Genetic Programming as an Analytical Tool for Metabolome Data

Richard J. Gilbert, Helen E. Johnson, Michæl K. Winson, Jem J. Rowland[†],
Royston Goodacre, Aileen R. Smith, Michæl A. Hall and Douglas B. Kell

Institute of Biological Sciences, University of Wales, Aberystwyth, Ceredigion SY23 3DD, UK
[†]Department of Computer Science, University of Wales, Aberystwyth, Ceredigion SY23 3DB, UK

rcg@aber.ac.uk, hej93@aber.ac.uk, mkw@aber.ac.uk, jjr@aber.ac.uk,
ars@aber.ac.uk, mah@aber.ac.uk dbk@aber.ac.uk

Corresponding Author: Richard Gilbert. Tel: +44 (0)1970 622353. Fax: +44 (0)1970 622354. http://gepasi.dbs.aber.ac.uk/rcg

Abstract

Genetic programming, in conjunction with advanced analytical instruments, is a novel tool for the investigation of complex biological systems at the whole-tissue level.

In this study, samples from tomato fruit grown hydroponically under both high- and low-salt conditions were analysed using Fourier-transform infrared spectroscopy (FTIR), with the aim of identifying spectral and biochemical features linked to salinity in the growth environment.

FTIR spectra are not amenable to direct visual analysis, so supervised machine learning was used to generate models capable of classifying the samples based on their spectral characteristics. The genetic programming (GP) method was chosen, since it has previously been shown to perform with the same accuracy as conventional data modelling methods, but in a readily-interpretable form.

Examination of the GP-derived models showed that there was a small number of spectral regions that were consistently being used. In particular, the spectral region containing absorbances potentially due to a cyanide/nitrile functional group was identified as discriminatory. The explanatory power of the GP models enabled a chemical interpretation of the biochemical differences to be proposed. The combination of FTIR and GP is therefore a powerful and novel analytical tool which, in this study, improves our understanding of the biochemistry of salt tolerance in tomato plants.

Introduction

The metabolome is a generic term for the total biochemical composition of a cell or tissue sample at any given time. Recent advances in DNA sequencing have lead to an explosion in the number of known gene sequences, but the majority of these new genes have never been characterised experimentally, and most have completely unknown functions within the cell. By investigating the changes in the metabolome of biological systems under different conditions, it is hoped that previously undescribed metabolic processes or pathways may be uncovered, leading to functional assignments for many of the newly-discovered genes within the genomic databases. This area of biology, termed functional genomics, will be a major focus of study over the next decade.

In order to study the metabolome of biological samples, new analytical techniques need to be developed. A typical metabolome study seeks to detect changes in the levels of a few specific biochemicals against a background of more than a thousand other cellular components. To address this, analytical instumentation is being developed which is capable of measuring biochemical signatures from whole-tissue or whole-organism samples. This typically results in datasets comprising measurements of many hundreds or thousands of variables. To complicate this task further, the identities of the particular biochemicals to be monitored are frequently unknown at the outset. The power of GP to select variables from high dimensional data and to form interpretable predictive models gives it a unique advantage in the analytical interpretation of metabolomic data.

Over the past two decades, the tomato as a crop has increased in popularity. Consequently, much research has been aimed at improving the economic viability of tomato production and post-harvest stability. Environmental stress, such as high salt concentration, is one of the main parameters limiting crop production. The tomato cultivar Edkawy is potentially salt-tolerant as it grows in the El-Bosaily area of North Egypt, where the soils are saline sands. Edkawy has already been studied in terms of salt tolerance and previous literature provides evidence that this tomato variety may have salt tolerant attributes[1,2]. In this study, Edkawy plants were cultivated using a hydroponic drip irrigation system, allowing precise control of the nutrient conditions within the root zone, including the salinity level. The aim of the study was to identify biochemical constituents (*biomarkers*) within the fruit tissue which are discriminatory for salt-grown tomato plants, and hence to contribute to the understanding of the fundamental biological mechanisms potentially underlying salt tolerance in tomato plants. This in turn may lead to rational improvements in the quality of tomato fruit grown in conditions of high salinity.

