The detection of caffeine in a variety of beverages using Curie-point pyrolysis mass spectrometry and genetic programming

Royston Goodacre* and Richard J. Gilbert

Institute of Biological Sciences, University of Wales, Aberystwyth, Ceredigion, UK SY23 3DD. Tel: +44 (0)1970 621947; Fax: +44 (0)1970 622354; E-mail: rrg@aber.ac.uk

Received 8th February 1999, Accepted 14th May 1999

Freeze dried coffee, filter coffee, tea and cola were analysed by Curie-point pyrolysis mass spectrometry (PyMS). Cluster analysis showed, perhaps not surprisingly, that the discrimination between coffee, tea and cola was very easy. However, cluster analysis also indicated that there was a secondary difference between these beverages which could be attributed to whether they were caffeine-containing or decaffeinated. Artificial neural networks (ANNs) could be trained, with the pyrolysis mass spectra from some of the freeze dried coffees, to classify correctly the caffeine status of the unseen spectra of freeze dried coffee, filter coffee, tea and cola in an independent test set. However, the information in terms of which masses in the mass spectrum are important was not available, which is why ANNs are often perceived as a 'black box' approach to modelling spectra. By contrast, genetic programs (GPs) could also be used to classify correctly the caffeine status of the beverages, but which evolved function trees (or mathematical rules) enabling the deconvolution of the spectra and which highlighted that *m*/*z* 67, 109 and 165 were the most significant masses for this classification. Moreover, the chemical structure of these mass ions could be assigned to the reproducible pyrolytic degradation products from caffeine.

Introduction

Pyrolysis mass spectrometry (PyMS) is a method which can be used to produce a biochemical fingerprint of complex biological material under study. However, the interpretation of these multivariate fingerprints is not possible by simple visual inspection and has usually been carried out by conventional 'unsupervised' chemometric tools such as principal components analysis, discriminant analysis and hierarchical cluster analysis. More recently the development of artificial neural networks (ANNs) has provided an alternative, 'supervised' learning method[1] and although ANNs have been shown to be effective tools for the identification, discrimination and quantification of foodstuffs from PyMS data (for examples, see refs. 2–8) the information in terms of which masses in the mass spectrum are important is not *readily* available, and ANNs are often perceived as a 'black box' approach to modelling spectra. There is a need therefore to deconvolve such complex spectra and use systems that produces 'rules' that are readily comprehensible.

A genetic algorithm (GA) is an optimisation method based on the principles of Darwinian selection.[9–12] A population of individuals, each representing the parameters of the problem to be optimised as a string of numbers or binary digits, undergoes a process analogous to evolution in order to derive an optimal or near-optimal solution. The parameters stored by each individual are used to assign it a *fitness*, a single numerical value indicating how well the solution using that set of parameters performs. New individuals are generated from members of the current population by processes analogous to asexual and sexual reproduction.

Asexual reproduction, or *mutation*, is performed by randomly selecting a parent with a probability proportional to its fitness, then randomly changing one or more of the parameters it encodes. The new individual then replaces a less-fit member of the population, if one exists. Sexual reproduction, or *crossover*, is achieved by randomly selecting two parents at a rate proportional to their fitnesses, and generating two new individuals by copying parameters from one parent, and switching to the other parent after a randomly-selected point. The two new individuals then replace less fit members of the population as before. The above procedure is repeated, with the overall fitness of the population improving at each generation, until an acceptably-fit individual is produced.

A genetic program (GP) is an application of the GA approach to derive mathematical equations, logical rules or program functions automatically.[13–15] Rather than representing the solution to the problem as a string of parameters, as in a conventional GA, a GP uses a tree structure. The leaves of the tree, or *terminals*, represent input variables or numerical constants. Their values are passed to *nodes*, at the junctions of branches in the tree, which perform some numerical or program operation before passing on the result further towards the root of the tree. Mutations are performed by selecting a parent and modifying the value or variable returned by a terminal, or changing the operation performed by a node. Crossovers are performed by selecting two parents and grafting sub-trees at randomly selected nodes within their trees. The new individuals so generated again replace less-fit members of the population.

Within our laboratory the application of GPs to the deconvolution of PyMS spectra has been very successful. We have been able to find particular masses which were characteristic of the adulteration of orange juice with sucrose in the range 0.5–20 g L^{-1};[15] the masses chosen were found to be characteristic of orange juice rather than sucrose and these were *m*/*z* 68 and 119, which had non-linear and linear relationships, respectively, with the level of adulteration. Studies on the classification of oral bacteria belong to the *Eubacterium* genus showed that specific masses from the breakdown products of polysaccharide (*m*/*z* 58), lipids (*m*/*z* 61), pentose anhydrosugar (*m*/*z* 85) and proteins (*m*/*z* 91) could be used to identify hospital oral abscess isolates successfully.[16] Finally, GPs have been used to assess the physiological state of a wide range of *Bacillus* species and to elucidate that sporulated bacteria possess the

dipicolinic acid biomarker, which has a highly characteristic pyrolysate fragment at *m*/*z* 105,[17,18] which is a pyridine ketonium ion $C_6H_3ON^+$.

