

# Dynamic Range and Mass Accuracy of Wide-Scan Direct Infusion Nano-electrospray Fourier Transform Ion Cyclotron Resonance Mass Spectrometry-Based Metabolomics Increased by the Spectral Stitching Method

Andrew D. Southam,<sup>†</sup> Tristan G. Payne,<sup>‡</sup> Helen J. Cooper,<sup>†</sup> Theodoros N. Arvanitis,<sup>‡</sup> and Mark R. Viant<sup>\*†</sup>

School of Biosciences, and Department of Electrical, Electronic and Computer Engineering, School of Engineering, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK

Direct infusion nano-electrospray Fourier transform ion cyclotron resonance mass spectrometry (DI nESI FT-ICR MS) offers high mass accuracy and resolution for analyzing complex metabolite mixtures. High dynamic range across a wide mass range, however, can only be achieved at the expense of mass accuracy, since the large numbers of ions entering the ICR detector induce adverse space-charge effects. Here we report an optimized strategy for wide-scan DI nESI FT-ICR MS that increases dynamic range but maintains high mass accuracy. It comprises the collection of multiple adjacent selected ion monitoring (SIM) windows that are stitched together using novel algorithms. The final SIM-stitching method, derived from several optimization experiments, comprises 21 adjoining SIM windows each of width  $m/z$  30 (from  $m/z$  70 to 500; adjacent windows overlap by  $m/z$  10) with an automated gain control (AGC) target of  $1 \times 10^5$  charges. SIM-stitching and wide-scan range (WSR; Thermo Electron) were compared using a defined standard to assess mass accuracy and a liver extract to assess peak count and dynamic range. SIM-stitching decreased the maximum mass error by 1.3- and 4.3-fold, and increased the peak count by 5.3- and 1.8-fold, versus WSR (AGC targets of  $1 \times 10^5$  and  $5 \times 10^5$ , respectively). SIM-stitching achieved an rms mass error of 0.18 ppm and detected over 3000 peaks in liver extract. This novel approach increases metabolome coverage, has very high mass accuracy, and at 5.5 min/sample is conducive for high-throughput metabolomics.

Metabolomics involves the measurement of multiple small molecules within a biological sample to generate a unique metabolic profile. These profiles can be compared directly between different biological phenotypes, and as metabolites represent the

end products of complex cellular control systems, such analyses can give insight into upstream gene and protein activity.<sup>1</sup> This approach has been applied to many fields, notably toxicology, drug design, disease diagnosis, and quality control of food substances.<sup>2–6</sup> Reproducibility of the measurement system is critical, with nuclear magnetic resonance (NMR)-based methods proving very robust.<sup>6,7</sup> However, due to its superior sensitivity over NMR, mass spectrometry (MS) represents an attractive detection method for metabolomics.<sup>8,9</sup> Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS) is a particularly powerful tool for complex mixture analysis due to its ultrahigh mass resolution and mass accuracy<sup>10,11</sup> and has been applied successfully to a small number of metabolomics studies.<sup>12–16</sup> In principle, this analysis method enables the empirical formulas of many low molecular weight metabolites to be unambiguously identified based upon

- (1) Fiehn, O.; Kopka, J.; Dormann, P.; Altmann, T.; Trethewey, R. N.; Willmitzer, L. *Nat. Biotechnol.* **2000**, *18*, 1157–1161.
- (2) Griffin, J. L.; Shockcor, J. P. *Nat. Rev. Cancer* **2004**, *4*, 551–561.
- (3) Nicholson, J. K.; Connelly, J.; Lindon, J. C.; Holmes, E. *Nat. Rev. Drug Discovery* **2002**, *1*, 153–161.
- (4) Rischer, H.; Oksman-Caldentey, K. M. *Trends Biotechnol.* **2006**, *24*, 102–104.
- (5) Stoughton, R. B.; Friend, S. H. *Nat. Rev. Drug Discovery* **2005**, *4*, 345–350.
- (6) Viant, M. R.; Rosenblum, E. S.; Tjeerdema, R. S. *Environ. Sci. Technol.* **2003**, *37*, 4982–4989.
- (7) Nicholson, J. K.; Lindon, J. C.; Holmes, E. *Xenobiotica* **1999**, *29*, 1181–1189.
- (8) Dunn, W. B.; Ellis, D. I. *Trends Anal. Chem.* **2005**, *24*, 285–294.
- (9) Villas-Boas, S. G.; Mas, S.; Akesson, M.; Smedsgaard, J.; Nielsen, J. *Mass Spectrom. Rev.* **2005**, *24*, 613–646.
- (10) Hendrickson, C. L.; Emmett, M. R. *Annu. Rev. Phys. Chem.* **1999**, *50*, 517–536.
- (11) Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S. *Mass Spectrom. Rev.* **1998**, *17*, 1–35.
- (12) Cooper, H. J.; Marshall, A. G. *J. Agric. Food Chem.* **2001**, *49*, 5710–5718.
- (13) Fard, A. M.; Turner, A. G.; Willett, G. D. *Aust. J. Chem.* **2003**, *56*, 499–508.
- (14) Stentiford, G. D.; Viant, M. R.; Ward, D. G.; Johnson, P. J.; Martin, A.; Wei, W. B.; Cooper, H. J.; Lyons, B. P.; Feist, S. W. *Omicron: J. Integrative Biol.* **2005**, *9*, 281–299.
- (15) Aharoni, A.; Ric de Vos, C. H.; Verhoeven, H. A.; Maliepaard, C. A.; Kruppa, G.; Bino, R.; Goodenowe, D. B. *Omicron: J. Integrative Biol.* **2002**, *6*, 217–234.
- (16) Brown, S. C.; Kruppa, G.; Dasseux, J. L. *Mass Spectrom. Rev.* **2005**, *24*, 223–231.

\* Corresponding author. Phone: +44-(0)121-414-2219. Fax: +44-(0)121-414-5925. E-mail: M.Viant@bham.ac.uk.

<sup>†</sup> School of Biosciences.

<sup>‡</sup> Department of Electrical, Electronic and Computer Engineering, School of Engineering.

mass measurements alone. Furthermore, if isotope information is used in conjunction with accurate mass measurements, identification of peaks over a wide mass range can be achieved with even greater certainty.<sup>17</sup> This is of considerable importance in metabolomics, as the identification of unknown metabolites remains one of the biggest analytical challenges.

