Drug Development and Industrial Pharmacy, 26(3), 275–296 (2000)

RESEARCH PAPER

Multivariate Methods in Developing an Evolutionary Strategy for Tablet Formulation

Jon Gabrielsson,*,† Åsa Nyström,‡ and Torbjörn Lundstedt*,§

Pharmacia and Upjohn, Structure-Property Optimization Center, Uppsala, Sweden

ABSTRACT

The aim of this study was to develop a new strategy for choosing excipients in tablet formulation. Multivariate techniques such as principal component analysis (PCA) and experimental design were combined in a multivariate design for screening experiments. Of a total 87 investigated excipients, the initial screening experiments contained 5 lubricants, 9 binders, and 5 disintegrants, and 35 experiments were carried out. Considering a reduced factorial design was used, the resulting PCA and partial least squares (PLS) models offered good insight into the possibilities of tablet formulation. It also offered solutions to the problems and clearly gave directions for optimum formulations. Further, it offered several alternatives for achieving quality formulations. Additional experiments conducted to validate and verify the usefulness of the model were successful, resulting in several tablets of good quality. The conclusion is that a multivariate strategy in tablet formulation is efficient and can be used to reduce the number of experiments drastically. Combining multivariate characterization, physicochemical properties, experimental design, multivariate design, and PLS would lead to an evolutionary strategy for tablet formulation. Since it includes a learning strategy that continuously incorporates data for new compounds and from conducted experiments, this would be an even more powerful tool than expert systems.

* To whom correspondence should be addressed.
† Present address Research Group for Chemometrics, Umeå University, SE-901 87 Umeå, Sweden.
‡ Present address Umetri AB, Box 7960, SE-907 19, Umeå, Sweden.
§ Present address Melacure Therapeutics AB, Uppsala Science Park, SE-75183 Uppsala, Sweden.

A MULTIVARIATE STRATEGY IN TABLET FORMULATION

The selection of excipients in tablet formulation can be a troublesome and time-consuming process. The number of excipients from which to choose is continuously growing, in pace with development of new forms of drug delivery. Today, the galenic scientist must rely on his or her experimental experience or findings documented by others when choosing suitable excipients that will bring the tablet the desired qualities.

There is a need for a more general strategy in tablet formulation, a strategy that can incorporate new findings with existing knowledge and proven formulations. We believe that combining multivariate characterization, experimental design, multivariate design, and partial least squares (PLS) would enable the development of such a strategy.

There is a history of use of multivariate tools in preformulation work, combinatorial libraries, and synthesis and drug development (1–6). Multivariate methods are also widely used for problem-oriented investigations (7). Although not to the same extent, others have also applied multivariate tools in pharmaceutical work (8).

The first step of a multivariate strategy is characterization of the excipients. By applying multivariate data analysis to the descriptive data, important information is extracted in a few variables. These latent variables can then be used as continuous variables in the ensuing experimental design.

Experimental design is then applied to investigate the experimental domain in a limited number of experiments. In experimental design, all parameters are changed simultaneously in a controlled manner. Evaluation of the experimental results enables estimations of the effects of excipient properties on different aspects of tablet quality. The resulting model will also be open for continuous update as new excipients surface.

Predictions and the validity of the model are then confirmed by additional experiments.

STRATEGY

The new strategy consists of three steps:

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

in that order. Multivariate data analysis has an important role in all steps.

Step 1: Multivariate Characterization

The purpose of characterizing the excipients in multiple variables in step 1 is to get as complete a picture as possible of the qualities of the excipients. Then, with a principal component analysis (PCA) of all the descriptive variables, the principal properties can be used as continuous variables in the experimental design. This way, the selections of excipients for the test runs in the design are no longer ''either-or'' options, but choices along continuous scales. Discrete variables are best avoided since they bring complications to the design in the form of an increased number of necessary experiments and difficulties in interpreting the results.

Step 2: Multivariate Experimental Design

It is crucial for a successful design that the excipients chosen for the test runs from the PCA are not too similar. The larger the distances between the compounds along the principal components, the bigger the differences in behavior can be expected. Therefore, excipients from every class were chosen from both ends of every principal property scale to enclose all possible diversions, thereby spanning the widest possible experimental region. It is important to emphasize the need for testing also the mixes of excipients that are presumed to give poor results. Thereby, the direction toward improved quality in the experimental region is pinned down.

Step 3: Evaluation, Validation, and Optimization

The evaluation of the results from the design and the optimization again include multivariate data analysis. With PCA and PLS projections to latent structures, the mass of information is handled efficiently and can be used for several aims. Primarily, the outcome of the tested formulations is thoroughly viewed. Predictions may be made regarding how to obtain or maintain a certain quality and how to find alternative routes to that goal. Specific demands on tablet quality aimed to fit new active substances can be added to the design and be used to predict tailor-made formulations with the aid of soft independent modeling of class analogy (SIMCA).

EXPERIMENTAL

All binders and disintegrants were analyzed with infrared (IR) spectroscopy (450–4400 cm^{-1}) using a Perkin Elmer 16PC FTIR (Fourier transform infrared) instru-

ment. All samples were prepared from 5 wt% analyte and the rest potassium bromide. Spectra were split into 1976 variables and corrected with multiplicative scatter correction (MSC) in Preefix 1.0 (Umetri AB, not commercially available) before the multivariate data analysis.

All excipients, with the exception of lactose and carbomer, which were deemed impossible to sift, were passed through a sieve, and no particles larger than 200 μm were used in the experiments. All samples were weighed on a Mettler AM100 scale and mixed with a Turbula mixer for at least 10 min.

The tablets were compressed using a Korsch EKO tableting machine equipped with a computerized system for collection of compressing data. All samples were compressed at the same pressure, 150 ± 5 MPa. The punch distance was kept constant, 3 mm, and the pressure was achieved through varying the sample weights. The relationship between sample weight and pressure was assumed to be linear in that region. For every batch, 18 tablets were produced.

Depending on the efficiency of the lubricant, the recorded pressure varied due to differences in friction between the punch and material stuck on the die wall. The standard deviation for the pressure (pressure STD) was calculated and used as a measure of the efficiency of the lubricant.

Of the tablets, 6 were evaluated for thickness and hardness to give their radial tensile strength. A Mahr thickness tester and a Pharma Test PTB 311 were used.

A standardized test for friability was performed on a Pharma Test PTFR-A for 6 of the tablets.

The remaining 6 tablets were tested for disintegration time in deionized water, 37°C, using a Pharma Test PTZ-E with integrated computer software. The maximum disintegration time was set at 75 min.

Alginic acid, aluminum oxide, hydroxyethylcellulose, and stearic acid calcium salt were from Sigma-Aldrich; lauric acid sodium salt was from Chemicon; aluminum hydroxide and stearic acid zinc salt were from Kebolab; and all other compounds were kindly provided by Pharmacia and Upjohn in Helsingborg and Uppsala, Sweden.

The experimental design (Table 1) was generated in Modde 4.0 (Umetri AB). All multivariate data analysis (Table 2) was performed in SIMCA-P 3.01 (Umetri AB).

METHODS AND THEORETICAL BACKGROUND

Principal Components Analysis

The development of modern analytical equipment provides the researcher with large quantities of data. Extracting useful information from the collected data becomes a new challenge for the researcher. Just looking at data tables in most cases is not enough. The application of PCA to large quantities of data provides a means for excellent overviews and the detection of trends, groupings, and outliers (9).

The objective of PCA is to describe the variation in data with a minimum of variables. Variables are often dependent on each other, and PCA shows the underlying structure in data. A set of data consists of N objects described by K variables. The principles of PCA are illustrated in two dimensions in Fig. 1. Equation 1 describes the mathematical expression of a PCA model.

$$x_{ik} = a_i + \mathrm{ws}_k \sum_{j=1}^{A} b_{ij} t_{jk} + e_{ik} \quad (1)$$

Here, x_{ik} denotes the scaled value of object i in variable k. Sometimes, it is necessary to transform data in a certain variable, prior to the investigation, to homogenize the spread of objects. The analysis corresponds to a least-squares fitting of a straight line ($A = 1$) or an A-dimensional hyperplane to the data in the k-dimensional variable space. The parameters a_i determine the center of the data set; ws_k is the scaling weight of variable k (usually unit variance); b_{ij} are the direction coefficients (one for each variable and component) of the line/hyperplane. For each object k, the parameter t_{jk} describes the position of the object projected down onto the model line/hyperplane. Hence, the t values (often referred to as the principal properties [PPs] or scores) can be used to relate objects to one another. Their distance from the model is expressed by the residual e_{ik} and is kept to a minimum using the least-squares technique.

How the original variables k are weighted together in a principal component (PC) is expressed by the loadings p_k. Depending on the scaling of the variable, the loading is given by the coefficient b_{ij}. 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 information to this PC. Together, the scores and loadings describe the principal components of the data set.

New components are fitted until enough variance in the data has been described. For this purpose, R^2 (Eq. 2) and Q^2 (Eq. 3) are used. SS_{tot} is the total sum of squares, and SS_{res} is the residual sum of squares. R^2 is a measure of how much of the variance in X can be explained by the model.

$$R^2 = \frac{SS_{\mathrm{tot}} - SS_{\mathrm{res}}}{SS_{\mathrm{tot}}} \quad (2)$$

Table 1

Experimental Design in Principal Properties with the Amounts of the Excipients Included in a Mixture Design

