Journal of Applied Bacteriology 1994, **76**, 124–134

Rapid identification using pyrolysis mass spectrometry and artificial neural networks of *Propionibacterium acnes* isolated from dogs

R. Goodacre, M.J. Neal, D.B. Kell, L.W. Greenham[1], W.C. Noble[2] and R. G. Harvey[3]

Department of Biological Sciences, University of Wales, Aberystwyth, Dyfed, [1]Department of Pathology and Microbiology, School of Medical Sciences, University of Bristol, Bristol, [2]Department of Microbial Diseases, St John's Institute of Dermatology, St Thomas's Hospital, London, and [3]Godiva Referrals, Cheylesmore, Coventry, UK

4583/05/93: accepted 6 August 1993

R. GOODACRE, M.J. NEAL, D.B. KELL, L.W. GREENHAM, W.C. NOBLE AND R.G. HARVEY. 1994. Curie-point pyrolysis mass spectra were obtained from reference *Propionibacterium* strains and canine isolates. Artificial neural networks (ANNs) were trained by supervised learning (with the back-propagation algorithm) to recognize these strains from their pyrolysis mass spectra; all the strains isolated from dogs were identified as human wild type *P. acnes*. This is an important nosological discovery, and demonstrates that the combination of pyrolysis mass spectrometry and ANNs provides an objective, rapid and accurate identification technique. Bacteria isolated from different biopsy specimens from the same dog were found to be separate strains of *P. acnes*, demonstrating a *within-animal* variation in microflora. The classification of the canine isolates by Kohonen artificial neural networks (KANNs) was compared with the classical multivariate techniques of canonical variates analysis and hierarchical cluster analysis, and found to give similar results. This is the first demonstration, within microbiology, of KANNs as an unsupervised clustering technique which has the potential to group pyrolysis mass spectra both automatically and relatively objectively.

INTRODUCTION

Propionibacterium acnes is widely distributed on the adult human skin, hair, oropharynx and gastrointestinal tract (Willis 1977; Brook and Frazier 1991), and is considered to cause skin disorders and acne. The organism is found on the oily areas of the skin such as the scalp and forehead, and is present, to a lesser degree, on the drier skin areas like arms and legs (McGinley *et al.* 1978, 1980). Furthermore, it can be found at all levels of the pilo-sebaceous duct and on the skin surface (Holland *et al.* 1977; Kearney *et al.* 1984).

Harvey *et al.* (1993) were motivated by the idea that wild type human strains of *P. acnes* might be acquired from dogs and used a microdissection technique to investigate the flora of canine skin surfaces and hair follicles. They reported the recovery from seven (63·6%) of 11 dogs of an organism with cultural characteristics indistinguishable from *P. acnes*. The regional distribution of the organism was similar to that found on man, with higher numbers found on the trunk than on the feet. The number of organisms recovered and their distribution on the dog were consistent with it being part of the normal canine microflora. It was therefore of interest to enquire whether the same strains of *P. acnes* might reside on both human and canine hosts.

Pyrolysis is the thermal degradation of complex material in an inert atmosphere or a vacuum. It causes molecules to cleave at their weakest points to produce smaller, volatile fragments called pyrolysate (Irwin 1982). A mass spectrometer can then be used to separate the components of the pyrolysate on the basis of their mass-to-charge ratio (m/z) to produce a pyrolysis mass spectrum, which can then be used as a 'chemical profile' or fingerprint of the complex material analysed.

Within microbiology, this technique, called pyrolysis mass spectrometry (PyMS), has largely been applied to the characterization of bacterial systems (for reviews see Meu-

Correspondence to: Dr R. Goodacre, Department of Biological Sciences, University of Wales, Aberystwyth, Dyfed SY23 3DA, UK.

zelaar *et al.* 1982; Gutteridge 1987; Berkeley *et al.* 1990). In particular, because of its high discriminatory ability, PyMS has been successfully applied to the inter-strain comparison of a wide range of medically-important bacterial species and groups, including *Corynebacterium* (Meuzelaar *et al.* 1982), *Escherichia coli* (Goodacre *et al.* 1991), *Legionella* (Kajioka and Tang 1984), mycobacteria (Wieten *et al.* 1981a, b), salmonellas (Freeman *et al.* 1990) and streptococci (Magee *et al.* 1991), highlighting the usefulness of the technique in the detection of small differences between microbial samples. One of the major advantages that PyMS has over other diagnostic methods, such as ELISA (Chantler and McIllmurray 1987) and nucleic acid probing (Saano and Lindström 1990), is that it is rapid, both for a single sample and in the (automated) throughput of samples. Typical sample time is less than 2 min.

Within the last year artificial neural networks (ANNs) have been applied successfully to the discrimination of biological samples analysed by pyrolysis mass spectrometry (Goodacre *et al.* 1992; Chun *et al.* 1993; Freeman *et al.* 1993). ANNs have also been employed for the identification of bacteria by flow cytometry (Boddy and Morris 1993) and via biochemical characteristics (Rataj and Schindler 1991). This supervised learning technique has also been used quantitatively to analyse pyrolysis mass spectra in terms of the concentrations of target determinants (Goodacre and Kell 1993; Goodacre *et al.* 1993). ANNs are a well-known means of uncovering complex, nonlinear relationships in multivariate data (Rumelhart *et al.* 1986). The relevant principle is that one can acquire sets of multivariate data using PyMS (e.g. normalized intensities at 150 values of m/z) for bacteria whose identities are known, and train ANNs using the (known) identities as the desired outputs. Once the ANNs are trained they may then be exposed to unknown inputs (i.e. spectra) and will then immediately output the globally optimal best fit to the outputs, in this case in terms of which bacteria the unknown spectra best portray.

In this study we report that the combination of PyMS and ANNs of the above type was able rapidly to identify propionibacteria isolated from canine skin surfaces, from dogs showing no evidence of skin disease, as human wild type *P. acnes*. Furthermore, we investigated the ability of Kohonen artificial neural networks (KANNs) to classify pyrolysis mass spectra of strains of *P. acnes* isolated from dogs, using unsupervised learning, and demonstrate that results similar to those produced by classical multivariate data analysis approaches are obtained.

