International Journal of Pharmaceutics 418 (2011) 235–242

Contents lists available at ScienceDirect

International Journal of Pharmaceutics

journal homepage: www.elsevier.com/locate/ijpharm

A framework for *in-silico* formulation design using multivariate latent variable regression methods

Mark A. Polizzi*, Salvador García-Muñoz

Pfizer Global Research and Development, Eastern Point Road, Groton, CT 06340, USA

ARTICLE INFO

Article history:
Received 10 November 2010
Received in revised form 1 April 2011
Accepted 26 April 2011
Available online 5 May 2011

Keywords:
PCA
PLS
Mixture modeling
Blending

ABSTRACT

A comprehensive Quality by Design development paradigm should consider the impact of raw materials and formulation on the final drug product. This work proposes a quantitative approach to simultaneously predict particle, powder, and compact mechanical properties of a pharmaceutical blend, based on that of the raw materials. A new, two-step, multivariate modeling method, referred to as the weighted scores PLS, was developed to address the challenge of predicting the properties of a powder blend while enabling process understanding. The model validation exercise is shown along with selected practical applications. It is shown how the proposed *in-silico* model exhibits sufficient predictive power to be an important tool in the pharmaceutical development decision making process while requiring minimal experimentation and material usage.

1. Introduction

1.1. A Quality by Design perspective

The pharmaceutical industry is undergoing a shift in the way drug products and processes are being designed and operated. This transformation of the pharmaceutical development process is triggered by the guidance documents released by the Food and Drug Administration (FDA, 2000, 2009). These guidelines opened the option to design and validate a given process over a range of process conditions, referred to as a design space, as opposed to a fixed set of points. New drug application submissions wanting to take advantage of such an option are expected to provide evidence that the established design space will yield a product with safety and efficacy for the patient. This concept is widely known as Quality by Design (QbD). In principle, the scientific evidence required for a QbD submission would provide an understanding of the driving forces acting upon the complex network of interactions between materials, process, and drug product.

An important point to consider in a QbD exercise is how the raw material attributes impact a given process. This is especially true since QbD principles imply that quality should be built into the process through an upstream design approach, rather than by down-stream troubleshooting. As the choice of formulation is the initial design space decision, it is important to have a comprehensive approach for selecting appropriate materials. Therefore, a QbD development project should consider the choice of excipients that are to be mixed with the active pharmaceutical ingredient (API) into the final drug product as an added degree of freedom. Multiple aspects of this decision should be considered. Specifically, the effect of the materials selected on the safety (e.g. chemical stability API), efficacy (e.g. impact on a drug release profile) and processability of the drug product (Hamad et al., 2010).

This work addresses the latter problem through a quantitative model that predicts the impact of the initial materials on the powder, flow, and mechanical properties for a blended formulation. This model is a key step towards enabling the *in-silico* design of the drug product by including material selection as a decision variable.

1.2. Modeling of mixtures from pure component properties

In general, modeling and predicting the output of mixing multiple individual components is a common problem in the scientific arena. Often products do not conform to simple ideal mixing rules and therefore require more sophisticated techniques. For example, extensive research in thermodynamics has addressed the effect on non-ideal mixing of liquids and gasses (Smith et al., 2005). Also, more fundamental mixing theories have been developed in the design of alloys in the metal and semiconductor field (Kumar, 2003) and to the design of advanced polymers (Bernardo et al., 1996; Kolarik et al., 2000). One method to address mixture modeling is to use group contributions; where the resulting properties of the blend are calculated as a function of the contributing groups present in each of the individual components. An example of this approach

* Corresponding author. Tel.: +1 860 686 2766; fax: +1 860 686 7423.
E-mail address: Mark.Polizzi@Pfizer.com (M.A. Polizzi).

0378-5173/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.ijpharm.2011.04.064

is the UNIQUAC and UNIFAC methods to calculate thermodynamic properties (Smith et al., 2005). In the pharmaceutical field, Cao et al. (2008), used element and ion volume contributions methods to predict the true density of active pharmaceutical ingredients. Also, organic structural group contributions were used to predict the refractive index of pharmaceutical solids (Cao et al., 2009).

For pharmaceutical powder blend prediction, black-box models, like expert systems (Shao et al., 2007), neural networks (Takayama et al., 2003; Kachrimanis et al., 2003), and neuro-fuzzy logic (Shao et al., 2007) have been proposed. However, most of these proposals do not go beyond an academic exercise and/or lack the transparency needed to understand the mechanics behind the prediction. As such, the use of these black-box models is unacceptable as part of a Quality by Design exercise which requires science-based understanding of the system being modeled.

A more quantitative approach to predicting the properties of a mixture is to use some simple mixing rules, such as weighted averages of the properties of the individual ingredients (either mass weighted, or volume weighted). This type of calculation assumes certain linearity in the mixture process. An example of this approach is the work by Wu et al. (2005, 2010), where mixing rules were used in conjunction with ryskewitch-ductworth equations (i.e. a logarithmic mass weighted average) to predict the tensile strength of multi-component mixtures of pharmaceutical powders.

