AAPS PharmSciTech, Vol. 12, No. 4, December 2011 (© 2011)
DOI: 10.1208/s12249-011-9700-4

Research Article

Theme: Quality by Design: Case Studies and Scientific Foundations
Guest Editors: Robin Bogner, James Drennen, Mansoor Khan, Cynthia Oksanen, and Gintaras Reklaitis

De-risking Pharmaceutical Tablet Manufacture Through Process Understanding, Latent Variable Modeling, and Optimization Technologies

Koji Muteki,[1,2] Vidya Swaminathan,[1] Sonja S. Sekulic,[1] and George L. Reid[1]

Received 20 May 2011; accepted 19 September 2011; published online 4 October 2011

Abstract. In pharmaceutical tablet manufacturing processes, a major source of disturbance affecting drug product quality is the (lot-to-lot) variability of the incoming raw materials. A novel modeling and process optimization strategy that compensates for raw material variability is presented. The approach involves building partial least squares models that combine raw material attributes and tablet process parameters and relate these to final tablet attributes. The resulting models are used in an optimization framework to then find optimal process parameters which can satisfy all the desired requirements for the final tablet attributes, subject to the incoming raw material lots. In order to de-risk the potential (lot-to-lot) variability of raw materials on the drug product quality, the effect of raw material lot variability on the final tablet attributes was investigated using a raw material database containing a large number of lots. In this way, the raw material variability, optimal process parameter space and tablet attributes are correlated with each other and offer the opportunity of simulating a variety of changes *in silico* without actually performing experiments. The connectivity obtained between the three sources of variability (materials, parameters, attributes) can be considered a design space consistent with Quality by Design principles, which is defined by the ICH-Q8 guidance (USDA 2006). The effectiveness of the methodologies is illustrated through a common industrial tablet manufacturing case study.

KEY WORDS: design space; latent variable modeling; partial least squares; process optimization; quality by design.

INTRODUCTION

The Quality by Design (QbD) concept proposed by the Food and Drug Administration plays a significant role in the development process within the pharmaceutical industry (1,2). The concept of QbD was mentioned in the ICH Q8 guidance (1), which states that "quality cannot be tested into products, that is, quality should be built in by design."

A key notion is to obtain a design space that is defined as a "multidimensional combination and interaction of input variables (raw material attributes and process parameters) that have been demonstrated to provide assurance of quality" by the ICH-Q8 guidance (1). This entails definition of (1) the key quality attributes of the finished drug product, (2) drug product formulation space that comprises the active pharmaceutical ingredient (API) and pharmacologically inactive components in the formulation, referred to as excipients that are selected to provide the desired physical and mechanical attributes in the drug product, and (3) the manufacturing process/operating parameter space to consistently produce a quality product. MacGregor and Bruwer (3) have provided a framework for the development of design and control spaces that simultaneously considers the raw material characteristic space, process variable space, and quality attribute space. This is a key concept to maximize the batch-to-batch consistency of pharmaceutical drug products.

In pharmaceutical manufacturing processes, a main source of disturbance which could affect the drug product quality is the variability of the incoming raw materials. A critical challenge for QbD is how to construct a process control strategy which can compensate for the effect of raw material variability and stabilize drug product quality. There are basically two different control schemes: Feedback control scheme to adjust the process parameters based on the error between desired final tablet attributes and observed attributes. This approach can handle both known and unknown disturbances, but always involves a delayed response and can result in a reduced yield. The feedforward (FF) control scheme adjusts the process parameters depending on the change in the incoming raw material attributes. The FF approach can compensate for the effect of raw material lot variability without a delayed response as it is performed prior to making a batch. In the case where the effect of an unknown disturbance is relatively large, the control performance (prediction) may be decreased due to not accounting for this in the model. This study focuses only on a FF control

[1] Pfizer Global Research & Development, Eastern Point Rd, Groton, Connecticut 06340, USA.
[2] To whom correspondence should be addressed. (e-mail: koji.muteki@pfizer.com)

scheme as the key attributes of raw materials have been well-investigated and known.

The basic concept of FF control is shown in Fig. 1. With FF control, the effect of incoming raw materials can be compensated by suitably adjusting process parameters prior to actually manufacturing a batch, while without FF control, the effect of raw material variability appears in the (variability of the) final tablet attributes. There are two key issues to design FF control: (1) how to build a multivariate model from raw material attributes and process parameters to final tablet attributes; (2) how to find the best process parameters which can compensate for the effect of incoming raw material variability (within constraints of the manufacturing equipment and desired final tablet attributes) prior to manufacturing a batch.

Latent variable modeling (LVM) methodologies (4–6) such as partial least squares (PLS) and principal component analysis (PCA) can play a key role to solve the first modeling problem. LVM has been well established as a useful modeling tool to construct a multivariate relationship using existing datasets or databases. LVM effectively reduces the large number of original variable dimensions into a lower dimensionality subspace, and provides good estimates for the *Y* variable(s). In addition, due to the nature of reduction of variable dimensions, LVM allows the building of a good model even with relatively small numbers of drug product batches. Wold *et al.* (7,8) proposed nonlinear PLS algorithms, and Lakshminarayanan *et al.* (9) presented a novel methodology to generate accurate nonlinear models using genetic programming where the relationships are nonlinear. There are precedents applying LVM for pharmaceutical tablet development. Westerhuis *et al.* (10) published the PLS modeling for simultaneously relating formulation compositions (Microcrystalline cellulose/Klucel) and process parameters (added water%, batch time) to some tablet physical attributes such as crushing strength and disintegration time in a wet granulation process. Huang *et al.* (11) and Haware *et al.* (12) demonstrated the use of LVM (PCA, PLS) to represent the multivariate relationship between material attributes, process parameters, and final tablet attributes, as a QbD case study in pharmaceutical development. The effectiveness of LVM has been demonstrated through a variety of industrial applications for pharmaceutical tablet development.

