Journal of Industrial Microbiology, 13 (1994) 382–388

Published by The Macmillan Press Ltd

Application of pattern recognition to monitoring fermentations of *Bacillus amyloliquefaciens*

Susanne E. Keller, Diana S. Stewart and Steven M. Gendel

National Center for Food Safety and Technology, US Food and Drug Administration, 6502 S. Archer Rd, Summit-Argo, IL 60501, USA

(Received 30 March 1994; accepted 20 July 1994)

Key words: Pattern recognition; *Bacillus amyloliquefaciens*; Characterization; Classification; Fermentation

SUMMARY

Pattern recognition techniques were applied to analytical data to distinguish abnormal from normal microbial fermentations using *Bacillus amyloliquefaciens* as a model system. Patterns of fermentation end products during growth of *B. amyloliquefaciens* were obtained from HPLC analysis of broth samples. Data were also obtained from fermentations using other bacterial species, strains, and environmental conditions, and were compared with the model data set. The bacterial species cultured included *B. subtilus*, *B. licheniformis*, and *Escherichia coli*. Environmental variables included aeration and temperature. The chromatographic patterns were compared by using hierarchical cluster and principal component analysis to obtain a quantitative measure of their similarity and to establish the normal variability within a model data set. Statistical analysis of the data indicated that individual fermentations can be assigned to distinct clusters on the basis of their divergence from the model system. Altered environments and other species can be identified as outliers from the model set. These results show that pattern recognition analysis has direct applicability to monitoring fermentation processes.

INTRODUCTION

In fermentation, the safety and quality of the final product depend on the growth and purity of microorganisms present. Traditionally, monitoring of microbial fermentations has focused on individual parameters, such as oxygen consumption, pH, or the production of a product. Improper conditions that cause abnormal growth patterns can result in large changes in the rate of product formation. Therefore, methods to monitor the growth of the microorganism and to provide an early indication of possible complications, such as contamination, can be critical in quality assurance and safety assessment and in maximizing profit. Unfortunately, all critical factors involved in the growth of a particular microorganism or in the production of a particular metabolite may not be recognized.

Statistical methods have been applied for pattern recognition to classify processes and predict the outcome of a fermentation [7,11]. Variables include factors such as oxygen consumption or product concentration. Information concerning metabolites or by-products other than the desired end product was not obtained. Generally, data on a particular fermentation may not be quantitatively analyzed during evaluation. Lack of evaluation can be attributed to the sheer volume of information, some of which may not be critical to the outcome of a fermentation. Complete analysis of all available information would require appropriate tools or methods to identify important parameters and successfully predict fermentation results. The statistical methods used by Guthke and Rossman [7], and Saner and Stephanopoulos [11] could be applied to metabolic information as well as to discrete parameters to predict fermentation outcome. Statistical treatment of metabolic data produced during a fermentation may also reveal correlations previously unsuspected, such as those between a precursor and a desired end product. Such information could later be applied to improve product yield.

The use of information on metabolic intermediates and by-products for monitoring microbial fermentation requires a quick and accurate means of biochemical analysis, such as HPLC. In his work on identification of pathogenic *Bacillus cereus*, Zerfiridis [14] used HPLC to separate major extracellular proteins produced during growth and speculated that the quantities and types of the proteins might be useful as an identification or classification tool. Generally, protein profiles used to identify microorganisms have focused on intracellular proteins separated by gel electrophoresis [8].

HPLC elution patterns have not typically been used to establish a metabolic profile for identification or classification. Examination of elution patterns obtained from HPLC analysis can yield information about the status of a fermentation in a rapid and reproducible manner. In this work, the end products produced during growth of *Bacillus amyloliquefaciens* were analyzed by HPLC. The resulting elution patterns were used to create statistical models representing a desired 'ideal' fermentation. The models were then used to identify fermentations arising from altered organism or culture conditions.

Correspondence to: Susanne E. Keller, FDA, National Center for Food Safety and Technology, 6502 S. Archer Rd, Summit-Argo, IL 60501, USA.