Liquid chromatography MS (LC-MS)<sup>18–20</sup> and direct infusion electrospray MS (DI ESI MS)<sup>21–24</sup> are both commonly used in metabolomics. One major benefit of DI ESI MS is that the dependent axis corresponding to the mass-to-charge ratio ( $m/z$ ) is substantially more reproducible than in LC-MS, where chromatographic retention time ( $R_T$ ) will change as a function of column age and performance; e.g., a routine 1 ppm uncertainty in  $m/z$  in FT-ICR MS is equivalent to a 0.0018-s uncertainty in  $R_T$  for a typical 30-min chromatographic separation, a value that is currently unachievable. The reproducibility of the  $m/z$ -dependent axis value is important for spectral interpretation as well as for facilitating comparison of data sets collected during different experimental sessions. Although DI ESI MS benefits from a fundamentally more simple one-dimensional data set than the two-dimensional LC-MS data set, the availability of the time dimension in LC-MS can provide better discrimination of metabolites and aid in structural identification. Considering sample throughput, DI ESI MS offers shorter analysis times than conventional LC-MS, which is clearly an advantage. Conventional DI ESI MS, however, does suffer from ionization suppression and enhancement, which arises when particular analytes preferentially ionize over less polar metabolites in the complex mixture. The conventional method for reducing this problem is to utilize LC separation prior to MS analysis,<sup>25</sup> which removes salts from the complex mixture and reduces the number of compounds entering the mass spectrometer at any given time. However, by slowing the sample delivery rate to 200 nL/min using a chip-based DI nanoelectrospray source (NanoMate, Advion Biosciences), as reported here, ionization suppression has been shown to be substantially reduced compared to conventional DI electrospray ionization.<sup>26</sup>

For samples analyzed by direct infusion, the finite dynamic range of a MS detector will limit the observation to only the most intense ions. To identify lower abundance ions, the dynamic range, i.e., the ratio of the highest to lowest concentration metabolites detected, needs to be increased. High dynamic range can be achieved using DI ESI FT-ICR MS by increasing the number of ions that enter the ICR detector cell. This, however, can signifi-

cantly increase space-charge effects,<sup>27,28</sup> which degrade the mass accuracy to such an extent that derivation of unique empirical formulas for the observed peaks is no longer possible. In this paper, we report a strategy for wide-scan DI nESI FT-ICR MS that effectively increases the overall dynamic range of the mass spectrum, enabling observation of both low- and high-concentration biological metabolites with high mass accuracy. The approach is based upon the collection of multiple narrow, overlapping spectra (or “windows”), that are subsequently combined (or “stitched”) together using novel algorithms. A somewhat similar non-data-dependent “wide-scan” acquisition strategy was reported by Venable et al.,<sup>29</sup> based upon the sequential isolation and fragmentation of multiple  $m/z$  10 wide precursor windows using LC-MS/MS, which increased the signal-to-noise ratio in a proteomics study. In our approach, each window is analyzed using selected ion monitoring (SIM) mode, and consequently, each window is composed of relatively few ions to minimize space-charge effects. This approach produces a “SIM-stitched”, wide-scan mass spectrum with a significantly greater dynamic range, enabling many more metabolites to be detected, and with higher mass accuracy than is possible with existing methods. To maintain high mass resolution, relatively slow data acquisition is required (typically 1 s/transient) with many transients averaged to produce a single mass spectrum. Therefore, our method requires continuous infusion of each sample for a few minutes, which is achieved using a NanoMate chip-based nanoelectrospray system.<sup>30</sup> This system has previously been used for metabolomics<sup>31</sup> and is fully automated, using a new sample delivery tip and electrospray nozzle for each analysis to enable high throughput with no cross-contamination.

Stitching the multiple SIM windows together to produce a single contiguous, wide-scan spectrum is important for several reasons. First, the multivariate analysis of mass spectral fingerprints using widely utilized methods in metabolomics, such as principal components analysis, requires a single spectral fingerprint per biological sample. Second, methods that search for characteristic peak spacings across the entire spectral range (for example, Breitling et al.<sup>32</sup>) also require one contiguous data set. Third, the stitching algorithm is used to mass calibrate SIM windows that do not contain internal calibrants (see Supporting Information for details). Finally, stitching multiple data sets together substantially aids data handling, storage, and visual inspection of the mass spectral measurements.

The aims of this study are therefore to develop and test a novel analytical and bioinformatic approach for wide-scan, high dynamic range DI nESI FT-ICR MS analysis of complex biological mixtures of low molecular weight metabolites, which can achieve a maximum mass error of <1 ppm and a mass resolution of 100 000. Such a high mass resolution and mass accuracy is currently only offered by FT-ICR MS and is key for the acquisition of reliable

(17) Kind, T.; Fiehn, O. *BMC Bioinformatics* **2006**, *7*.

(18) Idborg-Bjorkman, H.; Edlund, P. O.; Kvalheim, O. M.; Schuppe-Koistinen, I.; Jacobsson, S. P. *Anal. Chem.* **2003**, *75*, 4784–4792.

(19) Plumb, R. S.; Stumpf, C. L.; Gorenstein, M. V.; Castro-Perez, J. M.; Dear, G. J.; Anthony, M.; Sweatman, B. C.; Connor, S. C.; Haselden, J. N. *Rapid Commun. Mass Spectrom.* **2002**, *16*, 1991–1996.

(20) von Roepenack-Lahaye, E.; Degenkolb, T.; Zerjeski, M.; Franz, M.; Roth, U.; Wessjohann, L.; Schmidt, J.; Scheel, D.; Clemens, S. *Plant Physiol.* **2004**, *134*, 548–559.

(21) Allen, J.; Davey, H. M.; Broadhurst, D.; Heald, J. K.; Rowland, J. J.; Oliver, S. G.; Kell, D. B. *Nat. Biotechnol.* **2003**, *21*, 692–696.

(22) Castrillo, J. I.; Hayes, A.; Mohammed, S.; Gaskell, S. J.; Oliver, S. G. *Phytochemistry* **2003**, *62*, 929–937.

(23) Dunn, W. B.; Overy, S.; Quick, W. P. *Metabolomics* **2005**, *1*, 137–148.

(24) Goodacre, R.; York, E. V.; Heald, J. K.; Scott, I. M. *Phytochemistry* **2003**, *62*, 859–863.

