Published in final edited form as:

J Proteome Res. 2014 February 7; 13(2): 506–516. doi:10.1021/pr400675n.

SILAC Peptide Ratio Calculator - A tool for SILAC Quantitation of Peptides and Posttranslational Modifications

Xiaoyan Guan¹, Neha Rastogi², Mark R. Parthun², and Michael A. Freitas^{3,*}

¹Department of Chemistry and Biochemistry, The Ohio State University, Columbus, OH 43210

²Department of Molecular and Cellular Biochemistry, The Ohio State University, Columbus, OH 43210

³Department of Molecular Virology, Immunology and Medical Genetics, The Ohio State University, Columbus, OH 43210

Abstract

This paper describes an algorithm to assist in relative quantitation of peptide post-translational modifications using Stable Isotope Labeling by Amino acids in Cell culture (SILAC). The described algorithm first determines the normalization factor and then calculates SILAC ratios for a list of target peptide masses using precursor ion abundances. Four yeast histone mutants were used to demonstrate the effectiveness of this approach for quantitation of peptide post-translational modifications changes. The details of the algorithm's approach for normalization and peptide ratio calculation are described. The examples demonstrate the robustness of the approach as well as its utility to rapidly determine changes in peptide posttranslational modifications within a protein.

Keywords

Quantitative proteomics; global normalization factor; ratio/charge filter; peptide clustering; sliding time window

Introduction

Stable Isotope Labeling by Amino acids in Cell culture (SILAC) is a multiplex quantitative proteomic technique that uses metabolic labeling of proteins with isotopically heavy amino acids to determine changes in protein relative abundance¹. In a classic SILAC experiment, cells are metabolically labeled by culturing them in media that contains either light or heavy stable isotope-labeled amino acids. The light and heavy labeled cells are mixed followed by the protein extraction and proteolytic digestion. The digest is then analyzed by liquid chromatography tandem mass spectrometry (LC-MS/MS) to identify proteins and assess changes in protein abundance across sample treatments. For each proteolytic fragmentation, the mass spectra is represented by a light and heavy set of signals with a distinct mass difference that equals the incorporated label. By comparing the relative abundance of the light vs. heavy labeled signals, changes in protein relative amount can be measured.

*Address reprint requests to: Dr. Michael A. Freitas (Address: 906 Biomedical Research Tower, 460 West 12th Avenue, The Ohio State University, Columbus, OH 43210; Phone: 614-688-8432; Fax: 614-688-8675; freitas.5@osu.edu) or Dr. Mark R. Parthun (Address: 357 Hamilton Hall, 1645 Neil Ave, The Ohio State University, Columbus, OH 43210; Phone: 614-292-6215; Fax: 614-292-4118; parthun.1@osu.edu).

SUPPORTING INFORMATION AVAILABLE STATEMENT

The supporting materials including SI figure 1–3 and SI table 1 & 2 are available free of charge via the Internet at <http://pubs.acs.org>.

Software applications exist that facilitate protein SILAC data interpretation. These software applications (eg. MassMatrix ², SILACtor ³, Census ⁴, MaxQuant ⁵ and Uniquant ⁶) calculate protein relative abundances from the heavy and light abundances of the MS1 precursor ions of identified peptides. The SILAC ratios for each protein are aggregated from the multiple peptide IDs to generate the overall protein ratio. Other applications, such as Xpress ⁷, ASAPRatio ⁸ and MSQuant ⁹ determine the SILAC ratios using peak area integration of extracted ion chromatograms (XIC). SPeCtRA (SILAC Peptide Count Ratio Analysis) take a very different approach in which the MS² spectra are used for protein quantitation ¹⁰.

Peptide level changes are especially important for characterization of peptide modifications. Another widely used tool for SILAC ratio quantitation is MaxQuant. MaxQuant has the potential to be used for peptide level SILAC due to its feature detection capabilities. For SILAC experiments, MaxQuant will attempt to calculate a SILAC ratio for every detected feature. If the peptide is identified with the builtin Andromeda search engine, MaxQuant will then assign a peptide sequence and use the median ratio to approximate its SILAC value. Similarly, a protein's SILAC ratio is taken from the median ratio of all assigned peptides. The use of median ratios is appropriate for estimation of the SILAC ratio for a protein. However, a more robust strategy is required for peptide ratios.

In this paper a simple yet robust computational strategy is put forth to assist in determining peptide SILAC ratios for features (either identified or unidentified) in LC-MS/MS data sets. The paper describes both the important aspects of normalization factor estimation and peptide ratio calculation. Comparisons are made with MassMatrix and MaxQuant to emphasize the complementary use of the algorithm alongside database search and feature detection algorithms. The primary utility of this algorithm is directed at the accurate and precise SILAC quantification of peptide posttranslational modifications. An implementation of this approach is presented wherein changes in histone PTMs were quantified in a model yeast system.

