



ELSEVIER

Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

Control Engineering Practice ■ (■■■■) ■■■-■■■

CONTROL ENGINEERING  
PRACTICE[www.elsevier.com/locate/conengprac](http://www.elsevier.com/locate/conengprac)

# Factors affecting on-line estimation of diastereomer composition using Raman spectroscopy

Sze-Wing Wong<sup>a</sup>, Christos Georgakis<sup>a,\*</sup>, Gregory Botsaris<sup>a</sup>,  
Kostas Saranteas<sup>b</sup>, Roger Bakale<sup>b</sup>

<sup>a</sup>System Research Institute for Chemical & Biological Processes and Department of Chemical & Biological Engineering, Tufts University, Medford, MA, USA

<sup>b</sup>Chemical Process Research and Development, Sepracor Inc., Marlborough, MA, USA

Received 21 July 2006; accepted 2 November 2006

## Abstract

This paper addresses the estimation of fractional solid composition of two diastereomers during crystallization. The estimation is obtained through a Partial Least Square (PLS) model that utilizes on-line Raman spectroscopy and additional process information such as temperature and slurry density. Several PLS models are developed that incorporate conditions that either neglect or account for variability in the additional process variables. It is shown that the model that incorporates both temperature and slurry density is the most accurate.

© 2006 Elsevier Ltd. All rights reserved.

**Keywords:** Calibration; Estimation; Prediction method; Regression analysis; Spectra correlation

## 1. Introduction

The manufacturing of pharmaceuticals often involves separation of enantiomers, which are chiral molecules that are mirror images of each other. Since the physical properties of both enantiomers ( $R$  = right handed and  $S$  = left handed) are the same, a traditional separation method such as crystallization by seeding is feasible but it becomes very sensitive to the experimental conditions (Elsner, Menendez, Muslera, & Seidel-Morgenstern, 2005; Qian & Botsaris, 1997). Thus, the pharmaceutical industry usually relies on the achiral synthesis of the enantiopure product or the reaction between the enantiomer with a chiral acid/base to form different diastereomers. Diastereomers are molecules that have two or more chiral centres and are not mirror images of each other. The resulting diastereomers have different physical properties such as solubility and often crystallize into different crystal structures. With these differences, crystallization can effectively separate the

desired product with very high purity (Jacques, Collet, & Wilen, 1981; Wilen, Collet, & Jacques, 1977).

To meet government regulations, the purity of the final product is of crucial importance. Thus, the capability of on-line monitoring of the optical purity of the crystals present at a given time will help develop a robust crystallization process. The traditional methods used to determine the diastereomer composition in both solid and/or liquid phase are by chiral HPLC method or with the use of polarimeter to determine the chirality in solution phase. However, both of the analytical methods cannot be conducted on-line and required sample preparation. On-line Raman spectroscopy is suitable for this application since Raman can detect the lattice vibrations corresponding to the translational and rotational motion of the entire molecule within the lattice structure of the crystal (Ferraro, 1971). As a result, Raman spectroscopy is capable of differentiating similar molecules with different crystal lattice structures. To the best of the authors' ability, there was no previous work with Raman differentiating the composition of the diastereomer in crystallization. However, several authors have demonstrated the ability of monitoring the changing compositions of two different

\*Corresponding author. Tel.: +1 617 627 2573; fax: +1 617 627 3991.  
E-mail address: [Christos.Georgakis@tufts.edu](mailto:Christos.Georgakis@tufts.edu) (C. Georgakis).

crystals on-line during a solvent-mediated polymorphic transformation with Raman (Berglund, Wang, Wachter, & Antosz, 2000; Hu, Liang, Myerson, & Taylor, 2005; Ono, Horst, & Jansens, 2004; O'Sullivan, Barrett, Hsiao, Carr, & Glennon, 2003; Scholl, Bonalumi, Vicum, Mazzotti, & Muller, 2006). In addition, chemometric techniques can be applied to the Raman spectra to differentiate slight peak shifts and to remove noise from the signals off-line (Falcon & Berglund, 2004; Rades, Pratiwi, Fawcett, & Gordon, 2002; Starbuck et al., 2002).

The collection of data through an immersion probe and its connection to the detector via a fiber optic cable, allows analysis of solid phase composition in real-time. The returned Raman signal depends on the amount of inelastic scattering from the solids detected by the analyzer within the detection zone. As a result, the relative Raman intensity corresponding to the diastereomers in the slurry will be impacted by a number of other solid-state factors besides the primary one of particle characteristics. Several authors have hypothesized that the intensity of the returned Raman signal during crystallization of different polymorphs may be a function of particle size and shape (O'Sullivan et al., 2003; Zhou, Wang, Ge, & Sun, 2002). This is based on the assumption that the returned Raman signal primarily comes from a volume close to the surface of the crystal. Additionally, slurry density defined as the total mass of crystals per unit volume (mg/mL), may be another solid-state factor since the number of crystals inside the detection zone will influence the intensity of the returned Raman signal. Furthermore, the Raman spectrum will be affected by the relative amount of solvent and solids in its path. Consequently, slurry density should impact the returned Raman signal intensity. A recent study showed that Raman spectra along with multivariate analysis were capable in predicting both the slurry density and percent composition of an anhydrous/monohydrate system (Caillet, Puel, & Fevotte, 2006). However, the authors did not investigate whether temperature has any effects, as all the experiments were run isothermally.

For a batch cooling crystallization process, temperature changes constantly and affects slurry density directly since the solubility of the solute is a function of temperature. In addition, temperature can alter Raman band areas by changing the sample density and therefore the number of analyte molecules in the sampled volume (Pelletier, 2003). Hence, in the present work, we examine whether the information provided by Raman spectroscopy is sufficient or whether it needs to be complemented by additional process measurements in order to provide an accurate estimation, through a Partial Least Square (PLS) model, of the solid composition of one of the two diastereomers involved in the production of an active pharmaceutical ingredient, denoted here as compound *A*. A previous study showed that the incorporation of temperature into a single chemometrics calibration improves model accuracy in predicting solute concentration with the use of IR (Togkalidou, Fujiwara, Patel, & Braatz, 2001). In this

study, temperature, slurry density, and the entire Raman spectrum were used in the modeling task. PLS regression was used to quantify the composition of the diastereomeric mixture. Four different models are examined and compared for their prediction/estimation accuracy.