Fourier-transform infrared spectroscopy [3] is a physico-chemical analytical technique, which uses the vibrational characteristics of chemical bonds within molecules to obtain a 'fingerprint' spectrum with features defined by the functional chemical groups within the sample. This form of analytical technique is therefore able to give quantitative information about the total biochemical composition of a sample. A thin layer of the biological sample to be analysed is illuminated in the infrared to obtain an *interferogram* (produced by splitting an infrared beam of light, extending the path length of one half by reflecting it off a movable mirror, and recombining the beams optically). Chemical groups within the sample absorb specific frequencies of light within the interferogram due to 'resonance' with their vibrational motions, the precise frequencies absorbed being related to the energies specific to the vibrational modes of each chemical group. The information encoded in the reflected/absorbed light is then recovered by performing a Fourier-transform on the detected signal. The FTIR spectrum so obtained comprises 882 variables, each of which indicates the level of absorbance at a particular frequency of infrared light.

A readily accessible interpretation of such extremely high-dimensional spectra, also known as *hyperspectral data*, is often very difficult to obtain. Conventional analysis of data of this form falls into two types. The first type, *unsupervised* learning methods, includes principal components analysis (PCA), discriminant function analysis (DFA) and hierarchical cluster analysis (HCA), and seeks to form separable clusters in the data by performing mathematical transforms derived from the variables within the dataset without reference to known classes. The second type, *supervised* learning methods, includes partial least squares (PLS), multivariate rule induction (MRI), inductive logic programming (ILP) and artificial neural networks (ANNs), and seeks to refine a model based on the accuracy of its predictions for a set of examples with a known class structure. Although widely used, none of these methods provide models which are readily interpretable in a chemical sense.

Genetic programming [4-6] is an evolutionary technique which uses the concepts of Darwinian selection to generate and optimise a desired computational function or mathematical expression. GP is a

supervised learning method, and consequently requires a set of training examples to form predictive models that can then be applied to the classification of a set of previously unseen test samples. It has previously been shown that GP performs at least as well as conventional predictive modelling methods for analysing hyperspectral data [7].

Recently, hybrid GA-GP systems have been described [8,9] which are able to produce accurate predictive models whilst minimising their complexity by enforcing constraints on their functional form and expression length. However, a full GP system was chosen for use in this study because the precise mathematical form and complexity of predictive models able to classify tomato fruit tissue samples based on their FTIR spectra were unknown at the outset.

Methods

Plant Cultivation

The plants were grown in a hydroponic open-drip irrigation system, using perlite as an inert substrate. The use of a hydroponic system is ideal for studies into plant physiology as it allows complete control over the nutrients applied to the plants. The system was arranged to facilitate saline and control treatments. The capacity of this system was 120 plants with 60 replicates per treatment. All plants were irrigated with complete liquid fertiliser and supplementary sodium chloride (4000 ppm) was applied to the saline treated plants.

Fruit Tissue Preparation

Twenty fully ripe (at stage 10 on the OCDE tomato ripening chart) Edkawy fruits were harvested. Fruit were selected for uniformity to maximise homogeneity between samples. Ten fruit were taken from salt-grown plants, and ten from control plants. The seeds and skin were removed, the outer pericarp was crushed using a press, and kept on ice. The extract was homogenised using a Polytron blender at speed 5 for 1 minute. After homogenisation, 1ml aliquots of the sample were placed in Eppendorf tubes, and snap-frozen in liquid N_2. These were stored at -70°C until needed.

FTIR Spectroscopy

Ten replicate 5µl samples of each of the 20 fruit tissue samples were applied to wells drilled on a sandblasted aluminium plate, arranged to minimise the effects of artifactual trends in the data. Prior to analysis, the samples were oven-dried at 50 °C for 30 min. The plate was loaded onto the motorised stage of a reflectance thin-layer chromatography (TLC) accessory attached to a Bruker IFS28 FTIR spectrometer (Bruker Ltd.) equipped with a mercury-cadmium-telluride (MCT) detector cooled using liquid N_2.

