RESEARCH ARTICLE

Protecs, a comprehensive and powerful storage and analysis system for OMICS data, applied for profiling the anaerobiosis response of *Staphylococcus aureus* COL

Stephan Fuchs[1], Henry Mehlan[1], Harald Kusch[1], Alexander Teumer[2], Daniela Zühlke[1], Matthias Berth[3], Carmen Wolf[1], Thomas Dandekar[4], Michael Hecker[1], Susanne Engelmann[1] and Jörg Bernhardt[1,3]

[1] Institute for Microbiology, Ernst-Moritz-Arndt-University Greifswald, Greifswald, Germany
[2] Interfaculty Institute for Genetics and Functional Genomics, Ernst-Moritz-Arndt-University Greifswald, Greifswald, Germany
[3] DECODON GmbH, Biotechnikum, Greifswald, Germany
[4] Department of Bioinformatics, Bavarian Julius-Maximilians-University Würzburg, Biocenter, Würzburg, Germany

Received: June 5, 2009
Revised: March 5, 2010
Accepted: May 25, 2010

Broad functional genomic studies call for comprehensive and powerful data repositories for storage of genome sequences, experimental information, protein identification data, protein properties and expression values. The better such data repositories can integrate and display complex data in a clear and structured way the more biologically meaningful conclusions or novel hypotheses can be derived from extensive omics data sets. This work presents the web accessible database system Protecs and how it was used to support analysis of 50 samples drawn from four *Staphylococcus aureus* cultivations under anaerobiosis. Protecs incorporates findings from visualization science, *e.g.* micro charts and heat maps in the user interface. Its integrated tools for expression data analysis in combination with TIGR Multi Experiment Viewer were used to highlight similar gene expression profiles in single or multiple experiments based on the continuously updated *S. aureus* master gel. Raw data analysis results are available online at www.protecs.uni-greifswald.de. Our meta-study revealed that *S. aureus* responds in different anaerobiotic experimental setups (growth without oxygen; growth without oxygen but with supplemental pyruvate and uracil; growth without oxygen but with NO_3^-; growth without oxygen but with NO_3^- and without functional *nreABC* genes) with a general anaerobiosis response. Among others, this response is characterized by an induction of fermentation enzymes (PflB, Ldh1, SACOL0135, SACOL0660) as well as the response regulator SrrA. Interestingly, especially genes with a high codon adaptation index highly overlap with anaerobically induced genes.

Keywords:
Anaerobiosis / Bioinformatics / Database / Data management / Functional genomics / *Staphylococcus aureus*

1 Introduction

Omics research opens up a global view on cell performance and physiology and calls for fully integrated data repositories, which store data and, even more importantly, make them available to the scientific community and to the public at large.

Especially in the publicly accessible sector, strong efforts enabled researchers to access data collections of general use. This started with first sequence and publication directories like GenBank [1, 2], PubMed [3, 4] or UniProt [5, 6]. in recent years, more and more specialized databases complement the information that is provided by the general ones. Influenced by the high throughput sequencing and functional characterization efforts, which started in the mid-1990s, databases regarding specialized data sets of distinct properties of biological objects became accessible. Among

Correspondence: Dr. Jörg Bernhardt, Institute for Microbiology, Ernst-Moritz-Arndt-University Greifswald, Jahnstrasse 15, 17487 Greifswald, Germany
E-mail: Joerg.bernhardt@uni-greifswald.de
Fax: +49-3834-864202

Abbreviations: CAI, codon adaptation index; **MIAPE,** minimum information about a proteomics experiment; **TMEV,** TIGR Multi Experiment Viewer

them, databases for 3-D structures of proteins (PDB) [7, 8], horizontally transferred genes in prokaryotes (HTG-DB) [9, 10] as well as specific databases summarizing all aspects of biological entities of one species or a related set of species (*e.g.* databases of the Genolist family [11]) have been developed. The annual database January issue of *Nucleic Acid Research* regularly reviews these developments [12].

Initially, Protecs was planned for the species-centric representation of omics data related to specific physiological contexts for the gram-positive model bacteria *Staphylococcus aureus* and *Bacillus subtilis*. Protecs is the direct successor of Sub2D and Staph2D [13–15], established in the mid-1990s to disseminate the master gel information from 2-D gel-driven proteome research *via* WWW to interested researchers [16]. A short time after the initiation of Sub2D and Staph2D, it became obvious that the relatively simple data model of the original system needed a refurbishment to fulfill the needs of modern omics research. Lab book (experiment), genomic, transcriptomic, proteomic and metabolic information need to be stored and linked to each other for keeping an overview on the resulting complexity of data (for a simplified data model, see Supporting Information Fig. S1). Besides maintaining and storing data, such a system has to interconnect the data with external databases. Therefore, Protecs complies with the rules for establishing federated database systems [17].

Protecs is a powerful tool for (i) archiving, (ii) managing, (iii) linking, (iv) presenting and (v) analyzing omics data collections. Within Protecs, these data can be updated on the fly with incoming data or analyzed with new analysis tools and techniques. These functions not only allow linking of ''old'' experiments with newer ones or with subsequently generated complementing findings, but are also crucial to be prepared for future challenges. As an example, in the recent years, we were especially interested in stress-related responses on the transcriptome or proteome level in *S. aureus* [14, 18–22]. Protecs can now help us to determine the expression pattern of, for example, ''housekeeping proteins'' constitutively expressed regardless of the nature of the stressing agent. Only the combination of data from different experimental conditions yields a core set of proteins necessary for the survival of the organism in general. A review of ''older'' experimental data and their complementation by new findings, *e.g.* new identifications of protein spots on the master gel, can greatly enhance the impact and scientific significance of the analyzed data. Furthermore, the interlinked analysis of different experimental projects as supplied by Protecs allows the definition of stimulons and regulons at a higher level of coverage.

To demonstrate this, we complemented data from anaerobiosis-related proteomic projects published previously [23, 24] with a high amount of extended *S. aureus* master gel protein identification data (unpublished data). Protein synthesis data of *S. aureus* belong to samples grown in the presence of oxygen or nitrate as terminal electron acceptors or under fermentative conditions [23] with and without pyruvate and uracil [25]. Since *S. aureus* also faces low oxygen availability when invading different tissues [26] (*e.g.* anaerobic conditions were found in abscesses [27]), the results are potentially relevant with respect to pathogenic mechanisms.

In this study, we show how Protecs integrates complex data to comprehensively characterize omics data sets both on a global and on a single-gene (product) level across several experimental setups and thus helps to understand the physiological adaptations of *S. aureus* during the transition from aerobic to different oxygen-limited conditions.

2 Materials and methods

2.1 Caché and Caché Server Pages

Caché and Caché Server Pages form the technical backbone of Protecs. Caché is an object database management system established by Intersystems and is based on a post-relational data model [28]. It provides object and SQL access to the database, as well as allowing direct manipulation of Caché's underlying data structures (http://www.intersystems.com/cache/index.html). It is widely used in biomedical environments.

Internally, Caché stores data in multidimensional arrays capable of carrying hierarchically structured data. Caché is based on MUMPS database programming language and includes Multivalue data access (aka PICK). In most applications, however, object and/or SQL access methods are used.

Caché ObjectScript, Caché Basic or T-SQL are used to develop the application business logic. External interfaces include native object binding for C++, Java, EJB, ActiveX and .NET. Caché supports JDBC and ODBC for relational access as well as XML and web services.

Caché Server Pages technology allows tag-based creation of web applications that dynamically generate web pages, typically using data from a Caché database.

2.2 Protecs

Protecs establishes the appropriate bioinformatics infrastructure for functional genomic environments on top of Caché. Besides data storage and data administration, methods for global data analysis and data integration have been adapted or developed. Protecs stores data on (i) genome level, (ii) RNA level, (iii) protein level and (iv) metabolites for different species and a variety of experimental and physiological conditions. Furthermore, data on (v) protein localization, (vi) on regulatory interactions and (vii) on the experimental setup can be stored.

File interfaces allow the export and import of data to and from external applications, *e.g.* TIGR Multi Experiment Viewer (TMEV). Moreover, Delta2D, a software for the analysis of 2-D electrophoresis gels [29], can directly work as a Protecs client in a bi-directional way. For web-based

outputs, different Java applets visualize 2-D gel as well as DNA array data. Human as well as machine- or software-based interaction with Protecs has been made available by web, database and several software interfaces.

2.3 Data visualization

Data visualization for multidimensional data sets poses the challenge of representing vast amounts of gene (product) and expression data in an easily understandable and amenable manner but without losing informational content. With the aim of using minimal screen space, we implemented web pages that dynamically create heat maps [30] and minimalistic micro charts [31] from the underlying data of interest (see also Section 3). For a contextual interpretation of numeric gene/protein parameters, we decided to use micro chart-based histograms.

2.4 Strains and growth conditions

S. aureus COL wild-type and a derived *nreABC* mutant strain have been used in all experiments (Table 1). Cells were grown in synthetic medium until oxygen limitation was implemented [24] at an optical density of 0.5 at 500 nm. At this point, reference sample t_0 was harvested. Additional sampling points were 10, 20, 30 and 60 min anaerobiosis and an aerobically grown control (60 min aerobiosis). For a more detailed description of cultivation and experimental setups (see references in Table 1).

2.5 Radioactive labeling and 2-DE

Newly synthesized proteins were pulse labeled by incorporation of ^{35}S-methionine. 2-D SDS-PAGE was used to separate intracellular proteins. Gels were dried and detected by using a phosphorimager (Typhoon, GE, Freiburg, Germany) as described earlier [23]. Resulting images and detailed information on cultivation, sampling and technical parameters are provided at Protecs (www.protecs.uni-greifswald.de).

2.6 Spot detection

For all the experiments, spot detection and quantitation was updated according to the spot modeling approach [29] implemented in Delta2D since version 3.6. The resulting spot consensus patterns were manually optimized using Delta2D's spot editing tool within each experiment and then transferred to each gel of the corresponding experimental sampling series. This procedure ensures 100% spot matching. Spots covering more than one identification label are indicated in the quantitation tables (Supporting Information Tables S1–S4).

2.7 Protein identification and update

Proteins were identified by MS as described previously [23]. Briefly, spots were cut from gels and were transferred onto 96-well microtiter plates. Digestion with trypsin and subsequent spotting of peptide solutions onto the MALDI targets were performed by an Ettan Spot Handling Workstation (GE Healthcare, Little Chalfont, UK). MALDI-TOF-MS analyses of spotted peptide solutions were carried out on a Proteome-Analyzer 4700 (Applied Biosystems, Foster City, CA, USA). Spectra were recorded in reflector mode (mass range: 900–3700 Da). After calibration, peak lists were created using the ''peak to MASCOT'' script of the 4700 Explorer software. Peak lists were analyzed by using the MASCOT search engine (Matrix Science, London, UK), GPMAW 4.1 (Lighthouse data). Peptide mixtures that yielded at least twice a Mowse score of at least 50 and a sequence coverage of at least 30% were regarded as positive identifications. Proteins that failed to exceed the 30% sequence coverage cut-off were subjected to MALDI-MS/MS. Identification results were integrated into a master gel with 677 identified spots, which was used for spot identification on all other images of the performed experiments.

