AAPS PharmSciTech, Vol. 9, No. 2, June 2008 (© 2008)
DOI: 10.1208/s12249-008-9049-5

Research Article

Visualizing Solvent Mediated Phase Transformation Behavior of Carbamazepine Polymorphs by Principal Component Analysis

Fang Tian,[1] Thomas Rades,[1] and Niklas Sandler[2,3]

Received 4 May 2007; accepted 19 January 2008; published online 21 February 2008

Abstract. The purpose of this research is to gain a greater insight into the hydrate formation processes of different carbamazepine (CBZ) anhydrate forms in aqueous suspension, where principal component analysis (PCA) was applied for data analysis. The capability of PCA to visualize and to reveal simplified structures that often underlie large data sets are explored. Different CBZ polymorphs were dispersed separately in aqueous solution, and then recovered and measured by FT-Raman spectroscopy. PCA was employed for visualizing the dynamics of the phase transformation from each CBZ polymorph to the dihydrate (DH). As a comparison to PCA visualization, the transformation process of each CBZ polymorph was quantified using PLS modeling. The results demonstrated that PCA has advantages in presenting the original data in terms of the differences and similarities, and also directly identify the statistical patterns in the data even when the data set is large. These advantages provided greater insight into the measured Raman spectra as well as the phase transformation process of CBZ polymorphs to the DH in aqueous environment.

KEY WORDS: carbamazepine; dihydrate; principal component analysis.

INTRODUCTION

Solvent mediated polymorphic transformations from the anhydrate to the hydrate form of a drug are likely to be induced in certain pharmaceutical manufacturing processes e.g. wet granulation (1). Also recent studies with various pharmaceutical compounds have shown the influence of polymorphic conversions on the dissolution of drugs (2–4). The control of the polymorphic form is therefore of obvious importance in drug and dosage form development. Amongst a wide range of analytical techniques, Raman spectroscopy has attracted growing attention in recent years, not least due to its potential as a real time process analytical tool (5).

Meaningful data analysis is the key for improved understanding of any generated data in drug development. Recently the pharmaceutical industry has been encouraged to apply modern process monitoring tools to collect timely data for improved understanding of systems and processes consequently minimizing the risks for failures (6,7). In these cases the collected data will most often be multidimensional in nature and multivariate data analysis tools become useful if not indispensable in data interpretation and modeling. For qualitative purposes multivariate visualization such as principal component analysis (PCA) has large potential and may be extremely useful in increasing process understanding. PCA can visualize changes in variable values which often helps to give a better perception of any changes, whilst also reducing data dimensionality. Furthermore, PCA is a simple, non-parametric method of extracting relevant information from confusing data sets. Therefore, PCA has been called one of the most valuable results from applied linear algebra (8).

Carbamazepine (CBZ), an important anti-convulsant drug for epilepsy and trigeminal neuralgia treatment, has been in routine use for over 20 years (9). It is well known that CBZ in its anhydrate forms is unstable under humid conditions or in aqueous dispersions and converts rapidly to its DH when exposed to water. The conversion to the DH also leads to a significant decrease in the bioavailability of its marketed tablets (10–12). Since water is virtually ubiquitous as bulk water or water vapor, the conversion of CBZ to the DH is likely to be induced during all stages of its manufacturing process such as mixing, granulation, packaging etc. Therefore, a strict process monitoring and control of the existing form and amount of the drug substance in the formulation at each step of its manufacture is necessary to ensure the quality of the final products.

The main purpose of this study was to gain a greater insight into the hydrate formation processes of different aqueous dispersions of different CBZ anhydrate forms (polymorphic forms I, II and III) by examining data measured by Raman spectroscopy. PCA was applied for reducing dimensionality of the complex pharmaceutical data sets. Its capability to visualize and to reveal simplified structures that often underlie large data sets are explored.