The FTIR spectra were collected over a wavenumber range from 4000 cm^{-1} to 600 cm^{-1} under the control of an IBM-compatible personal computer using OPUS 2.1 software running under the IBM OS/2 Warp operating system. Spectra were acquired at a rate of 20 s^{-1}, and at a resolution of approximately 3.85 cm^{-1}. To improve the signal-to-noise ratio, 256 spectra were recorded and averaged for each sample. The complete dataset therefore comprised 200 averaged spectra, each containing 882 input variables. 100 spectra (50 from saline-grown and 50 from control fruit samples) were used by the GP as a training set to derive the models, and the remaining 100 spectra were used to test their predictive ability.

Genetic Programming

The GP implementation used in this study was capable of performing non-linear multivariate regressions with automatic variable selection. It was written in C, and was run on IBM-compatible PCs under Windows NT 4.0, and on DEC Alpha-based PCs under Linux 5.1.

The GP used the arithmetic operator functions *add*, *subtract*, *multiply*, and *protected divide* and a Boolean '*if greater than or equal to'* function. The *if* function returned a value of 1.0 if the first argument was greater than or equal to the second argument; 0.0 otherwise. To avoid possible numeric overflows, a protected *divide* function was used which returned a numerical value of 10^{15} for divisions with a denominator $\leq 10^{-15}$. Additional protection from floating-point errors was enforced by clipping the return value of each node into the range $\pm 10^{15}$.

Terminals comprised either floating-point constants (initialised randomly in the range –10.0 to 10.0) or input variables (corresponding to one of the 882 absorbance measurements which comprised each spectrum).

The GP generated initial individuals with random function trees of depth 2 to 6, and assessed their fitness using a scoring function that compared e_i (the model's estimate of the output for example i) with o_i (the experimentally-observed value) by calculating the root-mean-square error of prediction (RMSEP) for n training examples:

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{n} (o_i - e_i)^2}{n}}$$

The fittest individuals were those which gave the lowest RMSEPs for the training set examples.

Since the dataset contains two classes (fruit from plants grown under either saline or control conditions), class membership was defined in the training examples by assigning a target output value of 1.0 to members of the saline-grown class, and 0.0 to members of the control class. A correct classification assigned when the output value of the GP-derived rule was within 0.01 of the target output for any given spectrum.

GP-generated rules, if allowed to evolve unchecked, tend to become longer and more arithmetically complex as the evolution proceeds, a phenomenon known as *bloat* [10]. This increase in complexity reduces the ready interpretability of the expressions generated. To combat this, a penalty of $0.01 \times N$, where N represents the number of nodes in the function tree, was added to the fitness calculation. This ensured that, for a given RMSEP, a shorter tree would be chosen over a longer one. In addition, a maximum tree depth of 10, and a maximum node count of 100 was enforced during the evolution.

The size constraints on the GP rules meant that even the longest rules could use only a small subset of the available input variables comprising the dataset. The GP was therefore compelled to perform an automatic variable selection, resulting in predictive models with significantly lower dimensionality (*i.e.* using far fewer variables) than the dataset as a whole. The automatic variable-selection ability of the GP approach is one of the main benefits of using this as a predictive modelling method [11]. Since the GP-derived models are readily-interpretable, analysis of the selected variables can lead to a rationalisation of the mechanism underlying the model.

The GP used five demes (sub-populations) each of 7500 individuals. Every 10 generations, the best 5% of the individuals in each of four satellite demes replaced the worst 5% in a central deme. The best 5% from this deme then replaced the worst 5% in the satellite demes. This divergent evolution and migration strategy has been shown to be more effective at solving high-dimensional problems than a conventional single-population GP [12].

During each generation, 1500 new individuals were created by single-point mutation, and 3000 by single-point crossover. Parental selection was proportional to fitness, and new individuals were retained in the deme if their fitness was higher than that of the current worst individual.

