생활과학박사학위논문

Adaptive Neuro-Fuzzy Inference System-Applied QSAR for Antioxidant Activities of Food Phytochemicals

식품 유래 파이토케미컬의 항산화활성에 대한 적응형 뉴로-퍼지 추론시스템 적용 QSAR

2016년 8월

서울대학교 대학원

식품영양학과

진 창 호

Adaptive Neuro-Fuzzy Inference System-Applied QSAR for Antioxidant Activities of Food Phytochemicals

Advisor: Keum Taek Hwang, Ph. D

By

Changho Jhin

2016

Department of Food and Nutrition

Graduate School

Seoul National University

Abstract

Adaptive Neuro-Fuzzy Inference System-Applied QSAR for Antioxidant Activities of Food Phytochemicals

Department of Food and Nutrition

Seoul National University

Changho Jhin

Antioxidant is a chemical substance that significantly delays, reduces or prevents target molecules from oxidative stress-induced damages. One of the characteristics of phytochemicals is antioxidant activity. Dietary phytochemicals can contribute to human antioxidant system for scavenging free radicals. In addition to their physiological properties, phytochemical antioxidants affect stability, safety and quality of foods. Antioxidant activities of phytochemicals were evaluated in numerous studies. Although the structural characteristics of phytochemicals strongly affect the antioxidant activities, most of previous studies did not rationally explain the relationship between chemical structures and antioxidant activities in a physicochemical

manner. Therefore, structure-related analyses are needed to better understand and predict the antioxidant activities of phytochemicals. Quantitative structure-activity relationship (QSAR) is a methodology that quantitatively correlates chemical structural descriptors of molecules and their biological activities. Selecting chemical structural descriptors and modelling methods are important for developing reliable QSAR models. Hydrophobic constant, numerical structural variables, quantum chemical descriptors and 3D molecular interaction fields can be used as structural descriptors for QSAR. Among them, quantum chemical descriptors can explain chemical reactions such as redox reactions. Selecting QSAR modelling method is also important for QSAR studies. Common QSAR modelling methods are linear regression and partial least square methods. Recently, as an advanced modelling method, machine learning methods are applied in chemometric studies. Nonlinear relationship between variables can be fitted by optimisation process of machine learning methods. Adaptive neuro-fuzzy inference system (ANFIS), one of machine learning methods, is a QSAR modelling method for fitting nonlinear relationship with high accuracy.

The objective of this study was to develop accurate predictive QSAR models for antioxidant activities of prominent phytochemicals indigenous to foods, using quantum chemical descriptors. This study is composed of three parts. ANFIS-applied prediction models for antioxidant activities of anthocyanins (including anthocyanidins) and carotenoids were developed in the first and second parts of this study. In the third part, QSAR models for the

antioxidant activities of various phenolic compounds were developed, and the prediction efficiencies of linear regression and ANFIS-applied QSAR models were compared.

In the first part of the study, since anthocyanins have various chemical structures depending on pH of their solutions, quantum chemical descriptors of flavylium cation, quinoidal base, carbinol pseudo-base and chalcone structures of anthocyanins were calculated. Quantum chemical descriptors were calculated by semi-empirical PM6 and PM7 methods. Correlation analysis between calculated quantum chemical descriptors and antioxidant activities was performed to select descriptors for QSAR. Electron affinity and electronegativity of flavylium cation, and ionisation potential of quinoidal base were significantly correlated with radical scavenging activities of anthocyanins, and these descriptors were selected as independent variables for QSAR models. ANFIS-applied QSAR models were developed by using the selected quantum chemical descriptors. The developed ANFIS models had two triangular-shaped input fuzzy functions for each independent variable, and the models were optimised by backpropagation method. The constructed models using descriptors calculated by both PM6 and PM7 had good prediction efficiencies with Q square of 0.819 and 0.862, respectively.