(25) Siuzdak, G. *The Expanding Role of Mass Spectrometry in Biotechnology*; MCC Press: San Diego, CA, 2003.

(26) Hop, C.; Chen, Y.; Yu, L. J. *Rapid Commun. Mass Spectrom.* **2005**, *19*, 3139–3142.

(27) Ledford, E. B.; Rempel, D. L.; Gross, M. L. *Anal. Chem.* **1984**, *56*, 2744–2748.

(28) Zhang, L. K.; Rempel, D.; Pramanik, B. N.; Gross, M. L. *Mass Spectrom. Rev.* **2005**, *24*, 286–309.

(29) Venable, J. D.; Dong, M. Q.; Wohlschlegel, J.; Dillin, A.; Yates, J. R. *Nat. Methods* **2004**, *1*, 39–45.

(30) Schultz, G. A.; Corso, T. N.; Prosser, S. J.; Zhang, S. *Anal. Chem.* **2000**, *72*, 4058–4063.

(31) Boernsen, K. O.; Gatzek, S.; Imbert, G. *Anal. Chem.* **2005**, *77*, 7255–7264.

(32) Breitling, R.; Ritchie, S.; Goodenowe, D.; Stewart, M. L.; Barrett, M. P. *Metabolomics* **2006**, *2*, 155–164.

and meaningful metabolomics data. During this study, we have applied our SIM-stitching technique to a chemically defined mixture of poly(ethylene glycol) and amino acids, as well as to real biological samples (fish liver extracts). The chemically defined mixture enabled us to assess the mass accuracy of the FT-ICR MS as a function of the number of ions transferred into the ICR detector cell and to assess the mass accuracy of our SIM-stitching approach. The liver extracts were used to determine the ideal width of each SIM window and to assess the improvement in dynamic range by comparing the total number of peaks detected using our SIM-stitching approach with a leading commercial method.

## EXPERIMENTAL METHODS

**Poly(ethylene glycol) and Amino Acids (PEG&AA) Standard.** A solution of 0.0005% poly(ethylene glycol) (PEG) 200, 0.0005% PEG 600, and 0.5 mM concentrations of 10 amino acids (glycine, alanine, serine, proline, cysteine, aspartic acid, glutamine, phenylalanine, arginine, tyrosine) (Sigma-Aldrich, Dorset, UK) was prepared in 50:50 methanol/deionized water (both HPLC grade, J. T. Baker) with 0.25% formic acid (Fisher Scientific, Loughborough, UK).

**Preparation of Liver Extracts.** Frozen livers from a wild-caught species of marine flatfish called dab (*Limanda limanda*) were provided by the Centre for Environment, Fisheries and Aquaculture Science (Cefas, Weymouth, UK). Tissues were kept at  $-80\text{ }^{\circ}\text{C}$  until extraction. Each liver was homogenized in  $8\text{ }\mu\text{L}/\text{mg}$  (wet tissue mass) methanol and  $2.5\text{ }\mu\text{L}/\text{mg}$  water using a Precellys-24 ceramic bead-based homogenizer (Stretton Scientific Ltd.). Next,  $8\text{ }\mu\text{L}/\text{mg}$  chloroform (pesticide analysis grade, Fisher Scientific) and  $9.5\text{ }\mu\text{L}/\text{mg}$  water were added, the biphasic mixture was centrifuged ( $1500g$ ), and the upper (polar) and lower (non-polar) phases were isolated and frozen at  $-80\text{ }^{\circ}\text{C}$ . Prior to MS analysis, polar extracts were dried using a centrifugal concentrator (Thermo Savant, Holbrook, NY), resuspended in 3 times the volume of original solvent (50:50 methanol/water with 0.25% formic acid), and then centrifuged ( $5000g$ ). The nonpolar extracts were not used in this study.

**FT-ICR Mass Spectrometry.** Samples were analyzed using a hybrid 7-T Fourier transform ion cyclotron resonance mass spectrometer (LTQ FT, Thermo Electron Corp., Bremen, Germany) equipped with a chip-based direct infusion nanoelectrospray ionization assembly (NanoMate, Advion Biosciences, Ithaca, NY). Nanoelectrospray conditions comprised a  $200\text{ nL}/\text{min}$  flow rate, 0.5 psi backing pressure, and 1.65 kV electrospray voltage, controlled by ChipSoft software (version 6.4.3, Advion Biosciences). Various scan modes were used that we define as follows: “full scan” is a conventional data acquisition mode for wide  $m/z$  ranges, “SIM” is a conventional selected ion monitoring mode, and “wide-scan range” (WSR, Thermo Electron) is a new method designed to alleviate the loss of low mass ions when analyzing wide ranges.<sup>33</sup> WSR optimizes the transfer of ions from the linear ion trap to the ICR detector for full scans by segmenting the total  $m/z$  spectral range into an optimal number of smaller windows (typically 2–5 windows). This helps to minimize the loss of low mass ions due to time-of-flight effects. Mass resolution was

fixed at 100 000 (defined for an ion at  $m/z$  400) throughout. Automatic gain control (AGC, corresponding to the number of charges transferred from the front-stage ion trap to the ICR detector cell) targets were varied (discussed in Results), and the maximum ion trap fill time was set to 1 s throughout. Data were obtained either as processed mass spectra with associated peak lists (Xcalibur, version 2.0, Thermo Electron) or as transient files (i.e., scans recorded in the time domain), which were processed using custom-written MATLAB software, described below. The latter was required for all mass spectra subjected to the SIM-stitching algorithm.

**Processing of Transient Files.** Transient files were processed using custom-written MATLAB code (version 7, The MathWorks, Natick, MA) to generate a calibrated peak list (described in Supporting Information).

**Calibration and Calculation of Mass Errors.** Spectra were calibrated by one of two methods, either using RecalOffline (Thermo Electron) or custom-written MATLAB code (see Supporting Information). Both methods used the known accurate masses of at least three peaks within the mass spectrum for calibration. Calibrants used in the PEG&AA spectra are listed in Table S-1. For spectra of liver extracts, several known endogenous metabolites that were consistently present in 18 different fish (Table S-2) were used for internal calibration.<sup>34</sup> Mass errors for the PEG&AA spectra were calculated using all known peaks that were not used for calibration.

