MARCEL DEKKER, INC. • 270 MADISON AVENUE • NEW YORK, NY 10016

DRUG DEVELOPMENT AND INDUSTRIAL PHARMACY®
Vol. 29, No. 10, pp. 1053–1075, 2003

RESEARCH PAPER

Multivariate Methods in the Development of a New Tablet Formulation

Jon Gabrielsson, M.Sc.,[1] Nils-Olof Lindberg, Ph.D.,[2] Magnus Pålsson, M.Sc.,[2] Fredrik Nicklasson, Ph.D.,[2] Michael Sjöström, Ph.D.,[1] and Torbjörn Lundstedt, Ph.D.[3,*]

[1]Research Group for Chemometrics, Department of Chemistry, Umeå University, Umeå, Sweden
[2]Pharmaceutical R&D, Pfizer Consumer Healthcare, Helsingborg, Sweden
[3]Melacure Therapeutics AB, Uppsala, Sweden

ABSTRACT

The overall objective of this article is to use an efficient approach to find a suitable tablet formulation for direct compression. By using traditional approaches to statistical experimental design in tablet formulation, the number of experiments quickly grows when many descriptive variables or many excipients are included. To facilitate the screening process, a multivariate design, which allows a systematical evaluation of a large number of excipients with a limited number of experiments, was implemented. Formulations with acceptable values for disintegration time and crushing strength were obtained with some of the formulations in the present study. The multivariate experimental design strategy yielded PLS models that will be used to identify a region of interest for the optimization. The strategy is general and can be applied in many different areas of pharmaceutical research and development.

Key Words: Multivariate design; Principal properties; PCA; PLS; Excipient; Tablet formulation.

INTRODUCTION

Formulation of medical products (e.g., tablets) was previously performed mainly on the basis of the experience of the formulator often in combination with the approach of changing one separate factor at a time, but use of statistical experimental design in connection with commercial software has nowadays found widespread use.[1] Except for commercial software for statistical experimental design, formulators in some big pharmaceutical companies also have access to artificial neural networks and expert systems for formulation.

With traditional approaches to statistical experimental design the alternatives are either to use ordinary descriptive variables (e.g., viscosity, particle

*Correspondence: Torbjörn Lundstedt, Ph.D., Department of Organic Pharmaceutical Chemistry, Uppsala University, Box 574, SE-751 23 Uppsala, Sweden; Fax: +46-18-56-72-01; E-mail: torbjorn.lundstedt@melacure.com.

DOI: 10.1081/DDC-120025864
0363-9045 (Print); 1520-5762 (Online)

www.dekker.com

size, etc.) as design factors for each excipient class or to include the excipients in a qualitative factor for each excipient class. The drawback with the former alternative is that to keep the number of experiments on an acceptable level, only a few variables can be used for each excipient class. The problem with the latter alternative is that the number of necessary experiments drastically increases with many levels (i.e., excipients) in the qualitative design factor. In a previous study that recognized the need for a more general strategy in tablet formulation, the implementation of multivariate methods for both screening and optimization proved successful.[2] The study included multivariate characterization of lubricants, binders, and disintegrants. However, the study did not include a filler or an active ingredient.

The screening of the excipients and the resulting models will enable the identification of suitable excipients for the formulation. One of the advantages of this multivariate approach lies in the knowledge and material gathered during the screening phase. For instance, if any of the chosen excipients do not work in the ensuing optimization, there are alternatives with documented qualities to choose from.

To implement the multivariate design, the excipients have to be characterized in multiple variables. The descriptive variables can be literature data or measured values. To avoid problems with missing data, although principal component analysis (PCA) can handle missing data to a certain extent, the excipients in this study were characterized by FT-IR and NIR spectroscopy. The FT-IR and NIR spectra contain information about the chemical composition of the materials.[3,4]

The two spectroscopic methods form the basis for the multivariate characterization, which is an integral part of the multivariate design. To extract the essential information contained in the spectra, PCA is applied. The systematical information contained in the spectra is then extracted in a few principal components. Most of the spectral information is preserved, and the number of descriptive variables is considerably reduced, which is imperative because the aim is to implement an experimental design.

A statistical experimental design is applied to the principal properties to obtain a representative selection of excipients. By combining a multivariate design with a mixture design, understanding of how the amounts of the excipients influence the responses will be gained.

OBJECTIVE

The overall objective of the present article is to use an efficient approach to find a suitable tablet formulation for direct compression. To achieve this objective, a large number of excipients were screened with the aid of the chemometric tools. To facilitate the screening process, a multivariate design, which allows a systematical evaluation of the excipients with a limited number of experiments, was implemented. A multivariate design, a combination of PCA and statistical experimental design, ensures that most of the variation in excipient quality is covered and also significantly decreases the number of necessary experiments.

The results from the multivariate design were evaluated by using partial least squares projections to latent structures (PLS). PLS models, which can describe quantitative relationships between design factors and measured responses, were used for analysis of the results. Models and conclusions from the screening experiments will be used to work out suggestions for further experiments to obtain a formulation with suitable qualities (e.g., a disintegration time of specified duration, acceptable values for crushing strength, and other tablet properties).

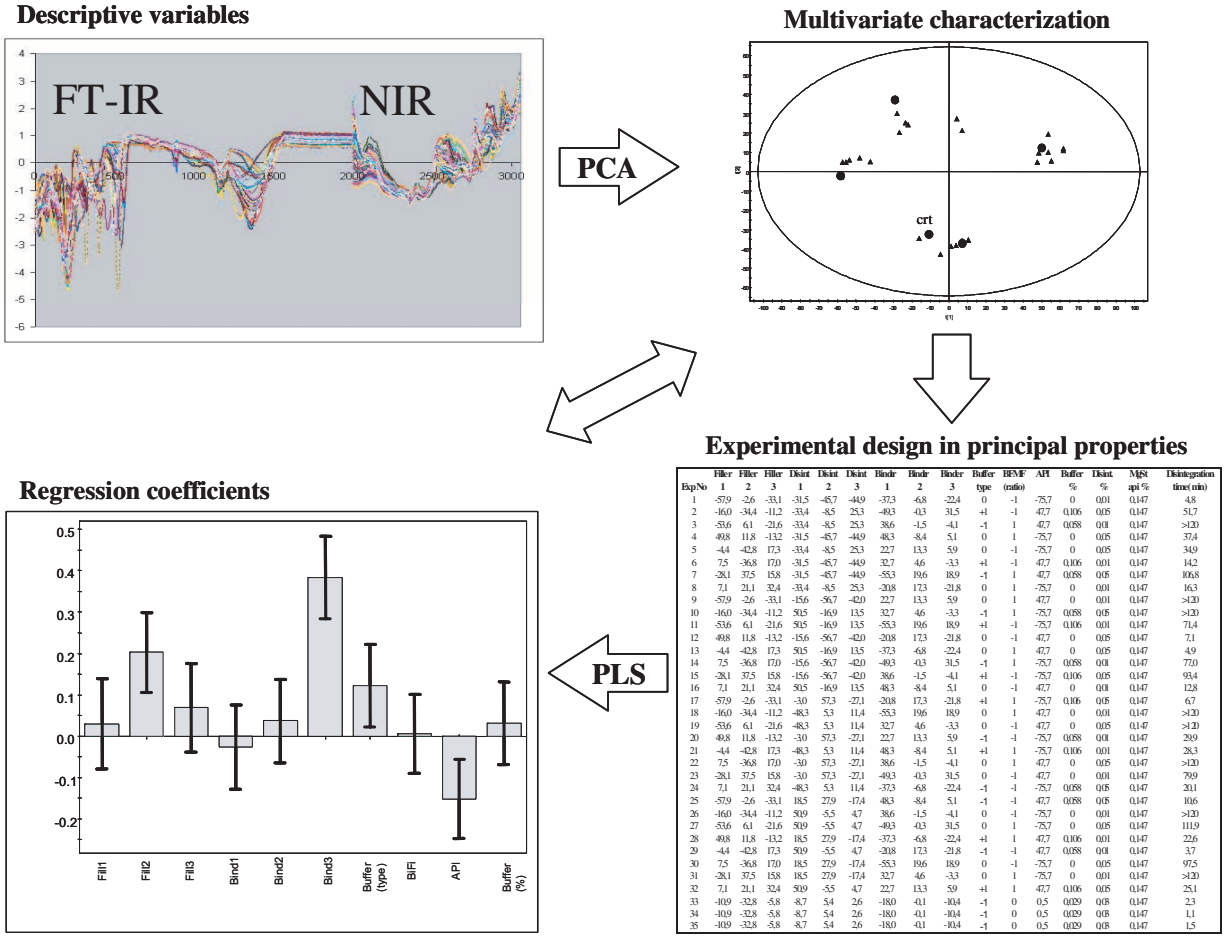
METHODS

A multivariate strategy has been applied in the objective to find a suitable formulation for a product. This strategy typically consists of three steps[2]:

(1) Multivariate characterization.
(2) Multivariate experimental design.
(3) Evaluation, validation, and optimization.

This work deals with the first two of those steps in detail but only covers the evaluation part of step three.

The first step of a multivariate strategy is characterization of the materials (Fig. 1). To have the excipients as well described as possible, they are characterized in multiple variables. The excipients are divided into different classes according to their potential use. For each of the excipient classes, separate PCA models are fitted. By applying PCA to the descriptive data, the important information is extracted in a few principal components (PC). The PCs are often referred to as latent variables or the principal properties (PP) of the data set. Each excipient is assigned a score value in each PC. Thus, the excipients can be compared and related to a



