Hard Modeling Methods for the Curve Resolution of Data from Liquid Chromatography with a Diode Array Detector and On-Flow Liquid Chromatography with Nuclear Magnetic Resonance Spectroscopy

Mohammad Wasim and Richard G. Brereton*

School of Chemistry, University of Bristol, Cantock's Close, BRISTOL BS8 1TS, United Kingdom

Received April 8, 2005

Hard modeling methods have been performed on data from high-performance liquid chromatography with a diode array detector (LC-DAD) and on-flow liquid chromatography with ^{1}H nuclear magnetic spectroscopy (LC-NMR). Four methods have been used to optimize parameters to model concentration profiles, three of which belong to classical optimization methods (the simplex method of Nelder−Mead, sequential quadratic programming approach, and Levenberg−Marquardt method), and the fourth is the application of genetic algorithms using real-value encoding. Only classical methods worked well for LC-DAD data, while all of the methods produced good results when LC-NMR data were divided into small spectral windows of peak clusters and parameters were optimized over each window.

1. INTRODUCTION

Compounds in chromatography can be separated either physically by optimizing experimental conditions such as the flow rate, pH, and concentration of the mobile phase[1,2] or mathematically by applying deconvolution methods.[3,4] Optimizing experimental conditions is selective and not a successful approach for all compounds in a chromatogram, especially in situations when compounds are isomers. The decomposition of multivariate evolutionary data into pure spectral and concentration profiles is called multivariate curve resolution (MCR). MCR is performed either by hard modeling,[5−11] where information is required about the peak shapes, or by soft modeling approaches,[12−14] where some constraints about the shape and nature of the concentration and spectral profiles are applied. For a successful curve resolution application, the data matrix must be bilinear, but practical situations where noise always exists and where there are other factors such as shifts in chromatographic or spectral peaks may result in a lack of bilinearity. In these cases, curve resolution becomes a challenge. There are many applications where hard and soft modeling techniques have been combined to obtain better results, for example, in kinetics,[15] the resolution of nuclear magnetic resonance (NMR) labile signals,[16] solvent effects on acid−base behavior,[17] and temperature-dependent kinetics.[18]

Liquid chromatography−nuclear magnetic resonance (LC-NMR) data are characterized by lower sensitivity (high noise level) and severely overlapped chromatographic peaks, which affect bilinearity adversely. Most of the mainstream rank determination and curve resolution methods do not produce good results on LC-NMR data.[19−22] However, LC-NMR data are also characterized by selective NMR resonances, which offer special features. ^{1}H NMR spectra of mixtures of compounds are characterized by several resonances; it is usually not difficult to find regions in the spectrum where only one compound resonates. Thus, two-way LC-NMR data have many spectral regions where there is no signal and only noise exists and others where signals from one or more than one compound resonate. The latter spectral regions can be separated into small data windows. These windows contain the same number of rows (elution times) as the original data but a reduced number of columns (frequencies). These data windows have less complexity and correspond to a reduced number of compounds and so are of lower rank than the full data set. Thus, applying curve resolution methods on spectral windows not only reduces the factor space but also improves the quality of the results.

Two curve resolution methods for on-flow LC-NMR data have been reported on the basis of the analysis of data windows.[22,23] The first[22] requires that each window be analyzed for its rank and its relationship with the compounds present in the whole data set. The second[23] is based on first estimating pure concentration profiles for all of the compounds and is called "constrained key variable regression" (CKVR). Finding pure concentration profiles in LC-NMR data is not a difficult procedure but is time-consuming in data exploration. In the present paper, we have tried to overcome the problem of finding pure concentration profiles by using hard modeling. Hard modeling has been used for the deconvolution of single chromatograms and for the curve resolution of multivariate data for a long time.[8] However, there are fewer papers on the resolution of two-way chromatographic data, and most are on kinetics. Moreover, there is no literature on the curve resolution of on-flow LC-NMR data using hard modeling. Hard modeling has been performed mostly by classical methods, which could lead to incorrect results easily by trapping the solution in local minima. Genetic algorithms (GAs), introduced by Holland[24] as alternative methods for optimization and searching, have been applied to a number of applications in chromatography such as the optimization of mobile phase composition,[25] the deconvolution of individual chromatograms,[26,27] multibatch high-performance liquid chromatography (HPLC),[28] and for the construction of response surfaces.[29] In our present

* Corresponding author tel.: +44-117-9287658; fax: +44-117-9251295; e-mail address: r.g.brereton@bris.ac.uk.

10.1021/ci050112f CCC: $33.50
Published on Web 04/05/2006

study, we apply classical optimization methods as well as GAs to resolve data of an eight-component mixture using liquid chromatography−diode array detection (LC-DAD) and on-flow LC-NMR. LC-DAD data have lower noise levels as well as smaller data sizes compared to those of LC-NMR data. A hard modeling method was reported on LC-DAD data in 1985[8] using a Gaussian function convoluted with a single-sided exponential, where only two unknown parameters per compound were optimized. Since then, there has been a lot of progress in the computational methods, as well as new peak shape functions, which have emerged. Our study will cover both aspects, that is, computational methods and different peak shape functions.

2. EXPERIMENTAL SECTION

2.1. Mixtures. *2.1.1. LC-DAD.* The mixture consists of 2,6-dihydroxynaphthalene (**I**) (98%, Avocado, Research Chemicals Ltd., Heysham, U. K.), 2,3-dihydroxynaphthalene (**II**) (98%, Acros Organics, Geel, Belgium, U. K.), diethyl maleate (**III**) (98%, Avocado), methyl *p*-toulenesulfonate (**IV**) (98%, Lancaster, Morecambe, U. K.), diethyl fumarate (**V**) (97%, Avocado), 1,2-diethoxybenzene (**VI**) (98%, Lancaster), 1,4-diethoxybenzene (**VII**) (98%, Lancaster), and 1,3-diethoxybenzene (**VIII**) (95%, Lancaster), each at 30 μM. The solvent was acetonitrile (HPLC grade, Rathburn Chemicals, Walkburn, U. K.) and water (HPLC grade, Rathburn) in a ratio of 80:20 (v/v).

2.1.2. LC-NMR. The eight-component mixture consists of **I**, **II**, **III**, **IV**, V, **VI**, **VII**, and **VIII**, each at 50 mM. The solvent was acetonitrile (HPLC grade, Rathburn) and deuterated water (Goss Scientific Instruments Ltd., Great Baddow, Essex, U. K.) in a ratio of 80:20 (v/v).

2.2. Chromatography. The chromatography was performed on a Waters (Milford, MA) HPLC system, which consisted of a 717 plus autosampler, a 600s controller, and a model 996 diode array detector with a model 616 pump. The mobile phase was acetonitrile (HPLC grade, Rathburn Chemicals, Walkburn, U. K.) and water (HPLC grade, Rathburn) in a ratio of 80:20 (v/v) for LC-DAD. The column was a C_{18} Waters Symmetry (100 × 4.6 mm, 3.5 μm) column. The mobile phase was acetonitrile (HPLC grade, Rathburn) and deuterated water (Goss Scientific Instruments Ltd., Great Baddow, Essex, U. K.) in a ratio of 80:20 (v/v) for LC-NMR. The flow rate was set at 0.2 mL min^{-1} and the injection volume was 50 μL for both experiments. A few drops of tetramethylsilane (Goss Scientific) were added in the LC-NMR experiment as a chemical shift reference.

2.3. Spectroscopy. A diode array detector was used in the LC-DAD experiment, which stores the data in the form of spectra in a machine-readable format.

For LC-NMR, a 4 m PEEK tube with a width of 0.005 in. was used to connect the eluent from the Waters HPLC instrument to a flow cell (300 μL) into an NMR probe on a 500 MHz NMR spectrometer (JEOL Alpha 500, Tokyo, Japan). For each spectrum, the spectral width was 7002.8 Hz and a pulse angle of 90° was employed (acquisition time, 1.1698 s) at a pulse delay of 2 s. Solvent suppression was performed by the DANTE sequence[30] on the acetonitrile singlet peak.

2.4. Pure Spectra. Pure spectra of all of the compounds were obtained using the same experimental conditions as described for the LC-DAD and LC-NMR experiments.