MATERIALS AND METHODS

Bacterial strains and culture medium

Strains were isolated from the foreheads of two dogs by methods previously described (Holland and Roberts 1974; Puhvel *et al.* 1975; Harvey *et al.* 1993). Briefly, strains were isolated from 6 mm punch samples taken from the foreheads of dogs immediately after euthanasia with intravenous pentobarbitone sodium (Euthatal, Rhône Mérieux, Dagenham). Samples were immediately placed into anaerobic transport medium. Each was homogenized in a sterile tissue grinder with 0·5 ml of nutrient broth. The homogenate was diluted 1/10 and inoculated on agar medium and incubated at 37°C for 5 d in an anaerobic jar fitted with cold catalyst and Gaspak (Oxoid). After incubation several well-isolated colonies were subcultured to ensure pure cultures. The canine bacterial isolates were coded Dog 1/1 (A, B, C, D, E and F) for strains isolated from dog 1 in the first biopsy; Dog 1/2 (A, B, C, D, E and F) in the second biopsy. Dog 2 was also sampled twice, and these strains were coded Dog 2/1 (A, B, C and D), and Dog 2/2 (A).

The reference *Propionibacterium* and human wild type strains included: *P. acnes* NCTC 737, WCN 12/1/81, WCN 12/1/93A, 216 A1, 222A and 74B; *P. avidum* 152 AXA; and *P. granulosum* K4, NCTC 11864, WCN 12/1/93B and 216 D.

The medium used to culture all the *Propionibacterium* strains contained (g l^{-1}): Tryptone Soya Broth (Oxoid), 30; yeast extract (Oxoid), 10; agar No. 1 (Oxoid), 10; and Tween 80 (Sigma), 10. For pyrolysis mass spectrometry strains were cultured anaerobically at 37°C for 5d.

In addition to the *Propionibacterium* genus several other bacterial strains representing other genera were also analysed by pyrolysis mass spectrometry; *Bacillus cereus* DSM 31, *Escherichia coli* W3110, *Klebsiella pneumoniae* (laboratory strain), *Pseudomonas aeruginosa* NCIB 8704 and *Staphylococcus aureus* NCTC 6571. These strains were cultured for 16 h at 37°C aerobically on LabM Blood Agar.

Sample preparation for pyrolysis mass spectrometry

Clean iron-nickel foils (Horizon Instruments Ltd, Ghyll Industrial Estate, Heathfield, E. Sussex) were inserted, with clean forceps, into clean pyrolysis tubes (Horizon Instruments), so that 6 mm was protruding from the mouth of the tube. After incubation, disposable plastic loops were used to remove some organisms from the top of one or more well-isolated colonies, avoiding the plate surface. These organisms were spread on 5 mm of a protruding foil to give a thin uniform surface coating. The samples were oven dried at 50°C for 30 min, then the foils were pushed into the tube using a stainless steel depth gauge so that they were 10 mm from the mouth of the tube. Finally, viton 'O'-rings (Horizon Instruments) were placed on the tubes. All bacteria were analysed in quadruplicate.

Pyrolysis mass spectrometry

The pyrolysis mass spectrometer used in this study was the Horizon Instruments PYMS-200X as described by Aries

et al. (1986). The sample tube carrying the foil was heated, prior to pyrolysis, at 100°C for 5s. The pyrolysate was generated in a vacuum by heating a ferro-magnetic foil carrying the sample with a 3 s radio-frequency (0·4 MHz) current through a pyrolysis coil which surrounds the sample-coated alloy foil. The foil and sample heated rapidly, within 0·5 s, to the temperature corresponding to the Curie-point of the iron-nickel foil. At this temperature, 530°C, the alloy ceased to exhibit ferro-magnetic properties and heating ended; on cooling below the Curie-point, inductive heating resumed, so that the foil-pyrolyser system acted as a thermostatic switch maintaining the sample at the Curie-point, until current ceased to flow through the pyrolysis coil. The pyrolysate then entered a gold-plated expansion chamber heated to 150°C, whence it diffused down a molecular beam tube to the ionization chamber of the mass spectrometer.

The pyrolysate was bombarded with low energy electrons (25 eV) producing both molecular and fragment ions (because low energy was used the majority carried only a single positive charge). Non-ionized molecules were deposited on a cold trap, cooled by liquid nitrogen. The ionized fragments were focussed by the electrostatic lens of a set of source electrodes, accelerated and directed into a quadrupole mass filter. The ions were separated by the quadrupole, on the basis of their mass-to-charge ratio, and detected and amplified with an electron multiplier. The mass spectrometer scans the ionized pyrolysate 160 times at 0·2 s intervals following pyrolysis. Data were collected over the m/z range 51–200, in one tenth of a mass-unit intervals. These were then integrated to give unit mass. Given that the charge of the fragment was unity the mass-to-charge ratio can be accepted as a measure of the mass of pyrolysate fragments. The IBM-compatible PC used to control the PYMS-200X was also programmed (with software provided by the manufacturers) to record spectral information on ion count for the individual masses scanned and the total ion count for each sample analysed.

Data analysis

To remove the effect of sample size differences the data from PyMS were normalized to a total ion count of 2^{16}.

Multivariate data analysis. The normalized data were then processed with the GENSTAT package (Nelder 1979) which runs under Microsoft DOS 5·0 on an IBM-compatible PC. This method has been previously described by MacFie and Gutteridge (1982) and Gutteridge *et al.* (1985). In essence, the first stage was the reduction of the data by principal components analysis (PCA) (Chatfield and Collins 1980; Causton 1987; Gutteridge 1987; Flury and Riedwyl 1988; Martens and Naes 1989; Everitt 1993), which is a well-known technique for reducing the dimensionality of multivariate data whilst preserving most of the variance. Data were reduced by keeping only those principal components (PCs) whose eigenvalues accounted for more than 0·1% of the total variance. Canonical variates analysis (CVA) then separated the samples into groups on the basis of the retained PCs and some *a priori* knowledge of the appropriate number of groupings (MacFie *et al.* 1978; Windig *et al.* 1983). The next stage was the construction of a percentage similarity matrix by transforming the Mahalanobis' distance between *a priori* groups in CVA with the Gower similarity coefficient S_G (Gower 1971). Finally, hierarchical cluster analysis was employed to produce a dendrogram, using average linkage clustering (Gutteridge *et al.* 1985).