For the second problem of FF control design, optimization technologies (13) such as nonlinear programming are deemed to be suitable tools. They efficiently determine the best process parameters which can maximize/minimize a given objective function (*e.g.*, closeness to the desired final attributes, raw material cost, *etc.*) with constraints (*e.g.*, incoming raw material lot attributes, feasible process parameter range, *etc.*), based on the constructed model. Some industrial applications to chemical product manufacturing processes using optimization technologies with LVM have been presented in the literature. Yacoub and MacGregor (14) demonstrated an optimization approach using a nonlinear PLS model to a polymer product design (overmolding injection process). This approach was effective in finding process operating conditions that can compensate for variations in both raw material and environmental factors to reduce the variability of the final product quality. Garcia-Munoz *et al.* (15) provided an LVM and optimization strategy to establish a design space for the wet granulation process manufacture of a pharmaceutical product, simultaneously taking into account the raw material properties and process parameters. Muteki *et al.* (16,17) presented an optimization approach which can simultaneously optimize 3 degrees of freedom in a general mixture manufacturing process: the selection of raw materials, the selection of the ratios in which to blend the raw materials, and the selection of process conditions used to manufacture the final product. It was successfully applied to industrial applications such as polymer product design (blend product of rubber, polypropylene, and oil) (16) and coke product design (18). However, there have been relatively few published industrial applications of optimization technologies for QbD studies in the pharmaceutical drug product development arena.

This paper presents a process optimization strategy using PLS models, which can compensate for the effect of raw material (lot-to-lot) variability to achieve desired final tablet attributes in pharmaceutical dosage form development (tablet manufactured with a dry granulation process). The approach involves building PLS models that combine raw material attributes and tablet process parameters and relates these to final tablet attributes. Prior to a manufacturing batch, the resulting models are used in an optimization framework to

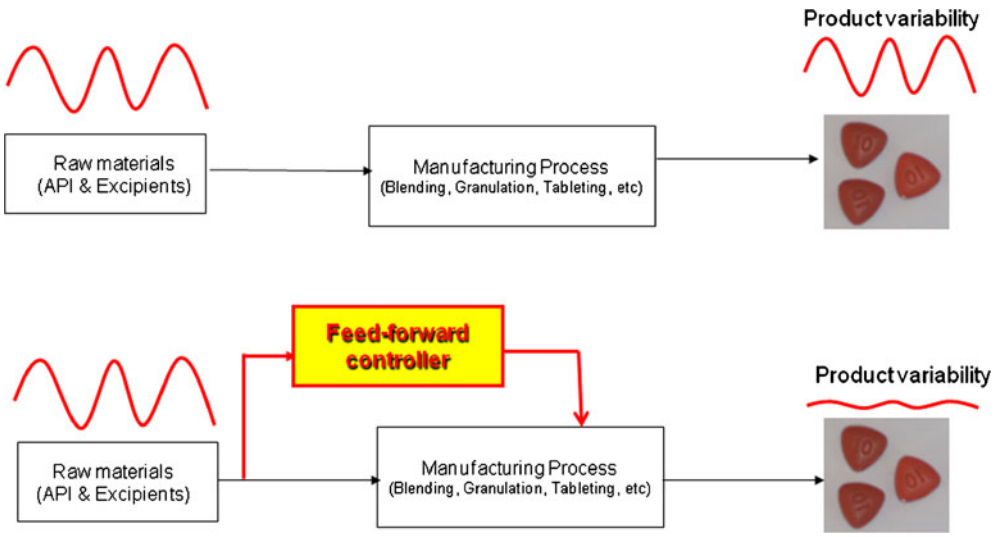

Fig. 1. Basic concept of feedforward control strategy

find optimal process conditions which can satisfy desired final tablet attributes while taking into account the attributes of specific raw material lots to be used in the batch. In order to reduce the impact of raw material (lot-to-lot) variability on drug product (lot-to-lot) consistency, the effect of variable raw material properties on the final tablet attributes was investigated using a raw material database based on a PLS model. Then, the optimal process parameter spaces taking into account the total raw material variability and manufacturing constraints to meet the desired final tablet attributes were found. The obtained process parameter space can be considered a QbD space as defined by the ICH-Q8 guidance (1). To implement the FF control, an in-house process simulator is presented. The effectiveness of the methodologies is illustrated through a common industrial tablet manufacturing case study.

This paper is organized as follows: In section "Materials and Methods," the generalized data structure used and the simultaneous PLS modeling taking into account both raw material attributes and process parameters on final tablet attributes are first described. A process optimization strategy using PLS is then presented. An industrial example of tablet development is described. In section "Results and Discussion," the result of PLS modeling, the raw material variability, and optimization simulation is demonstrated. An in-house process simulator for implementing the process optimization strategy is described.

MATERIALS AND METHODS

Partial Least Squares Modeling

In this section, the generalized data structure used is first shown, the simultaneous modeling using raw material attributes and process parameters on final tablet attributes is described, and the PLS modeling algorithm is briefly described.

Data Structure

The generalized data structure available in a general tablet development process is shown in Fig. 2. The data consist of raw material attributes database (X_{DB}) and data from batch manufacturing operations (X, Z, and Y). The latter consists of a ($M \times K$) X matrix consisting of K attributes measured on raw materials, a ($M \times J$) Z matrix consisting of J process parameters used during tablet manufacture and a ($M \times L$) Y matrix consisting of L attributes measured on the final tablets. M is the number of batches. X_{DB} is the ($N \times K$) database matrix containing attributes on all the available raw materials, including both lots used and not used in the past tablet batch manufacturing operations for a particular product. The X matrix is a subset of X_{DB}. Some attribute data in the X_{DB} matrix can be obtained from suppliers of the raw materials (Certificates of Analysis). Additional raw material characterization may be required to develop more robust relationships.

Simultaneous PLS Modeling Using Raw Material Attributes and Process Parameters on Final Tablet Attributes

The key question in modeling is how to build the model relating raw material attributes (X) and process parameters (Z) to the final tablet attributes (Y). The PLS model which simultaneously takes into account 2 degrees of freedom such as raw material attributes (X) and process parameters (Z) is generically expressed as:

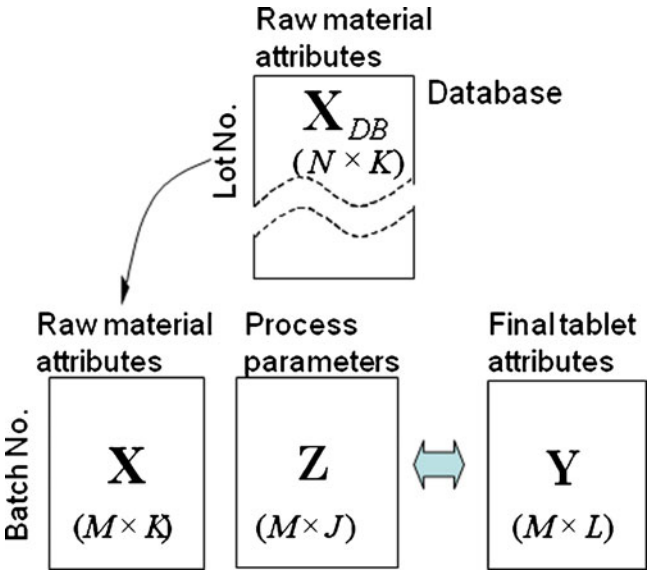

Fig. 2. Generalized data structure available in tablet development

$$Y = f(X, Z) + \varepsilon \tag{1}$$

where ε is the model error including measurement error. If one wants to consider the effect of formulation composition (R) as well as raw material attributes (X) and process parameters (Z), mixture-property PLS models (16) can simultaneously take into account the 3 degrees of freedom for a general chemical mixture product design and can be used.

