Stoichiometry determination of protein complexes by label-free quantitative mass spectrometry-based proteomics (Prot61)

Arne H. Smits and Michiel Vermeulen

Molecular Cancer Research
Division of Biomedical Genetics
University Medical Center Utrecht
Universiteitsweg 100, 3584 CG Utrecht, The Netherlands.

E-mail feedback to: A.Smits-4@umcutrecht.nl

Introduction

Most proteins assemble into multi-subunit complexes to perform their cellular functions. To understand the biological role of a protein of interest, it is therefore important to identify its protein-protein interactions (PPIs). Besides the qualitative identification of PPIs, the protein complex architecture is also of major importance. By distinguishing

core-subunits from substoichiometric complex subunits, key subunits and their intrinsic binding domains, enzymatic activities and/or regulatory functions can be identified.

Recent developments in quantitative mass spectrometry (qMS) have made it possible to screen for these protein-protein interactions in a comprehensive and unbiased manner (Vermeulen et al, 2008). Most recent qMS methods are based on label-free quantification, which do not rely on isotope labeling and are therefore ideally suited for PPI identification in any kind of tissue or cell type (Hubner et al, 2010). To determine the protein complex stoichiometry, information on the abundance of interactors needs to be gained. This used to be accomplished by spike-in of isotope labeled reference peptides, a laborious and expensive method. However, recently new computational methods have been developed that are able to approximate the abundance from qMS intensity, one of which is Intensity Based Absolute Quantification (iBAQ) (Schwanhausser et al, 2011).

We recently combined the label-free quantitative PPI identification method QUBIC with the iBAQ algorithm, which enables easy and fast abundance determination of the identified PPIs. Thereby, we are able to determine the stoichiometry of protein complex subunits and can easily distinguish core-subunits from substoichiometric, transient interactions (Smits et al, 2013).

In our approach, we make use of single-step affinity purification of GFP-tagged proteins of interest and an untagged control sample, followed by on-bead trypsin digestion and LC-MS/MS. In principle, our workflow can be adapted to any kind of single-step affinity purification. GFP-tagging is our method of choice, as tagging is optimized using bacterial artificial chromosomes (BACs) and tagged exogenous expression is near endogenous levels (Poser et al, 2008). Besides stable transfection using BACs, GFP-tagged proteins can also be expressed by other means, such as transient transfection. Additionally, the use of GFP-trap beads (Chromotek) enables highly efficient affinity enrichments without heavy and light chain contaminants (Rothbauer et al, 2008) (and EpiGeneSys protocol PROT44).

Once the GFP fusion protein is expressed in tissue or cell type of choice, the cells are collected and nuclear extract is obtained as described before (Dignam et al, 1983). This nuclear extract and that of the untagged control are applied to GFP-trap affinity purification. In order to perform statistics on the label-free quantification, GFP and control purifications are done in triplicate. After extensive washing, the remaining proteins are digested using trypsin, as described (Hubner & Mann, 2011). Next, the peptides are purified on StAGE-tips (Rappsilber et al, 2007) and applied to LC-MS/MS. Raw data are analyzed by MaxQuant (Cox & Mann, 2008) and PPI are identified using label-free quantification (Hubner et al, 2010). For each of the identified PPIs, we correct for background binding by subtracting the obtained iBAQ value in the control from the obtained iBAQ value in the GFP sample. These corrected values of abundance are scaled to one protein, most commonly the bait protein, which reveals the stoichiometries of interacting proteins relative to the protein scaled to one.

In this approach, iBAQ is used to estimate the abundance of identified interactors,based on the mass spec intensity of the tryptic peptides. Some peptides, however, can belong to multiple proteins, as these proteins are (partially) homologous. These peptides are assigned to the protein with the most unique peptides in the MaxQuant software, however, this leads to overestimation of the stoichiometry of this particular protein and

underestimation of the stoichiometry for the other proteins that give rise to this peptide. To circumvent this problem, we collapse the stoichiometries of proteins with shared peptides into a single value. Thereby, we lose resolution, as we cannot distinguish the stoichiometry of these proteins, yet we can be confident of the combined stoichiometry value.