[1] School of Pharmacy, University of Otago, Dunedin, New Zealand.
[2] AstraZeneca R&D, Pharmaceutical and Analytical R&D, Silk Road Business Park, Macclesfield, SK10 2NA, UK.
[3] To whom correspondence should be addressed. (e-mail: niklas.sandler@astrazeneca.com)

MATERIALS AND METHODS

Materials

CBZ form III (Alphapharm Pty Ltd, Glebe, Australia) was used as received. Form II was prepared by freeze-drying freshly prepared DH. DH was prepared from the CBZ as received from Sigma Chemical Company (St. Louis, MO, USA) by recrystallization from an ethanol–water mixture as reported by Krahn and Mielck (13). Two source materials were used for form I preparation: form I (1st batch) from Sigma Chemical Company (St. Louis, MO) and form I (2nd batch) from Sigma-Aldrich Chemie (Munich, Germany). All forms were confirmed by X-ray powder diffraction as reported previously (14).

The particle size of the different polymorphic forms was controlled by sieving to the range 180 to 250 μm (Test sieves, Endecotts Ltd, England).

Experimental Design of Hydrate Formation

Each CBZ anhydrate form (40 mg) was dispersed separately in triplicate into 2 ml distilled water at room temperature while stirring, and then recovered for analysis after several predetermined time intervals. The recovery is by pouring the dispersion onto two layers of filtration paper to remove excess water, and then scraping the resulting slurry from the filtration paper surface. The slurry samples were then transferred into three aluminium sample cups consecutively, and the Raman spectra were recorded immediately after filling the cups.

FT-Raman Spectroscopy

The FT-Raman instrument consisted of a Bruker FRA 106/S FT-Raman accessory (Bruker Optik, Ettlingen, Germany) with a Coherent Compass 1064-500N laser (Coherent Inc, Santa Clara, USA) attached to a Bruker IFS 55 FT-IR interferometer, and a D 425 Ge diode detector. Analysis was carried out at room temperature utilizing a laser wavelength of 1064 nm (Nd:YAG laser) and a laser power of 105 mW. Back-scattered radiation was collected at an angle of 180°. Samples were packed into an aluminum cup and a total of 16 scans was averaged for each sample at a resolution of 4 cm^{-1}. Sulfur was used as reference standard to monitor the wavenumber accuracy. OPUS™ 5.0 (Bruker Optik, Ettlingen, Germany) was used for all spectral analysis.

Preparation of Calibration Models and PLS Analysis

PLS analyses were performed using the Quant2 package that accompanies OPUS™ software (Bruker Optics, Germany). The spectral regions for calibration were selected, and all spectra were mean centered. Multiplicative scattering correction and/or first derivative calculation were used for preprocessing spectra if necessary to correct for sample packing and baseline differences. The calibration models were calculated using the PLS algorithm and cross-validation (one sample removed per cycle). Details of the method were reported in an earlier paper (15).

PCA Analysis

PCA modeling was performed with SIMCA-P+ (Version 11.0.0.0). Cross-validation was used to determine the optimum number of principal components (PCs). All the spectral data were normalized using OPUS™ software (Bruker Optics, Germany) and then used in the final PCA models.

Mathematical Models Applied to Spectral Data

PCA Analysis

In PCA, the dimensionality of the multidimensional data is reduced by transforming the original variables into new axes which will be laid along the directions of maximum variance of the data with the constraint that the axes are orthogonal. The orthogonal features are called principal components (PCs). The first principal component, PC1, in PCA always describes the largest variance with normalized coefficients applied to the variables used in the linear combinations, the second principal component, PC2, always has the largest remaining variance etc. The performance of the PCA can be mathematically described as $X = UV^{T}$. The two matrices U and V are orthogonal. The matrix U contains the original data in a rotated coordinate system. It is called the scores matrix. The matrix V is usually called the loadings matrix, and T (in superscript) indicates transposition. The loadings can be regarded as the weights for each original variable when calculating the principle components. A detailed description of PCA has been provided in several books and tutorials (16,17).