In the present study we investigate the ability of the genetic programming approach to classify whether a diverse range of beverages (instant and filter coffees, tea and cola) contain caffeine or are decaffeinated drinks, and to elucidate which pyrolysate fragments were characteristic for this separation.

Experimental section

Preparation of beverages

All beverages were purchased from a local supermarket. Details of the beverages are given in Table 1. The coffees and teas were prepared by adding 20 mL of hot water (≈80 °C) to 1 g of powder, grinds or leaf (any solids from grinds or leaf were removed by a simple centrifugation step). The cola drink was analysed directly.

Pyrolysis mass spectrometry

Five microlitre aliquots of the above samples were evenly applied on to iron-nickel foils to give a thin uniform surface coating. Prior to pyrolysis the samples were oven-dried at 50 °C for 30 min. Each sample was analysed in triplicate. For full operational procedures see refs. 19–21. The sample tube carrying the foil was heated, prior to pyrolysis, at 100 °C for 5 s. Curie-point pyrolysis was at 530 °C for 3 s, with a temperature rise time of 0.5 s. The data from PyMS were collected over the *m*/*z* range 51–200. These conditions were used for all experiments. Data were normalised as a percentage of the total ion count to remove the influence of sample size *per se*.

Table 1 Beverages used in this study

Type of drink	Brand name	Batch	Caffeine containing	Training or test set
Freeze-dried coffee	Nescafé	1	No	Training
Freeze-dried coffee	Nescafé	1	Yes	Training
Freeze-dried coffee	Nescafé Gold Blend	1	No	Training
Freeze-dried coffee	Nescafé Gold Blend	1	Yes	Training
Freeze-dried coffee	Lyons	1	No	Training
Freeze-dried coffee	Lyons	1	Yes	Training
Freeze-dried coffee	Kenco	1	No	Test
Freeze-dried coffee	Kenco	1	Yes	Test
Freeze-dried coffee	Nescafé	2	Yes	Test
Freeze-dried coffee	Nescafé	3	Yes	Test
Freeze-dried coffee	Nescafé Gold Blend	2	Yes	Test
Filter coffee	Kenco	1	No	Test
Filter coffee	Kenco	1	Yes	Test
Tea	Lyons	1	No	Test
Tea	Lyons	1	Yes	Test
Cola	Diet Coke	1	No	Test
Cola	Diet Coke	1	Yes	Test

Cluster analysis

The initial stage involved the reduction of the dimensionality of the PyMS data by principal components analysis (PCA).[22,23] PCA is a well known technique for reducing the dimensionality of multivariate data whilst preserving most of the variance, and Matlab was employed to perform PCA according to the NIPALS algorithm. Discriminant function analysis (DFA; also known as canonical variate analysis (CVA)) then discriminated between groups on the basis of the retained principal components (PCs) and the *a priori* knowledge of which spectra were replicates (therefore 17 classes represented by 51 spectra were used to form the models), and thus this process does not bias the analysis in any way.[24] These methods were implemented using Matlab version 4.2c. l (The MathWorks, Inc., Natick, MA, USA), which runs under Microsoft Windows NT on an IBM-compatible PC.

Multilayer perceptrons (MLPs)

When the desired responses (targets) associated with each of the inputs (spectra) are known then the system may be 'supervised'. The goal of supervised learning is to find a model that will correctly associate the inputs with the targets; this is usually achieved by minimising the error between the target and the model's response (output).

The input data to the standard back-propagation multi-layer perceptrons (MLPs)[25–30] contained either (1) the full replicate PyMS spectra (150 *m*/*z* intensities) or (2) the number of variables (masses) were reduced using product moment correlation (PMC). PMC is a method that uses linear transformations to decide which variables (*x*) are most strongly related to the output data (*y*) being modelled.[15] The first 20 mass inputs were used and ranked according to their PMC (most important first), and these were; 67, 109, 77, 63, 55, 82, 65, 165, 64, 62, 91, 75, 83, 84, 114, 116, 112, 107, 90, and 102.

Both input data types were partitioned into training and test sets (details are given in Table 1). For this binary classification problem the output data were encoded such that caffeine containing beverages were coded as 1 and decaffeinated drinks as 0.

In order to assess the reproducibility of the neural network approach 10 MLPs were trained using a user-friendly, neural network simulation program, NeuFrame version 3,0,0,0 (Neural Computer Sciences, Southampton, Hants), which runs under Microsoft Windows NT on an IBM-compatible PC.

Genetic programming

The GPs in this study were designed for classification purposes; that is to say, to be able to indicate correctly whether a beverage was caffeine containing or not. Therefore the GP implementations used here used four arithmetic node functions 'add', 'subtract', 'multiply', and 'protected divide' (where $n/0 = 1$), plus the function 'if-then-else'. All GP rules were derived using an in-house program[15] following a procedure similar to[31] which runs under Microsoft Windows NT on an IBM-compatible PC.

