ELECTRONIC SUPPLEMENTAL INFORMATION

In situ monitoring of enzyme-catalyzed (co)polymerizations by Raman spectroscopy

Matthew T. Hunley,[1] Atul S. Bhangale,[1,2] Santanu Kundu,[1] Peter M. Johnson,[1] Michael S. Waters,[1] Richard A. Gross,*[2] and Kathryn L. Beers*[1]

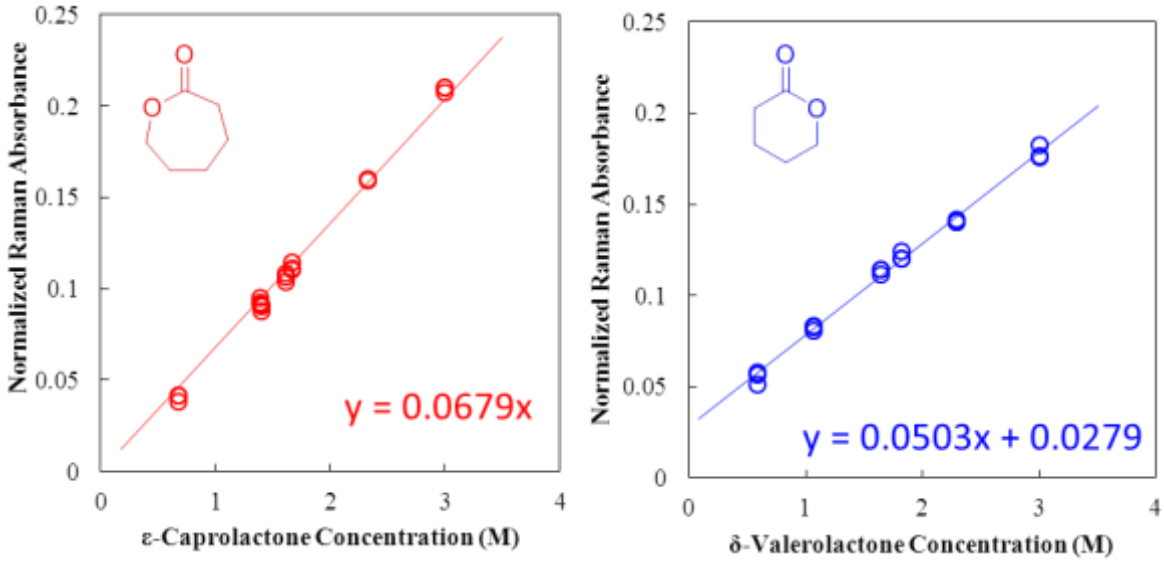

Figure S1. Calibration curves for determining the concentration of ε-caprolactone (ε-CL) and δ-valerolactone (δ-VL) by Raman. Measurements were performed on at least three different samples at each concentration, and all data points are shown.

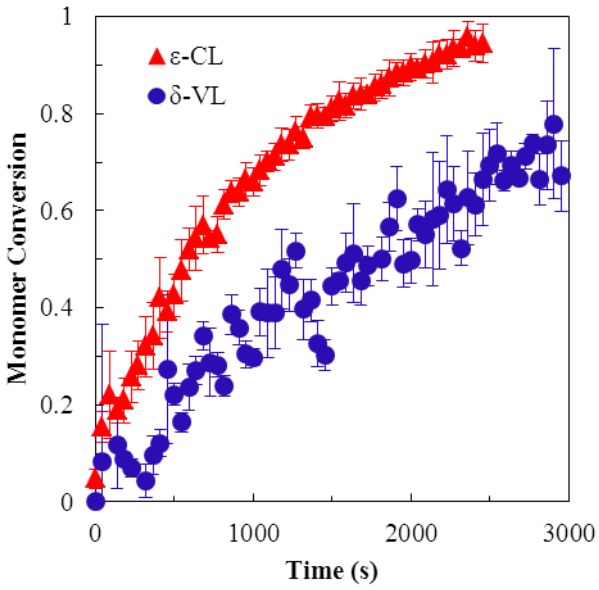

Figure S2. Fractional monomer conversion as a function of reaction time for the homopolymerizations of ε-CL and δ-VL. The error bars indicate one standard uncertainty based on measurements on three different samples.