Partial Least Squares Modelling

Partial least squares (PLS) modelling is a widely-used supervised learning technique which reduces the dimensionality of multivariate data by using *a priori* knowledge of which spectra were derived from plants grown under saline or control conditions to produce mathematical models comprising linear combinations of variables. For this study, we used a PLS modelling system written in house by Dr. Alun Jones.

Variable Analysis

An analysis was performed to investigate the correlation between the GP-selected input variables and the known class structure of the data. *Product moment correlation* (PMC) is a method which uses linear transformations to quantify which variables (x) are most strongly related to the output data (y) being modelled.

The PMC (R) ranges from -1 to +1, indicating a perfect negative to a perfect positive correlation R takes the sign of C_{xy}. A value of 0 indicates that x is uncorrelated with y. R for n examples be calculated as follows:

$$R = \frac{C_{xy}}{\sqrt{C_{xx} \cdot C_{yy}}}$$

where

$$C_{xy} = \left(\sum_{i=1}^{n} x_i \cdot y_i \right) - n \cdot \left(\bar{x} \cdot \bar{y} \right)$$

$$C_{xx} = \left(\sum_{i=1}^{n} x_i^{\,2} \right) - n \cdot \left(\bar{x} \right)^2$$

$$C_{yy} = \left(\sum_{i=1}^{n} y_i^{\,2} \right) - n \cdot \left(\bar{y} \right)^2$$

Quantum Mechanics and Infrared Spectral Analysis

The semi-empirical quantum mechanics program PM1, part of the HyperChem 5.1 molecular modelling package (HyperCube, Inc.) was used to calculate infrared vibrational spectra and molecular vibrational modes for potential metabolites identified during the analysis of the data. The infrared spectral analysis

package IR Mentor Pro 2.0 (Bio-Rad Laboratories) was used to suggest candidate chemical groups responsible for the particular spectral features selected as discriminatory by the GP models.

Results and Discussion

After a sufficient number of reproductive generations, the GP was able to derive expressions capable of correctly classifying the examples in both the training and test sets to an average accuracy approaching 90%. The final GP model was produced after fewer than 400 generations. The PLS model was also to separate the classes to a similar accuracy (Figure 1). However, the PLS model, in common with the other widely-used statistical modelling methods, does not provide readily-available information about which variables have been selected.

The rules from 30 independent GP runs used 112 of the 882 input variables. The dataset as a whole contained variables with PMC values ranging from 0.000332 to 0.4401. The GP-selected variables had PMC values ranging from 0.000642 to 0.4296, indicating that the GP selected variables with both high and low correlations with the known class structure. The two most widely-used variables (each found in five models) had a PMC value of 0.3055 and 0.2876, both reasonably well-correlated with the class structure. Although the most-correlated variable in the data set was not used, the GP selected an adjacent variable on two occasions.