In the second part of the study, quantum chemical descriptors of neutral and monovalent cationic carotenoids were calculated by semi-empirical PM6 and PM7 methods. Cross-product terms between quantum chemical descriptors of neutral and cationic carotenoids were calculated. The relationship between

quantum chemical descriptors of carotenoids and their radical scavenging activities was determined by correlation analysis. Ionisation energies of neutral and monovalent cationic carotenoids and the product of chemical potentials of neutral and monovalent cationic carotenoids were significantly correlated with their radical scavenging activities, and consequently these descriptors were used as independent variables of QSAR models. ANFIS-applied QSAR models were also developed for predicting and comparing radical scavenging activities of carotenoids. ANFIS-applied QSAR models were developed with two triangular-shaped input membership functions made for each of the independent variables and optimised by backpropagation method. High prediction efficiencies were achieved by the ANFIS-applied QSAR. R square values of the developed QSAR models with the variables calculated by PM6 and PM7 methods were 0.921 and 0.902, respectively. The results of this study demonstrated reliabilities of the selected quantum chemical descriptors and significance of QSAR models.

In the third part of the study with phenolic compounds, bond dissociation energy of O-H bond (BDE) was calculated by semi-empirical PM6 and PM7 quantum chemical methods. Previous studies reported negative correlations between the lowest BDE (BDE_{min}) and antioxidant activities. However, not only the lowest BDE but also BDEs of other hydroxyl groups may affect radical scavenging activities of phenolic compounds. In addition, calculated BDEs of phenol and enol groups were lower than those of alcohol groups. Considering these points, sum of reciprocals of BDE of enol and phenol

groups (X_{BDE}) was calculated as a new descriptor for QSAR models. Significant correlations were observed between X_{BDE} and antioxidant activities. X_{BDE} was introduced as a descriptor for developing QSAR models. The QSAR models by both of linear regression and AFNIS achieved high prediction accuracies. Among the developed models, ANFIS-applied models achieved better prediction accuracies than linear regression-applied models. The correlation coefficients of linear regression and ANFIS-applied QSAR models were 0.907 and 0.919, respectively. From these results, the proposed descriptor of X_{BDE} was confirmed to be an appropriate variable for predicting and analysing antioxidant activities of phenolic compounds. Also, the ANFIS could be applied to QSAR models to improve prediction accuracy.

In this study, appropriate chemical structural descriptors of food phytochemicals for QSAR models were calculated by rational processes. And accurate prediction models for antioxidant activities of the phytochemicals were developed by ANFIS-applied QSAR. The developed ANFIS-applied QSAR models could be used for establishing rational strategies for developing plant-derived functional foods and bio-materials with high antioxidant activities.

Key words: phytochemical, antioxidant activity, radical scavenging activity, QSAR, ANFIS.

Student number: 2011-23690

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ABTS: 2,2'-azino-bis(3-ethylbenzothiazoline-6-sulphonic acid)