The same 18 spectra (6 × 3 replicates) were used to evolve the GP as for the MLPs trained above; whilst the test set also consisted of 33 spectra. Again there were two sets of inputs to the GPs, and these were either the full PyMS spectra, or the PMC reduced set of 20 masses.

The GP used five independent sub-populations (demes) with a 5% migration every 10 generations. The deme size was set to 5000 individuals (therefore the population size was 25 000). The maximum number of generations was set to 250, although only between 5 and 100 were ever used before convergence of

the GP occurred. Convergence was achieved when the RMS (root mean squared) error between the training set estimates was within 0.1% of the real values. In order for relatively simple rules to be developed, the tree complexity was constrained by setting the maximum number of nodes used to 50 and the maximum depth of the trees to only five layers.

In order to facilate a rational interpretation, 100 GP rules were evolved by differing the random seeding for the initial trees. After evolution the calibrated GPs were challenged with the training set and the 33 spectra in the unseen test set.

Results and discussion

Pyrolysis mass spectra of Nescafé regular and decaffeinated coffees are shown in Fig. 1. The spectra are qualitatively identical although differences can be observed; for example, the caffeine containing spectrum contains a large peak at *m*/*z* 109 (marked with an asterisk) which is much smaller in the spectrum from the decaffeinated coffee.

The first stage was to perform discriminant analysis, as detailed above PCA was employed as a dimensionality reduction step and 10 PCs were extracted (accounting for 99.88% of the total variance) and the resulting score vectors were subsequently used as inputs to the DFA algorithm; the resulting ordination plot is shown in Fig. 2A, where only the replicate means are shown. This figure shows clearly that the first discriminant function (DF1) is all that is needed to separate the three types of beverages. DF2 separates the coffee beverages (including freeze dried and grinds) from tea and cola, and appears to be slightly important for differentiating between caffeine-containing (1) and decaffeinated (0) tea and cola, and to a much lesser degree for the different coffees. However, when the third DF is viewed against DF1 (Fig. 2B) very clear separation between the caffeine-containing and decaffeinated drinks was observed (as indicated in this figure). Caffeine is relatively abundant in coffee beans: 1–2% dry weight after roasting, and 4.5–5.1% in freeze-dried coffee,[32] which means in a 230 ml (8 oz) cup, instant coffee contains 60–85 mg caffeine, brewed coffee 65–120 mg, whilst decaffeinated coffee contains only 1–4 mg caffeine.

The next stage was to train MLPs, using the standard back-propagation algorithm, with the six normalised triplicate PyMS data from the training sets as the inputs, scaled for each input node such that the lowest mass was set to 0 and the highest mass to 1, and the output binary encoded so that a caffeine containing coffee was coded 1 and decaffeinated as 0. Furthermore, for the full spectral MLPs 8 nodes were used in the single hidden layer and this topology can be represented as a 150-8-1 MLP architecture; by contrast, the PMC reduced spectra used only 4 nodes, a 20-4-1 topology. Training was conducted 10 times (a) to observe whether this process was reproducible and (b) to use the 'committee' approach for prediction,[28] where the outputs from the ten MLPs were averaged. Training was stopped when the RMS error between the observed and expected outputs was 0.01, which typically took *ca*. 40 epochs. Initially MLPs were trained until the RMSEF (RMS error of formation) was 0.005 (0.5%), and their ability to generalise was assessed on the test set. It was found that MLPs trained until the RMSEF was 0.01 (1%) were still able to generalise well, and since these MLPs obviously took less time to train and were less likely to overfit the input data (*i.e.*, fitting to noise or the fitting of a model to outliers[33,34]), all MLPs were trained until the RMS was 0.01 (1%).

After training, the 150-8-1 and 20-4-1 MLPs were challenged with the training and test sets. Not surprisingly the caffeine

