CHAPTER 4

CONCLUSIONS

The chemometric procedures such as Principal Component Analysis (PCA), Principal Component Regression (PCR) and Partial Least-Square Regression (PLS) were developed by making use of the basic computer language. They involved – an analysis of food colorants in a mixture sample, the screening test for thalassemia by the flow injection and Bradford protein assay.

PCA is a multivariate statistical technique that can reduce the dimensionality of the data set. In this study, a number of the wavelength variables were reduced and expressed in the term of the first two principal components. Each food colorant can be differentiated by using the positioning in the two dimension space and the results show that PCA groups the data that have the similar patterns. The pattern of the thalassemia blood samples for the FI-grams in the flow injection experiments was also studied. Although the area between the negative and positive thalassemia blood samples can't be absolutely separated by the result from the two dimension score plot, they can be divided into the two zones of the data and it seems that the screening ability can be improved by varying the suitable area of the data and the interval of the time used in the calculation of the slopes.

PCR is the technique that performs a least-square regression of the new coordination variables resulting from the PCA procedure. In the calculation steps, it is necessary that the model with the best predictive ability concerns the minimum PRESS number. PCR have gained the advantage of using the PCA procedure. Only a

few selected new coordinated variables were used in the calculation steps in order to obtain the best predictive results. It was found that though peaks in the spectral data of the mixture of the red food colorants are the overlapping colorants in the yellow food colorants, PCR models can predict the concentration of the validation samples precisely.

PLS is an extension of the multiple linear regression model that enhances the relation between the data variables, dependent and independent. Like PCR, PLS needs to be optimized in order to find the model that obtains the lowest PRESS number. It can be seen that the results in the study show that the number of the suitable factors of the PCR are usually limited by the number of the training samples but the optimum number of the factors used in PLS is free from that limitation. However, the disadvantage of the PLS procedure is that for the more factors involved would take longer period of calculation process.

Both of the PCR and PLS procedures were applied for the determination of the Bradford protein in milk. In this experiment, the numbers of the extra validation samples (Cross-Validation (CV)) were needed in order to find the suitable areas in the SI-grams those providing model which gave the lowest PRESS. With the validation samples, the predictive results show that the areas that obtain the minimum of PRESS model of both the chemometric procedures are the areas containing the pH gradient and the PLS model shows the better predictive results than the PCR model. However, when they were applied to the real milk samples, the results obtained are different from the labeled real values. This could be due that the labeled values were obtained by the Kjeldahl method for total protein which would be different from Bradford

protein. It could also possibly be due to the turbidity data that can be applied for improving the predictive ability with the milk samples in the future work.