MATERIALS AND METHODS

B. amyloliquefaciens ATCC 23843, *Escherichia coli* ATCC 15922, *B. licheniformis* ATCC 10716, and *B. subtilus* ATCC 37015 cultures were obtained from the American Type Culture Collection (Rockville, MD, USA). *B. amyloliquefaciens* 1521 was obtained from the Institute of Applied Microbiology (Tokyo, Japan). Frozen stock cultures were inoculated into brain heart infusion broth (BHI; Difco Laboratories, Detroit, MI, USA) and incubated overnight at 37 °C. After incubation the cultures were transferred at a 1% inoculum level to 100 ml BHI in 500-ml baffled flasks and incubated at 37 °C with a shaker speed of 250 r.p.m. These parameters were the standard or control fermentation conditions. Environmental variations were simulated by altering shaker speed (100 and 400 r.p.m.), temperature (25 and 32 °C) and flask type (baffled vs non-baffled). Cultures were harvested when they reached approximately 10^9 cells ml^{-1}. The supernatant broths were passed through 0.2-μm pore size filters (Gelman Sciences, Ann Arbor, MI, USA) before injection into the HPLC system.

A Waters HPLC system (Milford, MA, USA) equipped with a Model 600E solvent delivery system, a Model 991 Photodiode Array Detector, and a Model 715 Ultra Wisp were used to analyze fermentation broths. The column used was a Waters Protein-Pak™ DEAE 8HR. The solvent system was a convex gradient (38 min) from an initial buffer of 20 mM Tris HCl (Sigma Chemical Co., St Louis, MO, USA), pH 8.2, to a final buffer of 20 mM Tris HCl, pH 8.2, plus 0.75 M NaCl; flow rate, 1.56 ml min^{-1}. The effluent was monitored at wavelengths of 200 to 350 nm. Absorbance data at 220 and 280 nm were extracted at 20-s intervals and imported to the statistical analysis software.

Amylase activity in broths was measured by a modification of the procedure of Bezbaruah et al. [3]. The reaction mixture contained 0.1 ml culture broth, 0.2 ml H_2O, 0.2 ml sodium phosphate buffer (0.1 M, pH 7.0). The reaction was initiated by adding 0.5 ml of a 1% potato starch solution (Sigma). After 5 min at room temperature, the reaction was stopped with 0.5 ml 1 N HCl. A 0.1-ml aliquot was transferred to a clean tube, 0.05 ml Lugol's solution and 0.05 ml 1 N HCl were added, and the volume was adjusted to 7.5 ml. The absorbance was measured at 610 nm. One unit (U) was defined as the amount of enzyme required to decrease the absorbance by 0.05 in 1 min at room temperature.

Data from HPLC profiles were analyzed by hierarchical cluster analysis (HCA), principal component analysis (PCA), and models developed by Soft Independent Modeling of Class Analogy (SIMCA) using Pirouette (Infometrix, Inc., Seattle, WA, USA), a multivariate data analysis software program. Data were preprocessed by autoscaling (data were mean-centered and then variance-scaled). SIMCA modeling and prediction were performed at the 95% confidence level.

RESULTS AND DISCUSSION

The typical HPLC elution pattern obtained from the standard fermentation broths had approximately 15–20 peaks (Fig. 1). Peaks on chromatographs from different fermentation broths could be identified as the same on the basis of their UV spectra and elution time. Some peaks from standard fermentation broths were also found in chromatographs of unused BHI media and were clearly related to the fermentation medium. Variations in peak size were observed visually; exact determination of the variability of each peak would require each peak of a chromatograph to be quantified and evaluated statistically. Evaluation of an individual chromatograph based on its overall similarity to others of previous fermentation broths, therefore, would be very difficult. Thus, to acertain whether a relationship existed between chromatograph patterns, all data were explored by the HCA algorithms in the Pirouette program.