Exp. No.	Lub 1	Lub 2	Bind 1	Bind 2	Bind 3	Dis 1	Dis 2	MLub (%)	MBind (%)	MDis (%)	Filler (%)
Screening experiments											
1	−1.68	−1.39	−13.63	−18.09	−2.65	10.95	34.25	0.020	0.15	0.15	0.680
2	2.34	−0.64	−13.63	−18.09	−2.65	5.55	−40.41	0.002	0.05	0.15	0.798
3	−4.12	0.86	−13.63	−18.09	−2.65	5.55	−40.41	0.002	0.15	0.05	0.798
4	2.44	1.19	−13.63	−18.09	−2.65	10.95	34.25	0.020	0.05	0.05	0.880
5	−1.68	−1.39	14.38	−0.24	−17.08	5.55	−40.41	0.020	0.05	0.05	0.880
6	2.34	−0.64	14.38	−0.24	−17.08	10.95	34.25	0.002	0.15	0.05	0.798
7	−4.12	0.86	14.38	−0.24	−17.08	10.95	34.25	0.002	0.05	0.15	0.798
8	2.44	1.19	14.38	−0.24	−17.08	5.55	−40.41	0.020	0.15	0.15	0.680
9	−1.68	−1.39	−30.04	25.18	−28.11	−72.91	1.19	0.002	0.05	0.05	0.898
10	2.34	−0.64	−30.04	25.18	−28.11	16.72	−13.73	0.020	0.15	0.05	0.780
11	−4.12	0.86	−30.04	25.18	−28.11	16.72	−13.73	0.020	0.05	0.15	0.780
12	2.44	1.19	−30.04	25.18	−28.11	−72.91	1.19	0.002	0.15	0.15	0.698
13	−1.68	−1.39	32.79	22.04	−13.70	16.72	−13.73	0.002	0.15	0.15	0.698
14	2.34	−0.64	32.79	22.04	−13.70	−72.91	1.19	0.020	0.05	0.15	0.780
15	−4.12	0.86	32.79	22.04	−13.70	−72.91	1.19	0.020	0.15	0.05	0.780
16	2.44	1.19	32.79	22.04	−13.70	16.72	−13.73	0.002	0.05	0.05	0.898
17	−1.68	−1.39	−13.83	−31.43	26.22	16.72	−13.73	0.002	0.05	0.05	0.898
18	2.34	−0.64	−13.83	−31.43	26.22	−72.91	1.19	0.020	0.15	0.05	0.780
19	−4.12	0.86	−13.83	−31.43	26.22	−72.91	1.19	0.020	0.05	0.15	0.780
20	2.44	1.19	−13.83	−31.43	26.22	16.72	−13.73	0.002	0.15	0.15	0.698
21	−1.68	−1.39	21.03	−8.21	11.25	−72.91	1.19	0.002	0.15	0.15	0.698
22	2.34	−0.64	21.03	−8.21	11.25	16.72	−13.73	0.020	0.05	0.15	0.780
23	−4.12	0.86	21.03	−8.21	11.25	16.72	−13.73	0.020	0.15	0.05	0.780
24	2.44	1.19	21.03	−8.21	11.25	−72.91	1.19	0.002	0.05	0.05	0.898
25	−1.68	−1.39	−56.20	35.81	19.74	5.55	−40.41	0.020	0.15	0.15	0.680
26	2.34	−0.64	−56.20	35.81	19.74	10.95	34.25	0.002	0.05	0.15	0.798
27	−4.12	0.86	−56.20	35.81	19.74	10.95	34.25	0.002	0.15	0.05	0.798
28	2.44	1.19	−56.20	35.81	19.74	5.55	−40.41	0.020	0.05	0.05	0.880
29	−1.68	−1.39	68.15	59.06	28.81	10.95	34.25	0.020	0.05	0.05	0.880
30	2.34	−0.64	68.15	59.06	28.81	5.55	−40.41	0.002	0.15	0.05	0.798
31	−4.12	0.86	68.15	59.06	28.81	5.55	−40.41	0.002	0.05	0.15	0.798
32	2.44	1.19	68.15	59.06	28.81	10.95	34.25	0.020	0.15	0.15	0.680
33	−0.32	−0.08	−20.87	−14.66	9.08	3.96	−10.71	0.011	0.10	0.10	0.789
34	−0.32	−0.08	−20.87	−14.66	9.08	3.96	−10.71	0.011	0.10	0.10	0.789
35	−0.32	−0.08	−20.87	−14.66	9.08	3.96	−10.71	0.011	0.10	0.10	0.789
Optimization screening											
1	2.24	1.18	−13.83	−31.43	26.22	5.55	−40.41	0.020	0.15	0.05	0.780
2	2.08	1.14	−13.83	−31.43	26.22	5.55	−40.41	0.002	0.15	0.15	0.698
3	2.08	1.14	5.81	−1.61	−13.79	3.96	−10.71	0.020	0.05	0.15	0.780
4	2.08	1.14	5.81	−1.61	−13.79	3.96	−10.71	0.002	0.05	0.05	0.898
5	2.08	1.14	−56.20	35.81	19.74	5.55	−40.41	0.002	0.05	0.15	0.798
6	2.08	1.14	−56.20	35.81	19.74	5.55	−40.41	0.020	0.05	0.05	0.880
7	2.24	1.18	32.79	22.04	−13.70	5.55	−40.41	0.002	0.15	0.05	0.798
8	2.08	1.14	32.79	22.04	−13.70	5.55	−40.41	0.020	0.15	0.15	0.680
9	−0.32	−0.08	−20.87	−14.66	9.08	3.96	−10.71	0.011	0.10	0.10	0.789
Pure optimization											
OP	2.08	1.14	32.79	22.04	−13.70	5.55	−40.41	0.001	0.24	0.30	0.460
OP1	2.44	1.19	32.79	22.04	−13.70	5.55	−40.41	0.001	0.12	0.17	0.714
OP4	2.08	1.14	−56.20	35.81	19.74	5.55	−40.41	0.001	0.12	0.17	0.714

Lub = lubricant; Bind = binder; Dis = disintegrant; MLub = amount of lubricant; Mbind = amount of binder; MDis = amount of disintegrants.

Table 2

Principal Properties Translated to Excipients and the Recorded Responses (with Lactose Filler)

Exp. No.	Exp. Name	Run Order	Lubricant	Binder	Disintegrant	Recorded Responses			
						Pressure STD	Tensile Strength (Nm^{-2})	Friability (%)	Disintegration Time (sec)
Screening experiments									
1	N1	12	AlOH	MeC	Gel	3.28	800243	1.16	76
2	N2	27	NaLaur	MeC	MicC	2.77	1540896	0.53	63
3	N3	26	AlO	MeC	MicC	3.68	1118980	0.83	105
4	N4	35	MgSt	MeC	Gel	0.53	969953	1.10	28
5	N5	23	AlOH	AlgNa	MicC	2.87	1243254	0.92	1202
6	N6	18	NaLaur	AlgNa	Gel	2.00	866312	1.26	2234
7	N7	6	AlO	AlgNa	Gel	3.04	607125	3.11	2004
8	N8	32	MgSt	AlgNa	MicC	0.73	1265818	0.60	3192
9	N9	31	AlOH	PMAc	PVPP	3.32	896443	1.22	30
10	N10	17	NaLaur	PMAc	CStar	0.66	872412	1.29	139
11	N11	16	AlO	PMAc	CStar	3.11	1195291	1.51	44
12	N12	5	MgSt	PMAc	PVPP	0.73	969956	0.79	18
13	N13	34	AlOH	Alg	CStar	3.05	1165934	1.02	109
14	N14	22	NaLaur	Alg	PVPP	0.64	721534	1.38	174
15	N15	25	AlO	Alg	PVPP	2.85	831119	1.38	46
16	N16	19	MgSt	Alg	CStar	0.57	1088176	1.05	63
17	N17	30	AlOH	Sucr	CStar	3.32	1134428	1.52	61
18	N18	28	NaLaur	Sucr	PVPP	0.58	833053	1.32	1.22
19	N19	8	AlO	Sucr	PVPP	1.59	721492	2.44	25
20	N20	9	MgSt	Sucr	CStar	0.62	1150234	1.15	80
21	N21	2	AlOH	Aca	PVPP	1.79	984543	1.16	455
22	N22	24	NaLaur	Aca	CStar	0.64	1072622	1.41	314
23	N23	14	AlO	Aca	CStar	2.52	1200443	1.28	735
24	N24	1	MgSt	Aca	PVPP	0.78	1088986	0.90	185
25	N25	21	AlOH	PEG3	MicC	1.16	1581754	0.48	376
26	N26	20	NaLaur	PEG3	Gel	0.64	655299	1.56	195
27	N27	7	AlOt	PEG3	Gel	0.94	1004529	1.08	489
28	N28	11	MgSt	PEG3	MicC	0.74	1293397	0.85	207
29	N29	10	AlOH	Carb	Gel	2.48	1523841	0.82	4501
30	N30	15	NaLaur	Carb	MicC	2.99	2823436	0.36	4501
31	N31	3	AlO	Carb	MicC	2.11	2135910	0.73	2446
32	N32	33	MgSt	Carb	Gel	0.64	2054662	0.46	4501
33	Crt1	13	CaCi	HPMC	HEC	3.09	935770	1.20	370
34	Crt2	4	CaCi	HPMC	HEC	2.32	862455	1.23	580
35	Crt3	29	CaCi	HPMC	HEC	2.29	980508	1.22	416
Optimization screening									
1	S1	9	CaSt	Sucr	MicC	0.76	970178	1.28	142
2	S2	3	ZnSt	Sucr	MicC	0.58	1474677	0.65	85
2	S3	7	ZnSt	Agar	HEC	0.90	517596	2.18	30
3	S4	5	ZnSt	Agar	HEC	0.51	808912	1.23	30
5	S5	6	ZnSt	PEG3	MicC	0.43	1433026	0.60	84
6	S6	4	ZnSt	PEG3	MicC	0.44	1189307	0.86	127
7	S7	8	CaSt	Alg	MicC	0.71	1003611	0.92	69
8	S8	2	ZnSt	Alg	MicC	0.98	1208109	0.64	56
9	CrtS	1	CaCi	HPMC	HEC	3.18	979260	1.02	241
Pure optimization									
	OP		ZnSt	Alg	MicC	0.72	1593600	0.58	43
	OP1		MgSt	Alg	MicC	0.86	1638208	0.43	46
	OP4		ZnSt	PEG3	MicC	0.59	1693232	0.43	310

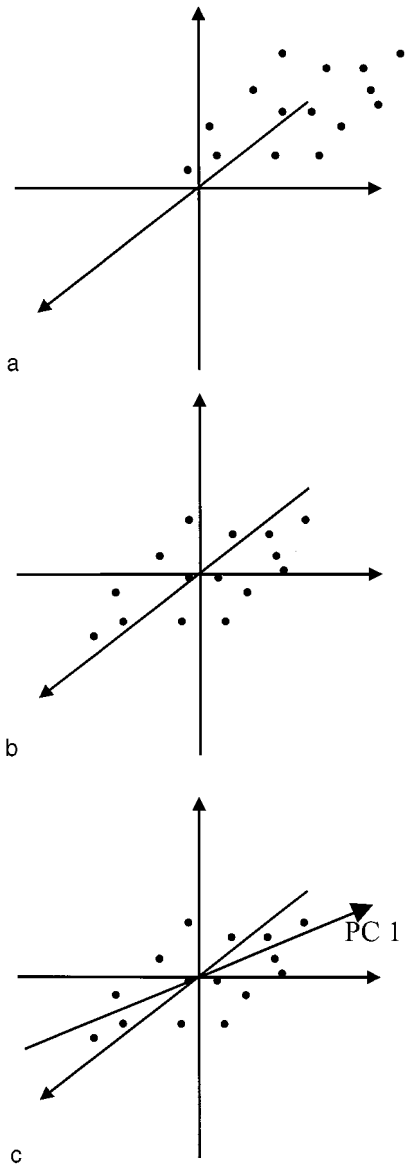


Figure 1. The objects can be regarded as a (a) swarm of points in the *K*-dimensional space. The first step of PCA is usually (b) to center the objects around the origin. The next step is (c) to draw a line, the first principal component, through the origin that describes as much as possible of the variation contained in the objects.

$$Q^2 = \frac{SS_{tot} - \text{PRESS}}{SS_{tot}} \quad (3)$$

The predicted residual sum of squares (PRESS) is calculated in PCA, and Q^2 explains how much of the variance in X can be predicted by the model.

Two principal components, t_1 and t_2, together span a mathematical plane, often referred to as a score plot. Objects are projected onto the plane to form a two-dimensional model of the data. One could say the t_1/t_2 score plot constitutes a window through which data can be viewed. Groupings and trends in data are easier discovered this way, and outliers can be detected.

Deviating objects in the collection of data can cause problems. Detecting and diagnosing abnormal objects, referred to as outliers, is of utmost importance. Whether it stems from a simple mistype or altered conditions in the reaction vessel, PCA offers the means for proper investigation. An outlier can be an object that does not fit very well into the model. The distance to the model is greater than what is accepted, and this is discovered by examining the residuals of that particular object. If an object lies far away from other objects in the score plot, it can also be suspected of being an outlier. Studying the factors that contribute to the scores of that object helps to decide in these cases. Outliers may cause problems when building a model of the data using PCA. Since the least-squares technique is used, an outlier might cause one of the principal components to run through or very close to it, resulting in a skew model.