ANFIS: adaptive neuro-fuzzy inference system

ANN: artificial neural network

BDE: bond dissociation energy

CADD: computer-aided drug design

DPPH: 2, 2-diphenyl-1-picryl-hydrazyl-hydrate

FIS: fuzzy inference system

HAT: hydrogen atom transfer

HOMO: highest occupied molecular orbital

IE: ionisation energy

LOOCV: leave-one-out cross-validation

LUMO: lowest unoccupied molecular orbital

MAE: mean absolute error

MLR: multiple linear regression

ORAC: oxygen radical absorbance capacity

QSAR: quantitative structure-activity relationship

RNS: reactive nitrogen species

ROS: reactive oxygen species

SAR: structure-activity relationship

SET: single electron transfer

SET-PT: single electron transfer-proton transfer

SLR: simple linear regression

SOMO: singly occupied molecular orbital

SPLET: sequential proton-loss electron-transfer

SVM: support vector machine

TEAC: trolox equivalent antioxidant capacity

Chapter 1

Introduction

1.1. Literature review

1.1.1. Antioxidant

Antioxidants are any chemical compounds that significantly delay, reduce or prevent target molecules from damage induced by oxidative stress (Halliwell, 2007). Oxidative stress can be occurred by increased level of free radicals with low antioxidant activity. Free radicals, such as reactive oxygen species (ROS) and reactive nitrogen species (RNS), are generated by intracellular mitochondrial activity (Cadenas and Davies, 2000) or external stress such as ultraviolet ray (Jackson, 1999; Logani and Davies, 1980). Antioxidants can be used as food preservatives for preventing food products from undesirable changes which occurred by oxidation. Lipid oxidation declines the quality of lipid-containing food complex by giving rancid odours of oxidation products (Gray, 1978). The lipid oxidation is inhibited by α-tocopherol, ascorbic palmitate and other antioxidants (Frankel, 1996). Del Carlo et al. (2004) reported a significant positive correlation between phenolic antioxidant content and oxidative stability of extra virgin olive oil. Frankel et al. (1996) reported that carnosic acid, carnosol and rosmarinic acid show high antioxidant activities on corn oil and oil-in-water emulsion. Antioxidants have also effects on physiological system. Small amount radicals of reactive oxygen and nitrogen species are required for normal physiological functions such as signal transduction (Valko et al., 2007). However, due to high reactivity of free radicals, physiological biomolecules can be damaged by

oxidative stresses induced by free radicals. ROS can cause DNA damaged, and the DNA damage has been postulated to promote carcinogenesis, coronary heart diseases and advancing age-related diseases (Cooke et al., 2003; Vaca et al., 1988). Also, free radicals induce damage to lipid, low-density lipoprotein, protein and nucleic acid, which are responsible for human diseases such as hypertension, atherosclerosis, cardiovascular diseases, Alzheimer diseases and cancer (Chen et al., 2012; Crichton et al., 2013; Förstermann, 2008). Thus, scavenging free radicals is important for preventing and treating diseases and health problems. Human antioxidant system is constituted of endogenous and exogenous components, which are classified into antioxidant enzymes, metal binding proteins and antioxidants (Dasgupta and Klein, 2014). Among them, antioxidant enzymes and metal binding proteins are mostly classified as endogenous components, while antioxidants are originated from both endogenous and exogenous ones. Epidemiological studies have been reported that consumption of fruits and vegetables, which are good sources of natural antioxidants, lowers risk and mortality of age-related chronic diseases (Asplund, 2002; Crichton et al., 2013; Podsędek, 2007; Rimm et al., 1996).

The antioxidant activities are mainly contributed by radical scavenging mechanisms of antioxidant molecules. Two main antioxidant mechanisms are hydrogen atom transfer (HAT) and single electron transfer (SET) (Prior et al., 2005). The radical scavenging activity by HAT mechanism is explained as following reaction:

$$X\cdot + AOH \rightarrow XH + AO\cdot$$

where X represents a free radical and AOH represents an antioxidant molecule and its hydroxyl group. As shown on the chemical equation, single hydrogen atom should be homolytic dissociated from the antioxidant molecule to scavenge radical by the HAT mechanism. Thus, the bond dissociation energy of O-H bond (BDE) is a crucial variable for estimating antioxidant activity. While, SET mechanism is explained by following equation:

$$X\cdot + AOH \rightarrow X^{-} + AOH\cdot^{+}$$

The mechanism is governed by ionisation energy (IE) which is the energy needed to dissociate an electron from a molecule. HAT and SET mechanisms can occur simultaneously in antioxidant reaction (Prior et al., 2005). Free radicals can be also scavenged by multi-step mechanisms which are single electron transfer-proton transfer (SET-PT) and sequential proton-loss electron-transfer (SPLET) (Klein and Lukeš, 2007; Musialik et al., 2009; Vagánek et al., 2012). The SET-PT mechanism occurs as following two-step reaction:

$$X\cdot + AOH \rightarrow X^{-} + AOH\cdot^{+}$$

$$X^{-} + AOH\cdot+ \rightarrow XH + AO\cdot$$

And the SPLET mechanism is explained by following equations:

$$X\cdot + AOH \rightarrow X\cdot + H^{+} + AO^{-}$$

$$X\cdot + H^{+} + AO^{-} \rightarrow X^{-} + H^{+}+ AO\cdot$$

$$X^- + H^+ + AO\cdot \rightarrow XH + AO\cdot$$

Basically, the net chemical equations of multi-step antioxidant mechanisms are the same as that of HAT mechanism. SET occurs in the first step of SET-PT and the second step of SPLET mechanism. Thus, multi-step antioxidant reactions are also governed by BDE and IE values.

