Available online at www.sciencedirect.com
SCIENCE DIRECT
Vibrational Spectroscopy 35 (2004) 131–137
VIBRATIONAL SPECTROSCOPY
www.elsevier.com/locate/vibspec

In situ bioprocess monitoring of *Escherichia coli* bioreactions using Raman spectroscopy

Harry L.T. Lee[a,*], Paolo Boccazzi[b], Nathalie Gorret[b], Rajeev J. Ram[a], Anthony J. Sinskey[b]

[a]*Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*
[b]*Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

Received 1 October 2003; received in revised form 9 December 2003; accepted 10 December 2003

Available online 1 February 2004

Abstract

We have demonstrated simultaneous concentration estimation of glucose, acetate, formate, lactate, and phenylalanine from in situ measured Raman spectra in *Escherichia coli* bioreactions. Attenuation due to light scattering from air bubbles and biomass was corrected by internally referencing the least-squares estimated concentrations to the estimated water concentration. Estimation accuracy was limited by errors in the physical model for the system, rather than noise. Eliminating model errors should enable shot-noise limited detection in the 0.1 mM range.

Keywords: Raman; Spectroscopy; In situ; Bioreactor; Bioprocess; Fermentation; Fermentor

1. Introduction

Bioprocesses are used to produce a wide variety of chemicals from alcohols, organic and amino acids, to insulin, enzymes, and antibiotics. The ability to monitor the chemical composition of a bioprocess is important for both bioprocess supervision for quality control, and bioprocess development where both the microbial strain and operating conditions are optimized. For process supervision, on-line data is necessary for timely control of process parameters and data reliability metrics are important for quality control. For bioprocess development, the flexibility to analyze a wide range of substrates and metabolites is important, and on-line measurements are useful for early identification of poor growth conditions, increasing overall experimental throughput.

In all cases, it is useful to be able to perform measurements in situ since it is not easy to automate sample removal and preparation. In addition, sample removal increases the risk of contamination and could perturb the physiology of the organisms. The inside of a bioreactor, however, is an extremely difficult environment for chemical sensors because all internal components must survive steam-sterilization at 120 °C for a few hours without adverse effect on their calibration. It is also challenging to develop chemical sensors specific to different analytes. For these reasons, vibrational spectroscopies and inference techniques to estimate the concentration of chemical components in bioprocess media are worthy of consideration for in situ measurements. Among the vibrational spectroscopies, Raman spectroscopy is promising for bioprocess monitoring applications due to the low interference from water, flexibility in choice of wavelength, and the availability of fiber optic probes.

The utility of Raman spectroscopy for the analysis of bioprocess media has been recognized by other authors. The majority have used implicit models [1–4], where reference measurements on a set of training samples are performed to acquire information about the system to be analyzed, and used to construct a model that implicitly accounts for any physical effects that influence the measured Raman spectra, allowing concentration estimates to be made on unprocessed samples. However, because the training samples only model the particular system to be analyzed, they are not generally applicable to other systems. While implicit models have utility and validity in bioprocess supervision applications, where the same bioreaction is to be monitored time and again for production purposes, they are not as useful in bioprocess development applications where bioreactions with variations in operating conditions, growth rates, and medium composition are typically run only once for screen-

* Corresponding author.
E-mail address: harrylee@mit.edu (H.L.T. Lee).

0924-2031/$ – see front matter © 2003 Published by Elsevier B.V.
doi:10.1016/j.vibspec.2003.12.015

ing. Thus, for bioprocess screening applications, explicit methods, based on physically modeling the system to be analyzed, are preferred. Explicit models have previously been used to estimate chemical concentrations in bioprocess media, where peak amplitudes were used to estimate the ethanol concentration in an anaerobic baker's yeast fermentation [5], and classical least-squares fitting was used to estimate glucose and spiked quantities of glutamine, lactate, and ammonia, in filtered samples of a bioreaction [6]. Background fluorescence was modeled by a linear function and variations in collected power were accounted for by normalizing the measured spectra to the water peak area. However, because limited analysis of the error was possible—due to lack of reference data [5] and over fitting by the addition of a free parameter to optimize the fit between the Raman and reference measurements [6]—and also because scattering was not significant [5] or removed [6] in these experiments, whether explicit methods can be successfully applied to in situ concentration estimation using Raman spectroscopy is an open question.