Table 1. Experimental parameters

	Experiment	Strain	Supplements[a)]	Stressor	Reference
A	Fermentation	COL	–	Oxygen limitation	[23]
B	Fermentation (supplemented with uracil and pyruvate)	COL	Uracil (5 mg/L)[b)] sodium pyruvate (2.5 mg/L)[b)]	Oxygen limitation	this work
C	Nitrate respiration (NO_3^- supplementation)	COL	Nitrate (8 mM)	Oxygen limitation	[23]
D	Δ*nreABC* background (NO_3^- supplementation)	COL Δ*nreABC*	Nitrate (8 mM)	Oxygen limitation	[24]

a) Cells were cultivated in synthetic medium as described previously [23].
b) According to [25].

Six hundred and seventy-seven spot labels (identifications) were manually transferred from the master gel to the experiment-specific fusion gel (proteome maps; [32]). Transferred spot labels were inspected and corrected if necessary. By this approach, 599 (fermentation), 527 (fermentation in the presence of uracil and pyruvate), 542 (nitrate respiration) and 566 (anaerobic growth of the Δ*nreABC* mutant) spot labels were successfully recovered in single experiments. In total, 465 expression profiles covering all experiments were considered.

In some cases, warping-based spot identification fails due to non-affine variations in spot positions. Involved spots had to be re-identified. If this was not possible, corresponding parts of expression profiles are indicated by empty cells within heat maps (101 expression profiles). Spot identification information of single experiments as well as Delta2D-based spot consensus patterns were uploaded to Protecs. They are publicly available by accessing the corresponding gel images (www.protecs.uni-greifswald.de).

2.8 Quantitation and statistical analysis

For quantitation, grey level integration within the spot boundaries of the spot consensus patterns was performed. For statistical analysis, we used Delta2D's relative spot quantitation data, which were normalized with respect to the total spot intensity of each gel. Expression values were standardized. This means expression profiles were made comparable by mean centering and variance scaling. Normalization and standardization was applied for every individual experiment. Further statistical analysis was performed with TIGR MeV 4.3 [33]. Clustering of samples and expression profiles was performed with Support Trees [34]. Generally, we used Bootstrapping to resample protein intensities and samples. Hierarchical clustering was performed by using complete linkage and Euclidean distance.

Statistically significantly changed spot quantities within expression profiles were determined by using one-way ANOVA within each experiment [35]. We used the non-parametric version based on a permutation test (1000 permutations) and an uncorrected alpha of 0.1. At least two biological replicates of each sampling time points were grouped. For the 30-min sampling point at fermentation (supplemented with uracil and pyruvate), one replicate experiment failed and was not available. Therefore, this time point was excluded from statistical analysis within the experiment. Proteins that show at least a 2.5 fold induction or repression are listed in Supporting Information spot Tables S1–S8.

Additionally, changes of expression level with respect to the unstressed sample at OD 0.5 were tested for each single time point with a *T*-Test [36–38] using Welch approximation of group variance and an uncorrected alpha of 0.1 based on all possible permutations. For an overall characterization of samples, PCA [39] was used.

3 Results

3.1 Database usage, setup and key contents

Protecs allows the integration of genomic information with descriptions of experimental setups and the corresponding transcriptome, (2-D gel-based) proteome and metabolome data.

3.1.1 Protecs URL, accessibility and entry page

Protecs is accessible *via* www.protecs.uni-greifswald.de. Directly from the entry page, public access (link "public

Figure 1. Homepage of Protecs. Protecs is accessible by using a web browser *via* www.protecs.uni-greifswald.de. By using the public account, one can access data that have already been published.

user'' for publicly available data) and restricted access for unpublished data are possible (Fig. 1).

3.1.2 Data access rules

A fine-grained role-dependent scheme of access rights separates data of several work groups. For every user profile, individual views on the data and rights (parameter and topic sheets) can be established.

3.1.3 Search and navigation

Stored information can be searched by using a search form (see Supporting Information Fig. S2). Queries can be restricted to single organisms or even strains. Experimental data (access *via* the cyan colored field in the navigation bar) are structured in a hierarchy that can be used to navigate, *e.g.* from an experimental setup (cultivation) to a specific sample or replicate.

3.1.4 Data import

Parsers import data from various sources (EMBL, GenBank UniProt) into the corresponding database sections.

Due to the flexible Caché-based data model that can be changed in runtime additional data tables referring to gene names/gene IDs can be imported directly *via* a web interface. It supports binary (+/−; 0/1; yes/no), numeric (1; 2.1; 3.7), string lists (intracellular; membrane bound; extracellular), internet links or free text (*e.g.* gene ontology definitions) and simplifies the extension of the data model for annotations.

2-D gel-based proteomic data can be imported by using the Delta2D (DECODON, Greifswald, Germany [29]) project upload wizard which associates gel image files, spot consensus data, spot quantities and spot identification labels with corresponding experimental lab book information. Similar tools are available for data import from MELANIE II (GeneBio, Geneva, Switzerland [40]) and Decyder (GE).

3.1.5 Data export

Raw data (gel/array images, primary quantitative data) can be downloaded in *.gel or *.tiff formats. The export features also include the extraction of array layout files as well as raw 2-D gel images in combination with xml-based spot boundary and spot identification files. Delta2D project structures for 2-D-based proteome analysis containing raw gel image data and derived information can be directly retrieved from Protecs for further analysis. For the export of gene expression data to third party statistics packages, Protecs supports html tables and tables in *.csv file format. For a data export in ''minimum information about a proteomics experiment'' (MIAPE) [41] or ''minimal information about microarray experiments'' (MIAME) [42] compliant formats, an appropriate JAVA program was developed (usage is possible upon request).

3.1.6 Storage of sequence data

Protecs includes an archive for sequence data and sequence-based gene/protein parameters, allowing the direct, integrative interconnection of sequence- and non-sequence-based information. *Staphylococcus aureus* COL DNA sequence data in Protecs are based on the current GenBank chromosome and plasmid entries [43] including start and end positions of sequence features, gene (product) names and comments. Open reading frames predicted by TIGR or UniProt that were missing in or differing from the GenBank prediction were added manually (*e.g. clpC*). Protecs presents genome sequence-related information on the gene and protein pages as shown in Fig. 2.

3.2 Specific database features

3.2.1 Codon adaptation index

The sequence-derived codon adaptation index (CAI) determines how well a gene sequence concurs with the codon usage of an organism. CAI values have been calculated on the basis of all genes or only the genes of the translational machinery as reference sets by using the algorithms favored by Eyre-Walker [44, 45] and Sharp and Li [46, 47].

Genes that play an important role in the survival and adaptation of an organism are likely to have a high CAI to ensure rapid translation in response to frequently occurring stimuli. Analysis of the top 100 genes with respect to CAI in *S. aureus* revealed that most of the corresponding gene products are involved in protein synthesis (Fig. 3). As expected, many of these genes encode ribosomal proteins. Other genes are related to fermentation and energy metabolism. For instance, three of four pyruvate dehydrogenase genes were found ($CAI_{PdhA} = 0.79$, $CAI_{PdhB} = 0.83$, $CAI_{PdhC} = 0.76$ was not among top 100, $CAI_{PdhD} = 0.81$). Genes encoding pyruvate formate lyase ($CAI_{PflB} = 0.84$), acetate/propionate kinase ($CAI_{AckA} = 0.8$) and, interestingly, tagatose 1,6-bisphosphate aldolase ($CAI_{LacD} = 0.83$) also show very high CAIs. Notably, virulence factors such as SsaA ($CAI_{SsaA} = 0.8$) and AgrD ($CAI_{AgrD} = 0.82$) are also encoded by highly codon-adapted genes (CAIs are given according to Eyre-Walker; ribosomal proteins served as reference set).

3.2.2 Replication-associated gene dosage

During exponential growth and ongoing replication, a part of the *S. aureus* genome may appear di- or

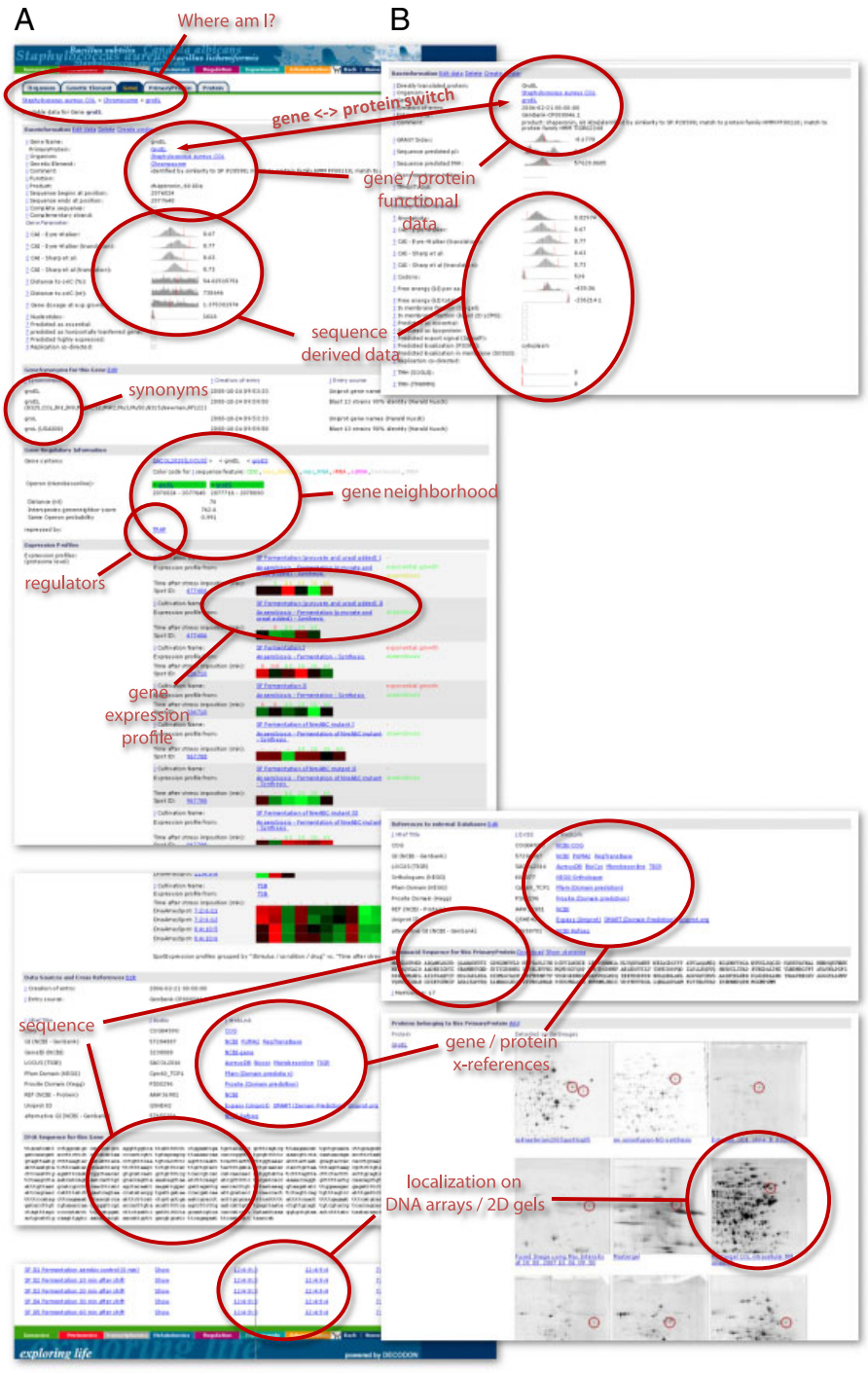


Figure 2. Protecs Gene/Protein Entry. Each gene/protein entry in Protecs contains a general gene/protein description, a variety of sequence derived parameters, regulatory data, several cross-references (x-references) to other data sources and the DNA/protein sequences. Columns A and B distinguish gene- (A) and protein-related (B) data.

