Absolute quantification of microbial proteomes at different states by directed mass spectrometry

Alexander Schmidt, Martin Beck, Johan Malmstrom, Henry H.N. Lam, Manfred Claassen, Dave Campbell, Ruedi Aebersold

Corresponding author: Ruedi Aebersold, ETH Zurich

Review timeline:

Submission date:	25 February 2011
Editorial Decision:	13 April 2011
Revision received:	13 May 2011
Accepted:	18 May 2011

Transaction Report:

(Note: With the exception of the correction of typographical or spelling errors that could be a source of ambiguity, letters and reports are not edited. The original formatting of letters and referee reports may not be reflected in this compilation.)

1st Editorial Decision   13 April 2011

Thank you again for submitting your work to Molecular Systems Biology. We have now heard back from two of the three referees who accepted to evaluate the study. Since the recommendations provided byy these two referees are very similar, I prefer to make a decision now rather than to delay the process further. As you will see, the referees find the topic of your study of potential interest and are supportive. They raise however a series of points and make several suggestions for modifications, which we would ask you to carefully address in a revision of the present work. The recommendations are very clear in this regard.

We would also kindly ask you to deposit the primary MS/MS data from the time course experiments in one of the major public databases and include the relevant identifiers in the Materials and Method section. We also greatly appreciate that you included key datasets in supplementary information. We would however kindly ask you to avoid using the PDF format for large tables and upload these files either as Excel tables or plain text (eg in tab delimited or CSV format).

*** PLEASE NOTE *** As part of the EMBO Publications transparent editorial process initiative (see our Editorial at http://www.nature.com/msb/journal/v6/n1/full/msb201072.html), Molecular Systems Biology will publish online a Review Process File to accompany accepted manuscripts. When preparing your letter of response, please be aware that in the event of acceptance, your cover letter/point-by-point document will be included as part of this File, which will be available to the scientific community. More information about this initiative is available in our Instructions to Authors. If you have any questions about this initiative, please contact the editorial office msb@embo.org .

Thank you for submitting this paper to Molecular Systems Biology and I look forward to reading your revised manuscript.

Best wishes,

Editor

Molecular Systems Biology

REFEREE REPORTS

Reviewer #1 (Remarks to the Author):

The manuscript by Schmidt et al. entitled "Absolute quantification of microbial proteomes at different states by directed mass spectrometry" describes the use of directed (in contrast to targeted) mass spectrometry experiments for the (absolute) quantification of 1680 proteins from L. interrogans. The paper is well written, nicely summarizing a large and impressive body of data. As such, the paper can almost be accepted as is. However, I have two (minor) criticisms which should be addressed prior to final acceptance:

1) Their definition of 'observable' is a rather artificial construct and is highly dependent on the analytical strategies used for the comprehensive analysis. As the data are sufficiently impressive as they are, I would significantly shorten the results section, where 'observable' is defined to remove the subjectivity of this aspect. Instead, mentioning that 1680 proteins (out of ~3600 gene products) were absolutely quantified is perfectly fine.

2) I am missing a more detailed description of Figure 4E in the Results section.

Reviewer #2 (Remarks to the Author):

This study uses a model pathogen Leptospira interrogans to investigate the capacity to perform absolute quantitation of a proteome in a consistent and reliable manner over multiple samples in a method that is compatible with high throughput analysis. The particular power of this quantitative method is that consistent proteome datasets can be generated and analyzed for consistent changes that occur due to altered environmental states.

1. The authors state that many of the 194 "missed" proteins are membrane proteins, which is a limitation of the study since oftentimes the alteration of expression of membrane proteins upon exposure to altered environmental states is of primary interest. Can the authors include a brief discussion of suggestions to overcome this limitation?

2. The extent of the proteome coverage obtained using such a posterior proteomics approach is impressive and by far represents the highest coverage obtained to date. It is also clear that without development of such an approach the feasibility of performing such an extensive analysis of a microbial genome would not be possible for the majority of research labs. However, the necessity of generation of a set of heavy labelled peptides could still be considered a limitation that increases cost.

3. Figure 4: The absolute protein concentrations obtained for Leptospira grown in the absence of any external factors (no antibiotic treatment, no FBS) should be included to show protein expression changes that are induced due to growth phase.

4. Pages 13-14: For much of the Leptospira genome, including the region of the genome encompassing the ribosomal genes, operon assignments derive from in silico predictions. The wording within the sections referring to "operon organizations" within Leptospira should reflect this fact.

