Drug Development and Industrial Pharmacy, 23(7), 639-645 (1997)

RESEARCH PAPER

Mixture Experimental Design Applied to Solubility Predictions

Dario Vojnovic and Daniela Chicco

Department of Pharmaceutical Sciences, University of Trieste, 34127 Trieste, Italy

ABSTRACT

The solubility of theophylline (as a model drug) was studied in four-component systems, using an a priori experimental strategy. Ethanol, polyethylenglycol, propylenglycol, and water were chosen as cosolvents. A reduced cubic model was postulated to describe the solubility as a function of mixture composition. A priori criteria in combination with an exchange algorithm were used to select, from a set of 31 candidate points, the optimal design with the least number of experiments. A weighting was assigned to each of the 31 experiments, on the basis of the cost, in order to obtain a design that would be optimal also from an economic point of view. Such a methodology made it possible to obtain, with the minimum number of experiments and with a low cost, a model which was validated and found suitable for accurate prediction of solubility.

INTRODUCTION

In the pharmaceutical industry, the formulator is faced with the problem of finding the right blend of excipients which leads to the optimal product. The effects of various excipient proportions on the formulation characteristics must be investigated. A casual experimental approach is, of course, inefficient and generally involves considerable expenditure, as well as waste of time. This can be avoided by a systematic experimental approach. Mixture experimental designs are useful tools in the optimization of pharmaceutical formulations. Such designs have been applied in the study of many formulations such as oral controlled-release tablets (1–3), mucoadhesive gels (4), and granulates manufactured in high-shear mixers (5), as well as in drug solubility studies (6–8).

In order to reduce the number of experiments and to verify, before performing any experiment, if a proposed design contains enough information about the response behavior in the experimental region, the so-called *a priori* criteria have been recently applied in different formulation studies (9,10). The objective of this work is to propose a strategy for reducing the number of experiments and the cost of a design by *a priori* evaluation. This strategy was used to study the solubility of theo-

phylline (as a model drug) in four-component systems, in order to obtain a mathematical model for the description and prediction of the solubility.

MATERIALS AND METHODS

Materials

Theophylline (anhydrous) was purchased from Prodotti Gianni (Italy). Ethanol, polyethyleneglycol 400 (PEG-400), and propyleneglycol were supplied from A.C.E.F. S.p.a. (Italy). Bidistilled water was used as a fourth component. All calculations were made using New Efficient Methodology for Research Using Optimal Design (NEMROD) software (11).

Solubility Determinations

In order to determine the solubility of theophylline in the different solvent mixtures, 30 ml of each mixture was placed in 50-ml flasks (Pyrex), and excess theophylline was added to obtain saturated solutions. The mixtures were shaken for 24 hr at 25°C. The solutions were then filtered through a 0.8-μm membrane filter (Millipore Corp.). The concentration of theophylline was then determined spectrophotometrically, after proper dilution, using a Perkin–Elmer spectrophotometer (Model 552). The absorbance was measured at 271 nm, and the concentration of theophylline in a given solvent blend was calculated based on standard curves obtained from solutions of theophylline in each blend.

Screening Experiments

Before building up any experimental strategy to study the solubility of theophylline as a function of the component proportions, it was necessary to estimate, first of all, the effect of each component on the response, i.e., to verify if each of the components affects the response, and to find the concentration interval by which the response is significantly altered.

The effect of a mixture component on the response can be studied by the use of axial design, which consists of points positioned on the so-called component axes (12). The points of the axial design for a four-component mixture are represented in Fig. 1, and their coordinates (the component proportions) are shown in Table 1.

The experimental data from the axial design (Table 1) were analyzed graphically by plotting the response values (Y) versus component proportions, as shown in Fig. 2. It is evident from the graph that each component has considerable effect on the response. Examining the plot, the values 0.1 and 0.7 were chosen as lower and upper limits for the proportion of each component. The resulting experimental domain maintains the shape of a regular tetrahedron (Fig. 3).