polyploidicly [48]. The copy number of genes depends on their distance to the chromosomal origin of replication. The absolute and relative distance between oriC and terminus of replication as well as the number of gene copies is given at Protecs' gene page.

3.2.3 Additional gene/protein parameters

Zhang and Lin [49] and Zhang *et al.* [50] summarized data from gene essentiality analyses [51, 52] in DEG, the database of essential genes, whose information for *S. aureus* was integrated into Protecs along with data on horizontally

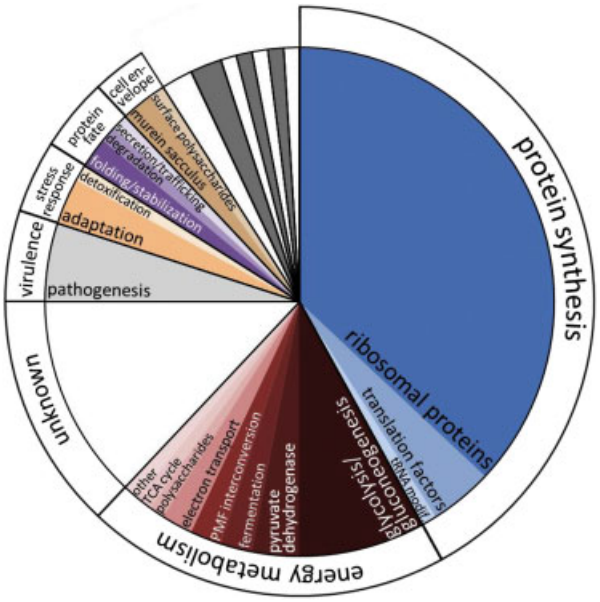


Figure 3. Codon adaptation index (CAI). Pie chart summarizes the functional categories of the top 100 highest codon-adapted genes in *S. aureus*. CAI calculation is based on Eyre-Walker. Functional categories are related to the TIGR annotation (http://www.tigr.org) and were adapted manually.

Table 2. Regulon references

Author	Regulon	PMID	Reference
Bischoff	SigB	15205410	[66]
Cassat	Agr; SarA	17005987	[67]
Cirz	LexA	17085555	[68]
Dunman	Agr; SarA	17384184	[69]
Herbert	GraRS	17676995	[70]
Hubscher	FemAB	17784943	[71]
Korem	TRAP	16177293	[72]
Kuroda	VraRS	12864861	[73]
Liang	SaeRS	16861653	[74]
Liang	ArlRS	16030243	[75]
Luong	MgrA	16484201	[76]
McCallum	TcaR	15126456	[77]
Michel	ClpP-influenced genes	16885446	[78]
Pane-Farre	SigB	16644280	[79]
Rogasch	SaeRS	17079681	[80]
Said-Salim	Rot	12511508	[81]
Sobral	MurF	17194794	[82]
Yang	CidR	16573694	[83]

transferred genes [10]. Probably, highly expressed [53] and co-directionally replicated and transcribed genes are tagged correspondingly.

Some only protein-related parameters were included into Protecs, *e.g.* GRAVY [54] and aromaticity index [55] as well as the predicted p*I* [56, 57] and molecular weight. They are useful for a theoretical prediction whether and where a protein will appear in a 2-D gel [58]. Furthermore, the number of codons, free energy (kJ) *per* amino acid and probable number of transmembrane helices calculated by SOSUI [59] and TMHMM [60] are shown. Tags indicating whether a protein was found in the membrane fraction (based on 1-D SDS gels or on liquid 2-D LC-MS) (Becher, personal communication), lipoprotein prediction data, predictions of export signals (SignalP) [61], predicted localizations in a defined cell compartment by PSORT [62, 63] and predictions whether a protein is localized in the membrane by SOSUI complement the protein information (Fig. 2).

To support interpretation of distinct, especially numeric parameter values of genes/proteins (*e.g.* CAI, gene dosage, GRAVY, p*I*, MW, *etc.*), a histogram-based representation of the entirety of all values with the highlighted individual value was implemented. This was done according to Tufte's [31] micro charts approach and is illustrated in Supporting Information Fig. S3. Micro charts have been developed to visualize data within text lines. For this purpose, a chart has been simplified to the minimal set of information, which is necessary for understanding the data. The presentation of comprehensive histograms enables the recognition of single gene/protein parameters in the context of the entirety of these parameters for all genes/proteins.

3.2.4 Gene and protein regulatory data

Gene and protein pages have been complemented by information on gene neighborhood (Fig. 2) and predicted association with operons (VIMSS [64, 65]), including intergenic distance, interspecies gene neighborhood score and the probability of an adjacent gene pair to be part of the same operon. Data from 32 global transcriptome and proteome studies (Table 2) were reviewed and included.

3.2.5 X-references

A general link constructor may cross-reference each gene and protein entry to a variety of other specialized or general data sources (Tables 3 and 4, x-references for *S. aureus* enolase gene/protein–*eno*/Eno).

3.2.6 Gene expression data

Expression data are available on the gene and protein pages (Fig. 2) as well as on the experimental design pages, so that likewise the expression of a gene of interest under a variety of stimuli or conditions or similar expression profiles within the same experimental context can be easily studied.

The data sets were not pre-filtered and are fully accessible and searchable by using tools integrated in the database, enabling unlimited searches in the whole data content without restrictions due to subjective or biased factors. For a more detailed analysis, raw expression data can be extracted from Protecs and analyzed with external tools (see Section 3.1.5).

Table 3. Gene x-references for *eno*

XRef Title	AccNo	WebLink
COG	COG0148G	COG
EC fourth level point	4.2.1.11	Biocyc; BRENDA; EBI-Intenz; EBI-PDB; ERGO; Expasy-NiceZyme; ExplorEnz; KEGG; NIST; PUMA2; UMBBD
EC fourth level slash	4/2/1/11	IUBMB
EC third level point	4.2.1	EBI-Intenz; EBI-PDB
GI (NCBI – GenBank)	57284304	NCBI; PUMA2; RegTransBase
GeneID (NCBI)	3238461	NCBI-gene
LOCUS (TIGR)	SACOL0842	AureusDB; Biocyc; Microbesonline; TIGR
Pfam Domain (KEGG)	Enolase_C	Pfam (Domain prediction)
Prosite Domain (Kegg)	PS00164	Prosite (Domain prediction)
REF (NCBI – Protein)	AAW36398	NCBI
Uniprot ID	Q5HHP1	SMART (Domain Prediction); Uniprot
Alternative GI (NCBI – GenBank)	57650118	NCBI-Refseq

Table 4. Protein x-references for Eno

XRef Title	ExtId	WebLink
COG	COG0148G	NCBI-COG
EC fourth level point	4.2.1.11	BioCyc; Brenda; EBI-IntEnz; EC-PDB; ERGO-lite; Expasy-NiceZyme; ExplorEnz; KEGG; NIST; PUMA2; UMBBD
EC fourth level slash	4/2/1/11	IUBMB
EC third level point	4.2.1	EBI-Intenz; EBI-PDB
GI (NCBI - GenBank)	57284304	NCBI; PUMA2; RegTransBase
LOCUS (TIGR)	SACOL0842	AureusDB; BioCyc; Microbesonline
Orthologues (KEGG)	K01689	KEGG-Orthologue
Pathway (KEGG)	sac00010	KEGG-Pathway
Pfam Domain (KEGG)	Enolase_C	Pfam (Domain prediction)
Prosite Domain (Kegg)	PS00164	Prosite (Domain prediction)
REF (NCBI – Protein)	AAW36398	NCBI
Uniprot ID	Q5HHP1	SMART (Domain Prediction); Uniprot
Alternative GI (NCBI – GenBank)	57650118	NCBI-Refseq

Gene/protein expression data are presented in different ways. One of them is numeric tables (not shown) of the original expression data. These tables are useful for the export, *e.g.* to GeneSpring (Agilent) or in plain text format for further analyses. Additionally, Protecs supports heat map (A) representation (Supporting Information Fig. S4) of standardized expression levels in a three-color gradient (green: relatively low expression; black: near average expression level; red: high expression). To avoid misinterpretations because of insufficient perception of color shades, the same data can be displayed using Tufte's [31] micro charts (B). Spots of interest can be shown within their corresponding gel image tiles (C).

3.2.7 Interpretation of gene expression profiles

To search similarly expressed genes/proteins, a correlation-based approach works on standardized expression values and can be initiated from the gene expression pages by defining a template expression profile (hyperlink ''find similar''). The resulting expression visualization starts with genes/proteins that correlate best with the template of interest and ends with inversely correlated entries. Selected parts of this list can be inspected directly by using the KEGG pathway mapper [84, 85] or the STRING co-occurrence tools [86, 87].

3.2.8 Archival storage of lab book data

The storage of experimental setups is structured with a four-level hierarchy. For bacterial cultures, the levels are defined as follows but can be adapted for other biological objects of interest: The organism/biological object under investigation is represented within the 1st level. The home project (2nd level) describes the scientific context the experiment was designed for. The cultivation (3rd level) originates from exactly the same bacterial culture. Sample points (4th level) represent the result of a sampling process at a defined time point and during an acting stimulus or treatment (Supporting Information Fig. S5). For data storage, we implemented the core experimental variables listed in Table 5.