Supervised learning using back-propagation artificial neural networks. All ANN analyses were carried out using a user-friendly, neural network simulation program, NeuralDesk (version 1.2) (Neural Computer Sciences, Lulworth Business Centre, Nutwood Way, Totton, Southampton), which runs under Microsoft Windows 3.1 on an IBM-compatible PC. To ensure maximum speed, an accelerator board for the PC (NeuSprint) based on the AT&T DSP32C chip, which effects a speed enhancement of some 100-fold, permitting the analysis (and updating) of some 400 000 weights per s, was used. Data were also processed prior to analysis with the Microsoft Excel 4.0 spreadsheet.

For training the ANN the inputs were the four normalized replicate pyrolysis mass spectra derived from the reference *Propionibacterium* spp. and the five bacteria representing non-*Propionibacterium* spp. (*Bacillus*, *Escherichia*, *Klebsiella*, *Pseudomonas* and *Staphylococcus* spp.), and were further normalized to lie in the range 0–1. For the outputs, *P. acnes* strains were coded as 1000, *P. avidum* as 0100, *P. granulosum* as 0010, and the non-*Propionibacterium* strains as 0001.

The algorithm used was standard back-propagation (Rumelhart *et al.* 1986), running on the accelerator board. This algorithm employs processing nodes (neurons or units), connected using abstract interconnections (connections or synapses). Connections each have an associated real value, termed the weight, that scale signals passing through them. Nodes sum the signals feeding to them and output this sum to each driven connection scaled by a 'squashing' function with a sigmoidal shape. This 'squashing' function $f = 1/(1 + e^{-x})$, where $x = \Sigma$ inputs (see also Kell and Davey 1992).

The training of the network thus consisted of the preparation of a set of pairs of patterns where one half of the pair is input to the network and the other represents the known or expected response. The stimulus pattern is

applied to the network, which is allowed to run until an output is produced at each output node. The differences between the actual output and that expected, taken over the entire set of patterns are fed back through the network in the reverse direction to signal flow (hence back-propagation) modifying the weights as they go. This process is repeated until a suitable level of error is achieved (Rumelhart *et al.* 1986; Wasserman 1989; Simpson 1990). In the present work we used a learning rate of 0·1 and a momentum of 0·9.

The structure of the ANN used in this study consisted of three layers containing 162 nodes made up of the 150 input nodes (normalized pyrolysis mass spectra), four output nodes (representing each of the identities), and one 'hidden' layer containing eight nodes (150-8-4). Each of the 150 input nodes was connected to the eight nodes of the hidden layer which in turn were connected to the output nodes. In addition, the hidden layer and output nodes were connected to the bias, making a total of 1244 connections, whose weights will be altered during training. Before training commenced the input and output nodes were normalized between 0 and + 1, and the connection weights were set to small random values, except the biases which were set to +1 (Wasserman 1989). Each epoch (one complete calculation in the network) represented 1244 connection weight updatings, per training pair, and a recalculation of the root mean squared (RMS) error between the true and desired outputs over the entire training set. A plot of the average RMS error *vs* the number of epochs represented the 'learning curve', and was used to estimate the extent of training. Training can be said to have finished when the network has found the lowest RMS, i.e. what was taken to be the global minimum on the error surface. Finally after training, when the average error had reached 0·005, the trained ANN was interrogated with the spectra from the 68 remaining samples (test data) as well as the 64 spectra of the samples used in the training set; the network then outputs its estimate (best fit) in terms of the identities of the sample.

Unsupervised learning with Kohonen artificial neural networks. KANNs provide an objective way of classifying data through self-organizing networks of artificial neurons. The algorithm forms a (self-organizing) map of the input data which retains information about the relationships between the items (Kohonen 1989; Hecht-Nielson 1990; Hertz *et al.* 1991). Networks were trained by presenting all of the neurons in the single layer with a pyrolysis mass spectrum (i.e. normalized ion counts in the range 51–200 m/z), and ascertaining which neuron has weights which most closely match the input mass spectral data. The weights of the winning neuron and those of its topological neighbours were then adjusted to provide an improved fit. The entire training data set, comprising 17 averaged normalized pyrolysis mass spectra derived from the *Propionibacterium* spp. isolated from the dogs, was repeatedly presented and learnt in this fashion. The size of the neighbourhood which was updated upon each presentation was gradually reduced as training proceeded, as was the size of the adjustment made to the weights. By using different sized networks it was possible to *force* the network to group the data items into different numbers of groups (Erwin *et al.* 1992). Networks on a square grid of four, nine, 16 and 25 nodes were used to group the samples. The networks were allowed to 'wrap around' so that they formed toroidal structures; this was in order to avoid the edge effects which otherwise tend to corrupt very small networks of this type. All experiments were run on a Sun SPARC10, the shortest execution time (for the four node network) was 10 s, and the longest execution time (for the 25 node network) was 4 min. The program used was written in the *C* programming language; no attempt was made to optimize the code.

RESULTS AND DISCUSSION

After the collection of pyrolysis mass spectra the first stage was to perform multivariate statistical methods using the GENSTAT package to establish the relationships between all the strains of propionibacteria. Each of the 11 *Propionibacterium* spp. and 17 canine isolates, each represented by four replicate spectra, were coded to give 28 individual groups; the resulting dendrogram is shown in Fig. 1. In this figure it can be seen that at 70% relative similarity the bacteria cluster into three groups and they have split mainly into the three species of *Propionibacterium* analysed: cluster 1 comprises all the *P. acnes* human wild type strains and all the strains isolated from dogs, with the exception of Dog 1/2 (A); cluster 2 consists of three of the four *P. granulosum* (216 D, K4, and WCN 12/1/93B) strains; finally, *P. granulosum* NCTC 11864 and *P. avidum* 152 AXA form a third cluster. These results indicate that the propionibacteria strains isolated from dogs are related to human wild type *P. acnes* except for the strain Dog 1/2 (A) which failed to cluster with any of the *Propionibacterium* spp. and is an outlier; it was therefore omitted from any further analyses.