5. Related to the prior point, the ribosomal "operon" results are surprising. How could one protein be upregulated and another downregulated when both proteins are predicted to be within an operon? The authors suggest post-transcriptional regulation may be a possibility, but in order to make such a suggestion the organization of the genes within an operon would have to be experimentally confirmed and the observed differences in abundance would have to be confirmed to not be due to differences in transcript levels. Unless the authors have performed these studies, they should eliminate such speculation from the interpretation of the data.

Minor:

1. Page 4, 7th line from the top of the page: should read "...and more dependent on the ...".

2. Page 11: Unlike Loa22, catalase and OmpL1 have not been confirmed to be virulence factors in Leptospira, and therefore should be referred to as potential virulence factors.

1st Revision - authors' response	13 May 2011

Please find enclosed the revised manuscript by Alexander Schmidt, Martin Beck, Johan Malmstrom, Henry H.N. Lam, Manfred Claassen, Dave Campbell and Ruedi Aebersold entitled "Absolute quantification of microbial proteomes at different states by directed mass spectrometry' for publication in Molecular Systems Biology.

We greatly appreciate your decision to allow us to submit a revised manuscript. We also thank the reviewers for their thoughtful and constructive criticisms. The reviewerís comments were addressed point-by point as below. Due to some delays with the Tranche Repository account activation, we will upload the primary MS/MS data to this public database as soon as the account is confirmed, which should be ready within the next days. We have already indicated a corresponding link in the Materials and Methods section. Furthermore, the large tables included in the supplementary information are now in Excel (.xls) format for better viewing. We have also added four highlights as 'bullet points' and a short 'standfirst text' to the main manuscript as well as an extended synopsis and 'thumbnail image' as separate files.

Reviewers Comments:

Reviewer #1 (Remarks to the Author):

The manuscript by Schmidt et al. entitled "Absolute quantification of microbial proteomes at different states by directed mass spectrometry" describes the use of directed (in contrast to targeted) mass spectrometry experiments for the (absolute) quantification of 1680 proteins from L. interrogans. The paper is well written, nicely summarizing a large and impressive body of data. As

such, the paper can almost be accepted as is. However, I have two (minor) criticisms which should be addressed prior to final acceptance:

1) Their definition of 'observable' is a rather artificial construct and is highly dependent on the analytical strategies used for the comprehensive analysis. As the data are sufficiently impressive as they are, I would significantly shorten the results section, where 'observable' is defined to

remove the subjectivity of this aspect. Instead, mentioning that 1680 proteins (out of ~3600 gene products) were absolutely quantified is perfectly fine.

We agree with the reviewer that the definition 'observable' proteome critically depends on the methods employed and might lead to some confusion. We therefore changed the following parts as suggested:

Page 2: '…5000 peptides, covering 75% of the L. interrogans observable proteins, can be…'

was changed to:

'...5000 peptides, covering 1680 L. interrogans proteins, can be...'

2) I am missing a more detailed description of Figure 4E in the Results section.

We want to thank the reviewer for this constructive criticism. We have changed the following parts in the corresponding result section on page 11 in order to better describe Figure 4E:

Page 11: '...K-means clustering. The generated cluster profiles were subjected to a pathway enrichment analysis (Huang et al, 2007) to generate a detailed picture of the pathways enriched or reduced in proteome abundance in response to the individual treatments (Fig. 4). Compared...'
 was changed to:

'...K-means clustering (Fig. 4A-D). The generated cluster profiles were subjected to an enrichment analysis of pathways (as present in the KEGG database (Kanehisa et al, 2010)) using the DAVID algorithm (Huang et al, 2007) to generate a detailed picture of the pathways significantly ($p<0.05$) enriched in response to the individual treatments (Fig. 4E). To better visualize the general regulation of the individual protein clusters, protein profiles showing up- (down-) regulation after 24 hours of treatment are indicated in red (blue). Compared...'

We also added the KEGG database reference in the Method section on page 19:
' ...analysis (using the KEGG database (Kanehisa et al, 2010)) of protein...'