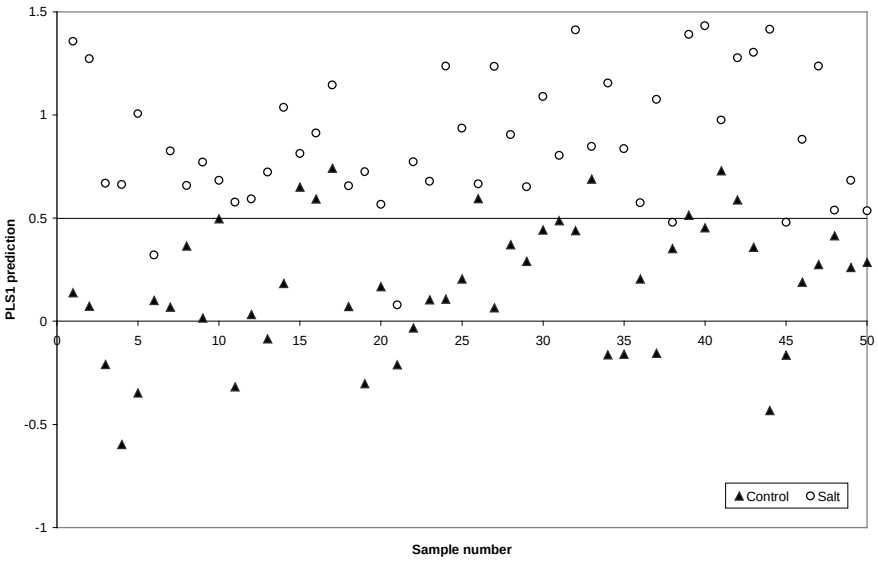


Figure 1

Partial Least Squares 1 (PLS1) was used to classify the samples. The PLS1 model was trained on 50 spectra. The chosen model used 11 components, which gave the minimum RMS error in the prediction values (0.3302) for a validation set of 50 previously unseen spectra. This model was then applied to 100 previously unseen spectra, comprising 50 control and 50 saline-grown samples. The plot below shows the model's predicted output values for this test set.

The GP-derived models used, on average, five variables in their predictive rules. No single variable could be used to classify the spectra with an accuracy approaching the 90% value of the GP-derived rules. Despite the reasonably high PMC values for most of the variables within the dataset PLS was unable to completely separate the two classes based on the spectral data (Figure 1). This indicates that the dataset does not contain enough information to allow a high degree of separation of the two classes to be made without using non-linear combinations of variables. The *if* operator was used in every GP rule, a clearly essential function for a classification problem which is not readily available to neural networks and the conventional statistical modelling methods.

The GP models were all different. The prediction accuracy ranged from 84.5% to 94% correct. Runs 9, 13, 18 and 27 produced remarkably similar models, with the same logical structure and using very similar variables:

Run 9: **IF (A_{2164}-A_{2245}) ≥ (A_{2060}-A_{2098}) THEN [Saline] ELSE [Control] (89% correct)**

Run 13: **IF (A_{2171}-A_{2245}) ≥ (A_{1963}-A_{2106}) THEN [Saline] ELSE [Control] (88% correct)**

Run 18: **IF (A_{2168}-A_{2257}) ≥ (A_{2029}-A_{2114}) THEN [Saline] ELSE [Control] (89.5% correct)**

Run 27: **IF (A_{2179}-A_{2230}) ≥ (A_{2025}-A_{2118}) THEN [Saline] ELSE [Control] (90.5% correct)**

In the above rules, A_n represents the measured absorbance at n wavenumbers. These rules may be indicative of the nature of the globally-optimal rule derivable from this dataset. The best performing rule, with a 94% predictive accuracy, used a similar logical construct and selected similar variables. However it included an additional term associated with a spectral feature at 2480 wavenumbers which enabled a better class prediction for some of the saline-grown samples which were incorrectly classified by the simpler rules:

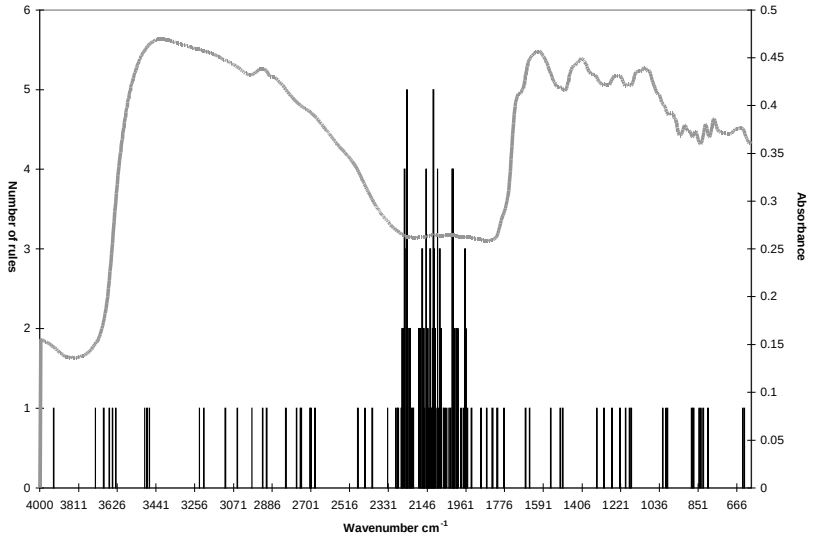
Run 28:

