during the digestion reaction, iodoacetamide was used to alkylate cysteines, and to select 'Carbamidomethyl (C)' as fixed modification.

13| Choose from the 'Database' box menu the protein sequence database desired to be used in the Mascot search. Select the 'Enzyme' specificity according to the enzyme used during protein digestion. Select the maximum number of missed cleavages a peptide can have in order to be found in the Mascot search.

▲ CRITICAL STEP Selections of 'Max.-labeled AAs', 'Enzyme' and 'Max.-missed cleavages' are not independent. For most efficient application of SILAC, the choice of labeled amino acids should coincide with the enzyme specificity. In that case, the maximum number of labeled amino acids should be one more than the number of maximum missed cleavages. We then recommend 3 and 2, respectively.

14| Specify the maximum mass deviation allowed for the fragment ions ('MS/MS tol.'). Units can be selected as Dalton (Da) or parts per million (ppm). For a calibrated LTQ, a tolerance of 0.5 Da is recommended²⁶. The maximum mass deviation for parent ion masses is determined by MaxQuant on the basis of the achieved mass accuracy and does not need to be specified.

15| Select the number of most intense peaks per 100 Da in which Quant will be retained after processing of MS/MS spectra ('Top MS/MS peaks per 100 Da') for the Mascot database search. By default we use six, which is a good balance between scoring correct fragment peaks and suppression of noise.

16| Set the size limit in megabytes for the .msm files that Quant creates, containing the processed MS/MS spectra ready for submission to Mascot ('Max. msm file size (MB)'). If a file exceeds this limit, it will be split into two or more parts. A maximum file size of 350 MB is recommended, as for larger files occasional crashes of Mascot searches have been observed. Splitting does not affect results, as Identify.exe assembles data from all Mascot searches.

17| Go back to the 'Raw files' tab, press the 'Start' button and wait for the program to finish. A popup will indicate that it is done. In the same folder where the original raw files are stored, Quant will create a folder for each of your raw files containing intermediate results of computations. In addition, a folder 'combined' will be generated, which will, after Quant has finished, contain output files ending with .par and .msm; .msm files contain the processed MS/MS spectra, whereas .par files carry the corresponding parameter settings for the Mascot searches. Depending on the SILAC setting, Quant will create one, three or four kinds of paired .par/.msm files (**Table 2**), which differ in the way in which SILAC labeling-related modifications are treated, either as fixed or as variable modifications. As the SILAC state of many isotope patterns is known earlier, the label modifications are treated as fixed modifications in the Mascot search. For unpaired isotope patterns, the search is done in the conventional way with variable label modifications.

? TROUBLESHOOTING



PROTOCOL

TABLE 2 | Quant output files for the submission to Mascot.

SILAC	MSM file	PAR file	Description
Singlets	test.iso_0.msm	test.iso.par	MS/MS on single isotope patterns. (No SILAC pair assembly is done)
Doublets	test.iso_0.msm	test.iso.par	MS/MS on single isotope patterns. SILAC labels are treated as variable modifications
	test.sil0_0.msm	test.sil0.par	SILAC state 'light.' No SILAC labels modifications are used
	test.sil1_0.msm	test.sil1.par	SILAC state 'heavy.' Heavy SILAC labels are treated as fixed modifications
Triplets	test.iso_0.msm	test.iso.par	MS/MS on single isotope patterns. SILAC labels are treated as variable modifications
	test.sil0_0.msm	test.sil0.par	SILAC state 'light.' No SILAC label modifications are used
	test.sil1_0.msm	test.sil1.par	SILAC state 'medium.' Medium SILAC labels are treated as fixed modifications
	test.sil2_0.msm	test.sil2.par	SILAC state 'heavy.' Heavy SILAC labels are treated as fixed modifications

MS/MS, tandem mass spectrometry; SILAC, stable isotope labeling by amino acids in cell culture.

Mascot: MS/MS ion search

18| Once Mascot Daemon is launched, click on the 'Task Editor' tab. Insert the name of the task (e.g., 'test sil0') into the field entitled 'Task.'

19| Upload one of the .par files in the 'combined' folder into the 'Parameter set' field. Upload the corresponding .msm file(s) by clicking on the 'Add Files...' button or by using drag&drop from an explorer window. Click on the 'Run' button. The display will switch to the 'Status' tab. Once the task is running, the corresponding task icon will change to a clock.