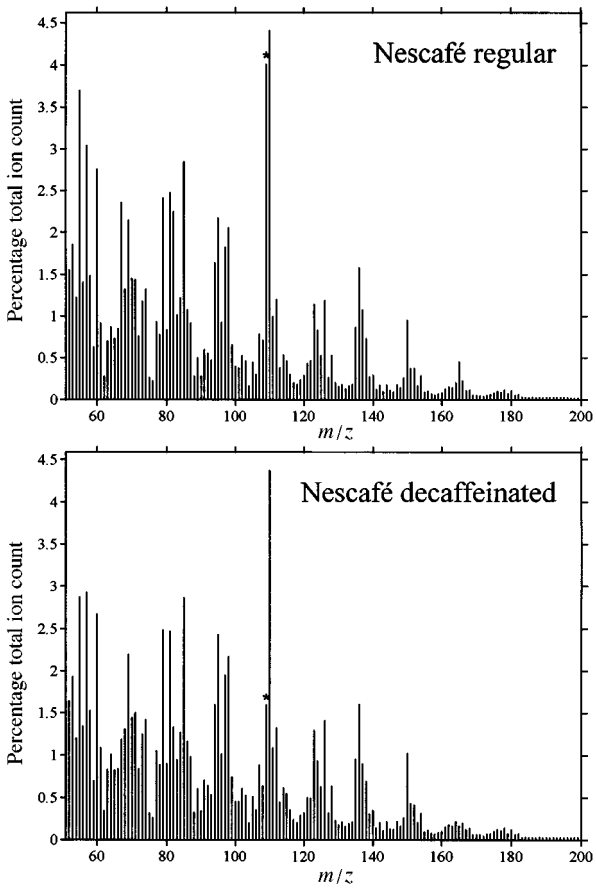

Fig. 1 Pyrolysis mass spectra of the freeze-dried coffee Nescafé regular and Nescafé decaffeinated.

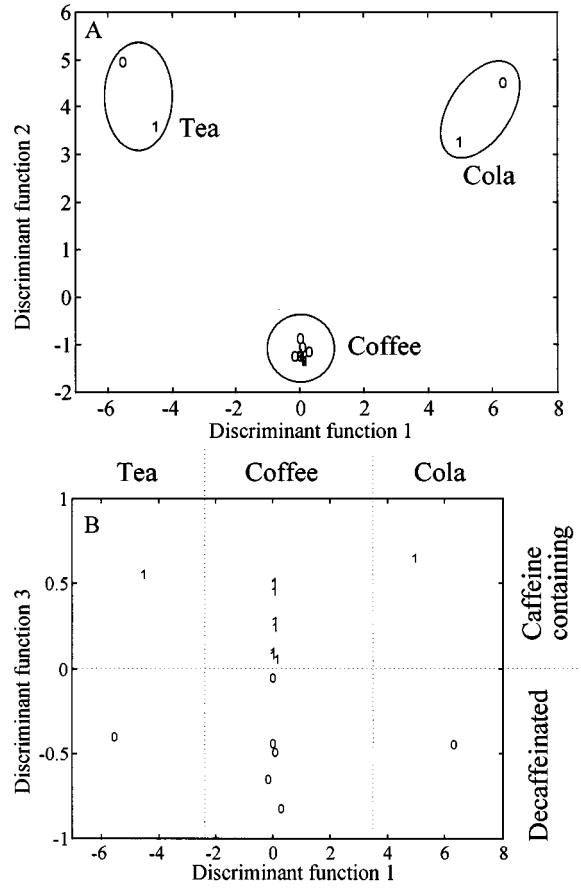

Fig. 2 Discriminant function analysis biplots based on PyMS data showing the relationship between the 17 different beverages. (A) DF1 *vs.* DF2, (B) DF1 *vs.* DF3. The numbers refer to whether the drink contains caffeine (1) or not (0).

status for the training set was correctly classified, but more importantly the 33 spectra from the unseen test set (11 beverages in triplicate) were also correctly identified. The training set contained only freeze-dried coffee (Nescafé, Nescafé Gold Blend, Lyons; Table 1), whilst the test set contained freeze-dried coffee from a different supplier (Kenco) as well as coffee grinds (Kenco filter coffee), tea (Lyons) and cola (Diet Coke). That the MLPs were able to correctly identify the caffeine status of beverages outside their knowledge realm (that is to say, to extrapolate) suggests that the MLPs extracted information from the PyMS spectra that was characteristic of caffeine *per se* rather than any subtle differences between the coffee brands. The question arises as to what that information is. However, although MLPs are a very powerful supervised learning method they are non-transparent and do not readily give any information in terms of which masses in the mass spectrum are important. Whilst it is true that the information used by the MLPs can nominally be found in their weights (the connections between the input, hidden and output layers); this information is very abstract and almost impossible to extract realistically, especially when these ANNs are *interconnected* and for the 150-8-1 and 20-4-1 MLPs contained 1217 and 89 weights, respectively.

In order to investigate which pyrolysates are important indicators of caffeine status one could simply subtract the normalised spectrum of a decaffeinated coffee from one which contains caffeine; this has been done for the Nescafé brand and the subtraction spectrum is shown in Fig. 3. It can be seen that ions which are indicative of caffeine are *m*/*z* 55, 67, 82, 109, 137 and 165. Subtraction spectra are very useful in this type of analysis; however the mass ions shown are not unique to the pyrolysis mass spectra of Nescafé regular (Fig. 1, top) and are readily observed in the PyMS spectra of Nescafé decaffeinated (Fig. 1, bottom). Therefore there is a need for a supervised analysis, operating in a similar fashion to MLPs, but which can be used to *deconvolve* such spectra by producing mathematical 'rules' that are readily comprehensible.

