Supplementary Figures

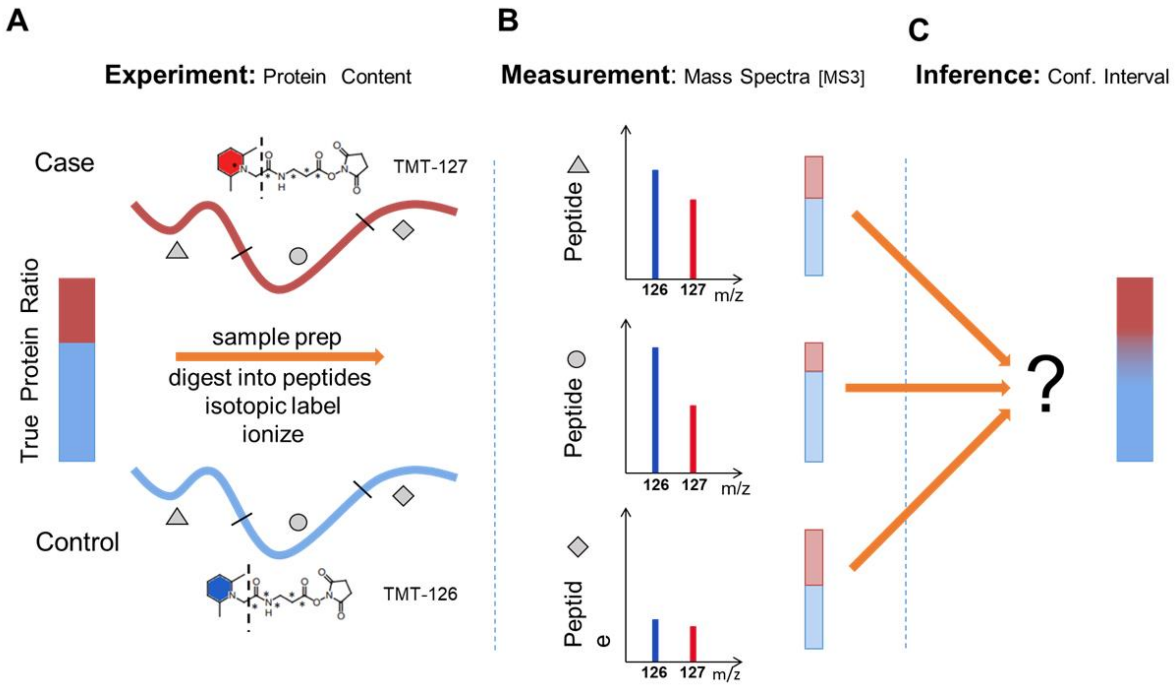

Figure S1. Overview of the challenge to integrate multiplexed proteomics measurements into confidence intervals. A) Multiplexed proteomics allows the comparison of protein abundance among multiple conditions. For simplicity, only two conditions are shown. A protein with true protein ratio (red-blue bar) is digested into peptides. The peptides are labeled with isobaric tags e.g. TMT to encode the different conditions. The peptides are combined and ionized and injected into the mass spectrometer. B) The different peptides derived from each protein results in separate spectra, which allow quantification using accurate multiplexed proteomics methods, e.g. MultiNotch MS3 or TMTc+. The relative intensity of the peptide signal can be used to quantify protein abundance. Each spectrum contains the information of relative abundance and signal i.e. the number of ions by which this ratio was measured. C) The challenge is how the information of different peptide intensity and agreement/disagreement between measured peptide ratios can be integrated to accurately reflect confidence intervals for the underlying true protein ratio.