20| Repeat Step 19 with the other files until all the searches have been submitted.

? TROUBLESHOOTING

21| Once the search is finished, expand a task node and click on the result node to see the name of the generated .dat file. There will be one .dat file for each .msm file searched in Mascot.

22| Using an internet browser, go the home page of your Mascot server.

23| In the address bar of the browser, append '/data' at the end of the address after '.../mascot'. For example, if the address of your Mascot server is http://hansi.biochem.mpg.de:2000/mascot, go to http://hansi.biochem.mpg.de:2000/mascot/data. See the MaxQuant website FAQ for further information.

24| Open the folder with the name of the date when the search was performed.

25| To save the .dat file, right click and select 'Save Target As...'. It is convenient (but not required) to save them in the combined folder containing the other files of the same project.

Identify.exe: identification and validation

26| Locate the program Identify.exe in the 'MaxQuant' folder on the local PC and start it by double clicking it.

27| In the 'Input files' tab, upload the .raw files that have been previously run with Quant using either the 'Select files' or the 'Select folder' button. The location of this tab, and the tabs described in later steps, is shown in **Figure 3**.

28| Upload the .dat files in the same way.

29| Select the appropriate Protein sequences file in the FASTA format.

▲ **CRITICAL STEP** On the local computer, one must have exactly the same sequence database (FASTA) file as used for the Mascot database search. Otherwise, Identify will report an error message and will not be able to finish.

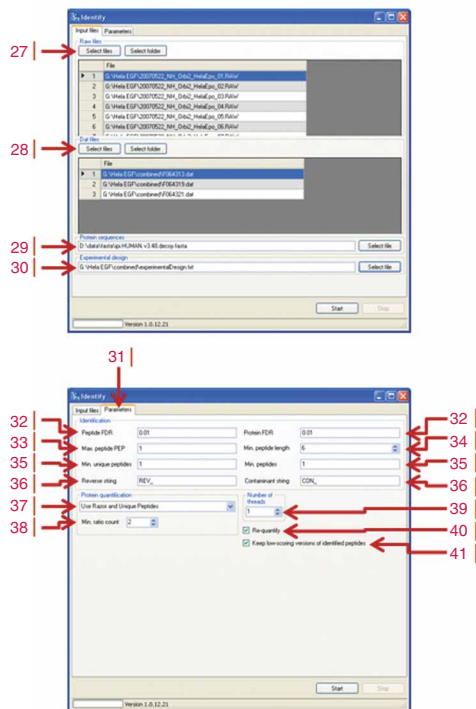


Figure 3 | The graphical user interface of the Identify module. The upper panel shows the 'Input files' tab, whereas in the lower panel, the 'Parameters' tab is shown. The positions in the graphical interface of the Identify module that correspond to numbered steps in the procedure are indicated by the step number.

BOX 1 | SPECIFYING THE EXPERIMENTAL DESIGN

To facilitate the inter-sample comparison of protein expression ratios or to assemble data into a specific form—for instance into a time series—the software needs to know which LC/MS runs belong to which ‘time point.’ This can be specified by uploading an experimental design file in Step 30. If no experimental design file is used, SILAC ratios and other protein information are provided as a whole for the entire data set. A template file for the experimental design is created automatically in the ‘combined’ folder. Open this template file in Excel, make appropriate changes and save it as a tab-delimited text file. If the ‘Experiment’ column is filled in, separate average ratios (and other information) will be reported for each different term used in the Experiment column. For instance, suppose one measures a six-point time course with five SILAC experiments by using the time zero sample always as the light state and the samples at five later time points in the heavy states. In addition, one may have separated proteins into 10 gel slices, resulting in 50 LC/MS runs. In this case one would fill the ‘Experiment’ column with five different terms, which would give rise to an expression ratio matrix with five columns. In addition, in the ‘Slice’ column denote the numbers 1–10 regarding which gel slice an LC/MS run corresponds to. Additional statistics on slice-specific identifications of proteins will then be provided, which might, for instance, be useful to detect isoforms differing in molecular weight. In the ‘Invert’ column, one can specify if the ratios originating from certain LC/MS runs should be inverted. This is useful when the labels have been swapped.