We therefore used the same training and test sets to supervise genetic programming to distinguish between the caffeine-containing and decaffeinated beverages. The same output encoding was used as before where a caffeine-containing coffee was coded as 1 and decaffeinated as 0. The evolution of the GPs was as detailed above, and in order to assess the reproducibility of the rules 100 separate GPs were evolved. A correct identity for decaffeinated beverage was scored as [rule output] < 0.05, whilst a caffeine containing beverage was scored as [rule output] > 0.95.

Table 2 shows the results of the GPs. For the PMC-reduced mass sectral data 37 of the 100 GPs were able to identify all beverages and extrapolate like the MLPs. As one would expect, the ability of the GP rules to identify the coffees correctly was much higher (86 for the freeze-dried coffee and slightly less accurate at 80 for the filter coffee) than the tea (51) or cola (64). Table 2 also gives details of the number of rules able to classify these various beverages, and a very similar trend was seen.

The function trees (rules) generated from the GPs that correctly identified all beverages were all slightly different and ranged from very simplistic ones like:

if (M67 + 0.71355) ≥ (8.08299 −5.71881)
then (3.99370/3.99370)
else (M55-M55)

which further simplifies to 'if *M67* ≥ 1.65063 then 1 else 0' to more complex ones like:

if ((M116 * M91) + M62) ≥ (M67 - M83)
then (1.39045 - 1.39045)
else (M165 / M165)

which simplifies to 'if ((*M116* * *M91*) + *M62*) ≥ (*M67* - *M83*) then 0 else 1'

Both rules use *m*/*z* 67 as an important mass, the first rule as a simple threshold value of whether this mass is greater than or equal to 1.65, whilst the second rule uses *m*/*z* 67 as the predominant mass but it was baseline-corrected using *m*/*z* 83. The ability of GP models to use mass ions with low discriminatory power for baselining purposes was observed in an earlier study which used GP to discriminate between sporulated and vegetative bacteria.[18] The threshold value which *m*/*z* 67 has to be greater than is provided by the calculations using *m*/*z* 62, 91 and 116. Since these are all the intensities of masses (rather than constant numerical values) it is likely that these act as internal calibrants and correct for any chemical or baseline differences between the beverages in the training set. It is likely that such rules would be more robust to chemical differences in a new test set than very simplistic rules.

The above two GPs rules from the PMC-reduced spectra and the other 35 function trees, which classified correctly all beverages in the independent test, were inspected and the frequency of the predominant masses for caffeine were plotted against the pyrolysis mass spectrum of pure caffeine (obtained from Sigma, Poole, Dorset, UK) (Fig. 4A). The PyMS spectrum of caffeine has intense, characteristic peaks at *m*/*z* 55, 67, 82, 109, 137 and 165, which is in agreement with electron impact MS studies on pure caffeine,[35] and with our own subtraction spectra (Fig. 3). The molecular ion for caffeine should be observed at *m*/*z* 194 but it is missing; this is because of the poor pyrolysate transfer of high molecular weight molecules in the expansion chamber and long (≈20 cm) molecular beam tube in our pyrolysis mass spectrometer. It can be seen that the GP selects *m*/*z* 67, 109 and 165 as being very discriminatory for this purine alkaloid and these masses were picked with a frequency of 18, 11 and 10 respectively. Whilst for the full spectral GPs frequencies of 10, 7 and 24 were observed for *m*/*z* 67, 109 and 165, respectively. For these genetic programs all but two trees used only one of these three mass ions when classifying the caffeine-status of the beverages; the other two used the product of *m*/*z* 67 and *m*/*z* 165. Whilst the other dominant masses in the

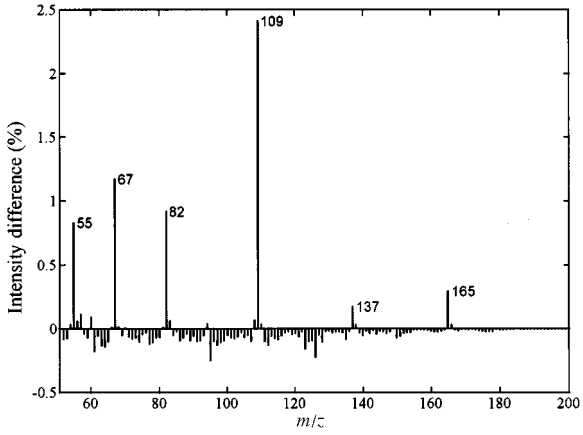

Fig. 3 Subtraction spectrum: Nescafé regular − Nescafé decaffeinated.

Table 2 Number of rules able to classify correctly all caffeine containing or decaffeinated beverages from the test set

	Number correct (out of 100 runs)	
Type of beverage	GPs evolved with 20 masses ranked according to their PMC values	GPs evolved with the full spectra (150 mass intensities)
All drinks	37	36
Freeze-dried coffee	86	96
Filter coffee	80	87
Tea	51	56
Cola	64	61

caffeine spectra of *m*/*z* 55, 82 and 137 were 'ignored'. The question arises as to why these latter peaks are uninformative as far as the GP is concerned since they do arise from the pyrolysis of caffeine.