The next stage was to investigate the variation of canine propionibacteria flora; therefore, only the bacteria isolated from dogs (omitting sample Dog 1/2 (A)) were analysed with GENSTAT. The results are displayed as canonical variates plots (CVA) (Fig. 2) and as a dendrogram (Fig. 3). When the first two canonical variates (Fig. 2a) are plotted, displaying 77% of the total variance, three groups can be seen, which are also evident in the dendrogram (Fig. 3). These are related to the origin of the isolates: group 1

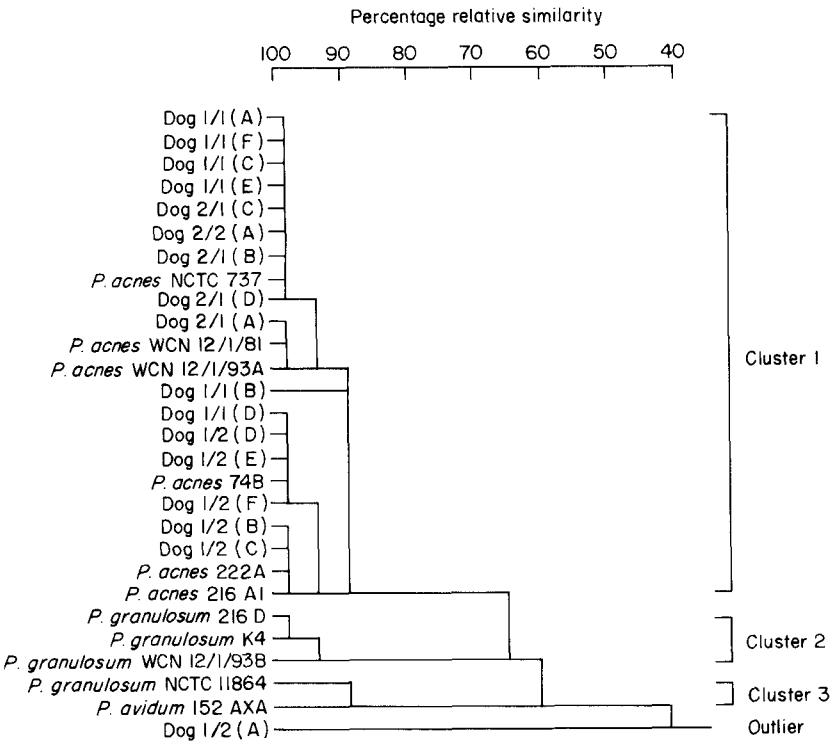

Fig. 1 Dendrogram representing the relationships between all the propionibacterial strains studied, based on pyrolysis mass spectrometry data analysed by GENSTAT

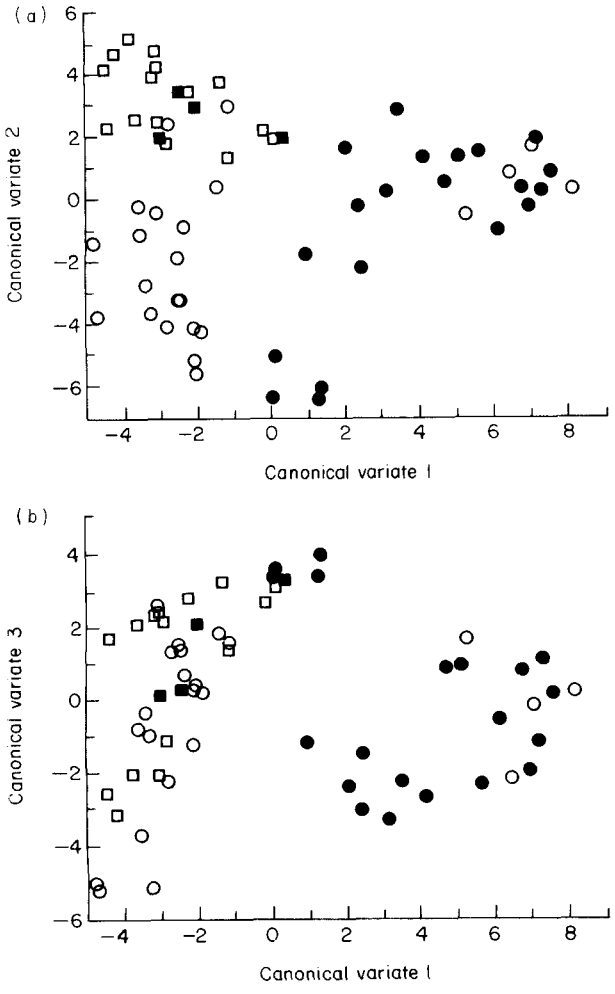

Fig. 2 Ordination plot based on pyrolysis mass spectrometry data analysed by GENSTAT showing the relationship between the propionibacterial strains isolated from dogs. The first and second canonical variates (a) accounted for 76·93% of the total variance, and the third (b) accounted for 9·85% (86·78% total). ○, Dog 1/1; ●, Dog 1/2; □, Dog 2/1; ■, Dog 2/2

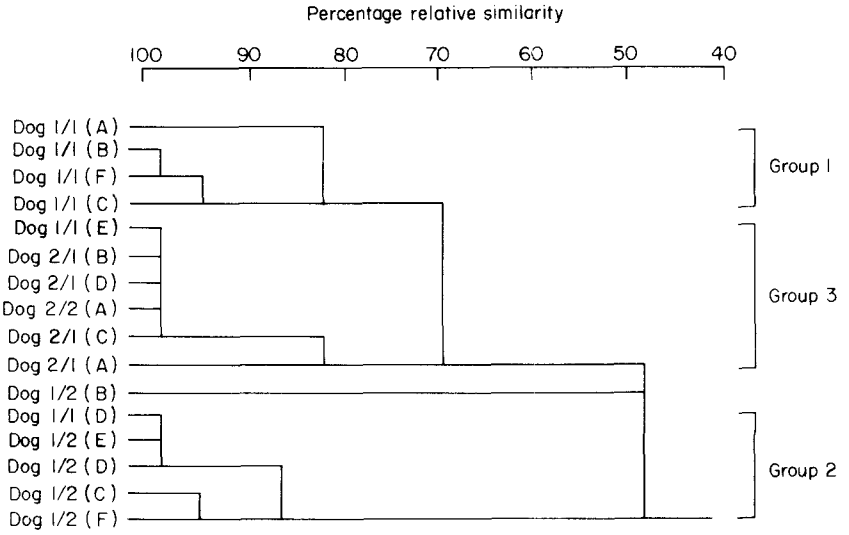

Fig. 3 Dendrogram representing the relationships between the canine isolates, based on pyrolysis mass spectrometry data analysed by GENSTAT