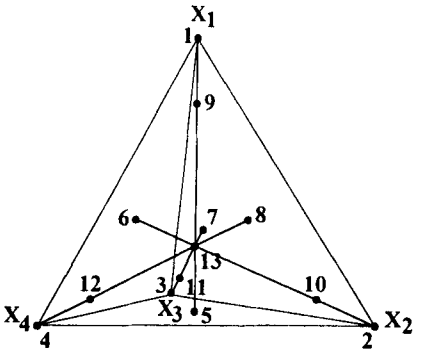

Figure 1. Representation of the experimental points of axial design in the mixture space for a four-component mixture. X_i = component proportions: X_1 = ethanol; X_2 = polyethyleneglycol; X_3 = propyleneglycol; X_4 = water.

Determination of Candidate Experimental Points

The experimental points within the restricted experimental domain were projected by joining two Scheffé's

Table 1

Component Proportions (X_i : see Fig. 1) Corresponding to the Points of the Axial Design, and Response Values

Exp.	X_1[a]	X_2[a]	X_3[a]	X_4[a]	Y (mg/ml)[b]
1	1.000	0.000	0.000	0.000	6.57
2	0.000	1.000	0.000	0.000	12.32
3	0.000	0.000	1.000	0.000	10.72
4	0.000	0.000	0.000	1.000	6.31
5	0.000	0.333	0.333	0.333	13.04
6	0.333	0.000	0.333	0.333	22.56
7	0.333	0.333	0.000	0.333	16.77
8	0.333	0.333	0.333	0.000	11.98
9	0.750	0.083	0.083	0.083	13.73
10	0.083	0.750	0.083	0.083	11.42
11	0.083	0.083	0.750	0.083	15.12
12	0.083	0.083	0.083	0.750	7.10
13	0.250	0.250	0.250	0.250	17.94

[a]All values are expressed as volume fractions of the mixture.
[b]Each response value is the mean of three replicates.

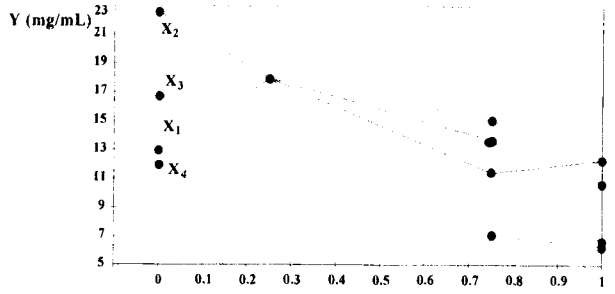

Figure 2. Plot of response values (Y) versus component properties (X_i). X_i = see Fig. 1.

mixture designs (13,14), thereby obtaining a set of 31 experimental points, as shown in Fig. 3. Since restrictions were set on the component proportions, the coordinates of the restricted experimental region were redefined in terms of "pseudocomponents" Z_i (15). The pseudocomponents and the component proportions corresponding to the 31 experimental points are shown in Table 2.

Choice of the Experimental Plan

On the basis of the results of screening experiments, a reduced cubic model [Eq. (1)] was postulated for the

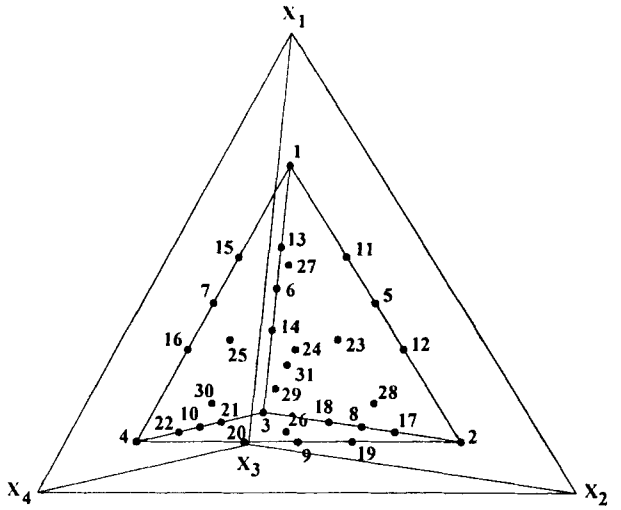

Figure 3. Representation of the experimental points within the experimental domain (internal tetrahedron). X_i = see Fig. 1. The numbers of the experimental points refer to those shown in Table 2.

description of theophylline solubility as a function of solvent composition.

