

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	iii
ABSTRACT (ENGLISH)	v
ABSTRACT (THAI)	vi
LIST OF TABLES	xii
LIST OF FIGURES	xiv
ABBREVIATIONS AND SYMBOLS	xvi
 CHAPTER 1 INTRODUCTION TO CHEMOMETRICS METHODS FOR DIAGNOSIS AND CALIBRATION	
1.1 Background	1
1.2 Important of chemometrics in some area of analytical chemistry	1
1.2.1 Diagnosis	2
1.2.2 Calibration	3
1.3 Research aim	3
1.4 References	5

CHAPTER 2 CLUSTERING ANALYSIS WITH UNBIASED METHODS TO DIFFERENTIATE CERVIX CANCER PATIENTS AND OVARIAN CANCER PATIENTS FROM NORMAL PERSONS USING PROTEIN CONTENT OF SPECIFIC PROTEOGLYCAN OBTAINED FROM FLOW INJECTION SYSTEM

2.1 Introduction	8
2.2 Chemometrics methodology	9
2.2.1 Box plot	9
2.2.2 Clustering methods	9
2.2.2.1 Hierarchical Clustering Analysis	9
2.3.1.2 <i>K</i> -means clustering	10
2.3 Experimental	10
2.3.1 The data	11
2.3.2 Apparatus	13
2.3.3 Chemometrics processing	14
2.3.3.1 Data observation	14
2.3.3.2 Samples groups identification by <i>k</i> -means and HCA clustering	14
2.4 Results and discussion	15
2.4.1 Dataset I: the cervix cancer-normal cases	15
2.4.1.1 Data observation	15
2.4.1.2 Samples groups identification by <i>k</i> -means clustering and HCA clustering	15

2.4.2 Dataset II: the ovarian cancer-normal cases	19
2.4.2.1 Data observation	19
2.4.2.2 Samples groups identification by <i>k</i> -means clustering and HCA clustering	19
2.5 Conclusion	22
2.6 References	22
 CHAPTER 3 COMBINATION OF HCA, K-MEANS CLUSTERING AND LDA FOR EVALUATION OF SIGNALS FROM OFT-ANALYZER TO THE PREDICTION OF PATIENT GROUP OF THALASSEMIA SCREENING	
3.1 Introduction	25
3.2 Chemometrics methodology	27
3.2.1 Extraction and distribution of data: Principal Component Analysis (PCA)	27
3.2.2 Clustering using Hierarchical Cluster Analysis (HCA) and <i>k</i> -means clustering	28
3.2.2.1 Hierarchical cluster analysis (HCA)	28
3.2.2.2 <i>K</i> -means clustering	29
3.2.3 Identification of group of signals of sample: Linear discriminate analysis (LDA)	29
3.3 Experimental	30
3.3.1 Osmotic Fragility Test (OFT) of blood samples	30
3.3.2 Apparatus	31
3.3.3 Chemometric methods and data analysis	31
3.3.4 Data pretreatment	32

3.3.5 Classification methods: <i>k</i> -means clustering and hierarchical clustering analysis (HCA)	34
3.3.6 LDA model for screening and visualization using principle component analysis	34
3.3.7 LDA training set selection and construction of LDA models	35
3.4 Results and discussions	35
3.4.1 Grouping of blood samples by hierarchical cluster analysis (HCA) and <i>k</i> -means clustering	35
3.4.2 LDA training set selection and construction of LDA models	42
3.5 References	51

CHAPTER 4 SIMULTANEOUS ASSAY OF EACH SUBSTANCE IN TERNARY MIXTURE OF FOOD COLORANTS VIA SPECTROMETRY

4.1 Introduction	57
4.2 Experimental	59
4.2.1 Apparatus	59
4.2.2 Reagents	59
4.2.3 Experimental procedure	59
4.2.4 Statistical techniques	61
4.2.4.1 Multivariate calibration techniques	61
4.2.4.2 Validation statistics terms	62
4.2.5 Data management	63
4.2.6 Optimization of model quality	64
4.2.6.1 Multivariate calibration method effect	64
4.2.6.2 Training set characteristic effect	65

4.2.6.2.1 Study of training set number effect when colorants concentration ranges was controlled	65
4.2.6.2.2 Study on training set number effect when colorants concentration ranges and colorants concentration ratios were controlled	69
4.2.6.3 Signal characteristics effect	69
4.3 Results and discussion	70
4.3.1 Multivariate calibration method effect	70
4.3.2 Training set characteristic effect	73
4.3.2.1 Study of training set number effect when colorants concentration ranges was controlled	73
4.3.2.2 Study on training set number effect when colorants Concentration ranges and colorants concentration ratios were controlled	78
4.3.3 Signal characteristics effect	80
4.4 Conclusions	83
4.5 References	84
CHAPTER 5 CONCLUSSIONS	86
CURRICULUM VITAE	89
THE RELEVANCE OF THE RESEARCH WORK TO THAILAND	92

