ORIGINAL RESEARCH

In silico Proteome-wide Amino aCid and Elemental Composition (PACE) Analysis of Expression Proteomics Data Provides A Fingerprint of Dominant Metabolic Processes

David M. Good [1,#], **Anwer Mamdoh** [1], **Harshavardhan Budamgunta** [1], **Roman A. Zubarev** [1,2,*]

[1] *Division of Physiological Chemistry I, Department of Medical Biochemistry and Biophysics, Karolinska Institute, SE 171 77 Stockholm, Sweden*
[2] *Science for Life Laboratory, SE 171 21 Solna, Sweden*

Abstract   Proteome-wide Amino aCid and Elemental composition (PACE) analysis is a novel and informative way of interrogating the proteome. The PACE approach consists of *in silico* decomposition of proteins detected and quantified in a proteomics experiment into 20 amino acids and five elements (C, H, N, O and S), with protein abundances converted to relative abundances of amino acids and elements. The method is robust and very sensitive; it provides statistically reliable differentiation between very similar proteomes. In addition, PACE provides novel insights into proteome-wide metabolic processes, occurring, *e.g.*, during cell starvation. For instance, both *Escherichia coli* and *Synechocystis* down-regulate sulfur-rich proteins upon sulfur deprivation, but *E. coli* preferentially down-regulates cysteine-rich proteins while *Synechocystis* mainly down-regulates methionine-rich proteins. Due to its relative simplicity, flexibility, generality and wide applicability, PACE analysis has the potential of becoming a standard analytical tool in proteomics.

* Corresponding author.
  E-mail: Roman.Zubarev@ki.se (Zubarev RA).
Current address: Department of Medicine, University of Wisconsin
  – Madison, Madison, WI 53706, USA.

Introduction

Modern proteomics analysis provides the identities and the relative abundance changes for thousands of proteins per a single LC–MS/MS experiment [1,2]. However, since many proteins have multiple functions and the exact function of many proteins is not yet known, this information is not always easy to rationalize. Pathway analysis [3,4] provides mapping of the proteome onto more than 160 known signaling pathways and dozens of metabolic pathways. Nonetheless, molecular

pathways are often overlapping and inter-related, such a mapping is rarely unequivocal. A similar problem plagues the popular gene ontology (GO) mapping. Ideally, an aggregate analysis of the proteome state would involve mapping onto a reasonably small number orthogonal, *i.e.*, non-overlapping and mutually independent, classification factors that have clear physico-chemical interpretations. Although mutually orthogonal ("extreme") pathways have been constructed for microorganisms [5,6], such constructs are usually artificial, *i.e.*, do not have clear counterparts at the molecular level.

However, methods to reduce the proteome to a manageable number of orthogonal entities do exist. For example, proteins can be broken down into their constituent amino acids (AAs). Since amino acids in protein sequences are, in general, not mutually interchangeable (the evidence for which is their survival of the evolutionary pressure), they represent an orthogonal set for global proteome analysis. And since all organisms try to minimize the "cost" of protein synthesis by adjusting their AA content to specific growth conditions [7], it is reasonable to assume that changes in these conditions will be reflected in the abundances of the component AAs. Thus, a proteome-wide AA composition analysis can provide an aggregate fingerprint characterizing the specific state of a given organism.

Unfortunately, the current methods for AA analysis all possess significant drawbacks. Edman degradation [8], for instance, is limited with regard to the size of polypeptide which can be interrogated. Meanwhile, acid hydrolysis [9,10] followed by quantification with either ninhydrin [11–13] or mass spectrometry (MS) [14–17] is limited by exposing proteins to harsh chemical treatment, which in turn completely destroys unstable AAs, *e.g.*, tryptophan. Even a short hydrolysis duration leads to deamidation of asparagine and glutamine to aspartic acid and glutamic acid, respectively [10,18].

As will be shown below, the AA and element analyses of whole proteomes can provide valuable information on the ongoing metabolic processes. Here, we present a novel, non-destructive method of performing such analysis on quantitative data obtained in expression proteomics experiments. The entire Proteome-wide Amino aCid and Elemental composition (PACE) analysis is performed *in silico*, and as it can be applied to previously acquired data, it can provide fresh insights from earlier results without a requirement of new experiments. In addition, this method is platform-independent, *i.e.*, can be used for data generated with any mass spectrometric, and even non-mass-spectrometric (*e.g.*, laser fluorescence or antibody-based) quantitative proteomics platforms.

What relevant biological insights can PACE mapping provide? At a very basic level, it can answer the question of whether two given proteomes are different better than any other known statistical method while providing a quantitative estimate of this difference and associated *P* value. PACE mapping also yields a fingerprint of the dominant metabolic processes and, in some cases, even reveals their character. For instance, PACE analysis confirms that single-cell organisms deprived of a single element (*e.g.*, sulfur) during growth exhibit depletion of this element in their proteins [7]. Analyzing both our own and published data with PACE, we investigated the question of whether this depletion is proteome-wide or is instead concentrated in a few highly abundant proteins. We also used PACE to reveal which AA residues get depleted and to what degree. Processes not involving nutrient depletion (*e.g.*, cold or heat stress) also leave a specific mark in the PACE

domain, which subsequently can be used as a fingerprint for their recognition. As a novel and informative way of interrogating the proteome, which combines relative simplicity, flexibility and wide applicability, PACE has the potential of becoming a standard analytical tool in proteomics.
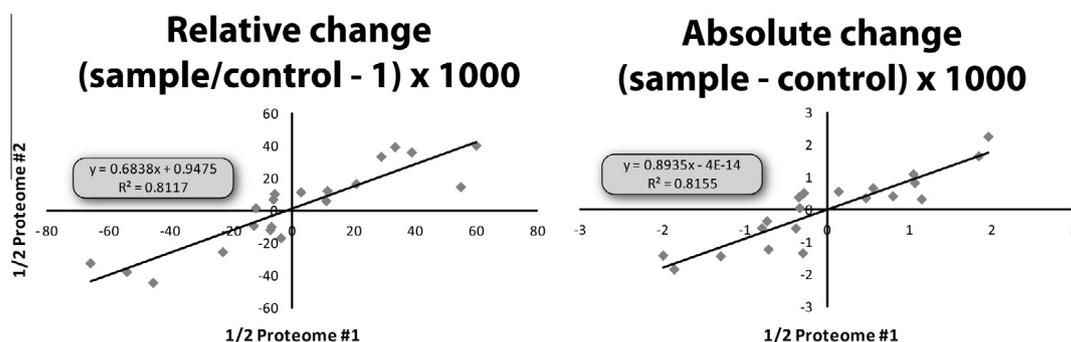
Results

Distribution of PACE signal in the proteome

Until very recently, proteomics analyses were unable to reveal the entire expressed proteome due to the high dynamic range of protein expression. Thus, in any real-life experiment, a subset of the total expressed proteome is sampled, representing the most abundant part of the proteome. To investigate whether the partial nature of the proteomics data affects the PACE diagram, we analyzed a "deep proteomics" (> 50% of the expressed proteome) literature dataset of the model cyanobacterium *Synechocystis* sp. PCC 6803 [19]. The total list of ~2000 quantified proteins was randomly split into two halves, and a PACE AA (**Figure 1**) and elemental histogram (Figure S1) were produced for each of the half-proteomes. The visual similarity between the two histograms is confirmed by correlation analysis (**Figure 2**; $R^2 \geqslant 0.8$ for both correlations). This example demonstrates that the PACE signal is distributed throughout the whole proteome, and the partial nature of real-life proteomics data does not affect the PACE analysis fatally.

