Oral Microbiology Immunology 2005: 20: 344–348
Printed in Singapore. All rights reserved

Oral Microbiology and Immunology

A searchable database for proteomes of oral microorganisms

Y. Nakano[1], Y. Shibata[1], M. Kawada[1], M. Kojima[1], H. Fukamachi[1], Y. Shibata[2,3], S. Okano[3], K. Matsushima[2], Y. Abiko[2,3], Y. Yamashita[1]

[1]Department of Preventive Dentistry, Kyushu University Faculty of Dental Science, Fukuoka, Japan, [2]Research Institute of Oral Science, [3]Department of Biochemistry and Molecular Biology, Nihon University School of Dentistry at Matsudo, Chiba, Japan

Nakano Y, Shibata Y, Kawada M, Kojima M, Fukamachi H, Shibata Y, Okano S, Matsushima K, Abiko Y, Yamashita Y. A searchable database for proteomes of oral microorganisms.
Oral Microbiol Immunol 2005: 20: 344–348. © Blackwell Munksgaard, 2005.

An online database of proteomes for two-dimensional electrophoresis (2DE) gel data was constructed and it is now freely accessible through a web-based interface. Proteins from three oral bacteria, *Streptococcus mutans* UA159, *Actinobacillus actinomycetemcomitans* HK1651, and *Porphyromonas gingivalis* W83, whose genome databases are freely available, were separated by 2DE, and protein spots were analyzed by matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) and identified. About 1000 spots from the gels of *P. gingivalis* W83 were extracted and analyzed by MALDI-TOF, and 330 proteins were identified. In addition, 160 of 240 spots of *A. actinomycetemcomitans* and 158 of 356 spots of *S. mutans* were identified. Information such as spot coordinates on the gels, protein names (predicted functions), molecular weights, isoelectroric points, and links to online databases, including Oral Pathogen Sequence Databases of the Los Alamos National Laboratory Bioscience Division (ORALGEN) and National Center for Biotechnology Information (NCBI) or The Institute Genomic Research (TIGR), were stored in tables accessible through the relational database management system MySQL on an Apache web server. To test for functionality of this database system, responses of *S. mutans* to environmental changes were analyzed using the database and 21 spots on the gel were identified as proteins whose expression had been increased or decreased by environmental pH change without in-gel trypsin digestion, protein extraction, or MALDI-TOF/TOF-MS (mass spectrometer) analysis. The identified proteins are agreement with those reported in previous papers on acid tolerance of *S. mutans*, demonstrating the usefulness of the system. This database is available at http://www.myamagu.dent.kyushu-u.ac.jp/~bioinformatics/index.html or http://www.bipos.mascat.nihon-u.ac.jp/index.html.

Key words: *Actinobacillus actinomycetemcomitans*; database; *Porphyromonas gingivalis*; proteome; *Streptococcus mutans*; two-dimensional electrophoresis

Y. Nakano, Department of Preventive Dentistry, Kyushu University Faculty of Dental Science, 3-1-1 Maidashi, Higashi-ku, Fukuoka 812–8582, Japan
Tel.: +81 92 6426423;
fax: +81 92 6426354;
e-mail: yosh@dent.kyushu-u.ac.jp
Accepted for publication May 3, 2005

Two-dimensional electrophoresis (2DE) is an effective method for proteome analysis and is widely used to analyze cellular responses to environmental changes. Proteome databases accessible through public networks are very useful for scientists working in the same area of research (1). However, matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) analysis of protein spots from 2DE gels is still a laborious and expensive technique. Public databases of 2DE gels showing the entire protein patterns of bacterial or eukaryotic cells would be helpful for identification of protein spots on the gels without resorting to in-gel trypsin digestion, protein extraction, or MALDI-TOF/TOF-MS(mass spectrometer) analysis. Nevertheless, only a limited number of resources are available in public databases, in particular in oral microbiology, although the sequence data of more than 200 genomes including several oral bacteria are accessible.

We constructed web-accessible 2DE databases containing data from *Streptococcus mutans*, *Porphyromonas gingivalis*, and *Actinobacillus actinomycetemcomitans* for oral microbial research using Apache (an httpd server), MySQL (a relational database management system), and PHP (the PreHypertextProcessor language). This system offers easy interactive access to users via World Wide Web, is cost-effective, easily constructed and can be maintained by a database administrator.