Partial Least Squares Model

In PLS, each component is obtained by maximizing the covariance between Y and all possible linear functions of X. This thus leads to components which are more directly related to variability in Y than are the principle components. The purpose of PLS regression is to predict Y from X and to describe their common structure. It performs a simultaneous decomposition of X and Y with the constraint that these decomposed components explain as much as possible of the covariance between X and Y. This step will then be followed by a regression step where the decomposition of X is used to predict Y. This can be described as $X = TP^{T}$ with T is called the scores matrix, P is the loadings matrix and T (in superscript) indicates transposition. The algorithms for calculating the PLS model eigenvectors and scores have been described in detail in several books and reviews (18–20).

RESULTS AND DISCUSSION

Understanding the Phase Transformation from Loading Plots in PCA

The Raman spectra of the initial CBZ anhydrate forms and DH are shown in Fig. 1a. The projection of spectral information for the first three principal components (PCs) is presented in Fig. 1b–d. The PCA model was built using the whole measured Raman region from 3,200 to 200 cm^{-1}.

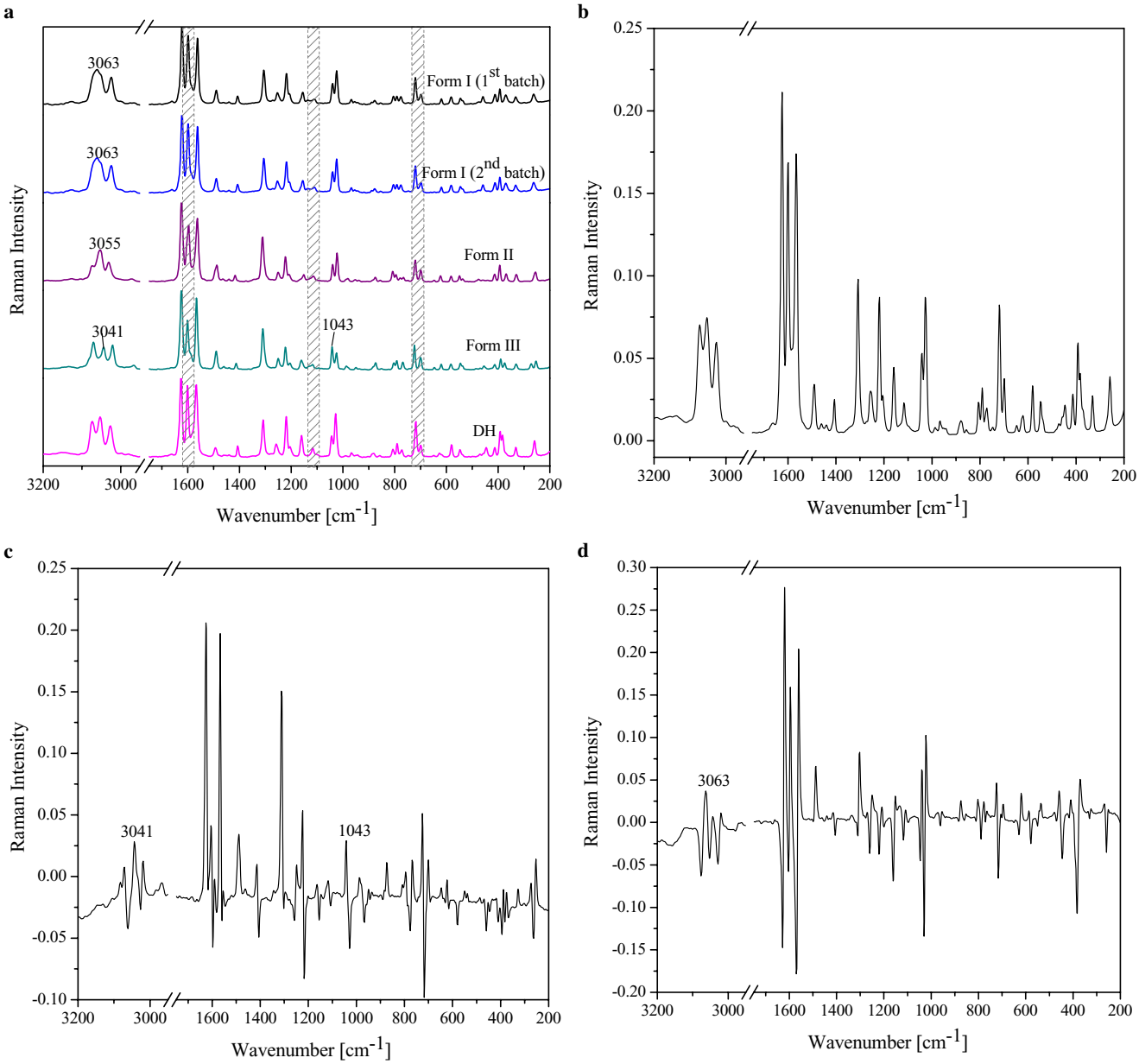

Fig. 1. Normalized Raman spectra of CBZ polymorphic forms (**a**, *shaded areas* showing CBZ bands which are directly related to its H-bonding with water) and plots comparing the loadings of the first three principal components (**b–d** respectively)

The two batches of form I showed a different morphology with form I (1st batch) mainly consisting of prism-like crystals and form I (2nd batch) of needle-like crystals. However, both of them showed identical Raman spectra. This has been described in detail in an earlier paper from our group (15). Also, it has to be mentioned that Raman is fairly insensitive to the polar groups of the molecule and therefore the CBZ vibrational bands which are directly related to its hydrogen bonding with water (shown in shaded areas in Fig. 1a) are mainly medium or weak Raman scatterers. For example, according to the CBZ vibrational bands assignment from O'Brien *et al.* in 2004 (21), the bands at around 1,600 cm^{-1} have been assigned to (N–H) amide II stretching vibrations, and those at around 1,116 cm^{-1} and 723 cm^{-1} are likely to be due to NH_2 rocking and (C–N–C) stretching vibrations respectively. Attempts have been made to detect changes of these bands during hydrate formation directly from the Raman spectra obtained, however, as expected this was found to be difficult.

