

Rapid identification of species within the *Mycobacterium tuberculosis* complex by artificial neural network analysis of pyrolysis mass spectra

R. FREEMAN, R. GOODACRE†, P. R. SISSON* J. G. MAGEE*, A. C. WARD and N. F. LIGHTFOOT*

Department of Microbiology, The Medical School and *Regional Public Health Laboratory, Newcastle-upon-Tyne and †Department of Biological Sciences, University of Wales, Aberystwyth

Summary. An artificial neural network (ANN) was trained to distinguish between *Mycobacterium tuberculosis* and *M. bovis* with averaged pyrolysis mass spectra from duplicate subcultures of four strains of each of these species, each pyrolysed in triplicate. Once trained, the ANN was interrogated with spectrum data from the original organisms (the “training set”) and from 26 other mycobacterial isolates (the “challenge set”) of the *M. tuberculosis* complex (MTBC). Eight strains of *M. bovis* and 13 of *M. tuberculosis*, whether sensitive or variously resistant to antituberculosis drugs, were identified in agreement with conventional identification. Four strains of “*M. africanum*” were identified as *M. bovis*. Of two atypical *M. tuberculosis* strains from South India, one was identified as *M. tuberculosis* and the other as *M. bovis*. Six strains of BCG proved heterogeneous; two gave equivocal identifications, three were identified as *M. bovis* and one was identified as *M. tuberculosis*.

Introduction

DNA probes are used increasingly for the rapid identification of mycobacterial isolates. However, currently available probes cannot differentiate between the species of the *Mycobacterium tuberculosis* complex (MTBC).^{1,2} Rapid high-performance liquid chromatography (HPLC) of mycolic acids from mycobacterial cultures also fails to differentiate species within the MTBC.³ Diagnostic discrimination between *M. tuberculosis* and *M. bovis* still relies on additional biochemical tests, usually for nitrate reduction and niacin production. The latter test involves the use of carcinogenic materials. Less hazardous alternatives involve either further subculture to test micro-aerophilicity or animal inoculation, which add several weeks to identification.

We report here the rapid identification of species within the MTBC by artificial neural network (ANN) analysis of pyrolysis mass spectrometry (PMS) data.

ANNs are used increasingly to analyse data comprising measurements of several parameters for each specimen. This type of multivariate or pattern data is found in PMS spectra, which comprise ion intensities for a range of masses, each reflecting some aspect of cell composition. ANNs comprise sets of multiply-linked simple computation units termed nodes (figs. 1 and 2). The spectrum data are applied as quantitative input signals to a series of input nodes, one for each parameter measured. The input node is connected

to each of a series of “hidden” nodes. Each node connection carries a connection weighting function that quantitatively modifies the signal. The signals received by the hidden nodes are summed, a bias is added and the total is transformed *via* a sigmoidal

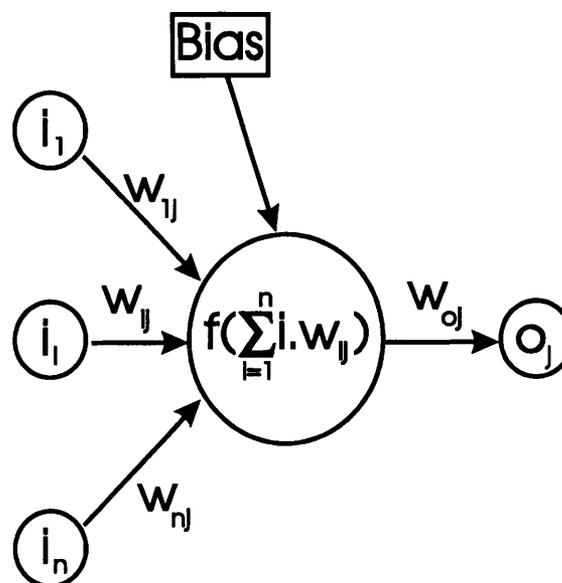


Fig. 1. Signal processing at an artificial network node. Weighted signal strengths ($i_n \times W_{nj}$) from the n nodes of the previous layer are summed and a bias signal is added. The sum (\times) is subjected to a sigmoidal squashing transformation, usually of the form $f = 1/(1 + e^{-x})$. The transformation product is then passed to the node(s) of the next layer, after multiplication by the appropriate connection weighting(s). In this case, the signal is passed to a single output node O_j , after multiplication by the connection weighting W_{oj} .