3.2.9 Generation of expression profiles from expression data

Gel-by-gel variation causes difficulties with spot to spot matching. We and others introduced dual channel imaging in combination with image warping [88–90] to compensate for positional fluctuations within the spot pattern. By introducing fusion images [32] and a project-wide spot consensus [29, 91, 92], we were able to solve the spot matching problem for related gel batches (Fig. 4).

For expression values from the same experiment/cultivation (*e.g.* Fig. 4, cultivation A), we are able to construct complete expression profiles. For temporally distant experiments (Fig. 4, comparison of cultivation A with B), problems with spot matching may persist. In such cases, a compensation of spot pattern variations by using image warping sometimes is not possible. Alternatively, we applied expression profile matching from different cultivations by protein identification. This unique feature of Protecs is illustrated in Fig. 4. For most identified protein spots, a complete inter-experimental profile was recorded.

Sorting of expression profiles by similarity in inter-experimental data sets highlights inversely expressed spots originating from the same protein. This is because profiles of the same protein are consecutively grouped in the expression table, and inverse profiles are dominated by correlating profiles of the same protein.

Table 5. Available experimental features for bacterial omics experiments in Protecs

Field	Explanation
Organism	Biological system under investigation
Home Project	Scientific contextually related data sets
Description	Free text explanation of home project
Cultivation	Belong to home projects and represents the bacterial culture
Author	Experimenter
Date	Date of experiment
Description	Free text explanation of cultivation
Genotype	Identifier
Medium	Growth medium
Strain/cell line	Identifier
Sample point	Sampling event
Date	Date of sampling event
Author	Experimenter who took the sample
OD	Value
OD-wavelength	Wavelength of OD measurement
Temperature	Value
Stimulus	Causes gene expression changes
Time after stimulus	Value
Growth phase	Identifier

3.3 Applied example: Physiology of anaerobiosis in *S. aureus*

Previously, adaptation of *S. aureus* COL to anaerobic conditions was investigated by using a proteomic approach [23]. In that study, 1073 different protein spots had been detected on the gels and their expression profiles had been determined. The synthesis of 357 protein spots was increased under fermentation conditions, while the synthesis of 478 spots was repressed. However, only 66 of the induced protein spots (18%) and 94 of the repressed protein spots (20%) had been identified until then.

Figure 4. Inter-experimental expression profile comparison. Inter-experimental expression profiling in gel based proteomics is often faced with the problem of non-removable image distortions especially for temporally distant gel images. Therefore, some expression profiles can be partially missing in multi-experiment setups (see Pkg, the profile of the last sub profile in experiment A; see Eno, the last sub profile in A).

Figure 5. Metabolic adaptation after oxygen limitation. Several missing expression profiles were added from the current master gel (underlined protein names). SdhB and SdhC are out of the analytical window. AckA, acetate kinase; AcnA, aconitate hydratase; Acs, acetyl-CoA synthetase; BudA1, α-acetolactate decarboxylase; BudB, acetolactate synthase; Ddh, D-lactate dehydrogenase (SACOL2535); Eno, enolase; FbaA, fructose-bisphosphate aldolase; FumC, fumarate hydratase; GapA1, glyceraldehyde-3-phosphate dehydrogenase; GlcK, glucokinase; GltA, citrate synthase; Icd, isocitrate dehydrogenase; IlvB, acetolactate synthase, large subunit, biosynthetic type; Ldh1, L-lactate dehydrogenase; Ldh2, L-lactate dehydrogenase; Mqo1, malate:quinone oxidoreductase; Mqo2, malate:quinine oxidoreductase; PckA, phosphoenolpyruvate carboxykinase; PdhA, pyruvate dehydrogenase complex E1 component, subunit; PdhB, pyruvate dehydrogenase component E1; PdhC, pyruvate dehydrogenase component E2; PdhD, pyruvate dehydrogenase component E3; PfkA, phosphofructokinase; PflB, pyruvate formate lyase; Pgi, glucose-6-phosphate isomerase; Pgk, phosphoglycerate kinase; Pgm, phosphoglycerate mutase; Pta, phosphate acetyltransferase; Pyc, pyruvate carboxylase; Pyk, pyruvate kinase; SACOL0111, acetoin reductase; SACOL0135, alcohol dehydrogenase; SACOL0241, alcohol dehydrogenase, zinc-containing; SACOL0660, alcohol dehydrogenase; SACOL1749, malic enzyme, flavoprotein chain; SACOL2301, formate dehydrogenase, subunit, putative; SdhA, succinate dehydrogenase; SdhB, succinate dehydrogenase, iron-sulfur protein; SdhC, succinate dehydrogenase, cytochrome b558 subunit; SucA, 2-oxoglutarate dehydrogenase component E1; SucB, 2-oxoglutarate dehydrogenase component E2; SucC, succinyl-CoA synthetase, subunit; SucD, succinyl-CoA synthetase, alpha subunit; TpiA, triosephosphate isomerase.

In the present study, the number of identified spots was increased and spot quantitation was improved by using current algorithms of Delta2D 4.0 for gel fusion, spot detection and intensive spot editing. Altogether, 1340 protein spots were detected. By using a recent reference map (master gel) containing 677 identified protein spots (unpublished data), 599 spot identifications were successfully transferred. Transfer of further 25 spot identifications is indicated as "under reserve" due to high gel to gel variation in the corresponding area (in Supporting Information Tables S1–S5, spot labels contain question mark).

The synthesis of 390 protein spots was increased during fermentation (2-fold threshold). One hundred thirty eight of these protein spots (35%) were identified and represent 110 different proteins The synthesis of 583 spots was repressed. Of these protein spots, 176 (30%) were assigned to identified spots on the master gel and represent 154 different proteins. Summarizing this, in the present study, we assigned 179 additional proteins that were induced (75) or repressed (104) after oxygen limitation (Supporting Information Tables S9 and S10).

3.3.1 Newly assigned expression profiles complement data of metabolic pathways

The newly identified proteins whose synthesis is regulated in dependence of oxygen availability provide deeper insights into the adaptation processes to anaerobic conditions (Fig. 5A). For instance, additional expression data for glycolytic enzymes, the pyruvate dehydrogenase complex and the tricarboxylic acid (TCA) cycle are now available: component 1 of the pyruvate dehydrogenase complex (PdhA), glucose-6-phosphate isomerase (Pgi) and fructose-bisphosphate aldolase (FbaA), components of the 2-oxoglutarate dehydrogenase. Most interestingly, the E1 (SucA) and the E2 subunit (SucB) of 2-oxoglutarate dehydrogenase were induced under anaerobic conditions in *S. aureus*. With respect to fermentation pathways, four additional enzymes were identified (Fig. 5B). Accordingly, formation of D-lactate should be increased under oxygen limited conditions as the synthesis rate of D-lactate dehydrogenase (Ddh) was more than 20-fold higher. Moreover, a zinc-containing alcohol dehydrogenase (SACOL0241) was found among the anaerobically induced proteins. In contrast, the synthesis of acetoine reductase (SACOL0111) as well as phosphate acetyltransferase (Pta) remained unchanged under these conditions.

3.3.2 Protecs opens up new opportunities for intra- and inter-experimental analysis

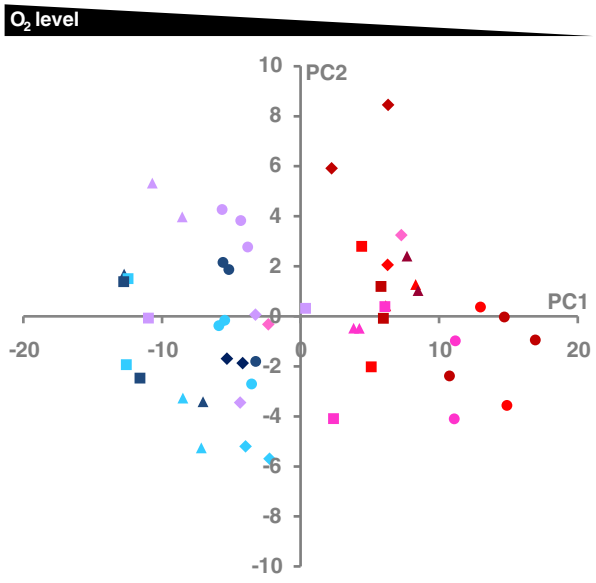
The response of *S. aureus* to anaerobic growth conditions was analyzed considering three different aspects: (i) the influence of nitrate as an alternative electron acceptor, (ii) the relevance of NreABC for global protein synthesis under anaerobic conditions and (iii) the influence of pyruvate and uracil on anaerobic adaptation. Hence, the anaerobiosis project consisted of four experiments: the wild-type grown anaerobically (i) in the presence and (ii) in the absence of nitrate, (iii) in the presence of uracil and pyruvate, and (iv) an *nreABC* mutant grown anaerobically in the presence of nitrate. All relevant experimental parameters, gel images, protein identifications and quantitation data of the different sampling points were imported into Protecs.

3.3.3 Inter-experimental sample analysis

Individual experiments were linked inter-experimentally *via* protein identifications. These inter-experimental expression profiles were used for an overall sample analysis by PCA. The resulting PCA plot shows the degree of similarity of all different samples (Fig. 6). Most samples were separated by the first principal component. Samples that were drawn from cultures grown under aerobic and microaerobic conditions are localized in the left quadrants within the 2-D PCA plot, whereas almost all anaerobic samples appear in the right quadrants. PCA clearly suggests that the oxygen level predominantly affects protein expression. In contrast, the addition of nitrate or uracil and pyruvate had no general effect on the synthesis pattern, as the corresponding samples were not separated by PCA.

3.3.4 Intra-experimental sample analysis

Within individual experiments, we screened for protein expression profiles that were significantly affected by oxygen



genotype	supplements	aerobic		increasingly anaerobic			
		0'	60'	10'	20'	30'	60'
wild type	-	▲	▲	▲	▲	▲	▲
wild type	uracil, pyruvate	◆	◆	◆	◆	◆	◆
wild type	nitrate	■	■	■	■	■	■
Δ*nreABC*	nitrate	●	●	●	●	●	●

Figure 6. PCA over anaerobiosis samples. Based on all detected expression profiles, samples of the proteomics meta study have been analyzed by PCA. The first principal component (PC1) can be interpreted as the influence of the oxygen level on protein synthesis pattern. In contrast, there was no further separation in relation to supplements or genotype.

limitation by using one-way ANOVA (TMEV). Subsequently, samples were clustered on the basis of significantly changed expression profiles using Support Trees (TMEV). Accordingly, expression profiles with highest induction rates under aerobic conditions (''aerobic cluster'') and those under anaerobic conditions (''anaerobic cluster'') could be clearly distinguished (Supporting Information Fig. S6).

Proteins of the ''anaerobic cluster'' belong to very similar functional groups (Supporting Information Tables S5–S8). As expected, most of these enzymes are involved in glycolysis and in fermentation pathways.

