Contents lists available at ScienceDirect

Chemical Engineering Research and Design

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

In-silico product formulation design through latent variable model inversion

Emanuele Tomba[a], *Massimiliano Barolo*[a], *Salvador García-Muñoz*[b,*]

[a] *CAPE-Lab – Computer-Aided Process Engineering Laboratory, Department of Industrial Engineering, University of Padova, via Marzolo 9, 35131 Padova PD, Italy*

[b] *Pfizer Worldwide Research & Development, 445 Eastern Point Road, Groton, CT 06340, USA*

A B S T R A C T

Two of the first important decisions to take in the development of a solid oral drug product are the selection of excipients that are to be mixed with the active pharmaceutical ingredient (API) in a commercial formulation, and the manufacturing route. This work proposes to use a latent variable model methodology presented in a previous work (Polizzi and García-Muñoz, Int. J. Pharm., 2011, 418, 235–242) to enable the *in-silico* design of new product formulations. A constrained optimization framework is used to invert the underlying model in order to select the best excipients and concentrations for a given API to ensure the achievement of a pharmaceutical blend with a desired profile of particle, powder and compact mechanical properties. The approach is verified by designing a new pharmaceutical formulation for direct compression, using an API that was previously formulated via a wet granulated process. The experimental results confirm the effectiveness of the method. The proposed methodology can act as an important tool to guide and accelerate the decision making process in pharmaceutical product development, while minimizing the required experimentation as well as the raw materials consumption. The approach can be extended to consider other constraints (or targets) such as stability, as long as there is a mathematical way to relate the targets (e.g., degradation extent) to the incoming formulation.

Keywords: Product design; Quality-by-design; Latent variable methods; Partial least squares; Model inversion

1. Introduction

The introduction of the Quality by Design (QbD) paradigm (ICH, 2009) aimed at changing the way in which the pharmaceutical industry managed product and process development activities. The fundamental principle of QbD is that the product quality should be *designed* into the process through an upstream approach, rather than be based on downstream troubleshooting. To apply the QbD paradigms in product development, it is therefore required to know and understand which are the driving forces acting upon the complex network of interactions between incoming and intermediate materials, processes and product quality. A correct understanding of the relationships between the variables acting on a system could be that of being able to describe them mathematically, in order to predict the behavior of the output variables upon changes on the inputs and take the appropriate control actions, if needed. This can be done if a model relating the raw material properties, the process parameters and the final product quality is available. The application of the QbD principles to pharmaceutical development can therefore be understood as the use of modeling techniques to support and accelerate product development and integrated product and process design activities (García-Muñoz and Oksanen, 2010).

In the development of a drug product, the first important decision to take is the choice of the formulation, namely the selection of the appropriate excipients that are to be mixed with an active pharmaceutical ingredient (API) into the final drug product. This choice is driven by constraints mainly related to the safety, the efficacy but also the processability of the drug product (Hamad et al., 2010).

* Corresponding author. Tel.: +1 860 7150578; fax: +1 860 6866380.
E-mail addresses: Salvador.Garcia-Munoz@pfizer.com, salvadorgarciamunoz@gmail.com (S. García-Muñoz).
Received 25 March 2013; Received in revised form 29 July 2013; Accepted 23 August 2013
0263-8762/$ – see front matter
http://dx.doi.org/10.1016/j.cherd.2013.08.027

In general, if a model describing the mixture properties from the properties of the excipients and the APIs involved in the formulation were available, it could be used to aid the selection of the best materials that ensure to obtain a mixture of desired properties. To this end, the model itself can act as a constraint to an optimization problem aiming at maximizing/minimizing an objective function, which may represent a product performance index. The optimization will then give as outputs the estimates of the best raw material types and amounts to be used in order to achieve the desired mixture (or blend) properties (Smith and Ierapepritou, 2010).

The use of deterministic models to describe these relationships would always be desirable, as deterministic models give a transparent representation of the physical phenomena acting on the system, explaining them from first principles. In the field of mixture modeling, some examples have been reported for the semiconductor industry (Kumar, 2003) and for polymeric blend design (Bernardo et al., 1996). The development of such models requires a detailed knowledge and understanding of the interactions between the materials entering the formulation and of their implications for the product properties. This may be burdensome to achieve in pharmaceutical product design, due to the variety of products and raw materials of different physical/chemical characteristics that may enter in a drug product formulation (e.g., APIs, fillers, binders, disintegrants), which can be difficult to manage in a deterministic modeling framework. For these reasons, formulators have often resorted to building models based on data obtained either from targeted experiments or from historical databases.