Finally, the use of multivariate latent variable regression methods (MacGregor et al., 2005) has been applied to mixture modeling of pharmaceutical powders. Particle size distribution (PSD) has been shown to predict powder flow (Mullarney and Leyva, 2009) within a partial least squares (PLS) regression framework. PSDs along with particle shape information was used also used to predict granule packing and flow behavior (Sandler and Wilson, 2010). Finally, a weighted average of the properties of individual components augmented with dry granulation process conditions were used within a PLS framework to predict selected granulation properties (Soh et al., 2008). These multivariate models have the advantage of providing understanding of the mechanics behind the prediction which enables process understanding. For other applications of multivariate methods to pharmaceutical processes, the reader is referred to the reviews by Workman et al. (2009) and Gendrin et al. (2008).

The work presented herein combines the weighted averaging of individual component properties (which we will refer to as "ideal-mixing") with the application of multivariate models in a new, two-step, process. The weighted scores PLS (WSPLS) takes scores from a principal component analysis (PCA) of individual components and then weights them prior to PLS regression against measured blend data. This WSPLS method is used to simultaneously predict the resultant particle, powder, and compact mechanical properties of blends of pharmaceutical powders with minimal experimentation and API usage (<50 g). The predictive abilities of the WSPLS model are contrasted with that of logarithmic weighted average.

2. Materials and methods

2.1. Dataset creation

Since 1996 the Pfizer Materials Assessment Laboratory (MAL) has developed an extensive database of physical properties of pharmaceutical powders. These data were gathered to create a blend prediction model which is representative of the range of materials typically seen in solid dosage form development. The material types in the database include individual excipients and active pharmaceutical ingredients APIs, as well as formulated blends and granulations containing multiple components. The physical characterization tests on the materials were performed in the powder and compacted state to discern and estimate material performance during processing. A subset of this database was gathered and used to create a blend prediction model. This subset included material properties from 64 excipients, 107 APIs, and 25 blends.

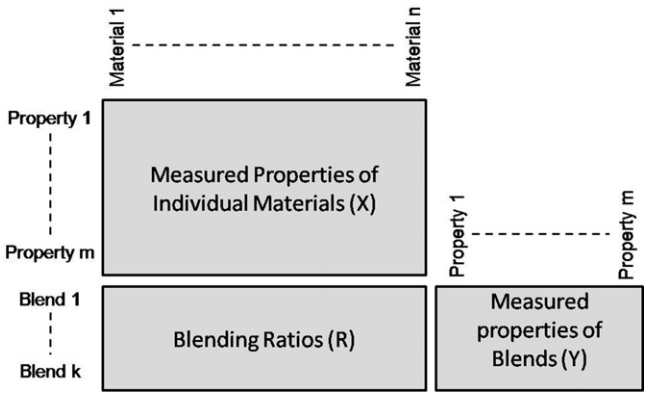

Fig. 1. Structure of mixture modeling data.

Examples of the powder-based material testing carried out included powder PSDs using a Sympactec/Rodos laser diffraction analyzer (Sympatec inc., Princeton, NJ), true densities via a helium pycnometer (Quantachrome inc., Boynton Beach, FL) and flow function coefficient using a Schulze Ring-Shear Tester. (Schultz, 1996) Several compact mechanical properties were also measured using standard methods. These properties included compression stress needed to achieve 0.85 solid fraction, compact tensile strength, and compact ductility via dynamic and quasi-static compact indentation (Hancock et al., 2002; Hiestand, 2003; Mullarney et al., 2003). Additionally, bonding/tableting indices (e.g. bonding index and elastic modulus) were directly calculated from the above properties and were also included in this analysis.(Hiestand and Smith, 1984; Hiestand, 1997) The bonding/tableting indices were included as they are known to capture interactions which have been deemed important for performance understanding (Hiestand, 1995). A complete list of the variables used in this effort are shown in Table 1. While all of these properties were included in the construction of the blend-prediction model presented herein, this report will focus on a few core properties which are considered most important for formulation design. These core properties, include compression stress, ductility, tensile strength (normal and compromised), brittle fracture index (loss in tensile strength with compromised compact), bonding index (tensile strength/ductility), viscoelasticity (dynamic ductility/quasi-static ductility), powder flow function coefficient, and particle size distribution. However, for a new blend, the model will provide a prediction of every property in the dataset which allows for the flexibility to focus on any material characteristic that may be important for a specific formulation or dosage form.

The material data outlined above were collected for individual excipients and APIs, as well as blends of both placebo and active formulations. The data were organized into an "L-shaped" mixture-modeling structure as shown in Fig. 1. The "*X*" matrix contains the properties of the individual materials (excipients and APIs). The "*R*" matrix consists of the mixing ratios that represent the compositions of each blended formulation. Finally, the "*Y*" matrix contains the physical property data for the tested blends. In order to create an appropriate dataset that adequately characterized the formulations, the dataset was refined using the following procedures.

- Materials in a blend were matched to the raw ingredients by lot number whenever possible. If a direct match by lot number was

Table 1
Material properties used in blend prediction.