**SIM-Stitching Algorithm.** The SIM windows, represented in the frequency domain, were individually calibrated when calibrants were available. The windows were aligned with peaks present in the overlap region between adjacent calibrated windows. The final wide-scan spectrum was obtained by averaging the resultant windows and removing regions of  $m/z$  5 from each end of each window (for reasons justified in Results). All processing was executed in MATLAB, as described in Supporting Information, which additionally includes an assessment of the stability of the SIM-stitching algorithm as the number of internal calibrants is varied. The MATLAB SIM-stitching code is available on request from the corresponding author.

**Peak Counting and Dynamic Range Measurement.** To virtually eliminate the effects of residual noise, all peak counting was performed using three spectra obtained from the same sample. A peak was only considered real if, within a sliding window of 1 ppm, exactly one peak appeared in each of these three spectra. The “true” peak position was then taken to be the mean of the three peaks. The dynamic range was calculated from peak lists that had been filtered using the above algorithm, as the intensity ratio between the strongest and weakest peaks.

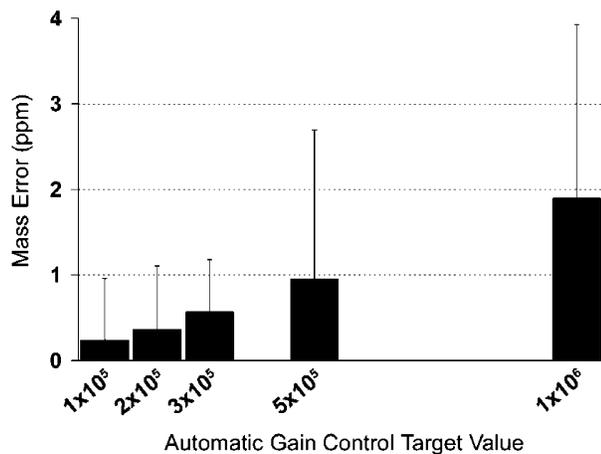
## RESULTS AND DISCUSSION

### Optimization of Number of Ions Transferred to ICR Cell.

Our first goal was to optimize the number of charges to be transferred from the front-stage ion trap to the ICR detector cell (using the AGC) in order to minimize space-charge effects and consequently maximize the mass accuracy. Since virtually all ions observed were singly charged (peaks occurred at integral mass values), the AGC target value was reflective of the actual number

(33) Muenster, H.; Sanders, M.; Warrack, B.; Lange, O.; Homing, S. In *53rd ASMS Conference on Mass Spectrometry*; San Antonio, TX, 2005.

(34) Yanofsky, C. M.; Bell, A. W.; Lesimple, S.; Morales, F.; Lam, T. T.; Blakney, G. T.; Marshall, A. G.; Carrillo, B.; Lekpor, K.; Boismenu, D.; Kearney, R. E. *Anal. Chem.* **2005**, *77*, 7246–7254.



**Figure 1.** Effect of the number of charges transferred to the ICR detector cell on the absolute mass accuracy of several known metabolites from a poly(ethylene glycol) and amino acid standard mixture. Solid bars denote the rms mass error, and error bars denote the maximum absolute mass error.

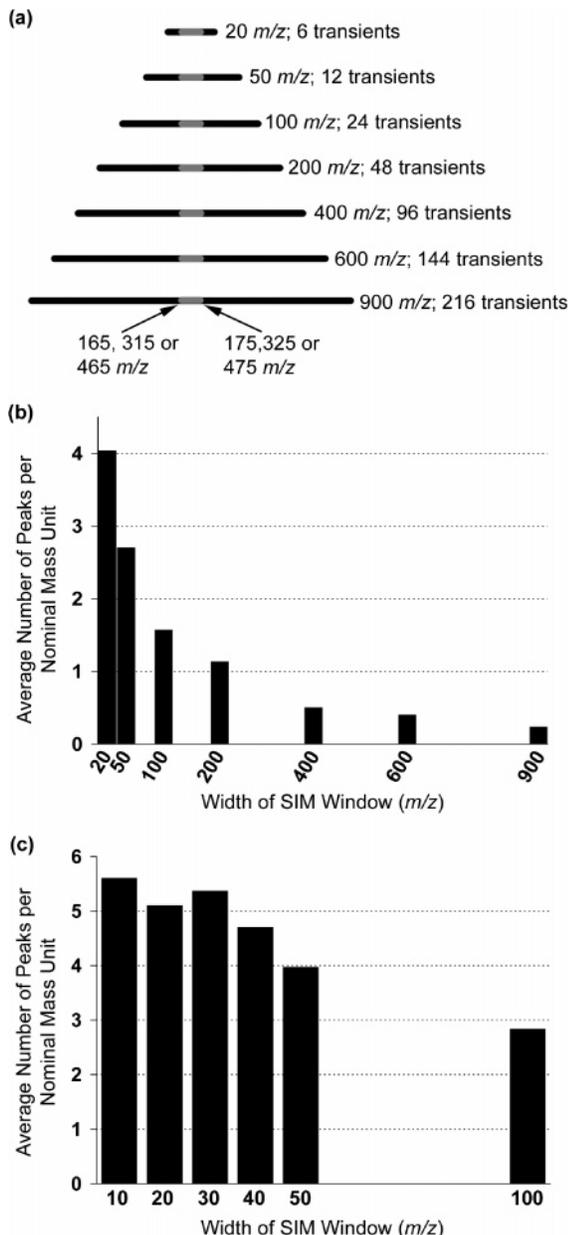
of ions being transferred. For all analyses, the maximum ion trap fill time (1 s) was never reached, thus guaranteeing each AGC target value was achieved. The PEG&AA standard was analyzed using AGC target values of  $1 \times 10^5$ ,  $2 \times 10^5$ ,  $3 \times 10^5$ ,  $5 \times 10^5$ , and  $1 \times 10^6$ . Each analysis comprised SIM scans (in triplicate) between  $m/z$  100–200, 300–400, and 500–600. All spectra were calibrated using RecalOffline with three internal calibrants, and mass errors were calculated on all remaining known peaks (18–24 peaks/replicate, including all three ranges). Increasing the AGC target caused an approximately linear increase in the root-mean-squared (rms) mass error (Figure 1). AGC targets of  $5 \times 10^5$  and  $1 \times 10^6$  resulted in unacceptably high maximum absolute errors (2.7 and 3.9 ppm, respectively), which would make peak identification difficult, and these settings will not be considered further. The three lowest AGC targets had somewhat similar maximum absolute mass errors; however, only the AGC target of  $1 \times 10^5$  yielded a maximum error below 1 ppm. An AGC target of  $1 \times 10^5$  also had the lowest rms error of 0.24 ppm, compared to equivalent values of 0.36 and 0.57 ppm for AGC targets of  $2 \times 10^5$  and  $3 \times 10^5$ , respectively. Closer analysis of the spectrum with an AGC target of  $1 \times 10^5$  revealed 47% of all measured peaks had an absolute error of  $<0.1$  ppm, while 90% were  $<0.3$  ppm. In comparison, the AGC target of  $2 \times 10^5$  had only 26 and 66% of all measured peaks of  $<0.1$  and  $<0.3$  ppm absolute error, respectively.