There are several advantages to the PLS models shown in Eq. 1: (1) It can provide information on which of the raw material attributes and process parameters have more relative impact on final tablet attributes; (2) With a new raw material lot/grade and process parameters which have never been used in the past, the PLS models can estimate the final tablet attributes *in silico* prior to any manufacturing or experimentation experience; (3) the built PLS models can be suitably applied in an optimization framework to find the best process parameters which can achieve desired tablet attributes subject to incoming raw material lot variations.

PLS Regression

PLS regression modeling has extensively been described in the literature (4–6), and only a brief description is given here. PLS regression is performed by projecting the X_{all}=[X, Z] and Yin Eq. 1 onto lower dimensional subspaces:

$$\begin{aligned} X_{all} &= T \times P^{T} + E \\ Y &= T \times Q^{T} + F \end{aligned} \tag{2}$$

where the columns of T are values of latent variables ($T = X_{all} \times W^{*}$) that capture most of the variability in the data; W^{*}, P, and Q are the loading matrices, and E and F are residual

matrices. The PLS loading matrices are obtained by maximizing the covariance between X_{all} and Y (5). Prediction of Y can be obtained from the PLS model as

$$\widehat{Y} = T \times Q^{\mathrm{T}} = X_{\text{all}} \times \left(W^* \times Q^{\mathrm{T}}\right) = X_{\text{all}} \times \widehat{B} \tag{3}$$

For any new raw material attributes and process parameters $x^T_{\text{all new}}$, one can compute new latent variable scores as $\tau^T_{\text{new}} = x^T_{\text{all new}} W^*$ and then predict final tablet attributes as $\hat{y}^T_{\text{new}} = \tau^T_{\text{new}} Q^T$. One can also compute two distance criteria to test the validity of the model for the new conditions. The Hotelling T^2 is expressed as

$$T^2 = \sum_{a=1}^{A} \frac{\tau^2_{\text{new},a}}{s^2_a} \tag{4}$$

where s^2_a is the variance of the a-th latent variable score vector in the matrix and A is the selected number of latent variables in the PLS model. It provides a measure of the distance from the center point in the latent space to the projection of the new observation onto the latent variable space. The squared prediction error (SPE) in the X space is expressed as

$$\text{SPE}_X = \sum \left(x_{\text{all new}} - \widehat{x}_{\text{all new}}\right)^2 \tag{5}$$

where $\widehat{x}^T_{\text{all new}} = x^T_{\text{all new}} W^* P^T$ is the predicted value of $x^T_{\text{all new}}$ estimated from the PLS model. The SPE provides a measure of the orthogonal distance (residual) of the new point from the latent variable space. A large residual implies that the PLS model is not valid in the region of $x^T_{\text{all new}}$, due to a breakdown in the correlation structure represented by the model.

Variable influence on projection (VIP) score (19) is a useful indicator to effectively find which variables/parameters have more impact on Y variables. The VIP score for the j-th variable can be calculated by Eq. 6. The average of squared VIP scores equals 1.0, and in general if VIP_j is greater than 1.0, the j-th variable has larger impact.

$$VIP_j = \sqrt{\frac{n \sum_{k=1}^{h} \left(q_k^2 t_k^t t_k \cdot \left(w_{jk}/\|w_k\|\right)^2\right)}{\sum_{k=1}^{h} \left(q_k^2 t_k^t t_k\right)}} \tag{6}$$

where q_k, t_k and w_k are the k-column vector of Q, T, W^*, n is the column number of X_{all} and h is the number of latent variables. The VIP score is particularly useful for the case where multiple Y variables are used for the modeling.

Process Optimization Strategy Using PLS Models

In this section, it is assumed that the required data are available and a PLS model between X_{all} and the final tablet attribute Y has been built. The objective is now to use this model to find the best process parameters z_{new} which can satisfy the desired final tablet attributes y_{des} with constraints subject to the incoming raw material lots x_{current}. The optimization objectives are expressed as:

Process condition →

$$\overbrace{Minimize \quad (y_{des} - \hat{y}_{PLS})^T \cdot W \cdot (y_{des} - \hat{y}_{PLS})}^{\textbf{Closeness to target final tablet attribute}}$$

$$\underbrace{\quad}_{z_{new}}$$

$$s.t.$$

$$\textbf{PLS estimation} \quad \begin{cases} \hat{y}_{PLS} = B^T_{PLS} \cdot x_{newall} \\ x_{newall} = \begin{bmatrix} x_{current} & z_{new} \end{bmatrix} \end{cases}$$

$$\textbf{Final attribute constraint} \quad \left\{ Lo_{PLS,l} \le \hat{y}_{PLS} \le Hi_{PLS,l} \right.$$

$$\textbf{Process parameter constraint} \quad \left\{ Lo_{z,j} \le z_{new,j} \le Hi_{z,j} \right.$$

$$\textbf{PLS model constraint} \quad \begin{cases} SPE_{new} = \sum_{i=1}^{i=(K+J)} (x_{newall} - \hat{x}_{newall})^2 \le \varepsilon \\ T_{new}^{\ 2} = \sum_{a=1}^{A} \frac{\tau^2_{new,a}}{s_a} \le T^2_{\max} \end{cases}$$

(7)

where the $(1\times J)$ z_{new} vector is the new process condition, the $(1\times L)$ y_{des} vector refers to the desired final attributes, W is a diagonal weighting matrix providing the relative importance of each final attribute, $Lo_{PLS,l}$ and $Hi_{PLS,l}$ are lower and upper limits on the final attribute $\hat{y}_{PLS}$, $Lo_{z,j}$ and $Hi_{z,j}$ are lower and upper limits on the new process conditions z_{new}, the $x_{current}$ is the attribute of the incoming raw material lots. The optimized variables are the new process conditions z_{new}.