## 2. Experimental section

### 2.1. Materials

HPLC grade solvents were used as received from commercial suppliers without further purification. The starting materials (racemic mixture of compound *A* and a pure enantiomer of a chiral acid, denoted here as *D*) that met the specifications were used as received from qualified suppliers without further purification.

### 2.2. Preparation of pure *S-D* diastereomer

A racemic mixture of compound *A* was reacted with a pure enantiomer of a chiral acid (denoted here as *D*) in a solvent mixture of well defined composition and the solution was heated and held at 5° above the saturation point to ensure complete dissolution. The solution was then slowly cooled and seeded with 2% by weight of the *S-D* diastereomer (desired product) at specific seeding temperature. The seeded slurry was cooled to a target solid recovery temperature. It was then followed by a filtration wash and drying step. The end product was analyzed by a chiral HPLC method resulting in optical purity of at least 97% of the *S-D* diastereomer.

### 2.3. Preparation of pure *R-D* diastereomer

The preparation of pure *R-D* diastereomer needed, for the calibration process, first involved the purification of the *R* enantiomer from the racemic mixture of compound *A*, followed by reaction with *D* to form the *R-D* diastereomer. Since the crystallization kinetic of the *R-D* diastereomer is slow compared with the *S-D* diastereomer, the purification step of the *R* enantiomers would ensure the optical purity of the *R-D* diastereomer. For the purification step, racemic free base of compound *A* was reacted with a pure enantiomer of another chiral acid, denoted as *L*, in solvent and the solution was heated and held at 5° above the saturated temperature to allow for complete dissolution. The solution was then slowly cooled and seeded with 2% by weight of the *R-L* diastereomer at a specified seeding temperature. The seeded slurry was cooled to a target isolation temperature and then followed by a filtration and drying step. The dry *R-L* crystals would then go through a free basing step to obtain the pure *R* enantiomer. Finally, the pure *R* enantiomer reacted with *D* and crystallized to form *R-D* diastereomer. The product was analyzed by a chiral HPLC method, resulting in optical purity greater than 98%.

#### 2.4. Raman spectroscopy

A RamanRxn1™ analyzer (Kaiser Optical System, Inc.) coupled with an immersion fiber optic probe was used for the in situ measurements. Raman spectra were recorded using NIR excitation radiation at 785 nm and the spectroscopy incorporates the TE-cooled CCD detector technology. All collected spectra were averaged over five accumulations collected over 8 s each. It should be noted that the collection time for all five spectra is about 1 min.

#### 2.5. Calibration experiments

In order to obtain spectra of a known amount of solids' concentration and diastereomeric composition, the solution was first pre-saturated with respect to both diastereomers at specified temperatures (Table 1). The saturated solution was prepared by adding excess amounts of compounds A (50% of *S* and 50% of *R*) and *D* in the solvent system and heated until dissolution. After nucleation occurred upon cooling, the slurry was under constant stirring for 4 h to ensure it reached equilibrium with both diastereomers at the specified temperature and finished by a filtration step. The saturated solution was kept in a jacketed round-bottom flask to maintain constant temperature.

Spectra of the standards were obtained for isothermal mixture. The spectra were collected at different temperatures (from 0 to 40 °C) and with different slurry densities (from 13.3 to 80 g/L) that reflected the range of the crystallization process conditions. A total of 65 standards were used with varying conditions (Table 1) and were divided into two groups—training and testing groups. A total of 55 standards were selected and used to construct the model while the remaining 10 standards were used to test the accuracy of the model. The 10 standards from the testing group were randomly selected to cover the whole experimental space. The Raman probe was inserted top-down into the 15 mL vial and all the spectra were collected

under constant stirring with magnetic stir bar to suspend the slurry.

#### 2.6. Dynamic crystallization experiment

With the use of a 250 mL Jacketed round-bottom flask, 5 g of compound A (racemic mixture) was reacted with 1.7 g of *D* in 238 mL of solvent mixture. The solution was kept under constant stirring with a magnetic stir bar, which is the same as the calibration experiment, and the jacket was connected to a temperature-controlled chiller. The Raman probe was inserted at a 45° angle into the reaction flask to ensure a representative sampling of the slurry and spectra were collected at 1-min intervals. The reason for placing the probe at a 45° angle instead of top-down as in the calibration experiment is because a condenser is needed to reduce solvent evaporation during the crystallization experiment.

The solution was first heated to 40 °C and held for 20 min to allow complete dissolution. It was then cooled to 0 °C at a rate of 5 °C/min and the temperature remained constant at 0 °C for 4 h. Since the crystallization kinetic of the *R-D* diastereomer is much slower than the *S-D*, a rapid cooling rate was applied in attempt to crash out both diastereomers simultaneously. After nucleation occurred, samples of the slurry were drawn for HPLC analysis. A total of six samples were drawn throughout the experiment.

The collected samples were first filtered and the mother liquor was sent for HPLC analysis for solute concentration and percent composition of the diastereomers in the solution phase. The wet cake did not go through solvent wash to avoid dissolution of the crystals during the wash. The cake was then weighed and vacuum dried at 40 °C overnight. The dry cake was again weighed and the solid contents sent for HPLC analysis for percent composition of the diastereomers in solid phase. Since the evaporated solvent contained residual of both diastereomers in the solution phase, the amount of solute in the evaporated solvent was calculated with the weight difference of the cakes and multiplied by the concentration of solute in solution from HPLC analysis. The relative composition of the evaporated residual diastereomers from the solvent was obtained from the chiral HPLC result in the mother liquor samples. The percent composition of the *S-D* diastereomer was then corrected with the residual solute from the solution phase.