$$Y = b_1Z_1 + b_2Z_2 + b_3Z_3 + b_4Z_4 + b_{12}Z_1Z_2 + b_{13}Z_1Z_3 + b_{14}Z_1Z_4 + b_{23}Z_2Z_3 + b_{24}Z_2Z_4 + b_{34}Z_3Z_4 + b_{123}Z_1Z_2Z_3 + b_{124}Z_1Z_2Z_4 + b_{134}Z_1Z_3Z_4 + b_{234}Z_2Z_3Z_4 \quad (1)$$

Since there are 14 coefficients to estimate, 14 is the minimum number of experiments required to solve the equation. The optimal design with 14 experiments and all the optimal designs with a number of experiments from 15 to 30, were selected by an exchange algorithm, using the NEMROD program (11).

Since another purpose of our experimental strategy was to find an experimental design which would be optimal also from an economic point of view, the cost had to be considered in the design construction. The cost of each experiment was determined using the price information in A.C.E.F.'s catalogue (A.C.E.F. S.p.a., Italy) and the experiments were then pondered on the basis of their cost using an equation which was already applied in the optimization of a granulate formulation (16). These weightings were inserted in the algorithm procedure so that the selection yielded the least expensive optimal designs.

The *a priori* criteria (A-criterion, variance function, and G-efficiency), already described and applied in our previous reports (10,16), were then determined for the optimal designs. Plot (a) in Fig. 4 shows that the design with 20 experiments is optimal with respect to D-criterion, since it has the highest value of the log det(M). The design with 20 experiments also has a good trace value in comparison with the other designs [Fig. (4b)]. Its variance function and G-efficiency are acceptable [Fig. (4c) and (4d)]. Although the designs with 16, 17, 18, and 19 experiments have higher G-efficiency values than the design with 20 experiments, their log det(M) and trace values are worse. This shows that a slight reduction in the number of experiments would lead to a remarkable loss of information. The design with 20 experiments was chosen, considering both the quality criteria and the number of experiments. These 20 experiments are marked by an asterisk in Table 2.

RESULTS AND DISCUSSION

The solubility of theophylline was determined in each of the selected mixtures and the coefficient estimates of the model were calculated by multilinear regression. The

Table 2

Pseudocomponents (Z_1 = ethanol, Z_2 = polyethyleneglycol, Z_3 = propyleneglycol, Z_4 = water) and Component Proportions (X_1 : see Fig. 1) Corresponding to the 31 Experimental Points

	Pseudocomponents				Original Components			
Exp.[a]	Z_1	Z_2	Z_3	Z_4	X_1	X_2	X_3	X_4
1*	1.000	0.000	0.000	0.000	0.700	0.100	0.100	0.100
2*	0.000	1.000	0.000	0.000	0.100	0.700	0.100	0.100
3*	0.000	0.000	1.000	0.000	0.100	0.100	0.700	0.100
4*	0.000	0.000	0.000	1.000	0.100	0.100	0.100	0.700
5*	0.500	0.500	0.000	0.000	0.400	0.400	0.100	0.100
6*	0.500	0.000	0.500	0.000	0.400	0.100	0.400	0.100
7*	0.500	0.000	0.000	0.500	0.400	0.100	0.100	0.400
8*	0.000	0.500	0.500	0.000	0.100	0.400	0.400	0.100
9*	0.000	0.500	0.000	0.500	0.100	0.400	0.100	0.400
10*	0.000	0.000	0.500	0.500	0.100	0.100	0.400	0.400
11*	0.667	0.333	0.000	0.000	0.500	0.300	0.100	0.100
12	0.333	0.667	0.000	0.000	0.300	0.500	0.100	0.100
13	0.667	0.000	0.333	0.000	0.500	0.100	0.300	0.100
14	0.333	0.000	0.667	0.000	0.300	0.100	0.500	0.100
15	0.667	0.000	0.000	0.333	0.500	0.100	0.100	0.300
16	0.333	0.000	0.000	0.667	0.300	0.100	0.100	0.500
17*	0.000	0.667	0.333	0.000	0.100	0.500	0.300	0.100
18	0.000	0.333	0.667	0.000	0.100	0.300	0.500	0.100
19*	0.000	0.667	0.000	0.333	0.100	0.500	0.100	0.300
20	0.000	0.333	0.000	0.667	0.100	0.300	0.100	0.500
21*	0.000	0.000	0.667	0.333	0.100	0.100	0.500	0.300
22*	0.000	0.000	0.333	0.667	0.100	0.100	0.300	0.500
23*	0.333	0.333	0.333	0.000	0.300	0.300	0.300	0.100
24*	0.333	0.333	0.000	0.333	0.300	0.300	0.100	0.300
25*	0.333	0.000	0.333	0.333	0.300	0.100	0.300	0.300
26*	0.000	0.333	0.333	0.333	0.100	0.300	0.300	0.300
27	0.625	0.125	0.125	0.125	0.475	0.175	0.175	0.175
28	0.125	0.625	0.125	0.125	0.175	0.475	0.175	0.175
29	0.125	0.125	0.625	0.125	0.175	0.175	0.475	0.175
30	0.125	0.125	0.125	0.625	0.175	0.175	0.175	0.475
31*	0.250	0.250	0.250	0.250	0.250	0.250	0.250	0.250