Property type (measurement system)	Parameter	Units
Particle size (Laser Diffraction)	Full particle size distribution (48 size regimes)	μm
Powder density (He Pycnometer)	Absolute density	(g/cc)
Powder flow (Schultz Ring-Shear Tester)	Flow function coefficient (FFC)	None
Compact mechanical properties and bonding indices	Compact solid fraction (SF)	None
	Compression stress (CS)	MPa
	Pendulum impact dent radius (PIDCR)	cm
	Dynamic indentation hardness (H0)	MPa
	Strain index (SI)	NA
	Reduced elastic modulus (REM)	GPa
	Quasi-static Indentation radius (QCR)	cm
	Quasi-static hardness (H10)	Mpa
	Shear modulus (SHEARM)	GPa
	Tensile strength (TSreg)	MPa
	Compromised tensile strength (TScomp)	MPa
	Brittle fracture index (BFI)	None
	Best-case bonding index (Bib)	None
	Worst-case bonding index (Biw)	None
	Brittle-viscoelastic bonding index (bBIv)	None
	Mod. brittle-viscoelastic bonding index (bBIvmod)	None
	Viscoelastic number (VENUM)	None

not possible the ingredients were chosen using a material deemed to be representative (i.e. an excipient of the same manufacturer and grade).
- Blends in "*Y*" that did not have >97% of ingredients appropriately matched in "*X*" were removed.
- Materials that were not characterized for particle size distribution using laser diffraction were removed as particle size is known to have an important effect on many of the other material properties.
- Materials with >30% missing data were removed.
- Particle size distributions were transformed from a 48 variable probability distribution by size to a quantile distribution with 200 variables.

Following these steps the dataset contained a total of 64 commonly used excipients, 103 APIs, and 25 blends (both placebo and active formulations) that were well characterized for mechanical, flow, and particle size characteristics. These data are representative of materials used in a wide variety of solid-oral dosage forms including both immediate and controlled release as well as tablet and capsule formulations.

2.2. Modeling methods

A novel "Weighted-Scores PLS" (WSPLS) method was developed whereby the *X* data structure was first transformed to the score space using principal component analysis (PCA). The scores (**T**) from this PCA model Eq. (1) were then linearly weighted by the blending matrix, *R* Eq. (2), prior to PLS regression against the blend data, *Y* Eq. (3). In this case the data required special treatments according to the testing method used. The APIs were all tested using smaller (0.5 g, $3/8'' \times 3/8'' \times 3/16''$) compacts compared to the excipients (5.0 g, $3/4'' \times 3/4'' \times 3/8''$) and this caused some differences between the correlations within the two distinct populations. Therefore, separate PCAs were constructed for the APIs and excipients (six principal components each). For each individual blend, the scores of the excipients used in that blend were weighted and added according to the corresponding row in *R* and this vector was concatenated with that of the scores of the API used in that blend (also weighted by *R*). This collection of concatenated vectors is referred to as the "weighted scores". These "weighted-scores" were then regressed against the measured blend properties using a PLS.

$$\mathbf{X} = \mathbf{TP}^{\mathrm{T}} + \varepsilon_X \tag{1}$$

$$t_{(n,a)} = \sum_i t_{(i,a)} r_{(n,i)} \tag{2}$$

$$\begin{aligned} \mathbf{X} &= \mathbf{TP}^{\mathrm{T}} + \boldsymbol{\varepsilon}_X \\ \mathbf{Y} &= \mathbf{TQ}^{\mathrm{T}} + \boldsymbol{\varepsilon}_Y \end{aligned} \tag{3}$$

The quantile particle size distribution (PSD) was multi-blocked to ensure that the predictions would not be artificially biased due to the relatively large number of variables (200). The weighting of the PSD was optimized by comparing the maximum average of the captured variance with varying block weights (from 0 to 100% weighing of the PSD block). With all the PLS regressions to the blend properties (*Y*) it was found that using 6 principal components (PCs) captured the maximum variance in the dataset without over-fitting. The model predictions were cross-validated using jack-knifing, whereby each sample was iteratively removed, the model was recalculated with the remaining samples, and the removed sample was projected into the new model. The cross validation term (Q^2 value) was calculated as a measure of the future predictability of the model.

2.3. Prediction of properties of new blends

New blends, which were not part of the training set, can be predicted as long as each of the individual components of the blend have been previously tested. For example, any additional API or excipient can be added to the model without requiring the addition of a blend containing that material. Typically a new material can be fully characterized for all of the required properties using <50 g of material.

Predictions were performed on blends containing common pharmaceutical excipients for immediate-release formulations and two Pfizer proprietary APIs (API-A and API-B). A third Pfizer API (API-C) was used to illustrate how the WSPLS can be used to guide formulation and process development with minimal experimentation. The blends were composed of 2:1 wt.% ratios of microcrystalline cellulose (Avicel PH102, FMC Biopolymer, Mechanicsburgh, PA) to lactose monohydrate (Fast Flow 316, Foremost Farms, Baraboo, WI) for API-A and API-C formulations or dicalcium phosphate (A-Tab, Innophos, Cranbury, NJ), for the API-B formulation. All the above formulation also contained 3 wt.% sodium starch glycolate (Explotab, JRS Pharma, Patterson, NJ) and 1 wt.% magnesium stearate (vegetable source, Mallinckrodt, Hazelwood, MO). These new formulations were not included in the model training set used to construct the model. The predicted properties

of these blends, including compact strength, powder flow, and particle size distribution, were then compared to the actual measured values to determine model usefulness.

2.4. Software

All of the data manipulation and PCA and PLS models were performed in MATLAB® (The Mathworks, inc., Natick, MA) using custom in-house code.