#### 2.7. Sampling error

Since the HPLC analysis for the percent composition of the diastereomers in solid phase was conducted after drying, the sampling error was primarily due to inhomogeneous mixing of the dry cake. The estimation of the sampling error was conducted by mixing a known amount of both diastereomers in a pre-saturated solution and followed the same filtration/drying step as described above. Six different samples were drawn from different locations

Table 1  
Experimental condition for standards

Fixed variable (°C)	Fixed variable (g/L)	Changing variable	# of samples
20	13.3	0–100% <i>S-D</i>	16
40	13.3	0–100% <i>S-D</i>	6
30	13.3	0–100% <i>S-D</i>	6
10	13.3	0–100% <i>S-D</i>	6
0	13.3	0–100% <i>S-D</i>	8
0	33.3	80–95% <i>S-D</i>	4
0	20	70–100% <i>S-D</i>	4
15	26.7	50–100% <i>S-D</i>	3
15	40	65–85% <i>S-D</i>	3
15	53.3	80–100% <i>S-D</i>	3
5	66.7	75–95% <i>S-D</i>	3
5	80	80–100% <i>S-D</i>	3

of the dry cake and the standard deviation of the relative sampling error was calculated. The procedure was repeated five times with different percent composition of both diastereomers and the average relative sampling error was estimated to be  $\pm 6\%$ .

## 2.8. Data pre-treatment

Data pre-treatment using the calculation of the first derivative of the spectral measurement with respect to the spectral frequencies was performed to correct any baseline offset caused by the scattering effect from the solids. The first derivative of all the spectra was computed with the Savitzky–Golay method using a second-degree polynomial fit with an eleven points window (Madden, 1978). It was found that the 11 points window was sufficient to achieve a high level of smoothing without attenuating the extrema in the data. The higher order derivative did not show any significant improvement in model accuracy and hence the first derivative of the spectral measurement was chosen as the only data pre-treatment step.

The data matrix,  $D_j$  ( $j = 1, 2, 3, 4$ ), of each of the four PLS models is composed of different combinations of measurements such as the entire Raman spectrum (spectrum range:  $75\text{--}3300\text{ cm}^{-1}$ ), temperature, and/or slurry density as shown in Eqs. (1)–(4). The percent composition of the diastereomers was the prediction or output variable of each of the PLS models

$$D_1 = [\text{Spectra } (S)], \quad (1)$$

$$D_2 = [\text{Temperature } (T) \text{ Spectra } (S)], \quad (2)$$

$$D_3 = [\text{Slurry Density } (D) \text{ Spectra } (S)], \quad (3)$$

$$D_4 = [\text{Temperature } (T) \text{ Slurry Density } (D) \text{ Spectra } (S)]. \quad (4)$$

In order to compare the contribution of each input variable equally, temperature ( $T$ ), slurry density ( $D$ ), percent composition in solids ( $P$ ), and spectra ( $S$ ) were first scaled to have zero mean and variance of one as shown in Eqs. (5)–(10).

$$\hat{T}_i = \frac{T_i - T_{avg}}{\sigma_T}, \quad (5)$$

$$\hat{D}_i = \frac{D_i - D_{avg}}{\sigma_D}, \quad (6)$$

$$\hat{P}_i = \frac{P_i - P_{avg}}{\sigma_P}, \quad (7)$$

$$F_i = \int f_i(w) dw, \quad (8)$$

$$F_{avg} = \frac{\sum_{i=1}^n F_i}{n}, \quad (9)$$

$$\hat{S}_i = \frac{f_i(w) - F_{avg}}{\sigma_F}. \quad (10)$$

Here  $\hat{T}$ ,  $\hat{D}$ ,  $\hat{P}$ , and  $\hat{S}$  represented the scaled values, obtained by subtracting the average values ( $T_{avg}$ ,  $D_{avg}$ ,  $P_{avg}$ , and  $F_{avg}$ ) and divided by the corresponding standard deviation ( $\sigma$ ). The spectral data, distributed over a frequency spectrum were scaled slightly different, using the average  $F_{avg}$  and the standard deviation  $\sigma_F$  of the area  $F_i$  under the spectral curve  $f_i(w)$ ; see Eqs. (8)–(10). It should be noted that  $f_i(w)$  is the spectrum function with respect to the wave number  $w$  that uniquely corresponds to the spectral frequency.

## 3. Result and discussions

### 3.1. Raman spectra of pure diastereomers

The Raman spectra of the pure diastereomer in Fig. 1 show that there was only a slight difference between the two diastereomers. The circled regions of spectra highlighted the slight peak shifts between the diastereomers. Hence, chemometric techniques need to be employed to account for the subtle differences in the whole spectrum. In addition, the spectra of the pure *S-D* diastereomer were compared at different temperatures (Fig. 2) and with different slurry densities (Fig. 3). While the relative intensity differs slightly with temperature and there was no peak shift observed due to temperature effect, the Raman spectra of different slurry density showed differences in peak positions and peak shape. The denser sample showed in Fig. 3 had more distinct shape peaks that resembled Raman spectrum of pure solid. The appearance of the shape peaks was an indication that the Raman analyzer detected higher amount of solids as slurry density increased.

### 3.2. PLS calibration model

Four PLS models were developed to investigate whether the additional process measurements of temperature and slurry density would improve the accuracy of the estimation model. Each of the PLS models used the same training set that incorporates conditions that either neglect or account for variability in the additional process variables (See [Chemometrics Software I](#); [Chemometrics Software II](#)). The data matrix of each PLS model Eqs. (1)–(4) included 55 data points from Table 1.