EspNo	Filler 1	Filler 2	Filler 3	Disint 1	Disint 2	Disint 3	Binder 1	Binder 2	Binder 3	Buffer type	BFM/F (ratio)	API	Buffer %	Disint. %	Mgst api %	Disintegration time(min)
1	-57,9	-2,6	-33,1	-31,5	-45,7	-44,9	-37,3	-6,8	-22,4	0	-1	-75,7	0	0,01	0,147	4,8
2	-16,0	-34,4	-11,2	-33,4	-8,5	25,3	-49,3	-0,3	31,5	+1	-1	47,7	0,106	0,05	0,147	51,7
3	-53,6	6,1	-21,6	-33,4	-8,5	25,3	38,6	-1,5	-4,1	-1	1	47,7	0,058	0,01	0,147	>120
4	49,8	11,8	-13,2	-31,5	-45,7	-44,9	48,3	-8,4	5,1	0	1	-75,7	0	0,05	0,147	37,4
5	-4,4	-42,8	17,3	-33,4	-8,5	25,3	22,7	13,3	5,9	0	-1	-75,7	0	0,05	0,147	34,9
6	7,5	-36,8	17,0	-31,5	-45,7	-44,9	32,7	4,6	-3,3	+1	-1	47,7	0,106	0,01	0,147	14,2
7	-28,1	37,5	15,8	-31,5	-45,7	-44,9	-55,3	19,6	18,9	-1	1	47,7	0,058	0,05	0,147	106,8
8	7,1	21,1	32,4	-33,4	-8,5	25,3	-20,8	17,3	-21,8	0	1	-75,7	0	0,01	0,147	16,3
9	-57,9	-2,6	-33,1	-15,6	-56,7	-42,0	22,7	13,3	5,9	0	1	47,7	0	0,01	0,147	>120
10	-16,0	-34,4	-11,2	50,5	-16,9	13,5	32,7	4,6	-3,3	-1	1	-75,7	0,058	0,05	0,147	>120
11	-53,6	6,1	-21,6	50,5	-16,9	13,5	-55,3	19,6	18,9	+1	-1	-75,7	0,106	0,01	0,147	71,4
12	49,8	11,8	-13,2	-15,6	-56,7	-42,0	-20,8	17,3	-21,8	0	-1	47,7	0	0,05	0,147	7,1
13	-4,4	-42,8	17,3	50,5	-16,9	13,5	-37,3	-6,8	-22,4	0	1	47,7	0	0,05	0,147	4,9
14	7,5	-36,8	17,0	-15,6	-56,7	-42,0	-49,3	-0,3	31,5	-1	1	-75,7	0,058	0,01	0,147	77,0
15	-28,1	37,5	15,8	-15,6	-56,7	-42,0	38,6	-1,5	-4,1	+1	-1	-75,7	0,106	0,05	0,147	93,4
16	7,1	21,1	32,4	50,5	-16,9	13,5	48,3	-8,4	5,1	0	-1	47,7	0	0,01	0,147	12,8
17	-57,9	-2,6	-33,1	-3,0	57,3	-27,1	-20,8	17,3	-21,8	+1	1	-75,7	0,106	0,05	0,147	6,7
18	-16,0	-34,4	-11,2	-48,3	5,3	11,4	-55,3	19,6	18,9	0	1	47,7	0	0,01	0,147	>120
19	-53,6	6,1	-21,6	-48,3	5,3	11,4	32,7	4,6	-3,3	0	-1	47,7	0	0,05	0,147	>120
20	49,8	11,8	-13,2	-3,0	57,3	-27,1	22,7	13,3	5,9	-1	-1	-75,7	0,058	0,01	0,147	29,9
21	-4,4	-42,8	17,3	-48,3	5,3	11,4	48,3	-8,4	5,1	+1	1	-75,7	0,106	0,01	0,147	28,3
22	7,5	-36,8	17,0	-3,0	57,3	-27,1	38,6	-1,5	-4,1	0	1	47,7	0	0,05	0,147	>120
23	-28,1	37,5	15,8	-3,0	57,3	-27,1	-49,3	-0,3	31,5	0	-1	47,7	0	0,01	0,147	79,9
24	7,1	21,1	32,4	-48,3	5,3	11,4	-37,3	-6,8	-22,4	-1	-1	-75,7	0,058	0,05	0,147	20,1
25	-57,9	-2,6	-33,1	18,5	27,9	-17,4	48,3	-8,4	5,1	-1	-1	47,7	0,058	0,05	0,147	10,6
26	-16,0	-34,4	-11,2	50,9	-5,5	4,7	38,6	-1,5	-4,1	0	-1	-75,7	0	0,01	0,147	>120
27	-53,6	6,1	-21,6	50,9	-5,5	4,7	-49,3	-0,3	31,5	0	1	-75,7	0	0,05	0,147	111,9
28	49,8	11,8	-13,2	18,5	27,9	-17,4	-37,3	-6,8	-22,4	+1	1	47,7	0,106	0,01	0,147	22,6
29	-4,4	-42,8	17,3	50,9	-5,5	4,7	-20,8	17,3	-21,8	-1	-1	47,7	0,058	0,01	0,147	3,7
30	7,5	-36,8	17,0	18,5	27,9	-17,4	-55,3	19,6	18,9	0	-1	-75,7	0	0,05	0,147	97,5
31	-28,1	37,5	15,8	18,5	27,9	-17,4	32,7	4,6	-3,3	0	1	-75,7	0	0,01	0,147	>120
32	7,1	21,1	32,4	50,9	-5,5	4,7	22,7	13,3	5,9	+1	1	47,7	0,106	0,05	0,147	25,1
33	-10,9	-32,8	-5,8	-8,7	5,4	2,6	-18,0	-0,1	-10,4	-1	0	0,5	0,029	0,08	0,147	2,3
34	-10,9	-32,8	-5,8	-8,7	5,4	2,6	-18,0	-0,1	-10,4	-1	0	0,5	0,029	0,08	0,147	1,1
35	-10,9	-32,8	-5,8	-8,7	5,4	2,6	-18,0	-0,1	-10,4	-1	0	0,5	0,029	0,08	0,147	1,5

Figure 1. A schematic view of the applied multivariate strategy. First spectra are acquired for the excipients. The PCA is then applied to extract the essential information and the excipients are related on a continuous principal property scale. Excipients are chosen from the score plots according to an experimental design to form the multivariate design. The PLS is used to relate the descriptive variables to the measured responses and the results are analyzed. With the PLS model predictions are made for further experiments to verify the model and obtain a formulation with suitable qualities.

continuous scale of PPs. The PPs are assumed to reflect real differences in excipient properties, and larger distances between excipients along the PCs reflect bigger differences in behavior.

The latent variables are continuous variables and can be used as ordinary design factors. This is one of the major improvements of using the multivariate approach. The selections of excipients for the experimental design are no longer "either-or" options, but choices along continuous scales (i.e., the design factors are quantitative instead of qualitative). This enables far more excipients to be evaluated in a limited number of experiments and is useful both when generating and evaluating the experimental design.

Multivariate experimental design is the application of statistical experimental design to the PPs of the excipients. Experimental design is applied to investigate the experimental domain (i.e., all possible combinations of excipients) in a limited number of experiments. Because the PPs represent a measure of excipient quality and the objective is to gain as much information as possible, the excipients chosen should not be too similar. By choosing the excipients according to an experimental design, spanning the widest possible experimental region, excipient diversity is ensured. It is important to emphasize the need for testing also the mixes of excipients that are expected to give poor results. It might at first seem as a waste of good material, but by obtaining diverse excipient properties, the direction toward improved quality in the experimental region is pinned down.

The results from the design are then evaluated with PLS, which efficiently handles the mass of information. The resulting model can be used for interpretation of the results and for predicting new formulations with suitable qualities. The regression coefficients can be used for identifying regions of interest in the latent variables. A new experimental design can be generated and implemented when the region of interest has been identified.

Predictions may be made regarding how to obtain or maintain a certain quality and how to find alternative routes to that goal. The resulting model will also be open for continuous update as new excipients become available. Excipients not included in the original material also can be classified with the aid of soft independent modeling of class analogy (SIMCA), and their expected properties can rapidly be assessed.[5] Additional experiments are then performed to confirm the predictions and the validity of the model.

Here follows a brief description of other chemometric methods used in the different applications covered in this article. For a more thorough theoretical background, there are several textbooks.[6–10]

Experimental Design

The objective of experimental design is to plan and conduct experiments in such a way that maximum information regarding the experimental domain is extracted in the fewest possible experiments.[11]

Full Factorial and Fractional Factorial Designs

The independent variables, often referred to as factors, are experimental variables that can be changed independently of each other. The experimental variables define the investigated area, the experimental domain. The response variables, the dependent variables, are measured results of the performed experiments.

The factors can be either quantitative or qualitative. A quantitative factor is a continuous variable that can take any number between predefined levels in the design. A qualitative factor, on the other hand, is called qualitative because the factor may only be varied at distinct levels such as present/not present, on/off, or Excipient A or B or C.

The experimental variables are given maximum and minimum values based on preexisting knowledge on the topic. In a full factorial experimental design, all factors are changed simultaneously. Not only does this cover the entire area of interest with as few experiments as possible, it also makes it possible to examine interaction effects.

The geometrical representation of the experimental design is a square with two factors and with three factors it is a cube, four factors make up a hypercube, and so on (Fig. 2). If k variables are investigated at two levels, the number of experiments in the full factorial design is given by the expression 2^k.

The number of experiments in an experimental design grows rapidly with an increasing number of variables. When dealing with many variables ($k > 5$), a full factorial design is not the best option, at least not for the purpose of screening. It is more appropriate to use only a fraction of the full factorial design, a fractional factorial design. Depending on the number of factors and experiments, the drawback is lost information caused by confounding of

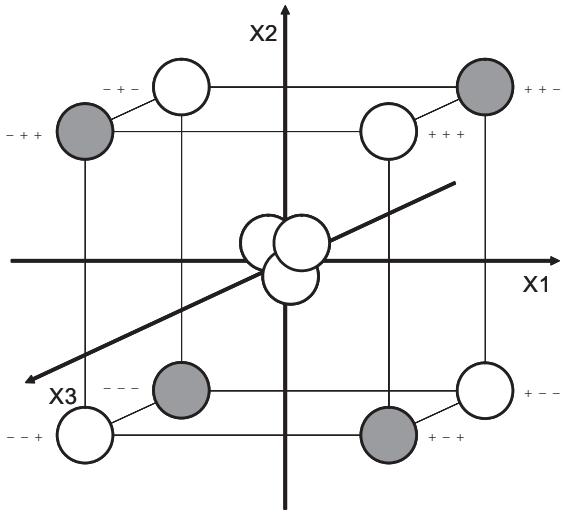


Figure 2. All combinations of the extreme values of the k factors are included as experiments in a full factorial design, illustrated by the spheres in the corners of the cube. Choosing a fraction of a full factorial design will give the maximum amount of variation possible with fewer experimental runs in a fractional factorial design. The white and gray spheres in the corners of the cube highlight different half fractions of the full factorial design in three variables. A center point, the middle value for each variable interval, is also included to detect curvature in the experimental region of the factorial design. It is recommended that the center point be repeated three times for statistic validation (e.g., for estimation of confidence intervals).

interaction effects with main and/or interaction effects. However, confounded effects can be resolved by performing additional experiments. The experiments are performed in random order to eliminate the influence of systematic errors.

Mixture Design

In a mixture design, the sum of all components add up to 100%. Mixture factors are expressed as the fraction of the total amount, and their experimental ranges lie between 0 and 1. This constraint means that the factors cannot be changed independently of one another.[12]

A mixture factor can be a formulation factor or a filler factor. Only one mixture factor can be defined as a filler factor. Formulation factors are the usual mixture factors used in formulations. The chemometric filler factor is a mixture component, usually of little interest, which makes up a large percentage of the mixture and is added at the end of a formulation to bring the mixture total to the desired amount. Presence of the described chemometric filler factor allows for the formulation factors to be changed independently of each other, with reservation for the influence that the varying amounts of filler might have on the mixture. To not confuse the chemometric term filler factor, which we have used in the experimental design, with the type of excipient called filler, the word filler will refer to the excipient throughout the text, unless otherwise stated.

Multivariate Design

Multivariate design, or experimental design in principal properties, as it is often referred to, is the combination of experimental design and PCA.[13] Instead of ordinary factors principal properties obtained from PCA are used as factors in the design (Fig. 3). This makes it possible to considerably reduce the number of experiments in the design and still obtain relevant information (e.g., which variables influence tablet quality).

The basis for the multivariate design is the descriptive variables that are part of the characterization. The descriptive variables can be measured values or relevant data taken from the literature. However, the descriptive variables must contain information relevant to the problem.

Evaluation of Designed Experiments

The analysis of designed experiments typically consists of three stages. The first step is to evaluate the raw data. This is done to check the distribution of the data and to verify that the replicate error is not larger than the total variation in the design points.

The second step is the calculation of models that link the design data to the recorded responses and the interpretation of these models. The fitting of a regression model to the experimental design and the recorded responses is usually done by using least squares techniques such as multiple linear regression (MLR) or PLS.

From the fitted regression model, conclusions can be drawn regarding how the factors influence the responses and how the influence changes over the experimental domain. The size and sign of the scaled and centered regression coefficients indicate the influence of the variables on the responses. The significance of the regression coefficients is determined by statistical tests. The statistical significance of the regression coefficient is indicated by the confidence interval, which is displayed as an error bar. Significant model terms are included in an equation that describes the experimental region as a plane of k dimensions.

The third stage is to use the obtained regression model. This involves predicting the optimum conditions or to find settings for an optimization design. This can be done in a number of ways: by response surface plots or by using the regression coefficients.

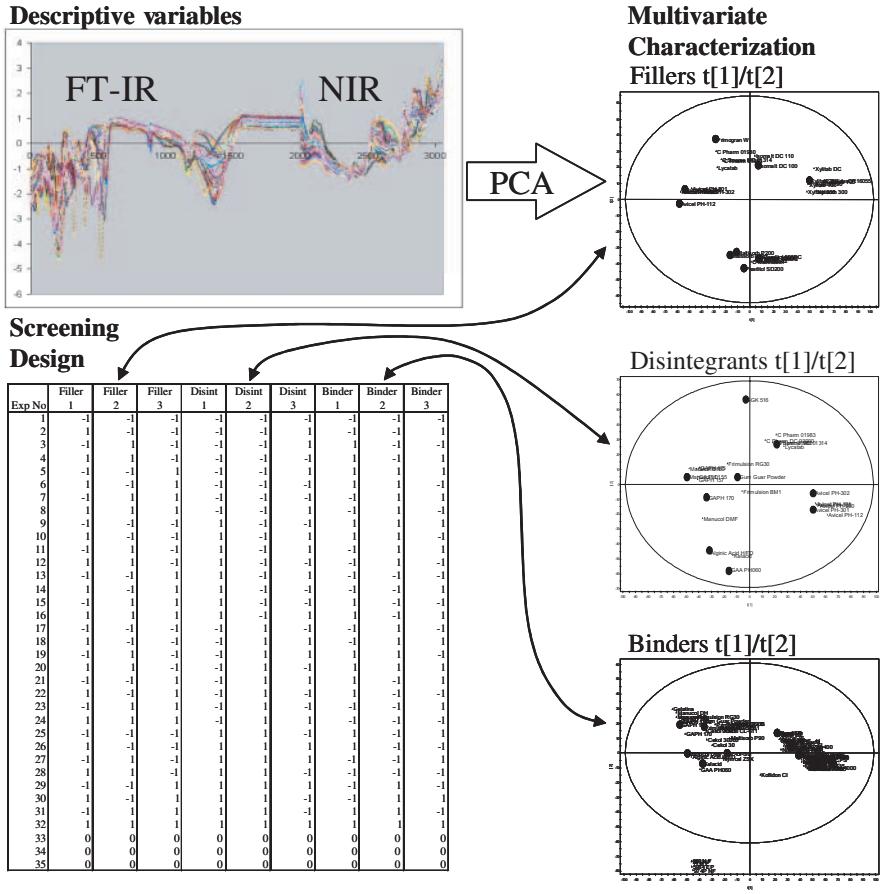
Additional experiments should be performed to confirm the validity of the model. It is important to remember that the model is valid only for the experimental domain.