Partial Least Squares

PLS takes PCA one step further as it deals with both *X* and *Y* data. *X* variables could be variables like reaction temperature and pH, and *Y* data responses in the form of yield. The underlying structures in *X* and *Y* data sets are related to each other (Fig. 2). Each object is represented in both *X* space and *Y* space. The first steps are the same for PLS and PCA. The first PLS component is a line in *X* space and *Y* space that passes through the swarms of points and the origin. It is calculated to approximate well the swarms of points in *X* and *Y* and to provide the best possible correlation between the projections. The projections of *X* and *Y* are connected and correlated through the inner relation (Eq. 4). The emphasis is both on the correlation between *X* and *Y* and on a good description of *X* and *Y*. This means that the projections of *X* and *Y* differ from the ones from PCA. The descriptive variables are also weighted together differently in the principal components.

The principal properties of the responses u_n are weighted together in the principal components as expressed by the loadings c_m. The loadings w_k describe the correlation between *T* and *U*, and thereby indirectly the descriptive variables and the responses. When examining the loadings in PLS, the *w*c* loading

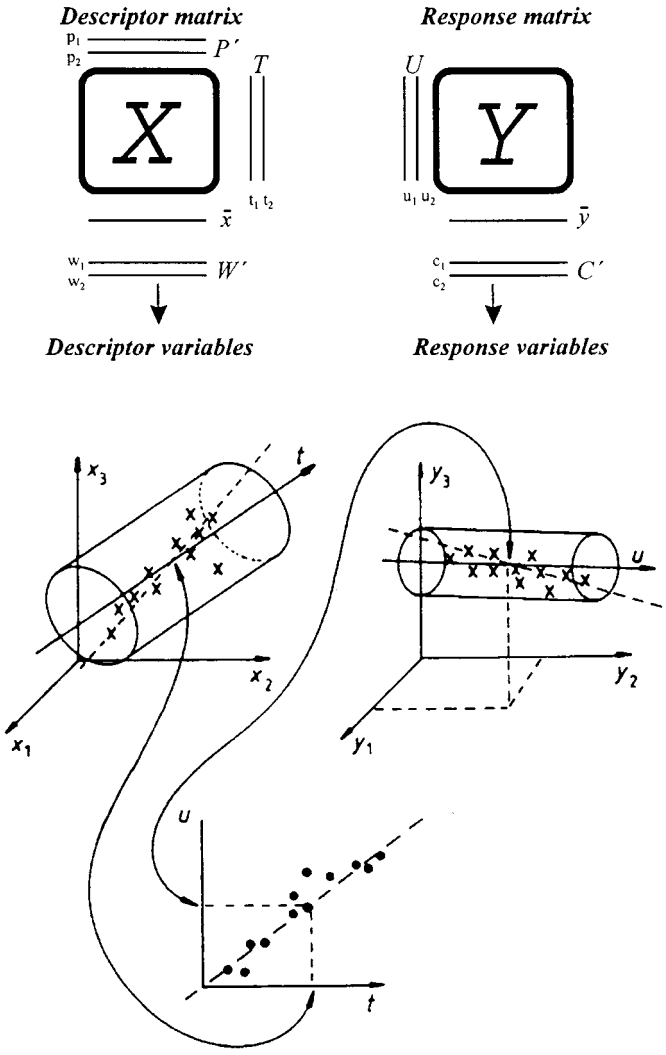


Figure 2. A PCA is performed for both the descriptive *X* and the response *Y* variables, represented by the dotted line. The best possible correlation between *X* and *Y* is calculated using the least-squares technique. Hence, the resulting PLS model is not always the best description of *X* and *Y*, but rather of the relation between them.

$$U = T + H \text{ (inner relation)} \quad (4)$$

$$X = 1 \cdot \bar{x} + T \cdot P' + E \quad (5)$$

$$Y = 1 \cdot \bar{y} + U \cdot C' + F \quad (6)$$

$$Y = 1 \cdot \bar{y} + T \cdot C' + G \quad (7)$$

plots are usually used. The *w*c* loading plots show the influence of the descriptive variables on the responses in the principal components. The correlation between *X* and *Y* is shown in the t_n/u_n score plots.

Using PLS, it is possible to make predictions in *Y* from *X* data. The *X* matrix is described by Eq. 5 and the *Y* matrix by Eq. 6.

The inner relation then gives Eq. 7.

Experimental Design

The objective of experimental design is to plan and conduct experiments to extract the maximum amount of information from the collected data in the fewest number

of experimental runs (10). The basic idea is to change all relevant factors simultaneously over a set of planned experiments and then connect and interpret the results using mathematical models.

Whether you want to optimize a reaction or perform a multivariate study on excipients, the variables that influence your responses are at the center of attention. The size and sign of the scaled and centered regression coefficients indicate the influence of the variables on the responses.

The variables X_n are given maximum and minimum values based on preexisting knowledge of the topic. All variables are changed simultaneously to cover the entire area of interest with as few experiments as possible. Another reason for this is that it makes it possible to examine interaction effects. A center point, the middle value for each variable interval, is also included to detect curvature in the experimental region.

In a full factorial experimental design (Fig. 3), all combinations of the extreme values are included as experiments. If there are k variables, the number of experiments is given by the expression 2^k. For statistic validation, one experiment, usually the center point, is repeated at least three times. The experiments are performed in random order to eliminate systematical errors.

The number of experiments in an experimental design grows rapidly with an increasing number of variables. When dealing with many variables ($k > 5$), for the purpose of screening, a full factorial design is not a realistic option. A reduced design, called a fractional factorial design, is a more appropriate choice. For a fractional factorial design, experiments are chosen to give the maximum amount of variation with fewer experimental runs. The drawback is lost information caused by confounding of main effects with interaction effects if the resolution is lower than 4.

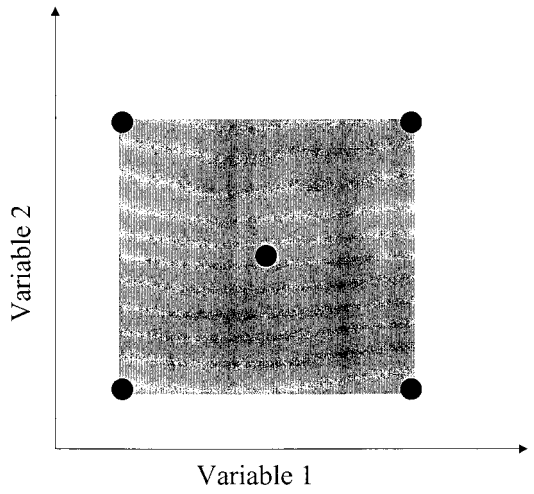


Figure 3. In experimental design, the experimental domain (shadowed) is covered in the fewest possible experiments. A center point is included to detect curvature.

Multivariate Design

Multivariate design (or experimental design in principal properties as it is often called) is a combination of experimental design and PCA (11). Instead of normal variables, principal properties are used in the design. This makes it possible to reduce the design considerably and still obtain relevant information (e.g., which variables influence tablet quality).

The IR spectra of a number of binders are described with two principal components (Fig. 4). The number of descriptive variables in the design can be reduced drastically without loss of information.

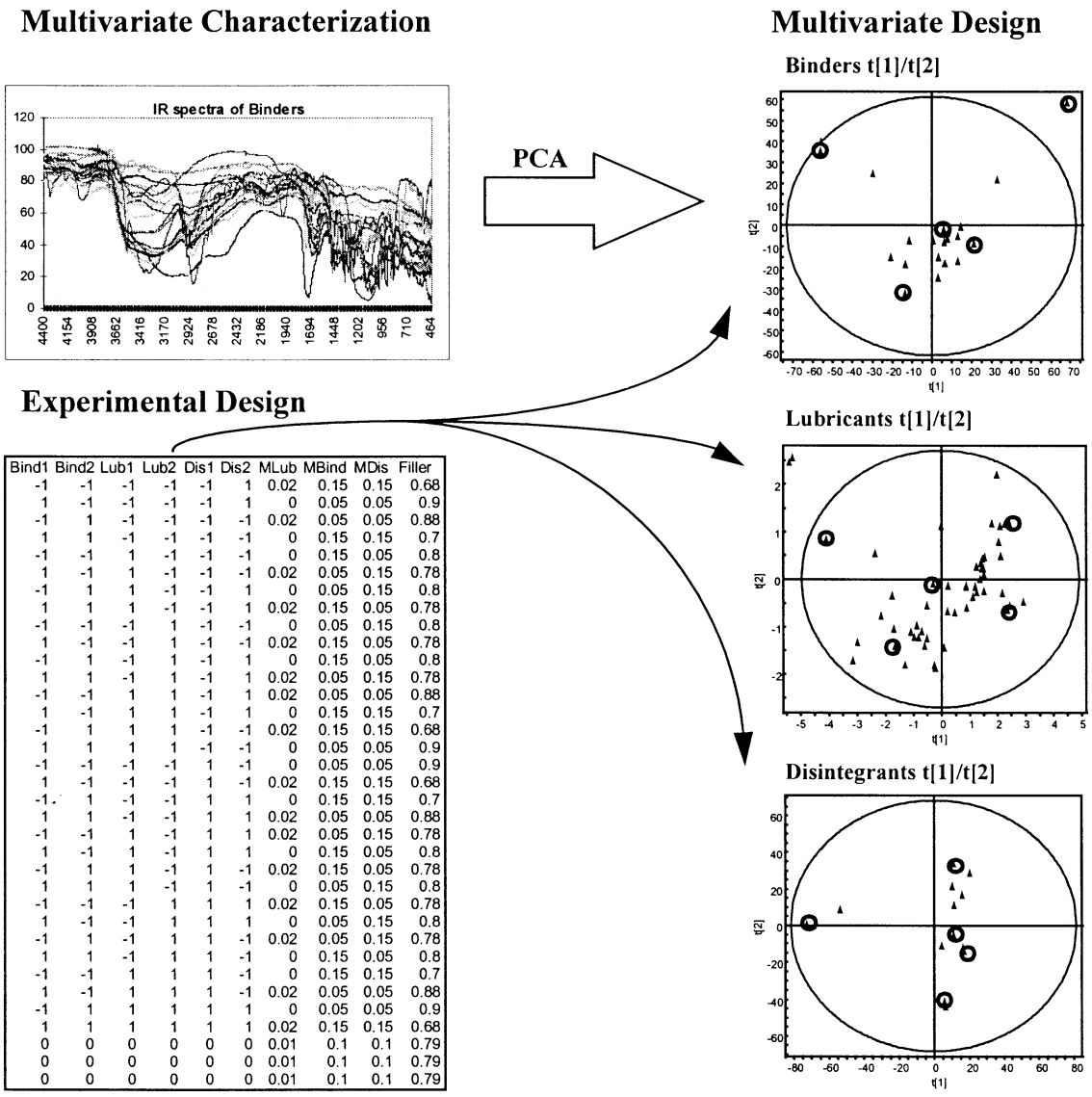
Mixture Design

In a mixture design, the sum of all components add up to 100%. Mixture factors are expressed as the fraction of the total amount (Fig. 4.). This means that they cannot be changed totally independent of one another. Their experimental ranges lie between 0 and 1. A mixture factor can be a formulation factor or a filler factor. Only one mixture factor can be defined as filler.