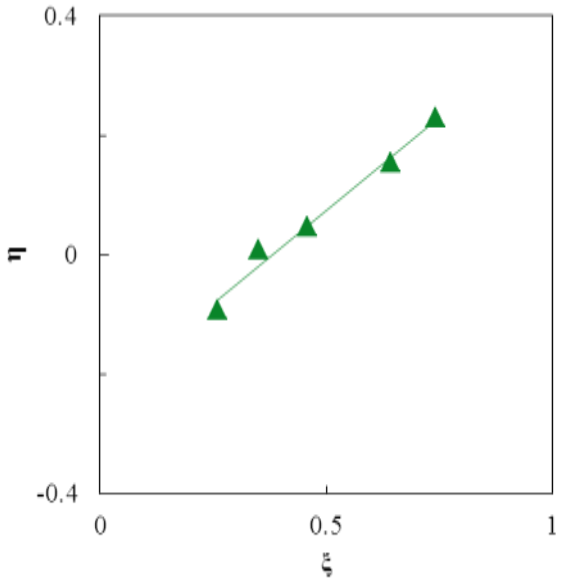

Figure S3. Kelen-Tudos plot to determine monomer reactivity ratios for the enzyme-catalyzed copolymerization of ε-CL (M_1) and δ-VL (M_2). The solid line represents the linear regression best fit to the data, $\eta = [r_{\varepsilon\text{-}CL}+r_{\delta\text{-}VL}/\alpha]\xi - r_{\delta\text{-}VL}/\alpha$ where $r_{\varepsilon\text{-}CL} = 0.38 \pm 0.06$ and $r_{\delta\text{-}VL} = 0.29 \pm 0.04$.

Experimental

Materials. Toluene, ε-caprolactone (97%), and δ-valerolactone (technical grade) were obtained from Sigma-Aldrich, distilled over CaH_2, and preserved over 0.4 nm activated molecular sieves prior to use. Novozyme 435 (N435) was obtained from Novozymes (Bagsvaerd, Denmark). To minimize deviations in experimental results due to variations in particle size of N435 beads, a 400 μm cutoff sieve was used to screen a narrow particle size distribution of 400 μm ± 50 μm. Sieved N435 beads were preserved over a desiccant in a vacuum dessicator and were used for all experiments.

Polymerizations. Reactions were performed in a 5 mL round bottom flask under constant magnetic stirring at 60 rad/s. The mass ratio of CALB to monomer was maintained at 1:100 for all reactions in both aliquot and *in situ* experiments. In a sample reaction, 100 mg N435 beads was added to the reaction flask under a blanket of argon. Toluene (2 mL) was transferred into the reaction flask via syringe under argon atmosphere. After the reaction flask was heated to the desired reaction temperature, ε-CL (1 mL) was added to the reaction via syringe under argon atmosphere. The Raman set up was aligned using the fine and coarse tunings of the adjustable moving stage. Control experiments were also performed using beads without enzyme (Lewatit VP OC 1600). Each reaction was performed in triplicate with Raman

spectra or NMR samples collected at specific time intervals. The reported conversions represent the average measured value and standard deviation at each time interval.

NMR Spectroscopy. ^{1}H NMR was recorded on a JEOL 270 MHz spectrometer. Monomer conversion was calculated as the ratio of integrated peak areas corresponding to the methylene unit adjacent the ester oxygen in the monomer (at 4.2 ppm for ε-CL and 4.3 ppm for δ-VL) and the polymer (at 4.0 ppm for both monomers, end group resonance at 3.6 ppm). The ratio of ε-CL to δ-VL in the polymer was determined by comparing the ratio of integrated peak areas corresponding to the methylene unit adjacent to the ester oxygen (at 4.0 ppm, end group resonance at 3.6 ppm) and the methylene unit of the γ-carbon of the polymerized ε-CL (at 1.4 ppm).