Table 1. Notations Used in the Study

$\mathbf{X}$	data matrix containing signals
$\mathbf{C}$	matrix of pure concentration profiles
$\mathbf{S}$	matrix of pure spectral profiles
$\mathbf{E}$	matrix of error
M	total number of rows
N	total number of columns
K	total number of compounds in data
$h(t)$	peak height at time t
t_r	retention time
H	peak height at time t_r
σ_0	peak shape parameters
σ_1	peak shape parameters
σ_2	peak shape parameters
p	number of unknown parameters
χ^2	least-squares-error function
$\bar{\chi}^2$	least-squares-error function with penalty term
$\hat{\mathbf{X}}$	predicted data matrix
$\hat{\mathbf{S}}$	matrix of predicted spectral profiles
$\mathbf{C_{mod}}$	matrix of modeled concentration profiles
ϵ	population size in genetic algorithms
Λ	a vector of random numbers
f	representation of a chromosome
μ	bounds in GAs
α	a parameter in Muhlenbein's mutation function
γ	a parameter in Muhlenbein's mutation function
$\mathbf{Y}$	a matrix in GA for rank and OPA
d	dissimilarity in GA for rank and OPA
SD	second difference
τ	guess value of retention time
q	a parameter for setting bounds
w	width of a Gaussian profile extrapolated to zero signal
u	standard deviation
$\bar{x}_n$	average intensity of nth column
$\bar{\mathbf{x}}$	average spectrum normalized to unit length
dis	a dissimilarity measure between two spectra
r	correlation coefficient between true and reconstructed spectra

2.5. Software. Data analysis was performed by computer programs written in Matlab by the authors of this paper except for the Fourier transformation and preprocessing of LC-NMR data, which was performed by in-house-written software called LC-NMR.[20]

3. DATA ANALYSIS

3.1. Data Preprocessing. The LC-DAD data was acquired from 200 to 375 nm with a digital resolution of 4.72 nm; because the signal-to-noise ratio was very low beyond 300 nm, the data was truncated after 300 nm. The data were acquired after every 1 s; the data size was 164 elution times × 22 wavelengths.

LC-NMR data were obtained in the time domain as free induction decays (FIDs), which were apodized, Fourier transformed, and phase corrected. A total of 256 FIDs were collected with digital resolutions of frequencies of 0.855 Hz. To remove errors due to quadrature detection, which results in the regular oscillation of intensity, a moving average of each frequency over every four points in time was performed in the chromatographic direction.[31] The data were truncated where there was no significant signal either on the time or on the frequency direction. The size of the whole LC-NMR data set was 141 elution times × 1409 frequencies.

3.2. Modeling of the Concentration Profile. A bilinear data matrix $\mathbf{X}$ is expressed as

$$\mathbf{X} = \mathbf{CS} + \mathbf{E} \tag{1}$$

where $\mathbf{C}$ is a matrix of concentration profiles, $\mathbf{S}$ is a matrix of spectral profiles, and $\mathbf{E}$ is a matrix of residuals (Table 1 provides a list of the notations used in this paper). This equation is derived from the Beer−Lambert law and assumes

that the concentrations of each component are within a region whereby the signal intensity is linearly related to the concentration. This equation can be expressed as a summation as follows. For a *K*-component system, each measurement is assumed as a sum of individual measurements, as shown in eq 2

$$x_{mn} = \sum_{k=1}^{K} c_{mk} s_{kn} + e_{mn} \quad (2)$$

where c_{mk} is the concentration of the compound k, s_{kn} is the intensity of one concentration unit of the compound at NMR frequency n, and e_{mn} is the residual error.

The purpose of multivariate curve resolution methods is to decompose data matrix **X** into **C** and **S**. In hard modeling, **C** is calculated by modeling concentration profiles and then **S** by ordinary least-squares regression (OLS).

$$\mathbf{S} = (\mathbf{C}'\mathbf{C})^{-1}\mathbf{C}'\mathbf{X} \quad (3)$$

There is no exact theoretical description of HPLC peak shapes,[2] but a number of empirical mathematical functions are available in the literature.[32−36] A peak shape known as the "polynomially modified Gaussian model" (PMG)[37] is reported to be a suitable function for almost every experimental peak shape, but it produces a baseline which increases outside the peak region. The PMG function for three peak shape parameters (PMG_2) is defined as

$$h(t) = H \exp\{-0.5(t - t_r)^2/[\sigma_0 + \sigma_1(t - t_r) + \sigma_2(t - t_r)^2]^2\} \quad (4)$$

where $h(t)$ is the peak height at time t; t_r is the retention time; H is the peak height at time t_r; and σ_0, σ_1, and σ_2 are the peak shape parameters. In our study, we also applied the PMG model with two peak shape parameters (PMG_1), which is defined as

$$h(t) = H \exp\{-0.5(t - t_r)^2/[\sigma_0 + \sigma_1(t - t_r)]^2\} \quad (5)$$

A new function, called the parabolic-Lorentzian modified Gaussian model (PLMG),[38] has been recommended to avoid baseline problems. The PLMG model has seven parameters; because the optimization of several parameters becomes difficult, it was not chosen for this study. Another function presented by Baeza-Baeza and Garcia-Alvarez-Coque[2] called the parabolic variance modified Gaussian model (PVMG) has the same number of unknown parameters as in the PMG_2 model. The PVMG model is described as

$$h(t) = H \exp\{-0.5(t - t_r)^2/[\sigma_0^2 + \sigma_1(t - t_r) + \sigma_2(t - t_r)^2]\} \quad (6)$$

The individual concentration profile $\mathbf{c}_k$ of each compound is modeled and assembled in $\mathbf{C}_{\mathbf{mod}}$ presuming the number of compounds in the data is known, equal to *K*.

$$\mathbf{C}_{\mathbf{mod}} = [\mathbf{c}_1\mathbf{c}_2...\mathbf{c}_K] \quad (7)$$

The matrix of predicted spectral profile $\hat{\mathbf{S}}$ is then calculated by OLS, and the predicted data matrix $\hat{\mathbf{X}}$ is calculated as

$$\hat{\mathbf{X}} = \mathbf{C}_{\mathbf{mod}}\hat{\mathbf{S}} \quad (8)$$

The modeled profiles are evaluated by calculating the least-squares error[8] χ^2, which is the difference between the original and the predicted data, weighted by the inverse of the number of degrees of freedom in the fit. The error function is the objective function and is calculated as

$$\chi^2 = \frac{1}{MN - (KN + PK)} \sum_{m=1}^{M}\sum_{n=1}^{N} (\hat{x}_{mn} - x_{mn})^2 \quad (9)$$

where P is the number of unknown parameters for a single profile, M is the number of rows in **X**, N is the number of columns in **X**, and K is the total number of concentration profiles modeled. Because negative intensities in $\hat{\mathbf{S}}$ have no physical meaning, an additional penalty is added to χ^2 to help the search converge quickly. The objective function, $\bar{\chi}^2$, with the penalty term is defined as

$$\bar{\chi}^2 = \chi^2\left(1 + 10\frac{\sum_{n=1}^{N}\sum_{k=1}^{K}|\hat{s}_{nk} < 0|}{\sum_{n=1}^{N}\sum_{k=1}^{K}|\hat{s}_{nk}|}\right) \quad (10)$$

In the optimization process, this objective function is minimized.

3.2.1. Optimization Methods. The optimization methods are divided in two categories.

(i) Classical Methods. The name "classical methods" is used to differentiate from GAs. Further details of each method are given in the next section. These methods can be applied to the whole data set as well as single chromatograms.

(ii) Genetic Algorithms. Genetic algorithms are computationally intense; therefore, it is not practical to apply GAs to optimize peak shape parameters for every chromatogram.

3.2.1.1. Classical Methods. The optimization functions in classical methods have been taken from the Optimization Toolbox of Matlab R12 (MathWorks, Inc.).

3.2.1.1.1. Optimization Using Whole Data Matrix. Two different optimization functions were applied.

fminsearch. This is an unconstrained nonlinear optimization function, also called as a direct method because it does not use numerical or analytical gradients. It finds the minimum of a function by optimizing several variables with an initial guess; lower and upper bounds are not required. It uses the simplex optimization algorithm of Nelder-Mead.[39,40]

fmincon. This finds a constrained minimum of a function of several variables, starting at an initial estimate. This is generally referred to as constrained nonlinear optimization or programming. It uses a sequential quadratic programming method, where a quadratic programming subproblem is solved at each iteration. An estimate of the Hessian of the Lagrangian is updated at each iteration using the Broyden−Fletcher−Goldfarb−Shanno formula.[41−42] A line search is performed using a merit function similar to that proposed in refs 43−45.

3.2.1.1.2. Optimization Methods for Single Profiles. *lsqcurvefit*. In this approach, the optimization was performed on individual chromatographic profiles. Each profile is

modeled using functions described in section 3.2, assuming each chromatogram represents a linear combination of the pure concentration profiles of individual compounds. *lsqcurvefit* is supplied in the Optimization Toolbox of Matlab R12 (MathWorks, Inc.). It performs nonlinear curve fitting using least squares and determines parameters that best fit the modeled profile and applies the Levenberg−Marquardt method.[46−48]

3.2.1.2. Genetic Algorithms. Genetic algorithms are a group of methods, each starting with a population of numbers, called chromosomes. The initial guesses constitute a set of possible solutions, contrary to classical methods, which start with one guess. The population is modified, like evolution modifies genes, with the help of three operations: selection, crossover, and mutation. The modification is performed iteratively; each iteration forms a generation. The modification stops when a preset number of generations are passed. A brief introduction to GAs is provided here; for more details, textbooks[49,50] and tutorial and review articles[51−54] are available.