Formulation factors are the usual mixture factors used in formulations, and they have specifically defined experimental ranges. A filler is a mixture component, usually of little interest, making up a large percentage of the mixture and added at the end of a formulation to bring the mixture total to the desired amount.

Soft Independent Modeling of Class Analogy

SIMCA is a method used to classify objects (12). Provided that most of the variables express a real similarity, a class can be well approximated by a PC model with few components. From this model, which is the basis of the SIMCA method, a tolerance interval (usually 95%) can be constructed around the PC hyperplanes (Fig. 5a). New objects are assigned to different classes depending on which tolerance ''cylinder'' they fit inside (Fig. 5b).



Bind1	Bind2	Lub1	Lub2	Dis1	Dis2	MLub	MBind	MDis	Filler
-1	-1	-1	-1	-1	1	0.02	0.15	0.15	0.68
1	-1	-1	-1	-1	1	0	0.05	0.05	0.9
-1	1	-1	-1	-1	-1	0.02	0.05	0.05	0.88
1	1	-1	-1	-1	-1	0	0.15	0.15	0.7
-1	-1	1	-1	-1	-1	0	0.15	0.05	0.8
1	-1	1	-1	-1	-1	0.02	0.05	0.15	0.78
-1	1	1	-1	-1	1	0	0.05	0.15	0.8
1	1	1	-1	-1	1	0.02	0.15	0.05	0.78
-1	-1	-1	1	-1	-1	0	0.05	0.15	0.8
1	-1	-1	1	-1	-1	0.02	0.15	0.05	0.78
-1	1	-1	1	-1	1	0	0.15	0.05	0.8
1	1	-1	1	-1	1	0.02	0.05	0.15	0.78
-1	-1	1	1	-1	1	0.02	0.05	0.05	0.88
1	-1	1	1	-1	1	0	0.15	0.15	0.7
-1	1	1	1	-1	-1	0.02	0.15	0.15	0.68
1	1	1	1	-1	-1	0	0.05	0.05	0.9
-1	-1	-1	-1	1	-1	0	0.05	0.05	0.9
1	-1	-1	-1	1	-1	0.02	0.15	0.15	0.68
-1	1	-1	-1	1	1	0	0.15	0.15	0.7
1	1	-1	-1	1	1	0.02	0.05	0.05	0.88
-1	-1	1	-1	1	1	0.02	0.05	0.15	0.78
1	-1	1	-1	1	1	0	0.15	0.05	0.8
-1	1	1	-1	1	-1	0.02	0.15	0.05	0.78
1	1	1	-1	1	-1	0	0.05	0.15	0.8
-1	-1	-1	1	1	1	0.02	0.15	0.05	0.78
1	-1	-1	1	1	1	0	0.05	0.15	0.8
-1	1	-1	1	1	-1	0.02	0.05	0.15	0.78
1	1	-1	1	1	-1	0	0.15	0.05	0.8
-1	-1	1	1	1	-1	0	0.15	0.15	0.7
1	-1	1	1	1	-1	0.02	0.05	0.05	0.88
-1	1	1	1	1	1	0	0.05	0.05	0.9
1	1	1	1	1	1	0.02	0.15	0.15	0.68
0	0	0	0	0	0	0.01	0.1	0.1	0.79
0	0	0	0	0	0	0.01	0.1	0.1	0.79
0	0	0	0	0	0	0.01	0.1	0.1	0.79

Figure 4. Overview of the principles of multivariate design. The experimental design is made in principal properties and substances that correspond to the min and max values are identified in the score plots. The amount of each excipient (MLub, MBind, MDis, Filler) in the formulation is included in a mixture design.

All objects outside of the tolerance intervals are classified as outliers.

RESULTS

The Multivariate Characterization

The excipients were divided into classes depending on their function in the formulation and were modeled in separate PCAs.

In the model for lubricants, 53 substances were described in six variables taken from the literature. Two principal components were calculated. Because of the nature of the literature, it was possible to predict early where in the PCA score plot the best-working lubricants should appear. One excipient from every quadrant of the score plot was selected for the experimental design, plus one close to the origin. From the top right corner of the diagram, which was the area for better performing lubricants, the choice fell on magnesium stearate instead of

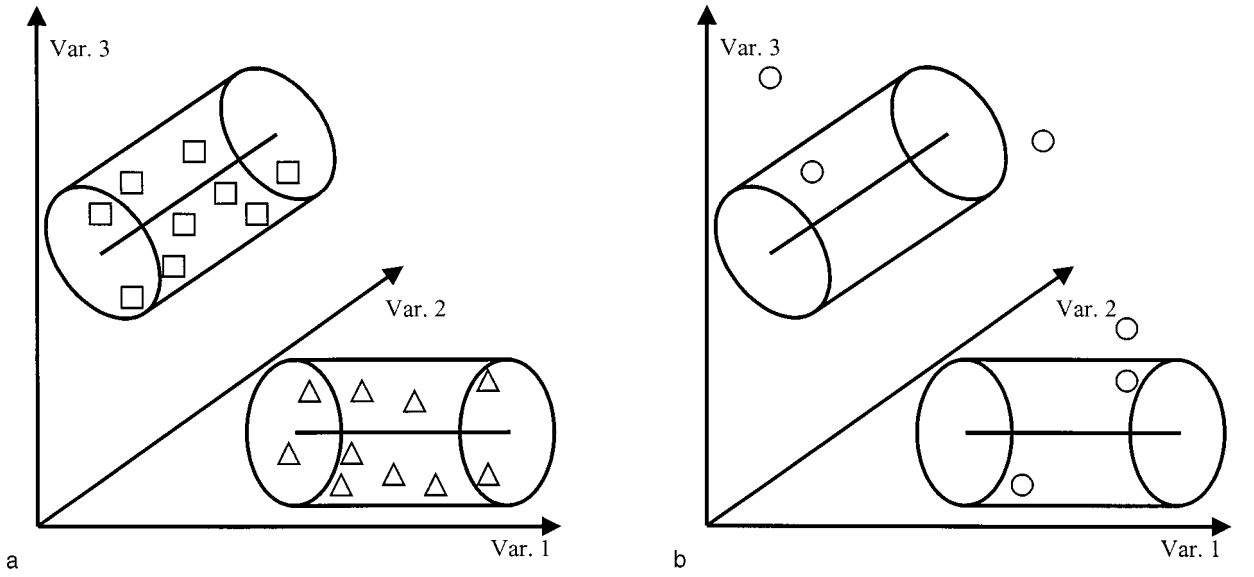


Figure 5. Illustration of the SIMCA method. Around one or more classes, a tolerance interval is constructed. New objects that fit inside the interval are said to belong to the class.

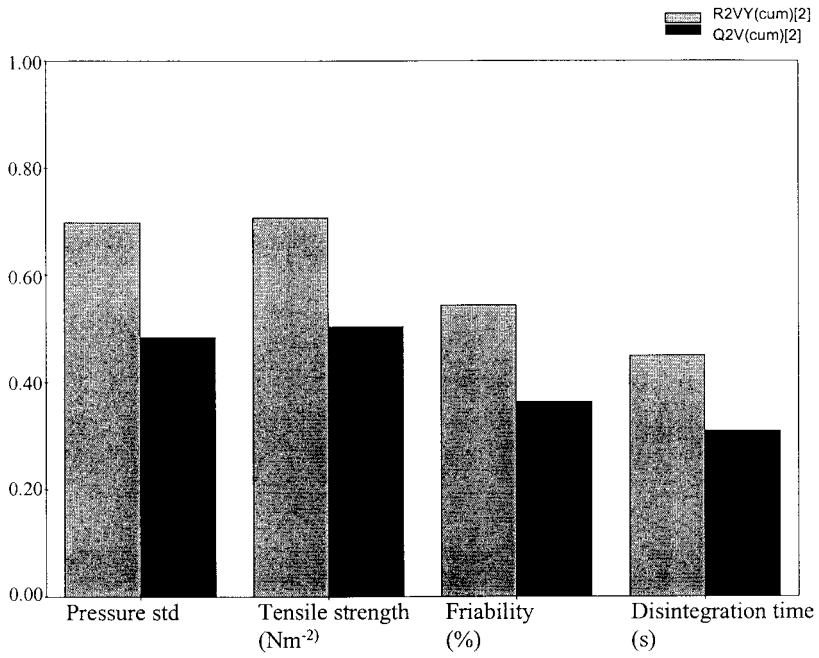


Figure 6. Summary overview plot for the recorded responses. Disintegration time is the response that has the least variance explained. The reason for this is mainly three of the tablets that never dissolved during the time allowed for the test. As different settings of experimental parameters yielded the same result, the effect on both the R^2 and the Q^2 is, of course, negative.

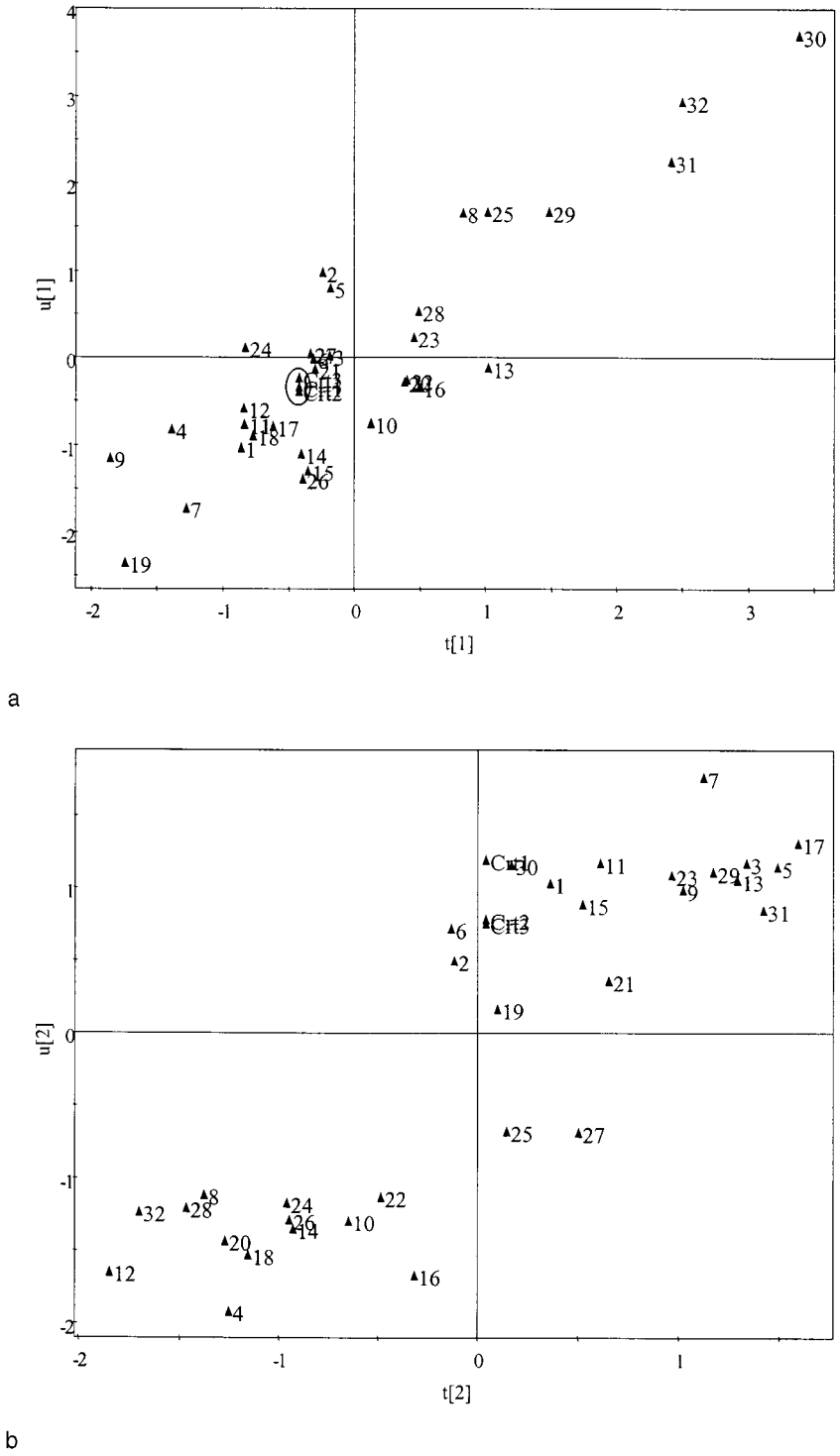


Figure 7. Score plots of the (a) first and (b) second PLS components. The first component shows a satisfactory linear relation. The center points (circled) lie close together near the origin in the score plot of the first PC. The second component reveals two groupings.