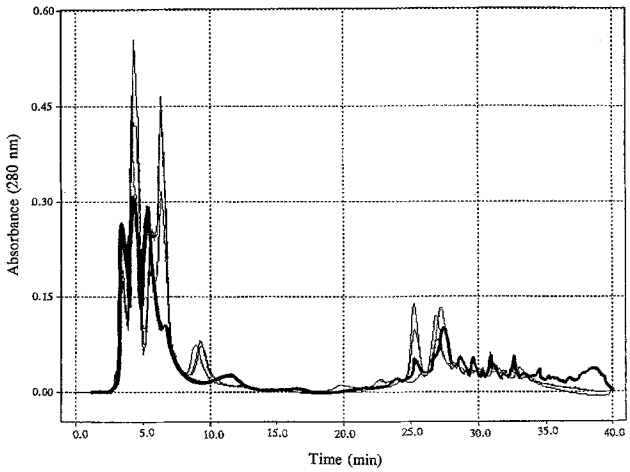

Fig. 1. HPLC elution pattern from fermentation broths. Typical elution patterns from three *B. amyloliquefaciens* 1521 fermentation broths are shown overlayed with the elution pattern of a control BHI broth (dark line).

Historically, there is a precedent for using HCA to classify microorganisms and examine fermentation products [5]. HCA is used extensively in numerical taxonomy, which is based largely on biochemical tests of microorganisms. These biochemical tests often measure or identify metabolites, e.g. the production of urease, expressed during the growth of an organism in a specific environment. Test results are combined and evaluated to determine taxonomic placement of a microorganism. The use of HPLC profiles of fermentation broths to classify microorganisms is based on one test: growth in a specific environment. The growth is described by the entire profile of biochemical compounds found. This classification procedure could be described as a series of biochemical changes characteristic of a particular bacterial species. The characteristic profile should be different for highly dissimilar species but similar for those closely related.

Fig. 2 shows the results of HCA of vectors from 40-min HPLC profiles of fermentation broths of bacteria grown under standard conditions. The HPLC profiles were obtained at 280 nm using 10-s intervals. HCA was performed by using an incremental link method (a centroidal clustering technique), which is described in the Pirouette software manual and uses a sum of squares approach to calculate the nearest cluster. The incremental link method is recommended when two classes or