- 30| Insert the experimental design file (optional; see **Box 1**).
- 31| Go to the ‘Parameters’ tab.
- 32| Choose the desired FDRs at the protein and peptide levels. For a comprehensive explanation, see ref. 6. The default value is in both cases 0.01 (1%).
- 33| Select the desired maximum posterior error probability, which is the probability of a false hit given the peptide identification score and length of peptides⁶. The default value is 1, corresponding to no additional filtering.
- 34| Set the minimum peptide length. Peptide shorter than the threshold will not be reported nor be considered for protein identification and quantification. Short peptides are usually not unique in the protein database and therefore not statically informative in any case.
- 35| Select the minimum number of unique and total peptides a protein group should have to be considered as identified and reported in the final table.
- 36| Specify the string for reverse and contaminant hits as used in the protein sequences file. We usually use ‘REV_’ and ‘CON_’, respectively.
- 37| Specify how the protein ratios will be calculated (‘Protein quantification’). When ‘Use all peptides’ is selected, the quantification is done on all peptides. With ‘Use unique peptides,’ only the peptides unique for that specific protein group are used for quantification. The ‘Use unique and razor peptides’ mode calculates ratios from unique and razor peptides. Razor peptides are non-unique peptides assigned to the protein group with the most other peptides (Occam’s razor principle).
- 38| Select the minimum ratio count for protein quantification. A protein with a lower number of quantified SILAC pairs/triplets will not be used for calculating the protein ratio. As ratio counts are reported in the output files, it is possible to filter the results in downstream applications (e.g., Excel).
- 39| Select the number of parallel threads.
- 40| Tick the ‘Re-quantify’ box if Identify should calculate the ratio for isotopic patterns not assembled in SILAC pairs by Quant. The shapes of the identified isotope pattern will be translated to the place in the m/z retention time plane, where its missing SILAC partner is expected and intensities will be integrated over these regions. This is particularly helpful for quantifying proteins with very high ratios, e.g., from pulldowns, where one of the SILAC partners is at or below noise level. However, in some cases where extreme ratios are expected, e.g., in incorporation studies, we do not recommend using ‘Re-quantify’.
- 41| If desired, check ‘Keep low-scoring versions of identified peptides.’ If checked, additional MS/MS spectra will be accepted beyond those that have individually passed the criteria for identification. This means that an MS/MS spectrum will be accepted even if its posterior error probability value is not sufficient, as long as the highest scoring peptide sequence for that spectrum has been identified with another MS/MS spectrum. This has no effect on the total number of identified peptide sequences or proteins. This option is particularly beneficial for complex experimental designs, as it increases the likelihood for each protein of finding a ratio in different samples or time points.
- 42| Press the Start button and wait for the program to finish. A popup will appear when it is done.

? TROUBLESHOOTING



PROTOCOL

TIMING

Computation times vary with sample complexity, richness of spectra and LC gradient length. Typical values for 72 LC/MS runs are Steps 1–17: 10 h; Steps 18–25: 4 h; and Steps 26–42: 2 h.

Almost all algorithmic parts of MaxQuant are parallelized and execution is automatically distributed to the number of specified computing cores. Although not tested rigorously, Windows clusters may enhance processing speed significantly. Disk access is likely also limiting and therefore solid-state disks may improve processing speed.

TRUBLESHOOTING

Troubleshooting advice can be found in **Table 3** and in the FAQ on the MaxQuant.org website. For more detailed assistance and troubleshooting with Mascot and Mascot Daemon, refer to the Mascot manual and support.

TABLE 3 | Troubleshooting table.

Step	Problem	Possible reason	Solution
17	Quant.exe crashes	Maximum file size of 2 GB may be exceeded	Record MS/MS spectra in the centroid mode
20	Quant.exe or Identify.exe crashes	Storage space may not be enough	Use an external disc as storage space
	Mascot crashes	.msm file may be too big in size	Use a maximal msm file size of 350 MB or even smaller
42	Identify.exe crashes	FASTA file used for Mascot search missing on your local computer	Copy the exact same to your local computer
		Wrong .dat files uploaded	Check the .dat file number of your 'task' in Mascot Daemon
	Identify.exe reports error of unknown protein ID	Wrong FASTA file uploaded	Use the same database for the Mascot search and Identify.exe
Any	Unforeseen	Unknown	First, double check the FAQ on MaxQuant home page. If still unresolved, describe the problems in the MaxQuant Google group at http://groups.google.com/group/maxquant-list

MS/MS, tandem mass spectrometry.