In this work, we will discuss our first results on in situ concentration estimation, using Raman spectroscopy and explicit models for concentration estimation. By modeling the effects of scattering, we have estimated the concentration of glucose, phenylalanine, acetate, formate, and lactate in two phenylalanine producing *Escherichia coli* bioreactions. In addition, an error analysis reveals noise is not currently limiting the concentration estimation performance, but rather effort should be focused on controlling systematic model errors.

2. Experiment and methods

2.1. In situ Raman measurements

Raman spectra were measured in situ from an *E. coli* culture grown in a 2.5 l stirred tank bioreactor (Chemap CMF100). The Raman spectra were collected using an InPhotonics Raman probe, which was adapted for insertion into a standard 19 mm bioreactor port using two custom made components. The first was a brass tube (1.43 in. O.D., 0.74 mm wall) to provide structural support, in which the Raman probe was inserted, with one end threaded at 15.7 threads/cm to enable precise axial adjustment of the focal point of the probe. The second was an anodized aluminum tube (1.88 cm O.D., 1.43 cm I.D.) with a magnesium acetate coating that had two outer o-ring seals and a recessed sapphire window (16 mm diameter, 1 mm thick) sealed to the end using epoxy (Tra-Bond 2143D). Upon insertion into a bioreactor port, the aluminum tube provided a sterile interface in which the Raman probe could be inserted. An external cavity laser, similar in design to [7], provided optical excitation at 785 nm resulting in 50–60 mW of power at the output of the probe. The collection fiber was butt coupled to the entrance slit of an Acton SpectraPro 300i spectrograph which was set to use a 600 g/mm grating with blaze wavelength of 1 μm and spectra were measured using a liquid nitrogen cooled back-illuminated, deep-depletion CCD camera (Roper Scientific Spec10:400BR) for maximal infrared sensitivity. Unless otherwise noted, the data was acquired for a total of 300 s in thirty 10 s intervals. Cosmic ray events were removed by comparing the data for each pixel across the thirty 10 s intervals, and setting points larger than two standard deviations from the mean equal to the mean. The final Raman spectrum was taken as the average of the cosmic ray cleaned spectra.

Two bioreactions were conducted using *E. coli* ATCC31883, a phenylalanine producing strain (US Patent 4,681,852), grown in defined medium at 37 °C. The medium composition is shown in Table 1, along with the expected concentrations of the bioreaction products. During the bioreaction, the pH was automatically controlled at 7.0 ± 0.05 by adding NaOH or HCl. The first bioreaction (BR1) was cultured for 22 h, with aeration provided by 1 VVM air (volume air per volume liquid per minute) and 500 rpm agitation. The nominal initial glucose concentration was 111 mM (20 g/l). The second bioreaction (BR2), was cultured for a total of 68 h, with aeration provided by 1 VVM air and 800 rpm agitation. The nominal initial glucose concentration was 167 mM (30 g/l) and an additional 90 mM (15 g/l) of glucose was added 39 h into the bioreaction in an attempt to induce more growth and phenylalanine production.

Table 1
Medium composition for *E. coli* ATCC31883 bioreactions and expected concentration of bioreaction products

Component	Concentration (mM)	Description
K_2HPO_4	61	Medium buffer, phosphorous and nitrogen sources
NaH_2PO_4	35	
NH_4SO_4	15	
NH_4Cl	70	
$MgSO_4 \cdot 7H_2O$	0.76	
$Ca(NO_3)_2 \cdot 4H_2O$	0.06	
Glucose	110–170	Carbon source
$(NH_4)_6(MoO_7)_{24}$	0.003	Trace elements
H_3BO_3	0.4	
$MnCl_2$	0.08	
$ZnSO_4$	0.01	
Ampicillin	0.285	Antibiotic
$FeCl_3$	0.14	Supplements
Tryptophan	0.2	
Tyrosine	0.5	
Thyamine	0.002	
	Expected concentration (mM)	
Acetate	100	Mixed acid fermentation products
Formate	100	
Lactate	100	
Succinate	100	
Phenylalanine	10	Product