EXPERIMENTAL PROCEDURES

Preparation of histone mutants

Lysine residues on the histones H3 and H4 that are subject to post-translational modification were mutated to glutamine or arginine. The mutations were constructed by site-directed mutagenesis from plasmid pMP3 that carries both wild type histone H3 and H4 genes (HHT2-HHF2) ¹¹. Plasmids containing histone mutations were shuffled into yeast strain UCC1111 (*Saccharomyces cerevisiae*), which is deleted for the endogenous histone H3 and H4 genes. Modified residues in histones H2A and H2B were altered by site directed mutagenesis of plasmid pQQ18. These plasmids were then shuffled into strain JHY205 ¹².

Cell preparation

For each wild type and histone mutant pair to be analyzed, strains were grown separately in duplicate cultures-one containing normal L-lysine and the other containing ¹³C₆, ¹⁵N₂-lysine. Cells were harvested by centrifugation in mid-log phase A₆₀₀ = 0.6 to 0.8. Equal weights of cells from each culture were mixed and histones were subsequently isolated.

Histone extraction and In-solution digestion

After the cells were harvested, histones were purified as described previously ¹³. Each sample was digested with Endoproteinase Arg-C (Sequencing grade from *Clostridium histolyticum*, Roche Applied Science) at a 1:50 enzyme:substrate ratio in 100 mM Tris-HCl

(pH 7.6). The digest was incubated at 37°C for 8 hours. The resulting peptides were desalted with a C₁₈ Zip-tip and eluted with 0.1% trifluoroacetic acid (TFA)/50% acetonitrile (ACN).

Protein identification

Eluting peptides were separated by reversed-phase HPLC (Dionex Ultimate 3000 capillary/nano HPLC system, Dionex, Sunnyvale, CA) and mass analyzed with a Thermo Fisher LTQ Orbitrap XL (ThermoFisher, San Jose, CA). Histones were separated on a 0.2 mm × 150 mm C₁₈ column (3 μm, 200Å, Michrom Bioresources Inc., Auburn, CA) at a flow rate of 2 μL/min with mobile phase A containing H₂O with 0.1% Formic acid (FA) and mobile phase B containing ACN with 0.1% FA. Using a 120 min gradient beginning with 2% mobile phase B, the phase B linearly increased to 5% in 10 min, from 5% to 15% in 20 min, from 15% to 30% in 45 min, from 30% to 50% in 15 min, and from 50% to 90% in 5 min. After washing at 90% B for 1 min, the column was equilibrated at 2% B for 24 min. A blank was run between each sample injection to check carryover and a bovine histone standard was run every 10 runs as a quality control. Proteins were identified and SILAC protein ratios were determined by the MassMatrix database search engine. All target peptides contained MS2 identifications in at least one data set. For data sets in which no corresponding MS2 spectra were obtained, the identity of the peptide was determined by the peptide's accurate mass, retention time, charge state distribution and pattern of modification.

Data Analysis

LC-MS/MS data files are converted to the mzXML format and then parsed by an in-house developed python script. The source code for this script is freely available at https://github.com/mafreitas/silac_peptide_calc. A detailed list of command line options and their usage is provided as a Supplemental Document. The mzXML filename, the zero-charge monoisotopic mass of the peptide, the mass of the label, the mass accuracy tolerance and noise rejection threshold are then provided as argument to python script that is used to calculate the normalization factor, cluster the peptide signals and determine the SILAC ratio. In detail, the program first scans the mzXML file to identify all matched pairs of signals across peptide isotopes (0–4) and charge states (+1–3). The SILAC ratios of resulting signal pairs are used to determine the normalization factor by finding the median, abundance weighted median, linear regression slope and robust linear regression slope. The algorithm can take additional command line options to restrict the charges when finding matched signal and limit the matched pair ratios to those that fall within a specified range when determining the normalization factor.

After finding the normalization factor, the program then iterates through all MS1 scan to find signal pairs that correspond to peptides on the target mass list. A sliding retention time window is used to cluster the peptide signals by time of elution. A cluster of peptide signals is comprised of signals pairs resulting from a peptide's isotope and charge distribution. The number of isotope peaks to include in the calculation and the max charge state can be specified in the algorithm. If the peptide sequence is provided, the algorithm will determine the max charges state automatically. A SILAC ratio is calculated for each distinct signal cluster by four different approaches. The first approach (referred to as total abundance) calculates the ratio by dividing the total heavy signal abundance by the light for the matched pairs in each cluster. Another approach takes the median matched pair ratio or the weighted median ratio (the weighted median adds additional weight to ratios calculated at the high abundance).

The ratio can also be determined by plotting the abundance of each matched pair and finding the slope using either a linear regression or a robust linear regression¹⁴. The linear regression is sensitive to the presence of outliers¹⁵. However, the robust linear regression,

also referred to as the Kendall-Theil robust regression, does not require normality of residuals and is not as strongly affected by outliers. The robust linear regression estimates the slope by comparing each data pair to all others in a pairwise fashion¹⁵. A data set of $n(x,y)$ pairs will result in $n(n-1)/2$ pairwise comparisons. For each of these comparisons, a slope is calculated. The median of all possible pairwise slopes is taken as the slope estimate b . If the data set contained more than 1,000,000 pairs of matched signals, the incomplete version of the method, called Theil's method, was used^{14b}. Slopes calculated by the incomplete robust linear regression have been proven to be a good alternative to the complete robust linear regression. The algorithm calculates the slope based on the abundances of matched pairs of peptides, which include the monoisotopic and up to 4 isotopic peaks. The default parameters used in this study are as follows: mass tolerance = 0.01 m/z , noise rejection threshold = 100,000 (arb units), sliding retention time window = 0.5 min, normalization ratio range = [0.5, 2], normalization charge state = 3. A detailed discussion of why these parameters were chosen is discussed below.