PCA was thus employed as a data analytical method in this study. It is always valuable to examine the loadings plot in PCA since it provides useful information for a clearer understanding of the data. The loadings plot can be used to tell which of the variables are most relevant for the different components in the data set, and identifying which variables characterize the different clusters (22). Therefore, the loadings plot of the first three principal components (shown in Fig. 1b–d) was firstly checked for a better understanding of the whole Raman dataset obtained during transformation. Interestingly, the first principal component PC 1 is only

limited to positive values and it is almost identical to the DH spectrum (Fig. 1a and b). This indicates a dominance of the DH spectrum in the spectra obtained during the whole hydration process. Therefore, it can be speculated the DH has already started forming in an early stage of the dispersion process and kept increasing while the hydration proceeds. The rapid DH formation of CBZ polymorphic forms upon moisture contact has been reported in several studies (23–25). The Raman spectra of the different CBZ anhydrates and the DH were fairly similar as demonstrated in Fig. 1a, and such high similarity among all the spectra might lead to the loading plot of PC 1 showing only positive values.

Both second and third principal components (PCs 2 and 3) accounted for a much smaller variation of the whole spectra compared to that of PC 1, with 1.7% and 1% respectively. The main variations explained by PC 2 (Fig. 1c) were noticed specifically for the bands at 3,041 and 1,043 cm^{-1}, which decreased while dehydration proceeds (shown in Fig. 2). Therefore, PC 2 mainly reflected the hydration process of form III. PC 3 mainly explain the spectral changes of form I where a characteristic band of form I at 3,063 cm^{-1} could be observed in Fig. 1d.

Visualization of the Phase Transformation Process

Data visualization using PCA has the advantage of presenting complex data in a simple form. This is exemplified in the three dimensional plot in Fig. 2, which represents the whole conversion process from different CBZ polymorphic forms and also the same form (form I) but with different morphology, to the DH respectively.