The radical scavenging activities can be evaluated by several assay methods. Oxygen radical absorbance capacity (ORAC) assay measures the amount of peroxyl radicals (Cao et al., 1993). The antioxidant power can be calculated by measuring decreased amount of peroxyl radicals, and the peroxy radical scavenging reaction is mainly governed by HAT mechanism (Prior et al., 2005). Stable radical molecules can be used for measuring antioxidant activities. 2,2'-Azino-bis(3-ethylbenzothiazoline-6-sulphonic acid) (ABTS) and 2, 2-diphenyl-1-picryl-hydrazyl-hydrate (DPPH) are stable radicals that can be used for measuring antioxidant power of antioxidant molecules (Osman, 2011; Re et al., 1999; Sharma and Bhat, 2009). ABTS and DPPH assays are quantified by measuring reduced amounts of ABTS and DPPH radicals, respectively, by antioxidant molecules.

1.1.2. Phytochemicals as antioxidants

Phytochemical is a compound originated from the Greek word 'phyto', which means plant, and 'chemical'. Phytochemicals are chemical substances derived from plant sources. One of important characteristics of phytochemicals is their antioxidant activity. As a part of plant defence system, antioxidant activity contributed by phytochemicals controls and defences oxidative stresses (Arora et al., 2002). Phytochemicals can be classified by their common skeleton structures (Figure 1.1), and some common classes of phytochemicals are as follows:

Flavonoids are distributed in various species of fruits, vegetables and traditional Chinese medicinal plants (Cai et al., 2004; Liao et al., 2008). The flavonoids have a common C6-C3-C6 carbon skeleton and are divided into several subclasses such as flavanols, flavonols, flavanones, flavones, isoflavones and so on, depending on their residues and bond orders. Flavonoids have been characterised by their high antioxidant activities. Torel et al. (1986) evaluated peroxy radical-inhibiting activities of flavonoids, suggesting morin is the most effective flavonoid species. Also, flavonoids have been known to retard oxidation caused by RNS (Haenen et al., 1997; Kerry and Rice-Evans, 1999).

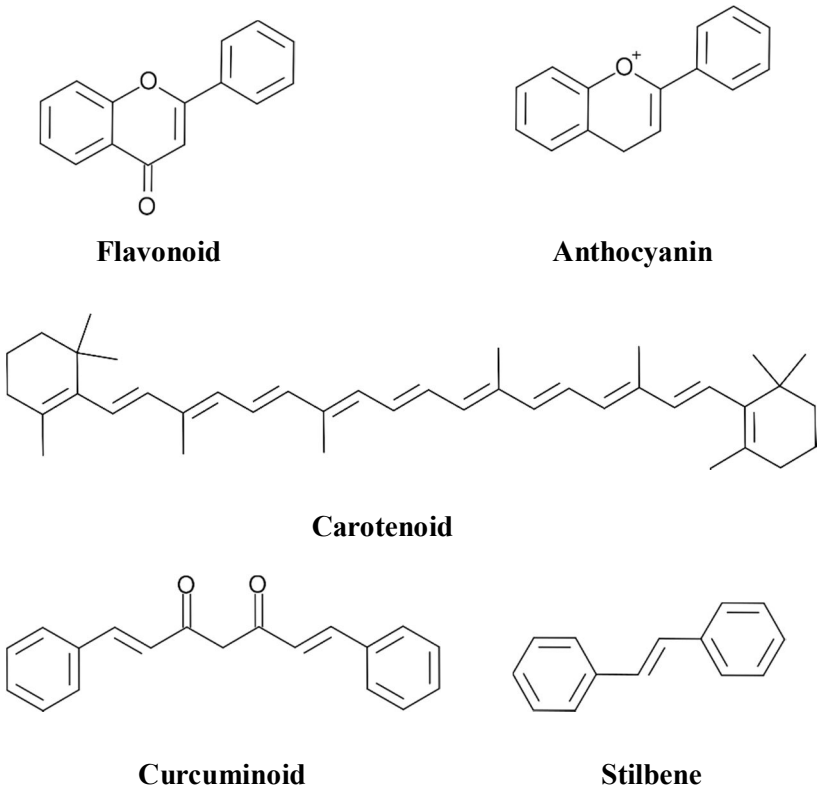

Figure 1.1. Skeleton structures of phytochemicals.