GAs start with a randomly created population of numbers, called chromosomes. The size of each chromosome depends on the number of unknown variables, called genes. Each variable is coded by either a binary, gray, or real-value method within the upper and lower limits of each variable. The binary and gray encoding consists of concatenated strings of variables containing 0s and 1s, while real-value encoding contains real numbers. The size of the population is a trade off between the computation time and the accuracy. Once a random population is created, each string is evaluated for an objective function and a fitness value is assigned to each chromosome. A termination condition is checked; if the termination criterion is not satisfied, the population is modified and a new population is created. The generation counter is incremented to indicate that one generation of GA is completed. Because real-value coding is applied in our study, the operations to modify a population are described in the context of real-value coding only.

3.2.1.2.1. Selection. The purpose of this operator is to make multiple copies of good chromosomes and eliminate bad chromosomes, while keeping the population size constant. There are many methods to select a chromosome, but we applied stochastic universal sampling (SUS).[55] In SUS, a uniformly distributed random number λ is generated, and then, a set of ϵ (population size) equispaced numbers is created:

$$\Lambda = \{\lambda, \lambda + 1/\epsilon, \lambda + 2/\epsilon, ..., \lambda + (\epsilon - 1)/\epsilon\} \quad (11)$$

The solutions corresponding to each member of Λ are chosen from the cumulative probability. Normally, solutions are scaled before the selection.

3.2.1.2.2. Crossover. This operator is mainly responsible for the creation of new chromosomes from the old ones, thus increasing the search space. In crossover, two chromosomes, called parents, are selected for "mating" and generate two new chromosomes called offspring. In our study, discrete recombination[56] is applied, as described below.

Suppose $F_1 = (f_1^1, ... f_k^1)$ and $F_2 = (f_1^2, ... f_k^2)$ are two chromosomes of length k that have been selected for crossover, where f_1^1 is the value of first variable in chromosome 1 and, similarly, f_1^2 is the value of the first variable in chromosome 2. The offspring $G_1 = (g_1^1, ... g_k^1)$ are generated so that g_i is a randomly (uniformly) chosen value from the set $\{f_i^1, f_i^2\}$. This process involves the rearrangement of values from the same variable in different chromosomes, so the offspring retain the values from the parent chromosomes for each variable, but they are randomly distributed among the two new chromosomes.

3.2.1.2.3. Mutation. This works in a manner similar to crossover by generating new chromosomes from the old ones, but it also keeps diversity in the population. It helps pull the solution out of local minima in solution space. The mutation operator in our study is Muhlenbein's mutation,[56] as described below.

Suppose $F_1 = (f_1^1, ... f_k^1)$ is a chromosome and $f_i \in [\mu_{1i},\mu_{2i}]$ is a gene to be mutated as f_i'. Muhlenbein's mutation produces a mutated gene as

$$f_i' = f_i \pm \text{range}_i \times \gamma \quad (12)$$

where range_i is the mutation range and is normally set to $0.5(\mu_{2i} - \mu_{1i})$. The + or − sign is chosen with a probability of 0.5 and

$$\gamma = \sum_{j=0}^{15} \alpha_j 2^{-j} \quad (13)$$

where $\alpha_j \in [0,1]$ is randomly generated with $p(\alpha_i = 1) = {}^1/_{16}$. Values in the interval $[f_i - \text{range}_i, f_i + \text{range}_i]$ are generated using this operator, with the probability of generating a neighborhood of f_i being very high.

It is known that binary-coded GAs require lot of alteration in the string for moving one solution to their neighborhood. This problem is overcome by gray coding. Even then, binary- and gray-coded GAs provide unstable results as compared to real-value coding.[49] Moreover, binary or gray coding requires the length of each variable to acquire any arbitrary precision in the solution. The string in binary and gray codings becomes very large if many decision variables are optimized. Both the precision and the number of variables increase the chromosome size and, hence, the complexity and computation time.

3.2.2. Objective Function in Optimization. The objective function is presented in eq 10, which is minimized to calculate the unknown parameters. In the optimization of the whole data set, H (maximum peak height) does not play an important role and, hence, is set equal to 1, but when single chromatograms are optimized using *lsqcurvefit*, H is optimized for each chromatogram.

3.2.3. Variable Space. All classical methods require an initial guess to start with, but GAs require only upper and lower bounds. The methods to obtain guesses and lower and upper bounds for the variables presented in eqs 4−6 are described below.

Retention Time. The guess for retention time was obtained by using a recent method, for the determination of data rank and key variables, developed by Wasim and Brereton using genetic algorithms,[57] as described below.

The method starts by finding two vectors from data matrix $\mathbf{X}$ to make a new matrix $\mathbf{Y}$

$$\mathbf{Y} = [\mathbf{x}_{n1},\mathbf{x}_{n2}] \quad (14)$$

where $\mathbf{x}_{n1}$ and $\mathbf{x}_{n2}$ are column vectors selected from data $\mathbf{X}$. A value d_2 is calculated by taking the determinant of the matrix

$$d_2 = \det(\mathbf{Y}'\mathbf{Y}) \tag{15}$$

Equation 15 is optimized in such a way as to maximize d_2 by varying column positions. The vectors selected in the process will produce the first two key variables. The GA search is then repeated for 3 to $K + 3$ components, where K is a value that should ideally exceed the total number of components expected; the value is not especially important so long as it is quite large. After searching for $K + 3$ components, a second difference is calculated by

$$\mathrm{SD}_k = \log(d_{k-1}) - 2\log(d_k) + \log(d_{k+1}) \tag{16}$$

where $k = 3$ to $K + 3$. The data rank is estimated as value for k where SD_k has a minimum value. If the vectors are selected on the time dimension, then the retention time is obtained from the spectra selected in the optimization method.

The initial guess for the retention time in optimization methods is the values obtained by the GA method. The lower and upper bounds are calculated as

$$\tau_k = t_k \pm q(t_0 - t_1) \tag{17}$$

where τ_k provides the lower and upper bounds of retention times, t_0 is the time when the first component starts appearing, t_1 is the retention time of the first component, and q is an adjustable parameter, which sets the range of the retention time. In our study, $q = 0.2$ was applied.

Peak Shape Parameter σ_0. The width w of a Gaussian profile extrapolated to zero signal height is defined as follows:[58]

$$w \simeq 4\sigma_0 \tag{18}$$

where σ_0 is the peak shape parameter, defined in eqs 4−6. The initial guess for σ_0 is calculated as

$$\sigma_0 = 2(t_0 - t_1)/4 \tag{19}$$

The lower and upper bounds are calculated as

$$\sigma_{0l} = \sigma_0 - q\sigma_0$$
$$\sigma_{0u} = \sigma_0 + q\sigma_0 \tag{20}$$

q is the same factor which is used in eq 17.

Peak Shape Parameters σ_1 and σ_2. These parameters are the most difficult ones to set. Their limits are based on simulation studies, which utilized the guessed values of retention times and σ_0. The lower limits for both parameters were set to zero, and the upper limit was 0.2 for LC-DAD and 0.1 for LC-NMR.

fminsearch and fmincon. These two functions from classical methods require initial guesses, but only *fmincon* accepts lower and upper bounds. In these methods, 100 random guesses are generated between the upper and lower limits, and the method is applied 100 times. The best guess was selected, which provided the lowest residual sum of squares (RSS)[59] defined as

$$\mathrm{RSS} = \sum_{m=1}^{M}\sum_{n=1}^{N}(x_{nm} - \hat{x}_{nm})^2 \tag{21}$$

where $\hat{\mathbf{X}}$ is calculated using eq 8. GAs create an initial guess but require lower and upper bounds and modify the solution within these limits.

lsqcurvefit. Because this function is applied to individual chromatograms, it requires that H be optimized in eqs 4−6. The lower bound for H is the minimum signal value, and the upper bound is the maximum signal height in a profile. Moreover, σ_0 has a lower limit of zero, and upper limit σ_{0u} is applied. After the optimization of parameters by *lsqcurvefit* in LC-DAD, the parameters are collected from all of the chromatograms and averaged out. In LC-NMR, the parameters cannot be averaged out because NMR resonances are selective and there are many channels where not all of the compounds show resonance, and at some frequencies, only noise exists. In this case, all of the parameters are calculated from only those chromatograms which show the highest peak height H.

3.3. Curve Resolution. The reconstruction of the spectral profiles in LC-DAD is performed by OLS, but in LC-NMR, two different methods are applied. As mentioned in section 1, good reconstructed spectral profiles for LC-NMR are obtained only by CKVR. OLS and CKVR are described below.

3.3.1. Ordinary Least Squares (OLS). After optimizing the parameters for all of the concentration profiles, eq 3 is applied to obtain reconstructed spectral profiles.

3.3.2. Constrained Least Squares Regression. CKVR was developed by Shen et al.[22] In the original method, concentration profiles using key variables are selected. Because the concentration profiles are not pure, it applies some correction procedures to remove the negative peaks. In this paper, the whole procedure of calculating pure concentration profiles is replaced by hard modeling methods; the remaining steps of the CKVR methods are described below.