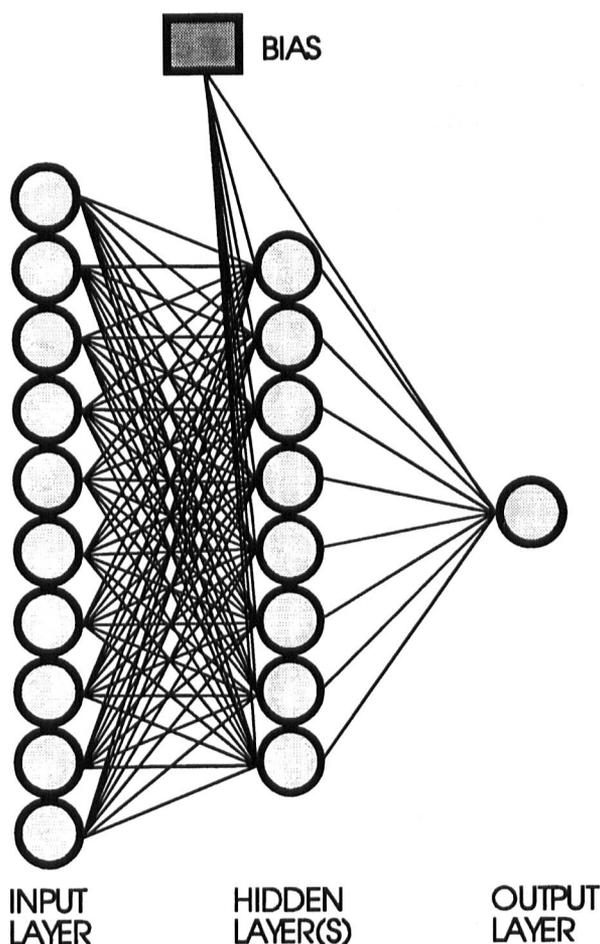


Fig. 2. An artificial neural network comprising 10 input nodes, one hidden layer of eight nodes, and a single output node. Nodes are represented by shaded circles, and the lines indicate passage of the node signals from the left to right of the diagram. Adjacent layers are fully connected, but other architectures are possible. In this study, the input layer comprised 150 nodes, one for each mass intensity from mass 51 to 200 of the pyrolysis mass spectrum. The hidden and output layers were as illustrated, and the output node signal was coded as 1 for an identification as *M. tuberculosis* or 0 for *M. bovis*.

activation function. Further connection weightings are applied, and the signal is passed to an output node, where the numeric result of the successive weighting, summation and transformation steps is displayed.

In an untrained ANN, the weightings are set to small random values. To train the ANN, data from a series of examples of each class of sample to be recognised is submitted. A back-propagation training program then alters the connection weightings successively over many adjustment cycles, optimising them until each class of samples gives a desired distinct output. The success of training is assessed by presenting further data from a challenge set. If the error level in identification of the challenge set is considered acceptable, then the ANN weightings can be recorded and recalled later for use in identification. ANN training requires much computation, and is usually performed on fast computers or transputers to speed processing. However, recall of the optimised weightings into an ANN and identification of a new unknown requires minimal computation, and is effectively in-

stantaneous. The principles of ANN analysis are described in detail elsewhere, as are specific microbiological applications of ANNs to PMS data.⁴⁻⁷

Materials and methods

Mycobacterial cultures and isolates

The strains examined comprised: eight clinical isolates of *M. bovis*; four of *M. africanum*; 13 of *M. tuberculosis* (four sensitive to all antituberculosis agents, nine resistant to one or more agents); two of the South India variant of *M. tuberculosis*; six clinical isolates of BCG; and one reference culture of *M. tuberculosis* H37 RV. These were cultured in duplicate on Middlebrook 7H10 Agar slopes (Difco Laboratories) for 13 days at 35°C to give a visible lawn of confluent growth. All strains had been identified previously by standard cultural and biochemical methods. All clinical isolates, including those of BCG, were from different patients.