(open circles) comprises strains from Dog 1/1; group 2 (closed circles) of Dog 1/2 isolates and Dog 1/1 (D); isolates from Dog 2 sampled on both occasions form the third group (open and closed squares), Dog 1/1 (E) also clusters with this group. When the third canonical variate, which accounts for 9·85% of the total variance, is viewed (Fig. 2b) only two clusters are observed and groups 1 and 3 coincide. This phenomenon is also reflected in Fig. 3 where the branch splitting groups 1 and 3 in the dendrogram occurs at 68% relative similarity compared to the split-off point of group 2 from groups 1 and 3 at 47%. CVA results can be interpreted statistically to discriminate populations based on the 95% tolerance region constructed around each population mean by the χ-squared distribution on two degrees of freedom (Krzanowski 1988). This area can be represented by drawing a circle of radius 2·448 canonical variates (CV) units. Therefore, because the group means in Fig. 2a are separated by more than 4·9 CV units the propionibacteria isolated from dogs appear to be three distinct strains of *P. acnes*.

It is not easy, by simple inspection of CVA plots alone (Fig. 2), to obtain a correct interpretation of the classification structure of the canine isolates and it is necessary simultaneously to examine the dendrogram produced by GENSTAT (Fig.3). In order to remove the somewhat arbitrary nature of the discrimination produced by these methods, it would be desirable to have a data analysis technique which would 'automatically' put the pyrolysis mass spectra of bacteria (or any other samples) into groups with no need for recourse to *a priori* information.

KANNs provide the desired automatic grouping of data. Initially, networks of four nodes were trained with the 17 averaged normalized pyrolysis mass spectra of the canine isolates. After training the 17 isolates separated into four groups (Table 1 and Fig. 4). The canine isolate 1/2 (A) was sufficiently different from all the other samples to be placed in a group on its own (group D), and was thus discarded as an outlier; this is in agreement with the previous GENSTAT analyses (Fig. 1). By retraining with larger networks (nine, 16 and 25 nodes), finer distinctions between samples were detected.

Group A, comprising mainly strains from Dog 1, is a robust group of isolates which retains its identity, apart from isolate 1/1 (A), until the size of the network reaches 25 nodes when it degrades into five sub-groups A^3 to A^7. This group contains all the members of group 2 from GENSTAT analyses (Figs 2 and 3) and also the isolates Dog 1/1 (A and F) and 2/1 (A). Group C from KANN

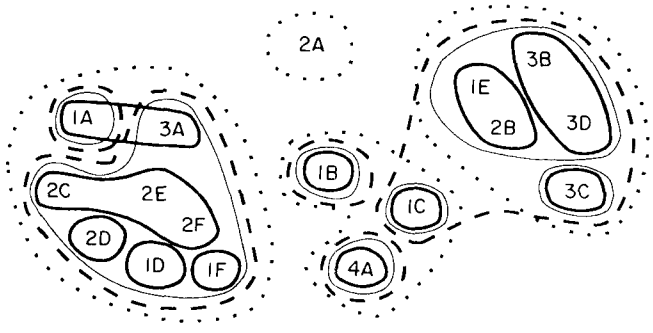

Fig. 4 Topological contour map of groups from Kohonen nets trained with pyrolysis mass spectrometry data of the canine isolates. ····, Groups from 2 × 2 network; – – –, groups from 3 × 3 network; ——, groups from 4 × 4 network; ——, groups from 5 × 5 network; 1, Dog 1/1; 2, Dog 1/2; 3, Dog 2/1; 4, Dog 2/2. Map is correct only in topology

Table 1 Groups produced by Kohonen artificial neural networks trained on pyrolysis mass spectral data of the canine isolates

2 × 2 network		3 × 3 network		4 × 4 network		5 × 5 network	
Group	Members*	Group	Members	Group	Members	Group	Members
A	1A, 1D, 1F, 2C, 2D, 2E, 2F, 3A	A^1	1D, 1F, 2C, 2D, 2E, 2F, 3A		1D, 1F, 2C, 2D, 2E, 2F, 3A	A^3	2C, 2E, 2F
		A^2	1A		1A	A^4	1A, 3A
						A^5	1D
						A^6	1F
						A^7	2D
B	1B, 1C, 4A	B^1	1B		1B		1B
		B^2	4A		4A		4A
				B^3	1C		1C
C	1E, 2B, 3B, 3C, 3D	C^1	1E, 2B, 3B, 3C, 3D, 1C	C^2	1E, 2B, 3B, 3D	C^4	1E, 2B
				C^3	3C		3C
						C^5	3B
						C^6	3D
D	2A	Strain Dog 1/2 (A) removed because it was an outlier					

* 1, Dog 1/1; 2, Dog 1/2; 3, Dog 2/1; 4, Dog 2/2.

analyses is also relatively robust (Fig. 4) and mirrors group 3 that was produced by GENSTAT (Fig. 3); it contains the canine isolates 2/1 (B, C and D), 1/1 (E) and 1/2 (B) but not strain 2/2 (A) which fell into group B (Table 1). This strain grouped with bacteria isolated from dog 2 sampled on the first occasion when analysed by GENSTAT (Fig. 3). The separation of the isolates sampled from dog 2, from different biopsies, may well be a useful differentiation and demonstrates the potential of KANNs to distinguish more finely discriminating features than the other linear data analysis techniques that were used.

Group B contains strains 1/1 (B and C) and 2/2 (A) which rapidly degrades (Table 1 and Fig. 4) and is less well defined. The membership of isolate 1/1 (C) drifts to group B when the network size is nine nodes; this change is due to failure of the topological ordering process of the network (Kohonen 1989), which may have been caused by the presence of distortions in the surface described by the network in weight space (Hertz *et al.* 1991; Erwin *et al.* 1992). In other words, the distribution of the inputs cannot perfectly map onto the distribution of the nodes due to the inappropriate connectivity of the latter. When 16 and 25 nodes were used, however, strain 1/1 (C) formed the single member group B^3.

KANNs grouped the samples in a similar way to statistical techniques. Although KANNs provide no truly quantitative information about the similarity of the samples within a group, they provide qualitative information about the groups present. This information is very useful for classification tasks. When the results are presented as a contour map (Fig. 4) they are easier to interpret, and provide less scope for subjective considerations as concrete boundaries between groups are drawn.