Detection of small differences between proteomes

To answer the question as to whether the observed proteome differences between two cellular states are statistically significant, one typically needs to use principal component analysis (PCA) or a similar statistical method to differentiate two groups, each consisting of multiple replicate analyses. In the absence of *a priori* knowledge of statistics associated with protein abundances (each protein being, strictly speaking, a separate statistical entity), there is no easy method to assign statistical significance to a difference, if only two proteomics datasets are available. However, this task becomes solvable with PACE analysis, as the following example demonstrates. In this example, a pair of measured proteomes (lists of ~500 protein identities and respective abundances; T1 and T2) represents two technical replicates of the same proteome B1, while a third measured proteome (B2) represents a separate biological replicate. The protein abundances of the same proteome analyzed repeatedly (technical replicates) are affected by random, statistically independent errors in the measured abundances of individual proteins, while non-identical but biologically similar proteomes (biological replicates) vary in a fundamentally different way, where abundances of the proteins within the same pathway are statistically linked. A simple comparison through the correlation coefficient *R* gives similar values when T1 and T2 are compared ($R^2 = 0.9999$) as well as for the similarity between T2 and B2 ($R^2 = 0.9989$), and provides no estimate for *P* values of the differences (Figure 2A). The failure of standard approaches to robustly differentiate between the biologically unique samples as compared to technical replicates of the same sample is further demonstrated by unsupervised PCA of the data (Figure 2A). Here, the PCA model yields a nonsensical negative Q2 value, illustrating the inability to separate these datasets from each other.

Figure 1   Robustness of the PACE method
The effect of randomly splitting the "sample" and "control" proteomes into two equal parts: the resulting PACE histograms of sample/control comparison are very similar.
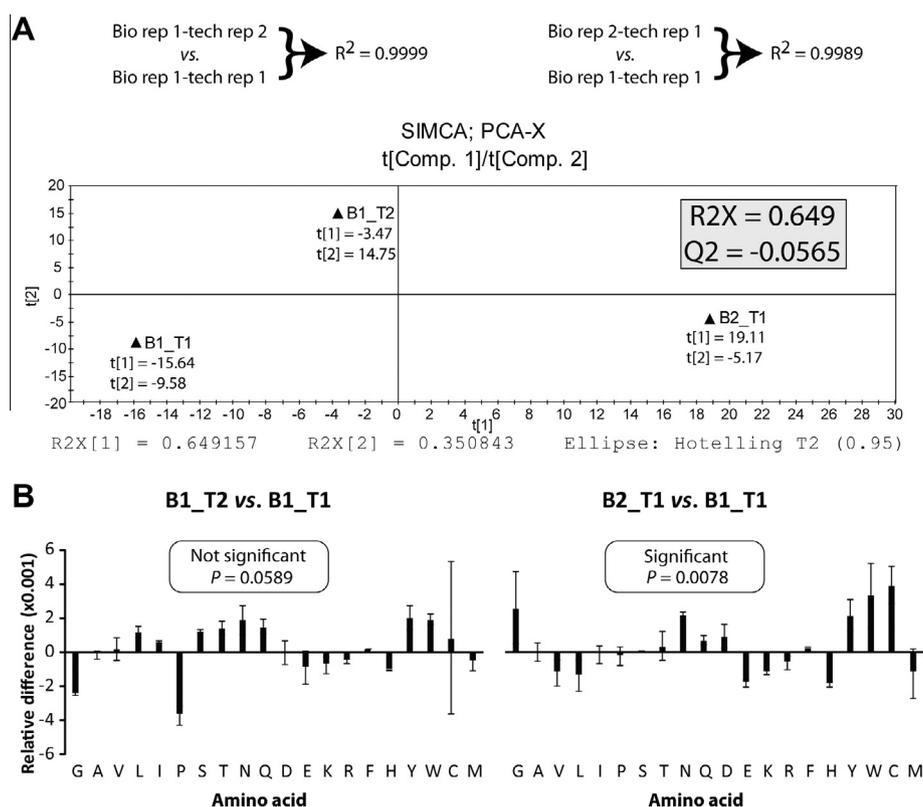
Figure 2   PACE detects minute differences
A. Principle component analysis (PCA) on three measured proteomes, of which two —Biorep 1-tech rep 1(B1_T1) and B1_T2— are technical replicates, and B2_T1 is another biological replicate. **B.** PACE analysis on the same data. left: B1_T1 *vs.* B1_T2; right: B1_T1 *vs.* B2_T1. PCA is unable to distinguish either the technical replicates or the biological replicates from each other with statistical significance, while upon performing PACE analysis, the biological replicates are able to be teased apart with statistical significance, thus illustrating the power of PACE to identify minute but real biological variability.

In contrast, PACE analysis of the same data allows a straightforward statistical testing of the T2/T1 and B2/T1 differences (Figure 2B). To illustrate the method of testing, imagine two measured proteome datasets, A and B, the comparison of which gives a PACE AA histogram A/B. Let us define the PACE "difference" D as a standard deviation of the 20 AA abundance values in A/B from zero. Since the null hypothesis is that A and B represent the same proteome, the true value of D is zero if the null hypothesis is accepted. Thus, the question

of whether A and B represent biologically different proteomes is reduced to testing whether $D_{A/B}$, which is the observed value of D, is consistent with its true value being zero. To address the latter issue, one needs to find the probability to obtain $D_{A/B}$ or larger value by pure chance, *i.e.*, to calculate P value. Assuming the half-normal distribution of D (assumption arising due to the fact that D is always non-negative), P value can be calculated as $P = 1 - \mathrm{erf}(D_{A/B}/[\pi^{1/2}D_m])$, where erf is the error function and D_m is the mean value of D. The latter quantity

can be estimated by repeated random permutation of the protein abundances between A and B (this method of randomization does not require *a priori* knowledge of the statistical properties of individual protein abundances). In the example above, $P \approx 0.06$ (no statistical significance) for the comparison between T1 and T2, whereas $P \approx 0.007$ (good statistical significance) between T1 and B2. Thus for T1 and T2 comparison, the null hypothesis (common origin) remains valid, while for T1 and B2 it should be rejected. Therefore, PACE analysis provides a statistical evaluation of small differences between just a few measured proteome datasets, in a situation where standard statistical methods fail.