The medium buffer components, including the phosphorous and nitrogen sources (Table 1) were steam-sterilized inside the bioreactor vessel for 90 min at 120 °C and 138 kPa, along with the aluminum Raman adapter tube with sapphire window, and the dissolved oxygen, pH, and temperature probes. The remaining medium components (Table 1) were filter sterilized and added after the bioreactor cooled overnight. The inoculum was prepared by streaking a glycerol stock culture on Luria-Bertani (LB) agar plates, and then transferring a single colony to an LB culture tube, which was subsequently incubated at 37 °C. When the culture reached an optical density (OD) of 1 at 600 nm (Spectronic 20 Genesys, Spectronic Instruments), 6.25 ml from the tube was added to a shake flask containing 118.75 ml of the defined medium (Table 1). The flask was shaken on a rotary table at 37 °C until the culture reached $OD_{600\,nm} = 1$ at which point, 125 ml from the shake flask was used to inoculate (5%, v/v) the bioreactor.

In order to perform HPLC (Agilent 1100 series) and cell density (OD) measurements, samples were taken from the bioreactor at intervals ranging from 30 min to a few hours, depending on the growth stage of the culture. The OD for the bioreactor samples was measured using a 650 nm diode laser through a 1 cm path length plastic cuvette and dilutions were made such that the resultant $OD_{650\,nm}$ was <0.5. Samples for HPLC were filtered (0.2 μm PVDF Acrodisk, Pall) and frozen using liquid nitrogen for later analysis.

2.2. Concentration estimation

Concentration estimates were based on the following assumptions: the total Raman scattering was a linear combination of pure component Raman spectra; there was interference from fluorescence with unknown, smoothly varying spectrum; scattering from solids and air bubbles primarily effected the amplitude of the signal—wavelength dependent scattering was treated as a higher order effect; noise in the measured spectrum was dominated by shot-noise; and there were background subtraction errors in the measurement of the pure spectra due to small differences in their measurement conditions, which led to pure spectrum model errors.

With these assumptions, the concentration estimations were made using the following algorithm. Raman spectra from pure samples of the medium components, the expected products, and expected background interferents were measured by filling a 5 ml cuvette to the brim such that the meniscus was flat, after which the Raman probe was positioned above the meniscus and Raman spectrum acquired. Alternatively, 150 μl of sample was held by its surface tension in a 0.1 in. diameter hole drilled through a stainless steel rod and the Raman spectrum was measured laterally. For aqueous solutions, the measured spectrum of water was subtracted to obtain the pure spectrum. The measured pure Raman spectra are shown in Fig. 1. These pure spectra were normalized to unit concentration and with basis functions for a fourth-order polynomial, formed the pure spectrum matrix,

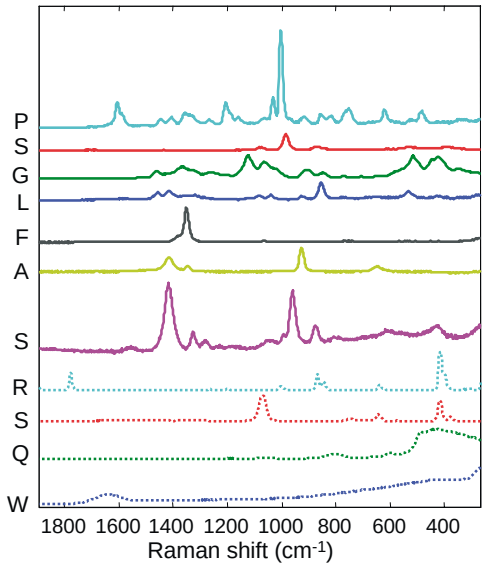

Fig. 1. Pure Raman spectra (solid) for 1 mM concentration of analytes, from top to bottom: phenylalanine, sulfate, glucose, lactate, formate, acetate, and succinate. The background components (dotted) are of arbitrary scale: room lights, sapphire, quartz, water.

K. The polynomial part of **K** served to model the background fluorescence and remove the influence of the polynomial part of any additive errors in the measured pure Raman spectra. Concentrations were subsequently estimated from the measured spectra in the least-squares sense by solving $\mathbf{s} = \mathbf{Kc}$ with non-negative constraint on concentrations. To correct for the dominant effects of scattering, all estimated aqueous concentrations are normalized to the estimated concentration of water, used as an internal standard, and multiplied by the concentration of water estimated from the pure spectrum measurements, before background subtraction. In this way, the method is robust to drift in laser power or coupling efficiency. Wavelength dependent scattering or errors in measuring the pure spectra were not corrected.