Procedure

The following procedure focuses on the computation analysis of the data obtained from mass spectrometry. It follows upon immunoprecipitation of the GFP-fusion protein and the control, as described in (Hubner et al, 2010; Rothbauer et al, 2008; Smits et al, 2013) and EpiGeniSys protocol PROT44.

1. GFP affinity purification of the GFP-fusion sample as well as an untagged control is done in triplicate, as described in (Hubner et al, 2010; Rothbauer et al, 2008; Smits et al, 2013). A detailed protocol can be found in EpiGeneSys protocol PROT44.

2. After GFP affinity purification and washes, the proteins are on-bead digested with trypsin (see comments 1,2).
   2.1. Wash the beads twice with 1 mL PBS (to get rid of any detergent!)
   2.2. Aspirate the supernatant
   2.3. Add 50 µL denaturation buffer
   2.4. Shake at room temperature for 20'
   2.5. Add iodoacetamide to a final concentration of 50 mM (light sensitive!)
   2.6. Shake at room temperature for 10' in the dark
   2.7. Add 0.25 µg Trypsin (Promega)
   2.8. Shake at room temperature for 2h
   2.9. Centrifuge 4' at 1500g and collect supernatant
   2.10. Add 50 µL denaturation buffer
   2.11. Shake at room temperature for 5'
   2.12. Centrifuge 4' at 1500g and collect supernatant
   2.13. Combine supernatants
   2.14. Add 0.1 µg Trypsin
   2.15. Incubate at room temperature over night
   2.16. Add 5 µL 10% trifluoroacetic acid

3. The peptides are desalted and concentrated using Stage tips, as described in (Rappsilber et al, 2007) and (Vermeulen, 2012) section 4.
   3.1. Fix C18 material in a 200 µL pipette tip
   3.2. Add 50 µL methanol, centrifuge 1500g for 4'
   3.3. Add 50 µL buffer B, centrifuge 1500g for 4'
   3.4. Add 50 µL buffer A, centrifuge 1500g for 4'
   3.5. Repeat 3.4.
   3.6. Apply on-bead digested sample, centrifuge400g for 15'
   3.7. Add 50 µL buffer A, centrifuge 1500g for 4'
   3.8. The peptides can be stored on the stage tip at 4°C
   3.9. Elute the peptides from the stage tip just before LC-MS/MS analysis by applying 30 µL buffer B

4. The peptides are analyzed by LC-MS/MS (LTQ Orbitrap Velos (ThermoScientific)). A detailed protocol can be found in (Vermeulen, 2012), section 8.

5. The data is analyzed using MaxQuant (Cox & Mann, 2008)
    5.1. Load the raw files
    5.2. Specify your experiment in the Experimental Design table. Label your triplicates differentially in the 'experiment' column, e.g. 'GFP_1', 'GFP_2' and 'GFP_3'
    5.3. Change the multiplicity to '1', since we use no isotope labels
    5.4. Load the appropriate Uniprot database
    5.5. Enable the 'label-free quantification' and 'iBAQ' options
    5.6. Start the analysis

6. Identify protein-protein interactions using Perseus (from MaxQuant software package), as described in (Hubner et al, 2010; Hubner & Mann, 2011)
    6.1. Load the 'proteingroups.txt' file into Perseus; *Generic upload*
        6.1.1. In the Expression tab, select the LFQ columns
        6.1.2. In the Numerical tab, select the iBAQ columns as well as the Peptides, Razor + Unique peptides and Unique peptides columns
    6.2. Filter Reverse and Contaminant hits; *Processing – Filter – Filter category*
    6.3. Transform the LFQ column in log2 scale; *Processing – Basic – Transform*
    6.4. Annotate your replicates; *Processing – Column annotation – Categorical annotate rows*, make sure that the replicates have exactly the same name
    6.5. Filter for proteins with at least three valid values in your replicates; *Processing – Filter – Filter valid values*, with *Mode:* 'In at least one group'
    6.6. Impute missing values by semi-random, low values; *Processing – Imputation – Replace missing values by normal distribution*, default settings
    6.7. Execute two sample test; *Analysis – Tests – Two sample test*, default settings
    6.8. Adapt the significance threshold by changing *FDR* and *s0* in the upper left corner.
        6.8.1. Since we do not expect any protein enriched in the control pull-down, try to set the significance thresholds such that no significant outliers appear in the left side of the scatterplot
        6.8.2. By adapting the *FDR*, the threshold line will shift along the y-axis
        6.8.3. By adapting the *s0*, the curve of the line will change. Higher *s0* values correspond with a higher weight factor for the ratio (x-axis)
    6.9. Once the optimal settings are found, export the table on the right; *right-click on the table – Export …*
    6.10. The scatter plot (volcano plot) can also be saved