The final SILAC ratios for each peptide were first normalized using the robust linear regression normalization factor. All normalization factors for all samples are provided in Supplemental Table 1. The histone H3 K79 and H3 K56 posttranslational modification changes are presented relative to the unmodified proteoform to account for any potential difference in H3 expression. The overall mean and standard deviation were determined by combining the sample-based statistics for each measurement assuming non-overlapping sub samples. The raw ratios and their calculated sample-based statistics are provided in Supplemental Table 2. The final relative ratios for methylation are compared pair-wise to the wild-type using an unpaired t-test. Data are reported in all tables and figures with error bars equivalent to ± 2 standard deviations.

RESULTS

Determination of SILAC Normalization Factors

Stable Isotope Labeling of Amino acids in Cell culture was used for relative quantitation of peptide post-translational modifications. To accurately determine the SILAC ratio for a peptide, one must first determine a reasonable normalization factor for each sample to correct for mixing bias. This SILAC Peptide Ratio Calculator (SPRC) algorithm considers all signal pairs in each and every scan that could correspond to a light:heavy labeled peptide. Using these pairs the sample's overall normalization factor can be estimated. Because this algorithm identifies all possible pairs of signals, the potential for false positives is high, which can affect the accuracy of peptide ratios. The following sections describe the strategies used to calculate the normalization factor and strategies to reduce false peptide signal pairs.

The first step in normalization is to identify all pairs of light:heavy signals in each and every full scan of the LC-MS/MS data set. Light:heavy pairs are defined as two signals (within the same scan) that have a mass difference equal to the label mass divided by charge within the specified mass tolerance. High mass accuracy MS1 is recommended to reduce false positive signal pairings. The charge can be fixed or varied over a specified range. Thus for a nominal label mass of 8 and charge from +1 to +3, any two signals separated by 8, 4 or 2.6 m/z in the same full scan are considered a "matched pair". From the set of all matched pairs, four different methods are then used to estimate the normalization factor: 1) the median ratio of all matched pairs, 2) the abundance weighted median ratio, 3) the linear regression slope of the light vs heavy abundance of all "matched pairs" and 4) the robust linear regression slope. The weighted median ratio adds more weight to the ratios calculated at the high abundance. Because the algorithm is opportunistic when finding all matched pairs, the quality of the

normalization factor can be improved by applying constraints to limit the matched pairs used for normalization.

Effect of Noise Rejection on Normalization

The first constraint used to improve the quality of normalization factors is the noise rejection threshold. In order for a matched pair to be included in the calculation of the normalization factor, both the light and heavy signals must exceed a specified noise rejection threshold. Figure 1A shows the normalization factor calculated over a range of noise rejection thresholds for a sample comprised of a light WT yeast strain mixed with a heavy labeled yeast mutant strain (H4 K16Q). Figure 1B shows data for a different sample: light WT yeast mixed with heavy labeled yeast mutant (H2B K21Q). In these figures the light abundance was plotted vs. the heavy abundance. Each point represents a matched pair of signals and their resulting SILAC ratio. Log-log plots were also generated for visualization of data points at lower abundance (Supplemental Figure 1). For both data sets, the noise rejection threshold was varied from 10,000 to 1,000,000 (arb units). As shown in Figure 1A, the four normalization factors agreed when the noise rejection threshold was set to 1,000,000. As the noise rejection threshold was lowered, the normalization factors diverged. As expected, linear regression was the most sensitive to random “matched pairs” at lower noise rejection thresholds and thus not deemed appropriate for the determination of the normalization factor when noise is present. In contrast, the robust linear regression and median methods were more resistant to noise. For the data set plotted in Figure 1B, increasing the noise rejection threshold did not lead to a convergence of the estimated normalization factors. The source of the divergence with this sample was the presence of polymer that led to matched pairs not of SILAC peptide origin, which confounded calculation of the normalization factor.

Effect of Ratio and Charge Filtering on Normalization

To address the confounding effect of false peptide match pairing and improve the robustness of the approach two normalization constraints were evaluated: 1) SILAC ratio filtering and 2) matched pair charge filtering. The effect of these constraints on the normalization factors is depicted in Figure 2 (log-log abundance scatter plots for these comparisons are also provided as Supplemental Figure 2). Ratio filters are a commonly used constraint to improve the estimation of normalization factors for microarrays¹⁶. In a similar manner, the SILAC ratio filter constrains the range of matched pair ratios used for the calculation of normalization factors. Constraining the ratios used for normalization has the added benefit of emphasizing the unchanging peptide population and limiting the acceptable range of error from sample mixing. The first column of Figure 2A–D shows the effect ratio filtering has on the normalization factors. For this example, the matched pair ratios used to estimate the normalization factor were varied from no ratio filter, ± 8 fold, ± 4 fold and ± 2 fold. As the allowed ratios were increasingly constrained, the normalization factors converged even in the presence of matched pairs from polymer contaminants (an example of a false-positive “matched pair” contributed by polymer contaminant is provided in Supplemental Figure 3).