The legend of Figure 4 was extended accordingly for clarification:

Figure 4. Pathway analysis of K-means clustered proteome changes. (A-D) The individual protein profiles (changes in copies per cell, log10) obtained were grouped by K-means clustering for each treatment, respectively, and each cluster was searched for significant overrepresentation of proteins belonging to specific pathways using DAVID (Huang et al, 2007). (E) Protein groups that are up-

regulated after 24 hours of treatment are shown in red, while clusters containing mostly proteins with reduced expression are indicated in blue.

was changed to:

Figure 4. Pathway analysis of K-means clustered proteome changes. (A-D) The individual protein profiles (changes in copies per cell, log10) obtained were grouped and numbered by K-means clustering for each treatment, respectively. Each cluster profile is indicated with the first letter of the treatment and its number. (E) All identified protein clusters were searched for significant overrepresentation of proteins belonging to specific pathways according to the KEGG database (Kanehisa et al, 2010) using DAVID (Huang et al, 2007), respectively. Protein groups that are up-regulated after 24 hours of treatment are shown in red, while clusters containing mostly proteins with reduced expression after 24 hours are indicated in blue.

Reviewer #2 (Remarks to the Author):

This study uses a model pathogen Leptospira interrogans to investigate the capacity to perform absolute quantitation of a proteome in a consistent and reliable manner over multiple samples in a method that is compatible with high throughput analysis. The particular power of this quantitative method is that consistent proteome datasets can be generated and analyzed for consistent changes that occur due to altered environmental states.

1. The authors state that many of the 194 "missed" proteins are membrane proteins, which is a limitation of the study since oftentimes the alteration of expression of membrane proteins upon exposure to altered environmental states is of primary interest. Can the authors include a brief discussion of suggestions to overcome this limitation?

We fully agree with the reviewer that membrane proteins are of high interest to study changes in proteome abundance at altered environmental states. As we wrote in the text, the corresponding group of 194 missed protein (many of which are annotated as transporters or membrane bound) were missed in the more labor intense multi-dimensional Off-gel-electrophoresis (OGE) LC-MS/MS experiment and only identified by the presented directed LC-MS/MS approach. The identification and quantification of membrane proteins is a particular strength of our directed LC-MS/MS strategy over the OGE-LC-MS/MS method. To make this issue more clear to the reader, we have changed the corresponding part in the text:

Page 6: '... Functional annotation revealed that many of the 194 protein hits missed by the OGE-LC-MS/MS approach are membrane proteins (Fig. S2), suggesting a decreased recovery of hydrophobic peptides after OGE. Conversely, the OGE-LC-MS/MS strategy showed an increased coverage of low abundant proteins ...'

was changed to:

'... Functional annotation revealed that many of the 194 protein hits exclusively identified by the directed (LC-only) LC-MS/MS approach and missed by the OGE-LC-MS/MS approach are membrane proteins (Fig. S2), suggesting a decreased recovery of hydrophobic peptides after OGE. Conversely, the OGE-LC-MS/MS strategy showed an increased coverage, particularly of low abundant proteins ...'

2. The extent of the proteome coverage obtained using such a posterior proteomics approach is impressive and by far represents the highest coverage obtained to date. It is also clear that without development of such an approach the feasibility of performing such an extensive analysis of a

microbial genome would not be possible for the majority of research labs.

However, the necessity of generation of a set of heavy labelled peptides could still be considered a limitation that increases cost.

We agree with the reviewer that the required heavy labeled peptides increase the overall cost of the approach. However, once synthesized, the peptides are very amount efficient and can be used as quantitative references in several thousand samples, thereby considerably reducing the costs per analysis, particularly, if the peptides are shared between labs. Nonetheless, since additional costs might be an obstacle for some studies, we have added the following sentence to the discussion section to express this issue more specifically:

Page 14: '... A current bottleneck is the dynamic range on the MS1 level that limits the approach to organisms of low to intermediate genomic complexity. Nevertheless, the combination ...'

was changed to:

' ... Current bottlenecks include the necessity of cost-intense heavy labeled reference peptides and the dynamic range on the MS1 level that limits the approach to organisms of low to intermediate genomic complexity. Nevertheless, new high-throughput methods to generate reference peptides, the combination...'

3. Figure 4: The absolute protein concentrations obtained for Leptospira grown in the absence of any external factors (no antibiotic treatment, no FBS) should be included to show protein expression changes that are induced due to growth phase.

We have done this experiment for 24 and 48 hours time points but did not include it in Figure 4, since we detected only very few significant changes in absolute protein levels for non-treated cells and therefore could not generate significant clusters of highly correlating protein profiles for pathway enrichment analysis. However, since it is of potential interest, we have added a figure in the supplemental section of the manuscript (Figure S9) showing that only a very small number of significant protein changes were detected in non-treated cells (6 go up and 7 down, representing less than 1% of all detected proteins after 24 hours, and 33 go down and 20 go up after 48h, representing around 3 % of all detected proteins). We also subjected the regulated proteins to a pathway enrichment analysis but could not get any significant hit. For comparison, 207 significant protein changes (around 14% of all detected proteins) were observed after 24h and 244 (around 16% of all detected proteins) after 48h of serum treatment. Enrichment analysis of the serum induced changes clearly detect the pathways displayed in Figure 4E as significantly enriched.