The accuracy of the four PLS models were first tested with the testing set (10 standards points), in which these 10 experimental data points were not used during the model development. The testing set included standard points taken from different days of the experiment and with varying experimental conditions such that the testing set would be representative of future samples (Table 3). It should be noted that each of the calibration experiments only offered a static view of crystallization. Slurry density

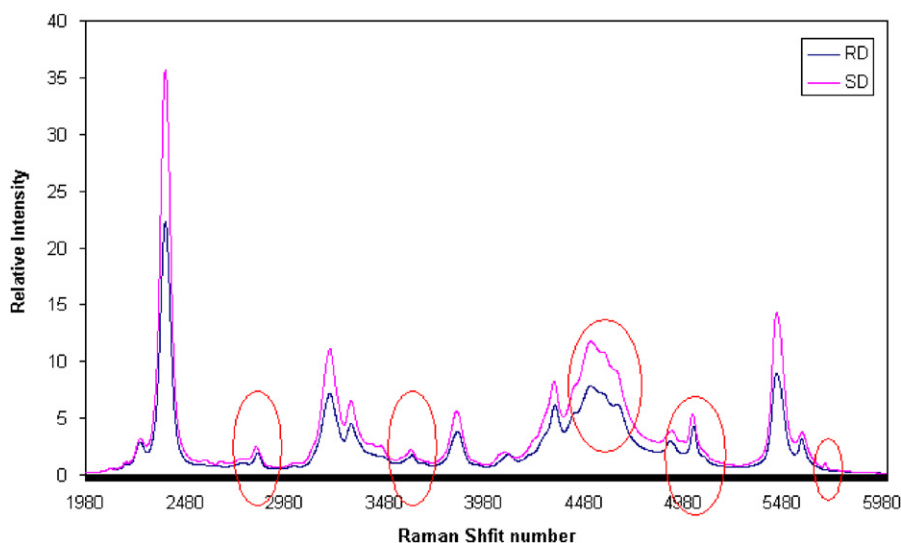


Fig. 1. Raman spectra of pure *R-D* and *S-D* at the same temperature and slurry density. The circles indicate different regions of slight peak shifts between the diastereomers.

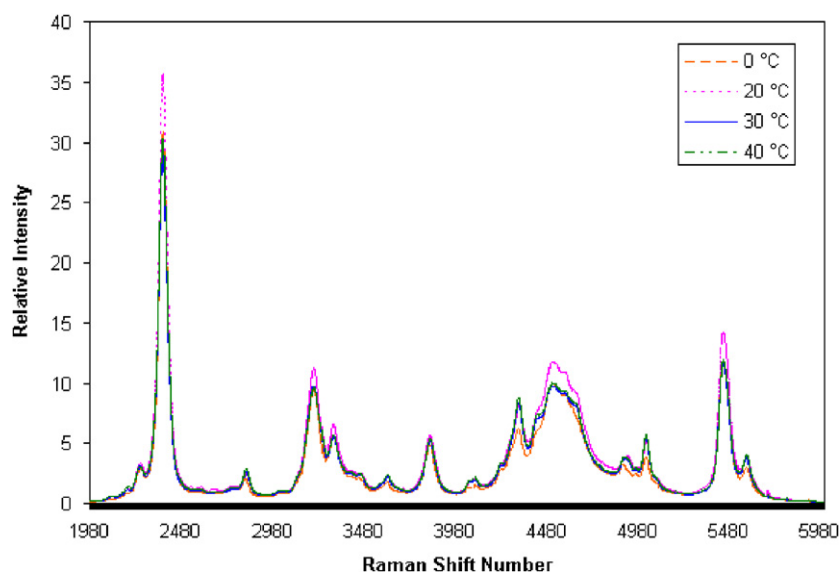


Fig. 2. Raman spectra of pure *S-D* with the same slurry density and different temperature.

and percent composition of both diastereomers remained constant throughout the experiment because the solution/solid phases of the system were in equilibrium. In order to further quantify the model accuracy, the models were then further compared with the result from the HPLC analysis of the dynamic crystallization experiment.

The estimation models were calculated by regressing the data of the training set with the percent composition using partial least square regression (PLS). Each training set was divided into four subgroups for cross-validation. The cross-validation method used three of the four subgroups to build the model and tested with the last subgroup as validation step. The process was repeated with all combinations of subgroups as training and testing sets. The number of Latent Variable (LV) selected for each PLS

model was based on the minimum prediction error ( $E_1$ ) from the cross-validation result, provided the percent of  $Y$  variance captured was more than 90%. The cross-validation result used  $E_1$  as the measure of performance (Fig. 4). In addition, the relative percent error ( $E_{2i}$ ) and root-mean-square of  $E_{2i}$  ( $E_3$ ) were also used to compare the different PLS models (Eqs. (11)–(13))

$$E_1 = \sqrt{\left( \sum_{i=1}^n (y_i^{es} - y_i^{ex})^2 \right) / n}, \quad (11)$$

$$E_{2i} = \left| \frac{(y_i^{es} - y_i^{ex})}{y_i^{ex}} \right| \times 100\%, \quad (12)$$

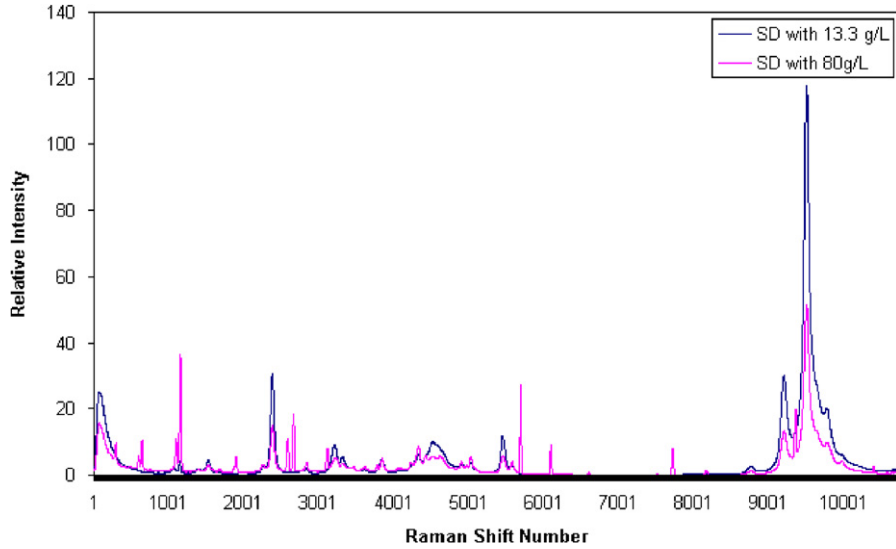


Fig. 3. Raman spectra of pure *S-D* with different slurry density.