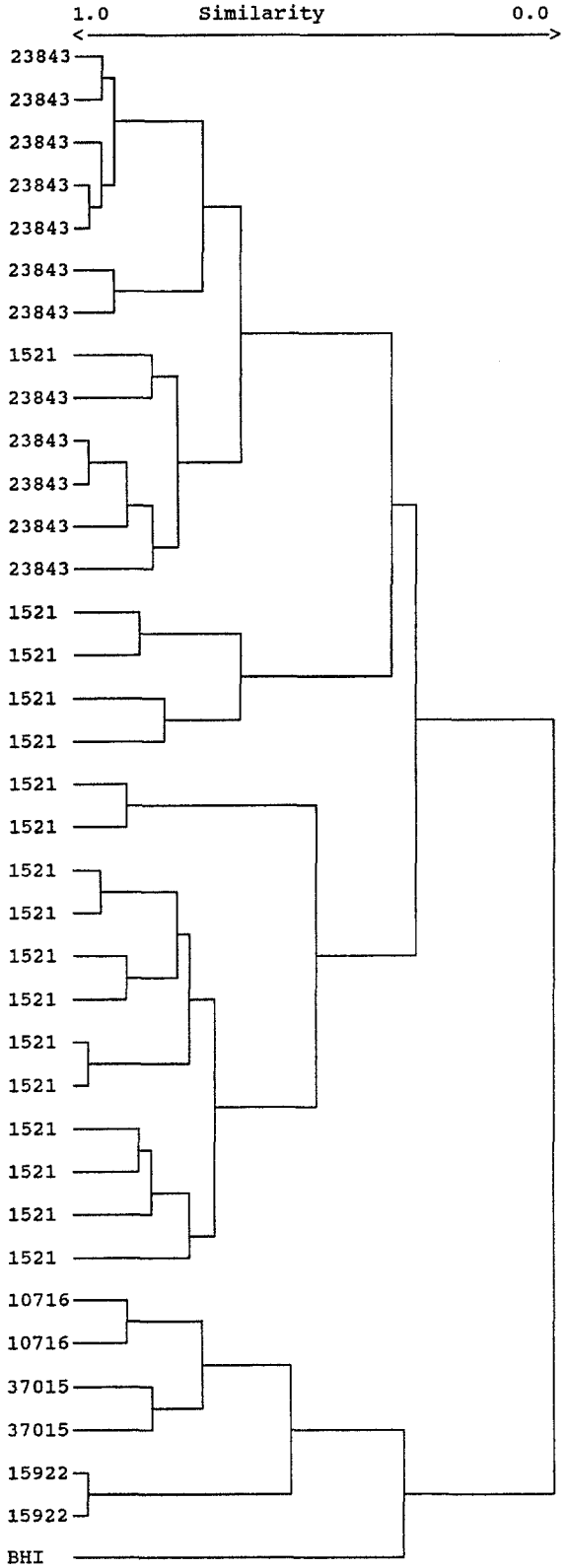

Fig. 2. Dendrogram of HPLC profiles from fermentation broths. *B. amyloliquefaciens* strains are designated as 23843 and 1521. Other designations are by ATCC number: 15922, *E. coli*, 10716, *B. licheniformis*, and 37015, *B. subtilus*. An uninoculated broth is designated BHI. All cultures were grown under the standard conditions. The distance scale is from 1.0 (identical) to 0.0 (no similarity).

clusters are not well separated, e.g. when strains of microorganisms are closely related. *B. amyloliquefaciens* strains 1521 and 23843 are closely related, whereas other *Bacillus* strains are less closely related, and *E. coli* is unrelated. Although some overlap of the broth samples occurred between *B. amyloliquefaciens* strains 1521 and 23843, both were clearly differentiated from other *Bacillus* strains, *E. coli*, and the sterile broth control (Fig. 2). The other *Bacillus* strains, *E. coli*, and the sterile broth control all comprised a separate node with no similarity to the *B. amyloliquefaciens* strains (1.0 represents complete similarity and 0.0 represents no similarity). The two nodes containing the two *B. amyloliquefaciens* strains were more dissimilar than one might expect, considering the relationship of the strains. Nonetheless, these results indicate that classification of a microorganism as to its relationship with other species is possible based on a single HPLC chromatograph elution profile.

Results of HCA from altered growth condition experiments for the two *B. amyloliquefaciens* strains are shown in Figs 3 and 4. Again, an incremental link method was used because the data were expected to reflect close groupings. Controls (designated C) from both strains tended to cluster together in the dendrograms. Overlap between altered conditions and standard conditions did occur; however, since all test samples were of the same species, this was not unexpected. Shaker speed (aeration) had a greater influence on growth products of the organisms than did temperature. For strain 23843, the higher shaker speed fermentations appeared to be more dissimilar. For strain 1521, a lower shaker speed was just as influential as a higher speed. This type of classification allows some visualization of fermentation similarity; however, it does not classify them as 'good' or 'bad' but only as different. Because the dendrogram distance scale is defined by the most dissimilar test sample used, the separation may be exaggerated when all test samples are related. However, since clustering occurred within the standard conditions, a model could be created to describe an ideal and expected variation. Such a model could be used to identify nonconforming fermentations. In a quality control program, this would not identify the cause of a particular deviation, but could flag nonconforming fermentations for closer evaluation.

B. amyloliquefaciens strains are often used to produce amylase. In this work, amylase activity was measured in all fermentation broths used for classification of test samples by HCA and PCA, and was used later for correlation with HPLC profile models. Such correlation would allow for classification or identification of high amylase-producing fermentations. The fermentation broths contained from 0 to 25 U ml^{-1} amylase activity. Generally, cultures grown under standard conditions produced higher levels of amylase than did those grown under altered conditions. In addition, under the culture conditions used, no amylase was found in cultures of strains other than *B. amyloliquefaciens*. Three classes were defined on the basis of U ml^{-1} amylase: low (0–5); medium (5–10); high (>10).