As the AGC target is decreased, the number of ions entering the ICR cell decreases; space-charge effects therefore also decrease and higher mass accuracy is achieved. Unambiguous peak identification requires high mass accuracy. For a singly charged ion of  $m/z$  500, there are two possible empirical formulas if the mass error is 0.1 ppm; however, this rises dramatically to 21 and 64 possibilities for errors of 1 and 3 ppm, respectively,<sup>17</sup> which highlights the importance of minimizing mass error. Furthermore *maximum* mass error, rather than rms error, should be used to assess mass accuracy since it better represents the capability of a mass spectrometry method for automated metabolite identification.<sup>35</sup> In theory, the AGC target would be set extremely low to achieve near perfect mass accuracy; however, lowering the AGC target reduces sensitivity and so a compromise

must be reached. For the SIM-stitching method, we chose the highest AGC target value that produced a sub-ppm maximum mass error, namely, 100 000 charges (singly charged ions), which also corresponds to Thermo's recommended AGC for a SIM scan.

**Optimization of SIM Window Size.** The second goal was to optimize the SIM window size to maximize the number of metabolites detected with an AGC target of  $1 \times 10^5$ . In the first "coarse" study, liver extract was analyzed using seven increasing window widths from  $m/z$  20 to 900, each sharing a common  $m/z$  10 region (Figure 2a). Considering that our final method requires extensive coverage of the metabolome, the narrower the windows, the more windows that would need to be acquired (and stitched together) to cover a wide mass range. Therefore, narrow windows received proportionally less signal averaging (number of transient scans) than wider ones in order for the total acquisition time for the entire mass range to remain constant (Table 1, Figure 2a). All spectra were acquired in SIM mode except for the  $m/z$  900 window (setting was unavailable so full scan was used). The entire study was conducted at three different  $m/z$  values centered at  $m/z$  170, 320, and 470 (each in triplicate), and for each experiment, peaks were then counted within the central  $m/z$  10 region that was common to all the spectra (Table 1). The average number of peaks detected increased as the window size became more narrow, with  $\sim 17$  times more peaks observed in the  $m/z$  20 versus 900 window even though the wider window was signal averaged for  $\sim 36$  times longer (Figure 2b). A "fine" study was then conducted in which liver extract was analyzed using six increasing window widths from  $m/z$  10 to 100 that shared a common  $m/z$  6 region. As before, the acquisition time (number of transient scans) of each window was proportionally altered. This was conducted at five different  $m/z$  values centered at  $m/z$  200, 250, 300, 350, and 400, and peaks were counted within the central  $m/z$  6 region that was common to all spectra (Table 1). Decreasing the window size from  $m/z$  100 to 30 showed  $\sim 2$ -fold increase in the average (of all five  $m/z$  values scanned) numbers of peaks detected. Window sizes of  $m/z$  30, 20, and 10 all achieved similar high peak counts of 5.4, 5.1, and 5.6 peaks per nominal mass unit, respectively (Figure 2c). These results show that reducing the window size increases the number of peaks detected, until a window of  $m/z$  30 is reached whereby any further size reduction shows no improvement in peak count. These findings can be rationalized by considering that (i)  $\sim 200$  ions of the same metabolite are required to generate a detectable signal in the ICR cell,<sup>11</sup> (ii) a total of only  $1 \times 10^5$  ions are transferred to the ICR cell as determined by the AGC, and (iii) the narrower the SIM window the fewer different metabolites are transferred to the cell and, therefore, the greater the likelihood that a low-abundance metabolite will be present in sufficient ion number to be detected. It appears, however, that the reduction in acquisition time for windows narrower than  $m/z$  30 counteracts any sensitivity benefit that would be gained by a further reduction of window size (i.e., there is insufficient signal averaging to increase the signal-to-noise ratio to detect more peaks). In conclusion, window sizes of  $m/z$  10, 20, and 30 were equally

(35) Herniman, J. M.; Langley, G. J.; Bristow, T. W. T.; O'Connor, G. J. *Am. Soc. Mass Spectrom.* **2005**, *16*, 1100–1108.



**Figure 2.** (a) Experimental design for the coarse study that optimized the SIM window size by counting the number of peaks within common  $m/z$  10 ranges for several different window sizes. Peaks were counted between  $m/z$  165–175, 315–325, and 465–475. Acquisition times are shown as the numbers of transient scans per window. (b) Effect of SIM window size (from  $m/z$  20 to 900) on the average number of peaks detected per nominal mass unit in fish liver extract (coarse study). (c) As for part b except SIM window size was varied from  $m/z$  10 to 100 (fine study).

efficient at detecting high numbers of peaks and would all be considered for our final SIM-stitching method.

**Ion Intensity across the SIM Window.** A phenomenon was observed (that we term an “edge effect”) whereby ion intensities at each end of a SIM window were severely reduced. A  $m/z$  30 SIM window containing two closely spaced peaks was analyzed multiple times, such that the two peaks appeared at different locations throughout the window. It was expected that the intensity ratio of the two peaks would remain constant regardless of their location in the SIM window. However, we discovered that as one of the peaks approached the low-intensity edge region then the