The aerobic clusters (Supporting Information Tables S5–S8) were also very similar between the individual experiments except for cells grown in the presence of pyruvate and uracil. Proteins with higher synthesis rates under aerobic conditions are mainly involved in the metabolism of nucleotides, amino acids, fatty acids and phospholipids but also in protein repair and degradation, translation and oxidative stress response.

3.3.5 Collecting similar expression profiles by pattern search

Expression profiles of classical anaerobically induced proteins like PflB, which are found in the ''anaerobic expression cluster'' of each experiment described above, can be used as a template for pattern searches for similarly regulated proteins. By this way, proteins of other fermentation pathways, glycolysis, isoleucine/valine biosynthesis pathway as well as regulatory proteins (*e.g.* SrrA) were retrieved. They form the anaerobiosis stimulon. Among proteins with contrary (anti-correlating) expression profiles (repressed under anaerobic conditions), we found several ribosomal proteins, proteins involved in nucleotide metabolism and the anti sigma factor RsbW that prevents *SigB*-directed transcription.

3.3.6 Fold change analysis of selected expression profiles

A detailed comparison of the different expression patterns based exclusively on a fold change threshold of 2.5 was performed (Supporting Information Tables S1–S8). As expected, synthesis rates of several protein spots representing fermentation enzymes such as alcohol dehydrogenase SACOL0135, formate acetyltransferase PflB and L-lactate dehydrogenase Ldh1 were highly increased in response to oxygen limitation. Interestingly, in the presence of uracil and pyruvate, the observed anaerobic induction for SACOL0135 and Ldh1 was clearly diminished. A similar phenomenon was observed for the zinc-containing alcohol dehydrogenase SACOL0660 (Supporting Information Table S2). In contrast, expression of formate acetyltransferase is increased when pyruvate and uracil are present under anaerobic conditions. Ldh2 synthesis was only increased under anaerobic conditions in the wild-type when pyruvate and uracil were not present (Supporting Information Table S3) and in the *nreABC* deficient mutant in the presence of nitrate.

Expression of the glycolytic operon encoding its own regulator GapR, glyceraldehyde 3-phosphate dehydrogenase (GapA1), phosphoglycerate kinase (Pgk), triosephosphate isomerase (TpiA), phosphoglycerate mutase (Pgm) and enolase (Eno) was anaerobically induced; however, the induction of several of the corresponding protein spots failed in the presence of pyruvate and uracil.

4 Discussion

4.1 Protecs in the environment of federated databases

Assuring public data accessibility is one of the major challenges in omics research. For genomic data, this has been realized with general web accessible resources (*e.g.* GenBank).

In the world of microbial model organisms, many organism-specific resources providing DNA sequence information have been set up. Among them are databases of the GenoList family (currently containing the original data sets from the sequencing efforts of, *e.g. B. subtilis* and *S. aureus*) [11], several resources at TIGR (www.JCVI.org) and Microbes Online (genomic and operon prediction information) [64].

A couple of these sequence-oriented resources also provide access to gene expression data. Exemplarily ''Gene expression omnibus'' [93] complements GenBank data at the NCBI and the array express (http://www.ebi.ac.uk/microarray-as/ae/ [94]) at the EBI. Very recently, Microbes Online also started to manage gene expression data of microorganisms. Transcriptome data became available especially due to the efforts of the data standardization consortium for microarray data, which culminated in the definition of the ''minimal information about microarray experiments'' standard [42].

Recent efforts, reviewed in *Nature Biotechnology* [41], reflect the work of the MIAPE consortium for the definition of standards in the proteome software and proteome experiment world. The protein mass spectrometric data standard definition is almost complete. Standards for proteome experiments and gel-based proteomics are still under discussion. For this reason, the OpenProteomicsDB (http://bioinformatics.icmb.utexas.edu/OPD/ [95]) mainly stores and maintains MS-based proteomic data sets in appropriate data formats. For the 2-D protein gel world, a few databases are available, which are storing 2-D gel associated spot identification data, among them Swiss2DPage [96], 2Dbase [97], Proteome 2D Page DB [98] and others. All

of them use proprietary data formats. For a portal of 2-D gel-based proteome databases, see http://world-2Dpage.expasy.org/ [99].

Standards for the metabolome world have been discussed and partly realized within the metabolomics standards initiative (http://msi-workgroups.sourceforge.net/) or systems biology standards initiative [100, 101]. These standards helped to setup the databases of the MetaCyc family whose data sets rely on genomic information and computer generated metabolic networks [102].

Wikis are a relatively new type of data repositories that were successfully introduced with Wikipedia – the online encyclopedia (www.wikipedia.org). Wikis are open systems and depend on the commitment of the public. Whether this concept is suitable for species-related data collections will be known for *Escherichia coli* (ecoliwiki.net) and *B. subtilis* (http://www.subtiwiki.uni-goettingen.de/) in the near future.

A summary view of data resources dealing with *S. aureus* data can also be found on the organism homepages of Protecs (Table 6).

As mentioned before, Protecs is for (i) archiving, (ii) managing, (iii) linking and (iv) analyzing omics data collections and for keeping them in an online available status. Furthermore, Protecs provides tools for data updating, interpretation, visualization and integration.

Protecs is used for organization of data in functional genomics projects related to human pathogenic fungi (Omnifung, http://www.omnifung.hki-jena.de/csp/protecs/login.csp [103–105] and marine microorganisms (Marine2D, http://www.marine2d.de/).

Within Protecs, organism-specific sequence data form the basis of any expression data collection. *Via* experimental information (lab book information storage) omics expression data (transcriptome and 2-D gel-based proteome) are linked to the organism's genome. Protecs' strength, the tight integration of heterogeneous data, makes it possible to complement any gene or protein entry with expression profiles of this gene/protein under all experimental conditions that have been analyzed and stored in Protecs so far. *Vice versa*, Protecs also supports the analysis of similarly expressed genes/proteins within one experiment by using a correlation-based grouping of expression profiles. Compared with other databases, only Protecs was able to generate quantitation tables based on spot identification that cover multiple experiments, which was the basis for the presented analysis.

For this study, different experiments were complemented with up-to-date identification data for 677 protein spots ensuring that the naming of all spots and numbering of multiple spots were consistent. Comparable experimental conditions and a similar protocol for radioactive protein labeling simplified analyses.

Table 6. *S. aureus* data resources

Resource	Content
AUGUR	LPXTG motifs
AUGUR	Signal peptides
AureusDB	Protein sequence information
BioCyc	Biochemical pathways
BONSAI	Terminator prediction
CBS	Genome information
DBD	Transcription factors
EBI	Gene ontology annotation
EFICAZ	Enzyme classification
EMBL	Genome information
G-tRNA-DB	tRNAs and codon usage
GECO	Genome information
Genolist	Genome information
HGT-DB	Horizontally transfered genes
IMG	GENOME information
KEGG	Biochemical pathways
MicrobesOnline	Genome information
NCBI	Genome information
NMPDR	General information
Pedant	Genome information
PSORT2.0	Localization prediction (file)
Pubmed	Genome publication
Pubmed	Strain specific literature
TIGR	Genome information
VIMSS	Operon prediction

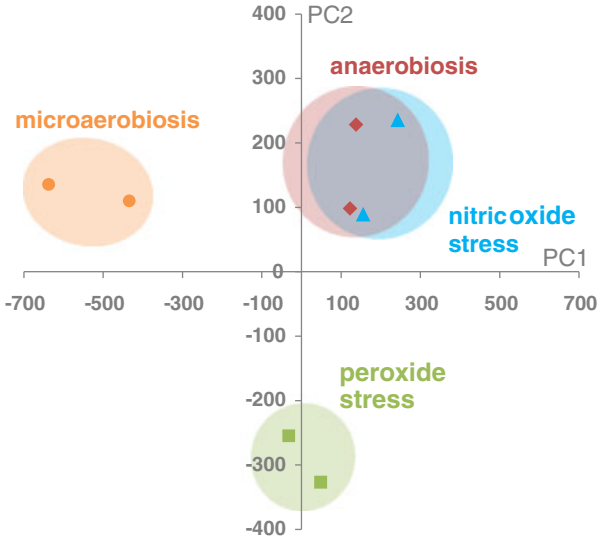


Figure 7. Global sample analysis for four different stimuli. Inter-experimental sample analysis can be used to classify new stimuli under investigation. Samples of *S. aureus* under 10 min anaerobiosis (orange – microaerobiosis [23]) and 60 min anaerobiosis (red [23]), 10 min nitric oxide stress (cyan [19]) and 10 min peroxide stress (green [22]) were analyzed by PCA. The plot demonstrates that samples related to anaerobiosis and nitric oxide stress were grouped together, whereas the other samples were clearly separated due to the stimulus they were exposed to.

4.2 A global view on *S. aureus* anaerobiosis physiology supported by Protecs

In the present study, we clearly demonstrated that improved analysis methods (spot detection, statistical analysis) help to achieve a new quality in data interpretation for already existing proteomic experiments. In particular, technical advances in MS resulted in an increased proteome coverage with many more identified protein spots. This provides a deeper understanding of already existing experimental data from bacteria that had been grown under defined conditions by assigning expression profiles of so far unidentified proteins to existing maps of metabolic pathways and regulatory networks.

Global data reanalysis was performed for the adaptational network of *S. aureus* triggered by anaerobic conditions. For that purpose, it was necessary to store experimental data in accessible and updatable formats as supported by Protecs. New spot identification data were provided by our current master gel. Unassigned spots (see results) may be explained due to the comparison of ^{35}S methionine-labeled protein synthesis patterns (snap shot of current protein synthesis) with our fluorescence stained master gel, which represents cumulative protein amounts until the sampling event. Among newly identified spots, we found the E2 subunit of the 2-oxoglutarate dehydrogenase that is induced under anaerobic conditions in *S. aureus*. This is in contrast to *E. coli*, where this enzyme is regulated by ArcAB and whose synthesis is strictly confined to high levels of oxygen [106].