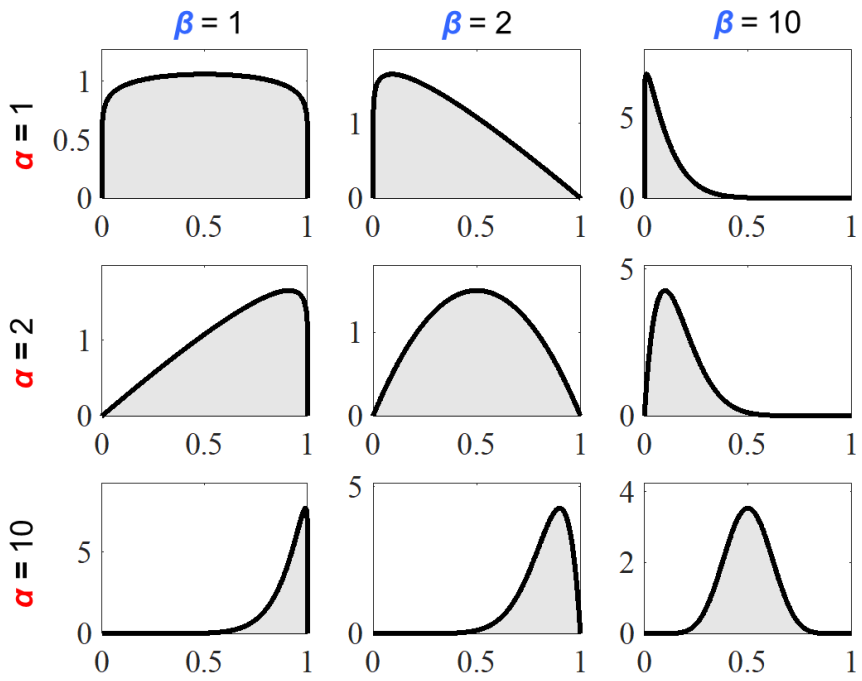

Figure S2. **The versatile beta distribution.** Illustration of various shapes of the PDF of a binomial fraction, that can be expressed by a Beta distribution with two respective parameters as illustrated. The diagonal corresponds to symmetric distributions with the mode at .5 but different confidence. Above the diagonal distributions skewed towards 0, below the diagonal – towards 1. Any fraction between zero and one can be therefore represented with various degree of confidence, using two parameters *α* and *β*.