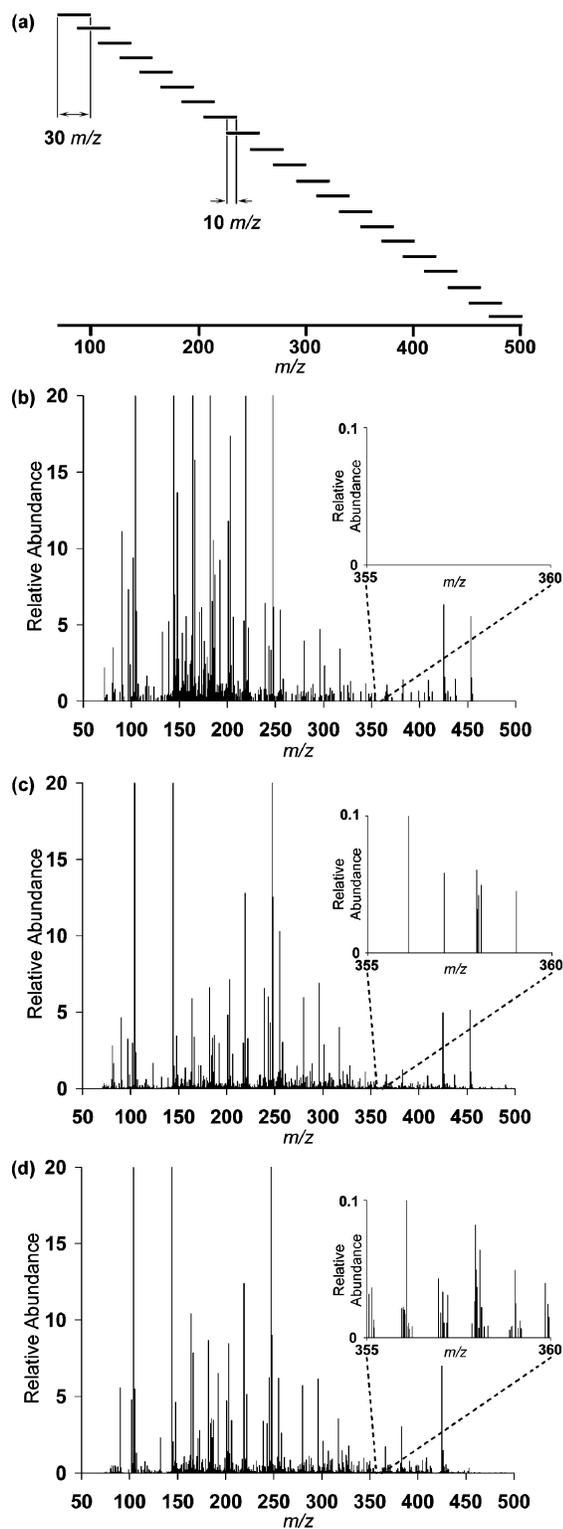
**Table 1. Window Widths and Acquisition Times (Given as Number of 1-s Transient Scans) Used in the Optimization of SIM Window Width Study<sup>a</sup>**

coarse study			fine study		
SIM window width	no. of transient scans acquired	common $m/z$ 10 ranges where peaks were counted	SIM window width	no. of transient scans acquired	common $m/z$ 6 ranges where peaks were counted
20	6	165–175,	10	5	197–203,
50	12	315–325,	20	10	247–253,
100	24	and	30	15	297–303,
200	48	465–475,	40	20	347–353,
400	96	for each	50	25	and
600	144	window	100	50	397–403,
900	216	width			for each window width

<sup>a</sup> Peaks were counted in the common  $m/z$  10 and 6 regions for the coarse and fine studies, respectively.

ratio of the two peaks altered dramatically (Figure S-3 Supporting Information). Edge effects were characterized at four mass ranges (near  $m/z$  84, 175, 350, and 700), which revealed the most dramatic effect at  $m/z$  84 (regions of  $m/z$  4 and 8 were lost at the lower and upper ends, respectively), less of an effect at  $m/z$  175 (regions of  $m/z$  4 and 3 were lost at the lower and upper ends, respectively), less effect again at  $m/z$  350 (region of  $m/z$  5 was lost at the upper end only), and no effect at  $m/z$  700 (Figure S-3). The  $m/z$  20 SIM windows showed signal loss very similar to  $m/z$  30 SIM scans (data not shown). Therefore, for the SIM-stitching method, we chose a  $m/z$  30 SIM window since it has the lowest percentage signal loss. Also, to compensate for this edge effect, adjacent SIM windows were overlapped by  $m/z$  10.

**Comparison of Mass Accuracies between SIM-Stitching and WSR Modes.** The SIM-stitching method was assessed against WSR mode to ensure that high mass accuracy was being achieved. The optimized SIM-stitching method comprised 21 adjacent  $m/z$  30 windows between  $m/z$  70 and 500, each overlapping by  $m/z$  10 to facilitate  $m/z$ -based stitching and to remove deleterious edge effects (Figure 3a). Each SIM window was acquired for 15 s with a single 15-s delay post electrospray initiation, giving a 5.5 min total analysis time; the AGC target was  $1 \times 10^5$ . PEG&AA standard was analyzed between  $m/z$  70 and 500 (in triplicate) by SIM-stitching and by WSR mode using AGC targets of  $1 \times 10^5$  (comparable with SIM-stitching method) and  $5 \times 10^5$  (Thermo recommendation); total acquisition times of 5.5 min were used for all three methods. For the SIM-stitching method, transient data for the 21 SIM windows were acquired, processed, and each internally calibrated using a single calibrant (16 calibrants in total, 5 used twice as they occurred in adjacent windows; Tables 2 and S-1) and then stitched together along the  $m/z$  and intensity axes. For WSR mode, transient data were acquired, processed, and internally calibrated using 10–14 of the 16 calibrants used previously (less calibrants used due to reduced sensitivity of WSR; Tables 2 and S-1). Mass accuracies were calculated using all known noncalibrant peaks (Table 2). Considering the three replicates for each method, SIM-stitching yielded the smallest average rms error, smallest maximum absolute error,



**Figure 3.** (a) Schematic of the optimized SIM-stitching method comprising 21 adjacent 30  $m/z$  SIM windows, each overlapping by  $m/z$  10 and acquired for 15 s, covering a total scan range of  $m/z$  70–500 and with an AGC setting of  $1 \times 10^5$  ions transferred to the ICR detector. Representative wide-scan FT-ICR mass spectra of a fish liver extract obtained using (b) WSR with AGC target of  $1 \times 10^5$ , (c) WSR mode with AGC target of  $5 \times 10^5$ , and (d) optimized SIM-stitching method. All spectra were normalized to the largest peak, and the main figures show a 5-fold zoom along the y-axis. Insets show the mass range between  $m/z$  355 and 360 using a 1000-fold zoom along the y-axis. The acquisition time for each method was 5.5 min to facilitate a direct comparison between the approaches.