This approach to scattering correction was verified in an experiment, similar to [8], where the concentration of isopropanol was least-squares estimated from Raman spectra measured from aqueous solutions with different concentrations of dispersed polystyrene spheres, and also with varying focal depth or propagation length through the turbid solution. Both the estimated concentration of water and isopropanol showed the expected exponential dependence on focal depth and normalization to the water concentration corrected the isopropanol concentration to the expected value, albeit with some error, especially for high OD, likely due to the above mentioned effects of scattering.

3. Results and discussion

Concentration estimates were performed using the in situ measured Raman data, however, on-line concentration esti-

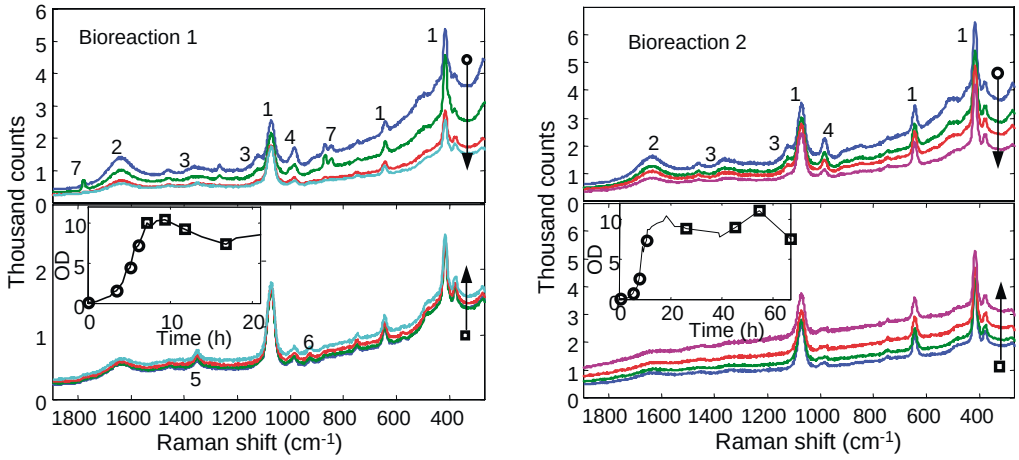

Fig. 2. Raman data at selected times with corresponding optical density curves for *E. coli* ATCC31883 bioreactions 1 and 2. Arrow indicates increasing time. Numbered spectral features correspond to: (1) sapphire; (2) water; (3) glucose; (4) medium sulfate; (5) formate; (6) acetate; (7) room lights.

mation (triangles in Fig. 3) was only performed during BR2. The on-line results suffered because the sapphire pure spectrum used in calibration was different from the one measured in situ due to changes in the Raman spectrum of the sapphire window after steam-sterilization. The primary concentration estimates shown (circles and squares in Fig. 3) therefore used a sapphire spectrum measured after removing the probe at the end of the bioreaction and hence, were off-line. Modification of the probe design to allow measurement of the sapphire spectrum after steam-sterilization without removing the probe, or the use of a different window material such as IR-quartz, will be necessary for improved on-line measurements.

Fig. 2 shows selected in situ measured Raman spectra from the two *E. coli* bioreactions described in Section 2 along with the corresponding growth curves in optical density units (the conversion to dry cell weight is approximately 0.5 g dcw/$OD_{650\ nm}$, from a single point calibration). The upper plots show the evolution of the Raman spectra during growth where the overall signal decreases due to increased scattering from the biomass. The lower plots show the evolution of the Raman spectra during stationary/death phase where the overall signal increases due to increased fluorescence background. This effect is more pronounced in BR2 due to the higher agitation speed.

Fig. 3 shows the Raman estimated concentrations versus the HPLC measured concentrations for glucose ((a) and (b) versus time), phenylalanine (c), acetate (d), formate (e), and lactate (f). Because Raman spectra were acquired at many more time points than samples for HPLC (20 for BR1 and 30 for BR2). Fig. 3 was made by linearly interpolating between HPLC measurements. The parameters for best-fit lines through the non-zero data points are shown in Table 2. Although they were used as a reference, HPLC measurements have their own associated error. Effects due to uncertainties in preparing the calibration samples and performing sample dilutions were controlled to less than 5%, however, for two samples in BR1 and three samples in BR2, the glucose concentrations measured by HPLC dropped 10–20% lower than the surrounding samples (Fig. 3b), which was unphysical since glucose is not produced by this strain. This appeared to be anomaly in the samples themselves since this deficiency persisted even after sample dilution. These data points were considered outliers and removed from the analysis.