The masses used by the 63 GP rules (trained with 20 *m*/*z* intensities) that failed to identify correctly the caffeine status of the beverages are shown in Fig. 4B. As well as choosing *m*/*z* 67, 109 and 165 (frequencies 30, 5 and 8, respectively), the mass ion at 82 was also selected 15 times as being a positive indicator of caffeine (and also its naturally occurring isotopes at 83, although to a lesser extent). This is a very interesting result because it indicates that the methylated pyrrole fragment ($C_3N_2H_3$–CH_3) at *m*/*z* 82 is not characteristic for caffeine in the wide range of beverages in the test set. Indeed, on inspection of the 15 GP rules that used *m*/*z* 82 as being characteristic, all 15 correctly assessed the caffeine status of the freeze dried coffees but failed to classify the filter coffee (only 4 correct), tea (all incorrect) and cola (13 correct). By contrast the 5 rules that used *m*/*z* 83 correctly assessed the filter coffee's caffeine status but correctly classified only 1 each of the freeze-dried, tea and cola drinks.

M/*z* 55 was not selected by any of the GPs as being characteristic of caffeine and on close inspection of the spectra from the pyrolysis of two coffees (Fig. 1) this ion is very intense in both the caffeine containing and decaffeinated coffee. This suggests that this low molecular weight ion is multi-component arising from the pyrolysis of many compounds, and so will not be very characteristic for caffeine. The other dominant peaks in the PyMS of pure caffeine comprise *m*/*z* 136 and 137 and these are also not used by the GP, and so may also have arisen from the pyrolysis of other material found in coffee. Indeed, when chlorogenic acids (obtained from Sigma) were pyrolysed we observed *m*/*z* 136 as the most abundant fragment (Fig. 5); it is likely therefore since freeze-dried coffee contains between 5 and 7.4% (dry weight) and roasted coffee between 1 and 5% chlorogenic acids,[32] that this would 'mask' the importance of *m*/*z* 136 and 137.

From the GP results shown in Fig. 4A it may be concluded that *m*/*z* 67, 109 and 165 are the most characteristic pyrolysis fragments for caffeine. However, these ions were also used in the GP rules that failed to classify all the beverages in the test set (Fig. 4B), and the question arises as to why this should be so. If the ion intensities for *m*/*z* 67 are normalised to the total ion count and plotted (Fig. 6) it can be observed that there is no overlap in the range between the caffeine containing (minimum for all drinks is 1.66) and decaffeinated drinks (maximum for all drinks is 1.37). Although the minimum ion intensity for *m*/*z* 67 in caffeine-containing drinks in the training set was also 1.66, the maximum ion intensity for the decaffeinated drinks in the training set was 1.20 (Table 3). Therefore GP rules that stated 'if $M67 \geq n$ then 1 else 0' (where *n* was between 1.21 and 1.36) would identify all the training set correctly but would fail to classify correctly all of the test set. Similar results were seen for *m*/*z* 165 but not for 109 (Table 3), although for *m*/*z* 109 the range of the test set values was outside the training set range for the maximum ion intensities in caffeine containing drinks and the minimum for the decaffeinated beverages (this was not observed for *m*/*z* 67 and 165).

Finally, the lack of overlap between the minimum ion intensities for masses in caffeine containing drinks and the maximum ion intensities for the decaffeinated drinks was

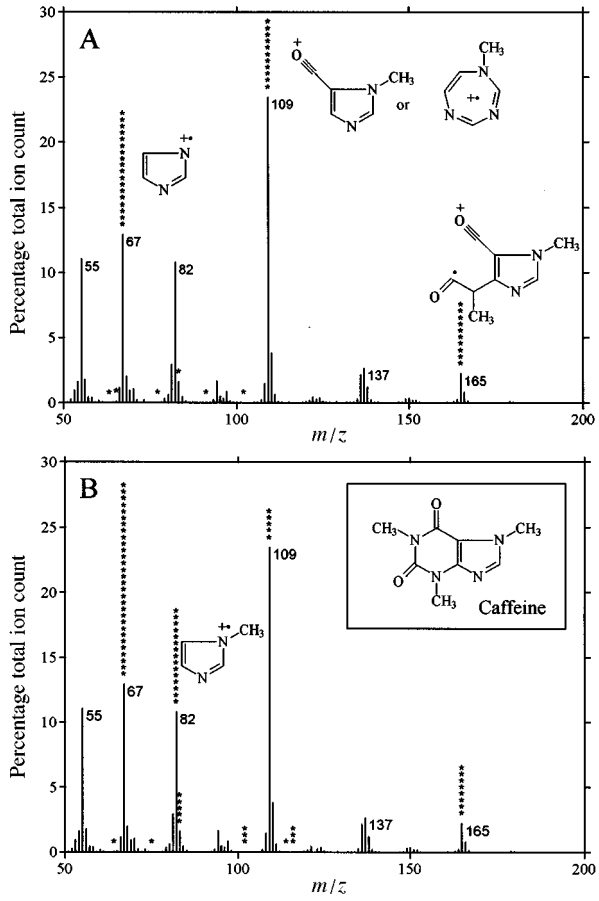

Fig. 4 Pyrolysis mass spectra of caffeine with genetic programming results overlaid from GPs evolved with 20 masses ranked according to their PMC values: (A) those GP results which correctly classified all beverages, (b) those results from GPs which failed to classify everything in the test set. Also shown are the chemical formulae of the significant pyrolysates chosen by the GPs. Asterisks indicate the frequency of each mass peak in the GP rules.