The constraints on SPE_{new} and Hotelling T^2_{new} forces the solution to lie in the space of the PLS model. The T^2_{max} value in the T^2_{new} constraint may be taken as the 95% or 99% limit on T^2 from the training data depending on how far from the training data one is willing to extrapolate. In order to allow the prediction greater flexibility due to compensation for the raw material variability, the values T^2_{max} of the Hotelling's T^2_{new} can be set to relatively large values (*e.g.*, it is recommended that the value corresponds to the 99% confidence limit). However, it should not be too large because nonlinearity or model lack of fit may impact predictions outside this region. The ε value in the SPE constraint can range from zero (perfect adherence to the model) up to some larger value such as the 95% limit on the SPE values from the training data. A larger ε value may be needed when constraints are being imposed on the elements of y_{new} and $x_{new\ all}$ to allow for the possibility of slight extrapolations of the model in order to satisfy these.

The objective function is a weighted measure of the estimation error between the desired tablet attributes and the estimated tablet attributes through the PLS model. Here, only one term is used in the objective function, but some other cost functions such as manufacturing process utility cost can be easily added, if needed.

Solutions to these nonlinear programming problems were obtained using the sequential quadratic programming approach (13,20). MATLAB was used for the execution. In the practical implementation, arbitrary diagonal weighting matrix W is suitably selected (for example, [1, 0; 0, 1], [100, 0; 0, 1]) to obtain multiple (various) solutions while still satisfying all the constraints. In this way, the operational design space (region) can be constructed. The detailed results will be described in section "Results and Discussion."

During practical computation, there would be two possible cases in which the above optimization formulation does not work: (1) The solution cannot be unique, that is, there remains unlimited number of solutions; (2) there is no solution satisfying all the constraints. In case 1, a unique solution can be obtained by modifying objective function or hard constraints (in Eq. 7) so as to further satisfy more severe requirements. For example, the weighting matrix W can be changed to prioritize more important tablet properties, and/or additional penalty terms such as cost function can be added into objective function, and/or the desired range of final tablet attributes can be narrowed to enhance their robustness. In case 2, inversely, by relaxing hard constraints, some feasible solution could be obtained. In most practical cases, there may still be a room to modify the hard constraints, because they tend to be often determined at conservative side during defining the optimization framework.

Industrial Tablet Development Example

The effectiveness of the above methodologies is illustrated through an industrial tablet manufacturing case study. These immediate release tablets are comprised of an API and four excipients (*i.e.*, microcrystalline cellulose (MCC), fast flo lactose (FFL), sodium starch glycolate (SSG), and magnesium stearate (MgSt)). The attributes of these pure raw materials modeled in the study were particle size distribution, water content, specific surface area, and acid content. Only the attributes of MCC, FFL, and MgSt have been taken into account in the final modeling and optimization stages, because they were shown to affect tablet attributes in a preliminary study undertaken to determine an appropriate set of raw material attributes for the final models (variable selection).

The tablet manufacturing process was a common dry granulation process consisting of blending, roller compaction, milling, and tableting unit operations. The process parameters evaluated in this study were roll force, roll speed, screen aperture size, blender load during lubrication, tablet compaction force and speed. The final tablet attributes used for this example were tablet hardness (strength) profile with compaction force, *in vitro* tablet dissolution, and tablet disintegration time.

The data matrix size and structure used in this study are described in Fig. 3 and consists of the database of batch operations and the database of raw material lots. The attributes of MCC, FFL, and MgSt lots used for the past manufacturing batches for this product are represented as X_{MCC}, X_{FFL}, and X_{MgSt}, respectively. The overall raw material database for MCC, FFL, and MgSt are shown as $X_{MCC\ DB}$, $X_{FFL\ DB}$, and $X_{MgSt\ DB}$, respectively. The data space of raw material lots used for the past operations (X_{MCC}, X_{FFL}, and X_{MgSt}) well covered that of all the possible lots ($X_{MCC\ DB}$, $X_{FFL\ DB}$, and $X_{MgSt\ DB}$). The range of process parameters (Z) has been widely investigated by a design of experiments. The high and low levels of each process parameter in the design of experiments were determined based on the maximum of the possible operational range. The tablet hardness profile, dissolution profile, and disintegration time are shown as Y_{HD}, Y_{Disso}, and Y_{DT}, respectively. A PLS model is built between the combined matrices (X_{all}=[X_{MCC}, X_{FFL}, X_{MgSt}, Z]) and the final tablet attributes (Y=[Y_{HD}, Y_{Disso}, Y_{DT}]). The total number of X and Y variables on the final tablet attributes is shown in Table I. Note one additional X variable (tableting speed) is included in the dissolution profile as compared to tableting strength and disintegration time.

One may consider that the intermediate quality attributes (*e.g.*, particle size distribution and density of lubrication process) would be useful to predict final tablet attributes. However, they were not included in the model as the purpose of modeling was to find the optimal process parameters of each unit operation prior to any manufacturing operations, rather than the prediction of final tablet attributes. As such, these intermediate quality attributes were not included in the model.

The total number of original variable dimensions of X_{all} and Y is generally much larger than that of the number of drug product manufacturing batches. This often occurs because the number of larger-scale batches tends to be small due to operation cost and the total number of variables (*e.g.*, raw material lots and attributes, process parameters, final tablet attributes) tends to be relatively large due to the