Sulfur assimilation by *Escherichia coli*

Sulfur is an essential nutrient and can be a growth-limiting factor in freshwater environments [7]. It is also unique among the six elements most important for life–C, H, N, O, S and P, in that it is mostly protein-related, which makes it most suitable for studying proteomics effects of element availability. Moreover, sulfur is unique among the five most protein-related elements–C, H, N, O and S, in that it is not found within the polypeptide backbone, but instead only in the side chains of two AAs – cysteine and methionine. Therefore, the impact due to changes in the availability of sulfur should be easily traceable not only in the element analysis, but also at the level of the AA content of the proteome.

Indeed, there is ample evidence in the literature of the impact that sulfur has on the proteome. In response to decreased sulfur levels in water, the cyanobacterium *Calothrix* sp. PCC 7601 initiates the production of a methionine- and cysteine- depleted form of its most abundant protein phycocyanin [7]. The cyanobacterium *Fremyella diplosiphon* behaves in a similar way. This response occurs over the physiological range of

sulfate concentrations likely to be encountered by the organism in its natural environment, which can be viewed as a form of environmental accommodation [20]. Although phycocyanin does not take part in sulfur fixation, its elevated expression is believed to affect the sulfur budget of cyanobacterial cells [5]. Other microorganisms, such as bacteria and yeast, can also respond to sulfur and carbon deprivation by reducing the number of sulfur and carbon atoms in the sulfur assimilatory pathway and carbon assimilatory pathway, respectively [21].

One question which has as of yet remained unanswered by previous research is whether sulfur deprivation affects the whole proteome, or depletion in methionine and cysteine is only observed in the most abundant protein(s). Another relevant question is to what extent each of these two AAs is affected. To answer these questions, we grew *E. coli* strain BL21 under conditions when low sulfur or low nitrogen concentrations started to reduce the growth rate (Figure 3). Proteomes of the microbes in their exponential growth phases were extracted and subjected to quantitative proteomics measurements. PACE analysis followed based on ∼500 quantified proteins.

Not completely unexpectedly [16,20], sulfur depletion led to an overall reduction of sulfur content in the proteome, while nitrogen depletion led to reduction of nitrogen (Figure 3B). At the AA level of analysis (left panel), the relative effects of sulfur starvation vary for cysteine and methionine, with cysteine being relatively more depleted. This effect can partially be explained by the fact that, in our PACE analysis, the N-terminal methionine has always been considered present, while in reality many proteins lack this residue. It is, however, unlikely that the observed large differences between the cysteine and methionine peaks are solely due to this phenomenon (*vide infra*). In addition, it is likely that the cysteine/methionine depletion is contained throughout the proteome, and not
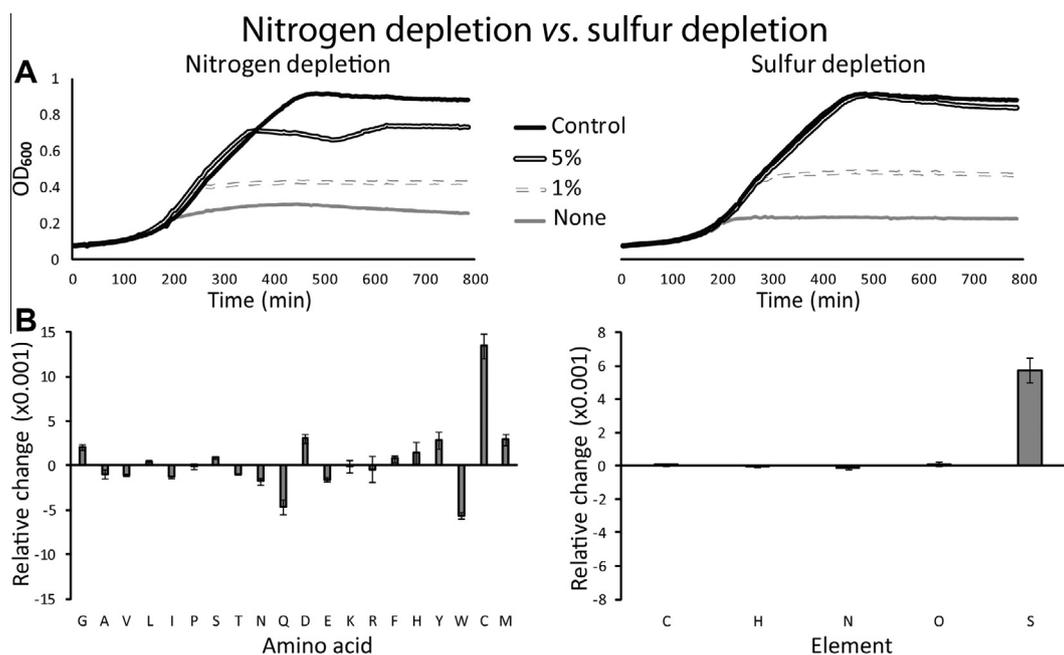
Figure 3   Effect of sulfur depletion and nitrogen depletion on *E. coli*
A. Growth curves of *E. coli* with respect to the level of nitrogen and sulfur content within their minimum growth media. **B.** PACE analysis of the observed proteome changes for nitrogen depletion *versus* sulfur depletion.

Nitrogen depletion *vs.* sulfur depletion
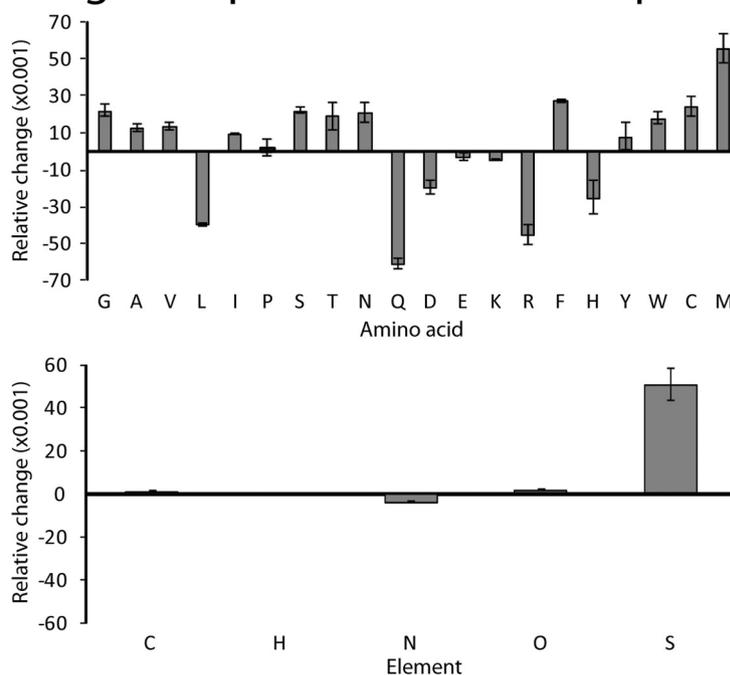
Figure 4 PACE analysis of sulfur depletion and nitrogen depletion on *Synechocystis*
PACE analysis of the observed proteome changes in *Synechocystis* resulting from depletion of sulfur as compared to depletion of nitrogen. The *P* value for sulfur depletion peak is 8×10^{-7}, while for nitrogen enrichment peak, *P* is less than 3×10^{-7} for the element domain.

simply in a few abundant proteins. If the latter were true, then the error bars would be much larger.