Mixture design of experiments (DOE) and response surface modeling techniques (Montgomery, 2005) have been among the first systematic multivariate approaches used to assist the design of the formulations of pharmaceutical products (Campisi et al., 1998). Due to the nature of the pharmaceutical formulations, the use of DOE techniques in the early stages of pharmaceutical development often requires performing a large number of experiments, as a large number of candidate materials and permutations have to be considered. To address this issue, multivariate design of experiments has been proposed (Wold et al., 1986). This approach combines DOE techniques with the multivariate analysis of databases of the available raw materials (e.g., through principal component analysis; PCA, Jackson, 1991), to select the most suitable for the experimental design. As an example, Gabrielsson et al. (2003) used a multivariate design to evaluate in a systematic way a large number of candidate excipients to include in a formulation, based on their similarity assessed through a PCA of their characterization data. Latent variable regression models (LVRMs) based on projection to latent structures (PLS, Höskuldsson, 1988) were then built on the data obtained from the experiments and used to test new formulations in order to obtain a product of desired disintegration time and crushing strength (Gabrielsson et al., 2004).

Indeed, one of the important characteristics a data-based model should have to support product formulation design is the ability in accurately predicting the mixture properties based on the raw materials data. Many efforts have been produced by researchers in finding methods able to give models with good prediction performances, often regardless of the model structure. For example, several contributions on the use of black box models to support the design of pharmaceutical formulations have been proposed (Rowe and Roberts, 1998). Learning techniques like neural networks (Agatonovic-Kustrin and Beresford, 2000; Takayama et al., 2003; Sun et al., 2003), neuro-fuzzy logic (Abraham et al., 2007; Landín et al., 2009), genetic programming (Barmpalexis et al., 2011) and expert systems (Shao et al., 2007) have been applied to build models based on historical or DOE formulation data, and then used in an optimization framework to suggest the experiments to be performed for the design of the product formulation. Despite these tools are very attractive and the results are promising, most black box models suffer from the lack of transparency in understanding the mechanisms behind predictions. Model design for black box models is essentially based on the accuracy of predictions rather than on the optimization of the model complexity. As a consequence, black box models parameters are often difficult to interpret and may not provide a scientific understanding of the system being modeled. This conflicts with the QbD requirements, for which the ability to predict has to reflect a high degree of process understanding (FDA, 2004).

Conversely, multivariate statistical methods like LVRMs combine the accuracy in prediction to a scientifically sound understanding and interpretation of both model building procedure and model parameters. Several studies have recently been published which demonstrate this (Soh et al., 2008; Sandler and Wilson, 2010). In most of the contributions that used LVRMs to model mixture data, the model is however limited to describe the relationships between a matrix of regressors, including only the fractions of the raw materials in the tested formulations, and a response matrix, including the properties of the mixture. This type of analysis does not properly include in the model the physical/chemical properties of all the raw materials that could possibly be included in the final formulation. Muteki and MacGregor (2007) addressed this issue by proposing an LV method called L-shaped PLS (LPLS), which includes in a unique LVRM framework the database of the physical/chemical characterization of all the available raw materials, the database on the fractions of each raw material in the historical formulations, and the properties of the obtained mixture. This modeling approach was then combined with LVM inversion techniques to successfully support the development of a new blend of rubbers (Muteki et al., 2006).

Recently, García-Muñoz and Polizzi (2012) recognized that, despite its effectiveness, the approach proposed by Muteki et al. (2006) did not consider those situations in which the desired product is obtained by mixing materials of different nature or which underwent different characterization procedures, as typical for example of APIs and excipients in pharmaceutical formulations. To address this issue, Garcia-Muñoz and Polizzi proposed the weighted-scores PLS method (WSPLS; García-Muñoz and Polizzi, 2012). This method was applied to predict the particle, powder and compact mechanical properties of pharmaceutical blends, starting from the raw materials properties and amounts, without resorting to extensive experimentation (Polizzi and García-Muñoz, 2011).

Following Polizzi and García-Muñoz (2011), in this paper we use their LVRM strategy to perform *in-silico* a new drug product formulation design. The strategy is based on an optimization framework for LVRM inversion (Tomba et al., 2012) which, starting from a given API, returns as output the set of excipient type and amounts most suitable to reach a blend of desired powder, flow and mechanical properties. Since the formulation design problem may have different objectives and constraints (e.g., the maximization of the API dosage, the minimization of the tablet weight, the choice of excipients of a given family), the strategy is developed in such a way as to allow the user to specify several different types of constraints,