3. Results and discussion

3.1. The use of multivariate modeling approaches

The simplest method for predicting blend properties from individual components is to employ ideal-mixing rules to estimate each individual property of the resulting blend. However, this method has some key limitations. First, ideal-mixing fails to capture any non-linear mixing effects (i.e. when the resultant property is not a simple weighted average of the properties of the components) on the properties of the blend. While this may be applicable to some properties such as particle size distribution, it is unlikely to work for others. For example, it has been shown that the bonding and tableting indices of binary mixtures exhibited non-linear relationships when brittle and plastic components were blended (Wurster et al., 1999; Majuru and Wurster, 1997). Often pharmaceutical blends are complex mixtures that contain 5 or more individual components. In these cases it is even more unlikely that these ideal-mixing predictions will apply. Furthermore, univariate analysis fails to capture the correlations and interactions between the materials properties. From a formulation design perspective this can become problematic when optimizing for a specific blend property may result in creating deficiencies in another.

Therefore a new, multivariate, "weighted-scores PLS" approach was developed which allows for simultaneous prediction of several properties of a new blend. This was accomplished by first transforming the data in X using PCA. This transformation was performed for two purposes: (1) to capture the complex relationship (correlation structure) between properties in the X space and (2) to allow for all materials (individual components and blends) to be positioned in a multivariate "design-space". In this case, when a new material is tested it can be readily compared to previous experience. This will allow the formulator to identify if a new material can be modeled based on past data as well as determine if it needs to be considered "special-case" for formulation development. These special-case materials, which fall outside of previous experience, may require specialized excipients or processing procedures to produce robust dosage forms. Following the component PCAs, the scores were then weighted by R, prior to a PLS regression, against the actual blend data.

Using this WSPLS approach the captured variance (R^2) for the training dataset was found to be high (>81%). This represents a significant improvement over the ideal-mixing method where the total R^2 value was found to be 62%. The WSPLS model was also cross-validated by iteratively removing each blend from the dataset, remodeling, and comparing the predictions to the measured values. The resulting Q^2 value was found to be very high (~80%). This indicates that the predictability of this model for new blends is quite robust. The R^2 per variable for several select powder and mechanical properties (Fig. 2) and the quantile particle size distribution (Fig. 3) were compared with that from ideal-mixing. This comparison shows the clear advantage of this multivariate approach over ideal-mixing rules. Of note, properties which are especially important from a solid-dosage form manufacturability standpoint, such as the pressure needed to achieve an acceptable compact (compression stress), compact strength (tensile strength and BFI), and powder flow function coefficient (FFC), are all particularly well understood (R^2 of 72–87%).

As an example, predicted vs. observed values for two material properties, compact tensile strength and powder FFC, are shown in Fig. 4. Excellent agreement can be seen between the model prediction and the measured properties. Some observations show the predicted values for blends where no actual measured values exist for that particular blend. In this multivariate model missing data was estimated through the known property correlations as captured in the PLS model loadings. This ability to handle missing data illustrates another important advantage of using a multivariate modeling approach vs. applying univariate mixing rules. This flexibility may be especially useful in that blend predictions are not limited if available API quantities are not sufficient to allow for complete testing. Taken together, the superior modeling performance and flexibility afforded by using WSPLS indicates this technique may have improved utility for pharmaceutical formulation development.

3.2. Prediction performance verification

Two example blend predictions were performed on formulations not contained in the original training dataset. The predictions for both formulations, the first containing 30 wt.% API-A (Table 2), and the second containing 23 wt.% API-B (Table 3), were compared to experimentally derived values. For each blend the predicted and measured blend properties are shown along with the experimental error (standard deviation or replicate compacts) and prediction error (model residuals). Furthermore, a rating category system was applied to each property. The categories include "attribute", "margina", "deficient", and "severely deficient". The ratings categories were developed for each property to give an indication of processing performance for solid-dosage form development. The particular ratings cutoffs were set based on published results and experience with proprietary compounds and formulations. Materials with similar ratings are considered to be similar in a real-world manufacturing environment. As seen in both examples, the actual and predicted values for the material properties show good agreement. The ratings categories of the predicted and measured values differed in only two cases. The brittle-fracture index is classified as an attribute for the API-A blend while the actual test exhibits marginal performance. In this case, the API performance falls very close to the ratings category boundary. The delineation falls within the combined prediction and experimental error. In these cases it may be best to identify this material as a "borderline attribute/marginal" case. For the API-B blend the viscoelasticity is classified to be marginal while the actual test shows the performance to be deficient. In this case, the difference is likely due to the relatively low predictive power for this particular property using the model (R^2 < 50%). This lack of predictability can be taken into account by using the R^2 per variable to guide prediction confidence. Each property can be weighted by its R^2 value in order to guide formulator's decisions. Overall, agreement with the actual blend performance is sufficient to make development decisions without further experimentation. The WSPLS model illustrates how an *in-silico* approach can be a useful tool for understanding and predicting real-life behavior of pharmaceutical blends.