Using this system, we analyzed proteins whose levels affected by environmental change to acidic pH.

Material and methods

Bacterial strains and culture conditions

P. gingivalis strain W83 was grown anaerobically (10% CO_2, 10% H_2, 80% N_2) in GAM broth (Nissui Medical Co., Tokyo, Japan) supplemented with hemin (5 μg/ml) and menadione (1 μg/ml) at 37°C. *A. actinomycetemcomitans* Y4 was grown in THY broth (per liter: 30 g of Trypticase soy broth [BBL Microbiology Systems, Cockeysville, MD], and 10 g of yeast extract [Difco, Detroit, MI]) at 37°C in a 5% CO_2 atmosphere. *S. mutans* UA159 was routinely subcultured and maintained in brain heart infusion (BHI, Difco) broth. Induction of the acid tolerance response was assessed by the incubation of *S. mutans* UA159 cells at pH 5.5 and pH 7.5 for 2 h in BHI broth, after growth to the mid-exponential phase (OD_{550} of 0.6) in BHI at 37°C with 5% CO_2.

2D gel analysis

Cell protein extracts from *S. mutans* cells for 2D gel electrophoresis were prepared using the FastPrep device (Savant Instruments, Holbrook, NY). The cells were suspended in lysis buffer that contained 7 M urea, 2 M thiourea, 40 mM Tris, 1 mM EDTA, 2 mM tributylphosphine, 1 mM PMSF, and 0.15% leupeptin. The suspension was transferred to a FastPrep BLUE tube (Savant Instruments). The tube was processed four times for 4 min at speed 6 in the FastPrep device. The supernatant containing the protein extracts was obtained by centrifugation and this supernatant was applied in sample cups under silicone oil at the anodic end of the gel strip.

The culture of *P. gingivalis* W83 was precipitated by adding 50% trichloroacetic acid (Sigma, St. Louis, MO) to a final concentration of 10% to prevent degradation of proteins by proteases or endopeptidases. The precipitate was centrifuged at $8500 \times g$ for 10 min at 4°C and then washed three times with ice-cold acetone. The harvested cells were suspended in lysis solution containing 7 M urea, 2 M thiourea, 4% 3-[(3-cholamidopropyl)-dimethylammonio]-1-propanesulfonate (CHAPS), 40 mM Tris, and 1 mM EDTA, and cells were lysed by ultrasonication and centrifuged for 15 min at $10{,}000 \times g$. The obtained supernatant was used as a protein extract for 2DE.

The culture of *A. actinomycetemcomitans* HK1651 was centrifuged at $5000 \times g$ for 10 min and the cells were washed twice with phosphate-buffered saline (PBS; 0.1 M NaCl, 33 mM Na_2HPO_4, 17 mM NaH_2PO_4; pH 7.4). The harvested cells were suspended in the same lysis and extract procedures as described above for *P. gingivalis* samples.

Two-dimensional gel electrophoresis was performed by using the Multiphor II (Amersham Biosciences, Little Chalfont, UK) system according to the instructions of the manufacturer. Protein components were quantified using the PlusOne 2-D Quant kit (Amersham Biosciences). Protein samples (200 μg) were separated by using immobilized 18-cm pH gradient (IPG) strips in pH ranges of 4.0–7.0, 4.5–5.5, 5.0–6.0, 5.5–6.7, or 6.0–10.0. The isoelectric focusing of the first dimension was initiated at 200 V and was gradually increased to 3500 V for 3000 Vh, remaining at 3500 V for 35 kVh at 15°C. After isoelectric focusing, the strips were equilibrated for two intervals of 15 min in equilibration buffer containing 6 M urea, 50 mM Tris-HCl, pH 8.8, 2% SDS, and 30% glycerol. For the first equilibration step, 1% (w/v) DL-dithiothreitol was added to reduce the proteins. Thereafter, the proteins were carbamidomethylated with 4% (w/v) iodoacetamide. The second dimension was separated on the basis of molecular mass using a 12.5% polyacrylamide gel at a constant voltage of 200 V. Migrated proteins were stained with Coomassie Brilliant Blue R-250. Two-dimensional gel electrophoresis was repeated for each strain independently seven times to confirm the reproducibility of the protein pattern. Only differences that were reproduced each time are described here. Gels were stained with Coomassie blue, and spots were detected and scanned and assisted by the PDQUEST 2-D GEL ANALYSIS Software (Bio-Rad, Laboratories, Richmond, CA). Differences of greater than 2-fold identified by the software were individually validated and are marked.