Other algorithms could also benefit from the use of ratio filters for normalization. The data in Figure 3 show the normalization factors determined by the use of MaxQuant for peptide identifications in the WT:WT sample with low polymer contamination (Figure 3A) and the WT:H2B K21Q mutant suffering from high polymer contamination (Figure 3E) as shown above. MaxQuant uses a normalization function based on peptide abundance and number of labels (Figures 3B–D & 3F–H). The MaxQuant WT:WT normalization shows good agreement across the abundances and number of labels (Figures 3B–D). Furthermore, the MaxQuant normalization function agrees well with the normalization factor calculated by SPRC (Log₂ normalization factor = 0.12) as shown by the solid horizontal line. However, outliers also affect MaxQuant’s normalization. For example the polymer contamination in

the WT:H2B K21Q sample caused a dramatic shift in the MaxQuant normalization at higher abundance resulting in inaccurate peptide ratios (Figures 3F–H). The agreement between MaxQuant and SPRC normalization factors at lower abundance (Log_2 normalization factors; $\text{SPRC} = -0.53$; median 1st quartile MaxQuant = -0.66) was very good.

Another approach that can help eliminate non-peptide matched pairs is charge filtering. Chemical and electrical noises are two sources of random matched pairs. Polymers are a common low-level contaminant in LC-MS that can lead to random signal pairing. The use of a charge filter was evaluated as a constraint to reduce chemical noise. The charge filter constrains the charge states used in identifying the matched pair. For example, a label of 8 Da and +3 charge filter would only find matched pairs of signals separated by 2.6 m/z. Non-peptide chemical noise in samples are dominated by singly charged ions and thus have integer m/z separations. In Figure 2, each column represents the use of a different charge filter. As expected the non-integer +2 and +3 charge filters were more effective at eliminating matched pairs due to contaminants than the +1 charge filter. Despite this improvement, the charge filter alone was not as effective at achieving a consensus normalization factor as the ratio filter. When the +2 or +3 charge filter was combined with the ± 2 fold ratio filter, consensus between the normalization factors was achieved.

The robustness of the combined charge and ratio filtering is best demonstrated when we significantly increase the likelihood of false signals by lowering the noise rejection threshold. The box plots shown in Figure 4 illustrate this quality of normalization across all constraints and the corresponding normalization factors are provided in Table 1. This comparison shows that the normalization factors calculated by the different means converge as the noise rejection threshold is increased. The ± 2 fold ratio filter is the most effective constraint and the combination of ratio and charge filtering at the higher noise rejection threshold results in excellent agreement across all normalization schemes.

Comparison of Normalization from Matched pairs vs. Identified Peptides

The approach described above was compared to more conventional normalization based on actual peptides identified by database search. Three replicates of a light WT: heavy WT sample and three replicates of a light WT: heavy H2B K21Q yeast mutant were analyzed by the MassMatrix database search software using the search parameters described in the methods section. The SILAC ratios for each identified peptide spectral match were in turn generated by MassMatrix's post-hoc SILAC quantitation software. Scatter plots of the light vs. heavy abundances for the matched pair approach compared to the database search identified peptide are shown in Figure 5. The average difference and standard deviation of the difference between predicted normalization factors is summarized in Table 2. This data shows excellent agreement between the matched pair generated normalization factors and normalization factors obtained from database search identified peptide spectral matches. This comparison demonstrated that the matched-pair approach can be used without prior database search to estimate the global normalization factor for a given LC-MS/MS experiment. Normalization factors by all schemes are provided in Supplemental Table 1.

Peptide Clustering

Once the global normalization factor is determined, the SILAC ratios for target peptides are determined. By providing a list of target masses, the algorithm will attempt to identify peptide signal clusters and calculate their SILAC ratios regardless of whether MS2 spectra were obtained and the peptide identified. This approach is often necessary for low abundant PTMs that may not be identified in every sample, but still have sufficient MS1 signal to determine an accurate ratio. It can also be used to mine SILAC data sets for novel peptides that can be verified in subsequent MS2 experiments using mass inclusion lists. To calculate

a peptide's SILAC ratio, the algorithm iterates through every MS1 scan, finding signals that correspond to the target peptide's m/z . The program identifies target peptide matched pairs for each isotope (0–4) and charge state (+1 to +N, where N = number of basic side chains +1). The target peptide's SILAC ratio is determined as described above for the global normalization factors from the matched pairs. To reduce false positive matches, three constraints are applied: noise rejection threshold, mass accuracy and retention time window.