3.3. Example application to pharmaceutical formulation development

The blend prediction model described above was applied to help determine initial blend formulations for a high-dose requirement API. Often these high-dose formulations require a balance to be struck between API loading (wt.%) and dosage-form size. On

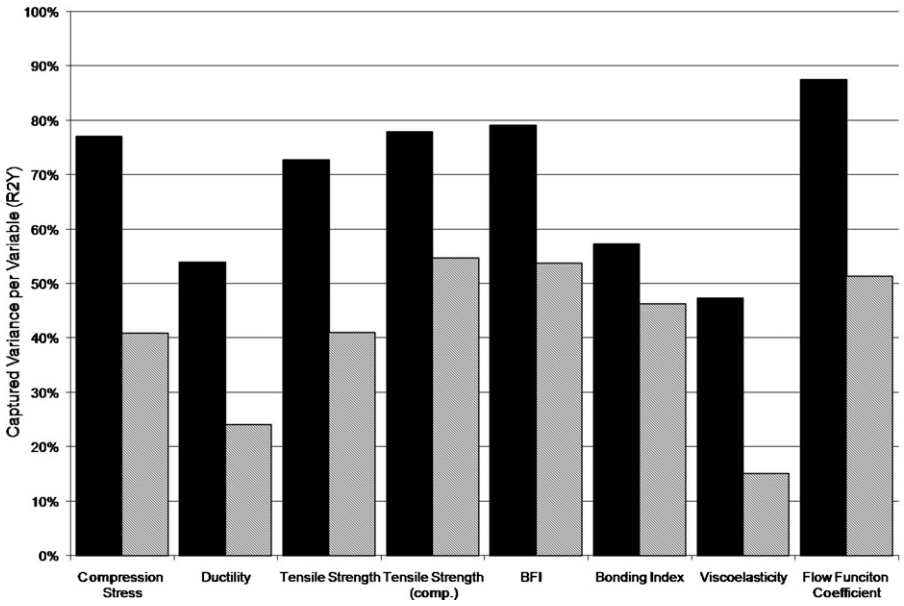

Fig. 2. Captured variance (R^2Y) per variable for selected blend properties using the WSPLS model (black bars) compared to simple volume weighting using ideal-mixing rules (shaded bars).

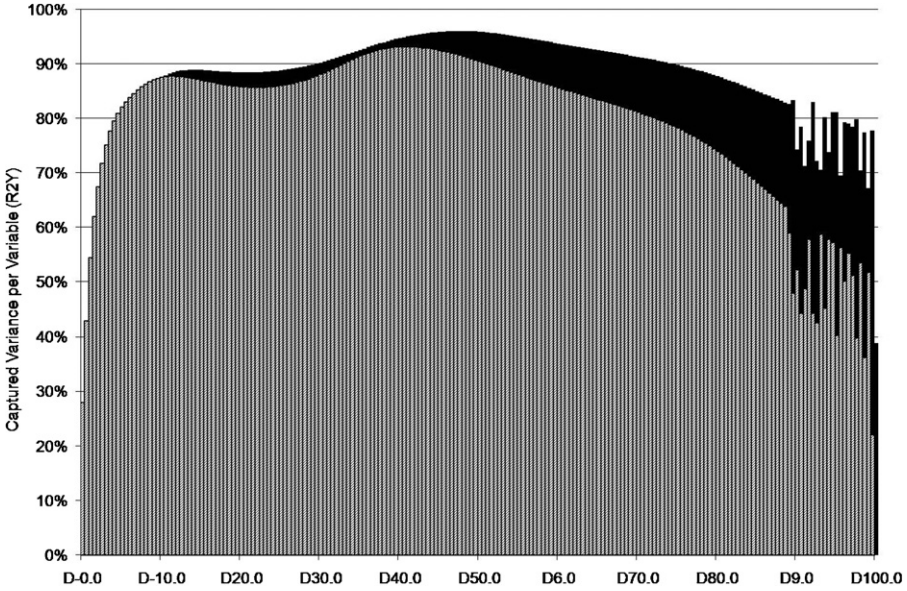

Fig. 3. Captured variance (R^2Y) per variable for quintile particle size distribution using the WSPLS model (black bars) compared to simple volume weighting using ideal-mixing rules (shaded bars).

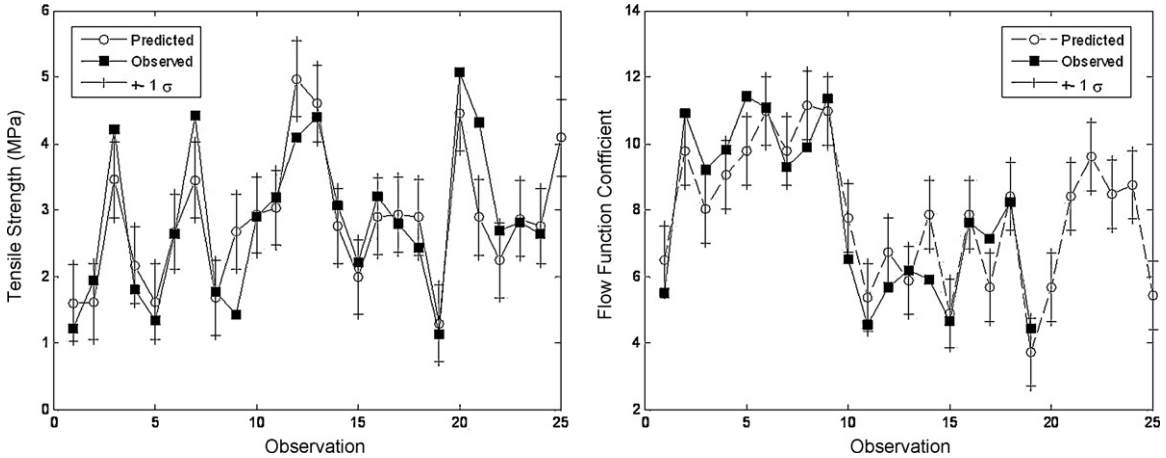

Fig. 4. Observed vs. predicted values for compact tensile strength (left) and powder flow function coefficient (right) for the 25 tested blends in WSPLS model.