Because clustering occurred within the data sets derived from the HPLC profiles (Figs 2–4), the data were further analyzed using PCA to develop SIMCA models to predict amylase production. The advantage of PCA is the ability to trans-

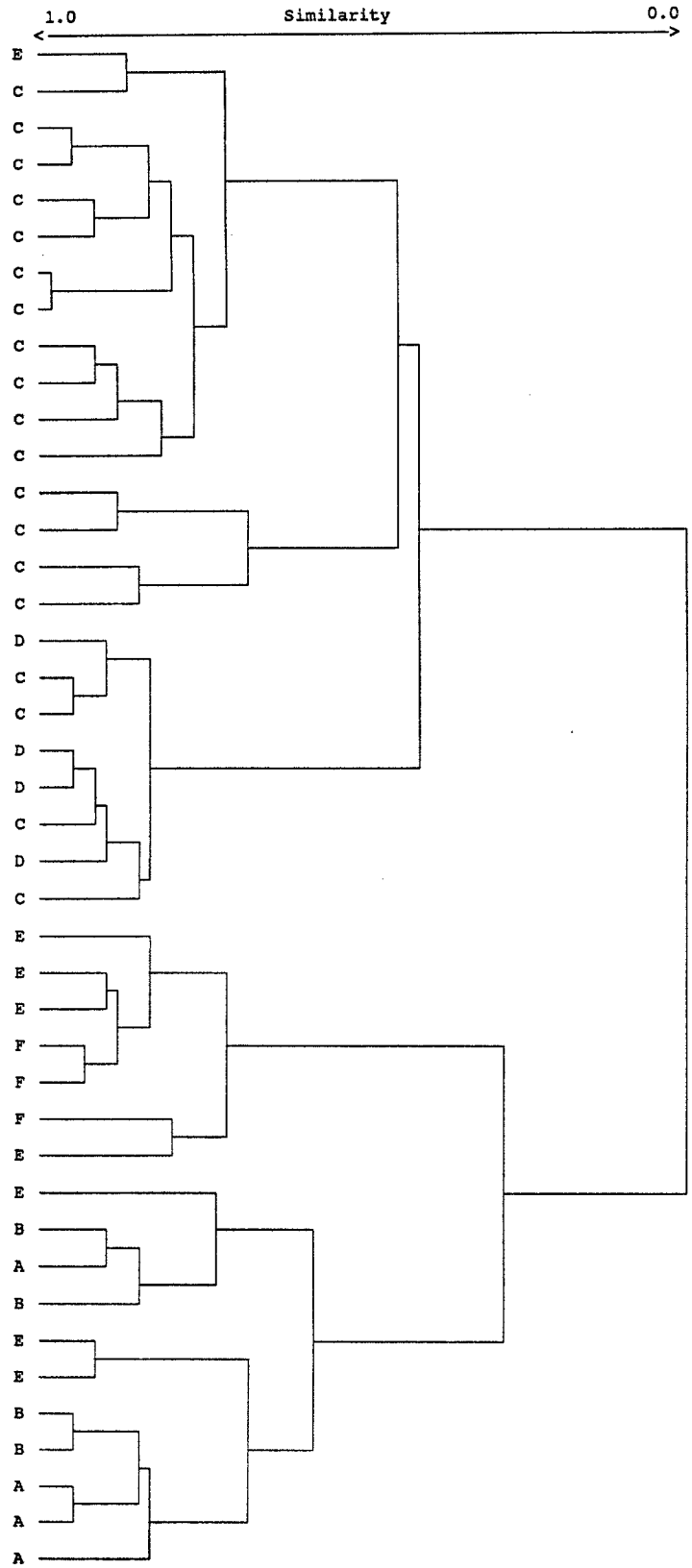

Fig. 3. Dendrogram of HPLC profiles from *B. amyloliquefaciens* 1521 grown under different temperatures and shaker speeds. A, 37 °C, 100 r.p.m.; B, 32 °C, 100 r.p.m.; C, 37 °C, 250 r.p.m. (standard condition); D, 32 °C, 250 r.p.m.; E, 37 °C, 400 r.p.m.; F, 32 °C, 400 r.p.m. The distance scale is from 1.0 (identical) to 0.0 (no similarity).