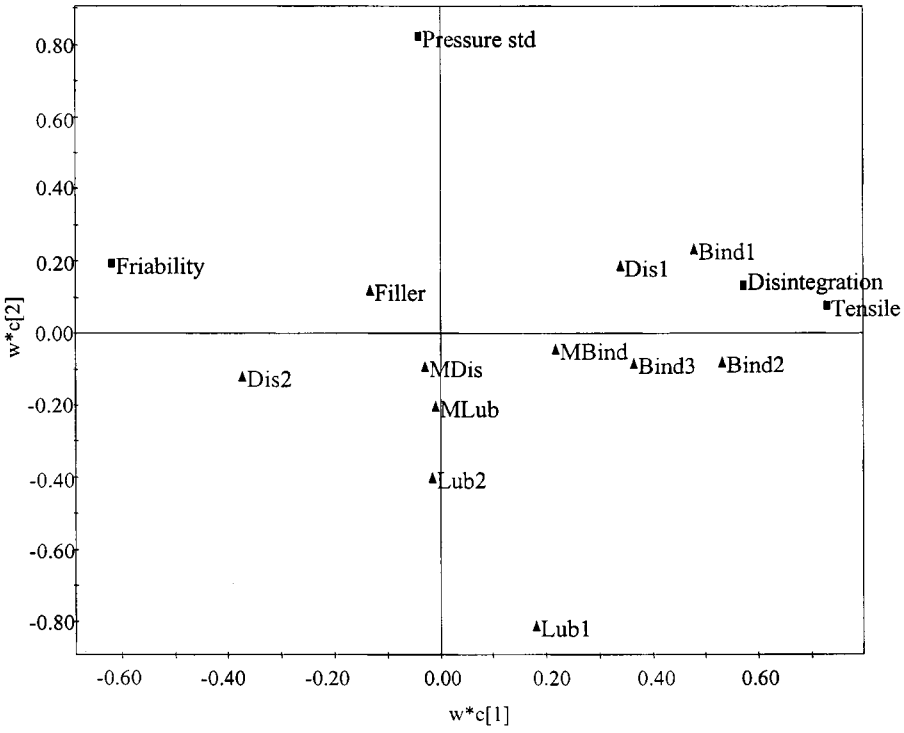


Figure 8. Loading plot of the first and second principal components. The first component describes the responses tensile strength, disintegration time, and friability. These responses are explained by the principal properties for the binder and the disintegrant (Bind1, Bind2, Bind3, Dis1, and Dis2). The variance in pressure STD is explained by the principal properties for the lubricants (Lub1 and Lub2) in the second component. The amount of binder is the only amount of excipient (MBind, MDis, MLub, and Filler) that affects tablet quality.

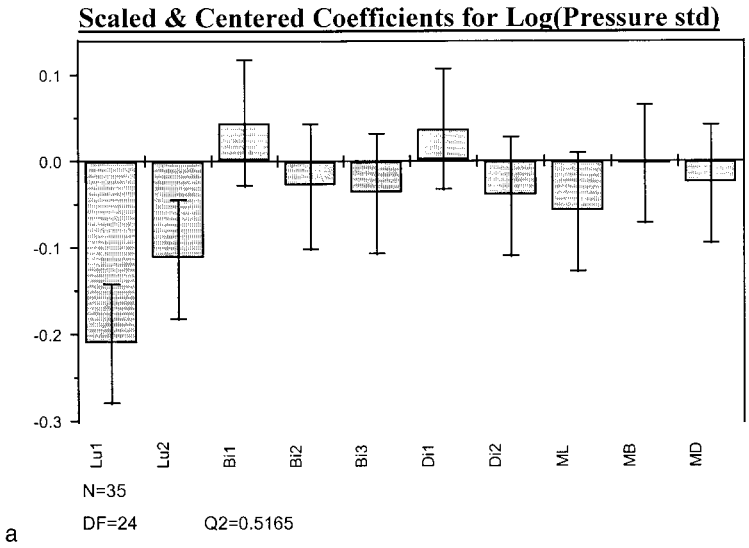
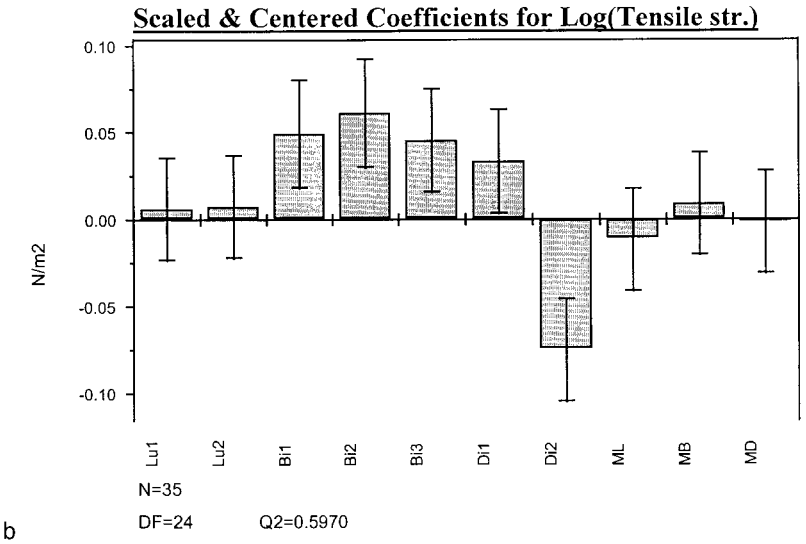
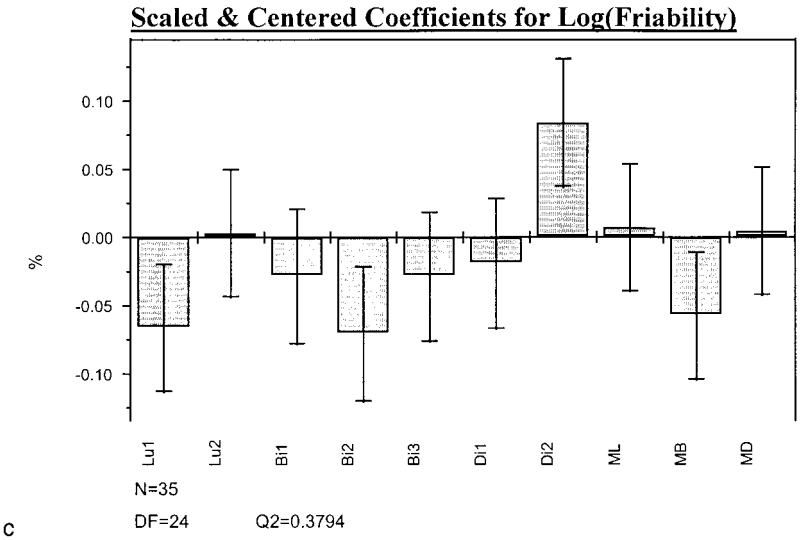


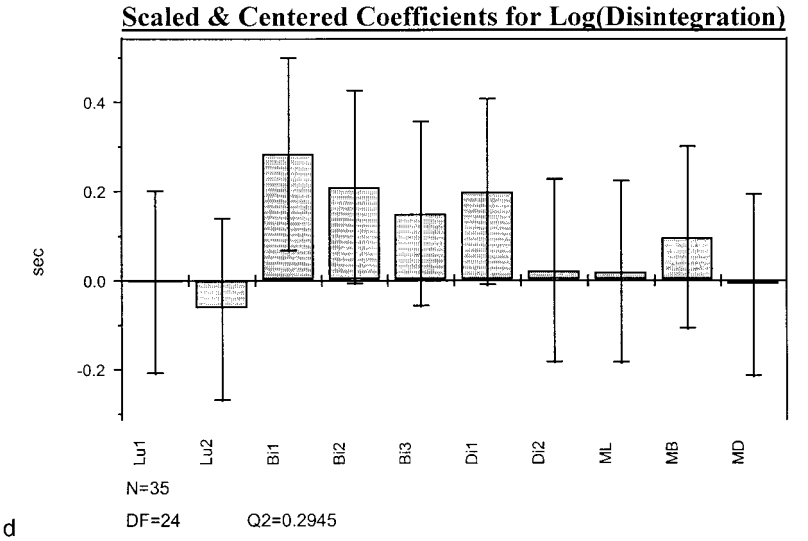
Figure 9. Coefficient plots for all responses. (a) Pressure STD is affected only by the principal properties for the lubricants (Lu1, Lu2). (b) All principal properties for the binder (Bi1, Bi2, Bi3), together with the second component for the disintegrant (Di2), have a statistically proven effect on tensile strength. (c) The first principal property for lubricants (Lu1), the second components for the binders (Bi2) and disintegrants (Di2), together with the amount of binder (MB) affect the friability of the tablet. (d) Only the first and second principal properties for the binders (Bi1, Bi2) have a statistically proven effect on the disintegration time.