(a) Determine all pure $\mathbf{C}$ profiles using hard modeling methods. The parameters for $\mathbf{C}$ profiles can be obtained either from the whole data set or by creating small windows in the data (called NMR peak clusters) and using those windows. The procedure for creating data windows (called peak clusters) is as follows.

(b) Locate regions of NMR peak clusters. The peak cluster regions are determined by calculating the standard deviation at each frequency[31] given below

$$u_n = \sqrt{\frac{\sum_{m=1}^{M}(x_{mn} - \bar{x}_n)^2}{M - 1}} \tag{22}$$

where u_n is the standard deviation, $\bar{x}_n$ is the average intensity of nth column, and M is the total number of rows in $\mathbf{X}$. The process not only selects the peak clusters but also reduces the number of variables. All variables where the signal-to-noise ratio is significant show higher standard deviations compared to the regions where there is only noise. A cutoff value of standard deviation can be used to select peak clusters.

(c) Determine the rank of each peak cluster. Two methods are applied in our study for rank determination. The first method is the orthogonal projection approach (OPA),[60−65] and the second is GA for rank analysis.[57] Because GA is applied only to those data sets whose rank is more than 2, it is used as a confirmatory method. OPA is described below, while GA for rank determination is described in section 3.2.3.

OPA is a stepwise approach and selects one pure or key variable in each step. The method is described here for chromatographic data using the time dimension for determining the pure variables. In the first step of OPA, it constructs a matrix **Y** as

$$\mathbf{Y}_m = [\bar{\mathbf{x}}\ \mathbf{x_m}] \tag{23}$$

where $\bar{\mathbf{x}}$ is the average spectrum normalized to unit length.

The dissimilarity (d_m) is calculated as

$$d_m = \det(\mathbf{Y}_m \mathbf{Y}_m') \qquad m = 1-M \tag{24}$$

The dissimilarity plotted against the time is called a dissimilarity plot, and the maximum is located. The first key variable corresponds to the maximum in this plot. In the second step, the average spectrum is replaced by the first reference spectrum corresponding to the first key variable. The second key variable is identified by locating a maximum point in the new dissimilarity plot, providing a second reference spectrum. In the third and later steps, reference spectra identified in the previous steps are added to **Y** until the dissimilarity plot exhibits only random noise. The total steps required by OPA are equal to one plus the number of compounds present in a data set. The reference spectra are the purest possible spectra with significant signal-to-noise ratios for each compound.

(d) Find the compounds associated with each peak cluster. The retention times obtained by OPA and GA are used to determine the relationship between peak clusters and compounds.

(e) Least squares is performed on each peak cluster by

$$\bar{\mathbf{S}} = (\bar{\mathbf{C}}'\bar{\mathbf{C}})^{-1}\bar{\mathbf{C}}'\bar{\mathbf{X}} \tag{25}$$

where $\bar{\mathbf{X}}$ is a reduced data matrix and contains only the peak cluster region, $\bar{\mathbf{C}}$ is a matrix containing only those concentration profiles which are associated with a peak cluster, and $\bar{\mathbf{S}}$ is a reduced spectral profiles matrix and covers only those frequencies which appear in a peak cluster region. One may estimate spectral coefficients using eq 25, and the coefficients of remaining analytes are set to zero.

3.3.3. Validation of the Reconstructed Spectral Profiles. 3.3.3.1. LC-DAD. The reconstructed spectral profiles are compared with the real spectra of pure compounds using a measure called dissimilarity:[66]

$$\text{dis} = \sqrt{1 - r} \tag{26}$$

where dis is the dissimilarity and r is the correlation coefficient calculated between the true and the reconstructed spectra.

3.3.3.2. LC-NMR. Because correlation coefficients are not good measures for LC-NMR spectra comparison,[22] the reconstructed spectra are visualized and each peak position and multiplicity is compared with the true LC-NMR spectra obtained from the pure compounds.

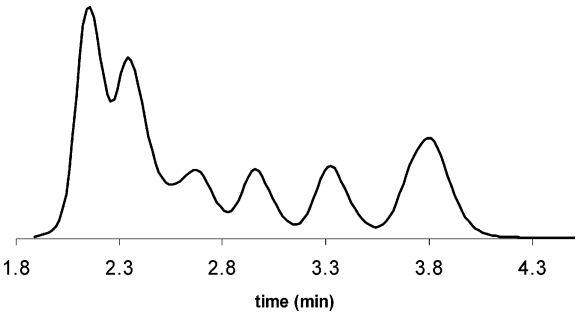

Figure 1. Sum of all chromatographic profiles of LC-DAD data of eight compounds.

Table 2. Parameters Used for GAs

parameters	coding
number of generations	5000
population size	20
selection probability (%)	90
crossover probability (%)	100
mutation probability (%)	1/variables
selection operator	stochastic universal sampling
crossover operator	discrete recombination
mutation operator	Muhlenbein's mutation

4. RESULTS AND DISCUSSION

Initially, three models, PMG_1, PMG_2, and PVMG, were applied, but PVMG was very difficult to use. It requires a very narrow range of bounds, which varies from chromatogram to chromatogram. Because it is not possible to set different bounds for different chromatograms, the PVMG model was abandoned after several unsuccessful tries. The parameters for the genetic algorithms used in our study are presented in Table 2.

4.1. LC-DAD. The sum of the chromatographic profiles of raw data is presented in Figure 1, where it is clear that at least six compounds are present in the data, the remaining two being embedded. The retention time of each compound in the data was estimated by OPA and GA using the method described in section 3.2.3. All of the estimations were performed using a fixed number of function evaluations for both models using classical optimization functions. The concentration profiles were calculated using the optimized parameters described in eqs 4 and 5. A comparison of concentration profiles created by all of the methods for both models is presented in Figure 2. It can be observed that GAs produced concentration profiles that differ from the others. All of the classical methods produced more or less similar profiles except for compound **III**, where there is big variation in all estimates. The results of all of the optimization methods are compared by calculating the dissimilarity measure, using eq 26, between the pure and reconstructed spectral profiles. The dissimilarity values for all of the compounds by each optimization method are summed and presented in Table 3. It is evident that the *fmincon* function performed better than the other methods. It also indicates that all of the methods produced better results for the PMG_1 model than for the PMG_2 model, except *lsqcurvefit*. A comparison of pure and reconstructed spectral profiles created by *fmincon* for the PMG_1 function are presented in Figure 3, which shows a slight difference between the spectra of compounds **III**; for

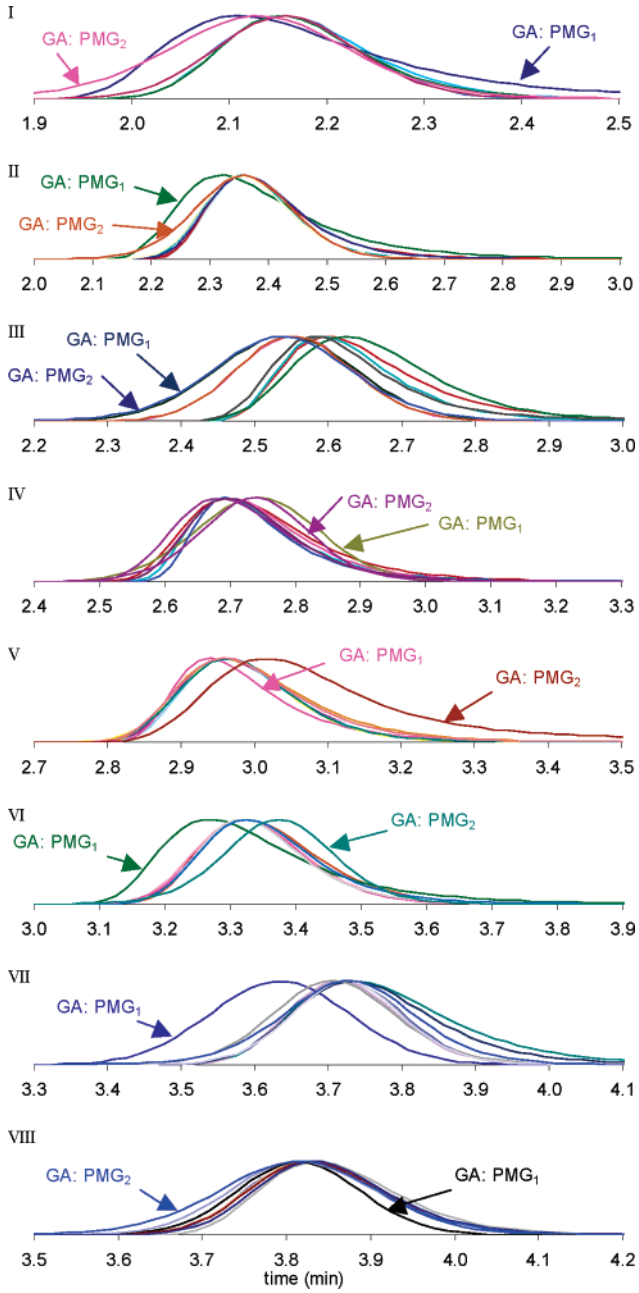

Figure 2. Comparison of concentration profiles calculated by different functions and the PMG_1 and PMG_2 models for LC-DAD data.