The Raman estimated concentrations are quite smooth, as indicated by the small scatter in data points, which suggests the concentration estimates were not noise limited. Also, for the two bioreactions, the performance of the Raman estimated glucose concentrations are similar, while the performance for organic acids and phenylalanine are widely different. The Raman estimates for glucose (Fig. 3a) consistently underestimate the HPLC reference measurements, except after the glucose spike in BR2 (squares), and the error is not simply a scale error, as indicated by the negative intercepts and near unity slopes of the best-fit lines (Table 2). In BR1 (circles), the three organic acids (Fig. 3d–f) show qualitative agreement—indicated by near unity correlation coefficients—with scale errors for acetate and formate, and an offset error for lactate (Table 2), but phenylalanine (Fig. 3c) was not detected. In contrast, for BR2 (squares), phenylalanine was detected with excellent correspondence with the HPLC measurements, however, lactate was not detected and acetic acid was severely overestimated, with low correlation coefficient (Table 2).

The observation of smooth Raman estimated concentrations is consistent with a first-order noise analysis of the algorithm. From the solution of $\mathbf{s} = \mathbf{K}\mathbf{c}$ given by $\mathbf{c} = \mathbf{K}^{+}(\mathbf{s}_0 + \Delta\mathbf{s})$, where $\mathbf{K}^{+}$ is the pseudoinverse of $\mathbf{K}$ and $\Delta\mathbf{s}$ the shot-noise dominated measurement noise, the shot-noise contribution to the concentration estimate is simply $\Delta\mathbf{c} = \mathbf{K}^{+}\Delta\mathbf{s}$. For noise dominated by a smoothly varying background, this is approximately $\Delta\mathbf{c} \approx \mathrm{diag}\,(\mathbf{K}^{+}\mathbf{K}^{+\mathrm{T}})^{1/2}\sigma_\mathrm{b}$ or $\Delta c_i = (\Sigma_j \mathbf{K}_{ij}^{+2})^{1/2}\sigma_\mathrm{b}$, where we have approximated the shot noise in each pixel to be independent of other pixels and Gaussian with variance, σ_b^2, equal to the