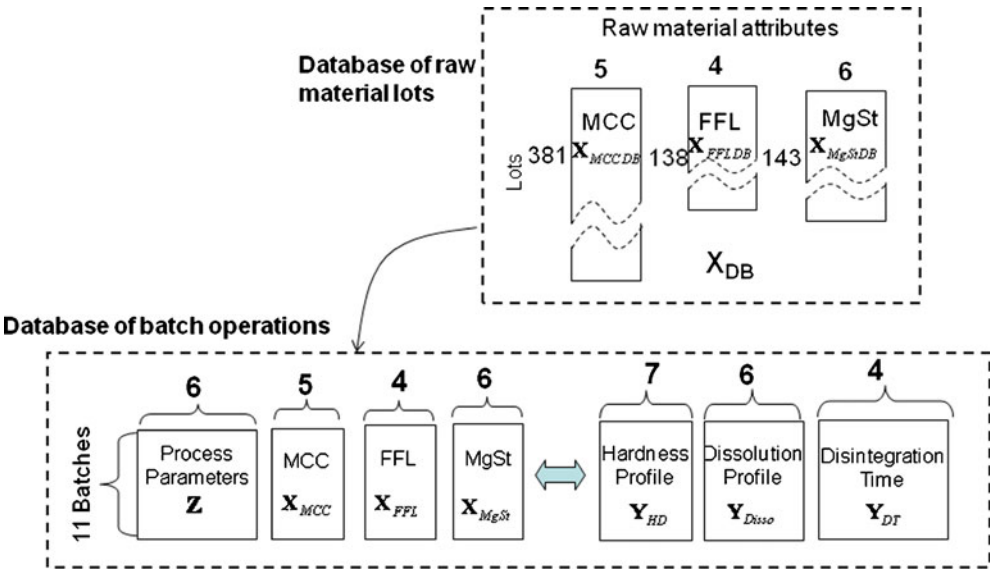

Fig. 3. Data structure used in the industrial example

complex nature of the drug product process. For such cases, common statistical tools such as ordinary least squares (OLS) will rarely work. The number of samples has to be (at least) two or three times larger than that of the number of columns/variables for successful OLS regression (even with independent X variables), or else, the OLS model would easily result in overfitting. PLS has the potential to overcome this problem by reducing the original dimensions into a much lower dimensionality subspace (*i.e.*, PLS components), and the PLS components may be small enough relative to the number of manufacturing batches. If so, the resulting PLS model would provide robust estimates.

RESULTS AND DISCUSSION

In this section, the result of PLS model ability is described first, the effect of all the possible raw material lot variability on tablet hardness profile is shown second, and the result of the optimal process parameter space and final tablet attribute space taking into account their total raw material lot variability is described last.

Result of PLS Model Ability

The ability of the PLS models to account for each tablet attribute is shown in Table I. The R^2 shows the fraction of the cumulative sum of squares of the response variables explained by the fitted model, and the Q^2 shows the cumulative sum of squares of prediction obtained from cross-validation. The number of PLS components for each model was selected so as to maximize the Q^2 value and avoid the overfitting problem. Both R^2 and Q^2 are generally important indicators for large samples of datasets.

As can be seen in Table I, the built PLS models provide good estimates of all the Y variables such as tablet hardness profile, dissolution profile and disintegration time (R^2 value of all the Y variables is more than 80%). The Q^2 value of the tablet strength results in being slightly lower while that of disintegration time and dissolution profile is sufficiently large. This occurs because the DOE data used in this study had independent relationship among X variables without the repeated batch samples (due to the limited data from drug development), and thus became sensitive to the result of cross-validation on the tablet strength. However, the prediction (validation) error of tablet strength was determined to be acceptable from another validation study (not shown in this manuscript). Furthermore, in the manufacturing (commercial) phase, the PLS model on tablet strength may require updating (model maintenance) as additional batch data is generated.

Notice that the number of PLS components used is much smaller than that of original dimensions of X_{all} and Y matrices. This indicates that the original variables were highly correlated to each other. Moreover, the number of PLS components is small relative to the number of manufacturing batches, while the total number of original variable dimensions of X_{all} and Y was much larger than the number of manufacturing batches. Therefore, the dimension reduction of PLS modeling was a significant advantage for this modeling study. As mentioned above, the data structure and

Table I. Result of PLS Modeling

	R^2(%)	Q^2(%)	The number of PLS components	The number of X variables	The number of Y variables
Tablet strength profile	89.8	35.4	4	20	7
In vitro dissolution profile	84.4	65.5	3	21	6
In vitro disintegration time	88.4	41.4	3	20	4

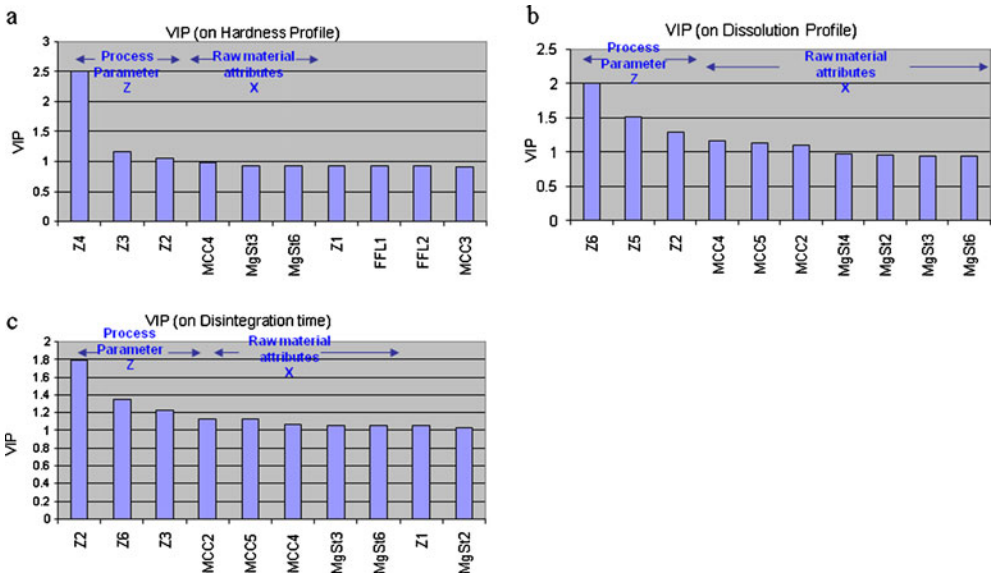

Fig. 4. Result of variable influence of projection (VIP): **a** VIP on hardness profile; **b** VIP on dissolution profile; **c** VIP on disintegration time (*Z1–Z6* process parameters, *MCC1–MCC5* MCC attributes, *FFL1–FFL4* FFL attributes, *MgSt1–MgSt6* MgSt attributes)

process nature of this study is a very common case in general pharmaceutical tablet development. The use of PLS modeling may not be just an option but an imperative for a QbD study which has to simultaneously take into account the relationship of many variables such as the raw material attributes, process parameters, and final tablet attributes with a relatively small number of drug product batches.

The PLS model can provide the information about which raw material attributes and process parameters have more impact on final tablet attributes. In order to recognize the overall key parameters, the results of the top ten VIP plots on the tablet hardness profile, dissolution profile, and disintegration time are shown in Fig. 4a–c, respectively. As mentioned previously, the process parameters and raw material attributes having a VIP>1.0 significantly contribute to the final tablet attributes. There are two key results: (1) in general, the process parameters contribute a larger effect on the final tablet attributes, compared to the raw material attributes; (b) the key process parameters and raw material attributes are different depending on the final attribute (*i.e.*, the tablet hardness profile, dissolution profile, and disintegration time). The VIP rankings on each final tablet attribute matched experimental observations.