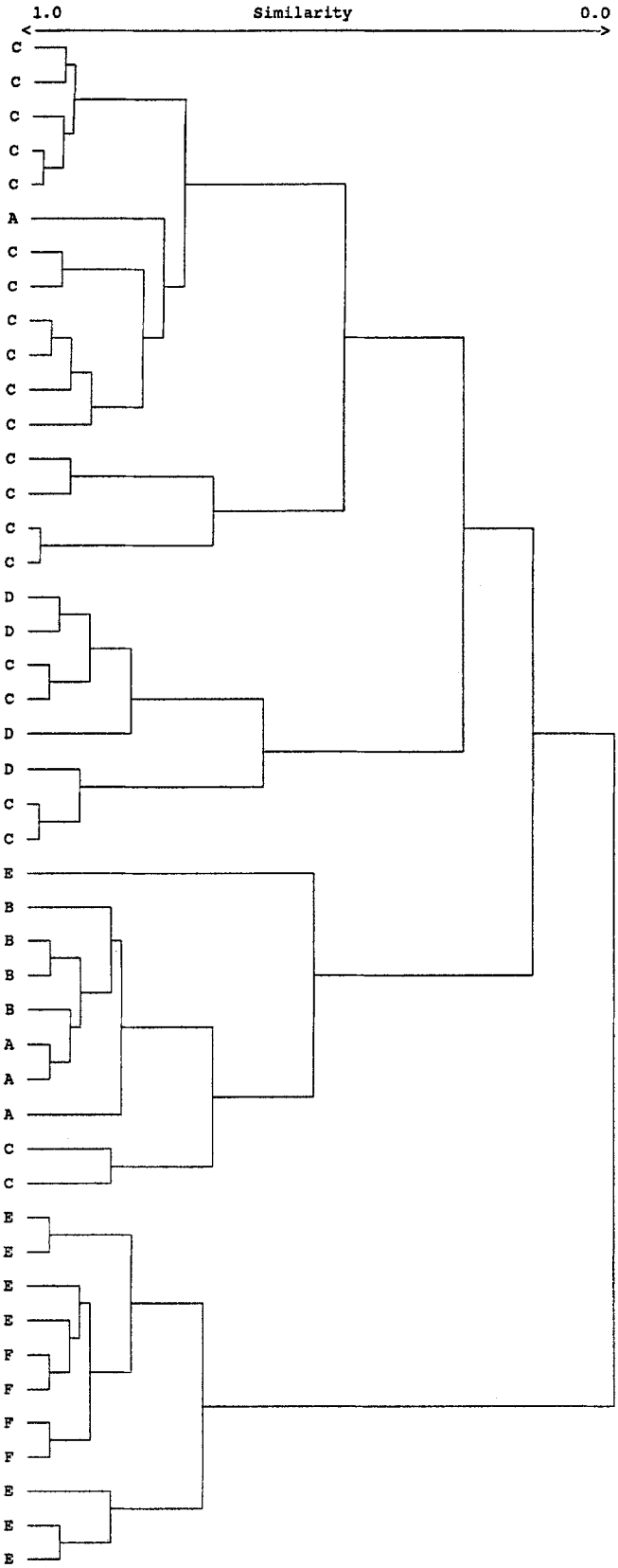

Fig. 4. Dendrogram of HPLC profiles from *B. amyloliquefaciens* 23843 grown under different temperature and shaker speeds. A, 37 °C, 100 r.p.m.; B, 32 °C, 100 r.p.m.; C, 37 °C, 250 r.p.m. (standard condition); D, 32 °C, 250 r.p.m.; E, 37 °C, 400 r.p.m.; F, 32 °C, 400 r.p.m. The distance scale is from 1.0 (identical) to 0.0 (no similarity).