Pyrolysis mass spectrometry

Each subculture was smeared, in triplicate, on to Ni-Fe foils (Horizon Instruments Ltd, Heathfield, Sussex) which were inserted into pyrolysis tubes (Horizon Instruments) and heated at 80°C for 5 min. Sample tubes were processed as a single batch on a Horizon Instruments PYMS 200X pyrolysis mass spectrometer, pyrolysing at 530°C for 4 s. For each sample, the pyrolysis sequence number, integrated ion counts at unit mass intervals from 51 to 200 amu and the total ion count were recorded on to floppy disk.

Data analysis

Spectra were normalised by an iterative technique to eliminate variations due to differences in the amount of sample pyrolysed. The normalised spectra from the three foils from each subculture were averaged. Normalised averaged spectra for 16 samples, comprising the duplicate subcultures of four strains of *M. bovis* and four strains of *M. tuberculosis* (the training set), were then used to train an ANN with the Neural Desk program (Neural Computer Sciences, Lulworth Business Centre, Totton, Southampton) running under Microsoft Windows version 3.1 on an IBM-compatible PC with an accelerator board (NeuSprint, Neural Computer Sciences) giving a speed enhancement of *c.* 100-fold. Data were submitted to the ANN via the Microsoft Excel 4.0 spreadsheet program.

The ANN comprised an input layer of the 150 normalised averaged ion intensities from the mass spectrum and one hidden layer of eight nodes. This was trained with the standard back-propagation algorithm, as previously described,⁸ with *M. tuberculosis* giving an output of 1 and *M. bovis* an output of 0. The effectiveness of training was expressed as the

root mean square (RMS) error between the actual and the desired outputs for the training set. Training was judged to be complete when the RMS error equalled 0.001.

The trained ANN was interrogated with the normalised averaged spectra for each of the duplicate subcultures of remaining samples (the challenge set) and the 16 spectra of the training set.

Results

Training of the ANN was accomplished in < 20 min and interrogation of the trained ANN with individual spectra was effectively instantaneous. The ANN correctly identified 13 clinical strains of *M. tuberculosis*, whether sensitive or resistant to common antituberculosis drugs, with scores of 0.865–1.00 (mean 0.984; SD 0.032) and the eight strains of *M. bovis* with scores of 0–0.070 (mean 0.002; SD 0.018). One subculture of *M. tuberculosis* H37 RV gave a code of 0, the output specified for *M. bovis*, and the other subculture gave an equivocal result of 0.49. Of the two South India variants of *M. tuberculosis*, both subcultures of one strain gave codes of 0 and the subcultures of the other gave codes of 1 and 0.95. The strains of *M. africanum* gave codes of 0–0.007 (mean 0.002; SD 0.003). The BCG strains gave codes of 1–0.999 (mean 0.294; SD 0.439). Three of these strains were unequivocally identified as *M. bovis*, one was unequivocally identified as *M. tuberculosis*. Equivocal or discordant scores were obtained for the duplicate subcultures of the two remaining BCG strains.

Discussion

Wieten *et al.*⁹ first noted that PMS could be used to differentiate mycobacteria within the *M. tuberculosis* complex, but the data processing strategies available at that time made interpretation difficult. We subsequently confirmed these observations by discriminant analysis.¹⁰ The current study shows that ANNs trained with PMS data can rapidly, reproducibly and ac-

curately differentiate species within the MTBC. Since PMS itself is a rapid method, full identification of cultures can be accomplished within a day. All cultures of *M. tuberculosis* and *M. bovis* were identified correctly by the ANN, irrespective of their susceptibility to antituberculosis agents. The consistent identification of *M. africanum* as *M. bovis* is compatible with longstanding observations that these species are synonymous.¹¹

The equivocal result for *M. tuberculosis* H37 RV further demonstrates that repetitively subcultured laboratory strains may differ significantly from recent clinical isolates.¹¹ The South India variants of *M. tuberculosis* are a highly atypical, heterogenous group that share at least one characteristic (sensitivity to thiophen-2-carboxylic acid hydrazide) with *M. bovis*.¹² The heterogeneity of BCG strains is well recognised¹³ and the variability of the ANN identification for this organism was not unexpected.

