CHAPTER 1
INTRODUCTION TO CHEMOMETRICS METHODS FOR DIAGNOSIS AND CALIBRATION

1.1 Background

Chemometrics is a knowledge which combines knowledge of mathematics, statistics, and logic to use for different aims. Chemometrics can be used for experimental design, investigation the relationship of interesting factor (e.g. group or concentration of analyte) and chemical data, and conservation of chemical system [1]. Chemometrics have more role on analytical chemistry field in this recent years [1, 2] . Some chemical problems need aid of chemometrics process to maximize quality of the model for answering those problems.

1.2 Important of chemometrics in some area of analytical chemistry

The conceptions to improve chemical analysis system is developed uncomplicated analysis systems which used to answer chemical problems with using low time consumption and generate low amount of wastes to provide good response [3]. Some problems need many analysis steps to answering or some problems give the unclear results. Chemometrics methods were used as choice to manage those situations. Interesting researches fields need chemometrics as the tool to gain the better results. In this thesis various chemometrics techniques were used for 2 interesting fields, namely, diagnostic and calibration.

1.2.1 Diagnosis

Disease Diagnosis need the performance of the fast correctly analysis method. Important idea for the diagnosis finding the specific variables which can be clearly separate healthy and unhealthy group apart. Diagnosis systems were developed based on analysis of selective chemical. Specific chemical can be used for increase efficiency of the diagnosis but it also increases cost of the analysis too. Many researches attempt to study alternative assay to generate response that can used to replace the expensive specific chemical in disease screening to reduce analysis cost. Some alternative methods were successfully used instead of the standard method but some methods just used for the screening with less sensitivity. Chemometrics combined with the alternative methods to increase sensitivity of many diagnosis methods. Various pattern recognition methods were used in this aim to increase confident of the screening. Natural of data distribution and group of analyte can be expressed by using unsupervised pattern methods such as Hierarchical Cluster Analysis (HCA) [4], self-organizing map [5] and Principal Component Analysis (PCA) [6, 7]. Various supervised pattern recognition like Support Vector Machine (SVM) [5] One vs. All (OVA) [5], Artificial Neuron Network (ANN) [6, 7], and logistic regression [7] were performed as diagnosis some diseases such as sleep apnoea [8], diabetes [9, 10], asthma [10, 11], cancer [5-7, 12-15] and thalassemia [4, 16].

1.2.2 Calibration

Applications of calibration of some analytes were hard and take time consumption because the analytes were contained in the mixture that make their analysis results were complicated, highly overlapped signals and cannot directly determined. Several methods were presented to separate each analyte from the mixture before determining. Overall processes from the chemical separation and analysis decreased sample though put and increased unwanted wastes. Multivariate calibration methods were popularly used in this field such as Principal Component Regression (PCR) [17, 18] and Partial Least Square (PLS) [17-21].

1.3 Research aim

Various chemometrics methods were used to evaluation the relationships of analytical signals and considering factors for the meaning of calibration work and screening work. This research involved clustering both unsupervised and supervised techniques. Three analytical problems were studied: unbiased clustering analysis of separating cervix cancer patients and ovarian cancer patients from normal persons using protein contents of specific proteoglycan obtained from a flow injection system; combination of various pattern recognition for thalassemia screening from flow system for Osmotic Fragility Test (OFT) for blood test in screening for thalassemia clustering and simultaneous calibration of food colorants from

spectrophotometic spectra of ternary mixtures by using multivariate calibration methods.

Multicomponent calibration methods, i.e. obtained Principal Component Regression (PCR), Partial Least Square 1 (PLS1) and Partial Least Square 2 (PLS2) were applied to determination of tartrazine, ponceau 4 R and indigo carmine in synthetic mixtures.

Hierarchical clustering analysis (HCA) and *k*-means clustering were used for evaluation of relative amount of protein content in specific proteoglycan per 100 mg total protein value as reported in reference [22]. Term of relative amount of protein content in specific proteoglycan per 100 mg total protein used for screening some kinds of cancers. In this thesis protein content in specific proteoglycan per 100 mg total protein of cervix cancer patient, ovarian cancer and healthy patients were selected to be used as data for the clustering. HCA and *k*-means clustering were used to find natural group of the value in the dataset. Discussions of correctly and incorrectly classification results of dataset of relative amount of protein content in specific proteoglycan per 100 mg total protein values of cervix cancer-healthy person and ovarian cancer-healthy person by the 2 clustering methods were used to explain confidence of the diagnostic.

The combination of *k*-means clustering, HCA and LDA were applied for the classification of unknown cases for thalassemia screening. Osmotic fragility test or OFT signals were used as signals for thalassemia screening and reported in 23. The dataset of OFT signals from reference 23 were used as dataset of this study. In this

study, confidence of the classification was increased by considering of various kinds of OFT signals.