The CBZ anhydrate forms are oriented mainly in one dimension but follow different clusters in the hydrate formation process to the DH. The transformation of form III appears to be gradual as the data points are spreading

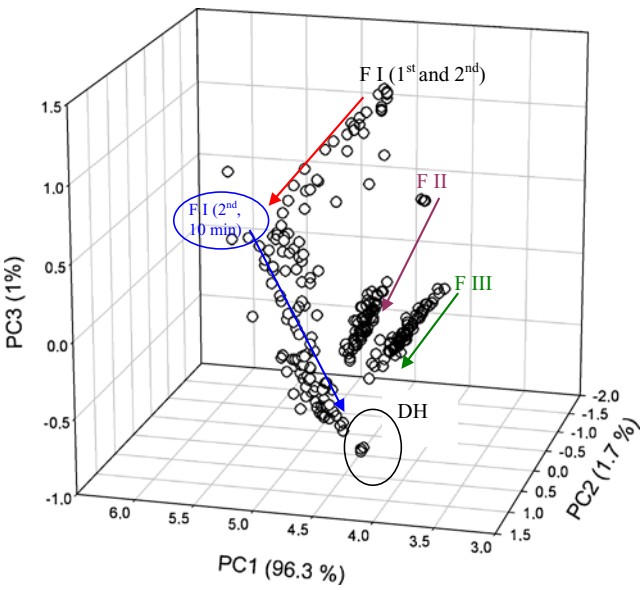

Fig. 2. Three dimensional plot showing the transformation process of CBZ polymorphic forms to the DH during dispersion. Form I with smaller surface area (*red arrow*), form I with larger surface area (*blue arrow*), form II (*purple arrow*), form III (*green arrow*), and DH (*in circle*)

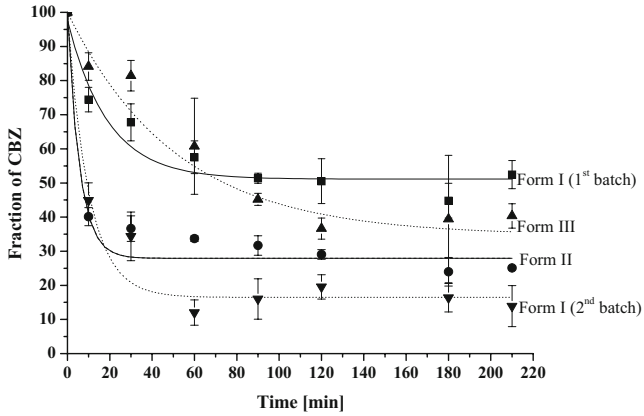

Fig. 3. Conversion of CBZ anhydrate forms quantified using Raman spectroscopy combined with PLS: (*filled triangles*) form I (1st batch); (*filled squares*) form III; (*filled circles*) form II; (*filled inverted triangles*) form I (2nd batch)

evenly along a line, but the transformation has not been completed after 210 min dispersion since there is still some distance between the 210 min data point and the cluster of the DH data. The transformation behaviour of form II is different to that of form III. The cluster points of the initial form II and that of form II dispersed for 10 min have a relative large distance to each other, indicating a rapid transformation in the first 10 min. This transformation however slowed down since the later data points exhibited a similar gradual spreading to the DH cluster. For the two batches of form I having different particle morphology, two main clusters could be observed which appear to have a turning point in their hydrate formation process although they have the exact same starting cluster. This is because of their identical initial Raman spectra but different hydrate formation kinetics, which has been reported earlier (15). Form I (1st batch) consists of prism like crystals and is demonstrating a more gradual hydrate formation process as indicated by the red arrow in Fig. 2. Form I (2nd batch) has a needle-like morphology, having a much larger surface area contributing to a much faster hydrate formation than form I (1st batch). Its rapid conversion can be observed following the blue arrow after the turning point in Fig. 2.

Comparison of PCA and PLS in Analyzing Polymorphic System

The transformation rate and extent of all the dispersed samples has previously been quantified by combining Raman with PLS (2,15). The quantified hydrate formation processes are presented in Fig. 3 below for comparison with those visualized from PCA.