Anthocyanins and their aglycones, anthocyanidins, are classified as subgroups of flavonoids that have a C6-C3-C6 carbon skeleton. Anthocyanins and anthocyanidins are water-soluble colour pigments, which are blue, purple and red depending on pH, and these pigments are mostly distributed in fruits. One of the characteristic features of these compounds is that these compounds contain an oxonium ion in flavylium cation structure, unlike other flavonoids. Another characteristic of anthocyanins and anthocyanidins is that their structures and colours vary depending on pH (Mazza and Brouillard, 1987). For this reason, the antioxidant activity of anthocyanins is strongly affected by pH of the system and their high antioxidant activities were observed on high pH (Borkowski et al., 2005). Also, high antioxidant activities of anthocyanins and anthocyanidins were observed against DPPH radical, methyl linoleate and low-density lipopolysaccharide (Kähkönen and Heinonen, 2003; Tsuda et al., 1996).

Carotenoids are fat-soluble phytochemicals, the most widely distributed colour pigments in natural products (Sanchez et al., 1999). Basic skeleton of carotenoids is a tetraterpenoid structure that contains 40 carbon atoms with unsaturated polyene chains. Carotenoids are classified into two subclasses, carotenes and xanthophylls. Carotenoids are constituted of only carbon and hydrogen atoms, while xanthophylls contain carbon, hydrogen and oxygen atoms. Carotenoids are found ubiquitously in plants and photosynthetic microorganisms since carotenoids defend oxidative stress from photooxidation and photosynthetic electron transfer system (Foyer et al., 1994;

Stahl and Sies, 2005). Antioxidant activities of carotenoids are mainly contributed by their polyene structure, and their antioxidant activities are governed by SET mechanism (Britton, 1995; Cantrell et al., 2003; Krinsky and Deneke, 1982). Antioxidant activities of various carotenoids have been reported, and lycopene is the most effective antioxidant among the tested carotenoids (Miller et al., 1996; Müller et al., 2011).

Other several classes of phytochemicals have antioxidant characteristics. Since antioxidant activities are known to be contributed by phenol group, phenol group containing phytochemicals such as phenolic acids, tannins, curcuminoids and stilbenoids have antioxidant activities (Cai et al., 2006; Hagerman et al., 1998; Müller et al., 2011; Priyadarsini et al., 2003).

1.1.3. Quantitative structure-activity relationship (QSAR)

1.1.3.1. Definition of QSAR

QSAR is a methodology that quantitatively correlates chemical structural descriptors of molecules and their biological activities such as enzyme inhibition constants, affinity constants, antioxidant activities and so on. Thus, a function for biological activity and chemical structural descriptor can be established:

$$\Phi = f(\mathrm{C})$$

where Φ and C represent biological/chemical activity and chemical structural descriptor, respectively. QSAR studies mainly focus on predicting biological/chemical activities, and comprehending and rationalising

biochemical mechanisms of the action.

1.1.3.2. Molecular descriptors

Various chemical structural descriptors can be used in QSAR models. It has been assumed that two independent studies by Hansch and Fujita (1964) and Free and Wilson (1964) are the very first studies of modern QSAR methodologies. Later, QSAR methodologies which were reported by two independent studies called *Hansch analysis* and *Free-Wilson analysis*.

As a physicochemical approach, *Hansch analysis* uses hydrophobic constants (π) of substituents as QSAR descriptors. The constant of π was calculated as following equation:

$$\pi = \log(\frac{P_X}{P_H})$$

where P_H is partition coefficient of a template molecule, and P_X is partition coefficient of a derivative. A two-phase octanol/water system has been generally used as a reference system for evaluating partition coefficient. Since partition coefficient represents hydrophobicity of a chemical compound, molecular interactions can be interpreted by the partition coefficient and derived descriptors. Biological systems such as cells and cell organelles are compartmented by biological lipid membranes of which structures were strongly affected by hydrophobic interactions. Since hydrophobic molecules are easily transported across biological lipid membranes, absorption, distribution, metabolism and excretion (ADME) properties of the molecules are governed by their hydrophobicity (Lipinski, 2002). Thus, the *Hansch*

analysis method have been applied to various *in vitro* and *in vivo* studies for analysing and predicting biochemical activities of chemicals including inhibition constants, affinities and toxicities.