***In situ* Raman Spectrometry.** A Raman Systems R3000HR Raman spectrometer equipped with a laser with excitation wavelength of 785 nm was used to collect spectral data. Raman spectra were collected sequentially, with each spectra averaged over 20 s and intervals between measurements of 90 s. For the batch reactor, a 38-mm aluminum cube was designed to hold a 5 mL round bottom flask and the external Raman probe. The metal block was placed on a heating plate to uniformly heat the reaction system. A thermocouple connected to a feedback controller was inserted into the reaction flask to maintain the temperature within ± 1 °C. The fiber optic Raman probe with a focal length of 5 mm was positioned to collect spectra inside the reaction vessel through the glass wall.

To obtain quantitative monomer concentrations by Raman, calibration curves were constructed for both ε-CL and δ-VL. The anti-symmetric ring-stretching peaks of ε-CL and δ-VL at 696 and 745 cm^{-1}, respectively, disappear during polymerization. The peak at 1003 cm^{-1}, corresponding to ring breathing of toluene, was taken as the internal standard. Two stock solutions of ε-CL and PCL in toluene were prepared with a concentration of 0.4 g/mL (equal to 3 M ε-CL). These stock solutions were mixed in various proportions to create a sufficient number of data points through a full range of monomer concentrations. Raman spectra for the mixtures of ε-CL and PCL were recorded and are displayed in **Figure S4**. By determining the ratio of peak areas at 696 cm^{-1} to that at 1003 cm^{-1} for different ε-CL concentrations, after deconvolution as described below, a calibration plot was generated **Figure S1a**. The calibration plot was constructed similarly for δ-VL, as shown in **Figures S1b and S5**.

These calibration plots were used to determine monomer conversion as a function of polymerization time.