Mass accuracy and the noise rejection threshold are two important parameters for elimination of false positive signal matches. An example of this effect is shown in Figure 6. In this figure the $\log_2$ ratio vs retention time is plotted for every match pair corresponding to the histone H3 peptide fragment EIAQDF⁷⁹KTDLR (mass = 1335.6903 Da) from the WT:H2B K21Q sample. This data clearly shows the number and retention time of matched signals is strongly dependent on the mass accuracy and noise rejection threshold. When the mass accuracy was set at 0.1 Da and the noise rejection threshold at 10,000 (arb units), random “matched pairs” were present throughout the entire data set. As the noise rejection threshold (y) increased from 10,000 to 100,000 (arb units), the number of random matches was reduced. By improving the mass accuracy (t) from 0.1 to 0.01 Da, the number of matched pairs was limited to two distinct clusters eluting at different times. This example shows that high mass accuracy and noise rejection threshold are very effective at eliminating random matched pairs.

The algorithm also used a sliding retention time window to separate peptide signals with different elution profiles. The sliding retention time window was specified by the user and corresponded to the maximum retention time between signals in a cluster. If this time was exceeded, matched pairs were assigned to a separate cluster of signals. The SILAC ratio for each cluster of signals was calculated separately. Figure 7 shows the effect the sliding window had on clustering signals for the target histone H3 peptide EIAQDF⁷⁹KTDLR. With the sliding window set at 1.0 min, signals for the unmodified H3 K79 peptide fragment and another set of matched signal pairs clustered together over the time window 41–43 min. The larger retention time window was, therefore, not able to distinguish between these two closely eluting signal clusters. When the sliding window was decreased to 0.5 min, the matched pairs were recognized as two different sets of signals. Peptide ratios were then calculated for clusters based on their set of matched pairs using the approaches described above. The MS/MS spectra for the cluster eluting at 41–41.5 min corresponded to the target H3 K79 peptide. To aid the interpretation and validation of the peptide ratios, the algorithm outputs a scatter plot, table of all light vs. heavy abundances, and the summary of ratios calculated by each method (w and w/o normalization).

Measuring Crosstalk Between the H4 N-terminal Tail and H3 K79/K56

The preceding paragraphs describe the approach used to generate SILAC peptide ratios. The importance of normalization and cluster of signals greatly improves the accuracy of the SILAC ratio. The real power of this approach lies in its ability to monitor changes in the abundance of modified peptides. In order to demonstrate this approach's utility we measured how histone modification sites in the H4 NH₂-terminal tail would affect H3 K79 methylation and H3 K56 acetylation. Four histone H4 double mutants were generated (H4 K5R/K12R; H4 K5Q/K12Q; H4 K8R/K16R; H4 K8Q/K16Q) and grown in SILAC media to introduce heavy lysine. The samples were mixed with an equal weight of wild type yeast grown in unlabeled media. The histones were extracted, digested with Arg-C and subjected to LC-MS/MS. Each sample was then analyzed by the SILAC peptide calculator to find ratios for H3 K79 methylation and H3 K56 acetylation. To compare the peptide level across samples, the raw ratios were normalized using the robust linear regression factor with +3 charge and ± 2 fold ratio filtering. The normalized SILAC peptide ratios for H3 K79 are shown in Figure 8A and listed in Table 3. Similarly, the normalized SILAC peptide ratios

for H3 K56 are shown in Figure 8B and listed in Table 4. Error bars correspond to ± 2 sample-based standard deviations based on the size and mean of each sample as described in the Experimental Procedures (an example calculation is provided in the supplemental tables).

This analysis identified residues on the NH₂-terminal tail of histone H4 that affect the pattern of methylation of histone H3 lysine 79. H4 lysine 5 and lysine 12 as well as H4 lysine 8 and lysine 16 were mutated in groups to arginine or glutamine to mimic the constitutively unacetylated and acetylated states, respectively. Mutating both H4 lysine 5 and 12 to glutamine led to an increase in mono-methylation and a decrease in tri-methylation of H3 K79. Mutating both H4 lysine 8 and 16 to glutamine also resulted in an increase in mono-methylation and di-methylation. However, double glutamine substitutions on H4 lysine 5 and 12 or H4 lysine 8 and 16 have no significant effect on the level of H3 K79 methylation. The analysis was also used to identify the effect of double mutations on the NH₂-terminal tail of histone H4 on the level of H3 K56 acetylation. Mutating H4 lysine 8 and 16 to glutamine led to a decreased level of H3 K56 acetylation while mutating H4 lysine 8 and 16 to arginine had no effect. Arginine or glutamine substitution on both H4 lysine 5 and 12 did not influence the level of H3 K56 acetylation.