Effect of Raw Material Lot-to-Lot Variability on the Tablet Hardness Profile

Based on the built PLS model, the effect of all the possible variability from excipient material lots on the final tablet attributes has been investigated using the excipient material lot database (X_{MCCDB} X_{FFLDB} $X_{\mathrm{MgSt\ DB}}$). The total number of combinations of excipient material lots in $X_{\mathrm{MCC\ DB}}$ $X_{\mathrm{FFL\ DB}}$ and $X_{\mathrm{MgSt\ DB}}$ was 7,518,654 (=381×138×143). For an efficient simulation, only the excipient lots that maximize variance in each excipient space (*e.g.*, PC1 *vs.* PC2 by PCA) were chosen. During the simulation, a nominal (fixed) process condition was used. In this simulation, the tablet hardness profile was primarily investigated as it has the biggest contribution to the performance of this drug product. The result of this lot-to-lot variability on the tablet hardness profile is shown in Fig. 5. The wider region (colored yellow) and the smaller region (colored green) correspond to the total variability of all the possible excipient material lots and the target region, respectively. Most batches resulted in being within the target region, but some batches were outside the target region. Therefore, a process optimization approach which takes into account the excipient lot-to-lot variability could provide additional process understanding for this case, could help to de-risk the overall manufacturing operations for this product and produce a more consistent drug product.

Result of Process Optimization

The effect of all the possible excipient lot-to-lot variability on the final tablet attributes (colored by yellow in Fig. 5) should be taken into account in the process optimiza-

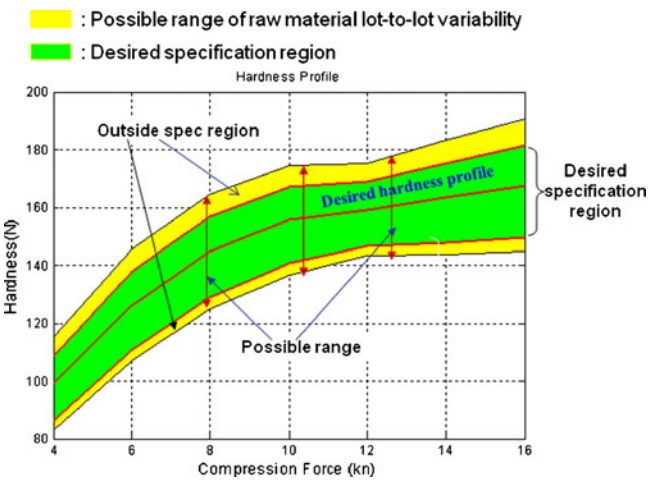

Fig. 5. Result of raw material lot-to-lot variability on tablet hardness profile

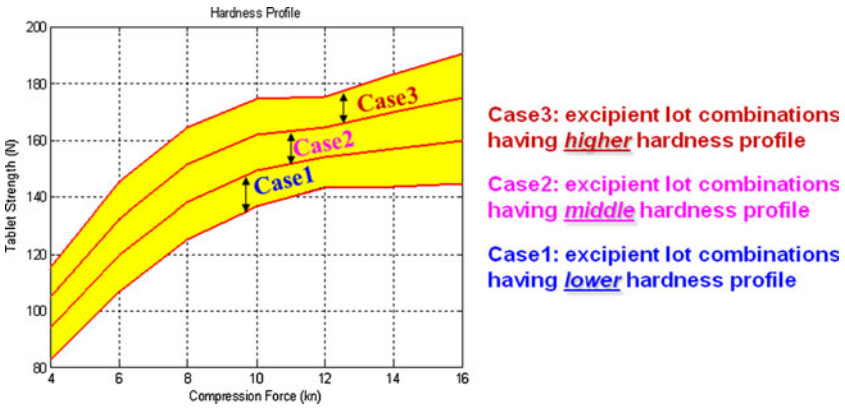

Fig. 6. Excipient lot combination cases corresponding to three levels of hardness profile (high, middle, low)

tion design. The range is first divided into three case regions (case 1, case 2, case 3) which corresponds to lower, middle, and higher tablet hardness profile, as shown in Fig. 6, in order to clearly demonstrate the change (transition) of optimal process parameters and final tablet attributes subject to the excipient lot variability. The upper region of case 3 and the lower region of case 1 shown in Fig. 6 fall outside the target region, as shown in Fig. 5, and are of particular interest. The excipient lot combinations corresponding to the above three cases were found through the simulation in the previous section, and were set as the constraints (x_{current} in Eq. 7) in the optimization framework. The centered line in the target region shown in Fig. 5 is set as the target tablet hardness profile y_{des} in Eq. 7. The other constraints such as process parameter ranges, dissolution profile, and disintegration time were suitably set.

The result of the process parameter (regions) optimization subject to the excipient lot combinations (corresponding to cases 1, 2, and 3) is shown in Fig. 7. Only the results of three key process parameters Z2, Z3, and Z6 are shown here, while the result of six process parameters is simultaneously obtained. The range of process parameter constraints is shown by the green line. Notice that the optimal process parameter region clearly changes depending on the excipient lot combination cases (cases 1, 2, and 3). This indicates and confirms that the process parameters can be changed to compensate for the effect of the excipient lot changes. The reason why the optimal process parameter regions (cases 1, 2, and 3) discretely change is because they are determined so as to minimize the objective function in Eq. 7 (*i.e.*, the error between the desired tablet hardness profile and the estimate) at each excipient combination case. If defining another objective function in the optimization framework (*i.e.*, another criteria), another optimal process parameter region will be obtained.

The result of tablet hardness profile, dissolution profile, and disintegration time using these optimal process parameters is simulated and shown in Fig. 8. All the desired final tablet attributes are well satisfied for all the possible excipient lot combinations (cases 1, 2, and 3), by selecting the optimal process parameters. This means that the process optimization has compensated for the effect of raw material lot variability and even excipient lots that fell outside of the target region with a fixed process produced acceptable tablets. This is a tangible benefit provided by the proposed process optimization strategy. There is still some room on the constraints of the desired dissolution profile and disintegration time, while there is not much on the tablet hardness profile (especially case 3). Such information enhanced the process understanding of design space during drug development.