form the data to reduce the amount of irrelevant information and identify relevant information. This approach is important in pattern recognition techniques. Pattern recognition has recently been used to classify mycobacteria on the basis of similarities in HPLC chromatographic elution patterns [4]. Similar techniques have been used to determine microbial contamination during fermentation. Elmroth et al. [6] used a series of chromatographic methods measuring fatty acids, amino acids, and carbohydrates to determine the presence of contaminating organisms in cultures of *Leuconostoc mesenteroides*. In the food industry, pattern recognition had been used successfully to classify a variety of food products, including wines, cheeses, and soy sauce [1,2,8,9,12]. These products were all identified by the pattern of biochemical compounds created by microorganisms during fermentation. Pattern analysis can also lead to the identification of important determinants of appropriate fermentation, e.g. cheddar cheese classification [10].

SIMCA models were developed for both *B. amyloliquefaciens* 1521 and 23843 based on levels of amylase activity found in the fermentation broth under various culture conditions. For the models, data at 280 nm were used and autoscaled before analysis. In a SIMCA model, a principal component model is created for each class. Cultures of unknown identity are then compared with the principal components for each class within the model. Three outcomes are possible in the test of an unknown with a SIMCA model: 1) the unknown falls into a pre-defined class; 2) the unknown will not fit into any class; 3) the unknown will fit more than one class. Confidence limits can be established for the outcome because the class decisions are based on statistical tests. Confidence limits were set at 95%. Results for the two strains are shown in Tables 1 and 2. Table 1 shows the distance between classes and gives an indication of the ability of the model to separate classes. The distance is another way to quantify class separation and is calculated based on the residuals between the classes. This calculation is described in the Pirouette software manual. A small value for between-class distance can indicate poor separation between classes. For the model based on strain 1521, all the classes appeared well separated. With the model based on 23843, little difference was observed between medium and high amylase producers. Use of this model for the 23843 strain would not clearly differentiate all three classes, but may still enable correct identification of poor amylase producers.

TABLE 1

Distance between classes as determined by models based on *B. amyloliquefaciens*

Class*	Low ($U\ ml^{-1}$)	Medium ($U\ ml^{-1}$)	High ($U\ ml^{-1}$)
B. amyloliquefaciens 1521			
Low	0	1.38	6.88
Medium		0	3.16
High			0
B. amyloliquefaciens 23843			
Low	0	5.19	6.94
Medium		0	0.67
High			0

*Classes are based on amylase activity ($U\ ml^{-1}$): low = 0–5; medium = 5–10; high = >10. SIMCA models were created by using HPLC profiles from one strain only.

TABLE 2

Summary of prediction results*

Model	B.A. 1521	B.A. 23843	Overall
B.A. 1521	97.8	44.4	78.6
B.A. 23843	21.7	84.4	79.6

*Percentage of correct identification for all test samples was based on amylase activity models. B.A. = *Bacillus amyloliquefaciens*. Overall = correct identification in low, medium, high, or no-fit classes of all test samples used (98 total).

Table 2 gives results of amylase class predictions for broth samples made with the *B. amyloliquefaciens*. Both models were partially successful in correctly classifying fermentation broth test samples from closely related strains. When the model based on strain 1521 was used to predict performance of strain 23843, the prediction matched the actual activity level only 44.4% of the time. When the model based on 23843 was used to predict amylase levels in strain 1521, the test samples were correctly identified only 21.7% of the time. In most cases where classes were misidentified, the incorrect class was either a lower amylase level or a no-fit designation. Both models correctly designated a no-fit category for all HPLC profiles from other *Bacillus* strains and *E. coli* (data not shown). As a result, the models gave an overall incorrect identification approximately 20% of the time.