MALDI-TOF/MS

Protein spots that differed in their expression profiles were semiautomatically extracted, digested with trypsin, and purified using a ProteomIQ Xcise In-Gel Digest kit (Proteome Systems, North Ryde, Australia) and a high-throughput gel-excise processor (Xcise, Shimadzu Co., Tokyo, Japan). Peptide mass analyses were performed using an AXIMA CFR (Shimadzu Co.) MALDI-TOF MS in reflectron mode.

Database searching

Proteins were identified from MALDI fingerprint data by using MASCOT for websearch (http://www.matrixscience.com) or MS-Fit (http://www.prospector.ucsf.edu/) against a public database (National Center for Biotechnology Information) or a locally installed protein prospector algorithm (http://www.prospector.ucsf.edu) with sequence data obtained from The Institute for Genomic Research (http://www.tigr.org) and Oklahoma University.

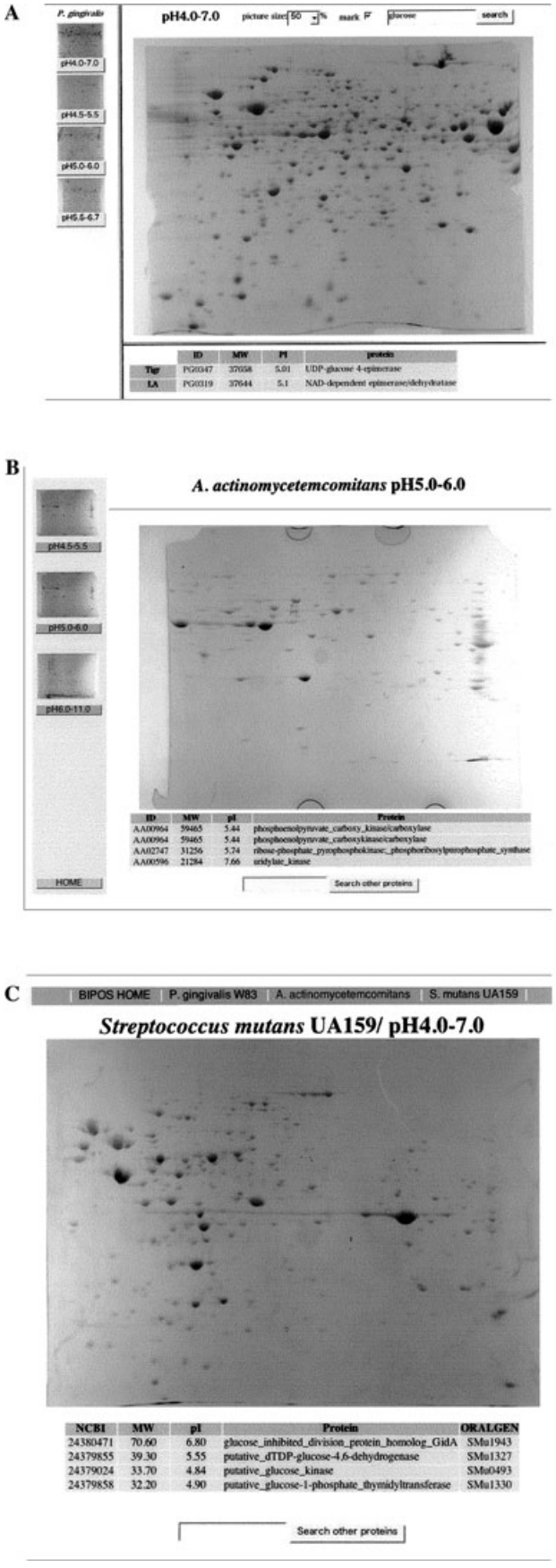
Software tools

For the software development of our database system, only open source software tools were applied. The program package consists of the relational database management system MySQL (http://www.mysql.com/), scripts written in PHP and HTML, and the GD graphic library (http://www.boutell.com/gd/) for manipulating images. Internet access is accomplished via an Apache Web server on a Linux platform. The values for molecular weight and isoelectric point (pI) for all proteins were calculated using the program PEPSTATS in EMBOSS (4).

