Fast-track to A Solid Dispersion Formulation Using Multi-way Analysis of Complex Interactions

JIAN X. WU,[1] FRANS VAN DEN BERG,[2] SØREN V. SØGAARD,[1] JUKKA RANTANEN[1]

[1]Department of Pharmacy, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen 2100, Denmark

[2]Quality and Technology, Department of Food Science, Faculty of Science, University of Copenhagen, Frederiksberg 1958, Denmark

Received 6 September 2012; revised 4 November 2012; accepted 14 November 2012

Published online 4 December 2012 in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/jps.23409

ABSTRACT: Several factors with complex interactions influence the physical stability of solid dispersions, thus highlighting the need for efficient experimental design together with robust and simple multivariate model. Design of Experiments together with ANalysis Of VAriance (ANOVA) model is one of the central tools when establishing a design space according to the Quality by Design (QbD) approach. However, higher order interaction terms are often significant in these ANOVA models, making the final model difficult to interpret and understand. As this is ordinarily the purpose of applying ANOVA, it poses an obvious problem. In the current study, the GEneralized Multiplicative ANOVA (GEMANOVA) model is proposed as an alternative for the ANOVA model. Two complex multivariate data sets obtained by monitoring the physical stability of a solid dispersion with image analysis and X-ray powder diffraction (XRPD) as responses were subjected to GEMANOVA analysis. The results showed that the obtained GEMANOVA model was easier to interpret and understand than the additive ANOVA model. Furthermore, the GEMANOVA model has additional advantages such as the possibility of readily including multivariate responses (e.g., an entire spectral data set), model uniqueness, and curve resolution abilities. © 2012 Wiley Periodicals, Inc. and the American Pharmacists Association J Pharm Sci 102:904–914, 2013

Keywords: GEMANOVA; PARAFAC; ANOVA; factorial design; design space; Quality by Design (QbD); solid dispersion; X-ray diffractometry; image analysis; multivariate analysis

INTRODUCTION

Enhancing solubility and release rate of poorly water-soluble drugs is an important formulation strategy for the pharmaceutical industry. Many formulation approaches have been suggested for increasing the bioavailability of the drug, among these, solid dispersion formulation is one of the promising methods.[1] Preventing the amorphous dispersed drug in polymer matrix from recrystallization is one of the key challenges related to solid dispersion formulations. During the solid dispersion formulation development phase, many factors originating from both formulation as well as processing can have decisive influence of the physical stability of the drug,[2,3] and the importance of evaluating the combined influence of these factors has been highlighted.[4,5] From a formulation perspective, numerous studies have investigated and found factors such as polymer-to-drug ratio, polymer type, and its molecular weight to have decisive influence for the drug physical stability.[6–8] When solid dispersion is prepared using the solvent evaporation method, several studies have highlighted the influence of various process parameters including solvent evaporation rate on solid dispersion physical stability.[2,4,9] With the many factors influencing on the solid dispersion physical stability, an emerging recognized challenge is related to the systematic Design of Experiments (DoE) planning, and minimizing the complexity of the following multivariate modeling step, so that understanding of the underlying phenomena and latent factors influencing on the drug's physical stability can be maximized.

The Quality by Design (QbD) approach is becoming an increasingly important element in drug development.[10–12] Establishing a design space by describing relationships between input variables that

Correspondence to: Jukka Rantanen (Telephone: +45-35-30-65-85; Fax: +45-35-30-60-30; E-mail: jukka.rantanen@sund.ku.dk)

Journal of Pharmaceutical Sciences, Vol. 102, 904–914 (2013)

have been identified in a risk assessment phase and their effect on the Critical Quality Attributes (CQAs) is one of the key elements in identifying a proper control strategy according to the QbD approach.[12,13] In the International Conference on Harmonisation (ICH) Q8 guideline, the design space is defined as: "*The multidimensional combination and interaction of input variables (e.g., material attributes) and process parameters that have been demonstrated to provide assurance of quality*."[14] It is well known that one of the central tools in establishing a robust design space is a proper DoE plan,[5,13,15] and when it comes to establishing a model based on the responses generated from the DoE, it is a commonly applied strategy to use Multiple Linear Regression (MLR) fitting of the DoE data set, followed by ANalysis Of VAriance (ANOVA) on the fitted coefficients. As an example, suppose a 2^3 full factorial design was generated consisting of three factors a, b, and c. In the classical ANOVA modeling, the response y is modeled using an additive polynomial model as in Eq. 1:

