CHAPTER 5
CONCLUSSIONS

In the studying of using chemometrics to evaluation signals, it was found that 3 different types of problems can be solved within proper condition.

Evaluation of WF6 value of data reported in the literature was performed by using *k*-means clustering and HCA. In this study, WF6 of healthy, ovarian cancer and cervix cancer cases were used as data. HCA expressed distribution of WF6 values and natural group in the dataset. From dendrogram of HCA reflected that the WF6 values in data of the 2 dataset (normal-cervix cancer and normal-ovarian cancer) can be separated into 2 groups. Number of k which needs to assign to *k*-means clustering was 2, the number of natural groups that found from HCA. Clustering of WF6 values by *k*-means clustering and HCA were done. It was found that some WF6 values were misclassified values. The misclassified WF6 values were considered to be the samples have to recheck the result by other techniques.

Thalassemia screening models were performed by the combination of *k*-means clustering, HCA and LDA of the dataset which reported in the literature. OFT signals of the dataset were adjusted. The working ranges of OFT signals were selected. Various types of OFT signals; OFT, 1 dimension slope, 13 dimension slopes and PCs of OFT; were used in this proposed. Training set samples of LDA were selected from the comparison of results of *k*-means clustering and HCA of 4 types of signals. From 8 models of clustering processes (from 2 methods of clustering of 4 types of OFT), the

cases which were clustered to be the same group with 8 of 8 of the clustering models were selected to used as training set samples. The rest of data when data of training set was leaved out were used as unknown samples. The comparison of LDA models which using 4 difference signals types were performed to predict the unknown. The unknown which classified to positive test or negative test of thalasemia with more than 2/4 of LDA models to be the same group was found. After the unknown cases were identified by chomometrics processes, their results were validated with hospital records. A criterion of the screening was set up by considering of ratio of clustering results from 8 models and their 1 dimension slope. All unknown cases within the boundary of classification can be successfully predicted.

Multivatiate calibration methods obtained PCA, PLS1 and PLS2 have ability to assay tatrazine, ponceau 4 R and indigo carmine in mixtures. The three colorant compounds can be determined without need of chemical separation. The spectra of mixtures of tattrazine, ponceau 4 R and indigo carmine can be used directly to input to the models and get the predicted concentration value of the three compounds. The multivariate calibration methods were demonstrated on limitation of dataset. Training set of each model were selected by use the criteria of concentration range of colorants and ratio of concentration of colorants. The models which obtained highest quality when validation processes were done were selected. Furthermore, it was found that when comparison of calibration algorithm of PCR, PLS1 and PLS2 shown no significantly difference. Time consumption of the calibration model was compared. Overall processes when PLS2 model was used were less than other model. Range of UV-VIS spectra and

interval of the spectra also studied to increase quality of model. The best condition of each parameters to calculate multivariate calibration model were 400-700 nm of spectra, 100 nm^{-1} interval of spectrum, colorants concentration ranges (4-20 ppm of T, P and I) and colorants concentration ratios (0.03-0.63, 0.3-4.0,0.3-4.0 and 0.3-3.0 for [T]/[P]/[I], [T]/[P] ,[T]/[I] ,[P]/[I], respectively). From validations, the spectrum of sample which obtained within working range with error ranges of prediction were ±1, ±1, and ±2 for T, P and I, respectively.

From the studying of multivariate classification, signals evaluation and classification by using chemometrics methods, the models can work well within the boundary of modeling. The chemometrics methods were applied to deal with complicated signals and unclear data and have probability to use in other problem in the future.