Results and discussion

Gel electrophoresis and identification of protein spots

Coomassie blue-staining of 2DE gels of whole-cell lysates of *P. gingivalis*, *A. actinomycetemcomitans*, and *S. mutans* is shown in Fig. 1. ORFs predicted from nucleotide sequences of the genomes were around 2000 in each bacteria and 271, 160, and 158 spots were identified from gels of *P. gingivalis*, *A. actinomycetemcomitans*, and *S. mutans*, respectively (Fig. 1). *P. gingivalis* proteins on pI 4.0–7.0 gel were separated again on narrower pI ranges. From these gels, 1020 spots were picked up in total, and 271 proteins were identified. Some spots did not yield any signals in TOF-MS analysis, and others had no matches to proteins in the databases. Calculation of the theoretical MW and the pI from the amino acid composition of the proteins predicted 645 proteins in the range of 4.5–6.5 in pI and 16,000–100,000 in MW, which were the detection limits observed in pH 4.0–7.0 gel strips and SDS-PAGE using a 12.5% acrylamide gel. Within this area, 446 spots were observed and 200 proteins were identified (Fig. 1A). More efficient protein identification may be achieved by in-gel digestion with other proteases because some proteins may be not suitable for cleavage by trypsin.



Database application and data presentation

A schematic overview of our proteome informatics approach for oral microbial research is illustrated in Fig. 2. Database users can gain access via the Hyper Text Transfer Protocol (http). When the demands for query arrive from the World Wide Web, PHP-script is activated and converts the demands into standard field values and also specific structured query language (SQL) commands to instruct MySQL to retrieve the data in the database. Moving the mouse pointer over the spot will display its predicted protein name, in addition to the calculated MW, pI, and hyperlinks to public databases, e.g. NCBI (http://www.ncbi.nlm.nih.gov/), TIGR (http://www.tigr.org/), or the Los Alamos Oral Pathogen Sequence Database (http://www.oralgen.lanl.gov/), which can be accessed by clicking on the spot. In addition, protein spots can be searched by name.

After 2DE and comparison of various gels, protein spots affected by environmental changes would be identified by finding the corresponding spots on the gel images on screen without extraction and MALDI-TOF MS analysis. This would greatly reduce the time and cost of proteome analysis of oral bacteria. Of course,

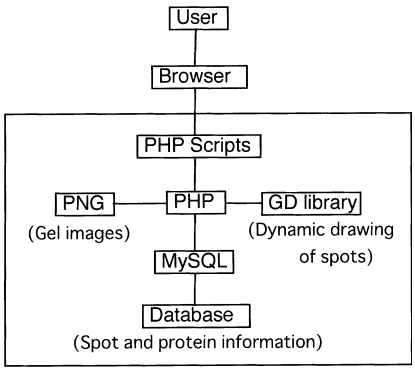


Fig. 2. Database architecture on the web server.

Fig. 1. Database interfaces for *P. gingivalis* (A), *A. actinomycetemcomitans* (B), and *S. mutans* (C) 2DE gels. Moving the mouse pointer over a spot will display its predicted protein name, as well as the calculated molecular weight, pI, and hyperlinks to public databases. In addition, protein spots can be searched by name. Each panel shows the result of protein name search; corresponding spots are marked with colored circles, tables under the gel displays IDs of NCBI and ORALGEN or TIGR, molecular weight, pI, and protein name.

spots that are not found at all on the gels used for the public database are produced by gene expression specific to other environments or conditions.

When new protein information for spots on the reference gels is found, it is added to the database. Relational database systems provide easy continuous updates for the database, even if they require the addition of new database fields and transferral of the new information for all proteins to the fields in one step.