[a]Experimental points selected by the exchange algorithm are indicated by asterisks (*).

coefficient estimates together with the experimental and predicted response values are shown in Table 3.

The analysis of variance gave satisfactory R^2 and R^2_{adj} values (R^2 = 0.995; R_{adj} = 0.987). The F ratio for the lack of fit of the model was 2.62, which is less than 27.91, the critical F ratio at α = 0.01 for significant lack of fit. Therefore, at a 99% confidence level, the model lack of fit turned out to be highly nonsignificant.

The response behavior over the whole experimental domain can be easily analyzed by response surfaces obtained from the model. Figure 5 shows some contour diagrams, obtained by Eq. (1), for different proportions of the fourth component (water).

Since each contour represents mixtures which give the same response value, the best mixture can be chosen according to the cost, availability of materials, and

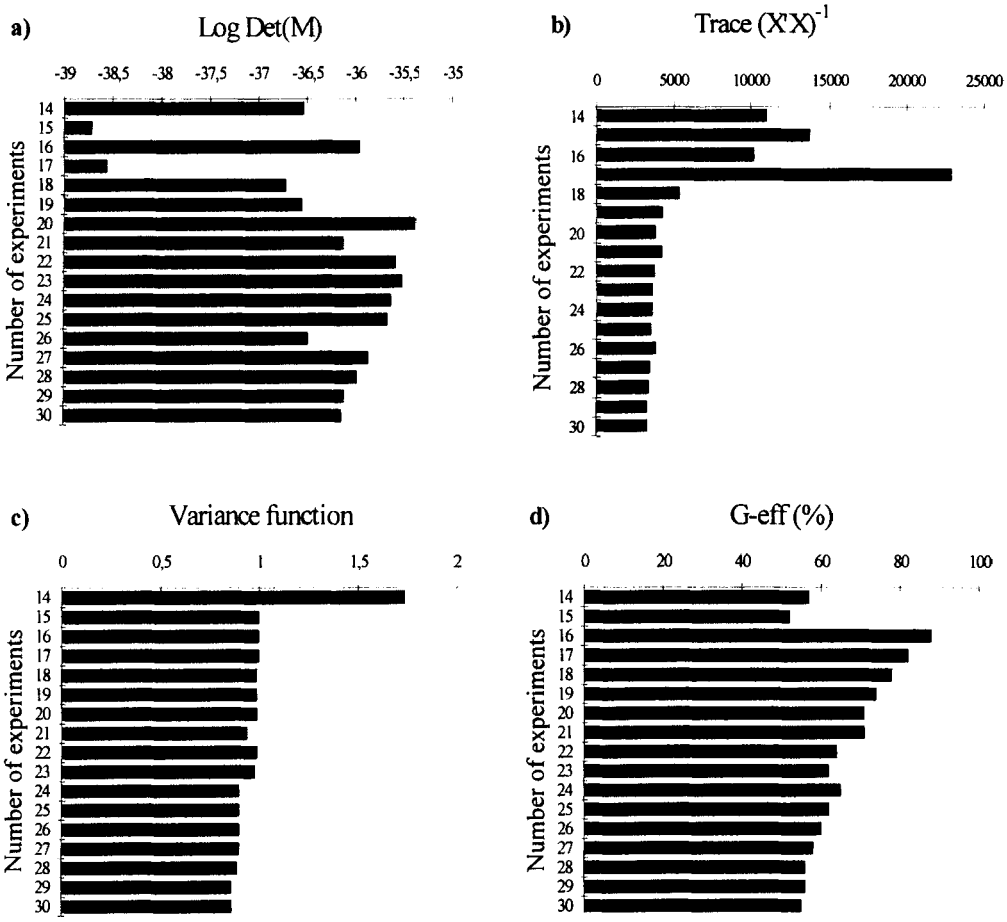

Figure 4. Plots of (a) log Det(M), where $M = (X'X)/N^p$(N = number of experiments contained in the matrix; p = number of model coefficients), (b) trace $(X'X)^{-1}$, (c) maximum variance function, (d) G-efficiency (in %) versus the number of experiments.

so on. For example, if a solubility of about 18 mg/ml is required, any of the different mixtures in contour 8 (Fig. 5) can be chosen.

In order to test the model lack of fit, the solubility of theophylline was measured at points other than the design points. The observed values and the values of the response predicted by the model at these points are shown in Table 4. As can be seen in this table, the experimental values are in good agreement with those predicted by the model.

CONCLUSIONS

The results of this study suggest that the application of the mixture experimental design, in combination with the *a priori* criteria and the exchange algorithm, enables one to define an experimental design with a reasonable number of experiments and to verify, before carrying out any experiment, if the design contains enough