BOX 2 | MAXQUANT OUTPUT TABLES

- **parameters.txt**—summary of parameters used for analysis
Version number of software used, threshold values used for identification and quantification, mode of quantification and so on
- **msScans.txt**—full scan summary
Information about the full scan details stored in the raw file including the fill time and total elapsed time, scan numbers, total ion current, dead time and so on
- **msmsScans.txt**—msms scan summary
Consists of the raw file parameters including the peak type, scan number, total ion count and sequence information wherever there was an identification event
- **allPeptides.txt**—full scan peptides details
Isotopic clusters, charge state, m/z values. Mass precision, retention time and whether the peptide is picked up for sequencing or not and so on
- **msms.txt**—summary of scan parameters for identified peptides
Similar to msmsScans file but restricted to identified peptides only
- **evidence.txt**—one stop master information
Peptide sequence, protein groups a peptide belong to, modification state, experiments in which peptide was identified, PTM score, mass error calibrated and un-calibrated, MS/MS count, raw file detail, gel slice/fraction and so on
- **peptides.txt**—concise non-redundant list of identified peptide sequences
Peptide sequence, proteins groups that contain the peptide, modification, miss cleavages, length of peptide, PEP values and so on. Identification scores for the best-identified version are displayed
- **modifiedPeptides.txt**—non-redundant list of identified peptide sequences with specific modifications
Peptide sequence, modification content, PTM score, intensity, Mascot score and so on
- **summary.txt**
Overall summary of whole analysis including identification success rate for individual raw files, the percentage of isotopic clusters picked up for sequencing and isotope patterns repeatedly sequenced
- **proteinGroups.txt**
Comprehensive list of identified proteins from the whole analysis, cross-references to various databases including Swissprot, Ensembl, Kegg, GO and so on

BOX 3 | INTEGRATED PROTEIN ANNOTATION

For the most common organisms, we automatically provide additional protein annotation in the output files. This is based on the mappings of Uniprot IDs to the respective specialized annotation IDs, as they are provided in the latest database dump of Uniprot. The Uniprot IDs for each protein are extracted from the FASTA header by matching the patterns SWISS-PROT:[ID_1;...;ID_n] and TREMBL:[ID_1;...;ID_n], as they occur, for instance, in the IPI FASTA files. In the proteinGroups.txt table, we report

- Gene Ontology (GO): GO IDs as well as their descriptive names are provided separately for biological process, molecular function and cellular component
- Pfam domain content: Occurrence of protein domain families represented by a hidden Markov model in the Pfam repository is indicated
- Membership of a protein in a KEGG pathway

ANTICIPATED RESULTS

Identify will create several tables as tab-delimited .txt files that contain the results of the MaxQuant analysis. See **Box 2** for a description of the different output tables. A document with detailed explanation of the columns can be downloaded from the MaxQuant website. Several annotation columns, such as Gene Ontology²⁷, Pfam²⁸ domain content and KEGG²⁹ pathway membership, are automatically shown in the table of proteins (**Box 3**).

A benchmark data set consisting of 72 LC/MS runs from epidermal growth factor-stimulated HeLa cells used in ref. 6 can be downloaded from <http://www.proteomecommons.org/data/show.jsp?id=7816>. Identification and quantification of more than 4,000 proteins as described in ref. 6 are expected. In another example, Pan *et al.*¹⁶ quantified the proteome of a cell line against primary cells of the same cell type. Furthermore, quantification of more than 4,000 proteins is expected from the analysis of the combined data set, which is also deposited at <http://www.proteomecommons.org>.

ACKNOWLEDGMENTS We thank all the other members of the Proteomics and Signal Transduction group for help with the development of MaxQuant. This work was supported by the Max-Planck Society and by the 6th Framework Program of the European Union (Interaction Proteome Grant LSHG-CT-2003-505520 and HEROIC Grant LSHG-CT-2005-018883).