$$\mathbf{IF}\left[\left(\frac{A_{3476}}{A_{3499}}\right)\times \mathbf{IF}\left(A_{2480} \geq A_{883}\right)\right] \geq \left[\left(A_{2268}-A_{2133}\right)\times\left(A_{1558}+A_{3638}\right)+A_{2017}-A_{2110}\right]$$

THEN [Saline] ELSE [Control] (94% correct)

An analysis of which input variables (in terms of absorbances at particular wavenumbers) were selected showed that there were a few regions of the spectra that were consistently being used to form the different models (Figure 2a). In particular, the spectral region covering 2270 to 1960 cm^{-1} was used by most of the rules, and all of the best-performing models were based on a few small but distinct features within this critical region. The absolute differences between saline and non-saline grown samples in this region are relatively small, and so would not have been selected in a direct visual analysis, for example by using a difference spectrum.

(A)



(B)

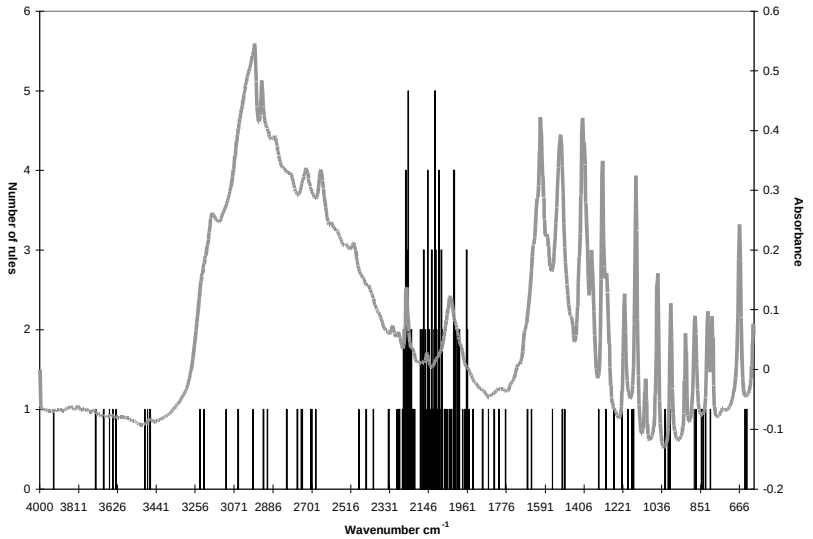


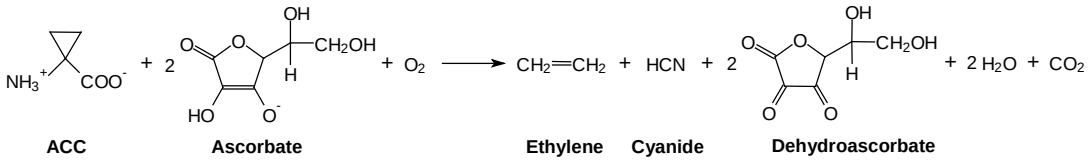
Figure 2

(A) The wavenumbers selected by the 30 GP rules are shown in reference to a spectrum averaged from the whole data set. The vertical lines represent the number of GP-derived rules that use particular wavenumbers to form a predictive model. The region from 2270 to 1960 wavenumbers is clearly important for producing good predictive models.

(B) The wavenumbers selected by the 30 GP rules are shown in reference to the FTIR spectrum of 0.1M β-cyanoalanine. The highly characteristic region identified by the GPs corresponds to the two distinctive peaks for the cyanide group in β-cyanoalanine.