Multivariate Methods

Extracting useful information from large volumes of data can be a great challenge for the researcher of today. Simply looking at data tables or examining one variable at a time is in most cases not enough. The multivariate projection methods PCA and PLS provide great means for grasping the complex relationships hidden in the data.

Principal Component Analysis

A data matrix **X** typically consists of N objects (rows) described by K variables (columns). In cases in



Exp No	Filler 1	Filler 2	Filler 3	Disint 1	Disint 2	Disint 3	Binder 1	Binder 2	Binder 3
1	-1	-1	-1	-1	-1	-1	-1	-1	-1
2	1	-1	-1	-1	-1	1	-1	-1	1
3	-1	-1	-1	-1	-1	1	-1	-1	-1
4	1	1	-1	-1	-1	-1	1	-1	1
5	-1	-1	1	-1	-1	1	1	1	1
6	1	-1	1	-1	-1	-1	1	1	-1
7	-1	1	1	-1	-1	-1	-1	1	1
8	1	1	1	-1	-1	1	-1	1	-1
9	-1	-1	-1	1	-1	-1	1	1	1
10	1	-1	-1	1	-1	1	1	1	-1
11	-1	1	-1	1	-1	1	-1	1	1
12	1	1	-1	1	-1	-1	-1	1	-1
13	-1	-1	1	1	-1	1	-1	-1	-1
14	1	-1	1	1	-1	-1	-1	-1	1
15	-1	1	1	1	-1	-1	1	-1	-1
16	1	1	1	1	-1	1	1	-1	1
17	-1	-1	-1	-1	1	-1	-1	1	-1
18	1	-1	-1	-1	1	1	-1	1	1
19	-1	1	-1	-1	1	1	1	1	-1
20	1	1	-1	-1	1	-1	1	1	1
21	-1	-1	1	-1	1	1	1	-1	1
22	1	-1	1	-1	1	-1	1	-1	-1
23	-1	1	1	-1	1	-1	-1	-1	1
24	1	1	1	-1	1	1	-1	-1	-1
25	-1	-1	-1	1	1	-1	1	-1	1
26	1	-1	-1	1	1	1	1	-1	-1
27	-1	1	-1	1	1	1	-1	-1	1
28	1	1	-1	1	1	-1	-1	-1	-1
29	-1	-1	1	1	1	-1	1	-1	-1
30	1	-1	1	1	1	-1	-1	1	1
31	-1	1	1	1	1	1	1	1	-1
32	1	1	1	1	1	1	1	1	1
33	0	0	0	0	0	0	0	0	0
34	0	0	0	0	0	0	0	0	0
35	0	0	0	0	0	0	0	0	0

Figure 3. The principles of the generation of a multivariate design. The excipients are characterized in multiple variables and the spectra are subjected to PCA. Excipients from each class are chosen from the respective score plots according to a fractional factorial design. The first experiment should, e.g., contain excipients with negative values in all three PCs (only two dimensions shown here). The original values are then substituted for the principal property values to form the multivariate design. In cases in which none of the excipients fulfill the criteria for a particular experiment, compromises have to be made. Excipients chosen for the multivariate design are encircled.

which a number of objects are described by many variables, the variables tend to be correlated to some extent. In PCA, this is used to describe the variation in data with a minimum number of calculated components.[14] The application of PCA to large quantities of data gives an excellent overview of the data matrix. The PCA model shows the underlying structure in the data and allows for the detection of trends, groupings and outliers. The principles of PCA are illustrated in Fig. 4. The values of the score matrix **T** can be used to relate and compare objects to one another. Similar objects will have similar score values and dissimilar objects will not.

How the original variables K are weighted together in a PC is expressed by the loadings matrix **P**. If the original variable is highly correlated with the PC it is given a high loading value. The loading values range from -1 to $+1$, and a loading value of 0 indicates that this variable does not contribute with information to this PC. The loadings also give information about how variables are correlated. A similar loading value indicates a positive correlation, and variables with loading values of opposite signs are negatively correlated to each other. Together, the scores and loadings describe the principal components of the data set.

The objects' distances to the model are expressed by the residual matrix **E** (i.e., the variation in the data that is not explained by the model). An outlier is an object that does not fit very well into the model (e.g., the distance to the model is greater than what is accepted).

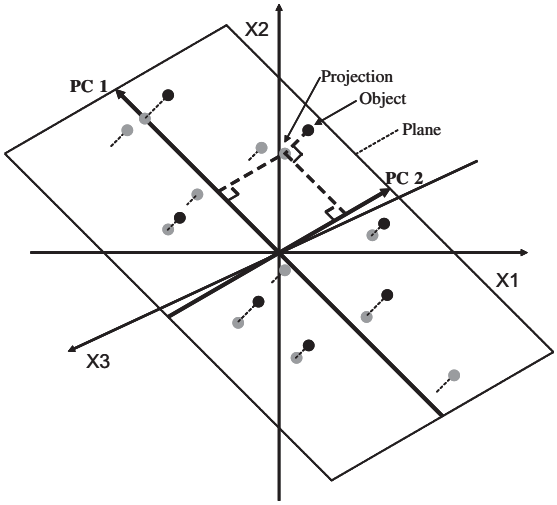


Figure 4. The PCA corresponds to a least squares fitting of a straight line ($A=1$) or an A-dimensional plane/hyperplane to the data in the K-dimensional variable space. In this case, the data are mean centered, and three variables are described by two PCs. The object is projected onto the mathematical plane described by the PCs, and the score value in each component is obtained by determining the distances from the origin to the projected object. The loadings are the direction coefficients of the fitted plane. The perpendicular distance between the object and the plane is the distance to the model.

When calculating the PCA model, the first PC will describe as much variation as possible. Components are fitted until enough of the variation in the data has been described. To decide the number of significant components, eigenvalue or R^2 and Q^2 are used. The eigenvalue should be larger than 2 for a significant component. The R^2 is a measure of how much of the variation in **X** that is explained by the model. Q^2 explains how much of the variation in **X** that can be predicted by the model.

The score vectors for two principal components (e.g., $\mathbf{t}_1$ and $\mathbf{t}_2$) together span a mathematical plane. By plotting the values in a score plot, groupings and trends in data and outliers can be detected.

Projections to Latent Structures

Partial least squares projections to latent structures (PLS) is to take PCA one step further as it deals with both **X** and **Y** data.[15] The underlying structures in **X** and **Y** data sets are related to each other. The first PLS component is calculated to well approximate the variation of objects in **X** and **Y** and to provide the best possible correlation between the projections. Each object is represented in both X-space (**T**) and Y-space (**U**). In the t/u scores plot, the correlation between X and Y is shown, which should be linear. The regression coefficients in PLS are a summary of the loading weights, and they can be used to interpret the model and to make predictions of **Y** from **X** data.

EXPERIMENTAL

Spectroscopic Characterization

A total of 105 excipient samples, 21 different excipients of varying quality types and from different producers, were considered for the new formulation. The samples were characterized by FT-IR (649–4500 cm^{-1}) (Mattson Fourier-transform 60AR instrument equipped with a Golden Gate Single Reflection Diamond ATR accessory) and NIR (400–2500 nm) (NIR Systems 6500 Spectrophotometer, FOSS) spectroscopy. The variables of the multivariate characterization are the digitized NIR (1050 variables) and FT-IR (1997 variables) spectra.

The excipients were divided into different classes according to their potential use as fillers, binders, or disintegrants (Appendices I, II, and III). Excipients can have dual functions. Batches of active pharmaceutical ingredient (API) from two manufacturers were also characterized.

The FT-IR and NIR spectra for the excipients in the different classes were pretreated separately with the standard normal variate transformation (SNV) and then combined to form the basis for the multivariate characterization.[16] The SNV transformation was performed to remove interferences from, e.g., sample packing.

The PCA models were calculated, and the different classes were described by three principal components, except for the active substance. There were samples from six batches of API, three from each of the two manufacturers, which were described by one principal component. The number of components was decided by the eigenvalue criteria.

Generation of the Multivariate Design

A fractional factorial design in 14 variables, 2^{14-9} design, and 35 experiments is the basis for the multivariate design used for the screening of the excipients. To form the multivariate design, excipients were chosen to correspond as well as possible to the

Table 1. Factor names, abbreviations, and a brief description of each of the factors in the design.

Number	Factor	Abbreviation	Explanation
1	Filler 1	Fill 1	The first principal property for the fillers
2	Filler 2	Fill 2	The second principal property for the fillers
3	Filler 3	Fill 3	The third principal property for the fillers
4	Binder 1	Bind 1	The first principal property for the binders
5	Binder 2	Bind 2	The second principal property for the binders
6	Binder 3	Bind 3	The third principal property for the binders
7	Disint 1	Dis 1	The first principal property for the disintegrants
8	Disint 2	Dis 2	The second principal property for the disintegrants
9	Disint 3	Dis 3	The third principal property for the disintegrants
10	Buffer type	BuffS	Buffer, −1 is NaCarb and 1 is NaGly
11	BFMF (ratio B/F)	BFMF	Ratio, −1 = binder/filler 30/70, 1 = binderfiller 70/30, 0 = 50/50
12	API score	API	Principal property for the active substance
13	Amount of buffer	BuffM	Amount of buffer
No factor	Amount of binder and filler		Amount of filler and binder in the design
14	Amount of disintegrant	DisM	Amount of disintegrant
No factor	Amount of MgSt and API		Fixed amount of lubricant (MgSt) and API

fractional factorial design (Fig. 3) (Appendix IV). The multivariate design is then ''translated'' into actual excipients for the work plan. The design is of resolution IV, which means main effects are not confounded with two-factor interactions.

The multivariate design includes 10 factors that describe the excipients and the active substance. The other four factors are part of or related to a mixture design with a filler factor (Table 1). The mixture design consists of only three constituents: buffer, disintegrant, and filler. The chemometric term filler factor in this case actually consists of both binder and what is pharmaceutically termed filler or diluent. Because the ratio between binder and filler is more interesting than the actual amount of each excipient, a separate factor describes the ratio (30/70 or 70/30) of binder and filler that make up the chemometric filler factor in the design. The amount of buffer, when present, was decided to be double the stoichiometric amount needed to neutralize the active substance. This means that the type of buffer will influence the amount of buffer, but this was still the best solution to obtain a manageable design for the screening experiments. One factor describes the type of buffer. This factor is set to 0 when no buffer is present in the formulation.

Selection of Excipients

The section of the basic design that is the basis for the selection of the excipients is shown in Fig. 3.

The first experiment should contain excipients with negative values in all three PCs. Looking closely at the $t[1]/t[2]$ score plots, one might be able to identify the filler as Avicel PH-112, the disintegrant as Alginic Acid H/FD, and the binder as Kelacid. Looking at the rest of the design, there are no excipients to match all the descriptions, so compromises had to be included. Maltisorb P90 was chosen as a substitute for a filler with a positive value in the first PC and negative values in the second and third (Appendix I). There were no disintegrants with a positive value in the first PC and negative values in the second and third or one with positive values in all three PCs, Alginic Acid PH060 and Avicel PH-302, respectively, were used instead.

Manufacturing and Characterization of Tablets

In Table 2, the general formulation of the tablets is summarized. The tablets were made by direct compression. Tablets were compressed at forces of 3, 6, and 9 kN. This article focuses on the results for tablets compressed at 6 kN. The active substance, and when included also the buffer, and all excipients except for the lubricant were weighed and sieved through a 1.68-mm sieve. The mixture was then blended for 8 min in a Turbula mixer. The lubricant was sieved separately through the same sieve and then added to the mixture and mixed for 2 min. The total mixing time was 10 min and the batch size was 500 g.