CONCLUSION

Pattern recognition techniques were used to distinguish normal and abnormal fermentation patterns of *B. amyloliquefaciens*. Both PCA and HCA enabled the classification of fermentation broth test samples on the basis of their HPLC profiles. Test samples were classified by strain differences (as shown by HPLC profiles of fermentation broth) and growth conditions. Models were then developed using SIMCA methods. The SIMCA model was first described by Wold in 1974 [13] and is included with enhancements in the commercially available Pirouette software package. This method was chosen for the classification of the *B. amyloliquefaciens* fermentation broths because it allows a designation of 'no fit' for divergent test samples that do not match any predefined class. During actual fermentations with *B. amyloliquefaciens*, an abnormal fermentation may diverge for a variety of reasons. Not all abnormal fermentations will have patterns already included in various models available. To be useful in a monitoring program for fermentation, a modeling system must be flexible enough to accommodate new conditions and patterns. This method was less useful for identifying closely related fermentations such as those that might be obtained when a new strain is introduced to

improve yields in a commercial operation. Nonetheless, the success of this method indicates that pattern recognition may be useful in monitoring the progress of a fermentation process, particularly if included in an overall quality control strategy.

ACKNOWLEDGEMENTS

The authors thank Fred Fry, Richard Albert and Mary Lou Tortorello, all of the US Food and Drug Administration (FDA), for critical evaluation and discussion. This publication was partially supported by a Cooperative Agreement, No. FD-000431, from the FDA and the National Center for Food Safety and Technology. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the FDA.

REFERENCES

1 Aishima, T. 1982. Discrimination and cluster analysis of soy sauce GLC profiles. J. Food Sci. 47: 1562–1567.

2 Ashima, T. and S. Nakai. 1987. Pattern recognition of GC profiles for classification of cheese variety. J. Food Sci. 52: 939–942.

3 Bezbaruah, R.L., B.K. Gogoi, R. Pillai and J.N. Nigam. 1991. Amylase producton by three *Bacillus* strains active at alkaline pH. J. Basic Microbiol. 31: 13–20.

4 Butler, W.R., K.C. Jost and J.O. Kilborn. 1991. Identification of mycobacteria by high-performance liquid chromatography. J. Clin. Microbiol. 29: 2468–2472.

5 Colwell, R.R. and B. Austin. 1981. Numerical taxonomy. In: Manual of Methods for General Bacteriology, Ch. 21 (P. Gerhardt, R.G.E. Murray, R.N. Costilow, E.W. Nester, W.A. Wood, N.R. Krieg and G.B. Phillips, eds), American Society for Microbiology, Washington, DC.

6 Elmroth, I., P. Sundin, A. Valeur, L. Larsson and G. Odham. 1992. Evaluation of chromatographic methods for the detection of bacterial contamination in biotechnical processes. J. Microbiol. Methods 15: 215–228.

7 Guthke, R. and R. Rossman. 1991. Fermentation analysis by clustering. Bioprocess Eng. 6: 157–161.

8 Hook, L.A., P.L. Bloch, R.W. Kohlenberger and P.A. Kinningham. 1987. Automated microbial identification system for computer-programmed analysis of radiolabeled protein banding patterns. Develop. Ind. Microbiol. 28: 149–160.

9 Kwan, W.O. and B.R. Kowalski. 1987. Classification of wines by applying pattern recognition to chemical composition data. J. Food Sci. 43: 1320–1323.

10 Pham, A.M. and S. Nakai. 1984. Application of stepwise discriminant analysis to high pressure liquid chromatography profiles of water extract for judging ripening of Cheddar cheese. J. Dairy Sci. 67: 1390–1396.

11 Saner, U. and G. Stephanopoulos. 1992. Application of pattern recognition techniques to fermentation data analysis. Fifth International Conference on Computer Applications and Fermentation Technology, March 29–April 2, 1992, Keystone, CO, USA.

12 Vasconcelos, A.P. and H.J. Chaves das Neves. 1989. Characterization of elementary wines of *Vitis vinifera* varieties by pattern recognition of free amino acid profiles. J. Agric. Food Chem. 37: 931–937.

13 Wold, S. 1976. Pattern recognition by means of disjoint principal components models. Pattern Recognition 8: 127–139.

14 Zerfiridis, G.K. 1990. Extracellular protein profile of *Bacillus cereus* strains. J. Food Prot. 53: 608–609.