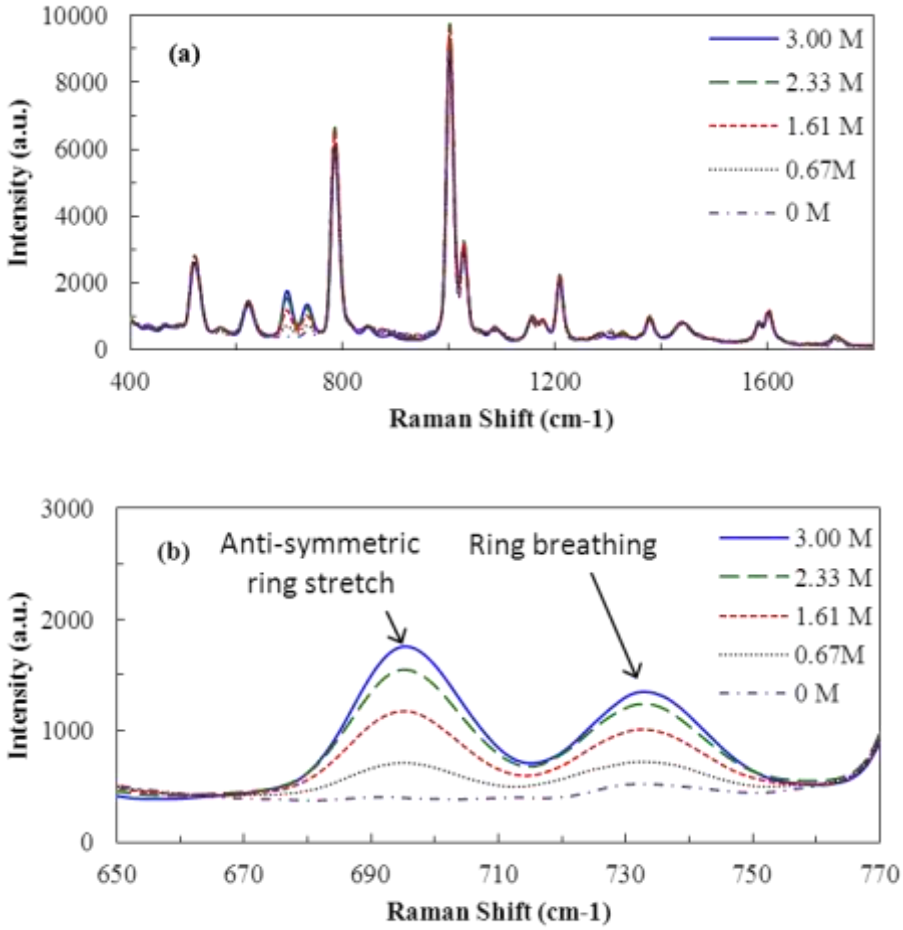

Figure S4. Raman spectra for mixtures of ε-CL and PCL for calibration. (a) Raman spectra over the wavenumber region 400 cm^{-1} to 1800 cm^{-1}. (b) Expansion of the wavenumber region 650 cm^{-1} to 770 cm^{-1}. The legend lists the molar concentrations of ε-CL in each sample.

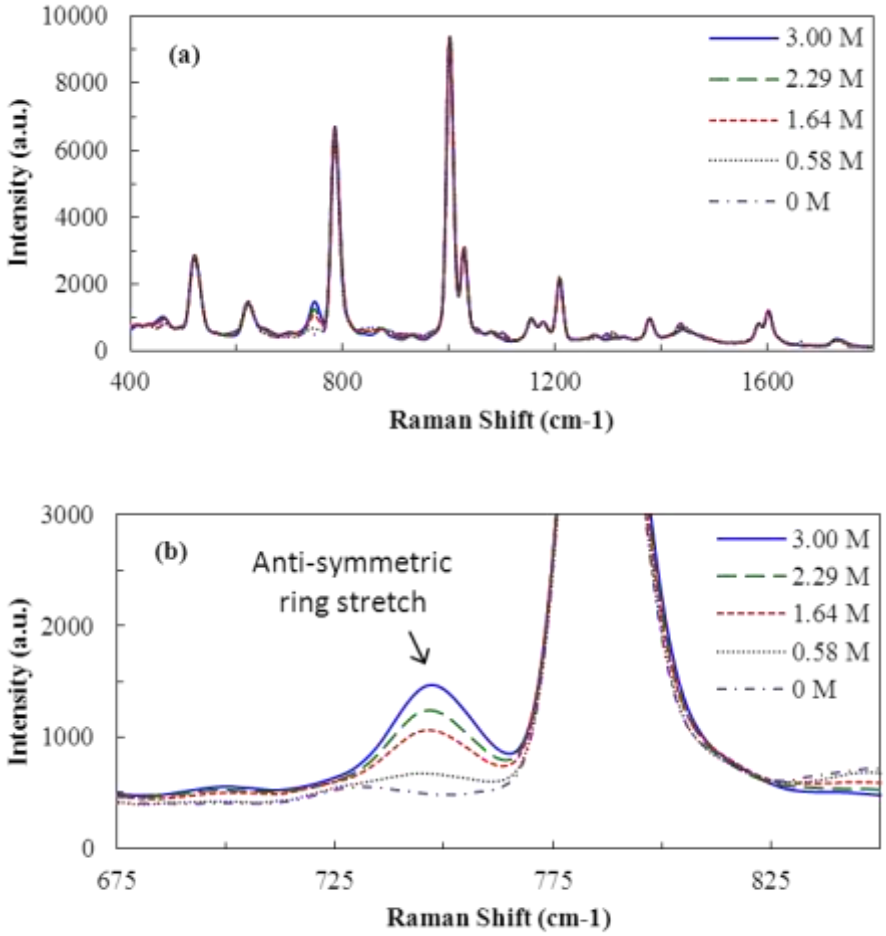

Figure S5. Raman spectra for mixtures of δ-VL and PVL for calibration. (a) Raman spectra over the wavenumber region 400 cm^{-1} to 1800 cm^{-1}. (b) Expansion of the wavenumber region 675 cm^{-1} to 850 cm^{-1}. The legend lists the molar concentrations of δ-VL in each sample.