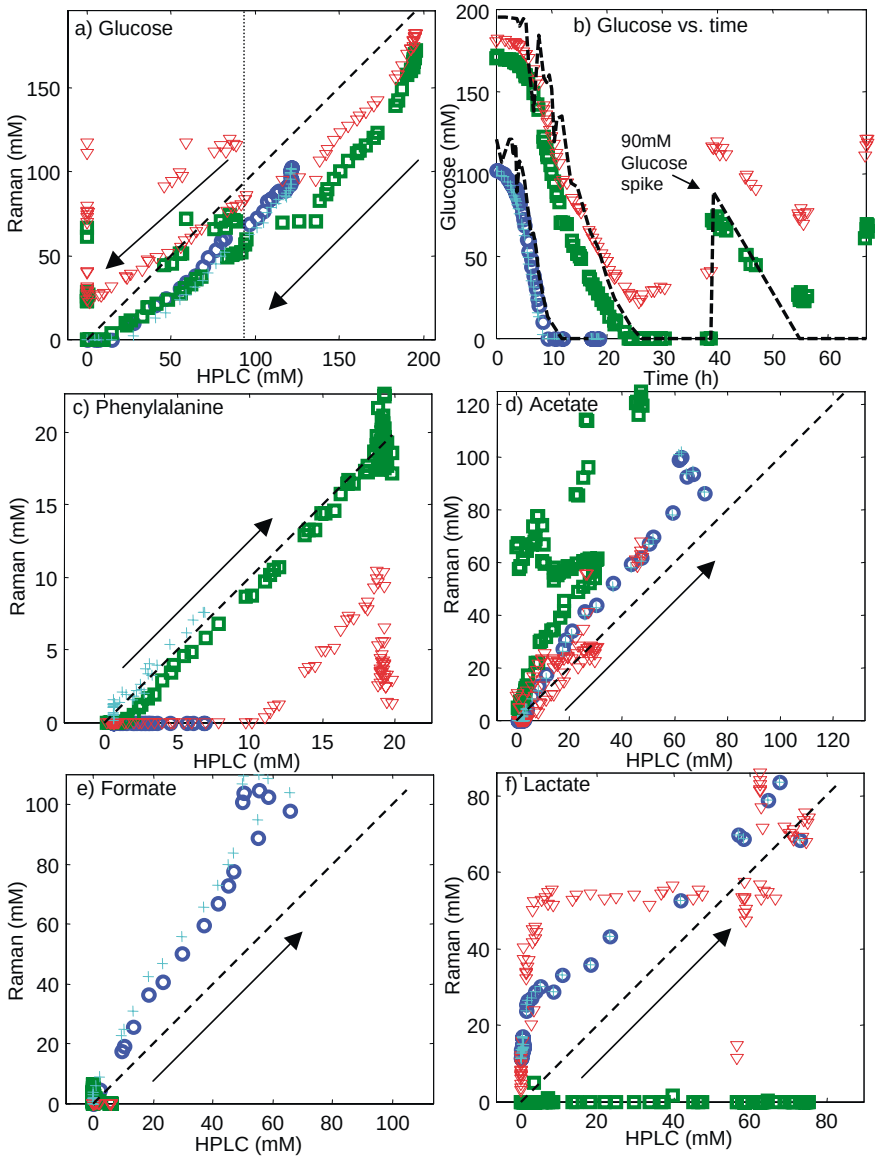

Fig. 3. Raman estimated concentrations vs. HPLC reference concentrations. Results are shown for BR1 (open circles), BR2 (squares), BR2 online (triangles), and BR1 ignoring data between 1032 and 1123 cm^{-1} at the sapphire spectrum peak (pluses). The dashed line is of unity slope and indicates perfect correspondence between Raman and HPLC. The arrow indicates increasing time. For BR2, a glucose spike of 90 mM was added as indicated by the dotted vertical line in (a). (b) shows the glucose concentration vs. time for BR1 and BR2 along with the HPLC measurements (dashed line).

average background level. If we further assume that the dominant contribution to the background is from water and include the normalization to the estimated water concentration, we can arrive at a lower bound on the shot-noise contribution to the estimated concentration:

$$\Delta c_i = \frac{1}{\sqrt{\mathrm{PT}}}\sqrt{\frac{(\mathbf{K}^{+}\mathbf{K}^{+\mathrm{T}})_{ii}\langle w\rangle}{G c_{\mathrm{w}}}}$$

where Δc_i (mM) indicates the noise contribution to the ith component, PT (mW s) is the power-integration time product, $\langle w\rangle$ (counts/(mW s)) the average of the Raman spectrum for water for unit power-integration time product, $\mathbf{K}^{+}$ ((mM mW s)/count) the pseudoinverse of $\mathbf{K}$ for unit power-integration time product, G (electrons/count) the CCD gain, and c_{w} the estimated water concentration which is normalized to 1 for no scattering.

Table 3 shows Δc_i calculated for the components of interest—using Raman data at the HPLC sample times—at unit power-integration time product, shown as theoretical shot-noise limited detection (SNLD). While the numbers are particular to the total collection efficiency for the optics used in this experiment, the relative values are general. Also shown, is the measured noise limited detection (MNLD) calculated from the standard deviation of the concentration estimate using the 30 individual 10 s time intervals for each data point, scaled by the square root of the water concentration to remove the effects of scattering. The reported num-