Table 2. The levels of filler, binder, disintegrant, and buffer were varied in the study as given in the table.

Component	Percentage (w/w)
API	13.7
Filler	21.0–59.0
Binder	21.0–59.0
Disintegrant	1.0–5.0
Buffer	0–10.6
Lubricant	1

The tablets were compressed on a Korsch EKO tabletting machine equipped with a computerized system (PuuMan, Portable Press Analyzer) for collection of compression data. Compression force and ejection force were recorded and are given as averages of 30 measurements. About 150–200 tablets were made from each batch.

The disintegration time (ErwekaZT 3–4) was recorded for six tablets; the average and RSD were calculated. The manner in which the tablets disintegrated also was noted, referred to as disintegration type. Twenty tablets were weighed (Mettler Toledo AT200), and the average and RSD for the weight were calculated. The crushing strength (Schleuniger 2E/205) was recorded for 10 tablets, and the averages and the RSD were calculated.

An overall impression of the adhesion of materials to the equipment during the tabletting was noted for each tablet formulation. Also the ratio between tapped and bulk density, the Hausner ratio, was determined for all powder mixes. The responses are presented in Appendix V.

Possible effects of compression force are also evaluated by including it as a constructed factor in a design (Appendix IV).

Statistical Analysis

All PCA models were calculated in Simca-P 8.0. The experimental designs were generated and analyzed by using Modde 5.0. All software was supplied by Umetrics AB, Umeå, Sweden.

RESULTS AND DISCUSSION

Multivariate Characterization

Included in the study are 100 excipient samples that were characterized by FT-IR and NIR spectroscopy. The excipients can have dual functions, and in this study there are 28 excipient samples that are potential fillers, 67 samples of potential binders, and 27 samples of potential disintegrants. Batches of API, three from each of two manufacturers, were also characterized. The PCA models are good; they generally describe about 80% of the variation in the spectral data.

The three classes are discernable in the $t[1]/t[2]$ score plot (Fig. 5). There is some natural overlap, because some excipients do not exclusively belong to just one class. Showing all of the relevant score plots would take up too much space in this publication. However, all score plots can be constructed from the values given in the appendices.

PCA of Fillers

Included in this model are 28 filler samples characterized by 3047 variables. The number of significant components was determined by using the eigenvalue criteria. The model has three principal components that explain 87% of the variation in the X data. All score values are given in Appendix I.

The score plot of the predicted scores for the fillers display several groupings (Fig. 3). The groupings contain excipients of similar qualities, often the same kind of excipient but with a different particle size. The center point is not as close to the center as desirable, but it is still the best candidate when all three PCs are taken into account. The excipients are rather well spread out in the score plot, which is desirable when making selections for the experiments.

There is only one combination that was not available; there is no excipient that has a positive value in the first PC and negative values in the other two. Maltisorb P90 (maltitol) with negative values in all three PCs was chosen as a substitute.

PCA of Binders

The PCA model includes 67 binder samples characterized by 3047 variables. The number of significant components according to the eigenvalue criteria is four. The fourth component only explained 6% of the variation in X and was, therefore, left out of the design. The three components included in the design explain 80% of the variation in the X data. All score values are given in Appendix II.

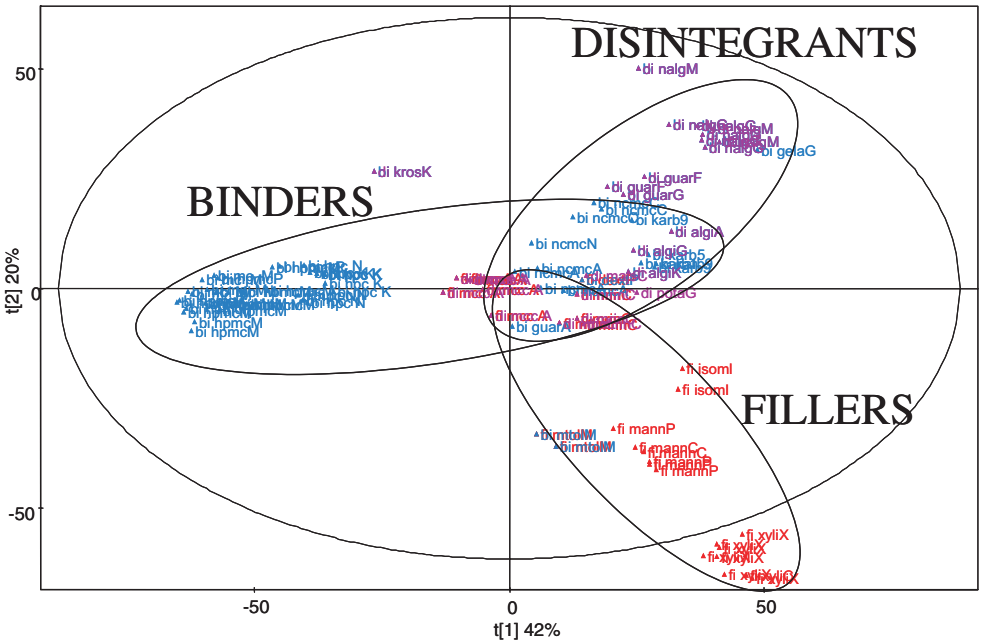


Figure 5. All excipients included in the study. The excipients are marked according to their potential use, fi for fillers, bi for binders, and di for disintegrants. The % explained variance of each PC is given next to the component number.

The deviating objects in the lower left of the $t[1]/t[2]$ score plot are carbomer (karbomer) batches and those were not considered for the design (Fig. 3). The material is divided into two parts in this plot; to the right are binders of cellulose type—methylcellulose (MC), hydroxypropyl cellulose (HPC), and hydroxypropyl methylcellulose (HPMC)—and to the left are binders of other types—carboxymethylcellulose sodium (Sodium CMC), Guar Gum, alginates, and alginic acids.

Four excipients from each half of the material have been chosen. The excipients are divided mainly by the first PC; in the second and third PCs, the excipients are well spread out (see Appendix II).

PCA of Disintegrants

The PCA model of the disintegrants includes 26 excipient samples characterized by 3047 variables. There are three significant components, according to the eigenvalue criteria. The model explains 88% of the variation in the X data. All score values are given in Appendix III. The disintegrants are well spread out in the $t[1]/t[2]$ score plot (Fig. 3). Also in the third PC, the excipients are well spread out. Yet, there were no disintegrants with positive values in all three or positive in the first and negative in the other two PCs.

PCA of API

Three batches of API from each of two manufacturers, S1 and S2, are included in this PCA model. They are characterized by 3047 variables. Only one PC was significant according to the eigenvalue criteria, but two components are depicted in the score plot to make the illustration easier. The one component model explains 74% of the variation in the X data.

The batches from S2 show less conformity than the batches from S1 (Fig. 6). The outcome of the experiments will determine whether these differences are large enough to affect the tabletting or the tablet quality. Because of limited availability, the extremes from the respective manufacturer were chosen.

Results of Screening Design

Disintegration time and crushing strength are the most important responses for this product. Acceptable qualities that are within the range of the product (disintegration time 15–30 min and crushing

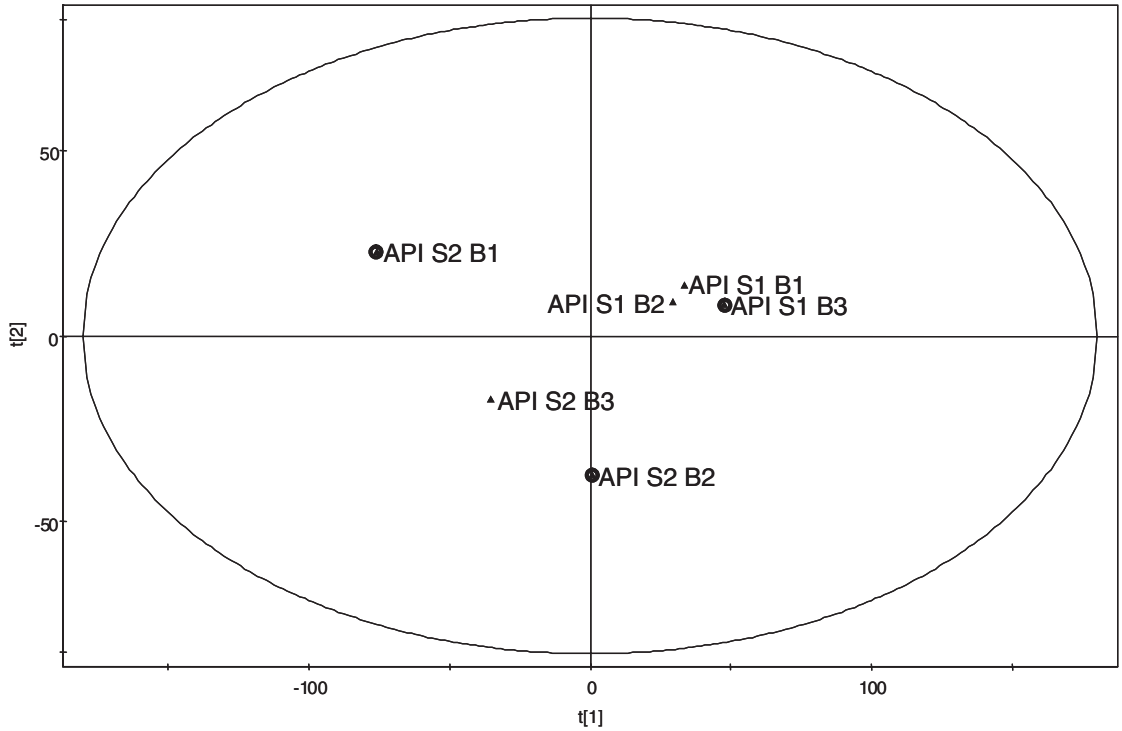


Figure 6. The $t[1]/t[2]$ score plot for the API. N.B. the second PC is not significant. Batches chosen for the multivariate design are encircled.

strength >6 kp) were obtained with some of the formulations.

PLS Model for Disintegration Time

The disintegration time was limited to 120 min. The PLS model for disintegration time contains only 27 experiments. The tablets from experiments N3, N9, N10, N18, N19, N22, N26, and N31 never disintegrated completely. They were excluded because these results severely impair the model. The loss of information from the exclusion of these experiments is a drawback. On the one hand, none of the combinations of excipients in the discarded experiments seem like interesting alternatives, but on the other hand this is the kind of vital information that should be included in the model. Attempts to find a common denominator for the formulations to elucidate the cause of this behavior do not give a straightforward answer. The response was log-transformed to be normally distributed.

The default PLS model for disintegration time is significant according to ANOVA and has no lack of fit. Four factors were still removed from this model (Fill1, Dis1, Bind2, and DisM). This did not improve the model with regards to R^2 or Q^2 values, but two of the regression coefficients (BuffS and API) became significant.

The model is significant according to ANOVA and has no lack of fit. The model has two PLS components and a R^2Y of 0.83 and a Q^2 of 0.69.

The center points (N33, N34, and N35) are located in the lower left of the $t[1]/u[1]$ score plot (Fig. 7). They are not at the center of either the X-space or the Y-space, and this is not ideal.

The main influence on the disintegration time is Bind3 (Fig. 8). The binders tested in this study with varying values in the third principal property have different and consistent disintegration times. The regression coefficient for the second PP for the fillers is also significant. The choice of buffer as well as the supplier of API also affects the disintegration time.

It was expected that the BFMF would influence the PLS model for disintegration time, but it did not. However, the intervals in which it was varied and the difference in excipient quality were simply too great. The conditions are not optimal for establishing a detailed quantitative relation, but they are sufficient for a screening study.