Table 3. Sum of Dissimilarity Calculated on LC-DAD Data for Optimization Functions

model	method	sum of dissimilarities
PMG_1	*fmincon*	0.308
PMG_2	*fmincon*	0.398
PMG_1	*fminsearch*	0.429
PMG_2	*fminsearch*	0.543
PMG_1	GA	0.572
PMG_2	*lsqcurvefit*	0.668
PMG_1	*lsqcurvefit*	0.689
PMG_2	GA	1.202

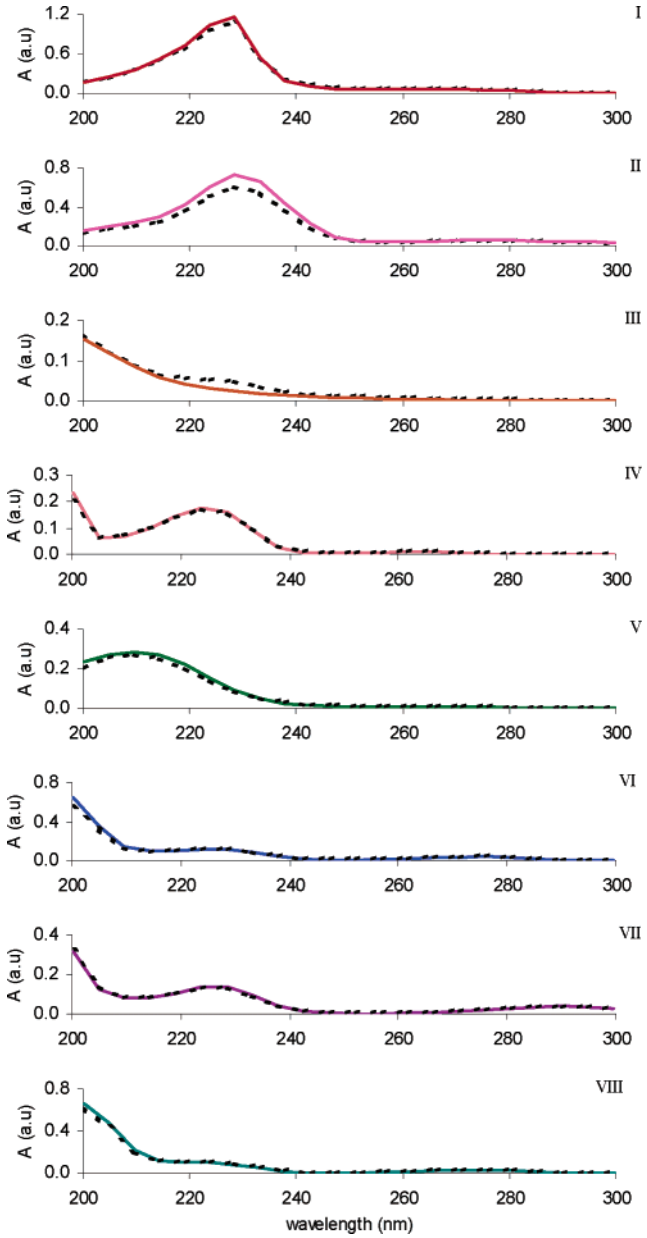

Figure 3. Comparison of reconstructed spectra (broken line) and pure spectra (solid line); the reconstructed spectra were obtained by *fmincon* using PMG_1 and concentration profiles scaled to a maximum height of 1.

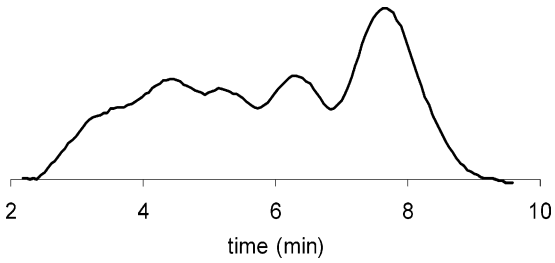

Figure 4. Sum of all chromatographic profiles of LC-NMR data of eight compounds.

the rest of spectra, the results were good. The results produced by GA for PMG_1 and PMG_2 show higher dissimilarity values, which does not indicate a good optimization.

4.2. LC-NMR. The sum of the chromatographic profiles for the LC-NMR data is presented in Figure 4; it is evident that at least five compounds are present in the data, and the remaining three are embedded. The pure NMR spectra are shown in Figure 5. Both OLS and CKVR require estimations of all of the concentration profiles. It has already been discussed in section 1 that LC-NMR data are characterized by distinctive regions of spectral peak clusters, which could

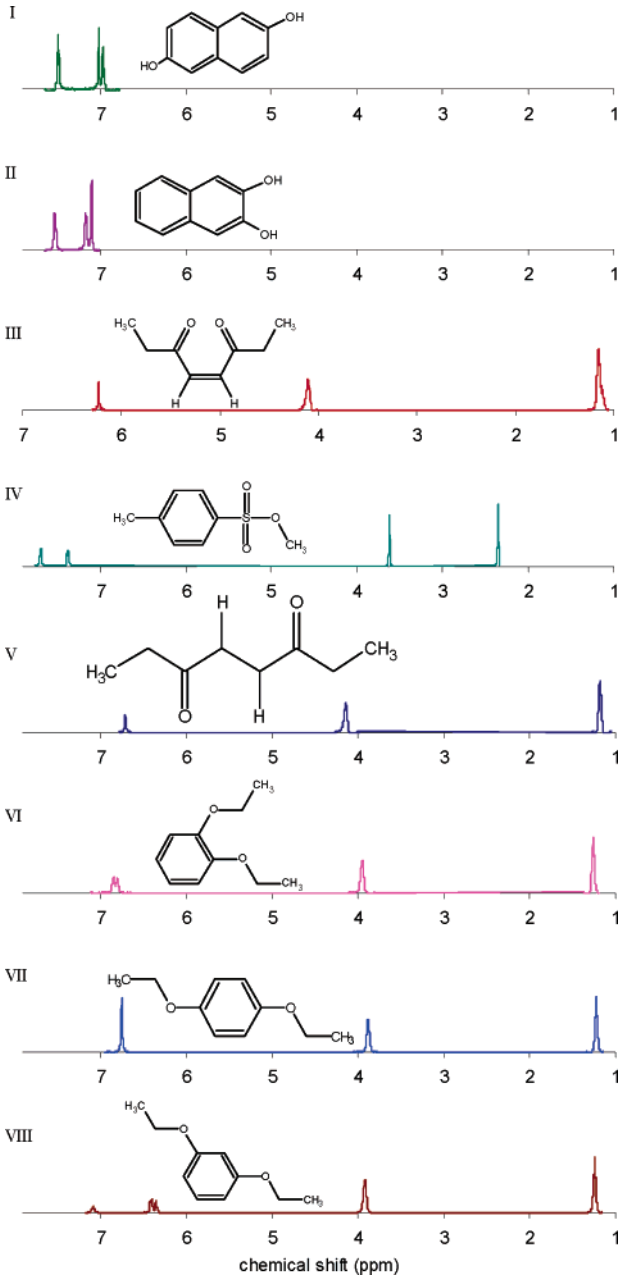

Figure 5. NMR spectra of pure compounds used in the LC-NMR study.

reduce the data complexity if the data analysis is performed on these spectral windows. The pure concentration profiles can be obtained by performing optimization over (a) the whole data set or (b) spectral peak clusters.

4.2.1. Optimization. 4.2.1.1. Optimization Using the Whole Data Set. Because this was a large data set, all of the methods (*fmincon*, *fminsearch*, *lsqcurvefit*, and GA) were slow to provide optimized parameters. Optimization of all of the parameters using the whole data set (eight compounds) is not only time-consuming but also prone to errors. The *lsqcurvefit* procedure produced very narrow concentration profiles for the first four compounds using PMG_1, and GA produced profiles which were widely different from the rest of the profiles.

4.2.1.2. Optimization Using Spectral Peak Clusters. The application of optimization methods on spectral peak clusters involves the creation of peak clusters as a first step.

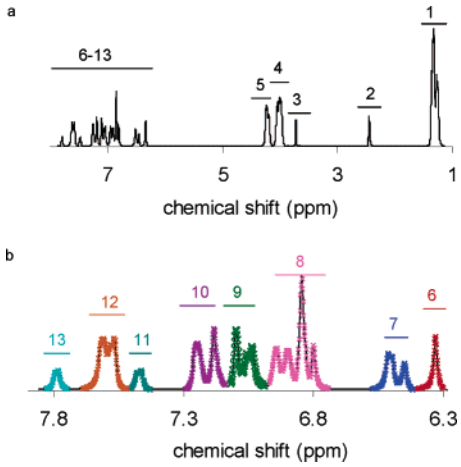

Figure 6. Plot of standard deviation as a function of chemical shift for LC-NMR data. (a) Regions of all peak clusters and (b) zoom of peak cluster region for 6−13.