$$y = \alpha_1 a + \alpha_2 b + \alpha_3 c + \alpha_4 ab + \alpha_5 ac + \alpha_6 bc + \alpha_7 abc + e \quad (1)$$

where $\alpha_{1\ldots}\ \alpha_7$ are the modeled coefficients and e is the error term. As demonstrated here, the effect of the main terms on the modeled response y can easily be understood. However, the presence of higher order interaction terms can make the model difficult to interpret for two reasons. First, the higher order interaction terms might be confounded. Second, suppose α_1 is positive and α_7 is negative. In this case, the interpretation of the modeled factor a can be confusing since varying a alone has a positive influence on y, whereas when a is varied together with b and c, the net effect of this part of the polynomial in Eq. 1 might well be negative. Although Hanrahan and Lu[15] advocated the use of the additive ANOVA model to estimate the design space, in practice, as emphasized by Smilde et al.[16] and other studies, the main factors are seldom independent of each other. This leads to the above-described complex interaction terms in the additive ANOVA, making it sometimes difficult to understand and interpret.[16,17] In the ICH Q8 guideline and, for example, in the study by Rathore,[18] it is emphasized that merely relating the input variables mathematically to CQAs when establishing a design space is not sufficient, since an essential part is to thoroughly understand the underlying model.[14,18] Hence, there is a need for developing models that are easy to understand in cases wherein the classical ANOVA leads to too complex models.

When the data for the design space arise from a DoE, it is by nature of a multi-way character, described by e.g. Bro[19] as several sets of variables measured in a crossed fashion, and hence such data are suitable for modeling by multi-way methods.[16] In a previous work by Bro,[19] it was demonstrated that PARAllel FACtor analysis (PARAFAC) is well suited for modeling DoE data from factorial and fractional factorial designs. The basic PARAFAC model structure is presented as an example in Figure 1. The Figure 1 illustrates that the DoE responses generated from a full factorial design can be approximated by multiplication of a set of the loading vectors. For example, suppose the same 2^3 full factorial design described above consisting of three factors a, b, and c, each varied at two levels $i = 1,2$, $j = 1,2$ and $k = 1,2$. Furthermore, suppose that the generated PARAFAC model consists of two components. A response element $x_{i,j,k}$ within the multi-way array $\underline{\mathbf{X}}$ generated using low level of a, b and high level of c (see Figure 1) can then be reproduced from the PARAFAC model according to Eq. 2, which exemplifies the model for one particular element of the multi-way array:

$$\underline{\mathbf{X}}_{(1,1,2)} = \mathbf{a_{1}}_{(1)} \times \mathbf{b_{1}}_{(1)} \times \mathbf{c_{1}}_{(2)} + \mathbf{a_{2}}_{(1)} \times \mathbf{b_{2}}_{(1)} \times \mathbf{c_{2}}_{(2)} + \underline{\mathbf{E}}_{(1,1,2)} \quad (2)$$

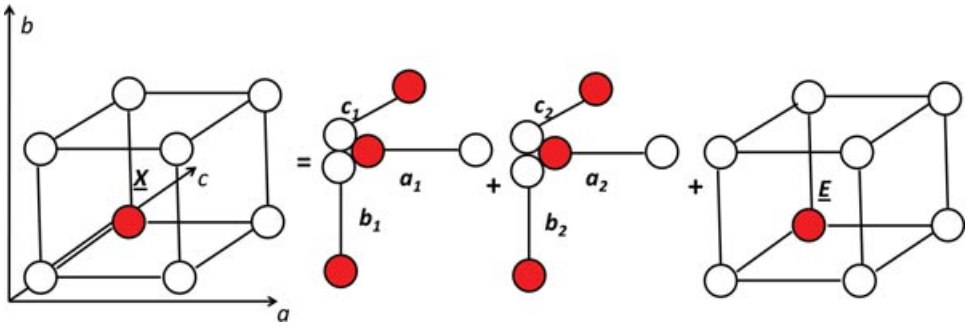

Figure 1. An example of the basic PARAFAC model structure. The responses from the Design of Experiment are arranged in the multi-way array $\underline{\mathbf{X}}$ to the left. In the illustrated example, the response colored in red in $\underline{\mathbf{X}}$ is approximated by multiplication of the elements colored in red belonging to loading vectors $\mathbf{a_1}$, $\mathbf{b_1}$, $\mathbf{c_1}$ then $\mathbf{a_2}$, $\mathbf{b_2}$, and $\mathbf{c}_2$ followed by summation. $\underline{\mathbf{E}}$ is the residual multi-way array.