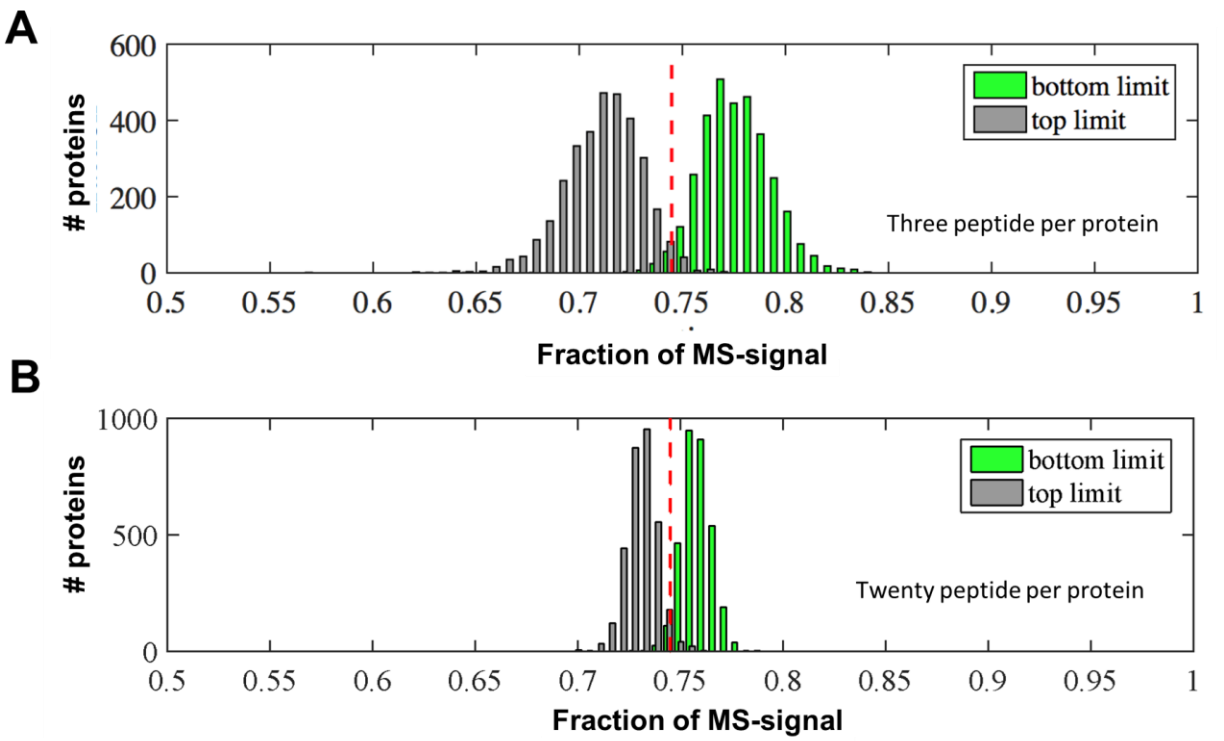

Figure S3: Summing up peptide signal works for the artificial protein case. Artificial proteins are generated by summing from several peptides from a sample in which all peptide ratios are identical. **A)** 3 peptides were selected per artificial protein. The correct mixing ratio for this sample is 0.745. The 95% confidence intervals are below bottom 2.69 % and above top limits 3.15 % of the time. **B)** 20 peptides were assigned per artificial protein. Bottom limit of 95% confidence intervals is 2.29 % above true ratio, and top limit is 2.47% of the time below true ratio.

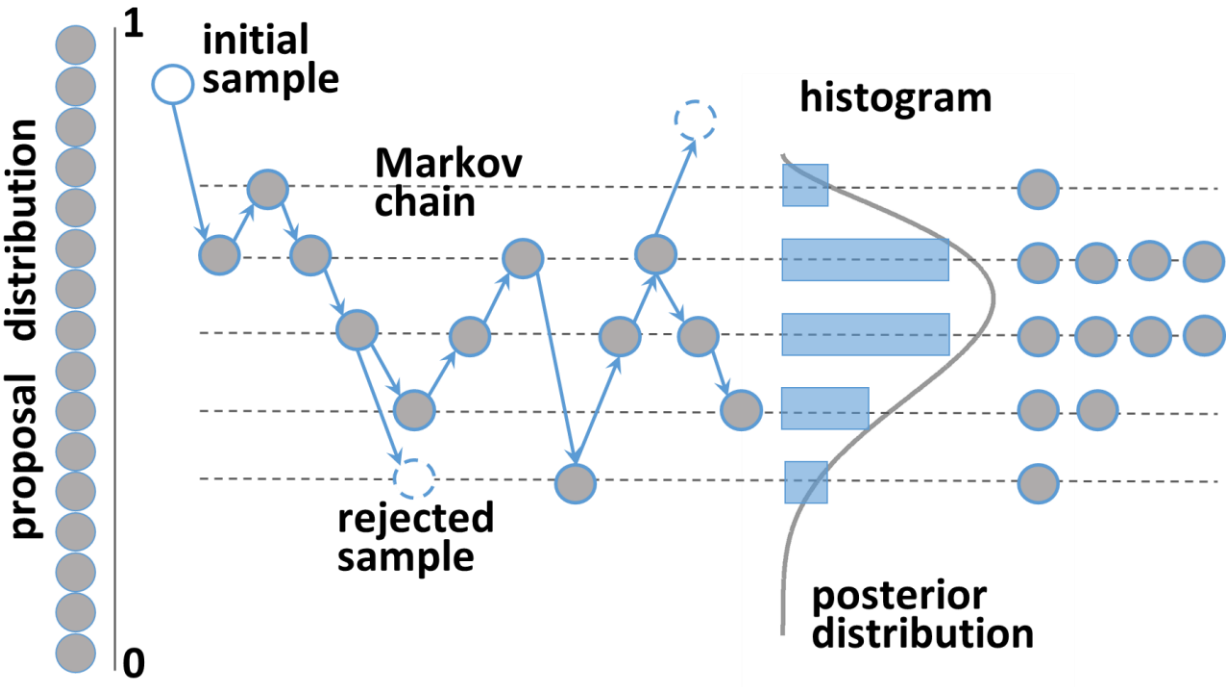

Figure S4: Intuitive explanation for the Monte Carlo Markov Chain method. We construct a large set of samples representing the distribution as a histogram. The process starts by taking samples from a proposal distribution (e.g. uniform) and constructing a chain of sequential samples selecting the next sample based on the likelihood of the current sample (and ignoring the rest of the history, therefore chain has the "Markov property").