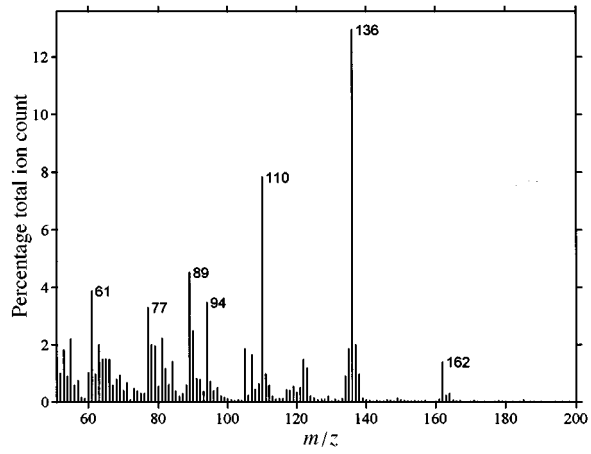

Fig. 5 Pyrolysis mass spectra of chlorogenic acid.

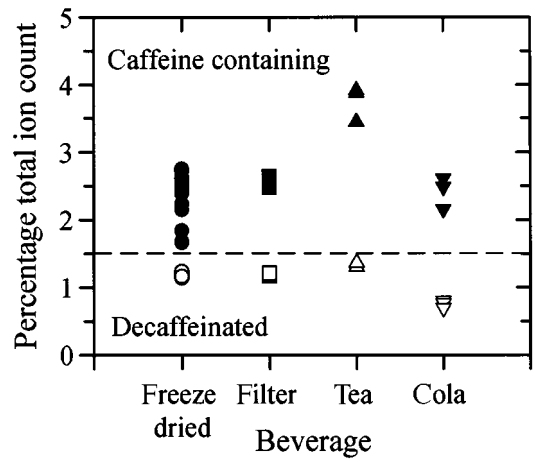

Fig. 6 The percentage of the total ion count for *m*/*z* 67 for freeze-dried and filter coffee, tea and cola.

Table 3 Details of the differences between the total percentage ion counts for the pyrolysis fragmentation of caffeine

	Ion intensities expressed as a percentage of the total ion count		
Mass ion	*Maximum* found in decaffeinated beverages	*Minimum* found in caffeine containing beverages	Difference[a]
55	3.85	2.81	−1.05
67[b]	1.37 (1.20)	1.66 (1.66)	0.29 (0.46)
82	2.24	1.6	−0.64
83	1.47	0.71	−0.75
109[b]	1.85 (1.85)	2.53 (2.53)	0.68 (0.68)
136	1.83	0.84	−0.99
137	0.94	0.47	−0.47
165[b]	0.19 (0.18)	0.23 (0.23)	0.04 (0.05)

[a] A negative value shows there is an overlap between the ion intensities for a particular mass between caffeine containing and decaffeinated beverages, whilst a positive value shows two discrete populations. [b] Values in brackets for *m*/*z* 67, 109 and 165 are from the training set only.

observed only for *m*/*z* 67, 109 and 165 (Table 3), which explains why these masses are very discriminatory and hence selected by the GPs.

In conclusion, this study demonstrates that the GP approach can be used to deconvolve pyrolysis mass spectra in terms of which masses are important for the separation between caffeine containing and decaffeinated drinks. Moreover, since the chemical structures of these ions are known (Fig. 4), this illustrates that GP is an essential chemometric tool for the *chemical* deconvolution of these spectra, which in this study identifies the reproducible pyrolytic degradation products from caffeine.

Acknowledgements

R. G. is indebted to the Wellcome Trust for financial support (grant number 042615/Z/94/Z). R. J. G. thanks the UK EPSRC for financial support.