In the nitrogen depletion, it is notable that not all nitrogen-rich AAs in the proteome are affected equally. For example, both lysine and arginine show no statistically significant difference between N and S starvations, while both glutamine and asparagine are quite depleted in nitrogen starvation as compared to sulfur starvation. This may be a manifestation of the fact that many *E. coli* strains preferentially catabolize these two AAs upon nitrogen starvation in glucose-ammonia minimal media [22].

Carbon/nitrogen assimilation by a cyanobacterium

Cyanobacteria are the only prokaryotes capable of oxygenic photosynthesis and they play a crucial role in the global carbon/nitrogen balance. Wegener et al. have performed a large-scale proteomic analysis of the widely studied model cyanobacterium *Synechocystis* sp. PCC 6803 under different environmental conditions [19]. We have PACE-analyzed their dataset of approximately 2000 proteins (53% of the predicted proteome) and their abundance changes in response to environmental stress. Most remarkable in the study was the impact of nitrogen deficit (shortage of nitrate) during growth. To account for the observed proteome changes, the authors suggested that the cyanobacterium resorts in these conditions to an unusual pathway in nitrogen accommodation.

As an alternative method to pathway analysis, nitrogen assimilation can be investigated through PACE analysis. In some microorganisms, proteins involved in the assimilation of carbon and sulfur are depleted in these respective elements

compared to the rest of the proteome. Therefore, Baudouin-Cornu et al. predicted that oligotrophic organisms could adapt to the permanent scarcity of an element by diminution of the content of that element in all proteins [22]. This prediction has been confirmed in yeast, which adapts to sulfur scarcity by reducing the content of sulfur-rich proteins in the proteome [23]. However, no net reduction of carbon in the proteome has been reported in yeast, due to its acute response to carbon limitation in relation to yeast limited by other nutrients (N, S or P) [22]. If the nitrogen effect in cyanobacterium is similar to the sulfur effect observed in yeast, one could predict that a nitrogen deficit should lead to down-regulation of nitrogen-rich proteins. To test this hypothesis and also to investigate the sulfur effect in an organism other than yeast, we performed PACE analysis of the dataset from Wegener et al. [19]. The elemental histogram (**Figure 4**) shows the proteome changes in the cyanobacterium grown on a nitrogen-depleted medium as compared to a sulfur-depleted medium. Here, the sulfur peak is strongly positive, while the nitrogen peak is significantly negative. The value of the latter on the arbitrary scale is 3.73, while random permutation of protein identities and abundances gives an average of 0.51. Assuming normal statistics, the *P* value of the nitrogen peak is less than 3×10^{-7}. Similarly, the *P* value for the sulfur depletion peak is 8×10^{-7}. Thus, the effect of down-regulation of sulfur- and nitrogen-rich proteins upon the corresponding starvation, which has been previously seen in yeast [22], exists in other organisms as well.

At the AA level, sulfur depletion affected methionine in the proteome much more significantly than cysteine, in contrast to the situation observed in *E. coli* (compare Figures 3 and 4). Nitrogen depletion caused the most significant down-regulation
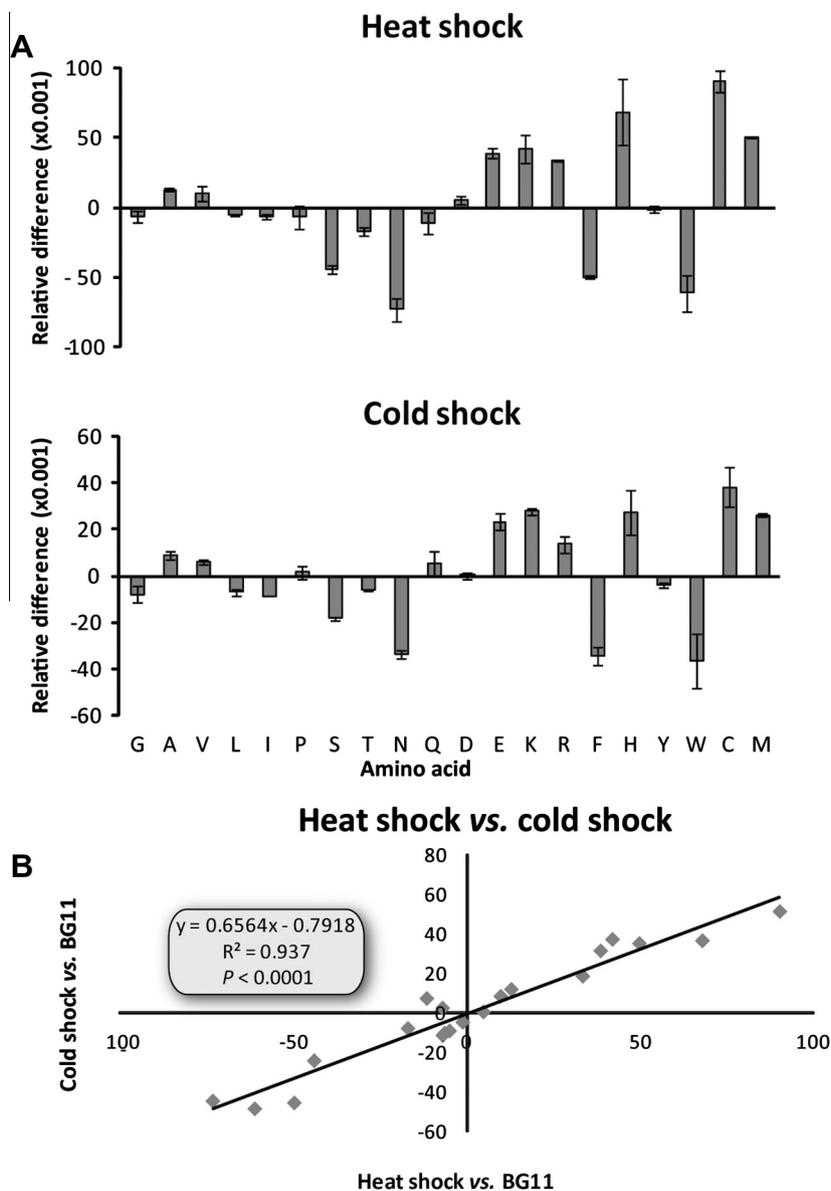
Figure 5   PACE elucidates similarities between heat shock and cold shock response

A. Comparison of PACE analyses of changes within the *Synechocystis* proteome due to heat shock and cold shock compared to standard growth conditions. **B.** Linear correlation between cold- and heat-shock responses in the AA space.

of glutamine (Q)- and arginine (R)-containing proteins, while lysine (K) remained unaffected and asparagine (N) content somewhat increased (Figure 4). Therefore, it appears that the scarcity of nitrogen in the media caused a shortage of arginine, an alternative source of nitrogen for cell growth [19]. Conversion of arginine into succinate also releases, besides glutamate and ammonia (which is also assimilated into glutamate), CO_2, whose carbon is then fixed by ribulose 1,5-bisphosphate carboxylase oxygenase (RuBisCO) [19]. This process may explain the observed excess of carbon-containing proteins under nitrogen starvation conditions (Figure 4).