Identification of the proteins affected by acid adaptation in *S. mutans*

To test the functionality of this database system, responses of *S. mutans* to environmental changes were analyzed by using the database, because many reports on proteomic studies of *S. mutans* have been published (2, 5–7), whereas there are not sufficient studies on *P. gingivalis* and *A. actinomycetemcomitans* to test the database. The 2DE protein profile of *S. mutans* UA159 after exposure to an acidic environment (pH 5.5) for 2 h was compared to that without such exposure (Fig. 3). Using the public database described above, we then attempted to identify the proteins whose synthesis was enhanced or reduced by environmental change. Twenty proteins showed a significant increase in response to the exposure, and one protein was decreased by the environmental change. The corresponding spots on the database were used for identification of these proteins, and all 21 proteins were easily identified (Table 1).

Len et al. reported changes in the expression of metabolic proteins in *S. mutans* when the growth pH was lowered from pH 7.0 to pH 5.0 at a steady state in glucose-limited anaerobic continuous culture (3). Nineteen of the 21 proteins mentioned above are listed in the literature (2, 3, 5–7) as the metabolic proteins whose expression is affected by pH change. The other two proteins corresponding to SMu0868 and SMu1097 in the Los Alamos Oral Pathogen Sequence Database are ATP-binding subunits of ATP-dependent Clp protease and ribosomal protein S1, respectively. The former is known to be induced at low pH (3). Further investigations of acid sensitive strains of *S. mutans* UA159 are currently in progress.

Extraction and identification of 21 proteins in polyacrylamide gel separated by 2DE requires costly and complex technology. Public databases that can be accessed from every computer connected to the Internet, simply by using a web browser, would be a very useful tool for proteomics of oral microorganisms. Our web pages for 2DE databases containing *S. mutans*, *P. gingivalis*, and *A. actinomycetemcomitans* are open to the public at http://www.myamagu.dent.kyushu-u.ac.jp/~bioinformatics/index.html or http://www.bipos.mascat.nihon-u.ac.jp/index.html.

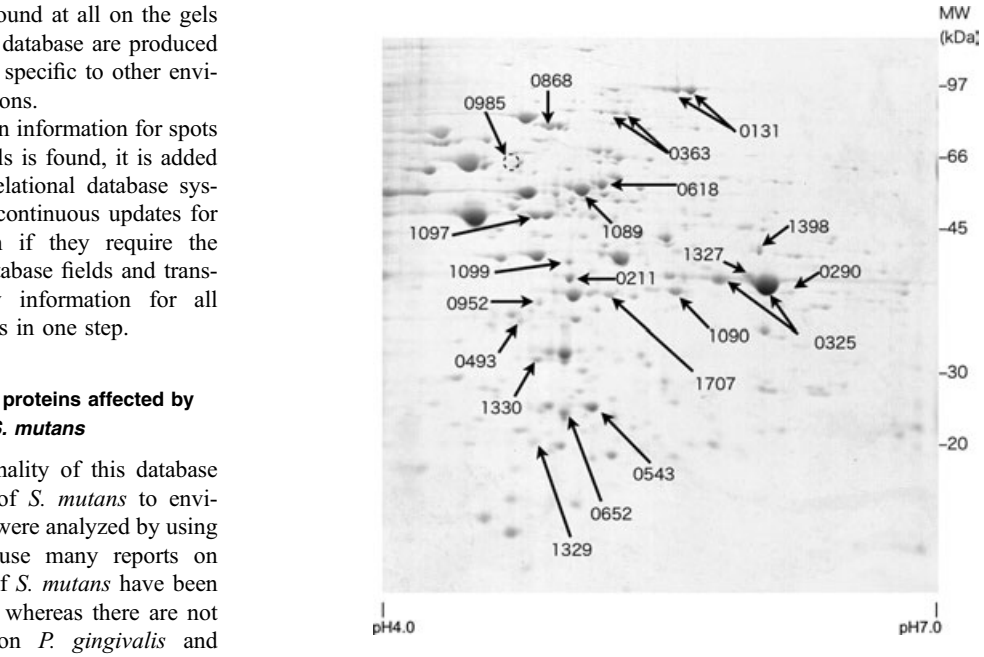


Fig. 3. Protein profiles of exponential-phase cells of *S. mutans* growing at pH 7.5 and rapidly subjected to low pH conditions for 2 h followed by extraction, separation by 2DE, and staining with Coomassie Brilliant Blue. The protein spots whose synthesis were enhanced or reduced by environmental change are indicated with arrows and numbered with the gene ID of the Oral Pathogen Sequence Database site at Los Alamos National Laboratory Bioscience Division. The circle (0985) indicates the position of protein spot found on the gel of the database but not on this figure.