b

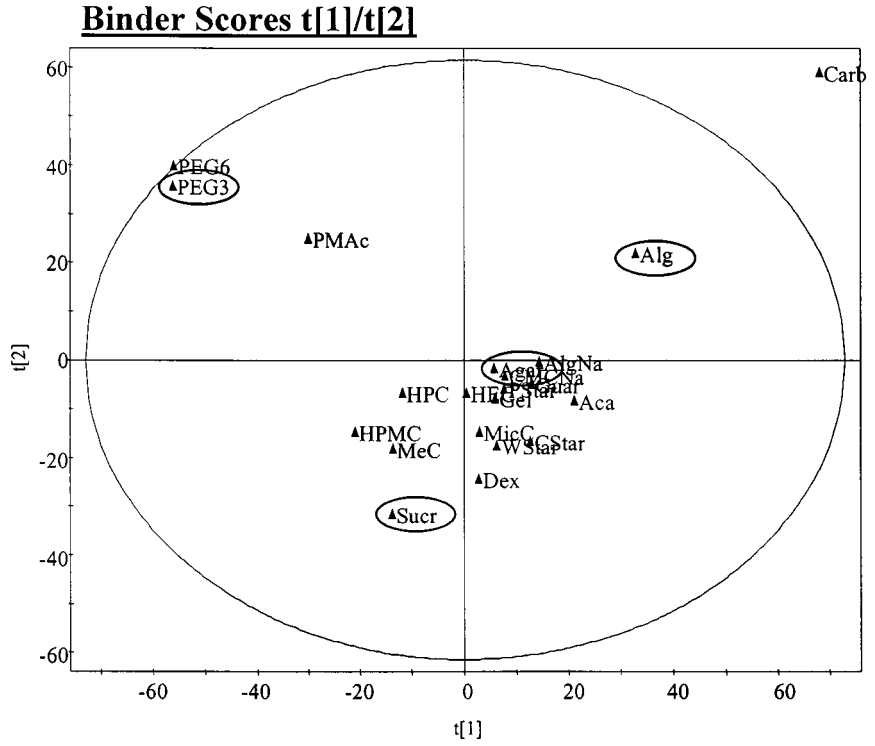


c

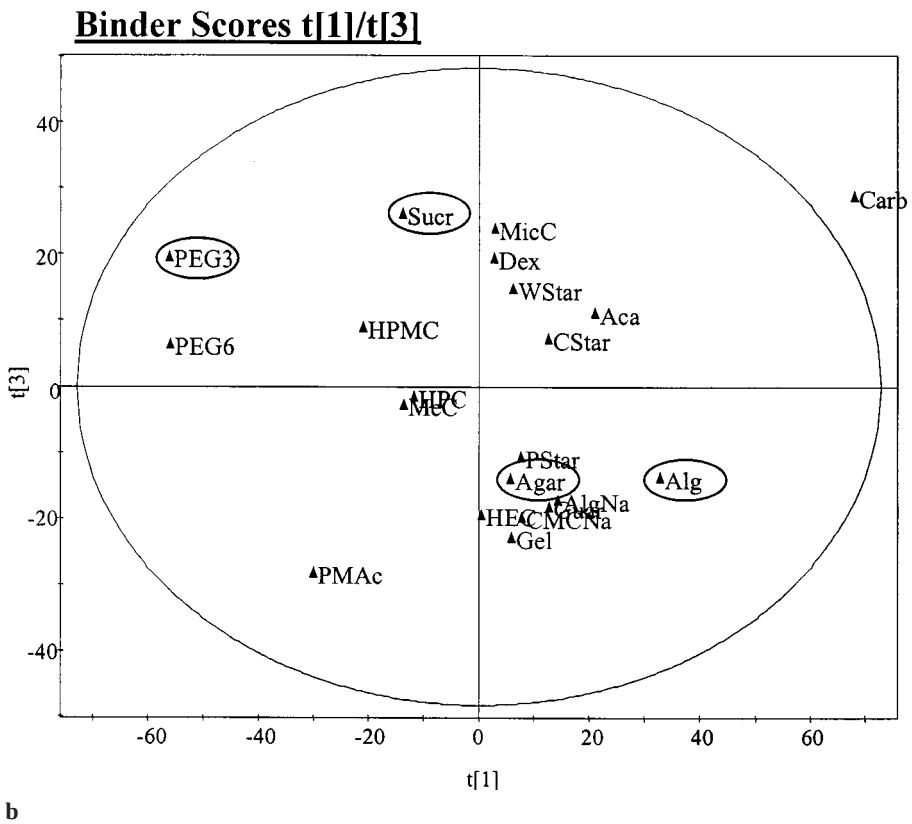


d

Figure 9. Continued

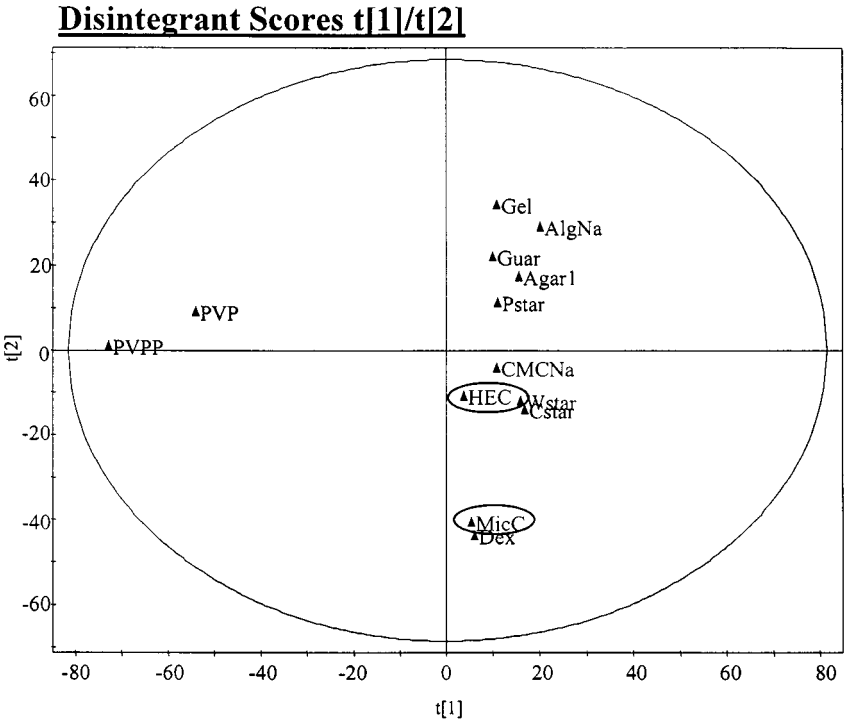


a

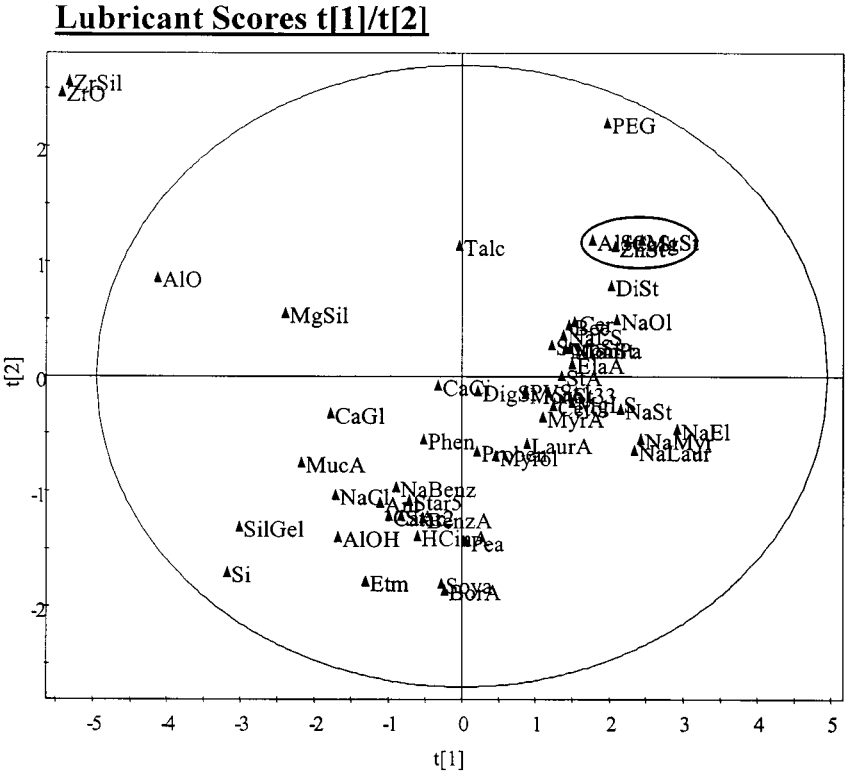


b

Figure 10. Excipients chosen for additional experiments are marked in the score plots.



c



d

Figure 10. Continued

polyethylene glycol (PEG) 3000. It was understood that the position of PEG3000 in the ''good'' corner of the plot was the result of an extreme molecular weight and not due to good lubricating quality. All five lubricant selections were added to the design with two principal properties, one for each principal component (Table 1, experimental plan).

As for the binder and disintegrant classes, which were subjected to IR analyses, there turned out to be an overlap between the two. Some compounds were listed in the literature both as binders and as disintegrants and presumably were able to work both ways.

The collected 21 compounds, of mainly binding qualities, needed three principal components in the PCA to cover fully the variations in quality. The selections for the design had to be eight compounds plus a center point (Table 1, experimental plan). Every selected binder thus had three principal properties in the design because of the three principal components.

Only 13 disintegrants were collected for the IR analyses. Two principal components were calculated in the PCA model. The poor spread of compounds did not allow for symmetric selections from every quadrant for the experimental design. Five disintegrants were selected anyhow, as far away from each other as possible. They all received two latent variables each from the principal components.

Partial Least Squares of Screening Experiments

Large differences between the smallest and largest value for the variables, especially disintegration time, warranted a log transformation of the response data. There are no significant interaction terms. The R^2 values [R^2Y(cum) pressure STD (0.70), tensile strength (0.71), friability (0.54), and disintegration time (0.45)] are good, and the Q^2 values [Q^2(cum) pressure STD (0.48), tensile strength (0.50), friability (0.36), and disintegration time (0.31)] show a highly significant model (Fig. 6). There are no outliers.

The very high tensile strength, low friability, and very long disintegration time of the tablets containing carbomer (objects 29–32) have a big impact on the first principal component (Fig. 7). These objects have a compressing effect on the rest, and the $t[1]/u[1]$ score plot shows centering of the objects around the origin.

Tensile strength, disintegration time, and friability separate the objects along the first component (Fig. 8). The score plot for the second component shows two groupings, and pressure STD is the response separating the objects (Fig. 7).

Tensile strength is positively correlated with disintegration time and negatively correlated with friability. The loading plot shows a positive correlation between the responses tensile strength and disintegration time and the first two principal properties of binder. Friability is positively correlated with the second principal property of the disintegrants. Pressure STD is negatively correlated with both principal properties of the lubricants.

New Experiments

Selecting Excipients for Additional Experiments

The multivariate design was generated in Modde, and evaluation of the screening experiments for further experiments was also carried out in Modde.

The regression coefficient plots show to what extent the variables in X influence the responses (Fig. 9). In selecting new experiments to improve tablet quality, the coefficient plots were carefully examined.

To decrease the pressure STD, a lubricant should have positive values in both of the principal components. Relating this information to the original score plots, a good lubricant would be found in the upper right corner (Fig. 10). To give high tensile strength, a binder with positive values in all three of the principal components is best suited. It should be combined with a disintegrant that has a positive value in the first PC and a negative value in the second. To minimize the friability, a disintegrant with a negative value in the second principal component should be combined with binder that has a positive value in the second PC and a lubricant that has a positive score in the first PC. A good lubricant keeps friability down as it provides the tablet with a smooth and undamaged surface. The amount of binder should also be kept high to ensure low friability. For tablets to dissolve as quickly as possible, a disintegrant with a negative value in the first PC should be combined with a binder having negative values in all principal components. Notable is that the disintegrants do not have a statistically proven effect on the disintegration time.