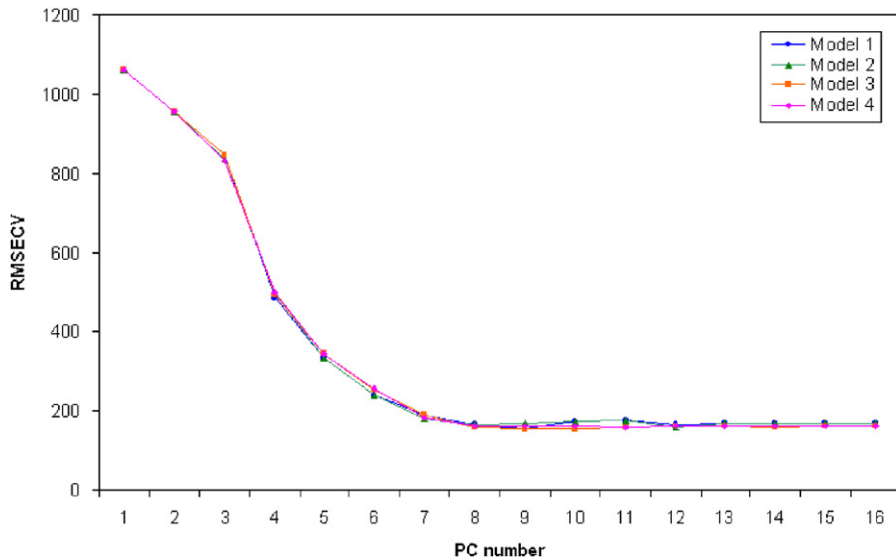


Fig. 4. RMSECV vs. number of latent variable for the PLS models.

$$E_3 = \sqrt{\left( \sum_{i=1}^n (E_{2i})^2 \right) / n}, \quad (13)$$

where  $n$  is the number of the total data points; the superscript,  $es$ , is the estimated value of  $y_i$  and the superscript  $ex$  is the experimental result of  $y_i$ .

From the results of the cross-validation, the number of LV was selected as described above. The detailed information of the models is summarized in Table 2. In addition, the predicted values,  $y_p$ , of  $Y$  (%*S-D*) were plotted against the measured values,  $y_m$ , of  $Y$  (%*S-D*) in Figs. 5–8 to check the accuracy of the models. It is observed that the data points all lie close to the diagonal line indicating small prediction error for all the models. Finally, the regression coefficients of Model 4 are plotted against the wavenumbers (Fig. 9) and the regression coefficients are further

Table 2  
Detailed information about four PLS models

	Model 1	Model 2	Model 3	Model 4
RMSECV	7.60	8.17	7.37	7.91
% $Y$ variance	96%	95%	95%	94%
# of LV	7	7	7	7
Process variables	$S$	$T, S$	$D, S$	$T, D, S$

compared with the first derivative of the Raman spectra of both diastereomers (Fig. 10). Although it is hard to observe the regression coefficients of temperature and slurry density in the figures, the regression coefficients of temperature and slurry density are 1.8 and 4.2, representatively, which illustrates the relative impact of temperature and slurry density have on the estimation model. The regression

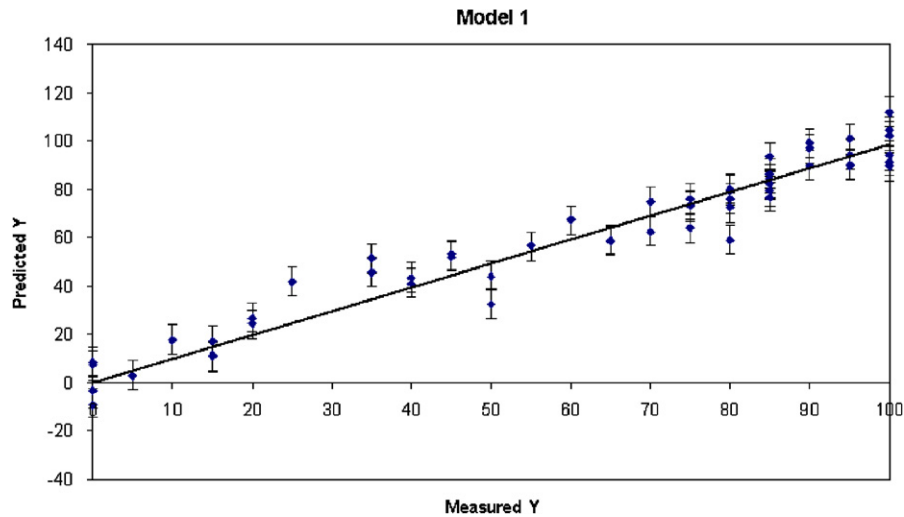


Fig. 5. Predicted value of  $Y$  vs. measured value of  $Y$  for Model 1.

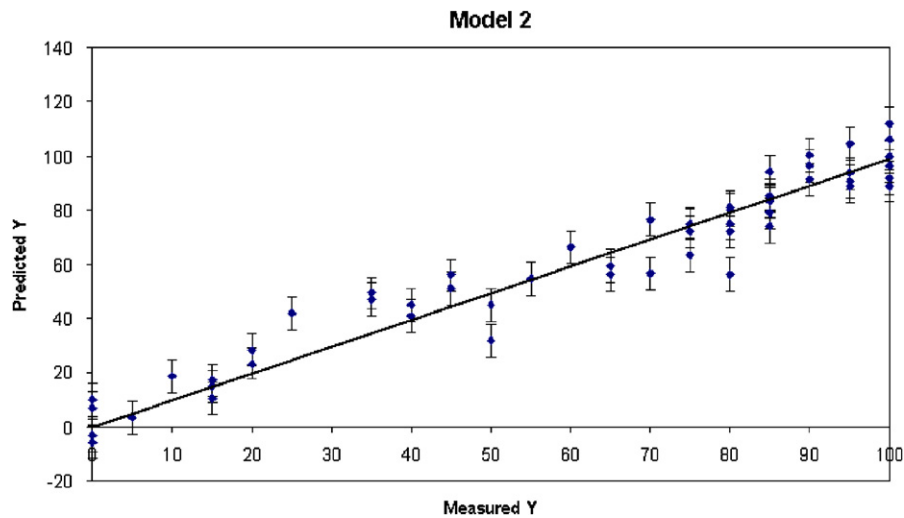


Fig. 6. Predicted value of  $Y$  vs. measured value of  $Y$  for Model 2.