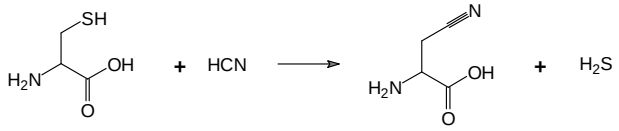
Quantum mechanics calculations showed that the only biochemically-reasonable functional groups that absorb strongly in this critical part of the IR spectrum are acetylenes ($R-C\equiv C-R'$) and cyanides or nitriles ($R-C\equiv N$), with the absorption due to a periodic stretching motion of the triple-bond. Acetylenes have a second characteristic vibration at approximately 3300 wavenumbers, a region unused by any of the GP-derived rules. If an acetylene group were responsible for the characteristic spectral features, this region would also be expected to be used by the rules. Therefore, the most likely candidate chemical moiety being identified by the predictive models as characteristic for tomatoes grown under saline conditions is a cyanide or nitrile group.

Cyanide, in the form of HCN, is formed in plants as a by-product during the conversion of the precursor 1-aminocyclopropane-1-carboxylic acid (ACC) to ethylene by the enzyme *ACC oxidase* [13]:



ACC **Ascorbate** **Ethylene** **Cyanide** **Dehydroascorbate**

Increased ethylene biosynthesis is known to occur in plants in response to stress and during climacteric fruit ripening [14]. It has previously been reported that tomato plants grown under saline conditions show enhanced ethylene production [15]. Hydrogen cyanide is an extremely toxic molecule, and plants have evolved a protective mechanism for its conversion to a less-harmful form. The cyanide produced during ethylene biosynthesis is rapidly converted to the less-toxic compound β-cyanoalanine by the enzyme *cyanoalanine synthase* (CAS):



Cysteine **Hydrogen cyanide** **Cyanoalanine** **Hydrogen sulphide**

The samples presented to the FTIR instrument were oven-dried in order to reduce the adverse effects of water in the spectra collected. This would also have driven off most of the other volatile compounds within the sample, such as free HCN. It is therefore a possibility that the compound being detected by the FTIR-GP analysis is β-cyanoalanine, or a related metabolite, which would be expected to remain in the dried samples. Quantum mechanical predictions of the spectral regions to which the cyanide group of β-cyanoalanine would contribute show a very high degree of correspondence with the critical regions selected by the GP models (Figure 2b). This is by no means conclusive evidence that β-cyanoalanine is the actual metabolite being detected as a discriminatory biomarker for salt-grown tomatoes, but it is consistent with this hypothesis.

From the observations and computational analyses, it seems reasonable to propose that tomato plants grown in conditions of high salinity produce enhanced levels of cyanide as a result of an increase in the production of the stress hormone ethylene. This is being detected by the analytical method described here, potentially in the form of β-cyanoalanine. It is possible that many of the toxic effects observed under saline-induced stress conditions are caused not by the salt *per se*, but by the concomitant increase in cyanide or a cyanide-containing compound as a result of increased ethylene biosynthesis. This proposed biochemical explanation is now subject to experimental verification using conventional biochemical techniques.

This study has shown that FTIR, in combination with GP, is a powerful new tool for the analysis of whole-tissue biological samples at the metabolome level. The technique is sensitive enough to detect changes in the levels of a single metabolite against the background of the entire cellular components, and can provide chemical information which can lead to the identification of the biochemicals which may be involved in metabolic processes under investigation. This method has the promise of becoming an extremely sensitive and discriminatory analytical tool which may be of crucial importance in the emerging field of functional genomics, and so help to advance the understanding of metabolic processes as yet unexplored by biological science.

Acknowledgements

RJG, JJR and DBK thank the UK EPSRC for financial support. MKW, JJR and DBK thank the UK BBSRC for financial support. HEJ and ARS and MAH acknowledge the financial support of the European Union INCO-DC programme. We thank Pat Causton and Dave Summers for their help in plant cultivation. We also thank Dr. Gary Salter for his assistance and encouragement.

References

[1] Mahmoud, M.H., El-Beltagy, A S, Helal, R M, Maksoud M A (1986) Tomato variety evaluation and selection for salt tolerance. Acta Horticulture 190, 559 - 565.