4.2.1.2.1. Creation of Spectral Peak Clusters. The determination of peak cluster regions was automated by calculating the standard deviation at each frequency, and only those variables that had a higher functional value than a preset lower limit were accepted. The method can be applied as a first step for peak cluster location, and then, selected regions are tuned by visual inspection. The peak cluster regions selected are presented in Figure 6, each region designated by a number.

4.2.1.2.2. Parameter Estimation. Optimization methods were applied to each peak cluster, and the parameters were averaged out over peak clusters belonging to the same compound. The relationship between peak clusters of individual compounds was built using retention time data. The concentration profiles obtained by all of the methods for the PMG_1 model are presented in Figure 7. It is evident that all of the methods produced more or less similar profiles. The results are better than those produced by using the whole data set.

4.2.2. Curve Resolution. 4.2.2.1. Parameters Based on the Whole Data Set. 4.2.2.1.1. OLS. OLS was applied and the spectral profiles calculated. The results were incorrect for all of the concentration profiles created and for both peak shape models.

4.2.2.1.2. CKVR. CKVR requires the rank of each peak cluster region and the cluster-to-compound relationship. The rank and cluster-to-compound relationship were determined by OPA and GA. Initially, OPA was performed, and then, the results were confirmed by GAs. In all cases where rank > 2 and a discrepancy existed between the results of OPA and GAs, the results of GAs were retained. Table 4 shows the results calculated by both methods.

The reconstructed spectral profiles were correct only for the *fmincon* function for the PMG_1 model. The parameters based on the PMG_2 model produced concentration profiles with an increasing baseline on both sides of the chromatographic peak and, hence, resulted in incorrect spectral profiles.

4.2.2.2. Parameters Based on Spectral Peak Clusters. 4.2.2.2.1. OLS. The results of OLS were incorrect for both peak shape functions and for all optimization methods.

4.2.2.2.2. CKVR. The spectral profiles produced by all of the functions using only the PMG_1 model were acceptable.

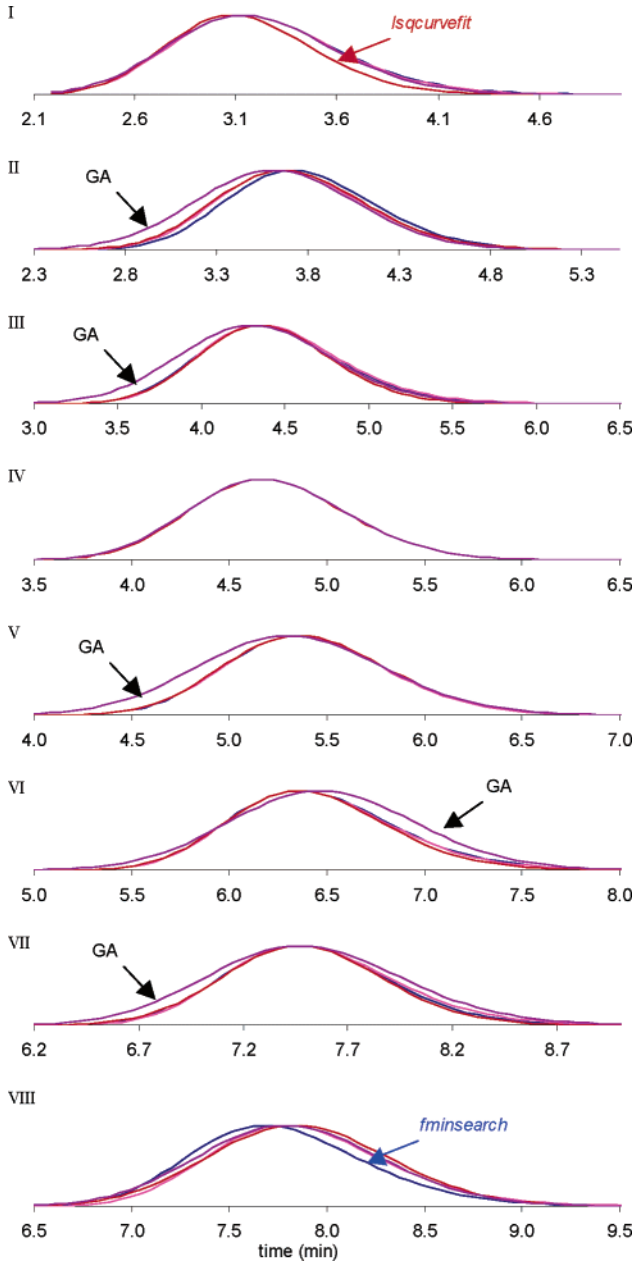

Figure 7. Comparison of concentration profiles calculated by different methods and the PMG_1 model averaged over LC-NMR peak clusters.

The only difference between reconstructed spectra by different functions was the NMR peak height; otherwise, peak positions and peak shapes were same. Because the PMG_2 model produced poor concentration profiles, the reconstructed spectra were incorrect. The reconstructed spectra by *fmincon* and the PMG_1 function are presented in Figure 8.

The GA model took the longest time for optimization, then *lsqcurvefit*, and the fastest were the *fmincon* and *fminsearch* functions. The optimized parameters calculated by GA and *fmincon* for rank 1 windows were, however, exactly the same. It suggests that GA can produce similar results but is a more computationally intense approach.

5. CONCLUSIONS

The low signal-to-noise ratio of LC-NMR data poses challenges to many curve resolution methods. In the present paper, it has been shown that the PMG_1 model works satisfactorily for the multivariate curve resolution of LC-DAD and LC-NMR data. Although *fminsearch*, *lsqcurvefit*, and GA produced reliable results, *fmincon* produced the best results in our study. This is mainly due to the bounded search space for optimization. The *lsqcurvefit* routine produced worse results, which is mainly due to peak shifting problems in LC-NMR, which weakens the assumption of linearity at each specific data point in the chromatogram. GA did not produce good results for LC-DAD data, especially for the PMG_2 function, which is mainly due to a large number of parameters that need to be optimized within a limited number of generations. Moreover, LC-NMR data contain a lower signal-to-noise ratio compared to that of LC-DAD data; the optimized parameters produced by PMG_2 for LC-NMR data were not meaningful, and only the PMG_1 model provided acceptable results.

Curve resolution for LC-NMR data requires the application of CKVR, which involves many steps. The correct curve resolution also needs the application of other chemometric methods for rank analysis and for key variable (retention time) selection. However, it is very important to obtain accurate results about the rank of the peak clusters and the retention time, which if performed by GA is a computationally intense but reliable method.

This paper has tackled a challenging problem of applying multivariate curve resolution to LC-NMR, a technique for which many historic approaches, developed primarily with DAD-HPLC in mind, break down and which requires the development of new data analytical techniques. Genetic algorithms, in general, have not been widely applied to the deconvolution of hyphenated chromatographic data, yet they show promise, especially using detector techniques for which there are potentially very characteristic variables (such as LC-NMR). The methodology in this paper works very well for the specific case studies represented in this paper. In particular, the data of Figure 4 represent a high level of overlap in LC-NMR, which was chosen as a mixture to indicate a problem that is quite hard to solve, although good signal-to-noise ratios are required, and the concentration levels are typical of those that can be studied successfully by on-flow NMR and the field strengths employed in this paper. When higher field strengths and newer probes, such as cryoprobes, are used, greater NMR sensitivity can be obtained.

Table 4. Key Variable Selection and Rank Analysis by GA[a]

	retention time (min)								
peak cluster	1	2	3	4	5	6	7	8	true rank
1			4.35		5.36	6.36	7.47	7.89	5
2				4.62					1
3				4.62					1
4						6.36	7.42	7.94	3
5			4.35		5.36				2
6			4.35						1
7								7.79	1
8					5.36	6.36	7.47		3
9	3.03								1
10		3.66						7.79	2
11				4.62					1
12	3.08	3.72							2
13				4.67					1

[a]OPA was used where rank < 3; the 13 peak clusters refer to Figure 6.