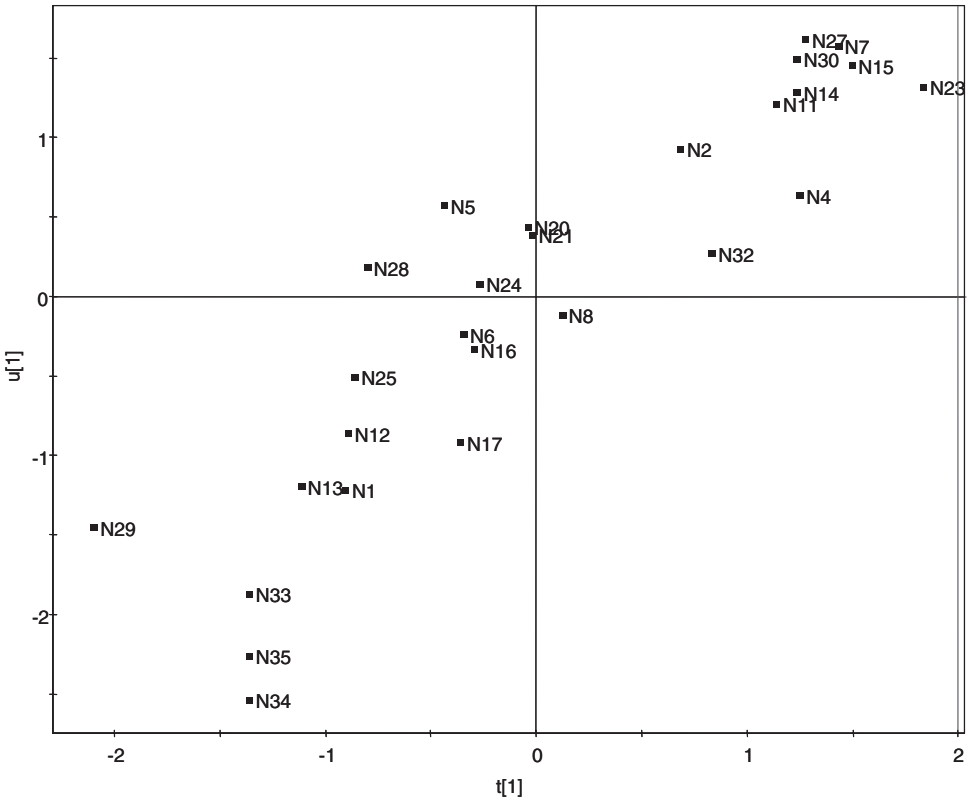


Figure 7. The $t[1]/u[1]$ score plot for the PLS model for disintegration time. The relationship between t and u is satisfactorily linear.

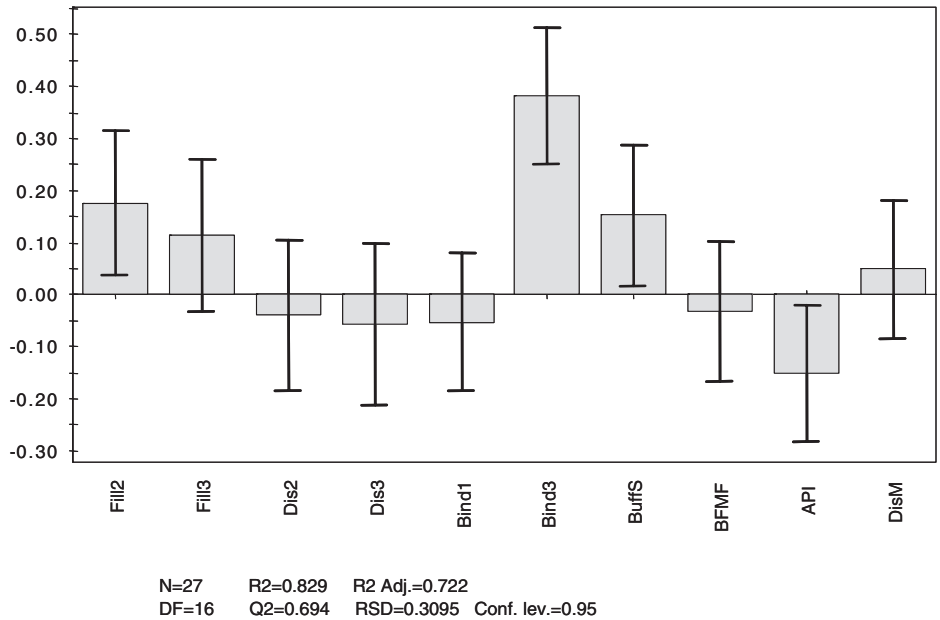


Figure 8. Regression coefficients plot for the PLS model for disintegration time. The error bars in the plot are the confidence intervals for the regression coefficients at the 95% level.

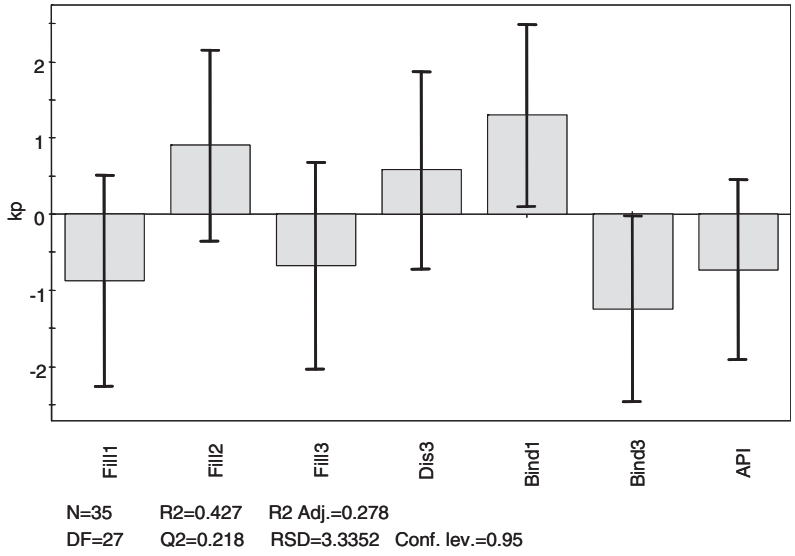


Figure 9. Regression coefficients plot for the PLS model for crushing strength. The model has two components.

The center points have very short disintegration times (~2 min). This low value of the center points indicates some non-linear relationship in the material.

PLS Model for Crushing Strength

The PLS model for crushing strength is significant according to ANOVA, but only after removal of several non-significant model terms. The model still has significant lack of fit. The model is still accounted for because it is such an important response when dealing with tabletting.

The choice of binder affects the crushing strength and the regression coefficients for the first and third PP's for binder are significant (Fig. 9). A positive value in the first and a negative value in the third binder PP will increase the crushing strength.

PLS Model for Ejection Force

All the tablet formulations had acceptable ejection forces. The PLS model is significant according to ANOVA and has no lack of fit.

There are several significant regression coefficients in the model for the ejection force (Fig. 10). The choice of both filler and binder, and maybe to a lesser extent the disintegrant, affects the ejection force. To decrease the ejection force, fillers with negative values in the first PP and positive in the second PP are possible candidates. A binder with a positive value in the first PP will also decrease the ejection force. The choice of disintegrant could be based on the first two PPs and a negative value in the first and a positive value in the second seems favorable.

The type of buffer should be sodium glycinate (coded +1 in the worksheet) and to decrease the ejection force further the larger ratio of binder to filler should be used (binder/filler 70/30 coded +1 in the worksheet).

PLS Model for Disintegration Type

The type of disintegration is graded between 0 and 2; 0 is bursting, 0.5 is bursting/erosion, 1 is erosion, 1.5 is erosion/swelling, and 2 is swelling. The tablets that swell generally have long disintegration times, if they disintegrate at all. Only one tablet disintegrated by bursting, object 34. Because it is one of the center points, object 34 has a large residual. Objects 34 and 26 were removed from the model because of their large residuals. The two component PLS model is significant according to ANOVA and has no lack of fit.

The PLS model has a R^2Y of 0.65 and Q^2 of 0.24 (Fig. 11). The model is significant but has a large discrepancy between the R^2 and Q^2 values.

With more formulations included in the model than in the disintegration time model, Bind3 still

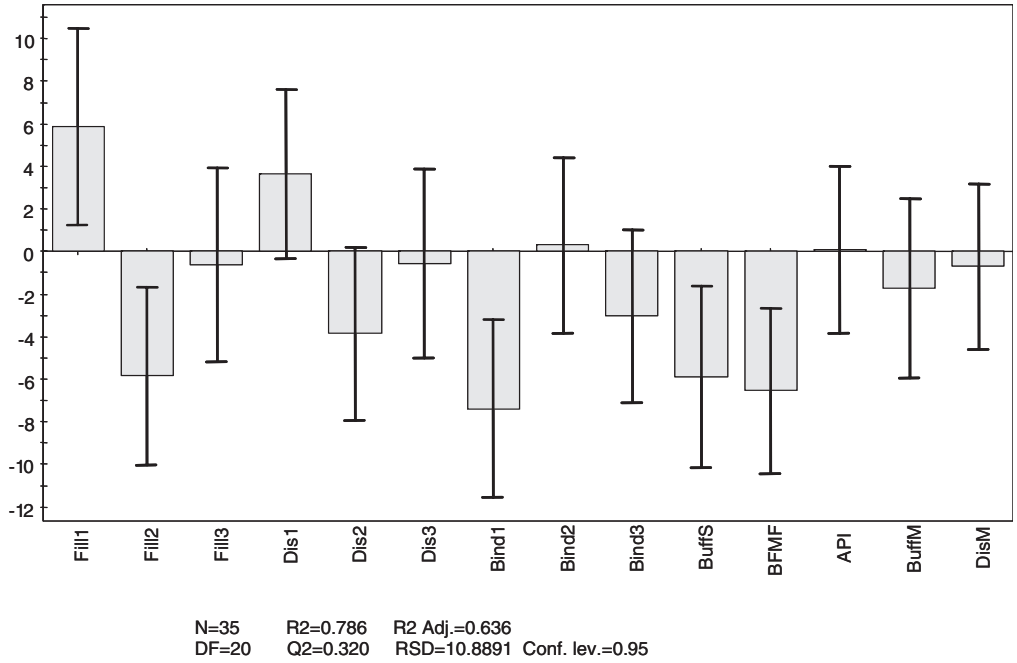


Figure 10. Regression coefficients from the PLS model for ejection force. The PLS model has two significant components.

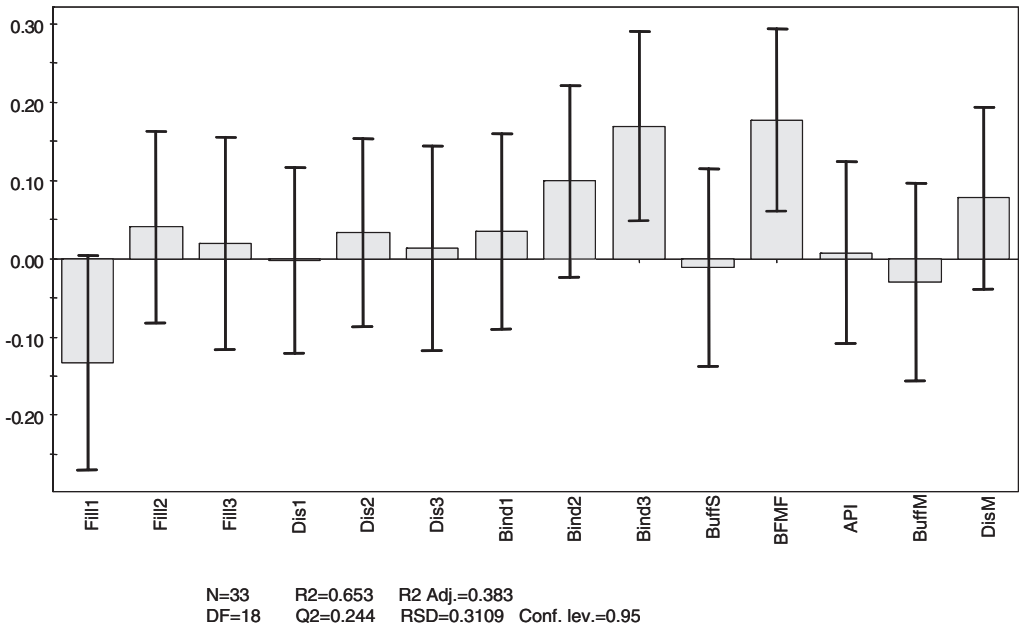


Figure 11. Regression coefficients plot for the PLS model for disintegration type.

dominates the influence on the disintegration. Formulations with a high value in Bind3 generally have more swelling than others. The high amount of binder in the BFMF also introduces more swelling. This indicates that the BFMF might influence the disintegration time, because swelling usually leads to a long disintegration time. The coefficient for the first PP for the fillers (Fill1) is very close to being significant.