The transformation rate and extent quantified for CBZ anhydrate forms are in good agreement with those visualized in the PCA plot (Fig. 2). As can be seen from Fig. 3 about 50% form III remained unconverted to the DH after 210 min dispersion whereas the remaining form II was only about 30%. Therefore, form II has transformed to a higher extent (about 10% more) to the DH after 210 min dispersion. This can be seen by the larger distance from the last form III data point to the DH compared with that distance in the data set of pure form II in Fig. 2. Also, one apparent difference in the

transformation process of form III and II is that form II transformed rapidly once contacting water. As seen in Fig. 3 only about 40% of form II remain after the first 10 min. This could also be visualized by PCA in the relatively large distances between the first few form II data points and the 10 min data points in Fig. 3.

Furthermore, the two batches of form I showed very different transformation processes. Form I (1st batch) had a similar gradual transformation to that of form III. The transformation stopped at about 55% CBZ form I remaining, which is again exhibited in PCA by the comparatively large distances between the data points in the cluster of form I dispersed for 210 min and DH in Fig. 2. However, form I (2nd batch) already transformed by about 45% to the DH after 10 min dispersion and only 15% CBZ were left after another 10 min of dispersion. This significant difference in the hydrate formation process of the two batches form I was visualized by PCA as two clusters orientated at two totally different dimensions as indicated in Fig. 2.

It is apparent that the information obtained from PCA in this study agreed well with that found from PLS earlier. It also has to be mentioned that although PCA is not designed for quantitative studies, the transformation rate and extent of different samples can still be directly, reliably and easily visualized from the PCA plots.

CONCLUSIONS

The results of this study show that Principal component analysis provides a deep insight into the transformation process of CBZ polymorphic forms to the DH in aqueous suspension through direct visualization of the Raman spectral data. It can be concluded that the use of PCA and PLS are complementary for an improved understanding of pharmaceutical systems. In this study the former enabled effective visualisation of the kinetics of a complex data set and the latter complemented with quantitative information. The results demonstrated that PCA has advantages in presenting the original data in terms of the differences and similarities, and directly identifies the patterns in the data even when the data set is large.