Observation of **Figures S4 and S5** shows that the ε-CL peak at 696 cm^{-1}, the δ-VL peak at 745 cm^{-1}, and the internal reference peak at 1003 cm^{-1} are each partially overlapped with at least one additional absorbance peak. Hence, it was necessary to deconvolute these absorbance bands. Due to the number of collected spectra per experiment, a MATLAB script was developed to fit Gaussian peaks to the Raman spectra automatically and report the peak area for each underlying peak. The program first corrected each spectrum with a two-point linear baseline and then calculated a residual error profile between the baseline-adjusted spectra and the predicted spectra from two Gaussian peaks. This residual error was then minimized through an unconstrained nonlinear optimization of the Gaussian function parameters. Due to the separation distance and magnitude changes between the confounding peaks, initial starting conditions could

quickly find the global minima. The program also compensated for the shift in background absorbance during polymerization by adjusting the baseline points to the lowest intensity within 10 cm^{-1} of the initial value. This change reduced the residual error at high conversion, since without this correction negative intensities could appear and cause smaller than expected Gaussian peak areas. A residual area error of less than 1 % was typical for the combined area of both peaks. For regions not confounded by the second peak, the error between the spectra and Gaussian fit was slightly higher but still less than 3% of the peak area. When calculating ε-CL peak area, initial baseline points were 660 cm^{-1} and 760 cm^{-1}, and the spectral range was set to the same values. For the toluene calibration peak (internal standard), baseline points were set at 945 cm^{-1} and 1220 cm^{-1}, and the fitted spectral range was 965 cm^{-1} to 1060 cm^{-1}. The δ-VL peak at 745 cm^{-1} overlapped a weak absorbance peak of toluene which was difficult to remove by deconvolution. However, the calibration curve accounts for the total integrated area of both peaks.

Gaussian Fitting Protocol. Both the peaks at 696 cm^{-1} and 1003 cm^{-1} had a second peak nearby which needed to be deconvoluted for the correct analysis of the ε-CL peak areas in the Raman spectra. Since only two peaks were convoluted for the ε-CL polymerization, separation of the peaks did not require the advanced deconvolution techniques that are needed for isolating multiple peaks from one single broad peak. A Gaussian peak fitting program was written to fit parameters based on unconstrained nonlinear optimization where the number of peaks was defined prior to the optimization process. The program found Gaussian peaks in the underlying spectra, as defined by the equation:

$$f = a \cdot \exp\left(-\frac{(x-\mu)^2}{2\sigma} \right) \quad (1)$$

where a is the peak height, μ is the center of the peak, and σ is related to the full width at half maximum of the peak. With these three parameters, the peak area can be calculated. For two peaks, six parameters must be fit to deconvolute the spectra successfully. Since spectra peaks and underlying background were altered due to intensity variations and composition changes as the polymerization proceeds, spectra were analyzed individually.

The program used for this work was developed in MATLAB, and the process is described in pseudocode below for adaption to other programming languages and spectra formats. The program requires manual input for four different parameters: the fitted spectral

range, spectra file location, peak locations, and baseline parameters. Only the first two parameters remain fixed during the program, as the latter two were optimized to account for the polymerization process. The manual input values are reported at the end of the program and can be loaded into the program to analyze different sets of data with equivalent initial parameters. For the Raman spectra program used in this work, each spectrum was in a separate file that were loaded sequentially, starting with the initial time point and iterating through every spectra file.

Using these input parameters and the loaded spectra, two steps were performed before the optimization began. First, the baseline points were checked to ensure that the defined wavenumbers were at the lowest point possible within 10 cm^{-1} of the initial input values. This was done to ensure that the baseline correction did not produce baseline corrected spectra with negative intensity. The second process formed an initial guess of (a, μ, σ) for each Gaussian peak. Both the peak height and center of the peak could be defined from the maximum point near the initial input values, and σ could be estimated from shape of the peak at 75% max height, where the peaks were minimally convoluted. With the correct baseline and defined initial Gaussian parameters, the unconstrained non-linear optimization could proceed. This process is not well suited to systems with a large number of peaks or highly convoluted peaks, which would require strict bounds on Gaussian parameters to prevent peak disappearance and large parameter error. These optimized parameters were used to calculate peak area, residual error and residual area for the current spectra. The calculated areas and error, along with the optimized (a, μ, σ) for each Gaussian peak, were reported at the end of the program for each spectra.

