Complex artificial neural networks for identification of lactic acid bacteria

Mareike Wenning, Nicole Büchl and Siegfried Scherer

Abteilung Mikrobiologie, Zentralinstitut für Ernährungs- und Lebensmittelforschung (ZIEL), Wissenschaftszentrum Weihenstephan, Technische Universität München, D-85350 Freising, Germany, E-mail: mareike.wenning@wzw.tum.de

Lactic acid bacteria (LAB) are a large and heterogeneous group of micro-organisms with high relevance for food. Many species are used for the production of fermented foods like cheese, yoghurt, sauerkraut, sourdough, salami etc.; however, LAB may also occur as food spoilers e.g. in beer or dairy products. Identification of LAB can be achieved by determining enzyme profiles for sugar assimilation, which is used in commercial identification kits, but molecular techniques like DNA sequencing or restriction analyses of DNA fragments provide more exact and reliable results. However, identification of LAB is complicated by high species richness as well as a high degree of heterogeneity and the presence of very closely related species at the same time. This leads in many cases to a restriction of identification schemes to a subgroup of the LAB taxon. Thus, to cover a large variety of LAB the application of several protocols may become necessary.

To construct an identification scheme applicable to as many LAB species as possible FT-IR spectroscopy in combination with artificial neural networks (ANN) was used. This combination had already proven to be a very powerful tool for difficult species identification and even typing of micro-organisms below the species level [1,2]. The incubation conditions for the microbes were 24 hours at 34°C on all peptone agar with tween. These enable growth of mesophilic and thermophilic strains as well as rods and cocci. 325 strains belonging to 92 species of the genera *Carnobacterium*, *Enterococcus*, *Lactococcus*, *Streptococcus*, *Pediococcus*, *Oenococcus*, *Lactobacillus*, *Leuconostoc* and *Weissella* served as reference strains. The majority of strains were identified by 16S rDNA sequencing, specific PCR, multiplex PCR, and amplified ribosomal DNA restriction analysis to ensure assignment of correct names. Of each strain at least six independent spectra were collected of which five were used for the training process of the ANNs and one for the internal validation of the resulting model. The identification procedure was divided into 33 single problems and, consequently, the final ANN consisted of 33 single nets. The internal validation of the ANN resulted in 98.8% correct identification. An external validation performed with 532 spectra of 83 unknown strains achieved 94% correctly identified spectra and only 1.3% were assigned to a wrong class.

Although, the obtained model may still be improved by including more reference strains, these good results demonstrate that FT-IR spectroscopy is a promising alternative to conventional and molecular tools for identification of LAB and powerful techniques for data treatment enable the solution of even very complex identification tasks.

References

[1] C. A. Rebuffo, J. Schmitt, M. Wenning, F. von Stetten, und S. Scherer. 2006. Reliable and rapid identification of *Listeria monocytogenes* and *Listeria* species by artificial neural network-based Fourier transform infrared spectroscopy. Appl. Environ. Microbiol. **72:** 994-1000.

[2] C. A. Rebuffo-Scheer, J. Schmitt, and S. Scherer. 2007. Differentiation of *Listeria monocytogenes* Serovars by Using Artificial Neural Network Analysis of Fourier-Transformed Infrared Spectra. Appl. Environ. Microbiol. **73**: 1036-1040.