Table 2
Actual and predicted properties for an immediate release formulation containing 30% API-A.

Property (unit)	Experimentally determined		WSPLS predicted	
	Value (Std dev)	Rating category	Value (95% CI)	Rating category
Compression stress (MPa)	90.8 (2.6)	Attribute	87.6 (12.5)	Attribute
Ductility (MPa)	447.2 (25.6)	Deficiency	590.3 (164.3)	Deficient
Tensile strength (MPa)	3.5 (7.0E-3)	Attribute	2.9 (0.5)	Attribute
Compromised tensile strength (MPa)	2.6 (4.9E-02)	Attribute	2.5 (0.5)	Attribute
Brittle fracture index (none)	1.6E-01	Marginal	7.9E-02 (3.2E-02)	Attribute
Bonding index (none)	7.9E-03	Marginal	4.7E-02 (1.8E-03)	Marginal
Viscoelasticity (none)	8.9	Attribute	8.7 (1.9)	Attribute
Flow function coefficient (none)	5.2	Marginal	5.3 (1.1)	Marginal

Table 3
Actual and predicted properties for an immediate release formulation containing 23% API-B.

Property (unit)	Experimentally determined		WSPLS predicted	
	Value (Std dev)	Rating category	Value (95% CI)	Rating category
Compression stress (MPa)	98.5 (1.8)	Attribute	95.6 (12.3)	Attribute
Ductility (MPa)	386.0 (40.8)	Marginal	342.1 (190.6)	Marginal
Tensile strength (MPa)	2.9 (5.4E-02)	Attribute	2.8 (0.4)	Attribute
Compromised tensile strength (MPa)	2.8 (4.4E-08)	Attribute	2.7 (0.4)	Attribute
Brittle fracture index (none)	2.8E-02	Attribute	1.7E-02 (2.7E-02)	Attribute
Bonding index (none)	7.6E-03	Marginal	4.6E-03 (1.1E-03)	Marginal
Viscoelasticity (none)	22.9	Deficient	13.7 (1.8)	Marginal
Flow function coefficient (none)	6.4	Attribute	6.9 (1.8)	Attribute

one hand, at high API loadings the blend properties may become deficient due to the undesirable characteristics of the API being prominent. This may result in poor robustness of the manufacturing process and unacceptable attributes of the drug product and intermediates such as increased potency variability, powder hang-ups, or softer tablets. However, decreasing the API loading is not always possible due to limits on tablet/capsule sizes. Equally, oversized oral dosage forms are not desirable as they can become difficult to swallow and may affect patient compliance, especially in particular populations such as young children and elderly individuals. Achieving the correct balance of these concerns may require producing and testing various formulations with a range of API loadings to achieve both acceptable blend properties and dosage form sizes. At the same time, especially in early drug-product development, it is important to reduce API use by minimizing experimentation. In these situations the use of *in-silico* predictions to facilitate rapid formulation design may be particularly useful.

In this example case, the WSPLS approach described above was applied to a high-dose (600 mg) direct compression (DC) tablet formulation. The goal was to screen potential formulations *in-silico* to help guide drug-product development without additional experimentation and material usage. The API was first tested using <50 g of material to determine the key compact, flow, and particle size material properties. Complete mechanical property characterization was not possible due to limited material availability. However, it was determined that the primary concern for this particular API was the deficient flow characteristics (FFC, Table 4). This property is considered particularly important for a formulation such as this since the DC process relies heavily on acceptable blend flow to achieve proper tablet weight control and content uniformity. Therefore, it was critical to gauge the upper limit of achievable API loading that would still result in acceptable blend FFC. This maximum loading would strike the best possible balance between manufacturability and tablet size.

Table 4
API mechanical and flow properties for API-C.

Property	API properties
Compression stress (MPa)	72.86 (Attribute)
Ductility (MPa)	NA
Tensile strength (MPa)	1.08 (Marginal)
Compromised tensile strength (MPa)	0.62 (Marginal)
Brittle fracture index (none)	3.70E-01 (Deficient)
Bonding index (none)	NA
Viscoelasticity (none)	NA
Flow function coefficient (none)	4.20 (Deficient)

The predicted mechanical properties of various formulations of API-C are shown in Table 5. As expected, the properties generally trend towards those of the pure API with increased loading in the formulation. It is important to note, that while the prediction errors seem significant compared to the measurement (generally 15–30%), the values are small relative to the differences in rating categories (as seen in Tables 2 and 3). Therefore, this model is a useful tool to guide formulation decision which produce desirable blend properties. As expected, the predicted FFC values were found to be less desirable with increasing API loading as shown in Fig. 5. At API loadings of >50% the predicted blend flow characteristics were no longer considered acceptable using the standard rating system (Mullarney and Leyva, 2009). Under these deficient flow conditions manufacturability of the tablet may suffer due to material hang-ups