[2] Mahmoud, M.H., Jones, R A, El-Beltagy, A S (1986) Comparative responses to high salinity between salt-sensitive and salt-tolerant genotypes of tomato. Acta Horticulture 190, 533 - 543.

[3] Winson, M.K., Goodacre, R., Woodward, A.M., Timmins, É., Jones, A., Alsberg, B.K., Rowland, J.J. and Kell, D.B. (1997) Diffuse reflectance absorbance spectroscopy taking in chemometrics (DRASTIC) A hyperspectral FT-IR based approach to rapid screening for metabolite overproduction. Analytica Chema Acta 348, 273 - 282.

[4] Koza, J.R. (1992) Genetic Programming: On the Programming of computers by Means of Natural Selection., pp. 819 MIT Press, Cambridge, MA.

[5] Koza, J.R. (1994) Genetic Programming II: Automatic Discovery of Reusable Programs., pp. 746 MIT Press, Cambridge, MA.

[6] Koza, J.R. (1995) Survey of Genetic Algorithms and Genetic Programming. In: Wescon® 95 : E2. Neural-Fuzzy Technologies and Its Applications, pp. 589 - 594 IEEE, San Francisco, California, USA.

[7] Gilbert, R.J., Goodacre, R., Woodward, A.M. and Kell, D.B. (1997) Genetic programming: A novel method for the quantitative analysis of pyrolysis mass spectral data. Analytical Chemistry 69, 4381-4389.

[8] Taylor, J., Winson, M.K., Goodacre, R., Gilbert, R.J., Rowland, J.J. and Kell, D.B. (1998) Genetic Programming in the Interpretation of Fourier Transform Infrared Spectra: Quantification of Metabolites of Pharmaceutical Importance. In: Genetic Programming 1998 (Koza, J.R. et al., Eds.) Morgan Kaufmann, Madison, Wisconsin, USA.

[9] Taylor, J., Goodacre, R., Wade, W.G., Rowland, J.J. and Kell, D.B. (1998) The deconvolution of pyrolysis mass spectra using genetic programming: application to the identification of some Eubacterium species. Fems Microbiology Letters 160, 237-246.

[10] Langdon, W.B., Poli, R (1998) Fitness causes bloat: mutation. In: EuroGP '98, Vol. 1391, pp. 37 - 48 (Banzhaf, W., Poli, R, Schoenauer, M, Fogarty, T C, Ed.) Springer, Paris, France.

[11] Gilbert, R.J., Goodacre, R., Shann, B., Rowland, J.J. and Kell, D.B. (1998) Genetic Programming-Based Variable Selection for High-Dimensional Data. In: Genetic Programming 98: Proceedings of the Third Annual Conference, pp. 109 - 115 (Koza, J.R., Banzhaf, W,

Chellapilla, K, Deb, k, Dorigo, M, Fogel, D B, Garzon, M H, Goldberg, D E, Iba, H, Riolo, R L, Ed.) Morgan Kaufmann, Madison, Wisconsin, USA.
[12] Whitlock, M.C., Barton, N H (1997) The effective size of a subdivided population. Genetics 146, 427 - 441.
[13] Peiser, G.D., Wang, T.T., Hoffman, N.E., Yang, S.F., Liu, H.W. and Walsh, C.T. (1984) Formation of Cyanide From Carbon-1 of 1-Aminocyclopropane-1- Carboxylic Acid During Its Conversion to Ethylene. Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences 81, 3059-3063.
[14] Hulme, A.C. (1970) The Biochemistry of Fruits and their Products. In: Food Science and Technology, Vol. 1 (Stewart, G.F., Chichester, C O, Galliver, G B, Morgan, A I, Mrak, E M, Scott, J K, von Sydow, E, Ed.) Academic Press, London.
[15] Mizrahi, Y. (1982) Effect of Salinity on Tomato Fruit Ripening. Plant Physiol. 69, 966 - 970.