PyMS and multivariate data analysis can be used to identify bacteria. The inclusion of suitable reference strains which fall into groups of unidentified isolates will allow their identification, a technique termed 'operational fingerprinting' (Meuzelaar *et al.* 1982). This process is often subjective because it relies on the interpretation of complicated CVA plots and dendrograms and the question therefore arises as to whether it is possible objectively to identify bacteria from their pyrolysis mass spectra.

ANNs were therefore trained with normalized ion intensities from the pyrolysis mass spectra from the training set (the three reference *Propionibacterium* spp. and the non-propionibacteria); the 16 bacteria, represented by 64 mass spectra, were coded at the output nodes as described above. The 150-8-4 ANNs were trained using the standard back-propagation algorithm, and the effectiveness of training was expressed in terms of the average RMS error between the actual and the desired outputs; examples of these 'learning curves' are shown in Fig. 5. Training was stopped after the average error had reached 0·005 and we interrogated using the pyrolysis mass spectral data from all bacteria from both the training and test sets. Training was effected three times, using randomized, small initial values for the starting weights; because the three curves were found essentially to superimpose, despite the randomized starting connection weights, it is clear that training was executed (i.e. the error surface in weight space was negotiated) in a rather repro-

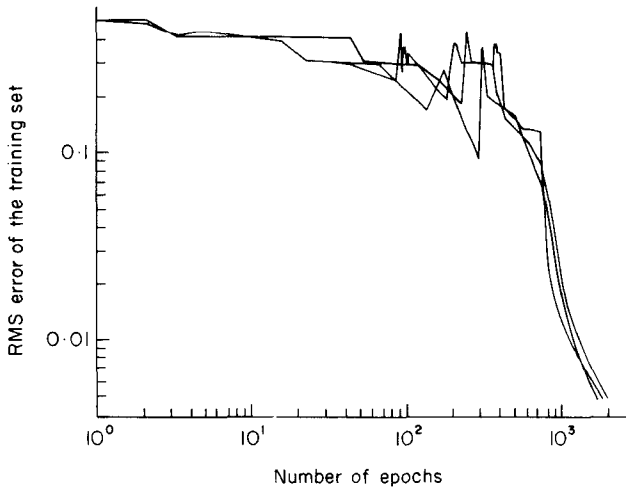

Fig. 5 The learning curve(s) for neural networks employing the standard back-propagation algorithm with one hidden layer consisting of eight nodes, and trained to assess the identity of the strains isolated from dogs

ducible and rapid manner. The actual time taken, using the NeuSprint accelerator card, to train to an RMS error of 0·005 was only 5–6 min.

These learning curves (Fig. 5) display some very interesting neurodynamics; for the first 1000 epochs the average error decreased from 0·5 to between 0·015 and 0·02. After approximately 1700–1900 epochs the average error reached 0·005 and training was stopped.

When training had ceased (i.e. as determined by the attainment of an average error of 0·005 over the entire training set) the network was interrogated with the normalized ion intensities of the pyrolysis mass spectra from the training and test sets. Not surprisingly, the network's estimate of the bacterial identity of the training set was the same as the known identities (Table 2). The results of the network's final analysis of the unknown test set (given as the average of the outputs for each of the four mass spectra replicates) are shown in Table 2. It is clear that *all* the canine isolates are identified non-subjectively as human wild type *P. acnes*, including the outlying strain 1/2 (A) observed in Figs 1 and 4. The 'correct' identification of strain 1/2 (A), typed as *P. acnes* previously using biochemical characteristics (Harvey *et al.* 1993), is perhaps surprising as we provided a fourth node of non-*Propionibacterium* spp. for the identification of *very different* isolates and one might have expected some outliers to appear there. The isolate Dog 1/1 (D) was scored as 0·54 at the first node on the output layer (denoting identity as *P. acnes*) and 0·47 at the third node (as *P. granulosum*). Although, the magnitude of these numbers is similar because the first node activation is greater than the third its identity is taken to be *P. acnes*.

Goodacre and Berkeley (1990) showed that PyMS was able to discriminate between bacteria that differed only in the presence or absence of antibiotic-resistance plasmids, illustrating the high discriminatory power of PyMS; this study also exemplifies this, but more importantly shows that isolates taken from the same dog in different specimens are separate strains of *P. acnes*. This demonstrates a within-animal variation, which may be because the propionibacterial microflora of Dog 1 was different in different follicles. Alternatively the original strain isolated in the first instance may have mutated, or may represent a 'contaminant' from the owner, veterinarian or laboratory worker. PyMS is a technique which highlights phenotypic differences between samples; in order to ascertain the exact phylogeny of the strains isolated from Dog 1 the homology of the DNA or 16S-rRNA of these organisms may have to be examined (Austin and Priest 1986).

Kohonen artificial neural networks, as an objective classification technique using unsupervised learning, were compared with canonical variates analyses and dendrograms produced by GENSTAT. So far as we are aware, this is the first time that KANNs have been applied to the classification of bacteria; this technique gave results similar to those produced by GENSTAT, except that KANNs were able to separate isolates from Dog 2 recovered from separate biopsy specimens. Furthermore, this classification method has the potential to group the pyrolysis mass spectra of bacterial (or any other) sample automatically.

Finally, the confirmation that the canine isolates are *P. acnes* is an important discovery and raises questions regarding the relationship between man and the dog, particularly if *P. acnes* is a member of the normal canine microflora, as seems to be the case (Harvey *et al.* 1993). The domestication of the dog and, in more recent times, the increasingly close physical relationships between man and the dog may have led to anthropozoonotic or zoonotic spread. Whatever the case, the organism appears to be well adapted to the dog and is not thought to be associated with any *canine* dermatoses (Muller *et al.* 1989), although this may reflect differences in follicular anatomy as much as host–pathogen relationships *per se*. The ability rapidly to identify these bacteria by PyMS and multivariate techniques will inevitably lead to further investigations to examine these interesting points. We therefore conclude that the combination of PyMS and various types of neural networks provides an objective, rapid and accurate discriminatory technique.