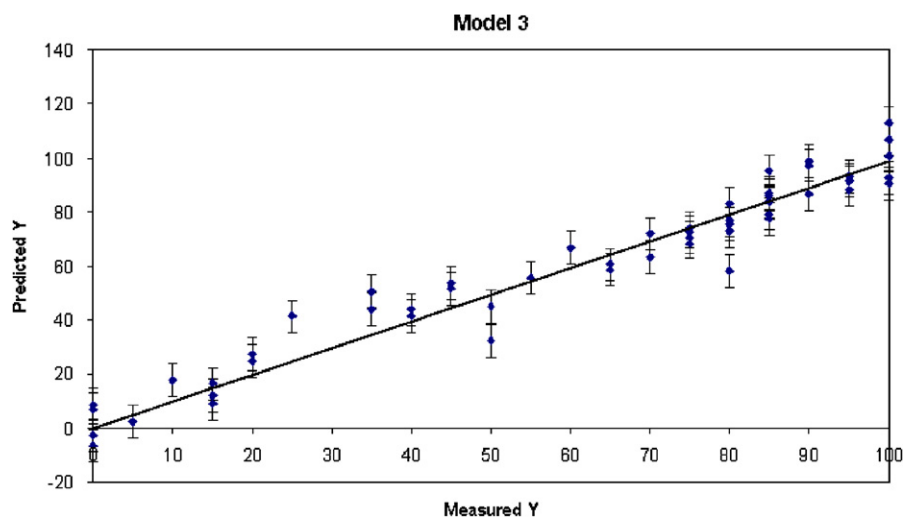


Fig. 7. Predicted value of  $Y$  vs. measured value of  $Y$  for Model 3.

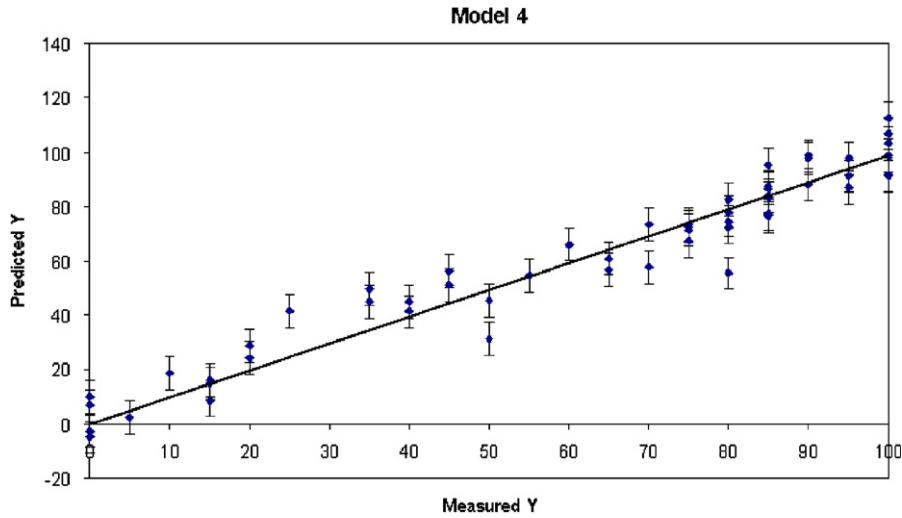


Fig. 8. Predicted value of  $Y$  vs. measured value of  $Y$  for Model 4.

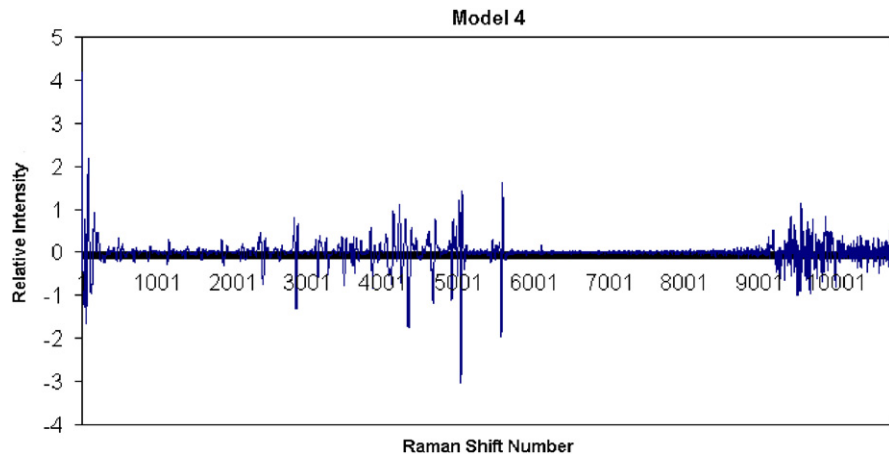


Fig. 9. The regression coefficients of Model 4.