The *Free-Wilson analysis,* first developed by Free and Wilson (1964) and then revised by Fujita and Ban (1971), is a simple and efficient QSAR methodology, which adapts numerical structural variables from chemical structures of compounds. QSAR descriptors for *Free-Wilson analysis* do not need to be calculated or derived by experiments. Descriptors for *Free-Wilson analysis* are generated by comparing the structures of compounds to structure of an arbitrarily chosen reference molecule. Each descriptor has value of 0, indicating absence of a particular substituent or structural feature, or 1, indicating its presence. Since the values of descriptors used in the *Free-Wilson analysis* can be directly recognised from the first look on structures of compounds, it is easy to establish and comprehend QSAR models. Thus, this method is useful for designing molecular structures for de novo synthesis. Despite the advantages of the *Free-Wilson analysis*, it has some disadvantages (Kubinyi, 1988). At least two essential different positions of substitution are needed to develop a QSAR model by the *Free-Wilson* method; therefore, the type of compounds is limited. A large data set is also necessary for developing a QSAR model, while only a small number of new analogues can be predicted by developed models. Another disadvantage is that the mechanism of action cannot be revealed easily since the *Free-Wilson analysis* is not based on physicochemical descriptors.

Rationalising chemical mechanisms of action is one of the main purposes of QSAR study. Biochemical activities of chemical compounds resulted from redox reactions are based on dissociation and formation of covalent or non-covalent bonds; therefore, the activities are closely related to quantum chemical descriptors. For this reason, via QSAR study with quantum mechanics-derived descriptors, the mechanism of action can be understood. Development of computational hardware and efficient software algorithms make it possible to perform quantum chemical calculation on personal computers. QSAR models for biochemical activities of chemical compounds, such as partition coefficient, solubility, vapour pressure and redox properties, were developed using quantum mechanics-derived descriptors in previous studies (Jönsson et al., 1999; Staikova et al., 2004; Yang et al., 2007; Zeng et al., 2007). The quantum mechanics-derived descriptors used in QSAR analysis are energy of the highest occupied molecular orbital (HOMO), the lowest unoccupied molecular orbital (LUMO), ionisation energy, electron affinity, chemical hardness, electrophilicity, electronegativity and so on. Energy states and electron densities of molecules can be calculated by several quantum chemical calculation methods of *ab initio*, density functional theory (DFT) and semi-empirical methods. *Ab initio* method, the most accurate quantum chemical calculation method, calculates target electron wave functions by linear combination of atomic orbitals. DFT method is an appropriate method for many-body systems with consideration of electron-electron repulsion. Semi-empirical method is the fastest quantum chemical

calculation method with several approximations such as neglect of diatomic overlap integral and electron correlations. In QSAR analysis, quantum chemical calculation process is a time-consuming process since descriptors of dozens or more compounds need to be calculated. Puzyn et al. (2008) suggested that semi-empirical method is an appropriate method for developing QSAR models because it is reliably accurate with a short calculation time.

3D structures of molecules can be used as descriptors for QSAR study. 3D-QSAR technique is a QSAR technique that utilises spatial structural characteristics as molecular descriptors. Biological activities are mainly altered by enzymes; therefore, ligand-enzyme interactions have crucial effects on the biological activities. Because of ligand selectivities of enzymes, inhibition and activation properties of chemical compounds cannot be derived from classical QSAR methods which do not involve 3D spatial structures. Comparative molecular field analysis (CoMFA), the first successful 3D-QSAR study, was performed for analysing binding properties of steroids to carrier proteins (Cramer et al., 1988). Van der Waals and coulombic forces between a target enzyme and compounds are calculated and used as descriptors in the CoMFA analysis. Total interaction energies, molecular lipophilic potentials, hydrophobic interactions and electron densities can be calculated from 3D molecular structures and used as descriptors (Kastenholz et al., 2000; Kellogg and Abraham, 1992). However, it has some disadvantages. Calculating 3D structural descriptors is a time-consuming

process, and due to this drawback, 3D-QSAR is less efficiently than other QSAR systems. Also, validity of 3D-QSAR model is significantly affected by structure optimisation and alignment procedures; therefore, a careful consideration is necessary for each step.