information for the postulated model. Moreover, the use of a pondering function allows the selection of a design that is optimal with regard to the number of experiments, the information provided about the response behavior, and the economic point of view. The experiments can also be pondered on the basis of difficulty and/or availability of raw materials.

The experimental work is sometimes hampered by technological and economic problems and, most impor-

Table 3

Results of Multiple Regression Analysis

Exp.	Response (Y) Experimental[a]	Response (Y) Predicted[b]	Regression Coefficient Values of Eq. (13)
1	15.78	15.72	
2	17.41	17.48	
3	14.37	14.37	b_1 (Z_1) = 15.72
4	9.91	9.88	b_2 (Z_2) = 17.48
5	16.65	16.51	b_3 (Z_3) = 14.37
6	17.05	17.12	b_4 (Z_4) = 9.88
7	18.27	18.34	b_{12} (Z_1Z_2) = -0.35
8	14.49	14.74	b_{13} (Z_1Z_3) = 8.29
9	13.87	13.98	b_{14} (Z_1Z_4) = 22.15
10	17.56	17.74	b_{23} (Z_2Z_3) = -4.74
11	16.00	16.23	b_{24} (Z_2Z_4) = 1.21
17	15.60	15.39	b_{34} (Z_3Z_4) = 22.46
19	15.27	15.22	b_{123} ($Z_1Z_2Z_3$) = -88.09
21	18.00	17.86	b_{124} ($Z_1Z_2Z_4$) = 28.98
22	16.35	16.36	b_{134} ($Z_1Z_3Z_4$) = 62.09
23	13.18	12.95	b_{234} ($Z_2Z_3Z_4$) = 45.29
24	18.22	17.99	
25	21.73	21.50	
26	17.92	17.69	
31[c]	17.92	18.18	
31[c]	18.00	18.18	
31[c]	17.94	18.18	

[a]Mean values of three replicates.
[b]Values calculated by Eq. (1).
[c]Replicate of point 31.

Table 4

Predicted and Experimental Values of Theophylline Solubility in Some Randomly Selected Points of the Experimental Domain. X_1 : See Fig. 1

Exp.	X_1	X_2	X_3	X_4	Y pred. (mg/ml)	Y exp.[a] (mg/ml)
1	0.100	0.350	0.100	0.450	16.86	16.98
2	0.450	0.350	0.100	0.100	16.66	16.72
3	0.100	0.350	0.450	0.100	15.67	15.64
4	0.100	0.233	0.333	0.333	18.27	18.35
5	0.333	0.233	0.100	0.333	18.81	18.70
6	0.333	0.233	0.333	0.100	16.12	16.07
7	0.275	0.175	0.275	0.275	18.95	18.76
8	0.625	0.125	0.125	0.125	17.86	17.94
9	0.125	0.625	0.125	0.125	16.49	16.69
10	0.125	0.125	0.625	0.125	16.92	17.05
11	0.125	0.125	0.125	0.625	16.87	16.95
12	0.300	0.130	0.270	0.300	19.56	19.45

[a]Mean values of three replicates.