Interpretation of the proteomics data at the level of individual proteins has been less than straightforward [19]. Classification of differentially regulated proteins according to known cellular functions yielded little insight, as the results were not correlated with observed physiological responses. Moreover,

a large number of proteins with unknown functions showed significant differential regulation during both depletion and recovery phases, as did many proteins associated with common housekeeping functions. Most proteins related to photosynthesis and pigment biosynthesis did not show significant changes in their abundance, although some proteins with several critical functions were differentially regulated. For example, heme oxygenase was down-regulated during nutrient depletion conditions [19]. This demonstrates one pitfall of straightforward interpretation of protein expression levels. That is, although the majority of environmental perturbations had little impact on levels of proteins involved in photosynthesis, the slow growth and chlorosis indicated that the efficiency of photosynthetic reactions was nevertheless significantly affected by these perturbations [19]. In contrast to that complex picture arising due to the intricacy of cellular mechanics and the limited
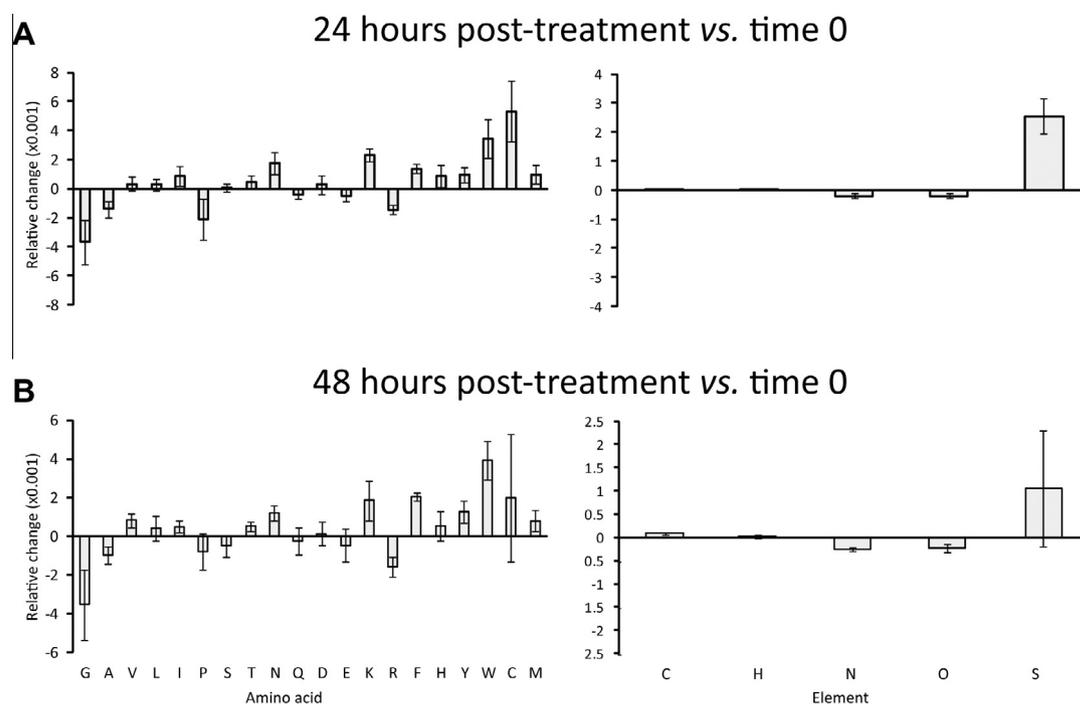
Figure 6   **PACE analysis of arginine deprivation on human carcinoma cell line A431**
The effects of arginine deprivation on sensitive human A431 epidermoid carcinoma cells 24 h (**A**) and 48 h (**B**) after growth in arginine-free media.

knowledge of the functional roles of proteins, PACE analysis provided an aggregate, easily interpretable view on the effect of nutrient deprivation on the proteome.

Fingerprinting of cellular response

Another important aspect of PACE analysis is to provide a fingerprint of the responses of an organism to varying environmental and/or other stresses. **Figure 5** demonstrates how the *Synechocystis* proteome responds to heat or cold stress as compared to normal growth in the control BG11 media. A striking similarity ($R^2 \sim 0.9$, corresponding to $P < 0.0001$) of the AA domain response to these two seemingly opposite stressors was revealed. This similarity is also observed on the elemental level (Figure S2). One may hypothesize that this could be the result of each of these stresses being thermal in nature. However, in *E. coli,* heat shock and cold shock protein are tightly controlled not to be expressed simultaneously [24]. Thus the similarity in the AA and elemental domains does not necessarily extend to the level of individual proteins. Therefore, the above PACE observation is intriguing and invites a more detailed research.

Effect of arginine deprivation on A431 human cells

Specific AA deprivation can selectively target subsets of human cancers. To study the effect of arginine deprivation, human A431 epidermoid carcinoma cells were exposed to varying time intervals with arginine-deprived media. **Figure 6** provides the first-ever view on the effect of such treatment on the proteomes after 24 h and 48 h of arginine deprivation. Not surprisingly, a significant drop in nitrogen is observed

for both depletion periods. Another expected result was the down-regulation of the proteins rich with arginine. Also as expected, and again supporting the robustness of PACE analysis, the AA response patterns for each of the time points are quite similar, with a relative change of each being in the same direction (either up- or down-regulated) within the experimental error.

Perhaps far more interesting than the expected results, however, are the responses of those AAs which do not seem to be affected by such deprivation. For example, though the overall level of nitrogen was reduced, only arginine was found to be down-regulated among the nitrogen-rich AAs. This speaks to the selectivity of arginine deprivation.