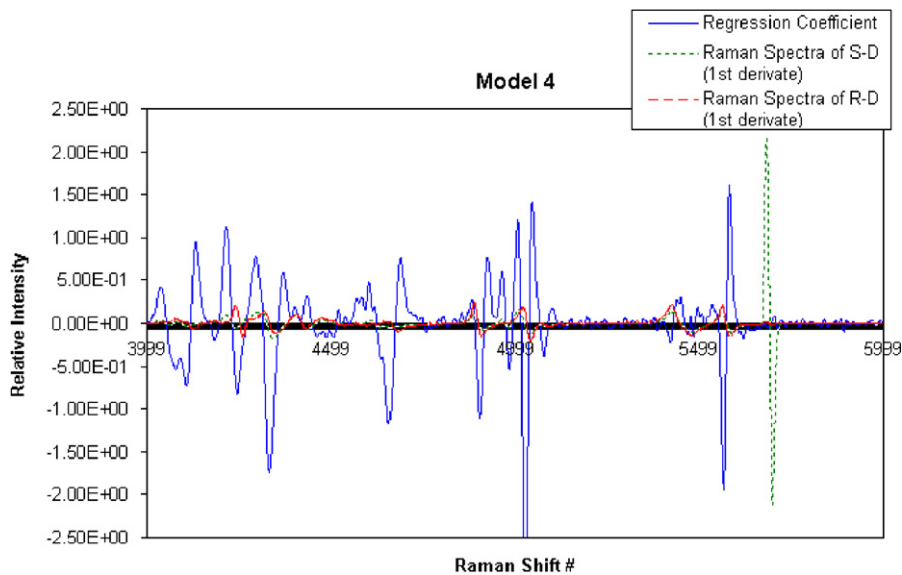


Fig. 10. Regression coefficient of Model 4 is plotted with the second derivate of Raman spectra of pure  $S$ - $D$  and pure  $R$ - $D$  diastereomers.

Table 3  
% Composition of SD diastereomer-validation result from the testing set

Sample	Measured	Model 1	Model 2	Model 3	Model 4
0°C-75%-20 g/L	75	74	75	76	76
0°C-95%-13.3 g/L	95	90	90	97	96
10°C-55%-13.3 g/L	55	62	62	60	60
15°C-65%-40 g/L	65	67	66	67	67
15°C-80%-53.3 g/L	80	81	82	82	83
15°C-100%-26.7 g/L	100	96	96	100	100
20°C-30%-13.3 g/L	30	31	30	30	29
20°C-50%-13.3 g/L	50	46	45	45	44
20°C-90%-13.3 g/L	90	91	88	89	87
40°C-60%-13.3 g/L	60	56	57	56	57

Table 4  
 $E_{2f}$  of the validation result from the testing set

Sample	Model 1	Model 2	Model 3	Model 4
0°C-75%-20 g/L	1.3	0.0	1.0	1.3
0°C-95%-13.3 g/L	5.3	5.3	1.9	1.1
10°C-55%-13.3 g/L	12.7	12.7	9.0	9.1
15°C-65%-40 g/L	3.1	1.5	3.6	3.1
15°C-80%-53.3 g/L	1.3	2.5	2.9	3.8
15°C-100%-26.7 g/L	4.0	4.0	0.1	0.0
20°C-30%-13.3 g/L	3.3	0.0	1.0	3.3
20°C-50%-13.3 g/L	8.0	10.0	9.3	12.0
20°C-90%-13.3 g/L	1.1	2.2	0.7	3.3
40°C-60%-13.3 g/L	6.7	5.0	6.9	5.0

Table 5  
Validation result from the testing set

	Model 1	Model 2	Model 3	Model 4
$E_1$	3.61	3.65	2.80	3.08
$E_3$	5.82	5.87	4.92	5.48

coefficient of slurry density also shows the strong effect it has on the estimation model.

The final step in assessing the accuracy of the PLS model was to calculate the prediction error with the set of 10 experiments that was set aside during model development (Table 3). The result showed that the PLS models along with Raman spectra that incorporated slurry density measurements or with the measurement of slurry density and temperature performed better than the models using only spectral data with or without temperature data (Tables 4 and 5).

### 3.3. Result from crystallization experiment

In order to further quantify their accuracy, the PLS models were used to predict the percent composition of the *S-D* diastereomer in a dynamic crystallization experiment. While the standards from the calibration set were pre-mixed slurries in 15 mL vials, the new crystallization

experiment was a time evolving or dynamic run in a 250 mL round-bottom jacketed flask. Six samples were drawn sequentially at different time instances after the onset of nucleation and were submitted for HPLC analysis. During this dynamic crystallization experiment, Raman and temperature data were collected and the slurry densities were calculated for the withdrawn samples.

According to the present understanding of the experimental system, the *R-D* diastereomer has a wider metastable zone and a slower growth rate than the *S-D* diastereomer. Besides assessing the accuracy of the different PLS models, the crystallization experiment aimed to investigate whether the *R-D* diastereomer would crystallize out simultaneously with the *S-D* diastereomer in an unseeded environment. The HPLC measurements revealed that, as nucleation occurred, both diastereomers initially crystallized out. However, due to the slower crystallization kinetics of the *R-D* diastereomer, the percent composition of the *S-D* diastereomer slowly increased over time.

The PLS model predictions were compared with the HPLC data as shown in Tables 6 and 7 and  $E_1$  and  $E_3$  were used as means to compare the models (Table 8). Although all four models performed quite accurately with the previously examined testing set, the accuracy in predicting the data of the new crystallization experiment is understandably smaller. The larger bias resulted in the prediction is possibly due to slight differences between the calibration and crystallization experiments. First of all, the percent composition of both diastereomers in solution phase is different in the two experimental setups. In the calibration experiment, the solution phase was first saturated with respect to both diastereomers, and as a result percent, the

Table 6  
HPLC analytical result vs. model prediction of the % composition of SD diastereomer

Sample	Measured	Model 1	Model 2	Model 3	Model 4
1	73.8	62.3	65.5	70.5	72.4
2	82.0	66.8	69.2	77.1	78.0
3	88.5	65.5	68.0	75.7	76.6
4	87.5	74.2	75.3	84.5	84.2
5	88.0	62.5	65.7	71.4	73.3
6	82.8	62.2	65.4	71.3	73.1

Table 7  
 $E_{2f}$  of the % composition of SD diastereomer prediction in crystallization experiment

Sample	Model 1	Model 2	Model 3	Model 4
1	15.6	11.3	4.5	1.9
2	18.6	15.6	6.0	4.9
3	26.0	23.2	14.5	13.4
4	15.2	13.9	3.4	3.8
5	29.0	25.4	18.8	16.8
6	24.9	21.1	13.9	11.7