Table 2
Parameters for best-fit lines through non-zero data points in Fig. 3

Component	Slope	Intercept	Correlation coefficient	Standard error	σ (slope)	σ (intercept)	Points
Glucose							
BR1	0.89	−13.4	0.9926	4.16	0.02	1.87	32
BR2	0.90	−21.9	0.9787	12.20	0.03	3.74	51
BR2 OL	0.80	11.31	0.9839	9.44	0.02	2.89	51
BR1 NS	0.90	−19.37	0.9815	6.75	0.03	3.02	32
Phenylalanine							
BR1	XX	XX	XX	XX	XX	XX	XX
BR2	1.05	−1.07	0.9929	1.00	0.01	0.20	82
BR2 OL	1.17	−12.55	0.9913	0.46	0.03	0.53	24
BR1 NS	1.08	0.16	0.9788	0.44	0.04	0.11	34
Acetate							
BR1	1.43	−0.79	0.9917	4.67	0.03	1.10	34
BR2	1.76	26.02	0.7451	21.25	0.18	3.83	82
BR2 OL	1.19	0.52	0.9098	7.32	0.06	1.32	82
BR1 NS	1.44	−1.32	0.9906	5.00	0.04	1.18	34
Formate							
BR1	1.73	0.46	0.9907	5.51	0.04	1.21	34
BR2	NA	NA	NA	NA	NA	NA	NA
BR2 OL	NA	NA	NA	NA	NA	NA	NA
BR1 NS	1.82	2.18	0.9907	5.81	0.04	1.28	34
Lactate							
BR1	0.90	17.01	0.9690	5.39	0.04	1.07	34
BR2	XX	XX	XX	XX	XX	XX	XX
BR2 OL	0.56	28.59	0.7220	15.82	0.06	2.59	82
BR1 NS	0.71	9.80	0.9825	3.17	0.02	0.63	34

XX indicates Raman failure (zero estimate). Labels correspond to bioreaction 1 (BR1), bioreaction 2 (BR2), bioreaction 2 online using the pre-autoclave sapphire spectrum (BR2 OL), and bioreaction 1 ignoring data between 1032 and 1123 cm^{-1} at the sapphire spectrum peak (BR1 NS). For BR2 glucose, the data points after the glucose spike were not included in the line fitting and for BR2 OL phenylalanine, the data points for HPLC measured concentrations greater than 18 mM were not included in the line fitting.

bers are the average standard deviation for measurements where the analyte of interest was present. As expected, the MNLD is greater than the calculated water background detection limits; by a factor of 2 for BR1 and a factor of 3 for BR2, due to contributions to the overall noise from the Raman spectra of the other medium components. Taking into account the true shot-noise calculated from the measured data, rather than only the background contribution from water, the correspondence between the SNLD and MNLD is improved with the measured detection limit slightly greater than the calculated limit due to factors including laser power drift and signal fluctuations due to scattering from bubbles.

These results indicate the error in the Raman estimated concentrations were due to systematic errors in the spectral model and not due to noise or signal degradation due to scattering from biomass or bubbles. If these model errors can be corrected, shot-noise limited detection in the range of 0.1 mM should be achievable. Model errors translate to concentration errors because in the physical model $\mathbf{s} = \mathbf{Kc}$, $\mathbf{K}$ should contain the true in situ measured pure component spectra, not the pure spectra that we have

Table 3
Shot-noise limited detection (SNLD) calculated using shot-noise from: water background only (W), bioreaction 1 data (BR1) and bioreaction 2 data (BR2); measured noise detection limit (MNDL) for BR1 and BR2 calculated from the standard deviation of 30 individual measurements; SNLD for 55 mW and 300 s integration time; and root mean squared deviation from HPLC measurements (RMSD-HPLC) for BR1, BR2, and BR2 online (BR2o)

Component	SNLD (mM (mW s)$^{1/2}$)			MNDL (mM (mW s)$^{1/2}$)		SNLD for 55 mW–300 s (SNR =1) (mM)	RMSE-HPLC (mM)		
	W	BR1	BR2	BR1	BR2		BR1	BR2	BR2o
Phenylalanine	5.6	9	14	XX	18	0.04	XX	1.1	9.1
Glucose	13	20	27	25	41	0.10	20	34	34
Lactate	24	43	53	53	XX	0.19	22	XX	30
Formate	14	22	NA	26	NA	0.11	14	NA	NA
Acetate	19	32	40	37	52	0.15	15	37	7

XX indicates a Raman failure (0 mM estimate).

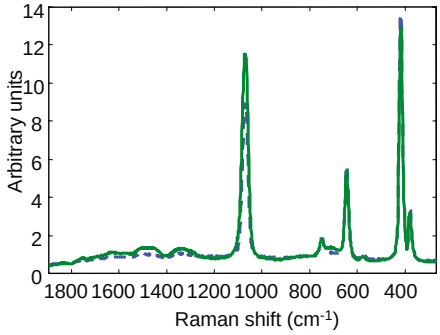

Fig. 4. Raman spectra for sapphire used in on-line (dashed) and off-line (solid) concentration estimates.

measured under clear conditions, which could also have some measurement error themselves.