Discussion

Searching for a mutually independent limited set of parameters with which to quantitatively characterize the difference(s) between proteomes, we have discovered that proteome-wide amino acid and elemental composition analysis (PACE-analysis) possesses the required features. Mapping the whole proteome onto 20 AAs provides a large parameter space and thus high specificity, while also exhibiting maximum sensitivity, *i.e.*, detecting statistically significant differences between two "identical" biological proteomes, which conventional methods based on individual proteins fail to uncover. Recently, Choi et al. have introduced an interesting approach to finding statistically significant differences in protein abundances that works with a small number of replicates [25]. The difference in the approaches is that Choi et al. assume that different proteins in the same proteome are statistically related, but they do not take into account the identities of individual proteins. In

PROTEIN ID	Control_Bio1_Tech1	Sample_Bio1_Tech1
6PGD_ECOLI	1120630024	1120631111
AAT_ECOLI	3057582429	3137489560
ACCA_ECOSM	478165597.9	478345067.1
ACCC_ECOLI	416814655.8	412346795.2
ACCD_ECOUT	324258680	324563434
ACEA_ECOLI	147378378.7	128978317.6
...	...	...
YOAB_ECOLI	108326165.5	164879797.4
YQHD_ECOLI	62367631.23	78985894.2
YQJD_ECOLI	21468713.71	22252545.6
YRDA_ECOLI	89277171.75	89548791.9
ZAPB_ECOSE	705868109.1	754897809.3
ZNUA_ECOLI	94998408.07	94254878.2

>sp|P00350|6PGD_ECOLI|6-phosphogluconate dehydrogenase, decarboxylating OS=Escherichia coli (strain K12) GN=gnd PE=1 SV=2
MSKQQIGVVGMAVMGRNLALNIESRGYTVSIFNRSREKTEEVIAENPG
KKLVPYYTVKEFVESLETPRRILLMVKAGAGTDAAIDSLKPYLDKGDIII
DGGNTFFQDTIRRNRELSAEGFNFIGTGVSGGEEGALKGPSIMPGGQK
EAYELVAPILTKIAAVAEDGEPCVTYIGADGAGHYVKMVHNGIEYGDM
QLIAEAYSLLKGGLNLTNEELAQTFTEWNNGELSSYLIDITKDIFTKKDE
DGNYLVDVILDEAANKGTGKWTSQSALDLGEPLSLITESVFARYISSLK
DQRVAASKVLSGPQAQPAGDKAEFIEKVRRALYLGKIVSYAQGFSQLR
AASEEYNWDLNYGEIAKIFRAGCIIRAQFLQKITDAYAENPQIANLLLAP
YFKQIADDYQQALRDVVAYAVQNGIPVPTFSAAVAYYDSYRAAVLPAN
LIQAQRDYFGAHTYKRIDKEGVFHTEWLD

G	A	V	L	I	P	S	T	N	Q	D	E	K	R	F	H	Y	W	C	M
41	51	29	40	35	16	25	22	20	21	27	34	28	20	17	4	25	4	2	7

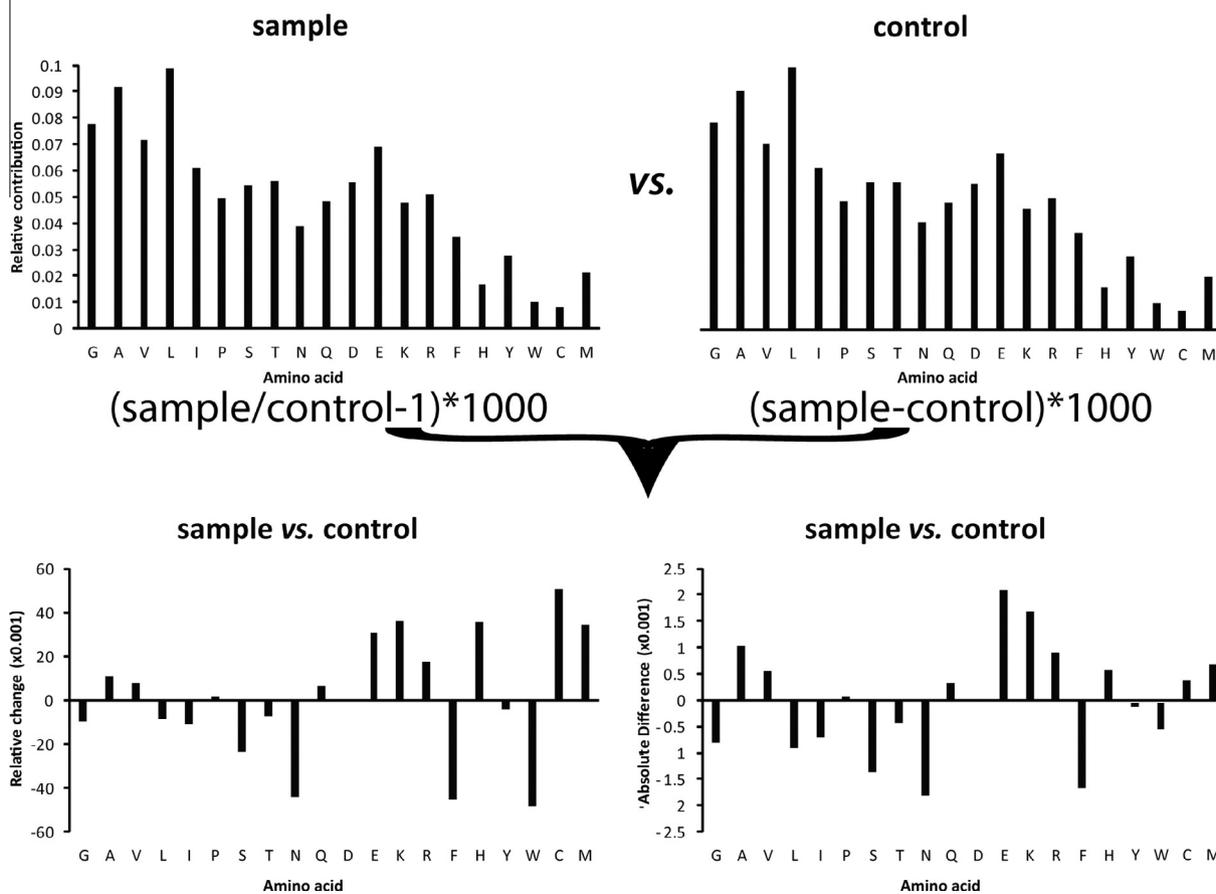| C	H	N	O	S
2311	3613	609	704	9

Figure 7 PACE work-flow
Shown here is a graphical description of the work-flow for PACE analysis. The quantitative proteomics data are loaded and protein sequences are identified in the corresponding protein database. For each sequence found, an array is created with the number of each AA or element contained within that protein. These arrays for all proteins are summed together, using as weighing factors for relative protein abundances in n-th power (scaling factor). The summed arrays for "sample" and "control" can then be compared, resulting in either a "relative" or "absolute" difference.

contrary, PACE analysis considers AA composition of each protein and explicitly utilizes intrinsic correlations between the abundances of proteins that share common compositional features. These two approaches are complementary, and a situation is conceivable (*e.g.*, when all protein abundances differ by less than 50%) when PACE can detect a difference that the approach of Choi et al. will miss.