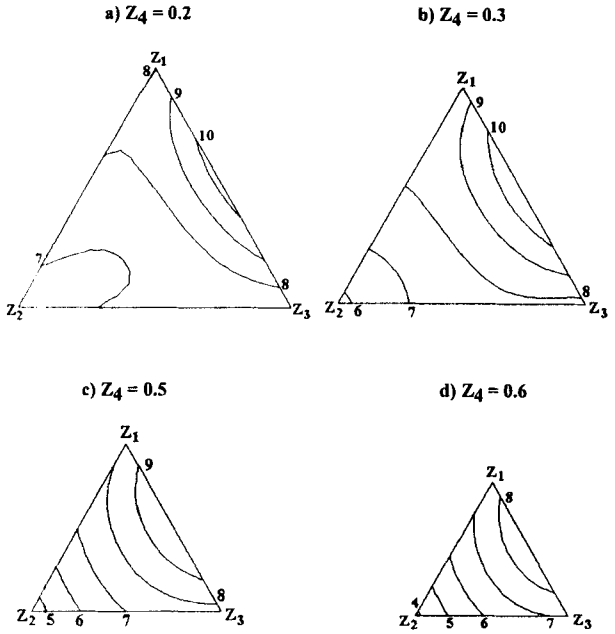

Figure 5. Contour diagrams of theophylline solubility as a function of mixture composition. Each contour represents a constant solubility value: contour (4): 13.46; (5): 14.64; (6): 15.02; (7): 17.00; (8): 18.18; (9): 19.37; (10): 20.55 mg/ml. Z_i = pseudocomponents (see Table 2).

tantly, by time restrictions. Therefore, the best experimental strategy is one that leads to a good knowledge of the studied phenomenon with the least expenditure of time and materials.

REFERENCES

1. A. D. Johnson, V. L. Anderson, and G. E. Peck, Pharm. Res., 10, 1092 (1990).
2. M. Hirata, K. Takayama, and T. Nagaj, Chem. Pharm. Bull., 40, 741 (1992).
3. A. Kristl, D. Vojnovic, R. Karba, A. Mrhar, F. Rubessa, and B. Zupani, Int. J. Pharm., 100, 33 (1993).
4. J. S. Chu, G. L. Amidon, N. D. Weiner, and A. H. Goldberg, Pharm. Res., 11, 1401 (1991).
5. D. Vojnovic, M. Moneghini, and F. Rubessa, Drug Dev. Ind. Pharm., 20, 1035 (1994).
6. S. T. Anik and L. Sukumar, J. Pharm. Sci., 70, 897 (1981).
7. M. A. Mustafa, A. M. Molokhia, and M. W. Gouda, J. Pharm. Sci., 70, 1172 (1981).
8. R. J. Bellotto, Jr., A. M. Dean, M. A. Moustafa, A. M. Molokhia, M. W. Gouda, and T. D. Sokoloski, Int. J. Pharm., 23, 195 (1985).
9. G. A. Lewis and M. Chariot, Drug Dev. Ind. Pharm., 17, 1551 (1991).
10. D. Vojnovic, M. Moneghini, and F. Rubessa, Drug Dev. Ind. Pharm., 21, 823 (1995).
11. D. Mathieu and R. Phan-Tan-Luu, NEMROD (New Efficient Methodology for Research Using Optimal Design) software version 2.01, L.P.R.A.I., Université d'Aix-Marseille, 1995.
12. J. A. Cornell, Technometrics, 17, 25 (1975).
13. H. Scheffé, J.R. Statist. Soc., B20, 344 (1957).
14. H. Scheffé, J.R. Statist. Soc., B25, 235 (1963).
15. J. A. Cornell, in *Experiments with Mixtures: Designs, Models and the Analysis of Mixture Data*, John Wiley, New York, 1990.
16. D. Vojnovic, M. Moneghini, and D. Chicco, Drug Dev. Ind. Pharm., 22(9&10), 997 (1996).