Published online at <http://www.natureprotocols.com/>

Reprints and permissions information is available online at <http://npg.nature.com/reprintsandpermissions/>

- Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature* **422**, 198–207 (2003).
- Zubarev, R. & Mann, M. On the proper use of mass accuracy in proteomics. *Mol. Cell. Proteomics* **6**, 377–381 (2007).
- Mann, M. & Kelleher, N.L. Special feature: Precision proteomics: the case for high resolution and high mass accuracy. *Proc. Natl Acad. Sci. USA* **105**, 18132–18138 (2008).
- Makarov, A. *et al.* Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. *Anal. Chem.* **78**, 2113–2120 (2006).
- Olsen, J.V. *et al.* Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol. Cell. Proteomics* **4**, 2010–2021 (2005).
- Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).
- Ong, S.E. *et al.* Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* **1**, 376–386 (2002).
- Mann, M. Functional and quantitative proteomics using SILAC. *Nat. Rev. Mol. Cell Biol.* **7**, 952–958 (2006).
- de Godoy, L.M. *et al.* Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature* **455**, 1251–1254 (2008).
- Bonaldi, T. *et al.* Combined use of RNAi and quantitative proteomics to study gene function in *Drosophila*. *Mol. Cell* **31**, 762–772 (2008).
- Selbach, M. *et al.* Widespread changes in protein synthesis induced by microRNAs. *Nature* **455**, 58–63 (2008).
- Graumann, J. *et al.* Stable isotope labeling by amino acids in cell culture (SILAC) and proteome quantitation of mouse embryonic stem cells to a depth of 5,111 proteins. *Mol. Cell. Proteomics* **7**, 672–683 (2008).
- Zanivan, S. *et al.* Solid tumor proteome and phosphoproteome analysis by high resolution mass spectrometry. *J. Proteome Res.* **7**, 5314–5326 (2008).
- Cox, J. & Mann, M. Is proteomics the new genomics? *Cell* **130**, 395–398 (2007).
- Schimmel, J. *et al.* The ubiquitin-proteasome system is a key component of the SUMO-2/3 cycle. *Mol. Cell. Proteomics* **7**, 2107–2122 (2008).
- Pan, C., Kumar, C., Bohl, S., Klingmueller, U. & Mann, M. Comparative proteomic phenotyping of cell lines and primary cells to assess preservation of cell type-specific functions. *Mol. Cell. Proteomics* **8**, 443–450 (2009).
- Hubner, N.C., Ren, S. & Mann, M. Peptide separation with immobilized pI strips is an attractive alternative to in-gel protein digestion for proteome analysis. *Proteomics* **8**, 4862–4872 (2008).
- Blagoev, B., Ong, S.E., Kratchmarova, I. & Mann, M. Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics. *Nat. Biotechnol.* **22**, 1139–1145 (2004).
- Vermeulen, M., Hubner, N.C. & Mann, M. High confidence determination of specific protein-protein interactions using quantitative mass spectrometry. *Curr. Opin. Biotechnol.* **19**, 331–337 (2008).
- Perkins, D.N., Pappin, D.J., Creasy, D.M. & Cottrell, J.S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551–3567 (1999).
- Kersey, P.J. *et al.* The International Protein Index: an integrated database for proteomics experiments. *Proteomics* **4**, 1985–1988 (2004).
- UniProt Consortium. The universal protein resource (UniProt). *Nucleic Acids Res.* **36**, D190–D195 (2008).
- Flicek, P. *et al.* Ensembl 2008. *Nucleic Acids Res.* **36**, D707–D714 (2008).
- Elias, J.E. & Gygi, S.P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* **4**, 207–214 (2007).
- Pevzner, P.A., Mulyukov, Z., Dancik, V. & Tang, C.L. Efficiency of database search for identification of mutated and modified proteins via mass spectrometry. *Genome Res.* **11**, 290–299 (2001).
- Cox, J., Hubner, N.C. & Mann, M. How much peptide sequence information is contained in ion trap tandem mass spectra? *J. Am. Soc. Mass Spectrom.* **19**, 1813–1820 (2008).
- Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29 (2000).
- Finn, R.D. *et al.* The Pfam protein families database. *Nucleic Acids Res.* **36**, D281–D288 (2008).
- Kanehisa, M. *et al.* KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* **36**, D480–D484 (2008).