Mapping the same dataset onto five bio-elements (C, H, N, O and S) reduces the specificity but provides clear insight into metabolic assimilation of nutrients, and can give important clues in the case of a deficit of a valuable element. Finally, PACE, being an *in silico* analysis, is applicable to a wide range of emerging and already published data, thus extending usefulness of such an approach.

Materials and methods

PACE analysis

The PACE approach is illustrated in Figure 7. In the simplest case, proteomics data contain a list of protein identities and their relative abundances As_i and Ac_i for proteomes of "sample" and "control", respectively. In order to avoid a systematic bias due to the differences in total protein amounts, total protein abundances in all proteomes are normalized to the same value prior to PACE analysis. Another required input is protein sequence database. For each protein i in the list, the PACE algorithm finds its AA sequence in the database and reduces it to an occurrence histogram of 20 AA residues, $(^1aa_i \ldots {}^{20}aa_i)$. Then, the occurrence histograms for individual proteins are summed together to a total histogram $(^1AA_i \ldots {}^{20}AA_i)$. Summation occurs with a weight W_i, i.e., $AA_i = W_i \cdot aa_i$, where

$$W_i = A_i^{1/n} \qquad (1)$$

Here, A is the relative abundance of protein, and n (> 0) is the power factor, whose function is to reduce the effect of large proteome dynamic range ($\geqslant 7$ orders of magnitude) and ensure that contribution of each protein to the total weight is not negligible. Typically, the value of n was in the range of 3–5, reflecting the dynamic range of the measured proteome. Note that in PACE analysis, proteins are not separated into up-/down-regulated and unchanged; all protein signals are utilized, regardless of their intensity or statistical significance, as statistical evaluation of the results is performed at a later stage.

The total histograms $(^1AA_{s/c} \ldots {}^{20}AA_{s/c})$ for "sample" and "control" are then compared in relative terms:

$$^jk_r = ((^jAA_s/^jAA_c) - 1) * 1000 \qquad (2)$$

as well as "absolute" terms,

$$^jk_a = (^jAA - {}^jAA_c) * 1000 \qquad (3)$$

and expressed in pro*mil* (×0.001). Each resultant dataset contains 20 numbers, both positive and negative, that show the change (relative or "absolute") of abundances for respective AAs in the proteome of "sample" and "control" compared to "control". A similar procedure is used for elemental composition analysis, with $^lE_{s/c}$ ($l = 1\ldots5$) replacing $^jAA_{s/c}$. The magnitudes and the error bars for the total histogram were calculated from a set of results, each obtained from PACE analysis of a unique "sample"–"control" pair of replicates. For instance, if there are two replicates for "sample" and "control", then the four pairwise comparisons (S1/C1, S1/C2, S2/C1 and S2/C2) will give a set of four values for each histogram column. The average of this set will be reported as the column magnitude, while standard error will be represented as its error bar.

E. coli growth and analysis

E. coli BL21 stock cells were cultured in M9 minimal media. To observe varying stress responses of the organism due to depletion of certain elements, specific forms of the M9 media deficient in carbon (glucose), nitrogen (NH_4Cl) or sulfur ($MgSO_4$) were employed: control (none of the elements depleted), 5% of control, 1% of control, and 0% (100% depletion). All samples were incubated in a Bioscreen C Automated Microbiology Growth Curve Analysis System (Growth Curves USA, Piscataway, NJ) with a growth time of 24 h at 39 °C. Growth was automatically recorded through use of optical density measurements taken at wavelength of 600 nm (OD_{600}). Three biological replicates were run for each condition.

At the end of culture, 3 mL of E. coli containing media were collected for each condition and spun down at ∼5000 g. The resulting pellet was rinsed with PBS and re-pelleted. Lysis and digestion were performed as outlined previously [26]. Briefly, lysis buffer (8 M urea, 75 mM NaCl, 50 mM Tris, one tablet of Complete Mini protease inhibitors cocktail [Roche Diagnostics, Bromma, Sweden] and 10 mM sodium pyrophosphate) was added in a volume ratio of ∼3:1 buffer to cell pellet. Samples were probe-sonicated on ice – 3×60 s with 90 s pause (6 s run, 3 s pause; amplitude 40%), vortexed and then centrifuged at 20,000 g for 20 min at 4 °C.

Protein concentration was determined using BCA assay (Thermo Scientific, Rockford, IL, USA) and 20 μg of each sample were taken for overnight trypsin digestion, following the method previously described [26]. Resulting peptides were cleaned using C_{18} Zip-Tips (Millipore, Billerica, MA, USA) and samples were analyzed by LC–MS/MS employing an EASY nLC (Thermo Scientific, Odense, Denmark) coupled to a Velos Orbitrap mass spectrometer equipped with electron transfer dissociation (ETD) [27,28] (Thermo Scientific, Bremen, Germany). Survey mass spectra were acquired at 60,000 resolving power and a data-dependent top-10 method was employed, with each precursor ion being fragmented by both ETD and collision-activated dissociation (CAD) in the linear ion trap, with subsequent detection there.

Resulting .raw data were converted to Mascot generic format (.mgf) files using in-house software and ETD spectra were cleaned [29,30] prior to database searching with Mascot. CAD and ETD spectra were not separated prior to searching against a concatenated version of the SwissProt E. coli database. The parameters employed were: peptide tolerance ± 10 ppm, fragment ion tolerance ± 0.6 Da, a maximum of three missed cleavages, fixed modification of carbamidomethyl on cysteine and a variable modification of oxidation on methionine. Search results were downloaded to a local computer as .dat files and subsequently filtered to a $< 1\%$ false discovery rate (FDR) using the target-decoy strategy [31]. These filtered files were then merged and the retention times for sequenced peptides were aligned using in-house software. This merged file was then re-searched in Mascot against a forward-only database. The resulting .dat file was used by the Quanti quantification algorithm [32] for label-free quantification with a minimum of three proteotypic peptides employed for calculation of abundance of each protein. PCA was performed using SIMCA-P + software (Umetrics, Sweden).

Arginine deprivation

A431 epidermoid carcinoma cells (ATCC; CRL-1555) have previously been shown to be sensitive to depletion of the conditionally-essential amino acid arginine [33], likely due to inactivity of argininosuccinate synthetase (ASS1). Therefore, we chose here to employ these cells under such deprivation conditions to provide a model system for investigating the ability of PACE to tease out important biological information from experiments focused on in vitro studies of human cell lines.

A431 cells were cultured in Dulbecco's Modified Eagle's Medium (DMEM; VWR, Solna, Sweden) supplemented with 10% fetal bovine serum (heat-inactivated at 57 °C for 1 h), 1% L-glutamine, 1% streptomycin/penicillin and 1% sodium pyruvate. Cells were cultured in a humidified atmosphere with 5% CO_2 at 37 °C. Arginine-depleted media was obtained by adding arginase (20 units for 40 mL of media). After cell splitting and establishing solid growth, their growth media were replaced with the depleted media (time = 0). Control cells were grown in full media. Cells were harvested after 24 and 48 h in the depleted media. Upon reaching ~75% confluency, cells were trypsin-released, rinsed with PBS and pelleted prior to lysis (lysis, sample clean-up and LC/MS/MS analysis were performed as described above).