LIST OF TABLES

Table	Page
2.1 Clustering results of database I from <i>k</i> -means clustering and HCA and confirmation of samples by comparison of both methods	17
2.2 Summary of clustering of WF6 values in dataset I by <i>k</i> -means clustering	18
2.3 Summary of clustering of WF6 values in dataset I by HCA	18
2.4 Clustering result of database II from <i>k</i> -means clustering and HCA and confirmation of samples by comparison of both methods	21
2.5 Summary of clustering of WF6 values in dataset II by HCA and <i>k</i> -means clustering	22
3.1 Clustering analysis of blood samples	40
3.2 Member in each group of clustering	42
3.3 Identification of blood samples by LDA models	43
3.4 LDA canonical function and characteristic of the LDA models	50
4.1 Composition of T, P and I for training set selection (colorants concentration ranges was controlled)	67
4.2 Composition of T, P and I for training set selection (colorants concentration ranges and colorants concentration ratios were controlled)	68
4.3 Composition of T, P and I of validation set	69
4.4 Quality of PCR PLS1 and PLS2 when using training set 01-07	75

4.5 Unknown prediction of PCR, PLS1 and PLS2 when using training set 01-07	77
4.6 Quality of PLS2 when using training set 0A-0E	78
4.7 Unknown prediction of PLS2 when using training set 0A-0E	81
4.8 Quality of PLS2 of training set 0D from varying of spectrum range and interval	82
4.9 Unknown prediction of PLS2 of training set 0D from varying of spectrum range and interval	83