ACKNOWLEDGEMENTS

RG, DBK and MJN are supported by the Biotechnology Directorate of the UK SERC, under the terms of the LINK scheme in Biochemical Engineering, in collaboration with Horizon Instruments, Neural Computer Sciences and

Table 2 Identity of the bacteria used in the test set and training set as judged by the artificial neural network

	Network's estimate			
	Propionibacterium			
Strain	*acnes*	*avidum*	*granulosum*	Non-*Propionibacterium* spp.
Training set:				
Propionibacterium acnes NCTC 737	1·00	0·00	0·00	0·00
WCN 12/1/81	1·00	0·00	0·00	0·00
WCN 12/1/93A	1·00	0·00	0·00	0·00
216 A1	1·00	0·00	0·00	0·00
222A	0·99	0·00	0·01	0·00
74B	0·99	0·00	0·00	0·00
P. avidum 152 AXA	0·00	0·98	0·00	0·01
P. granulosum K4	0·00	0·00	1·00	0·00
P. granulosum				
NCTC 11864	0·00	0·01	0·99	0·00
WCN 12/1/93B	0·01	0·00	0·99	0·00
216 D	0·00	0·00	1·00	0·00
Bacillus cereus	0·00	0·00	0·00	1·00
Escherichia coli	0·00	0·00	0·00	1·00
Klebsiella pneumoniae	0·00	0·00	0·00	1·00
Pseudomonas aeruginosa	0·00	0·00	0·00	1·00
Staphylococcus aureus	0·00	0·00	0·00	1·00
Test set:				
Dog 1/1 (A)	1·00	0·00	0·00	0·00
Dog 1/1 (B)	1·00	0·00	0·00	0·00
Dog 1/1 (C)	1·00	0·00	0·01	0·00
Dog 1/1 (D)	0·54	0·00	0·47	0·00
Dog 1/1 (E)	0·99	0·00	0·01	0·00
Dog 1/1 (F)	0·98	0·00	0·00	0·01
Dog 1/2 (A)	0·99	0·01	0·00	0·00
Dog 1/2 (B)	0·93	0·00	0·07	0·00
Dog 1/2 (C)	1·00	0·00	0·00	0·00
Dog 1/2 (D)	0·74	0·00	0·28	0·00
Dog 1/2 (E)	0·90	0·00	0·08	0·00
Dog 1/2 (F)	1·00	0·00	0·00	0·00
Dog 2/1 (A)	1·00	0·00	0·00	0·00
Dog 2/1 (B)	1·00	0·00	0·00	0·00
Dog 2/1 (C)	1·00	0·00	0·00	0·00
Dog 2/1 (D)	1·00	0·00	0·00	0·00
Dog 2/2 (A)	0·93	0·00	0·10	0·00

Zeneca plc. RGH was in receipt of a Clinical Studies Trust Fund Award awarded by the British Small Animal Veterinary Association.

REFERENCES

Aries, R.E., Gutteridge, C.S. and Ottley, T.W. (1986) Evaluation of a low-cost, automated pyrolysis-mass spectrometer. *Journal of Analytical and Applied Pyrolysis* **9**, 81–98.

Austin, B. and Priest, F. (1986) *Modern Bacterial Taxonomy*. pp. 74–81. Wokingham: Van Nostrand Reinhold.

Berkeley, R.C.W., Goodacre, R., Helyer, R.J. and Kelley, T. (1990) Pyrolysis-MS in the identification of micro-organisms. *Laboratory Practice* **39**, 81–83.

Boddy, K. and Morris, C.W. (1993) Neural network analysis of flow cytometry data. In *Flow Cytometry in Microbiology* ed. Lloyd, D. pp. 159–169. London: Springer-Verlag.

Brook, I. and Frazier, E.H. (1991) Infections caused by *Propionibacterium* species. *Reviews of Infectious Diseases* **113**, 819–820.

Causton, D.R. (1987) *A Biologist's Advanced Mathematics*. pp. 48–72. London: Allen & Unwin.

Chantler, S.M. and McIllmurray, M.B. (1987) Labelled-antibody methods for the detection and identification of microorganisms. *Methods in Microbiology* **19**, 273–332.

Chatfield, C. and Collins, A.J. (1980) *Introduction to Multivariate Analysis*. pp. 57–81. London: Chapman & Hall.

Chun, J., Atalan, E., Ward, A.C. and Goodfellow, M. (1993) Artificial neural network analysis of pyrolysis mass-spectrometric data in the identification of *Streptomyces* strains. *FEMS Microbiology Letters* **107**, 321–325.

Everitt, B.S. (1993) *Cluster Analysis*. London: Edward Arnold.

Erwin, E., Obermayer, K. and Schulten, K. (1992) Self-organizing maps: ordering, convergence properties and energy functions. *Biological Cybernetics* **67**, 47–55.

Flury, B. and Riedwyl, H. (1988) *Multivariate Statistics: a Practical Approach*. pp. 181–233. London: Chapman & Hall.

Freeman, R., Goodfellow, M., Gould, F.K., Hudson, S.J. and Lightfoot, N.F (1990) Pyrolysis-mass spectrometry (Py-MS) for the rapid epidemiological typing of clinically significant bacterial pathogens. *Journal of Medical Microbiology* **32**, 283–286.

Freeman, R., Goodacre, R., Sisson, P.R., Magee, J.G., Ward, A.C. and Lightfoot, N.F. (1993) Rapid identification of species within the *Mycobacterium tuberculosis* complex with an artificial neural network trained on pyrolysis mass spectral data. *Journal of Medical Microbiology* (in press).

Goodacre, R. and Berkeley, R.C.W. (1990) Detection of small genotypic changes in *Escherichia coli* by pyrolysis mass spectrometry. *FEMS Microbiology Letters* **71**, 133–138.

Goodacre, R. and Kell, D.B. (1993) Rapid and quantitative analysis of bioprocesses using pyrolysis mass spectrometry and neural networks: application to indole production. *Analytica Chimica Acta* **279**, 17–26.

Goodacre, R., Beringer, J.E. and Berkeley, R.C.W. (1991) The use of pyrolysis-mass spectrometry to detect the fimbrial adhesive antigen F41 from *Escherichia coli* HB101(pSLM204). *Journal of Analytical and Applied Pyrolysis* **22**, 19–28.

Goodacre, R., Kell, D.B. and Bianchi, G. (1992) Neural networks and olive oil. *Nature* **359**, 594.