Table 1. Differential expression of proteins by *S. mutans* grown at pH 7.5 or pH 5.5

Gene ID	Enzyme and gene	Regulation
SMu0131	alcohol-acetaldehyde dehydrogenase (*adhE*)	up
SMu0211	ketol-acid reductoisomerase (*ilvC*)	up
SMu0290	glycerol-3-phosphate dehydrogenase (*gpdA*)	up
SMu0325	glyceraldehyde-3-phosphate dehydrogenase; plasmin receptor (*gapC*)	up
SMu0363	pyruvate formate-lyase (*pfl*)	up
SMu0493	glucose kinase (*glk*)	up
SMu0543	phosphoglyceromutase (*pmgY*)	up
SMu0618	NADP-dependent nonphosphorylating glyceraldehyde-3-phosphate dehydrogenase (*gapN*)	up
SMu0652	triosephosphate isomerase (*tpiA*)	up
SMu0868	ATP-dependent Clp protease, ATP-binding subunit (*clp*)	up
SMu0952	phosphotransacetylase (*pta*)	up
SMu0985	phosphoglucomutase (*pgmA*)	down
SMu1089	pyruvate kinase (*pykA*)	up
SMu1090	6-phosphofructokinase (*pfk*)	up
SMu1097	ribosomal protein S1 (*rpsA*)	up
SMu1099	branched-chain amino acid aminotransferase (*ilvE*)	up
SMu1327	dTDP-glucose-4, 6-dehydratase (*rmlB*)	up
SMu1329	dTDP-4-keto-L-rhamnose reductase (*rmlC*)	up
SMu1330	glucose-1-phosphate thymidyltransferase (*rmlA*)	up
SMu1398	glycogen biosynthesis protein (*glgD*)	up
SMu1707	PTS system, mannose-specific component IIAB (*manL*)	up

Gene ID, gene names associated with the *S. mutans* genome at the Oral Pathogen Sequence Database site (http://www.stdgen.lanl.gov/oragen).

Acknowledgments

This work was supported in part by a Research Fellowship from the Japan Society for the Promotion of Science for Young Scientists 1507824 (H.F.), Grant-in aid for Scientific Research 16209063 (Y.A.) from the Japan Society for the Promotion of

Science, and Grant from 'Academic Frontier' Project for Private Universities: matching fund subsidy from MEXT (Ministry of Education, Culture, Sports, Science and Technology), 2001–05 (Y.A.).

References

1. Görg A, Weiss W, Dunn MJ. Current two-dimensional electrophoresis technology for proteomics. Proteomics 2004: **4**: 3665–3675.
2. Lemos JA, Burne RA. Regulation and physiological significance of ClpC and ClpP in *Streptococcus mutans*. J Bacteriol 2002: **184**: 6357–6366.
3. Len AC, Harty DW, Jacques NA. Proteome analysis of *Streptococcus mutans* metabolic phenotype during acid tolerance. Microbiology 2004: **150**: 1353–1366.
4. Rice P, Longden I, Bleasby A. EMBOSS: the European molecular biology open software suite. Trends Genet 2000: **16**: 276–277.
5. Welin J, Wilkins JC, Beighton D, Wrzesinski K, Fey SJ, Mose-Larsen P, et al. Effect of acid shock on protein expression by biofilm cells of *Streptococcus mutans*. FEMS Microbiol Lett 2003: **227**: 287–293.
6. Wen ZT, Burne RA. LuxS-mediated signaling in *Streptococcus mutans* is involved in regulation of acid and oxidative stress tolerance and biofilm formation. J Bacteriol 2004: **186**: 2682–2691.
7. Wilkins JC, Homer KA, Beighton D. Analysis of *Streptococcus mutans* proteins modulated by culture under acidic conditions. Appl Environ Microbiol 2002: **68**: 2382–2390.