We therefore conclude that the vast majority of the identified protein changes are specific to the individual treatments and that the impact of unspecific protein changes is negligible.

We revised the manuscript accordingly:

An additional supplemental figure showing the significant changes detected in cells without any treatment after 24 and 48 hours was added, including a figure legend (Figure S9). The numbers of the following supplemental figures were changed accordingly.

A sentence was added in the result part on page 9:

'... 100-fold. Protein abundance changes detected in the absence of any external factors or stimuli were negligible (Fig. S9).'

4. Pages 13-14: For much of the Leptospira genome, including the region of the genome encompassing the ribosomal genes, operon assignments derive from in silico predictions. The wording within the sections referring to "operon organizations" within Leptospira should reflect this fact.

We thank the reviewer for this important comment. We added the following text to the manuscript for clarification:

The following text on page 13:

'... First we asked if proteins that localize to the same operon in the genome have similar...'

was changed to:

'... First we asked if proteins that localize to the same (in-silico-predicted) operon in the genome (Dehal et al, 2010) have similar...'

5. Related to the prior point, the ribosomal "operon" results are surprising. How could one protein be upregulated and another downregulated when both proteins are predicted to be within an operon? The authors suggest post-transcriptional regulation may be a possibility, but in order to make such a suggestion the organization of the genes within an operon would have to be experimentally confirmed and the observed differences in abundance would have to be confirmed to not be due to differences in transcript levels. Unless the authors have performed these studies, they should eliminate such speculation from the interpretation of the data.

We agree with the reviewer that further follow-up experiments are required to support the conclusions drawn at the very end of the result section. Therefore, we removed the corresponding interpretation of the data and changed the text accordingly:

Page 14:

'... We thus conclude that although most proteins within an operon respond to regulation synchronously, bacterial cells seem to have subtle means to adjust individual proteins outside of the general trend. Such individual regulatory events are likely to utilize post-transcriptional mechanisms....'

was changed to:

'... This suggests that although most proteins within an operon respond to regulation synchronously, bacterial cells seem to have subtle means to adjust the levels of individual proteins or protein groups outside of the general trend, a phenomena that recently was also observed on the transcript level of other bacteria (G¸ell et al, 2009). ...'

Minor:

1. Page 4, 7th line from the top of the page: should read "...and more dependent on the ...".

We thank the reviewer for carefully reading the manuscript. We have corrected the spelling as suggested.

2. Page 11: Unlike Loa22, catalase and OmpL1 have not been confirmed to be virulence factors in Leptospira, and therefore should be referred to as potential virulence factors.

We thank the author for this important comment. We changed the text accordingly:

Page 11: '... For instance, virulence surface proteins with confirmed expression in vivo, like OmpL1 (Barnett et al, 1999) or Loa22 (Ristow et al, 2007) were clearly up-regulated on the protein level (both in cluster 'S-5'), but not differentially expressed on the mRNA level (Patarakul et al, 2010), underlining the importance of quantitative proteome studies. In fact, we found the concentration of these validated virulence factors Loa22 and OmpL1... contains virulence factors like catalase (Lo et al, 2010)...'

was changed to:

... For instance, the confirmed virulence surface protein Loa22 (Ristow et al, 2007) and the potential virulence factor OmpL1 (Barnett et al, 1999) with confirmed expression in vivo were clearly up-regulated on the protein level (both in cluster 'S-5'), but not differentially expressed on the mRNA level (Patarakul et al, 2010), underlining the importance of quantitative proteome studies. In fact, we found the concentration of these proteins Loa22 and OmpL1.... contains potential virulence factors like catalase (Lo et al, 2010)...'

Summary

We thank the editor and reviewers for the time and efforts on reviewing this manuscript. We believe that the revised manuscript has adequately addressed all the issues raised by the reviewers and it has been improved significantly. We are confident that the publication of the manuscript will benefit the scientific community by providing an efficient method for system-wide analysis of microbial proteomes. We hope the reviewers and you find the revised manuscript appropriate for publication in MSB.

We thank you for considering the publication.