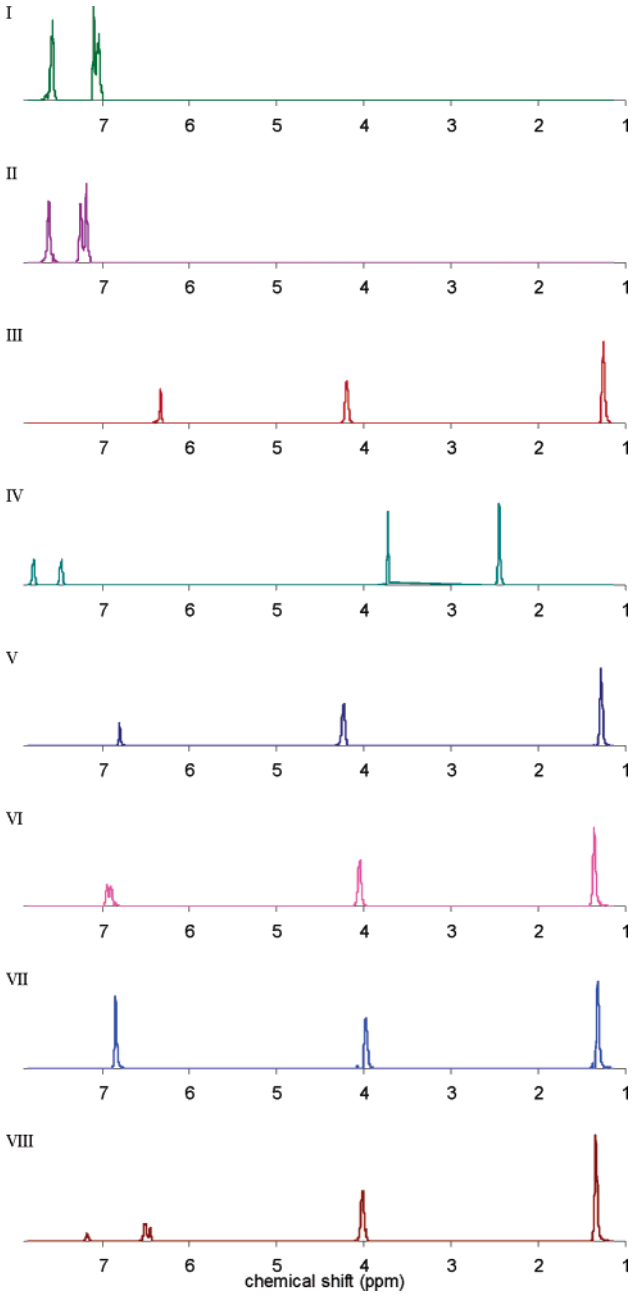

Figure 8. Reconstructed NMR spectra by *fmincon* and the PMG_1 model using the CKVR method on LC-NMR.

ACKNOWLEDGMENT

M.W. gratefully acknowledges the financial support provided by the Ministry of Science & Technology, Government of Pakistan, and the help and support from the PAEC, Pakistan. Both authors are thankful to Drs. C.Y. Airiau and M. Murray for helping in the LC-NMR experiments.

REFERENCES AND NOTES

(1) Lopez-Grio, S. J.; Vivo-Truyols, G.; Torres-Lapasio, J. R.; Garcia-Alvarez-Coque, M. C. Resolution Assessment and Performance of Several Organic Modifiers in Hybrid Micellar Liquid Chromatography. *Anal. Chim. Acta* **2001**, *433*, 187−198.

(2) Baeza-Baeza, J. J.; Garcia-Alvarez-Coque, M. C. Prediction of Peak Shape as a Function of Retention in Reversed-Phase Liquid Chromatography. *J. Chromatogr., A* **2004**, *1022*, 17−24.

(3) Malinowski, E. R. In *Factor Analysis in Chemistry*, 3rd ed.; Wiley: Canada, 2002.

(4) Bevington, P. R.; Robinson, D. K. In *Data Reduction and Error Analysis for the Physical Sciences*, 3rd ed.; McGraw-Hill: New York, 2003.

(5) Toft, J.; Kvalheim, O. M.; Libnau, F. O.; Nodland, E. Nonlinear Curve Fitting of Bilinear Data Using Orthogonal Projections for Rank Analysis. Applications to Gas Chromatography/Infrared Spectrometry and Variable Temperature Infrared Studies. *Vib. Spectrosc.* **1994**, *7*, 125−137.

(6) Maeder, M.; Zuberbuhler, A. D. Nonlinear Least Squares Fitting of Multivariate Absorption Data. *Anal. Chem.* **1990**, *62*, 2220−2224.

(7) Olive, J.; Grimalt, J. O.; Iturriaga, H. Resolution of Overlapping Peaks in Gas and Liquid Chromatography. Evaluation of Different Statistical and Empirical Functions. *Anal. Chim. Acta* **1989**, *219*, 257−272.

(8) Frans, S. D.; McConnell, M. L.; Harris, J. M. Multiwavelength Detection and Reiterative Least Squares Resolution of Overlapped Liquid Chromatographic Peaks. *Anal. Chem.* **1985**, *57*, 1552−1559.

(9) Sharaf, M. A.; Kowalski, B. R. Quantitative Resolution of Fused Chromatographic Peaks in Gas Chromatography−Mass Spectrometry. *Anal. Chem.* **1982**, *54*, 1291−1296.

(10) Knorr, F. J.; Harris, J. M. Resolution of Multicomponent Fluorescence Spectra by an Emission Wavelength-Decay Time Data Matrix. *Anal. Chem.* **1981**, *53*, 272−276.

(11) Aishima, T.; Nakai, S. Deconvolution of Gas Chromatographic Profiles for Multicomponent Model Mixtures and Essential Oils by the Simplex Algorithm. *Anal. Chim. Acta* **1991**, *248*, 41−50.

(12) Lawton, W. H.; Sylvestre, E. A. Self-Modeling Curve Resolution. *Technometrics* **1971**, *13*, 617−633.

(13) Geladi, P. Chemometrics in Spectroscopy. Part 1. Classical Chemometrics. *Spectrochim. Acta, Part B* **2003**, *58*, 767−782.

(14) Pere-Trepat, E.; Hildebrandt, A.; Barcelo, D.; Lacorte, S.; Tauler, R. Fast Chromatography of Complex Biocide Mixtures Using Diode Array Detection and Multivariate Curve Resolution. *Chemomem. Intell. Lab. Syst*. **2004**, *74*, 293−303.

(15) Haario, H.; Taavitsainen, V.-M. Combining Soft and Hard Modelling in Chemical Kinetic Models. *Chemomem. Intell. Lab. Syst*. **1998**, *44*, 77−98.

(16) Jaumot, J.; Vives, M.; Gargallo, R.; Tauler, R. Multivariate Resolution of NMR Labile Signals by Means of Hard- and Soft-Modelling Methods. *Anal. Chim. Acta* **2003**, *490*, 253−264.

(17) Marques, I.; Fonrodona, G.; Baro, A.; Guiteras, J.; Beltran, J. L. Study of Solvent Effects on the Acid−Base Behaviour of Adenine, Adenosine 3′, 5′-Cyclic Monophosphate and Poly(Adenylic) Acid in Acetonitrile−Water Mixtures Using Hard-Modelling and Soft-Modelling Approaches. *Anal. Chim. Acta* **2002**, *471*, 145−158.

(18) Juan, A. D.; Maeder, M.; Martinez, M.; Tauler, R. Application of a Novel Resolution Approach Combining Soft- and Hard-Modelling Features to Investigate Temperature-Dependent Kinetic Processes. *Anal. Chim. Acta* **2001**, *442*, 337−350.

(19) Wasim, M.; Hassan, M. S.; Brereton, R. G. Evaluation of Chemometric Methods for Determining the Number and Position of Components in High-Performance Liquid Chromatography Detected by Diode Array Detector and On-Flow ^{1}H Nuclear Magnetic Resonance Spectroscopy. *Analyst* **2003**, *128*, 1082−1090.

(20) Wasim, M.; Brereton, R. G. Determination of the Number of Significant Components in Liquid Chromatography Nuclear Magnetic Resonance Spectroscopy. *Chemom. Intell. Lab. Syst.* **2004**, *72*, 133−151.

(21) Wasim, M.; Brereton, R. G. Application of Evolving Factor Analysis to On-Flow LC-NMR Data Using Spectral Windows. *Chemom. Intell. Lab. Syst*. In Press.

(22) Shen, H.; Airiau, C. Y.; Brereton, R. G. Resolution of On-Flow LC/NMR Data by Multivariate Methods − A Comparison. *J. Chemom.* **2002**, *16*, 469−481.

(23) Shen, H.; Airiau, C. Y.; Brereton, R. G. Resolution of On-Flow Liquid Chromatography Proton Nuclear Magnetic Resonance Using Canonical Correlation and Constrained Linear Regression. *Chemom. Intell. Lab. Syst*. **2002**, *62*, 61−78.

(24) Holland, J. H. In *Adaptation in Natural and Artificial Systems*; MIT Press: Cambridge, MA, 1992.

(25) Vivo-Truyols, G.; Torres-Lapasio, J. R.; Garcia-Alvarez-Coque, M. C. A Hybrid Genetic Algorithm With Local Search: I. Discrete Variables: Optimisation of Complementary Mobile Phases. *Chemom. Intell. Lab. Syst*. **2001**, *59*, 89−106.