These anomalies may also reflect the fact that the training set contained only strains of classical *M. tuberculosis* and *M. bovis*. It must be emphasised that the results from ANNs are dictated by the content of the training set. Inclusion of substantial numbers of strains of the South India variant of *M. tuberculosis* may be unnecessary in the UK, where it accounts for < 2% of isolates; however, such strains would be essential components of the training set for an ANN to be used in identification of Asian strains. The ANN might then not only identify such strains as *M. tuberculosis* but also as the variant, provided that it had been trained with sufficient examples.

This small study shows the potential power of PMS and ANN analysis in mycobacterial characterisation within the MTBC. PMS has also been shown to distinguish rapidly and accurately between *M. xenopi* and *M. avium-intracellulare*.¹⁴ The possibilities of obtaining PMS spectra within a few hours of the receipt of a mycobacterial culture and identification by interrogation of a series of ANNs with this data may have important benefits for the diagnosis and treatment of mycobacterial diseases.

We thank Professor D. Kell for his help and advice with ANNs. R.G. is 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 Zeneca plc.

References

1. Heifets L. Gene-probe test should not be considered final in *Mycobacterium tuberculosis* identification. *J Clin Microbiol* 1989; **27**: 229.
2. Fries JWU, Patel RJ, Piessens WF, Wirth DF. Detection of untreated mycobacteria by using polymerase chain reaction and specific DNA probes. *J Clin Microbiol* 1991; **29**: 1774–1747.
3. Butler WR, Jost KC, Kilburn JO. Identification of mycobacteria by high-performance liquid chromatography. *J Clin Microbiol* 1991; **29**: 2468–2472.
4. Rumelhart DE, McLlland JL, and the PDP Research Group. Parallel distributed processing. Explorations in the microstructure of cognition. Cambridge, MA, MIT Press. 1986.
5. Wasserman PD. Neural computing: theory and practice. New York: Van Nostrand Reinhold. 1989.
6. Chun J, Atalan E, Ward AC, Goodfellow M. Artificial neural network analysis of pyrolysis mass spectrometric data in the identification of *Streptomyces* strains. *FEMS Microbiol Lett* 1993; **107**: 321–325.
7. Goodacre R, Kell DB. Rapid and quantitative analysis of bioprocesses using pyrolysis mass spectrometry and neural

- networks: application to indole production. *Analyt Chim Acta* 1993; **279**: 17-26.
8. Goodacre R, Kell DB, Bianchi G. Neural networks and olive oil. *Nature* 1992; **359**: 594.
 9. Wieten G, Haverkamp J, Meuzelaar HLC, Engel HWB, Berwald LG. Pyrolysis mass spectrometry: a new method to differentiate between the mycobacteria of the 'tuberculosis complex' and other mycobacteria. *J Gen Microbiol* 1981; **122**: 109-118.
 10. Sisson PR, Freeman R, Magee JG, Lightfoot NF. Differentiation between mycobacteria of the *Mycobacterium tuberculosis* complex by pyrolysis mass spectrometry. *Tubercle* 1991; **72**: 206-209.
 11. Goodfellow M, Wayne LG. Taxonomy and nomenclature. In: Ratledge C, Stanford J (eds) *The biology of the mycobacteria*, vol 1: Physiology, identification and classification. London, Academic Press. 1982: 471-521.
 12. Yates MD, Collins CH, Grange JM. "Classical" and "Asian" variants of *Mycobacterium tuberculosis* isolated in South East England 1977-1980. *Tubercle* 1982; **62**: 55-61.
 13. Osborn TW. Changes in BCG strains. *Tubercle* 1983; **64**: 1-13.
 14. Sisson PR, Freeman R, Magee JG, Lightfoot NF. Rapid differentiation of *Mycobacterium xenopi* from mycobacteria of the *Mycobacterium avium-intracellulare* complex by pyrolysis mass spectrometry. *J Clin Pathol* 1992; **45**: 355-357.