The above results (Figs. 7 and 8) can be considered as the suitable design space for a QbD study. The optimum

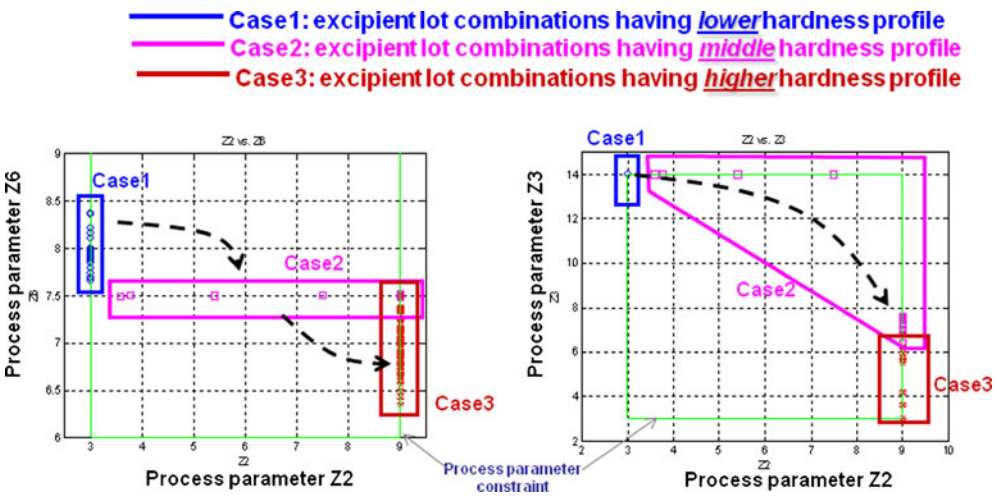

Fig. 7. Optimal process parameter region subject to excipient lot *Case 1–Case 3*

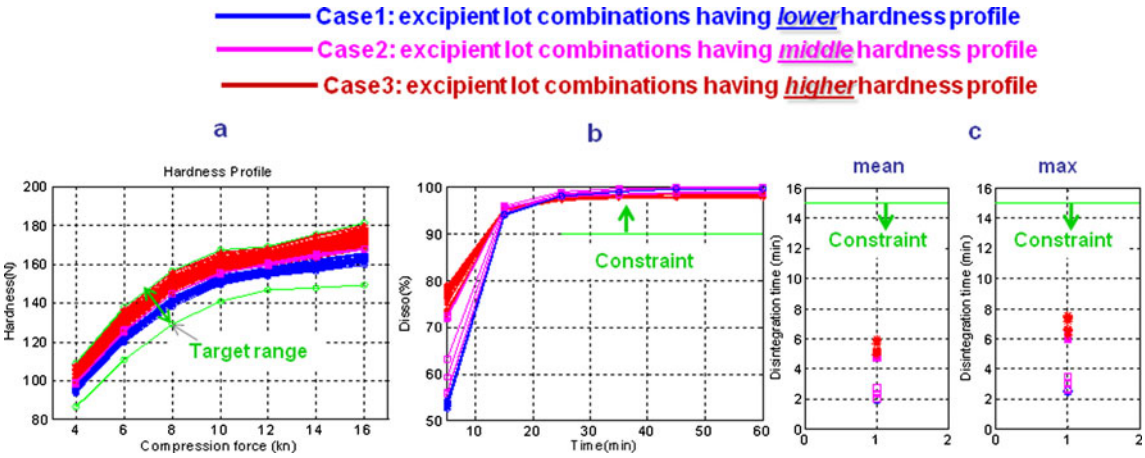

Fig. 8. Optimal process parameter region subject to excipient lot Case 1–Case 3: **a** Tablet hardness profile, **b** dissolution profile, **c** Disintegration time

process parameter space to achieve the desired final attributes may change with the variance in the raw material attributes. These studies provide a good "multidimensional and interactive relationship of raw material attributes and process parameters on final tablet attributes," which is defined by the ICH-Q8 guidance (1). The above methodologies using latent variable modeling and optimization technology can play a key role in presenting a suitable design space for a QbD study.

In-house Process Simulator

To aid operators in the selection of process parameters per excipient characteristics, an in-house process simulator for the proposed process optimization strategy has been developed using MATLAB. The simulator calculates the best process parameters within the selected constraints (*e.g.*, incoming raw material lots and process parameter ranges) to ensure desired final tablet attributes are obtained for every batch, prior to manufacturing a batch.

The screen shot of the process simulator is shown in Fig. 9. The column portions colored in blue and yellow show the input and output information, respectively. As input information, the incoming excipient lot (upper-left side of the screen) and the process parameter range constraints (lower-left side of the screen) can be selected. After pressing the "optimization" button, the optimal process parameters are provided in the yellow colored column (lower-left side of the screen), and the estimated final tablet attributes (tablet hardness profile, dissolution profile, and disintegration time) using the optimal process parameters are provided (right side of the screen).

The optimal process parameters can be predicted prior to a manufacturing operation through the process simulator. The simulator serves as a decision support system for the operators.

CONCLUSION

A latent variable modeling and process optimization strategy, which can compensate for the effect of raw material lot-to-lot variability to achieve the desired final tablet attributes, has been successfully applied to a common