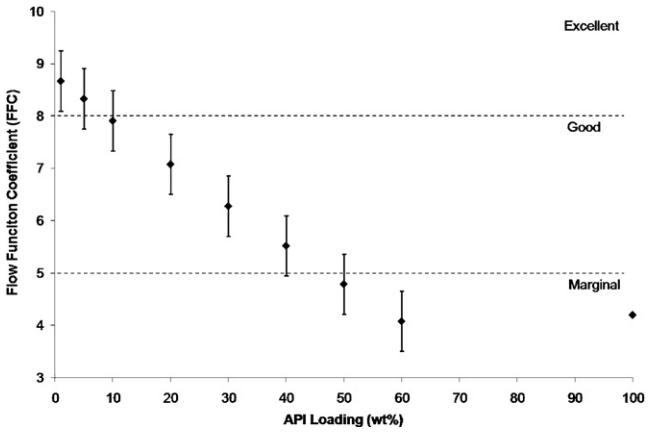

Fig. 5. Predicted FFC values for formulations with API loadings from 1 to 60 wt.% API-C along with the actual measured value of the pure API (shown at 100 wt.%) and ratings categories. Error bars represent ±1 standard deviation.

Table 5
Comparison of predicted properties at various API loadings of API-C.

Property	API-C loading (wt.%.)								Prediction error	
	1%	5%	10%	20%	30%	40%	50%	60%		
Compression stress (MPa)	89.3	87.8	85.8	82.0	78.4	75.0	71.7	68.5	±	12.46
Ductility (MPa)	646.7	636.2	622.6	596.1	570.8	546.5	523.1	500.6	±	164.28
Tensile strength (MPa)	2.9	2.9	2.8	2.7	2.6	2.5	2.4	2.3	±	0.54
Compromised tensile strength (MPa)	2.65	2.59	2.51	2.36	2.21	2.07	1.94	1.81	±	0.48
Brittle fracture index (none)	6.47E-02	6.77E-02	7.19E-02	8.01E-02	8.79E-02	9.54E-02	1.03E-01	1.10E-01	±	3.20E-02
Bonding index (none)	4.48E-03	4.45E-03	4.38E-03	4.24E-03	4.12E-03	3.99E-03	3.87E-03	3.76E-03	±	1.83E-03
Viscoelasticity (none)	8.49	8.56	8.64	8.78	8.93	9.06	9.19	9.32	±	1.92

during processing such as during hopper discharge. Furthermore, the poor flow of the DC blend into the tablet press die may result in inconsistent fill weights and produce tablets with high weight variability. Based on this analysis, a 40% API loading formulation was chosen based on the predicted FFC with consideration for the model prediction error. At this loading drug product development can proceed with good flow during manufacturing, while minimizing the required tablet size. This 40% loading translates to a minimum tablet weight of 1500 mg for this high dose DC formulation. Importantly the *in-silico* model yielded this conclusion rapidly without any additional experimentation or API usage. This example illustrates the utility of how model-based formulation design can be used to focus development efforts and result in minimized experimentation time and related resources.

4. Conclusions

A novel method for *in-silico* prediction of the performance of pharmaceutical powder blends has been developed. A two-step, multivariate modeling approach was created using historical physical data of APIs, excipients, and multi-component blends. The physical properties for each individual component were first transformed using a PCA technique to place them in a multivariate design space and capture property correlations. The scores from these PCA models were then weighted by the blending ratios prior to PLS regression versus actual measured blend properties. This method produced a complete prediction of all the material properties simultaneously which was shown to be superior to the prediction performance observed when applying linear ideal-mixing. An example of the application of this model was shown for a high-dose formulation where an optimal API loading was determined which best balanced the need for satisfactory blend performance along with acceptable tablet size. The ability to perform these formulation design analyses without the need for experimentation can be especially useful when API resources are limited (i.e. during early drug-product development). This approach represents a potential application of QbD principles to the selection of incoming materials and formulation through *in-silico* modeling.

In the future, additional API, excipient, and blend data will be added to the model as it is created. This will serve to further broaden the range of experience captured in the model while reducing prediction error. Additionally, an optimizer function will be added to this model whereby excipient types and concentrations will be chosen based on target blend properties for a given API. Since different dosage forms may have different product requirements (e.g. capsule vs. tablet, immediate vs. controlled release) the optimizer will allow for the target blend characteristics to be tailored to the specific goals of each project. Constraints can be added to ensure that materials that impart required functionality to the dosage form will be included. For example, the requirement that disintegrant and lubricant are used in the formulation will be built into the constraints so that dissolution and anti-sticking performance will not be compromised. Additional constraints can be added to choose preferred materials if the excipients chosen are not acceptable from a supply standpoint. With this system in place the development of new, more ideal, formulations can be achieved with little to no experimentation and less than 50 g of API.

Acknowledgments

The authors gratefully acknowledge Dr. Bruno Hancock, Dr. Mary am Ende, and Matthew Mullarney for their useful scientific discussion regarding this work.