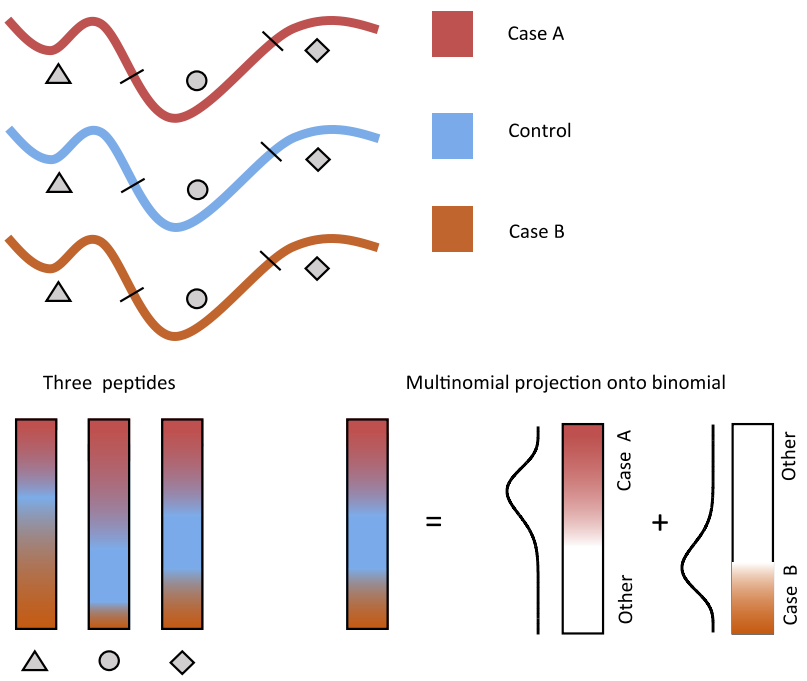

Figure S5. An illustration of joint multi-dimensional uncertainty representation for the case of a 3-plex projected into 2-plex.

Supplementary Material

Mathematical foundations

The functional form of this dependency for a coin toss represented by a binomial distribution is as follows. The mean of the binomial distribution with parameters **n** (number of tosses) and **p** (probability of success) is $m = np$ and the standard deviation is $s = \sqrt{np(1-p)}$ thus the coefficient of variation is $Cv = \sqrt{1-p}/\sqrt{np}$.

We fit a single parameter ***m*** as a multiplier to an S/N value ***s*** where ***n=ms*** to the binned data as illustrated in Figure 2B. The data of 10534 points is binned by 500 data points into 21 bins, and Cv is calculated for each bin the MATLAB's Nonlinear-Least-Squares fitting method is used which naturally produces not only most likely estimate but also the confidence intervals on the conversion parameter.

The confidence intervals are obtained from the Beta distribution by inverting the distribution in the following way. The likelihood function of the parameter **q** (Binomial probability of success) is a Beta distribution **q** ~ Beta (***α***, ***β***), where PDF of Beta is $\frac{x^{a-1}(1-x)^{b-1}}{B(a,b)}$

where $B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ defined via the Gamma function.

If S/N values for two channels are $\mathbf{k_1}$ and $\mathbf{k_2}$ for an instrument with a multiplier **m** we have a Beta function with parameters $\boldsymbol{\alpha = k_1 m}$ and $\boldsymbol{\beta = k_2 m}$. The values of **q** where the cumulative distribution function (CDF) of this function get to **(1-ξ)/2** and **(1+ξ)/2** respectively are the limits of the **100*ξ** percent confidence interval.

The multi-peptide case is modeled using the Bayesian inference language Stan as follows: the fraction in two channels for **i**-th peptide is q_i ~ Beta (***α***, ***β***), for peptides 1..K, then the number of events observed in the first channel between two channels which have a total of N_i events is n_i ~ Binom(q_i, N_i) where Binom(n | q,N) $\propto q^n (1-q)^{N-n}$

Implementation details

In addition to the peptide data file, which is expected in TSV or CSV format, our implementation of BACIQ takes the following parameters:

- the S/N conversion factor;
- the confidence level;
- the channels identity.

MCMC remarks

We illustrate the general intuition behind Monte-Carlo Markov Chain method in Figure S4. The key point is that in order to approximate the distribution which we cannot compute in a closed form, we construct a large set of samples representing the distribution as a histogram. The process starts by taking samples from a proposal distribution (e.g. uniform) and constructing a chain of sequential samples selecting the next sample based on the likelihood of the current sample (and ignoring the rest of the history, therefore chain has the "Markov property"). In the context of protein ratios estimation, we are essentially asking "what is the probability for the observed peptide data to have come from a situation where the protein fraction is **q** ?". If the probability is high, the sample will be accepted

and the next sample will be drawn not far from it, if it is low, the sample will likely get rejected.