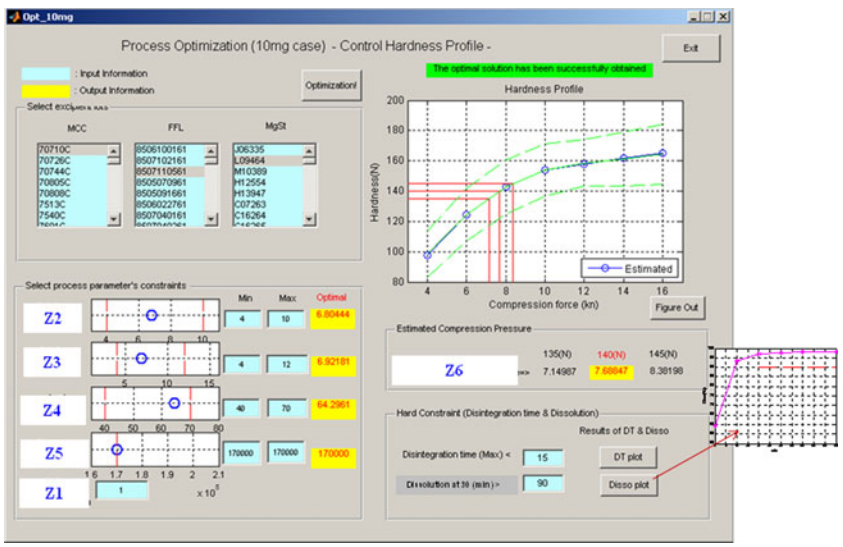

Fig. 9. Screen shot of process simulator for feedforward control

industrial manufacturing (pharmaceutical tablet) process. The PLS model simultaneously takes into account raw material attributes and process parameters and provided very good estimates on the final tablet attributes (tablet hardness profile, dissolution profile, disintegration time). The dimensionality reduction capability of PLS modeling was a significant advantage in this study, as the total number of original dimensions of X_{all} and Y is much larger relative to the total number of batches available. Many other common industrial cases would have a similar situation. The use of PLS modeling provides comparable advantages for QbD studies which have to simultaneously take into account the relationship of many variables such as the raw material attributes, process parameters, and final tablet attributes. The constructed PLS model provided useful information about which raw material attributes and process parameters have more impact on the final tablet attributes. Based on the PLS model, the total variability of raw material lots on the final tablet attributes was investigated using the raw material database, and then the optimal process parameter spaces (taking into account the total raw material variability to ensure desired final tablet attributes) was found. These results can be considered as a design space for a QbD study. This approach is able to demonstrate a multidimensional and interactive relationship of raw material attributes and process parameters on final tablet attributes, which is defined by the ICH-Q8 guidance (1).

The above methodologies using latent variable modeling and optimization technology can play a key role to demonstrate a suitable design space for a QbD study. An in-house process simulator for implementing the proposed process optimization strategy has been presented.

The modeling approach offers the following advantages over current approaches:

1. It provides the specific processing parameter combination(s) for a given combination of raw materials for manufacture of the product, whether in batch mode as discussed above, or in continuous processing mode;
2. Various combinations of raw material and processing variables can be simulated to assess its impact on finished product attributes. In effect, the entire operating space can be sampled (simulated) using the model. This capability is useful in the risk-based approach to managing quality risks; combinations of parameters that can result in unacceptable product are flagged prior to batch manufacture.
3. Its ability to accommodate and take into account variability in raw material on product attributes. It is during commercial manufacture that the effect of variability in input materials on product attributes becomes evident; it may be argued that the design space is fully defined and validated well into commercial manufacture. Accordingly, the preliminary model built from product development data can be refined and its accuracy improved with data obtained during commercial manufacture. Further areas of study are in the validation of the model through its evolution (refinement).

An additional benefit of the modeling approach and simulation is the ability to detect and track changes in excipient characteristics over time and to assess the effect of the changes on product quality. The information from the simulation can be used in setting meaningful specifications on incoming materials and in assessing excipients from new sources. A related point worth mention is that since there are a limited number of excipients for a given pharmaceutical dosage form and processing stream, the model and the understanding of raw material and manufacturing process may be translatable across products in a given category, thus enabling more efficient product development. The ultimate value of the modeling approach is that it helps us better understand the effect of raw material and the manufacturing process on the product.

ACKNOWLEDGMENTS

The authors wish to thank Jackson Pellett, Joseph Kushner, Jeff Moriarty, Lin Zhang, Andreas Muehlenfeld, Kirsten Reinheimmer in Pfizer Inc. for providing the data and useful discussion.

REFERENCES

1. USDA (2006), Guidance for industry: Q8 pharmaceutical development, Office of training and communication, division of drug information, HFD-240, center for drug evaluation and research, Food and Drug Administration, 5600 Fishers Lane, Rockville, MD 20857, USA.
2. Yu L.X. Pharmaceutical Quality by Design: Product and Process Development, Understanding and Control, Pharm. Res.Vol.25 (2008) No.4.
3. MacGregor JF, Bruwer MA. Framework for the development of design and control spaces. J Pharm Innov. 2008;3:15–22.
4. Martens H, Tormod N. Multivariate calibration, Wiley & Sons, (1991).
5. Höskuldsson A. PLS regression methods. J Chemom. 1988;2:211–28.
6. Burnham AJ, MacGregor JF, Viveros R. Frameworks for latent variable regression. J Chemom. 1996;10:31–45.
7. Wold S, Kettaneh-Wold N, Skagerberg B. Nonlinear PLS modeling. Chemom Intell Lab Syst. 1989;7:53–65.
8. Wold S. Nonlinear partial least squaers modeling: spline inner relation. Chemom Intell Lab Syst. 1992;14:71–84.
9. Lakshminarayanan S, Fujii H, Grosman B, Dassau E, Lewin DR. New product design via analysis of historical databases. Comput Chem Eng. 2000;24:671–6.
10. Westerhuis J, Coenegracht PMJ, Lerk CF. Multivariate modeling of the tablet manufacturing process with wet granulation for tablet optimization and in-process control. Int J Pharm. 1997;156:109–17.
11. Huang J, Kaul G, Cai C, Chatlapalli R, Hernandez-Abad P, Ghosh K, *et al.* Quality by design case study: an integrated multivariate approach to drug product and process development. Int J Pharm. 2009;382:23–32.
12. Haware RV, Tho I, Bauer-Brandl A. Multivariate analysis of relationship between material properties, process parameters and tablet tensile strength for α-lactose monohydrates. Eur J Pharm Biopharm. 2009;73(3):424–31.
13. Edgar T.F, Himmelblau D.M, Mautner D. Optimization of Chemical Processes, New York, 1988.
14. Yacoub F, MacGregor JF. Product optimization and control in the latent variable space of nonlinear PLS models. Chemom Intell Lab Syst. 2004;70(1):63–74.
15. Garcia-Munoz S, Dolph S, Ward HW II. Handling uncertainty in the establishment of a design space for the manufacture of a pharmaceutical product. Comput Chem Eng. 2010;34:1098–107.

16. Muteki K, MacGregor JF, Ueda T. Rapid development of new polymer blends: the optimal selection of materials and blend ratios. Ind Eng Chem Res. 2006;45(13):4653–60.
17. Muteki K, MacGregor JF, Ueda T. Mixture designs and models for the simultaneous selection of ingredients and their ratios. Chemom Intell Lab Syst. 2007;86(1):17–25.
18. Muteki K, MacGregor JF. Optimal purchasing of raw materials: a data-driven approach. AICHE J. 2008;54(6):1554–9.
19. Wold S, Johansson E, Cocchi M. 3D QSAR in drug design: theory, methods, and applications. Leiden: ESCOM; 1993. p. 523–50.
20. Boggs P.T, Tolle J.W. Sequential quadratic programming, *Acta Numerica*. 1995; pp.1–51.