REFERENCES

1. A. C. Jørgensen, J. Rantanen, M. Karjalainen, L. Khriachtechev, E. Rasanen, and J. Yliruusi. Hydrate formation during wet granulation studies by spectroscopic methods and multivariate analysis. *Pharma. Res.* **19**:1285–1291 (2002).
2. F. Tian, F. Zhang, N. Sandler, *et al*. Influence of sample characteristics on quantification of carbamazepine hydrate formation by X-ray powder diffraction and Raman spectroscopy. *Eur. J. Pharm. Biopharm.* **66**:466–474 (2007).
3. F. Tian, N. Sandler, J. Aaltonen, *et al*. Influence of polymorphic form, morphology and excipient interactions on the dissolution of carbamazepine compacts. *J. Pharm. Sci.* **96**:584–594 (2006).
4. J. Aaltonen, P. Heinänen, L. Peltonen, *et al*. *In-situ* measurement of solvent mediated phase transformations during dissolution testing. *J. Pharm. Sci.* **95**:2730–2737 (2006).
5. M. T. Islam, N. Rodriguez-Hornedo, S. Ciotti, and C. Ackermann. The potential of Raman spectroscopy as a process analytical technique during formulations of topical gels and emulsions. *Pharm. Res*. **21**:1844–1851 (2004).
6. Process Analytical Technology (PAT) initiative. The Food and Drug Administration (FDA) Website. Available at: http://www.fda.gov/cder/OPS/PAT.htm. Accessed December 17, 2003.
7. FDA, Guidance for Industry. PAT—a framework for innovative pharmaceutical manufacturing and quality assurance, *U.S. Food and Drug Administration (FDA),* Rockville, MD, USA. 2004:21.
8. J. Shlens. A tutorial on principal component analysis. Available at: http://www.cs.cmu.edu/~elaw/papers/pca.pdf. Accessed December 10, 2005.
9. R. Hartley, J. Aleksandrowicz, P. C. Ng, B. Mclain, C. J. Bowmer, and W. I. Forsythe. Breakthrough seizures with generic carbamazepine: a consequence of poorer bioavailability. *Br. J. Clin. Pract.* **44**:270–273 (1990).
10. Anon. Moisture hardens carbamazepine tablets, FDA finds. *Am. J. Hosp. Pharm*. **47**:958 (1990).
11. O. Shaheen, H. Mouti, M. Karmi, M. Dadala, and S. Othman. Comparative bioavailability of two brands of carbamazepine. *Curr. Ther. Res*. **45**:517–524 (1989).
12. C. Lefebvre, A. M. Guyot-Hermann, M. Draguet-Brughmans, and R. Bouche. Polymorphic transitions of carbamazepine during grinding and compression. *Drug Dev. Ind. Pharm*. **12**:1913–1927 (1986).
13. F. U. Krahn, and J. B. Mielck. Relations between several polymorphic forms and the dihydrate of carbamazepine. *Pharm. Acta. Helv*. **62**:247–254 (1987).
14. A. L. Grzesiak, M. D. Lang, K. Kim, and A. J. Matzger. Comparison of the four anhydrous polymorphs of carbamazepine and the crystal structure of form I. *J. Pharm. Sci*. **92**:2260–2271 (2003).
15. F. Tian, J. A. Zeitler, C. J. Strachan, D. Saville, C. K. Gordon, and T. Rades. Characterizing the conversion kinetics of carbamazepine polymorphs to the dihydrate in aqueous suspension using Raman spectroscopy. *J. Pharm. Biomed. Anal* **40**:271–280 (2006).
16. K. Y. Yeung, and W. L. Ruzzo. Principal component analysis for clustering gene expression data. *Bioinformatics* **17**:763–74 (2001).
17. I. S. Lindsay. A Tutorial on PCA. Maintained by Cornell University, USA. Available at: http://www.ucl.ac.uk/oncology/MicroCore/HTML_resource/PCA_1.htm. Accessed February 26, 2002.
18. K. R. Beebe, R. J. Pell, and M. B. Seasholtz, eds. *Chemometrics: A Practical Guide*. John Wiley & Sons, Inc. A Wiley-Interscience Publication, 1998, pp. 56–182.
19. H. Martens, and T. M. Naes. *Multivariate Calibration, Chapter 5*, Wiley, New York, USA, 1989.
20. B. M. Wise, N. B. Gallagher, and R. Bro. *PLS_Toolbox Version 3.5 (for use with MATLAB)*, Eigenvector Research, Inc, Manson, WA 98831, 2004.
21. L. E. O'Brien, P. Timmins, A. C. William, and P. York. Use of in situ FT-Raman spectroscopy to study the kinetics of the transformation of carbamazepine polymorphs. *J. Pharm. Biomed. Anal*. **36**:335–340 (2004).
22. T. Næs, T. Isaksson, T. Fearn, and T. Davies, eds. *A User-Friendly Guide to Multivariate Calibration and Classification*. Chichester: NIR Publications, 2002, pp. 39–46.
23. F. Tian, N. Sandler, K. C. Gordon, *et al*. Visualizing the conversion of carbamazepine in aqueous suspension with and without the presence of excipients: a single crystal study using SEM and Raman microscopy. *Eur. J. Pharm. Biopharm*. **64**:326–335 (2006).
24. E. Laine, V. Tuominen, P. Iivessalo, and P. Kahela. Formation of dihydrate from carbamazepine anhydrate in aqueous conditions. *Int. J. Pharm*. **20**:307–314 (1984).
25. J. T. Wang, G. K. Shiu, O. C. Ting, C. T. Viswanathan, and J. P. Skelly. Effects of humidity and temperature on *in-vitro* dissolution of carbamazepine tablets. *J. Pharm. Sci*. **82**:1002–1005 (1993).