Authors' contributions

DMG and RAZ designed experiments. DMG wrote the PACE software. AM performed *E. coli* experiments; HB and DMG performed human cell line experiments. DMG and RAZ wrote the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors claim no competing interests in the work presented here.

Acknowledgements

Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.gpb.2013.07.002.

References

[1] Graumann J, Hubner NC, Kim JB, Ko K, Moser M, Kumar C, et al. Stable isotope labeling by amino acids in cell culture (SILAC) and proteome quantitation of mouse embryonic stem cells to a depth of 5111 proteins. Mol Cell Proteomics 2008;7:672–83.

[2] de Godoy LM, Olsen JV, Cox J, Nielsen ML, Hubner NC, Frohlich F, et al. Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. Nature 2008;455:1251–4.

[3] Good DM, Zubarev RA. Drug target identification from protein dynamics using quantitative pathway analysis. J Proteome Res 2011;10:2679–83.

[4] Zubarev RA, Nielsen ML, Fung EM, Savitski MM, Kel-Margoulis O, Wingender E, et al. Identification of dominant signaling pathways from proteomics expression data. J Proteomics 2008;71:89–96.

[5] Schilling CH, Palsson BO. Assessment of the metabolic capabilities of Haemophilus influenzae Rd through a genome-scale pathway analysis. J Theor Biol 2000;203:249–83.

[6] Schilling CH, Schuster S, Palsson BO, Heinrich R. Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era. Biotechnol Prog 1999;15:296–303.

[7] Mazel D, Marliere P. Adaptive eradication of methionine and cysteine from cyanobacterial light-harvesting proteins. Nature 1989;341:245–8.

[8] Edman P. A method for the determination of amino acid sequence in peptides. Arch Biochem 1949;22:475.

[9] Braconnot HM. Sur la conversion des matières animales en nouvelles substances par le moyen de l'acide sulfurique. Ann Chim Phys Ser 2 1820;13:113–25.

[10] Burr GO, Gortner RA. The humin formed by the acid hydrolysis of proteins VIII. The condensation of indole derivatives with aldehydes. J Am Chem Soc 1924;46:1224–46.

[11] Moore S, Stein WH. Chromatographic determination of amino acids by the use of automatic recording equipment. Methods Enzymol 1963;6:819–31.

[12] Alterman MA, Hunziker P. Amino acid analysis: methods and protocols. Totowa, New Jersey: Humana Press; 2012.

[13] Cooper C, Packer N, Williams K. Amino acid analysis protocols. New York: Humana Press; 2001.

[14] Bordeerat NK, Georgieva NI, Klapper DG, Collins LB, Cross TJ, Borchers CH, et al. Accurate quantitation of standard peptides used for quantitative proteomics. Proteomics 2009;9:3939–44.

[15] Louwagie M, Kieffer-Jaquinod S, Dupierris V, Coute Y, Bruley C, Garin J, et al. Introducing AAA-MS, a rapid and sensitive method for amino acid analysis using isotope dilution and high-resolution mass spectrometry. J Proteome Res 2012;11:3929–36.

[16] Kato M, Takatsu A. Amino acid analysis by hydrophilic interaction chromatography coupled with isotope dilution mass spectrometry. Methods Mol Biol 2012;828:55–62.

[17] Mirgorodskaya OA, Korner R, Kozmin YP, Roepstorff P. Absolute quantitation of proteins by acid hydrolysis combined with amino acid detection by mass spectrometry. Methods Mol Biol 2012;828:115–20.

[18] Zubarev RA, Chivanov VD, Hakansson P, Sundqvist BU. Peptide sequencing by partial acid hydrolysis and high resolution plasma desorption mass spectrometry. Rapid Commun Mass Spectrom 1994;8:906–12.

[19] Wegener KM, Singh AK, Jacobs JM, Elvitigala T, Welsh EA, Keren N, et al. Global proteomics reveal an atypical strategy for carbon/nitrogen assimilation by a cyanobacterium under diverse environmental perturbations. Mol Cell Proteomics 2010;9:2678–89.

[20] Gutu A, Alvey RM, Bashour S, Zingg D, Kehoe DM. Sulfate-driven elemental sparing is regulated at the transcriptional and posttranscriptional levels in a filamentous cyanobacterium. J Bacteriol 2011;193:1449–60.

[21] Baudouin-Cornu P, Surdin-Kerjan Y, Marliere P, Thomas D. Molecular evolution of protein atomic composition. Science 2001;293:297–300.

[22] Bragg JG, Wagner A. Protein carbon content evolves in response to carbon availability and may influence the fate of duplicated genes. Proc Biol Sci 2007;274:1063–70.

[23] Fauchon M, Lagniel G, Aude JC, Lombardia L, Soularue P, Petat C, et al. Sulfur sparing in the yeast proteome in response to sulfur demand. Mol Cell 2002;9:713–23.

[24] Yamanaka K. Cold shock response in *Escherichia coli*. J Mol Microbiol Biotechnol 1999;1:193–202.

[25] Choi H, Fermin D, Nesvizhskii AI. Significance analysis of spectral count data in label-free shotgun proteomics. Mol Cell Proteomics 2008;7:2373–85.

[26] Good DM, Rutishauser D. Employment of complementary dissociation techniques for body fluid characterization and biomarker discovery. Methods Mol Biol 2013;1002:223–32.

[27] Syka JE, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. Proc Natl Acad Sci U S A 2004;101:9528–33.

[28] McAlister GC, Phanstiel D, Good DM, Berggren WT, Coon JJ. Implementation of electron-transfer dissociation on a hybrid linear ion trap-orbitrap mass spectrometer. Anal Chem 2007;79:3525–34.

[29] Good DM, Wenger CD, McAlister GC, Bai DL, Hunt DF, Coon JJ. Post-acquisition ETD spectral processing for increased peptide identifications. J Am Soc Mass Spectrom 2009;20:1435–40.

[30] Good DM, Wenger CD, Coon JJ. The effect of interfering ions on search algorithm performance for electron-transfer dissociation data. Proteomics 2010;10:164–7.

[31] Elias JE, Gygi SP. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nat Methods 2007;4:207–14.

[32] Lyutvinskiy Y, Yang H, Rutishauser D, Zubarev R. *In silico* instrumental response correction improves precision of label-free proteomics and accuracy of proteomics-based predictive models. Mol Cell Proteomics 2013;12:2324–31.

[33] Scott L, Lamb J, Smith S, Wheatley DN. Single amino acid (arginine) deprivation: rapid and selective death of cultured transformed and malignant cells. Br J Cancer 2000;83:800–10.