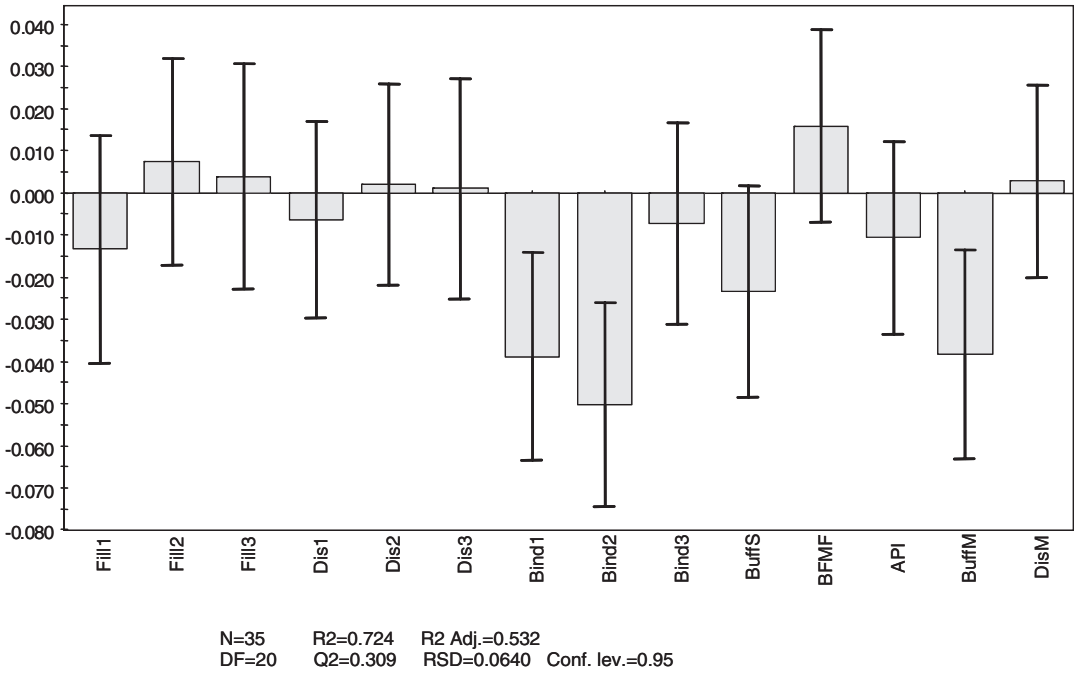


Figure 12. Regression coefficients plot for Hausner ratio. The PLS model for Hausner ratio has two significant components.

PLS Model for Hausner Ratio

The Hausner ratio is a measure of the flow properties for the powder mixes. A low value for the Hausner ratio means good flow properties for the powder mixture. The average Hausner ratio was 1.4, which is a high value.

The model is significant according to ANOVA, has no lack of fit and the residuals are normally distributed.

The PLS model has a R^2Y of 0.72 and a Q^2 of 0.30; the large discrepancy between the R^2 and the Q^2 values is indicative of a rather poor model (Fig. 12). To obtain good flow according to the Hausner ratio, a binder with positive values in the first and second PPs for the binders should be used. The mixtures with sodium glycinate buffer have lower Hausner ratio values than mixtures containing sodium carbonate. The buffer actually seems to improve the flow properties of the powder mixture, compared with the rest of the excipients as the Hausner ratio decreases with increased amount of buffer.

PLS Model for Adhesion

The adhesion was graded from 0 to 3 where 0 means no adhesion, 1 is adhesion, 2 is severe adhesion, and 3 implies very severe adhesion. Objects 26, 32, and 21 were excluded because of large residuals.

The PLS model has two components and is significant according to ANOVA. The center points were determined to have the same adhesion, so no lack of fit could be calculated.

The PLS model has two components and explains 40% of the variation in Y and has a Q^2 of 0.21 (Fig. 13). The model is not good and only tentative conclusions can be drawn. The buffer seems to cause adhesion of the materials. The type of buffer that causes more adhesion is sodium glycinate (coded +1 in the worksheet), but the main influence is the amount of buffer.

PLS Model for Evaluation of Compression Force

Generally, tablets were compressed at 6 kN. To evaluate the influence of the compression force on the responses, a design was generated with compression force (3, 6, or 9 kN) included as a factor (see Appendix IV). The results of these experiments are given in Appendix V.

Changing the compression force significantly affects only three of the responses: the crushing strength, the tablet height, and the ejection force.

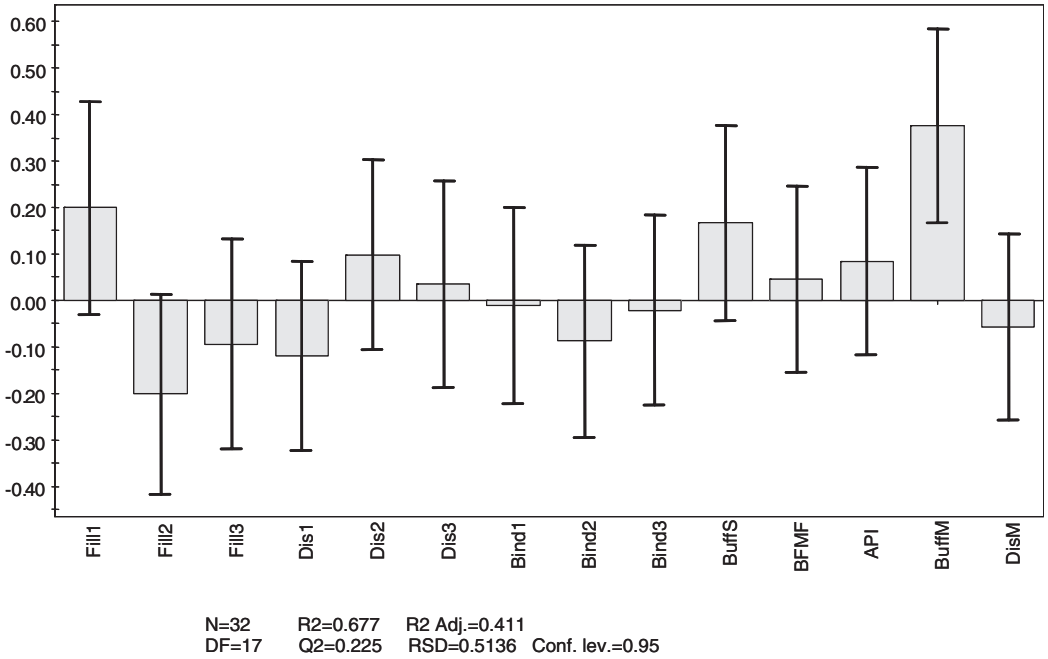


Figure 13. Regression coefficients for the PLS model for adhesion.

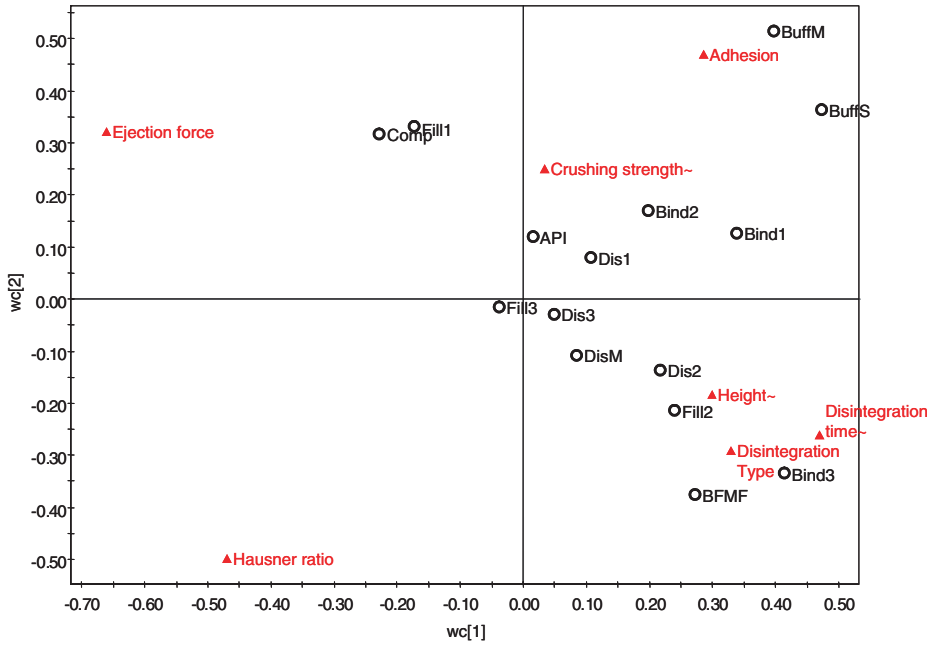


Figure 14. Loadings plot for the first two PLS components of the model with compression force (Comp) included as a factor. The responses are marked with solid triangles and the factors with open circles.

The correlations indicated in the wc[1]/wc[2] loadings plot have been verified in separate models (Fig. 14).

It is not surprising that the compression force is positively correlated with crushing strength (i.e., the crushing strength increases with increased compression force). Compression force is also somewhat positively correlated to ejection force. This is probably due to larger radial forces with increasing

compression force. For tablets compressed at 3, 6, and 9 kN, the average ejection forces are about 35, 45, and 50 N respectively. The model for tablet height has significant lack of fit, but the tablet height decreases with increased compression force.

CONCLUSIONS

Generally, the objects are well spread in the score plots, which enabled selections to be made without much compromising. The explained variation in X is typically around 80% for the PCA models that the design is based upon. Thus, most of the spectral information is accounted for in the multivariate characterization. This indicates that the multivariate characterization was successful. Further experiments will be performed to determine if the descriptive variables of the multivariate characterization contain enough information that is relevant for this application.

Formulations with acceptable values for disintegration time and crushing strength were obtained with some of the formulations.

The main influence on the disintegration time is the third PP for the binders and the second PP for the fillers. Different suppliers do not produce API of equal quality. The amount of disintegrant was probably too small compared with the varying amounts of binder and filler. The amounts of binder and filler also were varied in much too large intervals, considering the variety of binders and fillers included in the study. The conditions are not the best for establishing a detailed quantitative relation, but they are sufficient for a screening study. One conclusion from this study is that for a multivariate design such as this to be more reliable (i.e., to obtain more stable models), the amounts should be fixed or varied over somewhat narrower intervals. To determine the proper amounts of the excipients in the formulation (i.e., optimization), additional experiments will be performed.

The multivariate experimental design strategy for the screening study resulted in PLS models that were, according to cross-validation, generally good (disintegration time, ejection force) or significant but of less value (crushing strength, disintegration type, Hausner ratio, and adhesion). Consequently, conclusions or tentative conclusions regarding the influence from the formulation factors on the responses can be drawn. The models will be used to identify a region of interest for the optimization.

The strategy is general and can be applied in many different areas of pharmaceutical research and development.

APPENDIX

Appendix I. Score values for all fillers in the study.

ObsNum	ObsName	Type of filler	t[1] (52%)	t[2] (20%)	t[3] (14%)
1	**Pearlitol SD200**	**Mannitol**	**−4.4**	**−42.8**	**17.3**
2	Pearlitol 300DC	Mannitol	7.2	−35.9	15.5
3	**Pearlitol 400DC**	**Mannitol**	**7.5**	**−36.8**	**17.0**
4	Pearlitol 500DC	Mannitol	10.5	−35.6	15.0
5	C Mannidex	Mannitol	1.2	−38.8	15.1
6	**Avicel PH-101**	**Microcrystalline cellulose**	**−53.6**	**6.1**	**−21.6**
7	Avicel PH-102	Microcrystalline cellulose	−55.3	5.0	−24.3
8	**Avicel PH-112**	**Microcrystalline cellulose**	**−57.9**	**−2.6**	**−33.1**
9	Avicel PH-200	Microcrystalline cellulose	−57.2	4.9	−21.9
10	Avicel PH-301	Microcrystalline cellulose	−48.0	7.0	−26.0
11	Avicel PH-302	Microcrystalline cellulose	−42.3	5.1	−20.2
12	Xylitab DC	Xylitol	53.6	19.5	−9.5
13	Xylisorb 90	Xylitol	53.6	9.9	−19.3
14	Xylisorb 300	Xylitol	55.3	5.4	−14.6
15	Xylisorb 700	Xylitol	61.8	10.9	−8.4
16	Xylitab 100	Xylitol	48.5	9.6	−16.4
17	**Xylitab 200**	**Xylitol**	**49.8**	**11.8**	**−13.2**

(*continued*)

Appendix I. Continued.