Using the inter-experimental comparison approach supported by Protecs, global analyses of samples of different experimental origins became possible. The relatedness of different experiments can be easily visualized by PCA. The previously supposed similarity [19] between protein synthesis patterns of cells grown under fermentative conditions and those grown anaerobically in the presence of nitrate was impressively shown by this technique (Fig. 6). Moreover, we have demonstrated that the presence of uracil and pyruvate

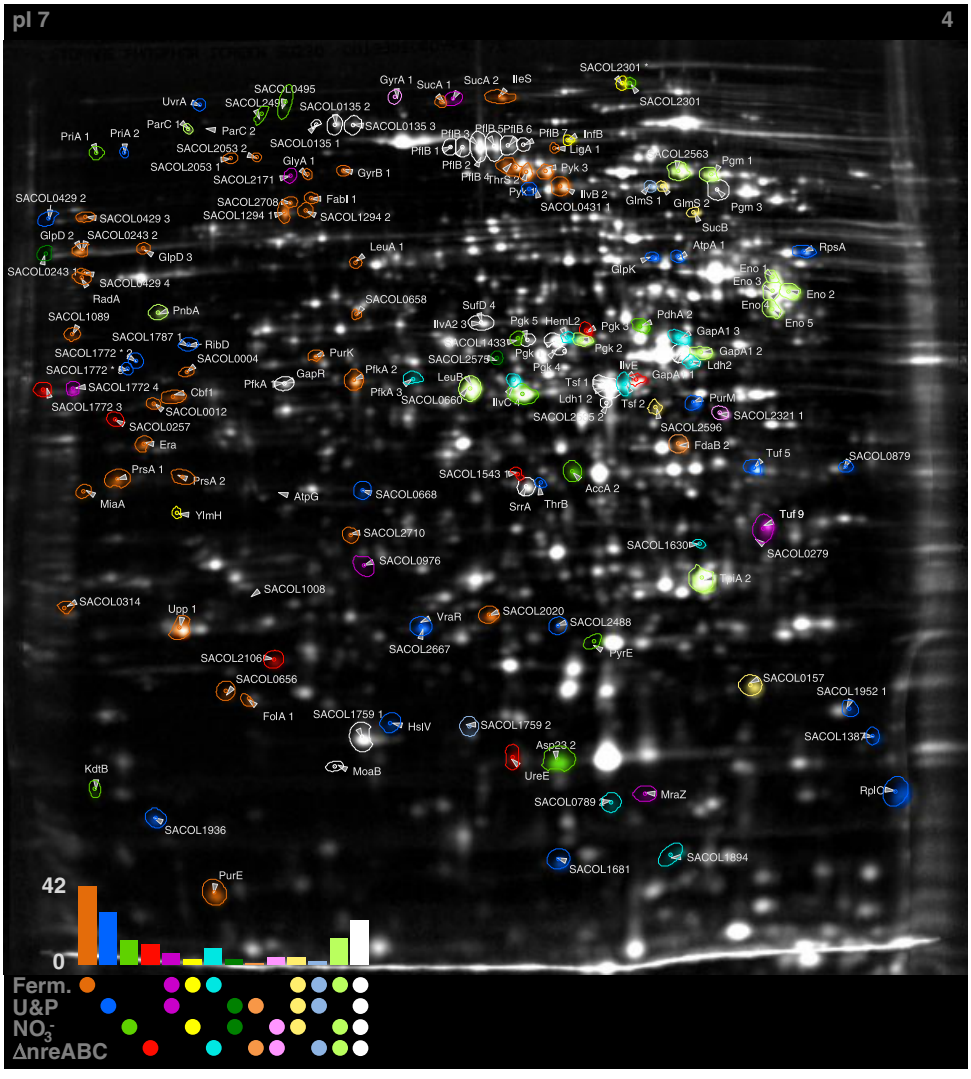


Figure 8. Summary of inter-experimental protein expression analysis. A union fusion gel was created based on all gels of the anaerobiosis project by using Delta2D software. Spots whose synthesis was at least 2.5 times induced within one or more individual experiment(s) were color-coded (see color legend). Additional numbers distinguish multiple spots of the same protein. Spots that were identified only once are indicated by an asterisk (*). Spot colors are explained in the color code table: circles represent occurrence of induction during fermentation at the absence (Ferm.) or presence (U&P) of uracil and pyruvate, nitrate respiration (NO_3) and anaerobic growth of the nreABC mutant in the presence of nitrate (Δ*nreABC*). Bars represent numbers of spots induced within the corresponding experiment(s). Spots were labeled according to the identification based on the TIGR annotation (http://www.tigr.org).

had only minor influence on the general response of *S. aureus* to oxygen limitation.

Global data analysis can benefit from collections of an increasing number of well prepared and analyzed proteomic experiments stored and made accessible by Protecs. In the future, such a library of proteomic experiments may be useful to classify protein patterns of cells that are grown under defined or even undefined conditions. Comparative pattern analysis with the still available anaerobiosis data may help to pinpoint common molecular signals and adaptation strategies, *e.g.* for bacteria grown in eukaryotic cell cultures or in tissues of the host.

Moreover, the consequent use of this library can provide new insights into the mode of action of new drugs or stimuli simply by matching drug-influenced expression profiles with similar protein expression profiles from well-characterized conditions. The following example illustrates this approach. Recently, NO stress in *S. aureus* has been shown to induce proteins involved in fermentation pathways rather than proteins involved in oxidative stress response [19, 22, 107].

By using PCA to compare protein expression profiles for oxygen limitation, peroxide stress and NO stress, this observation, derived from a very detailed and labor-intensive data analysis, can be confirmed. While in the PCA plot protein expression data for anaerobic conditions and NO stress are closely related to each other, protein expression data for peroxide stress are clearly separated by the second principal component (Fig. 7). This impressively shows that PCA represents indeed a suitable tool for a very rapid comparison of protein expression profiles of a given experiment with that of already existing experiments to find similar adaptational networks.

A subsequent more detailed analysis can then scrutinize the components that determine the similarity or the difference of these experiments with the ultimate aim to define common cellular signals. The more detailed analysis of all protein expression profiles from anaerobic conditions and NO_3^- stress revealed a set of proteins that were commonly induced (Fig. 8, white borders). This set of proteins includes the alcohol dehydrogenase SACOL0135, formate acetyltransferase PflB, L-lactate dehydrogenase Ldh1 and the response regulator SrrA. Interestingly, the corresponding genes are preceded by a conserved sequence motif [23, 78], which is similar to the binding site of Rex of *Streptomyces coelicolor* [107]. The role of the Rex regulator in *S. aureus* was previously unknown. Consequently, we started a detailed functional analysis of Rex in *S. aureus*. At present, we know that Rex is the main regulator for anaerobic gene expression that directly and indirectly represses a large set of genes involved in fermentation and anaerobic respiration by measuring the NADH pool within the cells [108]. This finding in turn explains the observation that genes involved in adaptation to anaerobic conditions were induced in *S. aureus* in response to NO stress. NO is known to preferentially act on cytochromes. Accordingly, high concentrations of NO induce a loss of respiration activity and a sudden increase of NADH, which deactivates Rex.

4.3 Codon adaptation of anaerobiotically genes and similarly regulated genes

A high adaptation of a gene to the preferred codon usage of an organism ensures rapid translation. Many genes whose products are involved in energy metabolism show high CAI values in *S. aureus*. Interestingly, genes encoding the pyruvate dehydrogenase and fermentation enzymes such as PflB are both among the 100 best-codon usage adapted genes. The pyruvate dehydrogenase connects glycolysis with the TCA cycle by converting pyruvate to acetyl-CoA accompanied by the reduction of NAD^+. This explains its importance and exclusive expression under aerobic conditions. The highly oxygen-sensitive PflB also uses pyruvate to initiate the mixed acid fermentation playing an important role during anaerobic growth. Rapid synthesis of both enzymes is probably necessary for *S. aureus* to survive on the one hand at oxygen-rich sites of mucous membranes and on the other at oxygen-limited areas within wounds and abscesses.

4.4 Conclusion and outlook

We demonstrated that Caché/Protecs is able to integrate sequence-derived data, lab book and gene expression data from the proteome level. Additionally, data from transcriptome and metabolome can be handled similarly. By using Protecs' ability to link inter-experimental expression profiles, we performed a combined analysis of independent anaerobiosis experiments with *S. aureus* COL.

From this analysis, we were able to complement gaps in anaerobiosis pathways by using expression data of newly identified protein spots. Furthermore, we were able to find that anaerobically expressed genes show high codon adaptation indices. While the current work was interested in the application of Protecs to ^{35}S-methionine-labeled protein patterns that reflect the synthesis rate of anaerobically induced proteins, we think about complementation of these data on protein level (staining) and protein stability (pulse chase labeling). Furthermore, Protecs should be able to help with a global integrated study on proteome, transcriptome and metabolome data.

Our study introduced Protecs as a functional genomics data repository. As a commercial product, Protecs can be purchased for local installation. Our freely accessible reference installation for *S. aureus*-related data can be found on the website www.protecs.uni-greifswald.de. For this Protecs reference instance, we invite members of the scientific community to provide us their omics data for integration, comparative analysis and online presentation.

By using Protecs, metastudies are not only limited to *S. aureus* but the usage of Protecs is also suitable for other organisms. Besides *S. aureus*, meanwhile, we have accumulated expression data for *B. subtilis*. Other researchers use Protecs for proteome data storage of human pathogenic fungi and marine microorganisms. With ongoing data integration and the continuously growing amount of expression data on a variety of omics fields, we hope to establish and to improve Protecs as a powerful library of proteome and other omics experiments that will decisively support functional genomics studies in the future.

This work was supported by grants of the German Research Foundation (Deutsche Forschungsgemeinschaft) in the framework of the Transregional Collaborative Research Initiative "Pathophysiology of Staphylococci in the Post-Genomic Era" TransRegio34 to M. H. (A1, C2); T. D. (A5); S. E. (B4) and J. B. (Z1) and by grants of the German Federal Ministry of Education and Research within the collaborative network MIMAS "Microbial Interactions in marine systems" to M. H. Additionally, the authors thank DECODON GmbH (Greifswald) for supporting them with current versions of Delta2D and with the Protecs system and for ongoing help and improvements. They are grateful to Professor Götz kindly providing the nreABC mutant. Furthermore, they thank Riad Schulz and Sebastian Baronick for their contribution during the development of the MIAPE compliant data export tool and all members of the Hecker group for their fair comments during Protecs adaptation and development. Last but not least, the authors are grateful to Holger Kock for proof reading the manuscript and to the editors and reviewers for their fair and helpful comments.

The authors have declared the following conflict of interest: M. B. and J. B. are working with DECODON, which supplies Protecs.

5 References

[1] Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., Sayers, E. W., GenBank. *Nucleic Acids Res.* 2009, *37*, D26–D31.

[2] Burks, C., Fickett, J. W., Goad, W. B., Kanehisa, M. *et al.*, The GenBank nucleic acid sequence database. *Comput. Appl. Biosci.* 1985, *1*, 225–233.

[3] Giglia, E., New Year, new PubMed. *Eur. J. Phys. Rehabil. Med.* 2009, *45*, 155–159.

[4] Walsh, J. H., PubMed Central: a boon to information access, a bane to quality control, or neither? *Gastroenterology* 1999, *117*, 1040.

[5] Uniprot, C., The Universal Protein Resource (UniProt) 2009. *Nucleic Acids Res.* 2009, *37*, D169–174.

[6] Apweiler, R., Bairoch, A., Wu, C. H., Barker, W. C. *et al.*, UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res.* 2004, *32*, D115–119.

[7] Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. F., Jr. *et al.*, The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* 1977, *112*, 535–542.