Taken together, the H4 K8Q/K16Q mutant increased the level of H3 K79 mono- and di-methylation and decreased the level of H3 K56 acetylation. According to our previous work ¹⁷, the glutamine substitution of H4 lysine 16 led to an increased level of H3 K56 acetylation while mutating H4 lysine 8 to glutamine resulted in no change in the level of H3 K56 acetylation. It has been proposed that the region of the H4 tail between lysine 8 and lysine 16 binds along a channel that is long enough to accommodate 6–7 amino acids. In addition, lysine 8 and lysine 16 are found to lie opposite two acidic patches at the ends of the channel, theoretically engaging in electrostatic interactions that hold the tail in place ¹⁸. Therefore, H4 lysine 8 and 16 play a significant role in stabilizing the histone tail-DNA interaction on the surface. Abrogating the positive charge of lysine 8 and lysine 16 by glutamine substitution may alter the binding between histone and DNA and further affect the binding of H3 K56 to histone acetyl transferase Rtt109 at the entry/exit point of the nucleosomal DNA.

CONCLUSION

A novel algorithm is presented that improves calculation of SILAC ratios for peptide post-translational modifications. The approach uses MS1 spectra to identify signal pairs that correspond to light:heavy peptides. Using the aggregate matched pair data, a normalization factor can be estimated. Robustness of normalization factors is improved by applying constraints to signal abundance, signal pairing and range of ratios. For this study a ± 2 fold ratio and +3 charge constraint provided robust estimation of normalization factors in excellent agreement with factors estimated from database search identified peptides. The peptide ratios for specified peptide targets can then be determined by clustering peptides based on mass retention time. There is no requirement to have identified a peptide prior to analysis, although it is recommended in order to verify identity. The approach could also be used in combination with algorithms that use feature detection to facilitate the discovery of novel unidentified peptide species. The utility of the approach for quantitation of peptide modifications was demonstrated by examining crosstalk between the NH₂-terminal tail of histone H4 and H3 K79/K56.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The study was funded by grants from the National Institutes of Health (R21 DK082634 to MAF and MRP, R01 CA107106 to MAF and R01 GM62970 to MRP) and support from the Ohio State University.

References

1. (a) Mann M. Functional and quantitative proteomics using SILAC. *Nat Rev Mol Cell Biol.* 2006; 7(12):952–958. [PubMed: 17139335] (b) Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M. Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. *Molecular & Cellular Proteomics.* 2002; 1(5):376–386. [PubMed: 12118079]
2. Xu H, Freitas MA. MassMatrix: A database search program for rapid characterization of proteins and peptides from tandem mass spectrometry data. *PROTEOMICS.* 2009; 9(6):1548–1555. [PubMed: 19235167]
3. Hoopmann MR, Chavez JD, Bruce JE. SILACTor: Software To Enable Dynamic SILAC Studies. *Analytical Chemistry.* 2011; 83(22):8403–8410. [PubMed: 21954881]
4. Park SK, Venable JD, Xu T, Yates JR. A quantitative analysis software tool for mass spectrometry-based proteomics. *Nat Meth.* 2008; 5(4):319–322.
5. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotech.* 2008; 26(12):1367–1372.
6. Huang X, Tolmachev AV, Shen Y, Liu M, Huang L, Zhang Z, Anderson GA, Smith RD, Chan WC, Hinrichs SH, Fu K, Ding SJ. UNiquant, a Program for Quantitative Proteomics Analysis Using Stable Isotope Labeling. *Journal of Proteome Research.* 2010; 10(3):1228–1237. [PubMed: 21158445]
7. Han DK, Eng J, Zhou H, Aebersold R. Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. *Nat Biotech.* 2001; 19(10):946–951.
8. Li, X-j; Zhang, H.; Ranish, JA.; Aebersold, R. Automated Statistical Analysis of Protein Abundance Ratios from Data Generated by Stable-Isotope Dilution and Tandem Mass Spectrometry. *Analytical Chemistry.* 2003; 75(23):6648–6657. [PubMed: 14640741]
9. Mortensen P, Gouw JW, Olsen JV, Ong S-E, Rigbolt KT, Bunkenborg J, Cox, Foster LJ, Heck AJR, Blagoev B, Andersen JS, Mann M. MSQuant, an Open Source Platform for Mass Spectrometry-Based Quantitative Proteomics. *Journal of Proteome Research.* 2009; 9(1):393–403. [PubMed: 19888749]
10. Parker SJ, Halligan BD, Greene AS. Quantitative analysis of SILAC data sets using spectral counting. *PROTEOMICS.* 2010; 10(7):1408–1415. [PubMed: 20104619]
11. Kelly TJ, Qin S, Gottschling DE, Parthun MR. Type B histone acetyltransferase Hat1p participates in telomeric silencing. *Mol Cell Biol.* 2000; 20(19):7051–8. [PubMed: 10982821]
12. Ahn SH, Henderson KA, Keeney S, Allis CD. H2B (Ser10) phosphorylation is induced during apoptosis and meiosis in *S. cerevisiae*. *Cell Cycle.* 2005; 4(6):780–3. [PubMed: 15970663]
13. Knapp AR, Ren C, Su X, Lucas DM, Byrd JC, Freitas MA, Parthun MR. Quantitative profiling of histone post-translational modifications by stable isotope labeling. *Methods.* 2007; 41(3):312–9. [PubMed: 17309842]
14. (a) Helsel, DR.; Hirsch, RM. Statistical methods in water resources. Elsevier; 1993. (b) Glaister P. Robust Linear Regression Using Theil's Method. *Journal of Chemical Education.* 2005; 82(10): 1472.
15. Hirsch, DRHaRM. USGS Statistical Methods in Water Resources. US Geological Survey, Techniques of Water-Resources Investigations. 2002; Chapter A3 Book 4.
16. (a) Padmanabhan B, Kataoka K, Umehara T, Adachi N, Yokoyama S, Horikoshi M. Structural similarity between histone chaperone Cialp/Asf1p and DNA-binding protein NF-kappaB. *J Biochem.* 2005; 138(6):821–9. [PubMed: 16428312] (b) Patterson TA, Lobenhofer EK, Fulmer-Smentek SB, Collins PJ, Chu TM, Bao W, Fang H, Kawasaki ES, Hager J, Tikhonova IR, Walker SJ, Zhang L, Hurban P, de Longueville F, Fuscoe JC, Tong W, Shi L, Wolfinger RD. Performance comparison of one-color and two-color platforms within the Microarray Quality Control (MAQC)