ObsNum	ObsName	Type of filler	*t*[1] (52%)	*t*[2] (20%)	*t*[3] (14%)
18	Xylitab 300	Xylitol	47.6	5.0	−23.4
19	C XylidexCR16055	Xylitol	61.8	11.9	−8.4
20	Lycatab	Maltodextrin	−26.8	20.1	19.2
21	**C Sperse MD 01314**	**Maltodextrin**	**−23.4**	**25.3**	**22.4**
22	C Pharm 01980	Maltodextrin	−27.9	30.1	26.4
23	C Pharm 01983	Maltodextrin	−22.0	24.3	28.3
24	**Maltisorb P90**	**Maltitol**	**−16.0**	**−34.4**	**−11.2**
25	**Maltisorb P200**	**Maltitol**	**−10.9**	**−32.8**	**−5.8**
26	Primogran W	Dextrin	−28.1	37.5	15.8
27	**Isomalt DC 100**	**Isomalt**	**7.1**	**21.1**	**32.4**
28	Isomalt DC 110	Isomalt	4.3	27.3	36.7

Fillers chosen for the experiments are given in bold letters. The contribution of each PC to the explained variance is given next to the component number.

Appendix II. Score values for the binders in the study.

ObsNum	ObsName	Type of binder	*t*[1] (51%)	*t*[2] (19%)	*t*[3] (9%)
1	Maltisorb P90	Maltitol	−15.9	10.2	−56.5
2	Maltisorb P200	Maltitol	−18.0	20.0	−54.9
3	Primogran W	Dextrin	−27.4	19.6	−24.7
4	**Pharmacoat 603**	**Hydroxypropyl methylcellulose**	**48.3**	**−8.4**	**5.1**
5	Pharmacoat 615	Hydroxypropyl methylcellulose	47.2	−6.3	1.5
6	Pharmacoat 904	Hydroxypropyl methylcellulose	29.8	7.2	0.8
7	Metolose 60SH-4000	Hydroxypropyl methylcellulose	49.0	−10.1	4.1
8	Metolose 65SH-400	Hydroxypropyl methylcellulose	42.1	−1.0	0.1
9	Metolose 65SH-4000	Hydroxypropyl methylcellulose	42.2	−2.2	−0.7
10	**Metolose 90SH-400**	**Hydroxypropyl methylcellulose**	**32.7**	**4.6**	**−3.3**
11	**Metolose 90SH-4000**	**Hydroxypropyl methylcellulose**	**38.6**	**−1.5**	**−4.1**
12	Methocel E3	Hydroxypropyl methylcellulose	43.0	−1.8	4.5
13	Methocel E50	Hydroxypropyl methylcellulose	47.9	−5.3	−0.5
14	Methocel E4M	Hydroxypropyl methylcellulose	47.3	−4.8	1.0
15	Methocel F50	Hydroxypropyl methylcellulose	37.4	1.2	1.4
16	Methocel F4M	Hydroxypropyl methylcellulose	48.0	−8.9	0.3
17	Methocel K100	Hydroxypropyl methylcellulose	30.7	3.3	1.3
18	Methocel K4M	Hydroxypropyl methylcellulose	36.5	2.7	−3.9
19	Pharmacoat 603,3 CPS	Hydroxypropyl methylcellulose	41.2	−4.1	7.1
20	Klucel EF	Hydroxypropyl cellulose	21.6	14.0	4.1
21	Klucel GF	Hydroxypropyl cellulose	22.2	12.7	7.5
22	**Klucel HF**	**Hydroxypropyl cellulose**	**22.7**	**13.3**	**5.9**
23	Klucel JF	Hydroxypropyl cellulose	24.4	10.2	7.5
24	Klucel LF	Hydroxypropyl cellulose	22.3	13.6	7.4
25	Klucel MF	Hydroxypropyl cellulose	23.5	9.6	7.6
26	Nisso-SSL	Hydroxypropyl cellulose	25.0	2.6	6.8
27	Nisso-SL	Hydroxypropyl cellulose	27.6	6.0	2.1
28	Nisso-L	Hydroxypropyl cellulose	27.1	4.8	1.2
29	Nisso-M	Hydroxypropyl cellulose	25.7	7.6	1.3
30	Nisso-H	Hydroxypropyl cellulose	26.5	8.8	0.6
31	Cekol 30	Sodium carboxymethylcellulose	−29.9	5.5	2.1
32	Cekol 2000	Sodium carboxymethylcellulose	−34.4	15.1	6.1

Appendix II. Continued.

ObsNum	ObsName	Type of binder	*t*[1] (51%)	*t*[2] (19%)	*t*[3] (9%)
33	Cekol 30000	Sodium carboxymethylcellulose	−34.0	9.2	6.7
34	Nymcel ZSX	Sodium carboxymethylcellulose	−21.9	−3.6	−8.2
35	**Ac-Di-Sol**	**Sodium carboxymethylcellulose**	**−18.0**	**−0.1**	**−10.4**
36	**Avicel RC-581[a]**	**Sodium carboxymethylcellulose**	**−20.8**	**17.3**	**−21.8**
37	Avicel RC-591[a]	Sodium carboxymethylcellulose	−24.7	17.5	−25.4
38	Avicel CL-611[a]	Sodium carboxymethylcellulose	−21.4	15.3	−26.2
39	Metolose SM-15	Methyl cellulose	43.2	−8.6	3.4
40	Metolose SM-400	Methyl cellulose	43.8	−3.4	−1.2
41	Metolose SM-4000	Methyl cellulose	41.3	−2.7	0.2
42	Methocel A15	Methyl cellulose	41.2	−5.8	4.5
43	Methocel A4C	Methyl cellulose	43.1	−6.7	1.8
44	Methocel A15C	Methyl cellulose	45.7	−10.1	1.7
45	Methocel A4M	Methyl cellulose	46.4	−10.5	2.5
46	Gum guar powder	Guar gum	−37.8	22.2	−6.2
47	Avicel CE-15	Guar gum	−13.7	19.7	−34.3
48	Frimulsion BM	Guar gum	−36.1	17.2	−5.5
49	Frimulsion RG30	Guar gum	−41.6	24.8	−2.2
50	5984 EP	Carbomer	−45.8	−76.8	−2.4
51	974P NF	Carbomer	−44.9	−79.3	4.6
52	980 NF	Carbomer	−45.9	−73.0	−4.2
53	981 NF	Carbomer	−46.3	−73.2	−3.4
54	1382	Carbomer	−43.9	−74.8	−2.0
55	Grindsted PH060	Alginic acid	−39.4	−11.0	−19.0
56	**Kelacid**	**Alginic acid**	**−37.3**	**−6.8**	**−22.4**
57	Alginic acid H/FD	Alginic acid	−46.3	−2.3	−13.6
58	Grindsted PH 150	Sodium alginate	−55.4	24.0	18.1
59	Grindsted PH 155	Sodium alginate	−55.4	22.1	15.6
60	**Grindsted PH 157**	**Sodium alginate**	**−55.3**	**19.6**	**18.9**
61	Grindsted PH 170	Sodium alginate	−51.3	13.1	19.6
63	Grindsted PH 175	Sodium alginate	−54.4	21.4	15.1
64	Manucol LD	Sodium alginate	−58.1	24.6	20.8
65	**Manucol DMF**	**Sodium alginate**	**−49.3**	**−0.3**	**31.5**
66	Manucol DH	Sodium alginate	−58.0	27.6	18.7
67	Kollidon Cl	Crospovidone	8.6	−14.4	53.3
68	Gelatina	Gelatin	−61.4	29.8	32.4

Binders chosen for the experiments are given in bold letters. The contribution of each PC to the expliained variance is given next to the component number.
[a]The excipient contains microcrystalline cellulose.

Appendix III. Score values for the disintegrants in the study.

ObsNum	ObsName	Type of disintegrant	*t*[1] (46%)	*t*[2] (23%)	*t*[3] (19%)
1	Avicel PH-101	Microcrystalline cellulose	52.2	−12.5	18.1
2	Avicel PH-102	Microcrystalline cellulose	52.8	−13.3	19.5
3	Avicel PH-112	Microcrystalline cellulose	61.6	−20.3	20.5
4	Avicel PH-200	Microcrystalline cellulose	54.8	−14.0	20.2
5	**Avicel PH-301**	**Microcrystalline cellulose**	**50.5**	**−16.9**	**13.5**
6	**Avicel PH-302**	**Microcrystalline cellulose**	**50.9**	**−5.5**	**4.7**
7	Lycatab	Maltodextrin	27.3	25.5	−21.2

(*continued*)

Appendix III. Continued.

ObsNum	ObsName	Type of disintegrant	*t*[1] (46%)	*t*[2] (23%)	*t*[3] (19%)
8	C Sperse MD 01314	Maltodextrin	21.5	28.0	−22.6
9	**C Pharm 01980**	**Maltodextrin**	**18.5**	**27.9**	**−17.4**
10	C Pharm 01983	Maltodextrin	21.3	33.6	−22.5
11	C Pharm DC 93000	Corn starch	12.5	29.8	−14.6
12	**Gum guar powder**	**Guar gum**	**−8.7**	**5.4**	**2.6**
13	Frimulsion BM	Guar gum	−5.7	−4.4	6.7
15	Frimulsion RG30	Guar gum	−17.4	14.0	3.3
16	**Grindsted PH060**	**Alginic acid**	**−15.6**	**−56.7**	**−42.0**
17	Kelacid	Alginic acid	−12.9	−47.4	−46.3
18	**Alginic acid H/FD**	**Alginic acid**	**−31.5**	**−45.7**	**−44.9**
19	Grindsted PH 150	Sodium alginate	−41.6	10.5	14.4
20	Grindsted PH 155	Sodium alginate	−40.6	5.4	10.1
21	Grindsted PH 157	Sodium alginate	−40.9	3.4	18.4
22	**Grindsted PH 170**	**Sodium alginate**	**−33.4**	**−8.5**	**25.3**
23	Grindsted PH 175	Sodium alginate	−39.6	11.3	12.6
24	**Manucol LD**	**Sodium alginate**	**−48.3**	**5.3**	**11.4**
25	Manucol DMF	Sodium alginate	−37.0	−22.7	50.6
26	Manucol DH	Sodium alginate	−47.7	10.8	6.6
27	**GGK 516**	**Potato starch**	**−3.0**	**57.3**	**−27.1**

Disintegrants chosen for the experiments are given in bold letters. The contribution of each PC to the explained variance is given next to the component number.

Appendix IV. The multivariate design with score values for all excipients and the fractional amounts of the formulation factors.

Exp. name	Run order	Filler 1	Filler 2	Filler 3	Disint 1	Disint 2	Disint 3	Binder 1	Binder 2	Binder 3
N1	32	−57.9	−2.6	−33.1	−31.5	−45.7	−44.9	−37.3	−6.8	−22.4
N2	15	−16.0	−34.4	−11.2	−33.4	−8.5	25.3	−49.3	−0.3	31.5
N3	23	−53.6	6.1	−21.6	−33.4	−8.5	25.3	38.6	−1.5	−4.1
N4	7	49.8	11.8	−13.2	−31.5	−45.7	−44.9	48.3	−8.4	5.1
N5	5	−4.4	−42.8	17.3	−33.4	−8.5	25.3	22.7	13.3	5.9
N6	18	7.5	−36.8	17.0	−31.5	−45.7	−44.9	32.7	4.6	−3.3
N7	11	−28.1	37.5	15.8	−31.5	−45.7	−44.9	−55.3	19.6	18.9
N8	14	7.1	21.1	32.4	−33.4	−8.5	25.3	−20.8	17.3	−21.8
N9	4	−57.9	−2.6	−33.1	−15.6	−56.7	−42.0	22.7	13.3	5.9
N10	12	−16.0	−34.4	−11.2	50.5	−16.9	13.5	32.7	4.6	−3.3
N11	33	−53.6	6.1	−21.6	50.5	−16.9	13.5	−55.3	19.6	18.9
N12	24	49.8	11.8	−13.2	−15.6	−56.7	−42.0	−20.8	17.3	−21.8
N13	9	−4.4	−42.8	17.3	50.5	−16.9	13.5	−37.3	−6.8	−22.4
N14	31	7.5	−36.8	17.0	−15.6	−56.7	−42.0	−49.3	−0.3	31.5
N15	6	−28.1	37.5	15.8	−15.6	−56.7	−42.0	38.6	−1.5	−4.1
N16	8	7.1	21.1	32.4	50.5	−16.9	13.5	48.3	−8.4	5.1
N17	25	−57.9	−2.6	−33.1	−3.0	57.3	−27.1	−20.8	17.3	−21.8
N18	35	−16.0	−34.4	−11.2	−48.3	5.3	11.4	−55.3	19.6	18.9
N19	21	−53.6	6.1	−21.6	−48.3	5.3	11.4	32.7	4.6	−3.3
N20	29	49.8	11.8	−13.2	−3.0	57.3	−27.1	22.7	13.3	5.9
N21	30	−4.4	−42.8	17.3	−48.3	5.3	11.4	48.3	−8.4	5.1
N22	26	7.5	−36.8	17.0	−3.0	57.3	−27.1	38.6	−1.5	−4.1
N23	27	−28.1	37.5	15.8	−3.0	57.3	−27.1	−49.3	−0.3	31.5
N24	2	7.1	21.1	32.4	−48.3	5.3	11.4	−37.3	−6.8	−22.4

Appendix IV. Continued.