[8] Dutta, S., Burkhardt, K., Young, J., Swaminathan, G. J. *et al.*, Data deposition and annotation at the worldwide protein data bank. *Mol. Biotechnol.* 2009, *42*, 1–13.

[9] Garcia-Vallve, S., Guzman, E., Montero, M. A., Romeu, A., HGT-DB: a database of putative horizontally transferred genes in prokaryotic complete genomes. *Nucleic Acids Res.* 2003, *31*, 187–189.

[10] Garcia-Vallve, S., Romeu, A., Palau, J., Horizontal gene transfer in bacterial and archaeal complete genomes. *Genome Res.* 2000, *10*, 1719–1725.

[11] Lechat, P., Hummel, L., Rousseau, S., Moszer, I., GenoList: an integrated environment for comparative analysis of microbial genomes. *Nucleic Acids Res.* 2008, *36*, D469–474.

[12] Galperin, M. Y., Cochrane, G. R., Nucleic Acids Research annual Database Issue and the NAR online Molecular Biology Database Collection in 2009. *Nucleic Acids Res.* 2009, *37*, D1–4.

[13] Hecker, M., A proteomic view of cell physiology of *Bacillus subtilis* – bringing the genome sequence to life. *Adv. Biochem. Eng. Biotechnol.* 2003, *83*, 57–92.

[14] Hecker, M., Engelmann, S., Cordwell, S. J., Proteomics of *Staphylococcus aureus* – current state and future challenges. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* 2003, *787*, 179–195.

[15] Ohlmeier, S., Scharf, C., Hecker, M., Alkaline proteins of *Bacillus subtilis*: first steps towards a two-dimensional alkaline master gel. *Electrophoresis* 2000, *21*, 3701–3709.

[16] Schumann, W., Ehrlich, S. D., Ogasawara, N., *Functional Analysis of Bacterial Genes*, Wiley, Chichester, NY 2000.

[17] Appel, R. D., Bairoch, A., Sanchez, J. C., Vargas, J. R. *et al.*, Federated two-dimensional electrophoresis database: a simple means of publishing two-dimensional electrophoresis data. *Electrophoresis* 1996, *17*, 540–546.

[18] Engelmann, S., Hecker, M., Proteomic analysis to investigate regulatory networks in *Staphylococcus aureus*. *Methods Mol. Biol.* 2008, *431*, 25–45.

[19] Hochgrafe, F., Wolf, C., Fuchs, S., Liebeke, M. *et al.*, Nitric oxide stress induces different responses but mediates comparable protein thiol protection in *Bacillus subtilis* and *Staphylococcus aureus*. *J. Bacteriol.* 2008, *190*, 4997–5008.

[20] Kohler, C., von Eiff, C., Liebeke, M., McNamara, P. J. *et al.*, A defect in menadione biosynthesis induces global changes in gene expression in *Staphylococcus aureus*. *J. Bacteriol.* 2008, *190*, 6351–6364.

[21] Kohler, C., Wolff, S., Albrecht, D., Fuchs, S. *et al.*, Proteome analyses of *Staphylococcus aureus* in growing and non-growing cells: a physiological approach. *Int. J. Med. Microbiol.* 2005, *295*, 547–565.

[22] Wolf, C., Hochgrafe, F., Kusch, H., Albrecht, D. *et al.*, Proteomic analysis of antioxidant strategies of *Staphylococcus aureus*: diverse responses to different oxidants. *Proteomics* 2008, *8*, 3139–3153.

[23] Fuchs, S., Pane-Farre, J., Kohler, C., Hecker, M., Engelmann, S., Anaerobic gene expression in *Staphylococcus aureus*. *J. Bacteriol.* 2007, *189*, 4275–4289.

[24] Schlag, S., Fuchs, S., Nerz, C., Gaupp, R. *et al.*, Characterization of the oxygen-responsive NreABC regulon of *Staphylococcus aureus*. *J. Bacteriol.* 2008, *190*, 7847–7858.

[25] Evans, J. B., Uracil and pyruvate requirements of anaerobic growth of staphylococci. *J. Clin. Microbiol.* 1975, *2*, 14–17.

[26] Coleman, G., Garbutt, I. T., Demnitz, U., Ability of a *Staphylococcus aureus* isolate from a chronic osteomyelitic lesion to survive in the absence of air. *Eur. J. Clin. Microbiol.* 1983, *2*, 595–597.

[27] Park, M. K., Myers, R. A., Marzella, L., Oxygen tensions and infections: modulation of microbial growth, activity of antimicrobial agents, and immunologic responses. *Clin. Infect. Dis.* 1992, *14*, 720–740.

[28] Kirsten, W., Ihringer, M., Kuhn, M., Rohrig, B., Rudd, A. S., *Object-Oriented Application Development Using the Caché Postrelational Database*, Springer, Berlin 2003.

[29] Berth, M., Moser, F. M., Kolbe, M., Bernhardt, J., The state of the art in the analysis of two-dimensional gel electrophoresis images. *Appl. Microbiol. Biotechnol.* 2007, *76*, 1223–1243.

[30] Eisen, M. B., Spellman, P. T., Brown, P. O., Botstein, D., Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA* 1998, *95*, 14863–14868.

[31] Tufte, E. R., *The Visual Display of Quantitative Information*, Graphics Press, Cheshire, CT 2001.

[32] Luhn, S., Berth, M., Hecker, M., Bernhardt, J., Using standard positions and image fusion to create proteome maps from collections of two-dimensional gel electrophoresis images. *Proteomics* 2003, *3*, 1117–1127.

[33] Quackenbush, J., Extracting meaning from functional genomics experiments. *Toxicol. Appl. Pharmacol.* 2005, *207*, 195–199.

[34] Graur, D., Li, W.-H., *Fundamentals of Molecular Evolution*, Sinauer Associates, Sunderland, MA 2000.

[35] Zar, J. H., *Biostatistical Analysis*, Prentice Hall, Upper Saddle River, NJ 1999.

[36] Dudoit, S., Yang, Y. H., Callow, M. J., Speed, T. P., Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. *Stat. Sin.* 2002, *12*, 111–139.

[37] Pan, W., A comparative review of statistical methods for discovering differentially expressed genes in replicated microarray experiments. *Bioinformatics* 2002, *18*, 546–554.

[38] Welch, B. L., The generalization of students problem when several different population variances are involved. *Biometrika* 1947, *34*, 28–35.

[39] Raychaudhuri, S., Stuart, J. M., Altman, R. B., Principal components analysis to summarize microarray experiments: application to sporulation time series. *Pac. Symp. Biocomput.* 2000, 455–466.

[40] Appel, R. D., Vargas, J. R., Palagi, P. M., Walther, D., Hochstrasser, D. F., Melanie II – a third-generation software package for analysis of two-dimensional electrophoresis images: II. Algorithms. *Electrophoresis* 1997, *18*, 2735–2748.

[41] Taylor, C. F., Paton, N. W., Lilley, K. S., Binz, P. A. *et al.*, The minimum information about a proteomics experiment (MIAPE). *Nat. Biotechnol.* 2007, *25*, 887–893.

[42] Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G. *et al.*, Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat. Genet.* 2001, *29*, 365–371.

[43] Gill, S. R., Fouts, D. E., Archer, G. L., Mongodin, E. F. *et al.*, Insights on evolution of virulence and resistance from the complete genome analysis of an early methicillin-resistant *Staphylococcus aureus* strain and a biofilm-producing methicillin-resistant Staphylococcus epidermidis strain. *J. Bacteriol.* 2005, *187*, 2426–2438.

[44] Eyre-Walker, A., Synonymous codon bias is related to gene length in *Escherichia coli*: selection for translational accuracy? *Mol. Biol. Evol.* 1996, *13*, 864–872.

[45] Eyre-Walker, A., The close proximity of Escherichia coli genes: consequences for stop codon and synonymous codon use. *J. Mol. Evol.* 1996, *42*, 73–78.

[46] Sharp, P. M., Li, W. H., An evolutionary perspective on synonymous codon usage in unicellular organisms. *J. Mol. Evol.* 1986, *24*, 28–38.

[47] Sharp, P. M., Li, W. H., The codon Adaptation Index – a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* 1987, *15*, 1281–1295.

[48] Neidhardt, F. C., Ingraham, J. L., Schaechter, M., *Physiology of the Bacterial Cell: A Molecular Approach*, Sinauer Associates, Sunderland, MA 1990.

[49] Zhang, R., Lin, Y., DEG 5.0, a database of essential genes in both prokaryotes and eukaryotes. *Nucleic Acids Res.* 2009, *37*, D455–458.

[50] Zhang, R., Ou, H. Y., Zhang, C. T., DEG: a database of essential genes. *Nucleic Acids Res.* 2004, *32*, D271–272.

[51] Forsyth, R. A., Haselbeck, R. J., Ohlsen, K. L., Yamamoto, R. T. *et al.*, A genome-wide strategy for the identification of essential genes in *Staphylococcus aureus*. *Mol. Microbiol.* 2002, *43*, 1387–1400.

[52] Ji, Y., Zhang, B., Van Horn, S. F. *et al.*, Identification of critical staphylococcal genes using conditional phenotypes generated by antisense RNA. *Science* 2001, *293*, 2266–2269.

[53] Karlin, S., Theriot, J., Mrazek, J., Comparative analysis of gene expression among low G+C gram-positive genomes. *Proc. Natl. Acad. Sci. USA* 2004, *101*, 6182–6187.

[54] Kyte, J., Doolittle, R. F., A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 1982, *157*, 105–132.

[55] Lobry, J. R., Gautier, C., Hydrophobicity, expressivity and aromaticity are the major trends of amino-acid usage in 999 *Escherichia coli* chromosome-encoded genes. *Nucleic Acids Res.* 1994, *22*, 3174–3180.

[56] Altland, K., IPGMAKER: a program for IBM-compatible personal computers to create and test recipes for immobilized pH gradients. *Electrophoresis* 1990, *11*, 140–147.

[57] Bjellqvist, B., Hughes, G. J., Pasquali, C., Paquet, N. *et al.*, The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. *Electrophoresis* 1993, *14*, 1023–1031.

[58] Eymann, C., Dreisbach, A., Albrecht, D., Bernhardt, J. *et al.*, A comprehensive proteome map of growing *Bacillus subtilis* cells. *Proteomics* 2004, *4*, 2849–2876.

[59] Hirokawa, T., Boon-Chieng, S., Mitaku, S., SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics* 1998, *14*, 378–379.

[60] Sonnhammer, E. L., von Heijne, G., Krogh, A., A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 1998, *6*, 175–182.

[61] Bendtsen, J. D., Nielsen, H., von Heijne, G., Brunak, S., Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* 2004, *340*, 783–795.