project. *Nat Biotech.* 2006; 24(9):1140–1150.(c) Peart MJ, Smyth GK, van Laar RK, Bowtell DD, Richon VM, Marks PA, Holloway AJ, Johnstone RW. Identification and functional significance of genes regulated by structurally different histone deacetylase inhibitors. *Proceedings of the National Academy of Sciences of the United States of America.* 2005; 102(10):3697–3702. [PubMed: 15738394]

17. Guan X, Rastogi N, Parthun MR, Freitas MA. Discovery of Histone Modification Crosstalk Networks by SILAC Mass Spectrometry. *Molecular & Cellular Proteomics.* 2013
18. (a) Dutnall RN, Tafrov ST, Sternglanz R, Ramakrishnan V. Structure of the histone acetyltransferase Hat1: a paradigm for the GCN5-related N-acetyltransferase superfamily. *Cell.* 1998; 94(4):427–38. [PubMed: 9727486] (b) Makowski AM, Dutnall RN, Annunziato AT. Effects of Acetylation of Histone H4 at Lysines 8 and 16 on Activity of the Hat1 Histone Acetyltransferase. *Journal of Biological Chemistry.* 2001; 276(47):43499–43502. [PubMed: 11585814]

Figure 1.

Effect of low abundant noise on the determination of global normalization factors. Each point in the scatter plot represents a SILAC ratio for a pair of matched signal in the LC-MS/MS data of A) WT:H4 K16Q sample and B) WT:H2B K21Q sample. Normalization factors were calculated for three noise rejection thresholds. The noise rejection threshold, denoted as “y” are: $y=10,000$, $y=100,000$; $y=1,000,000$ (arb Units).

Figure 2.

Effect of ratio and charge filtering on global normalization. Each point in the scatter plots represents a SILAC ratio from a light:heavy matched signal (LC-MS/MS data from the WT:H2B K21Q sample). The global normalization factors were determined after applying a ratio and/or charge state filter. In the rows A–D, a noise rejection threshold of 100,000 was applied prior to A) No ratio filter B) Filter ratios 8-fold change C) Filter ratios 4-fold change D) Filter ratios 2-fold change. In each row, a charge state filter was further applied, starting from left to right, no charge filter, +1 charge state filter, +2 charge state filter and +3 charge state filter.

Figure 3.

MaxQuant determined normalization factors demonstrating the effect of contaminant species on normalization. The WT:WT (A–D) and WT:H2B K21Q (E–H). The panels A and E are scatter plots of the light vs. heavy abundance for all identified features. Panel A is a clean sample with little contamination. Panel E shows how polymer contamination can lead to false identification of SILAC pairs. Normalization factors for both samples are shown in panels B–D and F–H. The horizontal lines represent the robust linear regression normalization factor for each sample. The effect of falsely identified SILAC pairs causes substantial changes in the MaxQuant normalization function at higher abundance (F–H).

Figure 4.

Boxplots summarizing the distribution of $\pm\log_2$ SILAC ratios for matched signals in the LC-MS/MS data from the WT:H2B K21Q sample. The matched signals were clustered under different noise rejection threshold, from 10,000, 100,000 to 1,000,000. A +3 charge state filter and a ratio filter were applied at each threshold. As the noise rejection threshold increased, the ratios converge. More importantly, with the ratio and charge state filter applied, the ratios of matched signals at the lower noise rejection threshold (10,000) converge similarly to those at the higher.

Figure 5. Comparison of normalization factors obtained by global signal normalization and peptides identified by database search. Plots are shown for three replicates of two SILAC samples (WT:WT and WT:H2B K21Q). 5A and 5C show the ratio calculation for global signal normalization. 5B and 5D show the global ratio calculation from the database search identified peptides. 2-fold ratio and +3 charge filter were used for global signal normalization.

Figure 6.