Friability and disintegration time are the most important responses when it comes to tablet formulation. From the loading plots, we know that friability and tensile strength are negatively correlated. A tablet with high tensile strength often has low friability. The problem of combining high tensile strength and short disintegration time was approached when planning for the follow-up experiments. Four different combinations of excipients

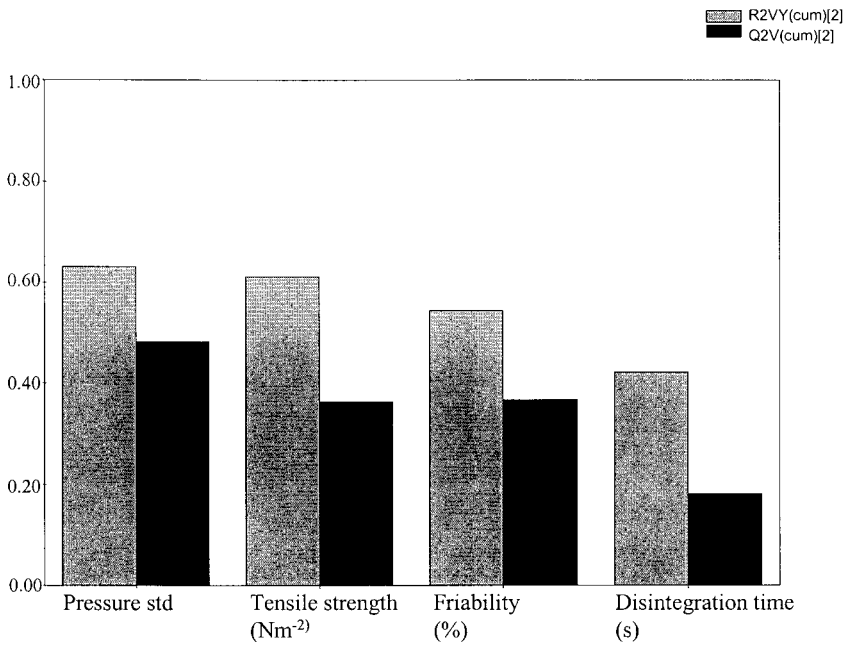


Figure 11. Summary overview plot for the model also including optimization experiments.

were chosen, based on the original score plots, for additional experiments. The amounts were kept at the same levels as in the screening experiments.

A lubricant with positive principal property values was used for all combinations: magnesium stearate, calcium stearate, aluminum stearate (Fig. 10).

Optimization Experiments

Experiment 1

A binder with positive values in the first two principal components was chosen to give high tensile strength at the expense of disintegration time. Carbomer is positive in all three components (Figs. 9a, 9b), but was not regarded as an option as in most cases it gives disintegration times well over what is considered acceptable. Alginic acid was chosen.

The perfect disintegrant would be one with negative values in both principal properties because it would shorten the disintegration time without too much loss of tensile strength. There is no such disintegrant in the original score plots (Fig. 9c). The best alternative was a disintegrant with a negative value in the second PC as it has the most positive effects on tensile strength and friability. Microcrystalline cellulose was chosen (Fig. 10).

Experiment 2

The binder was chosen to give a short disintegration time with as little influence on tensile strength as possible. Sucrose has negative values in the first two principal components and a positive value in the third (Figs. 9a, 9b). Once again, microcrystalline cellulose was chosen as the disintegrant to give high tensile strength and low friability.

Experiment 3

The excipients were chosen to represent a cross between experiments 1 and 2. Agar and hydroxyethylcellulose are both as close as possible to the origin in the binder and disintegrant score plots, respectively (Fig. 10).

Experiment 4

Choices were based on the Optimizer function in Modde 4.0 and represent another solution to the problem of how to achieve high tensile strength and disintegration time. Since a positive value in the first principal property for a binder is the biggest contributor to increasing disintegration time, that was kept negative. At the same time, it is not the most important principal property for tensile strength. Tensile strength was achieved through choosing

PEG3000, a binder with high values in the second and third principal properties. The disintegrant is the same as in experiments 1 and 2, microcrystalline cellulose.

The new experiments were called optimization screening experiments (Table 1). Additional experiments that either did not fit inside the extended worksheet or for which the amounts of the excipients were determined by the Optimizer were also carried out. They were called pure optimization experiments (Table 1).

Analysis of Screening and Optimization Experiments

Partial Least Squares of Screening and Optimization Experiments

The two-component PLS model of the screening experiments together with the optimization screening experiments improve the R^2 value [R^2Y(cum) pressure STD (0.63), tensile strength (0.61), friability (0.54), disintegration time (0.42)] for friability slightly, but actually decrease for the other responses (Fig. 11). The Q^2 values [Q^2(cum) pressure STD (0.48), tensile strength (0.36), friability (0.37), and disintegration time (0.18)] also decrease slightly for all responses except pressure STD and friability. There were no significant interaction terms. There are no outliers.

The additional experiments do not cause a big change in the score plot (Fig. 12). Most of the new tablets have a high tensile strength and low friability, which gives them positive values in the first component. In the loading plot, disintegration time is now separated from tensile strength (Fig. 13). The new experiments contain several examples of tablets with high tensile strength and short disintegration time, which might have brought about the change. The score plot for the second component again shows the grouping of objects due to the ability of the lubricant (Fig. 12). As expected, all additional experiments (except for the center point) are a part of the group with low values for pressure STD. The new objects also tend to place themselves to the left inside of that group. This is probably due to their rather short disintegration times.

Principal Components Analysis of Responses for Screening and Optimization Experiments

The two-component model has high R^2 values for all responses except disintegration time [R^2(cum) pressure STD (0.83), tensile strength (0.91), friability (0.90), and disintegration time (0.62)]. There are no outliers.

All tablets from the optimization experiments, except for the repeated center point, are, as expected, a part of the grouping containing efficient lubricants (Fig. 14). In addition to those, the group consists of a number of tablets containing lauric acid sodium salt and two objects with aluminum oxide and aluminum hydroxide (exp. 25 and 27). The last two were probably helped by the lubricating ability of PEG3000. These objects all have a pressure STD of less than 1, which is close to the standard deviation of the tablet machine. This means that no particles stuck to the die wall. S3 deviates from the rest due

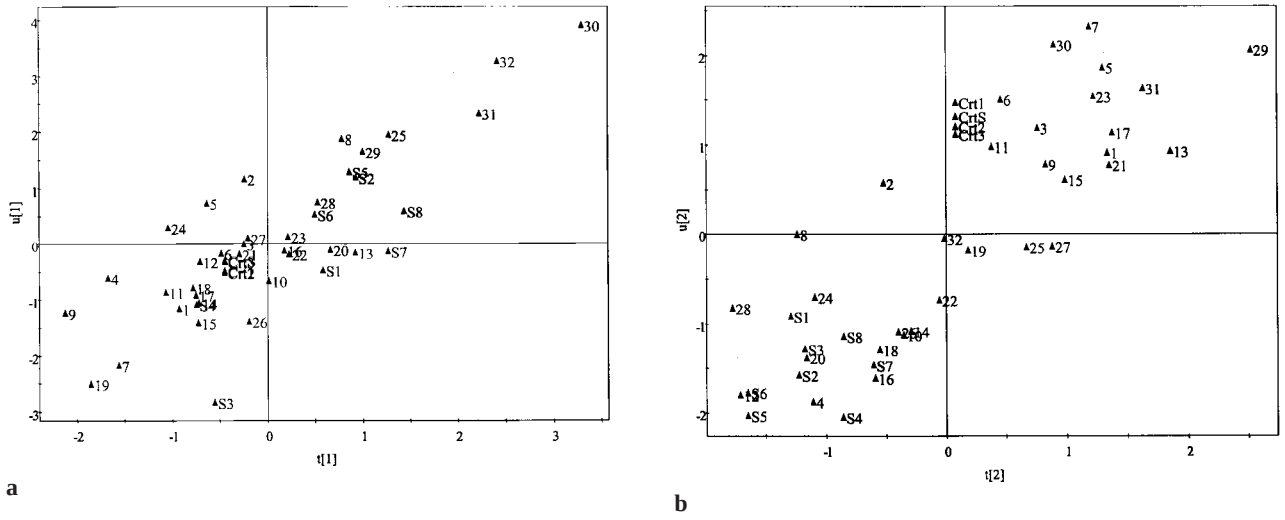


Figure 12. Score plots of the first and second PLS components. The optimization screening experiments (S) do not have a noticeable effect in the relation between X and Y in the first component.

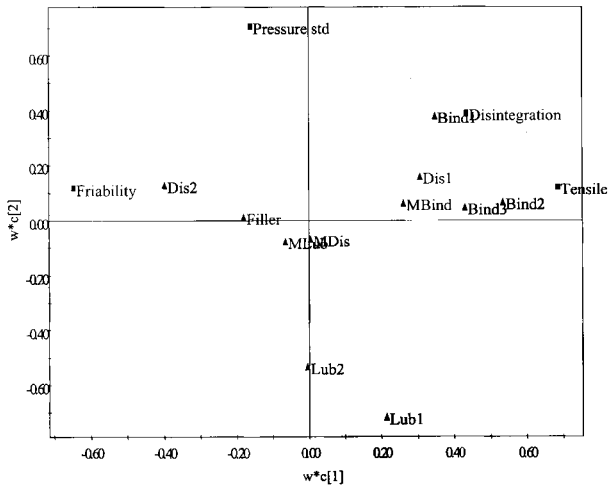


Figure 13. Loading plot for the first and second component.

to high friability. Object 2 has a pressure STD close to 3, but is still close to the grouping because of high tensile strength, low friability, and short disintegration time.

Soft Independent Modeling of Class Analogy

To get an idea of good tablet standards, six of the staff at Solid Formulation were asked to give values for three of the responses (Table 3). Pressure STD was set to 0.5 for all. These values and the responses from the screening and optimization experiments were used in a model created to illustrate the SIMCA method.

Table 3

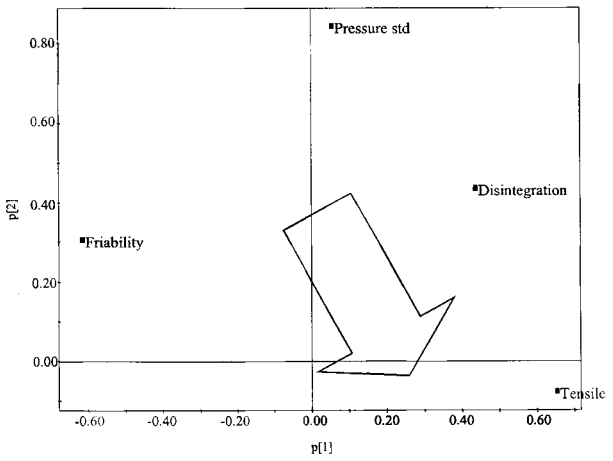
Personal Opinion of a Good Tablet

Staff	Pressure SD	Tensile Strength (Nm^{-2})	Friability (%)	Disintegration Time (sec)
Director 1	0.5	1,800,000	0.5	900
Director 2	0.5	1,500,000	0.5	600
Director 3	0.5	1,200,000	1	900
Engineer 1	0.5	1,900,000	0.5	300
Engineer 2	0.5	1,700,000	0.8	600
Engineer 3	0.5	1,500,000	1	1200

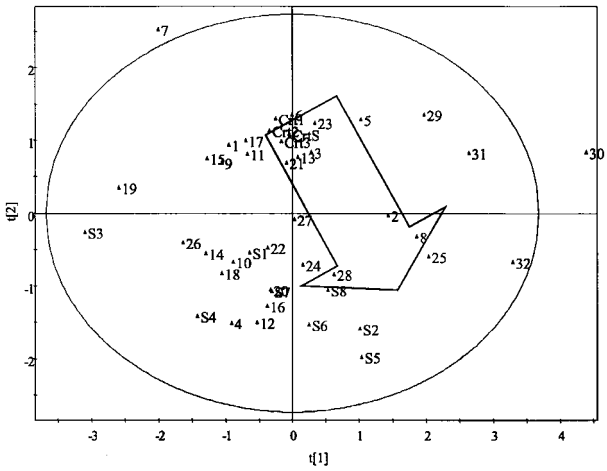
Six of the staff at Solid Formulation, Pharmacia & Upjohn, were asked to describe a good tablet in three responses: tensile strength, friability, and disintegration time. The pressure STD was set to 0.5 to enable this response to be included in the model. The numerical order refers to experience in the respective positions.