[62] Horton, P., Park, K. J., Obayashi, T., Fujita, N. *et al.*, WoLF PSORT: protein localization predictor. *Nucleic Acids Res.* 2007, *35*, W585–587.

[63] Nakai, K., Horton, P., PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. *Trends Biochem. Sci.* 1999, *24*, 34–36.

[64] Alm, E. J., Huang, K. H., Price, M. N., Koche, R. P. *et al.*, The MicrobesOnline Web site for comparative genomics. *Genome Res.* 2005, *15*, 1015–1022.

[65] Price, M. N., Huang, K. H., Alm, E. J., Arkin, A. P., A novel method for accurate operon predictions in all sequenced prokaryotes. *Nucleic Acids Res.* 2005, *33*, 880–892.

[66] Bischoff, M., Dunman, P., Kormanec, J., Macapagal, D. *et al.*, Microarray-based analysis of the *Staphylococcus aureus* sigmaB regulon. *J. Bacteriol.* 2004, *186*, 4085–4099.

[67] Cassat, J., Dunman, P. M., Murphy, E., Projan, S. J. *et al.*, Transcriptional profiling of a *Staphylococcus aureus* clinical isolate and its isogenic agr and sarA mutants reveals global differences in comparison to the laboratory strain RN6390. *Microbiology* 2006, *152*, 3075–3090.

[68] Cirz, R. T., Jones, M. B., Gingles, N. A., Minogue, T. D. *et al.*, Complete and SOS-mediated response of *Staphylococcus aureus* to the antibiotic ciprofloxacin. *J. Bacteriol.* 2007, *189*, 531–539.

[69] Dunman, P. M., Murphy, E., Haney, S., Palacios, D. *et al.*, Transcription profiling-based identification of *Staphylococcus aureus* genes regulated by the agr and/or sarA loci. *J. Bacteriol.* 2001, *183*, 7341–7353.

[70] Herbert, S., Bera, A., Nerz, C., Kraus, D. *et al.*, Molecular basis of resistance to muramidase and cationic antimicrobial peptide activity of lysozyme in staphylococci. *PLoS Pathog.* 2007, *3*, e102.

[71] Hubscher, J., Jansen, A., Kotte, O., Schafer, J. *et al.*, Living with an imperfect cell wall: compensation of femAB inactivation in *Staphylococcus aureus*. *BMC Genomics* 2007, *8*, 307.

[72] Korem, M., Gov, Y., Kiran, M. D., Balaban, N., Transcriptional profiling of target of RNAIII-activating protein, a master regulator of staphylococcal virulence. *Infect. Immun.* 2005, *73*, 6220–6228.

[73] Kuroda, M., Kuroda, H., Oshima, T., Takeuchi, F. *et al.*, Two-component system VraSR positively modulates the regulation of cell-wall biosynthesis pathway in *Staphylococcus aureus*. *Mol. Microbiol.* 2003, *49*, 807–821.

[74] Liang, X., Yu, C., Sun, J., Liu, H. *et al.*, Inactivation of a two-component signal transduction system, SaeRS, eliminates adherence and attenuates virulence of *Staphylococcus aureus*. *Infect. Immun.* 2006, *74*, 4655–4665.

[75] Liang, X., Zheng, L., Landwehr, C., Lunsford, D. *et al.*, Global regulation of gene expression by ArlRS, a two-component signal transduction regulatory system of *Staphylococcus aureus*. *J. Bacteriol.* 2005, *187*, 5486–5492.

[76] Luong, T. T., Dunman, P. M., Murphy, E., Projan, S. J., Lee, C. Y., Transcription Profiling of the mgrA Regulon in *Staphylococcus aureus*. *J. Bacteriol.* 2006, *188*, 1899–1910.

[77] McCallum, N., Bischoff, M., Maki, H., Wada, A., Berger-Bachi, B., TcaR, a putative MarR-like regulator of sarS expression. *J. Bacteriol.* 2004, *186*, 2966–2972.

[78] Michel, A., Agerer, F., Hauck, C. R., Herrmann, M. *et al.*, Global regulatory impact of ClpP protease of *Staphylococcus aureus* on regulons involved in virulence, oxidative stress response, autolysis, and DNA repair. *J. Bacteriol.* 2006, *188*, 5783–5796.

[79] Pane-Farre, J., Jonas, B., Forstner, K., Engelmann, S., Hecker, M., The sigmaB regulon in *Staphylococcus aureus* and its regulation. *Int. J. Med. Microbiol.* 2006, *296*, 237–258.

[80] Rogasch, K., Ruhmling, V., Pane-Farre, J., Hoper, D. *et al.*, Influence of the two-component system SaeRS on global gene expression in two different *Staphylococcus aureus* strains. *J. Bacteriol.* 2006, *188*, 7742–7758.

[81] Said-Salim, B., Dunman, P. M., McAleese, F. M., Macapagal, D. *et al.*, Global regulation of *Staphylococcus aureus* genes by Rot. *J. Bacteriol.* 2003, *185*, 610–619.

[82] Sobral, R. G., Jones, A. E., Des Etages, S. G., Dougherty, T. J. *et al.*, Extensive and genome-wide changes in the transcription profile of *Staphylococcus aureus* induced by modulating the transcription of the cell wall synthesis gene murF. *J. Bacteriol.* 2007, *189*, 2376–2391.

[83] Yang, S. J., Dunman, P. M., Projan, S. J., Bayles, K. W., Characterization of the *Staphylococcus aureus* CidR regulon: elucidation of a novel role for acetoin metabolism in cell death and lysis. *Mol. Microbiol.* 2006, *60*, 458–468.

[84] Aoki, K. F., Kanehisa, M., Using the KEGG database resource. *Curr. Protoc. Bioinformatics* 2005, Chapter 1, Unit 1.12.

[85] Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K. F. *et al.*, From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res.* 2006, *34*, D354–357.

[86] Jensen, L. J., Kuhn, M., Stark, M., Chaffron, S. *et al.*, STRING 8 – a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res.* 2009, *37*, D412–D416.

[87] Snel, B., Lehmann, G., Bork, P., Huynen, M. A., STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. *Nucleic Acids Res.* 2000, *28*, 3442–3444.

[88] Bernhardt, J., Buttner, K., Scharf, C., Hecker, M., Dual channel imaging of two-dimensional electropherograms in *Bacillus subtilis*. *Electrophoresis* 1999, *20*, 2225–2240.

[89] Bernhardt, J., Weibezahn, J., Scharf, C., Hecker, M., *Bacillus subtilis* during feast and famine: visualization of the overall regulation of protein synthesis during glucose starvation by proteome analysis. *Genome Res.* 2003, *13*, 224–237.

[90] Smilansky, Z., Automatic registration for images of two-dimensional protein gels. *Electrophoresis* 2001, *22*, 1616–1626.

[91] Hecker, M., Antelmann, H., Buttner, K., Bernhardt, J., Gel-based proteomics of Gram-positive bacteria: a powerful tool to address physiological questions. *Proteomics* 2008, *8*, 4958–4975.

[92] Voigt, B., Schweder, T., Sibbald, M. J., Albrecht, D. *et al.*, The extracellular proteome of *Bacillus licheniformis* grown in different media and under different nutrient starvation conditions. *Proteomics* 2006, *6*, 268–281.

[93] Edgar, R., Domrachev, M., Lash, A. E., Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* 2002, *30*, 207–210.

[94] Brazma, A., Parkinson, H., Sarkans, U., Shojatalab, M. *et al.*, ArrayExpress – a public repository for microarray gene expression data at the EBI. *Nucleic Acids Res.* 2003, *31*, 68–71.

[95] Prince, J. T., Carlson, M. W., Wang, R., Lu, P., Marcotte, E. M., The need for a public proteomics repository. *Nat. Biotechnol.* 2004, *22*, 471–472.

[96] Pasquali, C., Frutiger, S., Wilkins, M. R., Hughes, G. J. *et al.*, Two-dimensional gel electrophoresis of *Escherichia coli* homogenates: the *Escherichia coli* SWISS-2DPAGE database. *Electrophoresis* 1996, *17*, 547–555.

[97] Vijayendran, C., Burgemeister, S., Friehs, K., Niehaus, K., Flaschel, E., 2DBase: 2D-PAGE database of *Escherichia coli*. *Biochem. Biophys. Res. Commun.* 2007, *363*, 822–827.

[98] Pleissner, K. P., Eifert, T., Buettner, S., Schmidt, F. *et al.*, Web-accessible proteome databases for microbial research. *Proteomics* 2004, *4*, 1305–1313.

[99] Hoogland, C., Mostaguir, K., Appel, R. D., Lisacek, F., The World-2DPAGE Constellation to promote and publish gel-based proteomics data through the ExPASy server. *J. Proteomics* 2008, *71*, 245–248.

[100] Hucka, M., Finney, A., Sauro, H. M., Bolouri, H. *et al.*, The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* 2003, *19*, 524–531.

[101] Brazma, A., Krestyaninova, M., Sarkans, U., Standards for systems biology. *Nat. Rev. Genet.* 2006, *7*, 593–605.

[102] Karp, P. D., Riley, M., Paley, S. M., Pellegrini-Toole, A., The MetaCyc Database. *Nucleic Acids Res.* 2002, *30*, 59–61.

[103] Vodisch, M., Albrecht, D., Lessing, F., Schmidt, A. D. *et al.*, Two-dimensional proteome reference maps for the human pathogenic filamentous fungus *Aspergillus fumigatus*. *Proteomics* 2009, *9*, 1407–1415.

[104] Vodisch, M., Kniemeyer, O., Albrecht, D., Lessing, F. *et al.*, Two-dimensional reference map of the mycelial proteome and mitochondrial sub-proteome of *Aspergillus fumigatus*. *Mycoses* 2008, *51*, 422–422.

[105] Albrecht, D., Kniemeyer, O., Brakhage, A. A., Berth, M., Guthke, R., Integration of transcriptome and proteome data from human-pathogenic fungi by using a data warehouse. *J. Integr. Bioinformatics* 2007, *4*, 52.

[106] Park, S. J., Chao, G., Gunsalus, R. P., Aerobic regulation of the sucABCD genes of *Escherichia coli*, which encode alpha-ketoglutarate dehydrogenase and succinyl coenzyme A synthetase: roles of ArcA, Fnr, and the upstream sdhCDAB promoter. *J. Bacteriol.* 1997, *179*, 4138–4142.

[107] Richardson, A. R., Libby, S. J., Fang, F. C., A nitric oxide-inducible lactate dehydrogenase enables *Staphylococcus aureus* to resist innate immunity. *Science* 2008, *319*, 1672–1676.

[108] Pagels, M., Fuchs, S., Pané-Farré, J., Kohler, C. *et al.*, Redox sensing by a Rex-family repressor is involved in the regulation of anaerobic gene expression in *Staphylococcus aureus*. *Mol. Microbiol.* 2010, *79*, 1142–1161.