References

Bernardo, C.A., Cunha, A.M., Oliveira, M.J., 1996. The recycling of thermoplastics: prediction of the properties of virgin and reprocessed polyolefins. Polym. Eng. Sci. 36, 511–519.
Cao, X., Hancock, B.C., Leyva, N., Becker, J., Yu, W., Masterson, V.M., 2009. Estimating the refractive index of pharmaceutical solids using predictive methods. Int. J. Pharm. 368, 16–23.
Cao, X., Leyva, N., Anderson, S.R., Hancock, B.C., 2008. Use of prediction methods to estimate true density of active pharmaceutical ingredients. Int. J. Pharm. 355, 231–237.
Food and Drug Adminstration, 2000. ICH Q6A: test procedures and acceptance criteria for new drug substances and new drug products: chemical substances. Fed. Regist. 65, 83041–83063.
Food and Drug Adminstration, 2009. ICH Q8(R2) Pharmaceutical Development. Fed. Regist. 71.
Gendrin, C., Roggo, Y., Collet, C., 2008. Pharmaceutical applications of vibrational chemical imaging and chemometrics: a review. J. Pharm. Biomed. Anal. 48, 533–553.
Hamad, M.L., Bowman, K., Smith, N., Sheng, X., Morris, K.R., 2010. Multi-scale pharmaceutical process understanding: from particle to powder to dosage form. Chem. Eng. Sci. 65, 5625–5638.
Hancock, B.C., Carlson, G.T., Ladipo, D.D., Langdon, B.A., Mullarney, M.P., 2002. Comparison of the mechanical properties of the crystalline and amorphous forms of a drug substance. Int. J. Pharm. 241, 73–85.
Hiestand, E., 1997. Mechanical properties of compacts and particles that control tableting success. J. Pharm. Sci. 86, 985–990.
Hiestand, E., Smith, D.P., 1984. Indices of tabletting performance. Powder Technol. 38, 145–159.
Hiestand, E.N., 1995. Rationale for and the measurement of tableting indicies. In: Alderborn, G., Nystrom, C. (Eds.), Pharmaceutical Powder Compaction Technology. Marcel Dekker, Inc, New York.
Hiestand, E.N., 2003. Mechanics and Physical Principles for Powders and Compacts, 2nd ed. SSCI Inc, West Lafayette, USA.
Kachrimanis, K., Karamyan, V., Malamataris, S., 2003. Artificial neural networks (ANNs) and modeling of powder flow. Int. J. Pharm. 250, 13–23.
Kolarik, J., Fambri, L., Pegoretti, A., Penati, A., 2000. Ternary polymer blends: prediction of mechanical properties for various phase structures. Polym. Adv. Technol. 11, 75–81.
Kumar, V., 2003. Computational materials science: the emergence of predictive capabilties of material behaviour. Sadhana 28, 815–831.
MacGregor, J.F., Yu, H., Garcia-Munoz, S., Flores-Cerrillo, J., 2005. Data-based latent variable methods for process analysis monitoring and control. Comput. Chem. Eng. 29, 1217–1223.
Majuru, S., Wurster, D.E., 1997. The effect of compositions on the tableting indices of binary powder mixtures. Pharm. Dev. Technol. 2, 313–321.
Mullarney, M.P., Hancock, B.C., Carlson, G.T., Ladipo, D.D., Langdon, B.A., 2003. The powder flow and compact mechanical properties of sucrose and three high-intensity sweeteners used in chewable tablets. Int. J. Pharm. 257, 227–236.
Mullarney, M.P., Leyva, N., 2009. Modeling pharmaceutical powder-flow performance using particle-size distribution data. Pharm. Technol. 33, 126–134.
Sandler, N., Wilson, D., 2010. Prediction of granule packing and flow behavior based on particle size and shape analysis. J. Pharm. Sci. 99, 958–968.
Schultz, D., 1996. Flowability and time consolidation measurements using a ring shear tester. Powder Handl. Process. 8, 221–226.

Shao, Q., Rowe, R.C., York, P., 2007. Comparison of neurofuzzy logic and decision trees in discovering knowledge from experimental data of an immediate release tablet formulation. Eur. J. Pharm. Sci. 31, 129–136.
Smith, J.M., Van Ness, H.C., Abbott, M.M., 2005. Introduction to Chemical Engineering Thermodynamics, 7th ed. McGraw Hill.
Soh, J.L.P., Wang, F., Boersen, N., Pinal, R., Peck, G.E., Carvajal, M.T., Cheney, J., Valthorsson, H., Pazdan, J., 2008. Utility of multivariate analysis in modeling the effects of raw material properties and operating parameters on granule and ribbon properties prepared in roller compaction. Drug Dev. Ind. Pharm. 34, 1022–1035.
Takayama, K., Fujikawa, M., Obata, Y., Morishita, M., 2003. Neural network based optimization of drug formulations. Adv. Drug Deliv. Rev. 55, 1217–1231.
Workman, J., Koch, M., Lavine, B., Chrisman, R., 2009. Process analytical chemistry. Anal. Chem. 81, 4623–4643.
Wu, C.Y., Best, S.M., Bentham, A.C., Hancock, B.C., Bonfield, W., 2005. A simple predictive model for the tensile strength of binary tablets. Eur. J. Pharm. Sci. 25, 331–336.
Wu, C.-Y., Best, S.M., Bentham, A.C., Hancock, B.C., Bonfield, W., 2010. Predicting the tensile strength of compacted multi-component mixtures of pharmaceutical powders. Pharm. Res. 23, 1898–1905.
Wurster, D.E., Shingai, M., Euichaul, O., 1999. Prediction of Hiestand bonding indices of binary powder mixtures from single-component bonding indices. Pharm. Dev. Technol. 4, 65–70.