7. Calculate stoichiometry for the interacting proteins (e.g. in Excel)
    7.1. Select the interacting proteins
    7.2. Check for shared peptides between proteins
        7.2.1. Identify proteins that do not share any peptides with other proteins. The number of peptides and unique peptides are annotated for every interactor in the exported table. If the number of unique peptides for a particular protein is equal to total amount of peptides peptides, this protein does not share any peptides with other proteins.
        7.2.2. Identify proteins that share peptides with other proteins. These proteins have less unique peptides than total peptides. The other

proteins that share these peptides can be identified in the 'peptide.txt' file, a file containing all identified peptides which is generated by MaxQuant. In the column 'Gene Name' the protein names are presented that contain that specific peptide. Proteins annotated in the same row as your protein are proteins that share this peptide.

    7.2.3. Combine the iBAQ values of the proteins that share peptides. The iBAQ values can be found in your exported table from Perseus or the 'proteingroups.txt' file.

7.3. Calculate the average background binding, iBAQ value, from the control triplicate

7.4. Subtract this background binding from the individual GFP-AP replicates (iBAQ values)

7.5. Calculate the relative stoichiometry for every replicate by dividing the background corrected iBAQ values by the iBAQ value of, for example, the GFP tagged protein. Thereby, the GFP tagged protein will have a stoichiometry of 1, and the stoichiometry of all other proteins is expressed relative to the GFP tagged protein.

    7.5.1. This step may be optimized by dividing all iBAQ values by the iBAQ value of another protein. This can be a protein which is known to be monomeric

7.6. The stoichiometries in the different replicates can be plotted as a mean value with standard deviation.

To obtain a more comprehensive characterization of a single complex and to get more confidence about the obtained stoichiometry values for a particular protein complex, one can tag and purify different subunits from the same complex. In this case, the stoichiometry value of a common interacting subunit can be set to 1 and all other values can be scaled accordingly.

Materials and reagents

GFP immunoprecipitation: EpiGeniSys protocol PROT44
Mass Spectrometry: (Vermeulen, 2012) section 8

On-bead digestion
    PBS
    Iodoacetamide
    Trypsin
    Trifluoroacetic acid

Denaturing buffer	2M Urea 100 mM Tris-HCl pH7.5 10 mM DTT

Stage Tip purification
    C18 disk
    Pipette tip
    Methanol

Buffer A	0.5 % (v/v) acetic acid

| Buffer B | 0.5% (v/v) acetic acid |
| | 80 % (v/v) acetonitrile |

Computation analysis
Windows PC
      Quad-core processor
      16 GB RAM
      1 TB HDD
      (optional 128 GB SSD)

Software
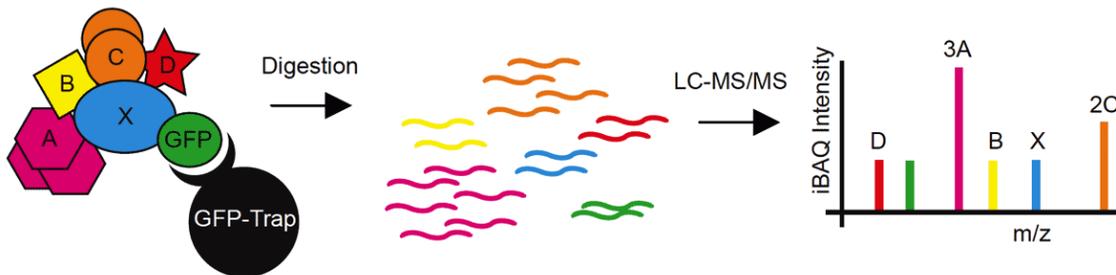   MaxQuant
      Perseus
      Excel