References

1 R. Goodacre and D. B. Kell, *Curr. Opin. Biotechnol.*, 1996, **7**, 20.
2 R. Goodacre, D. B. Kell and G. Bianchi, *Nature (London)*, 1992, **359**, 594.
3 R. Goodacre, D. B. Kell and G. Bianchi, *J. Sci. Food Agric.*, 1993, **63**, 297.
4 R. Goodacre, D. Hammond and D. B. Kell, *J. Anal. Appl. Pyrol.*, 1997, **40**/**41**, 135.
5 R. Goodacre, *Appl. Spectrosc.*, 1997, **51**, 1144.
6 G. J. Salter, M. Lazzari, L. Giansante, R. Goodacre, A. Jones, G. Surricchio, D. B. Kell and G. Bianchi, *J. Anal. Appl. Pyrol.*, 1997, **40**/**41**, 159.
7 E. Anklam, M. R. Bassani, T. Eiberger, S. Kriebel, M. Lipp and R. Matissek, *Fresenius' J. Anal. Chem.*, 1997, **357**, 981.
8 E. Anklam, M. Lipp, B. Radovic, E. Chiavaro and G. Palla, *Food Chem.*, 1998, **61**, 243.
9 J. H. Holland, *Adaption in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*, MIT Press, Cambridge, MA, 1992.
10 D. E. Goldberg, *Genetic algorithms in search, optimization and machine learning*, Addison-Wesley, Reading, MA, 1989.
11 M. Mitchell, *An Introduction to Genetic Algorithms*, MIT Press, Boston, MA, 1995.
12 T. Bäck, D. B. Fogel and Z. Michalewicz, *Handbook of Evolutionary Computation*, IOP Publishing/Oxford University Press, Oxford, 1997.
13 J. R. Koza, *Genetic Programming: On the Programming of Computers by Means of Natural Selection*, MIT Press, Cambridge, MA, 1992.
14 J. R. Koza, *Genetic Programming II: Automatic Discovery of Reusable Programs*, MIT Press, Cambridge, MA, 1994.
15 R. J. Gilbert, R. Goodacre, A. M. Woodward and D. B. Kell, *Anal. Chem.*, 1997, **69**, 4381.
16 J. Taylor, R. Goodacre, W. G. Wade, J. J. Rowland and D. B. Kell, *FEMS Microbiol. Lett.*, 1998, **160**, 237.
17 R. Goodacre, B. Shann, R. J. Gilbert, M. Timmins, A. C. McGovern, B. K. Alsberg, N. A. Logan and D. B. Kell. in Proc. *1997 ERDEC Scientific Conference on Chemical and Biological Defense Research, Aberdeen Proving Ground*, 1998, *ERDEC-SP-063*, pp. 257–265.
18 R. J. Gilbert, R. Goodacre, B. Shann, J. Taylor, J. J. Rowland and D. B. Kell. in *Genetic Programming 1998: Proceedings of the 3rd Annual Conference, Madison, WI, 1998*, ed. J. R. Koza, Morgan Kaufmann, San Francisco, CA, 1998, pp. 109–115.
19 R. Goodacre, M. J. Neal, D. B. Kell, L. W. Greenham, W. C. Noble and R. G. Harvey, *J. Appl. Bacteriol.*, 1994, **76**, 124.
20 R. Goodacre, A. N. Edmonds and D. B. Kell, *J. Anal. Appl. Pyrol.*, 1993, **26**, 93.
21 R. Goodacre, S. Trew, C. Wrigley-Jones, G. Saunders, M.J. Neal, N. Porter and D. B. Kell, *Anal. Chim. Acta*, 1995, **313**, 25.
22 D. R. Causton, *A Biologist's Advanced Mathematics*, Allen and Unwin, London, 1987.
23 I. T. Jolliffe, *Principal Component Analysis*, Springer-Verlag, New York, 1986.
24 B. F. J. Manly, *Multivariate Statistical Methods : A Primer*, Chapman & Hall, London, 1994.
25 D. E. Rumelhart, J. L. McClelland and The PDP Research Group, *Parallel Distributed Processing, Experiments in the Microstructure of Cognition*, MIT Press, Cambridge, MA, 1986.
26 R. Goodacre, M. J. Neal and D. B. Kell, *Anal. Chem.*, 1994, **66**, 1070.
27 P. J. Werbos, *The roots of back-propagation: from ordered derivatives to neural networks and political forecasting*, John Wiley, Chichester, 1994.
28 C. M. Bishop, *Neural networks for pattern recognition*, Clarendon Press, Oxford, 1995.
29 Y. Chauvin and D. E. Rumelhart, *Backpropagation: Theory, Architectures, and Applications*, Erlbaum, Hove, 1995.
30 B. D. Ripley, *Pattern recognition and neural networks*, Cambridge University Press, Cambridge, 1996.
31 A. Singleton, *Byte*, 1994, **19**, 171.
32 A. W. Smith, in *Coffee. Volume 1: Chemistry*, ed. R. J. Clarke and R. Macrae, Elsevier, London, 1985, pp. 1–41.
33 D. B. Kell and B. Sonnleitner, *Trends Biotechnol.*, 1995, **13**, 481.
34 R. Goodacre, M. J. Neal and D. B. Kell, *Zbl. Bakt. - Int. J. Med. M.*, 1996, **284**, 516.
35 H. Budzikiewicz, C. Djerassi and D. H. Williams, *Structure Elucidation of Natural Products by Mass Spectrometry. Volume 1: Alkaloids*, Holden-Day, Inc., San Francisco, CA, 1964.

Paper 9/01062I