Possible sources of model error are: error in pure spectra measurement or differences in actual in situ spectra compared to calibration spectra; distortion due to wavelength dependent scattering; temperature dependence of Raman probe filter transfer functions; or missing or extra basis spectra for minor components. Initial numerical experiments, where the errors were modeled and concentration estimates compared with those of the error-free model indicate that all of the above mentioned effects have the potential to influence the concentration estimates in a manner consistent with experimental observations. However, the large difference in results between the on-line and off-line Raman estimates for BR2 (triangles versus squares in Fig. 3) that are only due to small differences in the sapphire Raman spectrum (Fig. 4) indicates that pure spectrum errors are the most dominant. In fact, the phenylalanine estimate for BR1 (pluses in Fig. 3) recovered to agree with HPLC, with little effect on the other components, when the data between 1032 and 1123 cm^{-1}, which encompasses the largest differences in the sapphire spectrum, was not considered in the least-squares concentration estimates. While this result is suggestive, it is far from a solution since similar improvements in the lactic acid estimation for BR2 were not achieved using the same method. Additional experiments will be necessary to evaluate the most significant error contributions.

In summary, to make improvements, it will be necessary to physically model the processes that cause the in situ measured spectra to differ from the spectra measured under clear conditions, and with the appropriate transfer function—which will in general depend on independently measured parameters of the culture, such as pH or optical density—correct the pure component spectrum matrix **K** before performing concentration estimation. For example, if the pH were not controlled, the pure spectra would have to be measured over a range of pH values at calibration, since the Raman spectra for organic acids are modified when bound to a counter ion, and **K** would have to be populated with the pure spectra from the measured culture pH before each concentration estimate. Similarly, if wavelength dependent scattering proves to be important, we would need to determine the degree of distortion, using Mie theory for example, given an independent measurement of the optical density and populate **K** with distorted pure spectra relative to the calibration measurements where no scattering was present. While implicit methods may prove useful for extracting model parameters, we would not treat the entire bioreaction as a black box to avoid the disadvantages discussed in Section 1.

4. Summary

We have demonstrated simultaneous concentration estimation of glucose, acetate, formate, lactate, and phenylalanine from in situ measured Raman spectra in *E. coli* bioreactions using explicit models. Attenuation due to light scattering from air bubbles and biomass was corrected by internally referencing the least-squares estimated concentrations to the estimated water concentration. The observed estimation accuracy was limited by errors in the physical model for the system, rather than noise. With improved physical modeling and window materials with stable Raman spectrum, detection limits and sensitivity should approach shot-noise limited values, which are in the range of 0.1 mM for reasonable excitation power and integration times. Additional experiments and analysis will be necessary to identify and correct model errors and also to establish confidence intervals around the concentration estimates. With these developments Raman spectroscopy can be developed into a useful in situ and on-line bioprocess monitoring tool.

Acknowledgements

This work was supported by the DuPont MIT alliance.

References

[1] A.D. Shaw, N. Kaderbhai, A. Jones, A.M. Woodward, R. Goodacre, J.J. Rowland, D.B. Kell, Appl. Spectrosc. 53 (1999) 1419.
[2] S. Sivakesava, J. Irudayaraj, A. Demirci, J. Ind. Microbiol. Biotechnol. 26 (2001) 185.
[3] S. Sivakesava, J. Irudayaraj, A. Demirci, Proccess Biochem. 37 (2001) 371.
[4] A.C. McGovern, D. Broadhurst, J. Taylor, N. Kaderbhai, M.K. Winson, D.A. Small, J.J. Rowland, D.B. Kell, R. Goodacre, Biotechnol. Bioeng. 78 (2002) 527.
[5] T.B. Shope, T.J. Vickers, C.K. Mann, Appl. Spectrosc. 41 (1987) 908.
[6] Y. Xu, J.F. Ford, C.K. Mann, T.J. Vickers, J.M. Brackett, K.L. Cousineau, W.G. Robey, SPIE 2976 (1997) 10.
[7] A.S. Arnold, J.S. Wilson, M.G. Boshier, Rev. Sci. Instrum. 69 (1998) 1236.
[8] M. van den Brink, M. Pepers, A.M. van Herk, J. Raman Spectrosc. 33 (2002) 264.