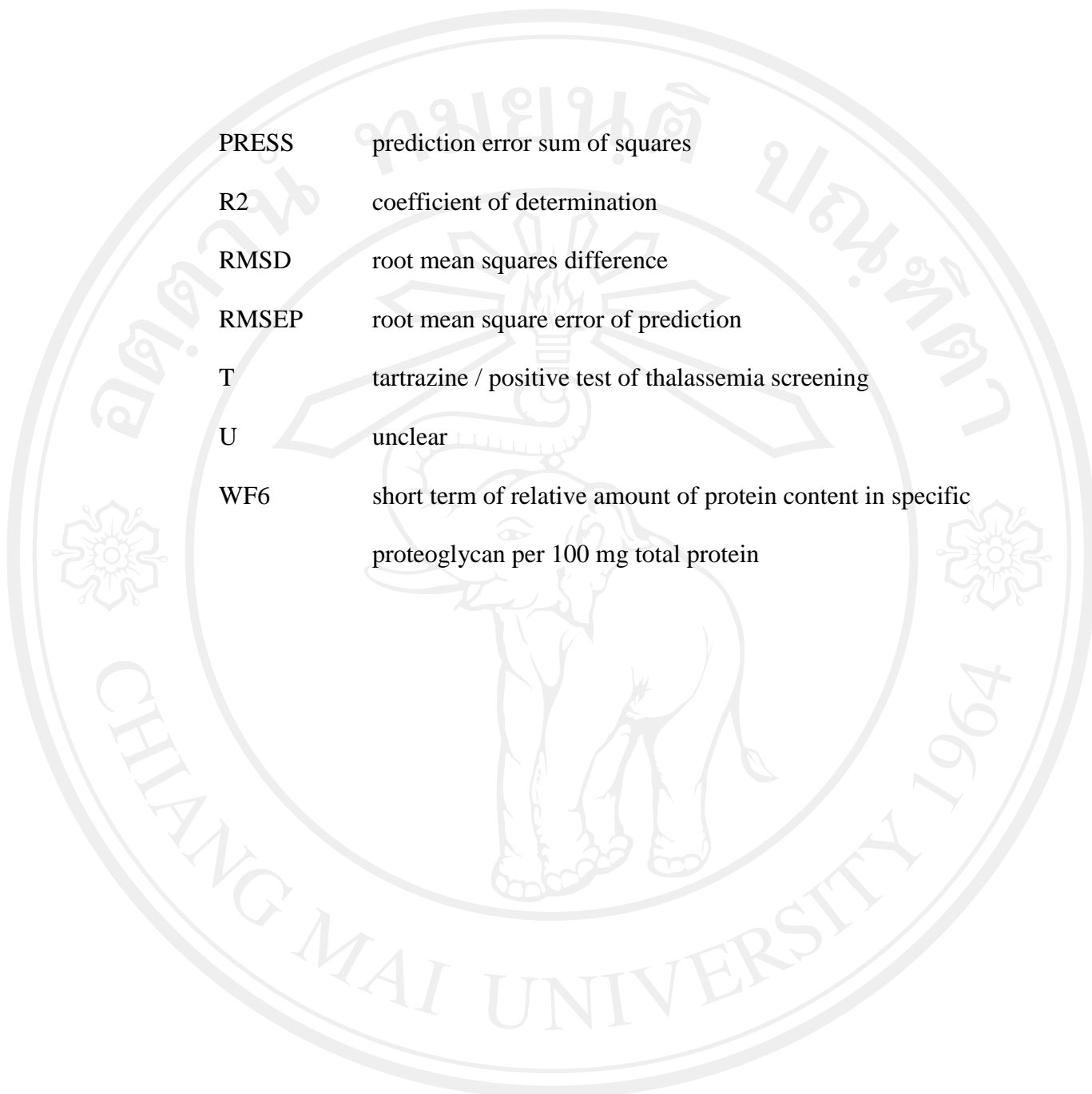
LIST OF FIGURES

Figure	Page
2.1 Chemometrics process of clustering of WF6	11
2.2 Relative amount of WF6-specific protein per 100mg total protein \pm SD of normal-cervix cancer cases (database I) and normal-ovarian cancer cases (database II)	13
2.3 Box test of database I. normal-cervix cancer	15
2.4 Dendrogram of database I (N was normal and CC was cervix cancer)	16
2.5 Box test of database II: normal-ovarian cancer	19
2.6 Dendrogram of database II (N was normal and OC was ovarian cancer)	20
3.1 Raw OFT signals of 73 cases obtained 21 positive test of thalassemia and 52 negative test of thalassemia	32
3.2 The selected range of OFT signals of blood samples	33
3.3 Dendrograms of blood samples; OFT, PCs of OFT, 1 dimension slope and 13 dimension slopes	38
3.4 Distribution of blood samples defined by groups of LDA models with ratio of prediction; OFT, PCs of OFT, 1 dimension slope, and 13 dimension slopes	45
3.5 Unreliable signals different from database	47
3.6 Overview of thalassemia screening by aid of clustering methods	49
4.1 Absorption spectra for solutions of 10 ppm T (1), 10 ppm P (2), 10 ppm I (3), and their mixtures of 10 ppm of T, 10 ppm of P, and 10 ppm of I (4), prepared in 10 ml of acetate buffer solution (pH 5.5) and deionized water (milli Q quality) recorded against a blank of deionized water	60

4.2 The number of factors and RMSD of PCR model, PLS1 model and PLS2 model for the resolution of the ternary mixture: T, P, and I of training set 07	71
4.3 RMSEP of T,P and I of PCR, PLS1 and PLS2 model when using different training set (training set 01-07)	72
4.4 RMSEP of T, P and I of PLS2 model when using training set 01-07	75
4.5 RMSEP of T, P and I of PLS2 model when using training set 0A-0E	79

ABBREVIATIONS AND SYMBOLS

CC	cervix cancer
HCA	hierarchical cluster analysis
I	indigo carmine
LDA	linear discriminant analysis
MT	positive test of thalassemia predicted by LDA
MN	negative test of thalassemia predicted by LDA
MU	unidentified case of thalassemia predicted by LDA
N	normal or negative test
OC	ovarian cancer
OFT	osmotic fragility test
P	ponceau 4 R
PCA	principal component analysis
PCR	principal component regression
PT	positive test of thalassemia clustered by <i>k</i> -means clustering and HCA
PN	negative test of thalassemia clustered by <i>k</i> -means clustering and HCA
PU	unidentified case of thalassemia clustered by <i>k</i> -means clustering and HCA
PLS 1	partial least square 1
PLS 2	partial least square 2

The background of the page features a large, faint watermark of the Chiang Mai University seal. The seal is circular, containing an elephant in the center, with Thai script and the English text 'CHIANG MAI UNIVERSITY 1964' around the perimeter.

PRESS	prediction error sum of squares
R ²	coefficient of determination
RMSD	root mean squares difference
RMSEP	root mean square error of prediction
T	tartrazine / positive test of thalassemia screening
U	unclear
WF6	short term of relative amount of protein content in specific proteoglycan per 100 mg total protein