The score plot from this model shows the extent to which the experiments match the demands of the experts (Fig. 15). Tablets that meet the requirements of the staff are found in their vicinity in the score plot.

Object 28 matches the demands of director 3. Engineer 3 can find many tablets that fulfill his requirements.



a



b

Figure 14. The loading plot and the score plot that display similar relationships between the responses that have been observed in previous models. By weighing all responses together in the loading plot, with an ideal tablet in mind, a ''direction'' can be drawn. By transferring this direction to the score plot, the area in which good tablets are found is evident. This model apparently lacks tablets of high quality.

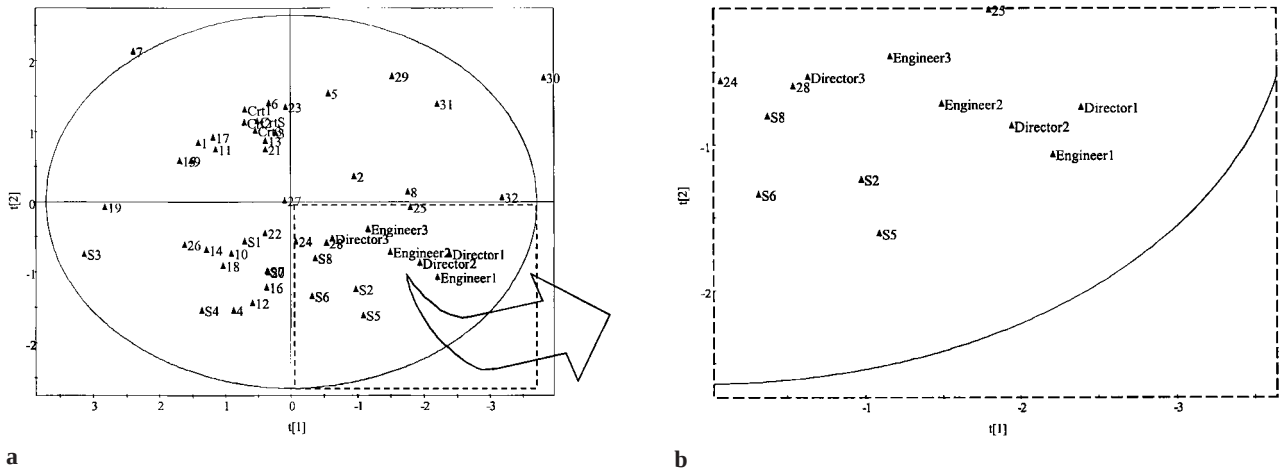


Figure 15. Score plot from the model containing screening and optimization experiments together with statements from the experts. Some of the objects representing the staff fill the void in the lower right of the score plot.

The others have demands that none of the tablets included in these experiments can match. Notable is that demands on tablet quality increase with higher rank, that is, the more experienced a person is in his or her position.

The pure optimization experiments meet even the toughest demands of the staff (Fig. 16). There are several tablets that fulfill the needs of everybody. Since friability is more important than hardness, OP1 and OP4 will do even for engineer 1 and director 1.

This method of classification could also be used as a means for designing a tablet with specific qualities, for example, a tablet with a very short disintegration time is required, while friability and tensile strength are less important. The disintegration time cannot exceed 1.5 min. Another demand might also be to cut costs and use as much filler as possible. As far as disintegration time goes, there is the possibility of choosing among experiments OP, OP1, S2, S5, S7, S8, and 12. These tablets have a friability of less than 1% as well. The next requirements, cost efficiency and a filler content of about 80%, were best accommodated by OP, S5, and S7.

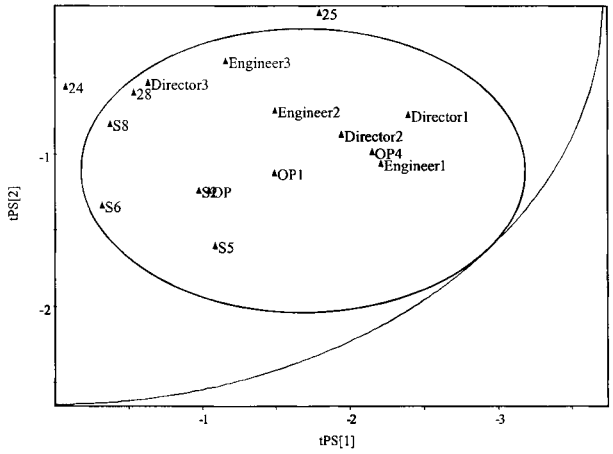


Figure 16. Score plot with *t* values for the pure optimization experiments (OP, OP1, OP4) predicted from the current model. The circle represents the area in which tablets of good quality are found.

Another application is in the development of a new drug. Knowledge about the active substance might prevent the use of certain excipients. Target values can be set as in the example above, and again, there is the possibility of choosing OP, OP1, S2, S5, S7, S8, or 12. The new active substance is very sensitive to water, and PEG3000 is known to be very hygroscopic. That leaves OP, OP1, S2, S7, S8, and 12 for further studies.

CONCLUSIONS

Screening Experiments Successful

The PCA and PLS models offered a good overview of the possibilities of tablet formulation. The greatest challenge lies in combining low friability and short disintegration time. The coefficient plots from the four-component PLS model generated in Modde 4.0 offered several alternatives for selecting excipients to make tablets of good quality. The coefficient plots were used for selecting excipients with good predicted properties for ad-

ditional experiments. The validity of the model was confirmed by making these formulations.

Information from the Principal Properties of the Excipients

The study offered much information about the excipients. Some was expected, and some was quite unexpected. Magnesium stearate is a proven lubricant. It was expected that good lubricants would be found in the vicinity of magnesium stearate in the score plot characterizing the lubricants. The results showed that a good lubricant has positive values in both principal properties. It was also shown that the first principal property is the most important one, which implies that useful lubricants could also be found in other parts of the score plot. Some polymers even make the lubricant obsolete.

In this study, the binder is the most important excipient in tablet formulation. As the binder has the biggest impact on tensile strength and disintegration time, the strategies for obtaining tablets of optimum quality were based on the choice of binder.

The difficulty in defining binders and disintegrants was further emphasized by this study. The disintegrants had no proven effect on disintegration time. However, as they did have an effect on tensile strength and friability, which are correlated with disintegration time, the possibility of distinguishing between the two classes becomes difficult. The SIMCA method can be used to predict the principal properties of an excipient in both the binder and the disintegrant models. Based on the IR spectra, it would then be possible to predict its potential in tablet formulation.

The model that included the opinions of the staff at Pharmaceutical Development provided an excellent overview of the qualities that are in demand in modern pharmaceutical work. There was clearly a region in which good tablets could be found. The pure optimization experiments proved to match even the toughest demands. This method can also be used in more complex problems (e.g., dissolution curves) when specific requirements need to be fulfilled.

Further Applications

The SIMCA method could be used for selecting formulations of desired qualities. Combined with a well-working Laboratory Information Management System (LIMS), containing information such as physicochemical properties and compatibility with known substances, a candidate drug could be combined quickly with excipients that do not impair performance. This would lead to rapid development of new formulations. The next step is multivariate characterization of a number of active substances, potential candidate drugs of different classes, to create a model to improve further predictions of compatibility.

Multivariate characterization of new excipients enables rapid assessment of their qualities and expected use. Combining multivariate characterization, physicochemical properties contained in LIMS, experimental design, multivariate design, and PLS would lead to an evolutionary strategy for tablet formulation. Since it includes a learning strategy that continuously incorporates data for new compounds and from conducted experiments, this would be an even more powerful tool than expert systems.

ACKNOWLEDGMENTS

The study was performed in two parts as examination projects during spring and fall 1997 under the supervision of Prof. Michael Sjöström, Organic Chemistry, Umeå University, Sweden. Pharmacia and Upjohn in Uppsala, Sweden, provided all necessary material and facilities during both projects, which we gratefully acknowledge. Tesfai Sebhatu, Mats E. Johansson, and Margareta Duberg of Solid Formulation, Pharmacia and Upjohn, Uppsala, were a great help in preparing for the experiments and teaching the basics of small-scale tablet production. Nils-Olof Lindberg at Pharmacia and Upjohn in Helsingborg provided literature references and chemicals, which was greatly appreciated.

REFERENCES

1. H. Leuenberger, W. Becher, A factorial design for compatibility studies in pre-formulation work, Pharm. Acta Helv., 50, 88–91 (1975).
2. R. Carlson, M. P. Prochazka, T. Lundstedt, Principal properties for synthetic screening: ketones and aldehydes, Acta Chem. Scand., B42, 145–156 (1988).
3. R. Carlson, M. P. Prochazka, and T. Lundstedt, Principal properties for synthetic screening: amines, Acta Chem. Scand., B42, 157–165 (1988).
4. T. Lundstedt, P. M. Andersson, S. Clementi, G. Cruciani, N. Kettaneh. A. Linusson, B. Nordén, M. Pastor, M. Sjöström, and S. Wold, Intelligent combinatorial libraries, in *Computer-Assisted Lead Finding and Optimization* (H. van de Waterbeemd, Ed.), Verlag Helvetica Chimica Acta, Basel, Switzerland, 1997, pp. 191–208.
5. P.M. Andersson, A. Linusson, S. Wold, M. Sjöström, T.

Lundstedt, and B. Nordén, Design of small libraries for lead exploration, in *Molecular Diversity in Drug Design* (R. Lewis and P. M. Dean, Eds.), Kluwer Academic, Dordrecht, The Netherlands, 1999, pp. 197–220.
6. T. Lundstedt, A QSAR strategy for screening of drugs—and predicting their clinical activity, Drug News Perspec., 4(8), 468–474 (1991).
7. N.-O. Lindberg and T. Lundstedt, Application of multivariate analysis in pharmaceutical development work, Drug Dev. Ind. Pharm., 21(9), 987–1007 (1995).
8. G. A. Lewis, D. Mathieu, and R. Phan-Tan-Luu, *Pharmaceutical Experimental Design, Drugs and the Pharmaceutical Sciences*, Vol. 92, Marcel Dekker, New York, 1998.
9. Bruce R. Kowalski, *Chemometrics—Mathematics and Statistics in Chemistry*, Reidel, Dordrecht, The Netherlands, 1984.
10. G. E. P. Box, W. G. Hunter, and J. S. Hunter, *Statistics for Experimenters*, Wiley, New York, 1978.
11. S. Wold, M. Sjöström, R. Carlson, T. Lundstedt, S. Hellberg, B. Skagerberg, C. Wikström, and J. Öhman, Multivariate design, Analytica Chimica Acta, 191 17-32, (1986).
12. B. Thelin, T. Lundstedt, R. Lundgren, A. Olsson, and S. Batra, Classification of Estradurin® batches: correlation between ^{31}P NMR and a biological duration test for batch approval, in *Chemometrics and Intelligent Laboratory Systems*, Elsevier, Amsterdam, 1995.