and smallest average standard deviation. WSR (AGC target of  $5 \times 10^5$ ) mode showed the largest errors (Table 2, Figure S-4), almost certainly resulting from increased space-charge effects in the ICR cell. The small increase in error in the WSR (AGC target of  $1 \times 10^5$ ) method versus SIM-stitching could be due to the lack of calibrants. However, when considering that WSR (AGC target of  $1 \times 10^5$ ) replicate 3 (using 10 calibrants) achieved mass errors similar to replicate 1 (using 14 calibrants), this suggests that increasing the number of calibrants above 10 has little benefit. Therefore, it is unlikely that addition of two more calibrants (to equal the 16 used in SIM-stitching) would further improve the mass accuracy. This suggests that the algorithm used to calibrate SIM-stitched data could be improving the mass accuracy. Each window comprising the SIM-stitched wide scan was calibrated using unique calibration parameters allowing for precise correction of local  $m/z$  shifts, while in WSR mode just one set of calibration parameters is used for the whole spectrum, likely explaining the slightly higher mass errors of the WSR mode.

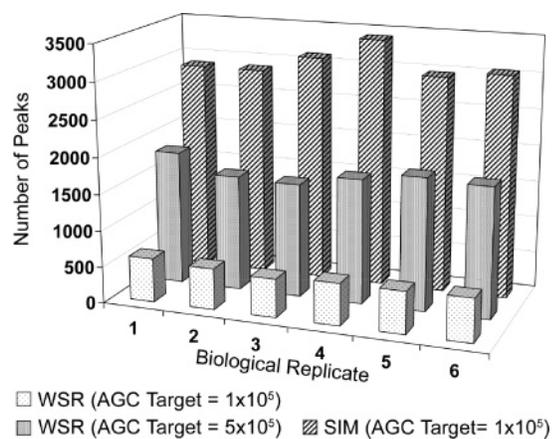
Comparison with the mass errors reported above in the Optimization of Number of Ions Transferred to ICR Cell section revealed the WSR (AGC target of  $1 \times 10^5$ ) data to have errors similar to the earlier data recorded at an AGC target of  $1 \times 10^5$  (demonstrating consistency), while the SIM-stitching method showed improved mass accuracy (likely due to a more efficient calibration method). The WSR (AGC target of  $5 \times 10^5$ ) data had lower errors than the earlier spectra recorded at AGC target of  $5 \times 10^5$ , which probably results from the use of more calibrants (14 compared to 3) and the exclusion of  $m/z$  500–600 (which contains the largest mass errors). In conclusion, SIM-stitching showed high mass accuracy and, in terms of maximum absolute error, was  $\sim 1.3$ -fold more accurate than WSR (AGC target of  $1 \times 10^5$ ) and 4.3-fold more accurate than WSR (AGC target of  $5 \times 10^5$ ). Reducing mass error from 1.0 to 0.1 ppm causes  $\sim 10$ -fold decrease in the number of potential empirical formulas for a peak at  $m/z$  500.<sup>17</sup> Therefore, halving the maximum absolute mass error from 1 (our original target) to 0.48 ppm will prove particularly advantageous for peak identification, which will be strengthened further by isotope information.

**Comparison of Numbers of Peaks Detected and Dynamic Range between SIM-Stitching and WSR Modes.** The final goal was to identify whether the number of detected peaks and the dynamic range could be increased using SIM-stitching compared to WSR. Six liver extracts were analyzed using SIM-stitching (Figure 3a), WSR mode (AGC target of  $1 \times 10^5$ ) and WSR mode (AGC target of  $5 \times 10^5$ ) between  $m/z$  70 and 500 (in triplicate). In each case, the total acquisition time was 5.5 min (including a 15-s data acquisition delay postelectrospray initiation) to facilitate a direct comparison between the methods. The  $m/z$  range of 70–500 corresponds to the highest density of low molecular weight endogenous metabolites. For all methods, transient data were collected and processed as described for the PEG&AA studies above, except that known endogenous metabolites were used as calibrants (Table S-2). Visual inspection of the spectra (normalized to the largest peak) with a 5-fold zoom along the y-axis revealed that WSR (AGC target of  $5 \times 10^5$ ) and SIM-stitching (Figure 3c, d) detected a more dense spread of peaks throughout the entire  $m/z$  70–500 range than WSR (AGC target of  $1 \times 10^5$ ) (Figure 3 b). A further 200-fold zoom within  $m/z$  355–360 (Figure 3b–d

**Table 2. Summary of Mass Errors for SIM-Stitching and WSR Methods (with AGC Targets of  $1 \times 10^5$  and  $5 \times 10^5$ ) Determined from Analysis of the PEG&AA Standard<sup>a</sup>**

scan type	replicate	$N_C$	$N_P$	rms error (ppm)	max absolute error (ppm)	SD of error (ppm)	overall		
							av rms error (ppm)	max absolute error (ppm)	av SD of error (ppm)
WSR ( $1 \times 10^5$ )	1	14	26	0.254	0.617	0.257	0.234	0.617	0.228
	2	14	26	0.199	0.461	0.194			
	3	10	20	0.248	0.613	0.232			
WSR ( $5 \times 10^5$ )	1	14	35	0.612	2.041	0.593	0.565	2.041	0.492
	2	14	35	0.519	0.915	0.441			
	3	14	35	0.564	1.408	0.443			
SIM stitched	1	16	39	0.179	0.423	0.161	0.181	0.475	0.165
	2	16	39	0.193	0.459	0.181			
	3	16	39	0.171	0.475	0.154			

<sup>a</sup> Each method was characterized in triplicate, and the numbers of internal calibrants ( $N_C$ ) and other known peaks used to calculate mass errors ( $N_P$ ) are shown.



**Figure 4.** Comparison of the total number of peaks detected between  $m/z$  70 and 500 for six different fish liver extracts using the optimized SIM-stitching method, WSR mode with AGC target of  $1 \times 10^5$ , and WSR mode with AGC target of  $5 \times 10^5$ .

insets) revealed SIM-stitching to detect significantly more peaks than WSR (AGC target of  $5 \times 10^5$ ), while WSR (AGC target of  $1 \times 10^5$ ) failed to detect any peaks.