1.4 References

1 D. L. Massart, B. G. M. Vandeginste, L. M. C. Buydens, S. d. Jong, P. J. Lewi, J. Smeyers-Verbeke, *Handbook of Chemometrics and Qualimetrics Parts A and B*. Editor, Elsevier, Amsterdam, The Netherlands, 1998.

2 M. Esteban, C. Ariño, J. M. Díaz-Cruz, *TrAC Trends in Analytical Chemistry* 2006, 25, 86-92.

3 G. Kate, *Talanta* 2004, 64, 1084-1090.

4 K.-Z. Liu, K. S. Tsang, C. K. Li, R. A. Shaw, H. H. Mantsch1, *Clinical Chemistry* 2003, 49, 1125-1132.

5 S. Ramaswamy, P. Tamayo, R. Rifkin, S. Mukherjee, C.-H. Yeang, M. Angelo, C. Ladd, . Reich, E. Latulippe, J. P. Mesirov, T. Poggio, W. Gerald, M. Loda, E. S. Lander, T. R. Golub, *Proceedings of the National Academy of Sciences* 2001, 98, 15149-15154.

6 Z. Zhang, H. Zhuo, S. Liu, P. d. B. Harrington, *Analytica Chimica Acta* 2001, 436, 281-291.

7 E. A. Hernández-Caraballo, F. Rivas, A. G. Pérez, L. M. Marcó-Parra, *Analytica Chimica Acta* 2005, 533, 161-168.

8 J. V. Marcos, R. Hornero, D. Álvarez, F. del Campo, C. Zamarrón, *Medical Engineering & Physics* 2009, 31, 971-978.

9 J. Yang, G. Xu, Q. Hong, H. M. Liebich, K. Lutz, R. M. Schmülling, H. G. Wahl, *Journal of Chromatography B* 2004, 813, 53-58.

10 R. Madsen, T. Lundstedt, J. Trygg, *Analytica Chimica Acta* 2010, 659, 23-33

11 S. Carraro, S. Rezzi, F. Reniero, K. Héberger, G. Giordano, S. Zanconato, C. Guillou, E. Baraldi, *American Journal of Respiratory and Critical Care Medicine* 2007, 175, 986-990.

12 M. Zhang, P. Tong, W. Wang, J. Geng, Y. Du, *Chemometrics and Intelligent Laboratory Systems* 2011, 105, 207-214.

13 V. Mrázová, J. Mocák, E. Varmusová, D. Kavková, A. Bednárová, *Journal of Pharmaceutical and Biomedical Analysis* 2009, 50, 210-215.

14 T. Imre, T. Kremmer, K. Héberger, É. Molnár-Szöllősi, K. Ludányi, G. Pócsfalvi, A. Malorni, L. Drahos, K. Vékey, *Journal of Proteomics* 2008, 71, 186-197.

15 A. Devos, L. Lukas, J. A. K. Suykens, L. Vanhamme, A. R. Tate, F. A. Howe, C. Majós, A. Moreno-Torres, M. van der Graaf, C. Arús, S. Van Huffel, *Journal of Magnetic Resonance* 2004, 170, 164-175.

16 M. Arjmand, M. Kompany-Zareh, M. Vasighi, N. Parvizzadeh, Z. Zamani, F. Nazgooei, *Talanta* 2010, 81, 1229-1236.

17 J. J. B. Nevado, J. R. Flores, G. C. Peñalvo, *Analytica Chimica Acta* 1997, 340, 257-265.

18 J. J. Berzas-Nevado, C. Guiberteau-Cabanillas, A. M. Contneto-Salcedo, R. Martin-Villamuelas, *Anal. Lett.* 1999, 32, 1879-1898.

19 G. K. Karaoglan, G. Gumrukcu, M. U. Ozgur, A. Bozdogan, B. Asci, *Anal. Lett.* 2007, 40, 1893-1903.

20 H. Wold, H. Martens, *Multivariate Calibration Problems in Chemistry Solved by PLS*. Editor, A. Ruhe, B. Kagstrom (Eds.), Heidelberg, 1983.

21 D. M. Haaland, E. V. Thomas, *Analytical Chemistry* 1988, 60, 1193-1202.

22 S. K. Hartwell, K. Pathanon, D. Fongmoon, P. Kongtawelert, K. Grudpan, *Anal Bioanal Chem* 2007, 388, 1839-1846.

23 S. Khonyoung, S. K. Hartwell, J. Jakmunee, S. Lapanantnoppakhun, T. Sanguansermsri, K. Grudpan, *analytical sciences* 2009, 25, 819-824.