Goodacre, R., Edmonds, A.N. and Kell, D.B. (1993) Quantitative analysis of the pyrolysis-mass spectra of complex mixtures using artificial neural networks: application to casamino acids in glycogen. *Journal of Analytical and Applied Pyrolysis* **26**, 93–114.

Gower, J.C. (1971) A general coefficient of similarity and some of its properties. *Biometrics* **27**, 857–874.

Gutteridge, C.S. (1987) Characterization of microorganisms by pyrolysis mass spectrometry. *Methods in Microbiology* **19**, 227–272.

Gutteridge, C.S., Vallis, L. and MacFie, H.J.H. (1985) Numerical methods in the classification of microorganisms by pyrolysis mass spectrometry. In *Computer-assisted Bacterial Systematics* ed. Goodfellow, M., Jones, D. and Priest, F. pp. 369–401. London: Academic Press.

Harvey, R.G., Noble, W.C. and Lloyd, D.H. (1993) Distribution of propionibacteria on dogs: a preliminary report of the findings on 11 dogs. *Journal of Small Animal Practice* **34**, 80–84.

Hecht-Nielsen, R. (1990) *Neurocomputing*. pp. 138–147. Massachusetts: Addison-Wesley.

Hertz, J., Krogh, A. and Palmer, R.G. (1991) *Introduction to the Theory of Neural Computation*. pp. 236–244. California: Addison-Wesley.

Holland, K.T. and Roberts, C.D. (1974) A technique for sampling micro-organisms from the pilo-sebaceous duct. *Journal of Applied Bacteriology* **37**, 289–296.

Holland, K.T., Cunliffe, W.J. and Roberts, C.D. (1977) *Acne vulgaris*: an investigation into the number of anaerobic diphtheroids and members of the Micrococcaceae in acne and normal skin. *British Journal of Dermatology* **96**, 623–626.

Irwin, W.J. (1982) *Analytical Pyrolysis: A Comprehensive Guide*. New York: Marcel Dekker.

Kajioka, R. and Tang, P.W. (1984) Curie-point pyrolysis-mass spectrometry of *Legionella* species. *Journal of Analytical and Applied Pyrolysis* **6**, 59–68.

Kearney, J.N., Harnby, D., Gowland, G. and Holland, K.T. (1984) The follicular distribution and abundance of resident bacteria on human skin. *Journal of General Microbiology* **130**, 797–801.

Kell, D.B. and Davey, C.L. (1992) On fitting dielectric spectra using artificial neural networks. *Bioelectrochemistry and Bioenergetics* **28**, 425–434.

Kohonen, T. (1989) *Self-Organization and Associative Memory*. Berlin: Springer-Verlag.

Krzanowski, W.J. (1988) *Principles of Multivariate Analysis: A User's Perspective*. pp. 369–400. Oxford: Oxford University Press.

MacFie, H.J.H. and Gutteridge, C.S. (1982) Comparative studies on some methods for handling quantitative data generated by pyrolysis. *Journal of Analytical and Applied Pyrolysis* **4**, 175–204.

MacFie, H.J.H., Gutteridge, C.S. and Norris, J.R. (1978) Use of canonical variates in differentiation of bacteria by pyrolysis gas-liquid chromatography. *Journal of General Microbiology* **104**, 67–74.

Magee, J.T., Hindmarch, J.M. and Nicol, C.D. (1991) Typing of *Streptococcus pyogenes* by pyrolysis mass spectrometry. *Journal of Medical Microbiology* **35**, 304–306.

Martens, H. and Naes, T. (1989) *Multivariate Calibration*. New York: John Wiley.

McGinley, K.G., Webster, G.E. and Leyden, J.J. (1978) Regional variations of cutaneous propionibacteria. *Applied and Environmental Microbiology* **35**, 62–66.

McGinley, K.G., Webster, G.E., Ruggieri, M.R. and Leyden, J.J. (1980) Regional variations in density of cutaneous propionibacteria: correlations of *Propionibacterium acnes* populations with sebacious secretions. *Journal of Clinical Microbiology* **12**, 672–675.

Meuzelaar, H.L.C., Haverkamp, J. and Hileman, F.D. (1982) *Pyrolysis Mass Spectrometry of Recent and Fossil Biomaterials*. Amsterdam: Elsevier.

Muller, G.H., Kirk, R.W. and Scott, D.W. (1989) *Small Animal Dermatology*. pp. 244–248. Philadelphia: WB Saunders.

Nelder, J.A. (1979) *Genstat Reference Manual*. University of Edinburgh: Scientific and Social Service Program Library.

Puhvel, S.M., Reissner, R.M. and Amirian, D.A. (1975) Quantification of bacteria in isolated pilosebaceous follicles in normal skin. *Journal of Investigative Dermatology* **65**, 525–531.

Rataj, T. and Schindler. J. (1991) Identification of bacteria by a multilayer neural network. *Binary* 3, 159–164.

Rumelhart, D.E., McClelland, J.L. and the PDP Research Group (1986) *Parallel Distributed Processing. Experiments in the Microstructure of Cognition.* Cambridge, Massachusetts: MIT Press.

Saano, A. and Lindström, K. (1990) Detection of rhizobia by DNA–DNA-hybridization from soil samples: problems and perspectives. *Symbiosis* 8, 61–73.

Simpson, P.K. (1990) *Artificial Neural Systems.* Oxford: Pergamon Press.

Wasserman, P.D. (1989) *Neural Computing: Theory and Practice.* New York: Van Nostrand Reinhold.

Wieten, G., Haverkamp, K., Meuzelaar, H.L.C., Engel H.B.W. and Berwald, L.G. (1981a) Pyrolysis mass spectrometry: a new method to differentiate between the mycobacteria of the 'Tuberculosis' complex and other mycobacteria. *Journal of General Microbiology* 122, 109–118.

Wieten, G., Haverkamp, K., Engel, H.B.W. and Berwald, L.G. (1981b) Application of pyrolysis mass spectrometry to the classification and identification of mycobacteria. *Reviews of Infectious Diseases* 3, 871–877.

Willis, A.T. (1977) *Anaerobic Bacteriology: Clinical and Laboratory Practice.* London: Butterworths.

Windig, W., Haverkamp, J. and Kistemaker, P.G. (1983) Interpretation of a set of pyrolysis mass spectra by discriminant analysis and graphical rotation. *Analytical Chemistry* 55, 387–391.