Exp. name	Run order	Filler 1	Filler 2	Filler 3	Disint 1	Disint 2	Disint 3	Binder 1	Binder 2	Binder 3
N25	28	−57.9	−2.6	−33.1	18.5	27.9	−17.4	48.3	−8.4	5.1
N26	3	−16.0	−34.4	−11.2	50.9	−5.5	4.7	38.6	−1.5	−4.1
N27	20	−53.6	6.1	−21.6	50.9	−5.5	4.7	−49.3	−0.3	31.5
N28	1	49.8	11.8	−13.2	18.5	27.9	−17.4	−37.3	−6.8	−22.4
N29	10	−4.4	−42.8	17.3	50.9	−5.5	4.7	−20.8	17.3	−21.8
N30	19	7.5	−36.8	17.0	18.5	27.9	−17.4	−55.3	19.6	18.9
N31	34	−28.1	37.5	15.8	18.5	27.9	−17.4	32.7	4.6	−3.3
N32	17	7.1	21.1	32.4	50.9	−5.5	4.7	22.7	13.3	5.9
N33	13	−10.9	−32.8	−5.8	−8.7	5.4	2.6	−18.0	−0.1	−10.4
N34	16	−10.9	−32.8	−5.8	−8.7	5.4	2.6	−18.0	−0.1	−10.4
N35	22	−10.9	−32.8	−5.8	−8.7	5.4	2.6	−18.0	−0.1	−10.4

Exp name	Buffer type (coded)	BFMF	API batch	Amount buffer	Amount bind and fill	Amount disint.	Amount of MgSt API	Comp. force (kN)*
NI	None (0)	−1	S2 B1	0	0.843	0.010	0.147	3.0
N2	NaGly (+1)	−1	S1 B3	0.106	0.697	0.050	0.147	3.0
N3	NaCarb (−1)	1	S1 B3	0.058	0.785	0.010	0.147	9.0
N4	None (0)	1	S2 B1	0	0.803	0.050	0.147	8.9
N5	None (0)	−1	S2 B1	0	0.803	0.050	0.147	9.1
N6	NaGly (+1)	−1	S1 B3	0.106	0.737	0.010	0.147	8.9
N7	NaCarb (−1)	1	S1 B3	0.058	0.745	0.050	0.147	3.1
N8	None (0)	1	S2 B1	0	0.843	0.010	0.147	2.7
N9	None (0)	1	S1 B3	0	0.843	0.010	0.147	3.3
N10	NaCarb (−1)	1	S2 B1	0.058	0.745	0.050	0.147	3.0
N11	NaGly (+1)	−1	S2 B1	0.106	0.737	0.010	0.147	8.9
N12	None (0)	−1	S1 B3	0	0.803	0.050	0.147	8.6
N13	None (0)	1	S1 B3	0	0.803	0.050	0.147	9.0
N14	NaCarb (−1)	1	S2 B1	0.058	0.785	0.010	0.147	8.6
N15	NaGly (+1)	−1	S2 B1	0.106	0.697	0.050	0.147	3.2
N16	None (0)	−1	S1 B3	0	0.843	0.010	0.147	3.0
N17	NaGly (+1)	1	S2 B1	0.106	0.697	0.050	0.147	9.0
N18	None (0)	1	S1 B3	0	0.843	0.010	0.147	9.1
N19	None (0)	−1	S1 B3	0	0.803	0.050	0.147	3.2
N20	NaCarb (−1)	−1	S2 B1	0.058	0.785	0.010	0.147	3.3
N21	NaGly (+1)	1	S2 B1	0.106	0.737	0.010	0.147	3.3
N22	None (0)	1	S1 B3	0	0.803	0.050	0.147	2.9
N23	None (0)	−1	S1 B3	0	0.843	0.010	0.147	8.8
N24	NaCarb (−1)	−1	S2 B1	0.058	0.745	0.050	0.147	8.5
N25	NaCarb (−1)	−1	S1 B3	0.058	0.745	0.050	0.147	9.2
N26	None (0)	−1	S2 B1	0	0.843	0.010	0.147	9.6
N27	None (0)	1	S2 B1	0	0.803	0.050	0.147	3.0
N28	NaGly (+1)	1	S1 B3	0.106	0.737	0.010	0.147	3.3
N29	NaCarb (−1)	−1	S1 B3	0.058	0.785	0.010	0.147	3.1
N30	None (0)	−1	S2 B1	0	0.803	0.050	0.147	3.3
N31	None (0)	1	S2 B1	0	0.843	0.010	0.147	8.7
N32	NaGly (+1)	1	S1 B3	0.106	0.697	0.050	0.147	8.5
N33	NaCarb (−1)	0	S2 B2	0.029	0.794	0.030	0.147	6.2
N34	NaCarb (−1)	0	S2 B2	0.029	0.794	0.030	0.147	6.2
N35	NaCarb (−1)	0	S2 B2	0.029	0.794	0.030	0.147	5.9

*The effect of the compression force was evaluated by using a constructed variable because it was not included as a factor in the design.

Appendix V. The results of the screening study have headers in bold letters.

Exp. name	**Ejection force** (N)	**Disintegration time** (min)	**Disintegration type** (–)	**Crushing strength** (kp)	**Adhesion** (–)	**Hausner ratio** (–)	Crushing strength (kp)*	Ejection force (N)*	Tablet height (mm)*
N1	48	4.8	1	13.7	0	1.52	7.4	54	2.69
N2	45	51.7	1.5	5.4	2	1.36	3.1	28	2.6
N3	30	>120	2	10.4	0	1.46	11.9	21	2.56
N4	33	37.4	1	10.5	0	1.44	12	33	2.57
N5	42	34.9	1.5	4.5	0	1.29	6.3	54	2.56
N6	44	14.2	1	5.8	2	1.21	7.2	79	2.47
N7	47	106.8	2	2.3	0	1.41	0	43	2.6
N8	30	16.3	1	11.9	0	1.44	5	29	2.55
N9	11	>12.0	0.5	4.2	0	1.31	2.7	9	3.07
N10	46	>120	2	10.1	1	1.45	6	41	2.69
N11	34	71.4	1.5	10.0	0	1.32	11.2	38	2.29
N12	83	7.1	1	7.3	0	1.46	8.9	98	2.26
N13	48	4.9	1	5.4	0	1.56	8.1	54	2.43
N14	63	77.0	1.5	1.8	0	1.51	3.3	74	2.32
N15	21	93.4	1.5	5.4	1	1.45	3	19	2.7
N16	59	12.8	1	10.9	0	1.38	5	58	2.66
N17	29	6.7	1.5	12.1	1	1.3	13.2	28	2.39
N18	48	>120	2	2.0	1	1.49	4	49	2.29
N19	26	>120	1.5	14.4	0	1.47	10.2	32	2.63
N20	39	29.9	1	4.7	1	1.28	3.4	31	2.73
N21	19	28.3	1	9.4	0	1.42	6.9	20	2.74
N22	24	>120	2	6.8	0	1.46	4	30	2.83
N23	28	79.9	1.5	3.1	0	1.54	4.3	22	2.32
N24	51	20.1	1	7.4	0	1.49	8.4	67	2.32
N25	31	10.6	1	12.5	0	1.43	14.5	26	2.42
N26	68	>120	2	6.5	2	1.46	9.2	75	2.4
N27	39	111.9	2	5.0	0	1.51	1.7	35	2.64
N28	46	22.6	1	4.9	2	1.4	1	34	2.94
N29	74	3.7	1	9.2	0	1.37	5	45	2.55
N30	69	97.5	1.5	4.7	0	1.43	2	43	2.59
N31	29	>120	2	9.1	0	1.54	11.7	29	2.61
N32	14	25.1	1.5	2.7	3	1.22	3.6	19	2.86
N33	69	2.3	1	0.9	1	1.51	0.9	69	2.69
N34	66	1.1	0	0.9	1	1.52	0.9	66	2.66
N35	60	1.5	1	0.0	1	1.56	0	60	2.71

*Results for the tablet formulations that were also compressed at 3 or 9 kN.

REFERENCES

1. Gabrielsson, J.; Lindberg, N.-O.; Lundstedt, T. Multivariate methods in pharmaceutical applications. J. Chemometrics **2002**, *16* (3), 141–160.
2. Gabrielsson, J.; Nyström, Å.; Lundstedt, T. Multivariate methods in developing an evolutionary strategy for tablet formulation. Drug Dev. Ind. Pharm. **2000**, *26* (3), 275–296.
3. Gustafsson, C.; Nyström, C.; Lennholm, H.; Bonferoni, M.C.; Caramella, C.M. Characteristics of hydroxypropyl methylcellulose influencing compactibility and prediction of particle and tablet properties by infrared spectroscopy. J. Pharm. Sci. **2003**, *92* (3), 460–470.
4. Pasikatan, M.C.; Steele, J.L.; Spillman, C.K.; Haque, E. Near infrared reflectance spectroscopy for online particle size analysis of powders and ground materials. J. Near Infrared Spectrosc. **2001**, *9* (3), 153–164.
5. Thelin, B.; Lundstedt, T.; Lundgren, R.; Olsson, A.; Batra, S. Classification of Estradurin® batches: correlation between 31P NMR and a biological duration test for

batch approval. Chemom. Intell. Lab. Syst. **1995**, *27*, 135–145.
6. Box, G.E.P.; Hunter, W.G.; Hunter, J.S. *Statistics for Experimenters—An Introduction to Design*, *Data Analysis*, *and Model Building*; John Wiley & Sons, Inc.: New York, 1978.
7. Martens, H.; Naes, T. *Multivariate Calibration*; John Wiley & Sons, Inc.: New York, 1989.
8. Jackson, J.E. *A User's Guide to Principal Components*; John Wiley & Sons, Inc.: New York, 1991.
9. Eriksson, L.; Johansson, E.; Kettaneh, N.; Wikström, C.; Wold, S. *Design of Experiments—Principles and Applications*, Umeå, 2000; ISBN 91-973730-0-1, available from Umetrics, www.umetrics.com.
10. Eriksson, L.; Johansson, E.; Kettaneh-Wold, N.; Wold, S. *Multi- and Megavariate Data Analysis—Principles and Applications*, Umeå, 2001; ISBN 91-973730-1-X, available from Umetrics, www.umetrics.com.
11. Lundstedt, T.; Seifert, E.; Abramo, L.; Thelin, B.; Nyström, Å.; Pettersen, J.; Bergman, R. Experimental design and optimization. Chem. Intell. Lab. Sys. **1998**, *42*, 3–40.
12. Eriksson, L.; Johansson, E.; Wikström, C. Mixture design—design generation, PLS analysis, and model usage. Chem. Intell. Lab. Sys. **1998**, *43*, 1–24.
13. Wold, S.; Sjöström, M.; Carlson, R.; Lundstedt, T.; Hellberg, S.; Skagerberg, B.; Wikström, C.; Öhman, J. Multivariate design. Anal. Chim. Acta **1986**, *191*, 17–32.
14. Wold, S.; Esbensen, K.; Geladi, P. Principal component analysis. Chemom. Intell. Lab. Systems **1987**, *2*, 37–52.
15. Wold, S.; Ruhe, A.; Wold, H.; Dunn, W.I. The collinearity problem in linear regression. The partial least squares approach to generalized inverses. SIAM J. Sci. Statist. Comput. **1984**, *5*, 735–743.
16. Barnes, R.J.; Dhanoa, M.S.; Lister, S.J. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. Appl. Spectrosc. **1989**, *43* (5), 772–777.