1.1.3.3. QSAR modelling methods

For developing robust and reliable QSAR models, selecting appropriate modelling methods is important. Various modelling methods should be applied for the QSAR models depending on the size and characteristics of the variables. Various modelling methods can be used for QSAR model development. Representative modelling methods are linear regression, partial least-square and machine learning.

Linear regression method is the most common and fundamental problem solving technique. Linear regression is a method for modelling relationship between a dependent variable and explanatory independent variables (Cruciani et al., 1990; Kubinyi, 1997). The simplest case of linear regression, which has only one independent variable, is called simple linear regression (SLR), a linear regression with multiple independent variables called multiple linear regression (MLR) (Snedecor and Cochran, 1967). Because QSAR models generally has multiple descriptors, MLR is commonly used in chemometric studies. Due to its simplicity, transparency, reproducibility and ease of interpretation, many studies have used MLR method for developing QSAR models; from the very first modern QSAR studies of *Hansch analysis* (Hansch and Fujita, 1964) and *Free-Wilson analysis* (Free and Wilson, 1964)

to the latest QSAR studies (Liaw and Svetnik, 2015). Basically, the result from MLR analysis can be interpreted by the sign and the size of each regression coefficient. The plus or minus sign of each coefficient represents a positive or negative correlation to dependent variable. However, interpretations of MLR models are not always correct when there is a collinearity between variables such as multiple mechanisms-related activities (Yee and Wei, 2012). Thus, MLR is not an appropriate method for interpreting non-linear relationship between variables, and the descriptors used in MLR should be independent each other without collinearities. Despite these disadvantages, MLR remains as the most popular QSAR modelling method because of its merits, ease of development and interpretation, and its comparability to previous studies.

Partial least squares or projection to latent structure (PLS) method is a popular modelling method that is the second most used in QSAR studies (Wold et al., 2001; Yee and Wei, 2012). PLS assumes a linear relationship between an explanatory matrix of independent variables and a matrix of dependent variables. Latent factors were derived respectively from independent and dependent variables, and the aim of PLS is to find a multidimensional direction between latent factors. There is no issue on collinearity between dependent variables since PLS uses only one explanatory-latent factor for interpreting biochemical activities. PLS method is usually applied to 3D-QSAR models, while MLR method cannot be applied due to its great number of independent explanatory variables, typically several

thousands (Verma et al., 2010). To sum up, a great number of descriptors are projected to small-sized latent factor as a representative vector for PLS, and this projection makes easier to solve collinearity problems and develop QSAR models with good prediction efficiencies. However, since PLS is a method for finding a relationship between latent variables, the correlation between each independent variable and the dependent variable cannot be directly interpreted.

Due to the complexity of a relationship between structural descriptors and biochemical activities, the relationship might not be defined by classical mathematical QSAR modelling methods including MLR and PLS. Machine learning technique is recently introduced in QSAR field. As a part of computer science, machine learning is defined as the ability to develop self-improvement algorithms without explicit programming (Michalski et al., 2013). One of the main features of machine learning is that nonlinear relationships between descriptors and activities can be mapped by using machine learning. Machine learning methods empirically develop models and construct problem-solving algorithms by training process with observed data sets. By using machine learning, the number of input variables which are introduced in QSAR models can be reduced. Major machine learning methods used in QSAR study area are support vector machine (SVM) and artificial neural network (ANN).

SVM is a method for classifying and regressing independent explanatory variables based on structural risk minimisation principle from statistical learning theory (Vapnik, 2013). This method was originally designed for

binary classification of given data (Cortes and Vapnik, 1995). In addition, SVM for regression was suggested for developing prediction models (Smola and Vapnik, 1997). Xue et al. (2006) reported a QSAR study to distinguish toxic chemical compounds, and Liew et al. (2010) identified effective inhibitors for phosphoinositide 3-kinases by applying SVM. SVM method is a relatively simple and effective machine learning method to discover active compounds with only a few user-defined parameters and lower risk of model overfitting compared to artificial neural network (ANN) models. ANN is also a machine learning method that can be applied in QSAR modelling. The structure of ANN models is basically constituted of processing elements with weighted multi-link-network connections, imitating human cognitive systems. Each processing element is constituted of a function or a group of functions for data processing. ANN-applied model and its connection weights are adjusted and optimised by repetition of training procedure for minimising errors and updating the ANN-applied model. There are various ways to optimise ANN including gradient descent and Quasi-Newton methods. ANN-applied QSAR models can achieve high prediction accuracies (Agatonovic-Kustrin and Beresford, 2000; Byvatov et al., 2003; Niculescu, 2003). However, ANN-applied models are prone to overfitting which lowers prediction efficiency; therefore, validation processes are needed to avoid overfitting (Tu, 1996).