Figures

Figure 1: Schematic representation of the experiment. The abundance of the proteins in the complex can be determined by normalization of the measured mass spectrometric intensities.

Reviewer comments:

Reviewed by: by Angus Lamond
Wellcome Trust Centre for Gene Regulation and Expression, College of Life Sciences, University of Dundee, UK

The protocol by Smits and Vermeulen describes procedures that enable stoichiometric characterization of protein complexes using GFP-tagging, immunoprecipitation and label-free quantitative mass spectrometry. Drawing from several well-established approaches, this protocol offers several advantages, including the flexibility of performing the analysis in tissues, high specificity, and relatively quick sample throughput. Below are listed several considerations that may assist other groups in adapting this protocol for use in their laboratories:

(1) The use of other proteases, in addition to trypsin, (such as Lys-C and/or chymotrypsin), can be helpful for increasing peptide coverage, enables the use of stronger denaturing conditions and improves digestion efficiency. This is especially important for analysis of strongly basic proteins with a high Arg + Lys content, such as histones.

(2) The appropriate amount of trypsin to be added to ensure efficient digestion will vary depending on the amount of immunoprecipitated protein. We therefore advise adding steps to quantify the total protein level recovered in the IP and to check that digestion has gone to completion.

References

Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature biotechnology* **26:** 1367-1372

Dignam JD, Lebovitz RM, Roeder RG (1983) Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic acids research* **11:** 1475-1489

Hubner NC, Bird AW, Cox J, Splettstoesser B, Bandilla P, Poser I, Hyman A, Mann M (2010) Quantitative proteomics combined with BAC TransgeneOmics reveals in vivo protein interactions. *The Journal of cell biology* **189:** 739-754

Hubner NC, Mann M (2011) Extracting gene function from protein-protein interactions using Quantitative BAC InteraCtomics (QUBIC). *Methods* **53:** 453-459

Poser I, Sarov M, Hutchins JR, Heriche JK, Toyoda Y, Pozniakovsky A, Weigl D, Nitzsche A, Hegemann B, Bird AW, Pelletier L, Kittler R, Hua S, Naumann R, Augsburg M, Sykora MM, Hofemeister H, Zhang Y, Nasmyth K, White KP, Dietzel S, Mechtler K, Durbin R, Stewart AF, Peters JM, Buchholz F, Hyman AA (2008) BAC TransgeneOmics: a high-throughput method for exploration of protein function in mammals. *Nature methods* **5:** 409-415

Rappsilber J, Mann M, Ishihama Y (2007) Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nature protocols* **2:** 1896-1906

Rothbauer U, Zolghadr K, Muyldermans S, Schepers A, Cardoso MC, Leonhardt H (2008) A versatile nanotrap for biochemical and functional studies with fluorescent fusion proteins. *Mol Cell Proteomics* **7:** 282-289

Schwanhausser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, Selbach M (2011) Global quantification of mammalian gene expression control. *Nature* **473:** 337-342

Smits AH, Jansen PW, Poser I, Hyman AA, Vermeulen M (2013) Stoichiometry of chromatin-associated protein complexes revealed by label-free quantitative mass spectrometry-based proteomics. *Nucleic acids research* **41:** e28

Vermeulen M (2012) Identifying chromatin readers using a SILAC-based histone peptide pull-down approach. *Methods in enzymology* **512:** 137-160

Vermeulen M, Hubner NC, Mann M (2008) High confidence determination of specific protein-protein interactions using quantitative mass spectrometry. *Curr Opin Biotechnol* **19:** 331-337