Table 8  
Error between models and experiment

	Model 1	Model 2	Model 3	Model 4
$E_1$	18.9	16.4	10.1	9.0
$E_3$	22.2	19.1	11.7	10.3

composition of the diastereomers remained constant throughout all the calibration experiments. On the other hand, as nucleation occurred in the dynamic crystallization experiments, the percent composition of the diastereomer in solution phase changes constantly until the system reaches equilibrium. Secondly, the mixing environment between the calibration and crystallization experiments is also different. Although both experiments used the same agitation method, the different mixing volume and the different Raman probe position potentially affects the Raman spectrum, which further contribute to the larger bias. Nevertheless, the superiority of Model 4, which includes slurry density and temperature as inputs, is clearly demonstrated. It is observed that when compared Model 1 and Model 4 in terms of  $E_1$  and  $E_3$ , the accuracy of Model 4 has improved by over 100%. The improvement in model accuracy by incorporating temperature and slurry density into the model can be explained by the crystallization dynamic. When nucleation occurred, there was only a thin layer of slurry in the solution. As the experiment progressed, more material crystallized out and the system tried to reach equilibrium between the two solids. It should also be noted that Model 3 utilizing only slurry density and spectra data had a RMSE of 10.1 for the crystallization experiment. This clearly confirmed that the inclusion of slurry density into the PLS model improves the accuracy of estimation.

#### 4. Conclusion

Four PLS models were constructed with 65 calibration standards using Raman spectra, temperature and/or slurry density data. The models were further tested and compared against data from a larger scale 250 mL dynamic crystallization experiment. This paper has shown that in situ Raman spectroscopy is capable of differentiating diastereomers in a crystallization slurry, provided the changing process parameters of temperature, and most importantly, slurry density are included in the calibration model. In addition, the spectral data acquisition time is quite less than 1 min and the estimation of the diastereomer composition is even faster which makes monitoring the changing composition of diastereomer during crystallization in real time indeed very possible. The implementation of a simple feedback controller onto the system can be done if on-line density measurement is made available. Because slurry density is not easily measured on-line

without a sampling loop, it is essential to find an alternative on-line measurement from which to infer slurry density.

#### Acknowledgment

The financial support from Sepracor Inc. in the form of an internship to Sze-Wing Wong and the availability of their experimental facilities is greatly appreciated. Special thanks to Camo Inc. for providing a free license of The Unscrambler software.

#### References

- Berglund, K. A., Wang, F., Wachter, J. A., & Antosz, F. J. (2000). An investigation of solvent mediated polymorphic transformation of progesterone using in situ Raman spectroscopy. *Organic Process Research and Development*, 4, 391–395.
- Caillet, A., Puel, F., & Fevotte, G. (2006). In-line monitoring of partial and overall solid concentration during solvent-mediated phase transition using Raman spectroscopy. *International Journal of Pharmaceutics*, 307, 201–208.
- Chemometrics Software I. PLS\_Toolbox 3.5 by Eigenvector Research, Inc., Manson, WA.
- Chemometrics Software II. The Unscrambler from Camo Inc., Trondheim, Norway.
- Elsner, M. P., Menendez, D. F., Muslera, E. A., & Seidel-Morgenstern, A. (2005). Experimental study and simplified mathematical description of preferential crystallization. *Chirality*, 17, S183–S195.
- Falcon, J. A., & Berglund, K. A. (2004). In situ monitoring of antisolvent addition crystallization with principle components analysis of Raman spectra. *Crystal Growth and Design*, 4, 457–463.
- Ferraro, J. R. (1971). *Low-frequency vibrations of inorganic and coordination compounds*. New York: Plenum Press.
- Hu, Y., Liang, J. K., Myerson, A. S., & Taylor, L. S. (2005). Crystallization monitoring by Raman spectroscopy: Simultaneous measurement of desupersaturation profile and polymorphic form in flufenamic acid system. *Industrial and Engineering Chemistry Research*, 44, 1233–1240.
- Jacques, J., Collet, A., & Wilen, S. H. (1981). *Enantiomers, racemates, and resolution*. New York: Wiley.
- Madden, H. H. (1978). Comments on the Savitzky–Golay convolution method for least-squares fit smoothing and differentiation of digital data. *Analytical Chemistry*, 50, 1383–1386.
- Ono, T., Horst, H. T., & Jansens, P. J. (2004). Quantitative measurement of the polymorphic transformation of L-glutamic acid using in situ Raman spectroscopy. *Crystal Growth and Design*, 4, 465–469.
- O'Sullivan, B., Barrett, P., Hsiao, G., Carr, A., & Glennon, B. (2003). In situ monitoring of polymorphic transitions. *Organic Process Research and Development*, 7, 977–982.
- Pelletier, M. J. (2003). Quantitative analysis using Raman spectrometry. *Applied Spectroscopy*, 57, 20A–42A.
- Qian, R. Y., & Botsaris, G. B. (1997). A new mechanism for nuclei formation in suspension crystallizers: The role of interparticle forces. *Journal of Chemical Engineering Science*, 52, 3429–3440.
- Rades, T., Pratiwi, D., Fawcett, J. P., & Gordon, K. C. (2002). Quantitative analysis of polymorphic mixtures of ranitidine hydrochloride by Raman spectroscopy and principal components analysis. *European Journal of Pharmaceutics and Biopharmaceutics*, 54, 337–341.
- Scholl, J., Bonalumi, D., Vicum, L., Mazzotti, M., & Muller, M. (2006). In situ monitoring and modeling of the solvent-mediated polymorphic transformation of L-glutamic acid. *Crystal Growth and Design*, 4, 881–889.

- Starbuck, C., et al. (2002). Process optimization of a complex pharmaceutical polymorphic system via in situ Raman spectroscopy. *Crystal Growth and Design*, 2, 515–522.
- Togkalidou, T., Fujiwara, M., Patel, S., & Braatz, R. D. (2001). Solute concentration prediction using chemometrics and ATR-FTIR spectroscopy. *Journal of Crystal Growth*, 231, 534–543.
- Wilens, S. H., Collet, A., & Jacques, J. (1977). Strategies in optical resolutions. *Tetrahedron*, 33, 2725–2736.
- Zhou, G., Wang, J., Ge, Z., & Sun, Y. (2002). Ensuring robust polymorph isolation using in situ Raman spectroscopy. *American Pharmaceutical Review*, Winter 2002.