(26) Shao, X.; Chen, Z.; Lin, X. Resolution of Multicomponent Overlapping Chromatogram Using an Immune Algorithm and Genetic Algorithm. *Chemom. Intell. Lab. Syst.* **2000**, *50*, 91−99.

(27) Vivo-Truyols, G.; Torres-Lapasio, J. R.; Garrido-Frenich, A.; Garcia-Alvarez-Coque, M. C. A Hybrid Genetic Algorithm With Local Search II. Continuous Variables: Multibatch Peak Deconvolution. *J. Chromatogr., A* **2001**, *59*, 107−120.

(28) Vivo-Truyols, G.; Torres-Lapasio, J. R.; Caballero, R. D.; Garcia-Alvarez-Coque, M. C. Peak Deconvolution in One-Dimensional

Chromatography Using a Two-Way Data Approach. *J. Chromatogr., A* **2002**, *958*, 35−49.
(29) Nikitas, P.; Pappa-Louisi, A.; Papageorgiou, A.; Zitrou, A. On the Use of Genetic Algorithms for Response Surface Modeling in High-Performance Liquid Chromatography and Their Combination With the Microsoft Solver. *J. Chromatogr., A* **2002**, *942*, 93−105.
(30) Morris, G. A.; Freeman, R. Selective Excitation in Fourier Transform Nuclear Magnetic Resonance. *J. Magn. Reson*. **1978**, *29*, 433−462.
(31) Airiau, C. Y.; Shen, H.; Brereton, R. G. Principal Component Analysis in Liquid Chromatography Proton Nuclear Magnetic Resonance: Differentiation of Three Regio-Isomers. *Anal. Chim. Acta* **2001**, *447*, 199−210.
(32) Nikitas, P.; Pappa-Louisi, A.; Papageorgiou, A. On the Equations Describing Chromatographic Peaks and the Problem of the Deconvolution of Overlapped Peaks. *J. Chromatogr., A* **2001**, *912*, 13−29.
(33) Li, J. Development and Evaluation of Flexible Empirical Peak Functions for Processing Chromatographic Peaks. *Anal. Chem*. **1997**, *69*, 4452−4462.
(34) Di Marco, V. B.; Bombi, G. G. Mathematical Functions for the Representation of Chromatographic Peaks. *J. Chromatogr., A* **2001**, *931*, 1−30.
(35) Papai, Z.; Pap, T. L. Determination of Chromatographic Peak Parameters by Non-Linear Curve Fitting Using Statistical Moments. *Analyst* **2002**, *127*, 494−498.
(36) Li, J. Comparison of the Capability of Peak Functions in Describing Real Chromatographic Peaks. *J. Chromatogr., A* **2002**, *952*, 63−70.
(37) Torres-Lapasio, J. R.; Baeza-Baeza, J. J.; Garcia-Alvarez-Coque, M. C. A Model for the Description, Simulation, and Deconvolution of Skewed Chromatographic Peaks. *Anal. Chem*. **1997**, *69*, 3822−3831.
(38) Caballero, R. D.; Garcia-Alvarez-Coque, M. C.; Baeza-Baeza, J. J. Parabolic-Lorentzian Modified Gaussian Model for Describing and Deconvolving Chromatographic Peaks. *J. Chromatogr., A* **2002**, *954*, 59−76.
(39) Lagarias, J. C.; Reeds, J. A.; Wright, M. H.; Wright, P. E. Convergence Properties of the Nelder−Mead Simplex Method in Low Dimensions. *SIAM J. Opt.* **1998**, *9*, 112−147.
(40) Morgan, S. L.; Deming, S. N. Simplex Optimization of Analytical Chemical Methods. *Anal. Chem*. **1974**, *46*, 1170−1181.
(41) Fletcher, R.; Powell, M. J. D. A Rapidly Convergent Descent Method for Minimization. *Comput. J*. **1963**, *6*, 163−168.
(42) Goldfarb, D. A Family of Variable Methods Derived by Variational Means. *Math. Comput.* **1970**, *24*, 23−26.
(43) Han, S. P. A Globally Convergent Method for Nonlinear Programming. *J. Opt. Theor. Appl.* **1977**, *22*, 297−309.
(44) Powell, M. J. D. In *Numerical Analysis, Lecture Notes in Mathematics*; Watson, G. A., Ed.; Springer-Verlag: New York, 1978; Vol. 630.
(45) Coleman, T. F.; Li, Y. On the Convergence of Reflective Newton Methods for Large-Scale Nonlinear Minimization Subject to Bounds. *Math. Programming* **1994**, *67*, 189−224.
(46) Levenberg, K. A Method for the Solution of Certain Problems in Least-Squares. *Q. Appl. Math*. **1944**, *2*, 164−168.
(47) Marquardt, D. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *SIAM J. Appl. Math*. **1963**, *11*, 431−441.
(48) More, J. J. In *Numerical Analysis, Lecture Notes in Mathematics*; Watson G. A., Ed.; Springer-Verlag: New York, 1977; Vol. 630, pp 105−116.
(49) Michalewicz, Z. In *Genetic Algorithms + Data Structures = Evolution Programs*; 3rd ed.; Springer-Verlag: Berlin, 1996.
(50) Goldberg, D. E. In *Genetic Algorithms in Search, Optimization and Machine Learning*; Addison-Wesley: Reading, MA, 1989.
(51) Lucasius, C. B.; Kateman, G. Understanding and Using Genetic Algorithms Part 1. Concepts, Properties and Context. *Chemom. Intell. Lab. Syst*. **1993**, *19*, 1−33.
(52) Hibbert, D. B. Genetic Algorithms in Chemistry. *Chemom. Intell. Lab. Syst*. **1993**, *19*, 277−293.
(53) Lucasius, C. B.; Kateman, G. Understanding and Using Genetic Algorithms Part 2. Representation, Configuration and Hybridization. *Chemom. Intell. Lab. Syst*. **1994**, *25*, 99−145.
(54) Leardi, R. Genetic Algorithms in Chemometrics and Chemistry. *J. Chemom*. **2001**, *15*, 559−569.
(55) Baker, J. E. In *Proceedings of the Second International Conference on Genetic Algorithms*; Grefenstette, J. J., Ed.; Lawrence Erlbaum Associates: Hillsdale, NJ, 1987; pp 14−21.
(56) Muhlenbein, H.; Schlierkamp-Voosen, D. The Distributed Breeder Genetic Algorithm. *J. Evol. Comput.* **1993**, *1*, 25−49.
(57) Wasim, M.; Brereton, R. G. Determination of Number of Significant Components and Key Variables Using Genetic Algorithms in Liquid Chromatography−Nuclear Magnetic Resonance Spectroscopy and Liquid Chromatography−Diode Array Detector. *Chemom. Intell. Lab. Syst.* In press.
(58) Harris, D. C. In *Quantitative Chemical Analysis*, 5th ed.; W. H. Freeman & Company: New York, 1999.
(59) Brereton, R. G. *Chemometrics: Data Analysis for the Laboratory and Chemical Plant*; Wiley: Chichester, U. K., 2003.
(60) Sanchez, F. C.; Khots, M. S.; Massart, D. L.; De Beer, J. O. Algorithm for the Assessment of Peak Purity in Liquid Chromatography with Photodiode-Array Detection. *Anal. Chim. Acta* **1994**, *285*, 181−192.
(61) Sanchez, F. C.; Khots, M. S.; Massart, D. L. Algorithms for the Assessment of Peak Purity in Liquid Chromatography with Photo-diode-Array Detection. Part II. *Anal. Chim. Acta* **1994**, *290*, 249−258.
(62) Sanchez, F. C.; Toft, J.; Bogaert, B. V. D.; Massart, D. L. Orthogonal Projection Approach Applied to Peak Purity Assessment. *Anal. Chem*. **1996**, *68*, 79−85.
(63) Sanchez, F. C.; Vandeginste, B. G. M.; Hancewicz, T. M.; Massart, D. L. Resolution of Complex Liquid Chromatography−Fourier Transform Infrared Spectroscopy Data. *Anal. Chem*. **1997**, *69*, 1477−1484.
(64) Tauler, R.; Barcelo, D. Multivariate Curve Resolution Applied to Liquid Chromatography−Diode Array Detection. *Trends Anal. Chem*. **1993**, *12*, 319−327.
(65) Sanchez, F. C.; Rutan, S. C.; Garcia, M. D. G.; Massart, D. L. Resolution of Multicomponent Overlapped Peaks by the Orthogonal Projection Approach, Evolving Factor Analysis and Window Factor Analysis. *Chemom. Intell. Lab. Syst*. **1997**, *36*, 153−164.
(66) Frenich, A. G.; Galera, M. M.; Vidal, J. L. M.; Massart, D. L.; Torres-Lapasio, J. R.; Braekeleer, K. D.; Wang, J.-H.; Hopke, P. K. Resolution of Multicomponent Peaks by Orthogonal Projection Approach, Positive Matrix Factorization and Alternating Least Squares. *Anal. Chim. Acta* **2000**, *411*, 145−155.

CI050112F