For the next spectra in the series, the process is similar except that the initial guess for each Gaussian peak can be weighted with the fitted parameters from the previous spectra. The optimized parameters for μ and σ did not vary significantly from the previous spectra to the current spectra, so these values were blended with the initial guess from the current spectra for an improved initial guess. This weighting did not affect the optimized result for the spectra shown in this work, but did slightly improve the processing time. For a set of 30 spectra, the overall program would take approximately one minute to perform baseline corrections, fit each spectrum for Gaussian peaks, and report peak areas.

Psuedocode

```
K = the number of spectra files
GP = Gaussian fit parameters (a,μ,σ)
GFit = optimized Gaussian parameters (a,μ,σ)
Peaks = peak locations
Intensity = spectra and associated wavenubmers
read BaselineRange, SpectraRange, Peaks (all in wavenumbers)
from I = 1 to K
      load Intensity from spectra file I
      check and move BaselineRange to minimum Intensity within ± 10
      truncate Intensity to BaselineRange
      BaseSpec = linear baseline subtraction of Intensity
      compute GP(a) from BaseSpec at Peaks
      compute GP(μ) from Peaks values
      compute GP(σ) from width at half max (at GP(a)/2)
      if I > 1
            GP(σ) = GFit(σ) from I-1
            GP(a,μ) = Weight*GP(a,μ) * + (1-Weight)* GFit (a,μ) from I-1
      end
      GFit = optimization using GP as initial guess
            Minimize sum square of Error over SpectraRange
            Error = BaseSpec minus Gaussian peaks calculation from GP
      end
      compute PeakAreas from GFit
      compute ResidualError
end
print GFit, PeakAreas, ResidualError for I = 1 to K
```

Reactivity Ratios. Reactivity ratios were determined using the Fineman-Ross linearization technique and the equation

$$G = r_1 F - r_2, \tag{2}$$

where $G = X(Y\text{-}1)/Y$, $F = X^2/Y$, $X = [M_1]/[M_2]$, and $Y = d[M_1]/d[M_2]$. For each reaction composition, the initial monomer consumption ratio ($d[M_1]/d[M_2]$) was determined from the initial slope (slope of the data at the highest concentrations, corresponding to the start of the polymerization) of the plot of [ε-CL] versus [δ-VL], as shown in **Figure S3**. The values of G

and *F* are plotted in **Figure S4** for each reaction composition, and a linear regression line was used to determine r_1 and r_2. To accurately represent the error structure of the calculated reactivity ratios, the 95% unbiased joint confidence interval (JCI) ellipse was calculated for each set of experiments. The joint confidence region is expressed by the following inequality,

$$SSR(r_1, r_2) \leq SSR(\hat{r}_1, \hat{r}_2)\left[1 + \frac{p}{n-p} F\left(p, n-p, 1-\alpha\right)\right], \quad (3)$$

where $SSR(r_1,r_2)$ is the sum of the squares of the residuals as defined below, $\hat{r}_1$ and $\hat{r}_2$ are the set of reactivity ratios that minimize the sum of squares of the residuals (determined by linear regression), *p* is the number of fitting parameters, *n* is the number of experimental data points, α is used to choose the confidence interval (α = 0.05 for the 95% JCI), and $F(p, n\text{-}p, 1\text{-}\alpha)$ is the F-distribution level at the 1-α confidence level. To calculate the 95% JCI, a short MATLAB algorithm was written to calculate the *SSR* over a wide range of values for r_1 and r_2, where *SSR* is defined as

$$SSR(r_1, r_2) = \sum_{i=1}^{n} (G_i - G(F_i, r_1, r_2))^2 , \quad (4)$$

where *n* is the number of experimental points, G_i is the experimentally determined value for a given value of F_i, and $G(F,r_1,r_2)$ is the value of *G* for a given value of *F*, r_1, and r_2. The boundary values to the region defined by **Equation 4** were plotted to illustrate the JCI.