For each method, the total number of peaks was counted (between  $m/z$  70 and 500) for each of the six liver extracts (Figure 4), and the dynamic ranges were determined. The average peak counts for the six liver extracts were 3046 for SIM-stitching, 575 for WSR (AGC target of  $1 \times 10^5$ ), and 1719 for WSR (AGC target of  $5 \times 10^5$ ). SIM-stitching detected on average 5.3- and 1.8-fold more peaks than both WSR methods (AGC targets of  $1 \times 10^5$  and  $5 \times 10^5$ , respectively), even though each SIM window received 21 times less signal averaging than the full-scan WSR spectra. Even when compared to an experiment where five times as many ions enter the ICR detector cell (i.e., WSR with AGC target of  $5 \times 10^5$ ), the SIM-stitching method still detected ~44% additional ion species. The relative standard deviation (RSD) of the number of peaks detected across the six fish livers was 7% for SIM-stitching, 5% for WSR (AGC target of  $1 \times 10^5$ ), and 7% for WSR (AGC target of  $5 \times 10^5$ ). This showed that each method was able to generate consistent peak counts when genetically different fish from the same species were analyzed and highlights the reproduc-

ibility of all methods. RSDs of peak intensities were calculated for the same 17 known internal calibrants (Table S-2) for each of the three analysis methods across triplicate analyses of one fish (only 17 internal calibrants were common to all three analysis methods due to reduced sensitivity of the WSR methods). The mean and maximum RSD values were 8.1 and 16.5% for the SIM-stitching method, 11.4 and 23.2% for WSR (AGC target of  $1 \times 10^5$ ), and 11.3 and 17.9% for WSR (AGC target of  $5 \times 10^5$ ), which further emphasizes the reproducibility of all three methods but, more importantly, highlights another advantage of the SIM-stitching method. Specifically, that by acquiring the metabolic data using narrow SIM windows the reproducibility of the intensity profiles is increased relative to wide-scan methods. The mean dynamic range of 16 061 for the SIM-stitched data was ~22-fold greater than for WSR with an AGC target of  $1 \times 10^5$  (mean dynamic range of 726) and ~4.3-fold greater than for WSR with an AGC target of  $5 \times 10^5$  (mean dynamic range of 3684), which mirrored the results from the peak count study. This increase in dynamic range is clearly visible in Figure 3b–d.

When acquiring data as a single large window (e.g.,  $m/z$  70–500), which for a complex biological mixture will comprise a large number of different ion species, a considerable percentage of the AGC target value will be occupied by the most highly abundant ions and there is less chance that low-abundance species will achieve the ~200 ion detection threshold.<sup>11</sup> However, when analyzing a  $m/z$  30 SIM window, only ion species within the specified range are transferred to the ICR detector and all other ions are excluded by isolation waveforms. This greatly reduces the number of individual ion species, meaning that the AGC target value is no longer dominated to the same extent by highly abundant ions, which allows low-abundance species to reach the detection threshold. In effect, SIM-stitching allows more ions per nominal mass unit to transfer into the ICR detector; e.g., AGC target of  $1 \times 10^5$  when scanning a  $m/z$  30 window is ~3334 ions per nominal mass unit, whereas the higher AGC target of  $5 \times 10^5$  when scanning  $m/z$  70–500 is only ~1163 ions per nominal mass unit. Therefore, compared to WSR (AGC target of  $5 \times 10^5$ ), SIM-stitching increases the number of ions scanned per nominal

mass unit by  $\sim 3$ -fold, yielding greater sensitivity and consequently increased peak detection (1.8-fold higher) and dynamic range. The increase is significantly larger when compared to WSR (AGC target of  $1 \times 10^5$ ), where SIM-stitching increases the number of ions scanned per nominal mass unit by  $\sim 15$ -fold, yielding significantly increased peak detection (5.3-fold higher). The reduced signal averaging that is inherent to the SIM-stitching method apparently does not offset this sensitivity increase too significantly.

## CONCLUSIONS

The results demonstrate that when analyzing a wide mass range ( $m/z$  70–500) by SIM-stitched DI nESI FT-ICR MS, peak detection can be increased on average 5.3- and 1.8-fold and maximum mass error decreased by 1.3- and 4.3-fold compared to commercial wide-scan range mode (with AGC targets of  $1 \times 10^5$  and  $5 \times 10^5$ , respectively, the latter being the recommended setting). Using SIM scan mode, detection sensitivity was enhanced by collecting the data as a series of narrow, overlapping windows of width  $m/z$  30. The increased detection sensitivity allowed ion transmission to the ICR cell to be reduced to  $1 \times 10^5$  ions, thus reducing space-charge effects and allowing high mass accuracy. Mass accuracy was enhanced further by the calibration method whereby each of the 21 overlapping SIM windows was individually, internally calibrated prior to stitching, which allowed correction of local  $m/z$  shifts. The maximum absolute mass error was shown to be 0.48 ppm. When compared to conventional DI ESI MS methods, the SIM-stitching method can provide a significant increase in the number of metabolites detected, thereby increasing coverage of the metabolome. In addition, it maximizes the major strengths of FT-ICR MS, those of high mass accuracy and

resolution, thereby facilitating improved peak identification via the calculation of empirical formulas. The coupling of a NanoMate to the LTQ FT enables a fully automated analysis in  $\sim 5.5$  min/sample, which is conducive for high-throughput analyses. This SIM-stitching approach has wide applicability to the measurement of any complex chemical mixture by FT-ICR MS. In particular, for biological samples, this new method could benefit both metabolomics and proteomics.

## ACKNOWLEDGMENT

A.D.S. thanks the BBSRC and Cefas for a Ph.D. CASE studentship (under Cefas Seedcorn contract DP195), T.G.P. and T.N.A. thank the EPSRC for a Ph.D. studentship and DTA support, H.J.C. thanks the Wellcome Trust (074131), and M.R.V. thanks the NERC for an Advanced Fellowship (NER/J/S/2002/00618). This work was partly supported by NERC grant NE/D002508/1. The authors gratefully acknowledge technical support from Mark Baumert and Mark Allen (Advion Biosciences) and Bernard Delanghe (Thermo Electron Corporation). We also thank Grant Stentiford, Brett Lyons and Steve Feist (Cefas) for the fish liver samples. A.D.S. and T.G.P. contributed equally as first authors to this work.

## SUPPORTING INFORMATION AVAILABLE

Descriptions of transient preprocessing, calibration, and SIM-stitching algorithms. This material is available free of charge via the Internet at <http://pubs.acs.org>.

Received for review December 27, 2006. Accepted March 19, 2007.

AC062446P