Effect of mass accuracy and noise rejection threshold on peptide clustering of H3 EIAQDF⁷⁹KTDLR in the WT: H2B K21Q sample. As the noise rejection threshold (y) increased from 10,000 to 100,000 (arb units), majority of the noise were excluded. Furthermore, when the mass accuracy (t) improved from 0.1 to 0.01 Da, the number of signal cluster was limited to two. Notably, one is the peptide cluster of peptide and the other is noise signal.

Figure 7.

Effect of the width of the sliding retention time window on the ratio of peptide H3 EIAQDF⁷⁹KTDLR in the WT:H2B K21Q sample. A) The large retention time window of 1.0 min collected two peptide clusters eluting closely in the same window. The inner small figure shows the retention time (RT) of these two peptide clusters. B&C) A smaller window of 0.5 min effectively distinguished these two peptides into individual clusters, yielding more accurate SILAC ratios.

Figure 8.

A) SILAC ratios for the H3 K79 peptide EIAQDF⁷⁹K(Me₀₋₃)TDLR and B) SILAC ratios for the H3 K56 peptide FQ⁵⁶K(Ace₀₋₁)STELLIR across four yeast double mutants. WT is from plasmid pMP3 that carries both wild type histone H3 and H4 genes (HHT2-HHF2). To compare the peptide level between mutants, raw H/L ratios of each peptide are normalized using the filtered robust linear regression normalization factor. Error bars correspond to 2 folds of the sample based standard deviation (see Experimental Procedures and Supplemental tables for description and example).

Normalization factor of matched pairs generated under no filter, +3 charge and/or 2-fold ratio filter for the WT: H2B K21Q data set. The matched signals were clustered under different noise rejection threshold, from 10,000, 100,000 to 1,000,000.

Table 1

Threshold		10,000			100,000			1,000,000					
Filter		n/a	+3	ratio	+3 & ratio	n/a	+3	ratio	+3& ratio	n/a	+3	ratio	+3& ratio
RLR		1.99	1.93	1.03	0.99	2.85	3.07	0.86	0.80	1.47	0.87	0.79	0.75
LR		0.27	0.43	0.84	0.78	0.35	0.67	0.83	0.78	0.53	0.67	0.81	0.78
Median		1.29	1.23	1.03	1.02	2.15	2.06	0.93	0.88	1.34	0.90	0.81	0.77
AW Median		0.56	0.69	0.81	0.79	0.80	0.79	0.80	0.77	0.81	0.79	0.79	0.77
Average		1.03	1.07	0.93	0.90	1.54	1.65	0.86	0.81	1.04	0.81	0.80	0.77
Std. Dev		0.77	0.66	0.12	0.13	1.16	1.14	0.06	0.05	0.44	0.10	0.01	0.01

RLR = Robust Linear Regression, LR = Linear Regression, AW Median = Abundance Weighted Median

Table 2

Average difference and standard deviation of the difference between the global normalization factor and normalization factor based on peptides identified by database search.

<u>Global normalization vs. Normalization based on peptide identified only</u>		
	Mean of Difference	Std of Difference
RLR Regression	−0.055	0.217
LR Regression	0.070	0.152
Median	−0.048	0.173
Abundance Weighted Median	−0.045	0.053

SILAC ratios for the H3 K79 peptide EIAQDF⁷⁹K(Me₀₋₃)TDLR across four yeast double mutants. Each ratio is the average of three biological replicates. A t-test was performed between the wild type and each mutant. Peptide ratios with a p-value 0.05 are reported as significant.

Table 3

	H3 K79Me ₀			H3 K79Me ₁			H3 K79Me ₂			H3 K79Me ₃		
	average	Std. Dev		average	Std. Dev	p-value	average	Std. Dev	p-value	average	Std. Dev	p-value
WT1	0.89	0.09		1.00	0.06	-	0.98	0.12	-	0.91	0.13	-
H4 K5,12R	1.03	0.22		1.37	0.12	0.40	1.57	0.07	0.20	0.76	0.08	0.81
H4 K5,12Q	0.99	0.11		1.66	0.16	0.022	1.43	0.10	0.11	0.60	0.07	0.017
H4 K8,16R	0.89	0.18		0.92	0.12	0.74	0.99	0.14	0.85	1.21	0.19	0.068
H4 K8,16Q	0.92	0.17		2.18	0.26	0.020	1.61	0.23	0.019	0.65	0.08	0.43

SILAC ratios for the H3 K56 peptide FQ⁵⁶K(Ace₀₋₁)STELLIR across four yeast double mutants. Each ratio is the average of three biological replicates. A t-test was performed between the wild type and each mutant. Peptide ratios with a p-value 0.05 are reported as significant.

Table 4

	H3 K56Ace ₀		H3 K56Ace ₁		p-value
	average	Std. Dev	average	Std. Dev	
WT1	0.91	0.16	0.95	0.16	-
H4 K5,12R	0.99	0.09	1.09	0.15	0.77
H4 K5,12Q	0.92	0.07	1.06	0.10	0.22
H4 K8,16R	1.13	0.13	1.01	0.19	0.61
H4 K8,16Q	1.14	0.07	0.74	0.06	0.0042