1.1.4. Adaptive neuro-fuzzy inference system (ANFIS)

ANN-applied models, developed by imitating neural networks system in human brain, can be an alternative predictive system for nonlinear relationship between variables. ANN contains several series of mathematical equations used for learning, training and predicting processes (Agatonovic-Kustrin and Beresford, 2000). As one of ANN-applied models, ANFIS is an ANN-applied fuzzy inference system (FIS), which is based on fuzzy logic which is a form of many-valued logic (Figure 1.2). In contrast to Boolean logic, of which output is only 0 or 1, the output of fuzzy logic can have various values (Hájek, 1998). Therefore, continuous quantitative variables can be processed easily by applying fuzzy logic. In addition to fuzzy logic, the structure of FIS has a deffuzification layer for drawing single quantitative value. Data processing procedure of ANFIS is initialised by passing explanatory variables to input membership fuzzy functions of FIS, and the responses from input membership fuzzy functions are connected to output

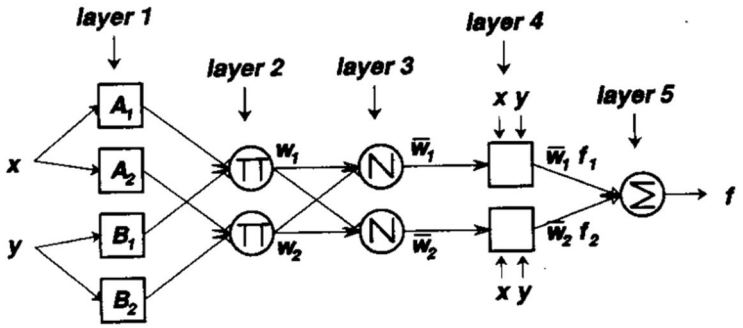

Figure 1.2. Basic structure of adaptive neuro-fuzzy inference system (ANFIS). Retrieved from Jang et al. (1993).

membership function by if-then rules. If-then rules and the level response

from each of the if-then rules are adjusted by output membership fuzzy functions of FIS, so that the boundaries of the rules are unsharpened (Jang, 1993). ANFIS has a multilayer structure, and connection weights between the layers are optimised by training procedure. As a result of training procedures, ANFIS-applied model can be used for predicting and analysing activities of molecules with high accuracy. Buyukbingol et al. (2007) developed a ANFIS-applied prediction model for N-methyl-d-aspartate receptor antagonists, and the developed model has satisfactory prediction accuracy. Afiuni-Zadeh and Azimi (2010) also developed QSAR models for A1 receptor antagonists, and the ANFIS-applied model had higher prediction accuracy than linear regression and radial basis function models.

1.2. Objectives

As mentioned above, human antioxidant system is constituted of endogenous and exogenous components. As exogenous antioxidants, dietary phytochemicals can contribute to human antioxidant system for scavenging free radicals. In addition to physiological effects, antioxidant characteristic of phytochemicals affects stability, safety and quality of food. Antioxidant activities of various phytochemicals were evaluated in thousands of previous studies, and the phytochemicals have been demonstrated as effective antioxidants. Although there have been numerous studies that evaluated antioxidant activities of phytochemicals, most of previous studies did not rationally explained the relationship between chemical structures and

antioxidant activities. However, for better understanding the antioxidant mechanism of phytochemicals, and for developing advanced bio-materials, structure related approaches are needed. The structure-related approach can be an effective and efficient strategy for discovering active compounds for functional foods. Furthermore, accurate prediction models are need to be developed to predict and evaluate antioxidant activities of phytochemicals and their derivatives. For example, antioxidative plant extracts can be directly used as ingredients for functional foods. However, the activities of plant-derived functional foods can be increased by converting natural phytochemicals to more effective derivatives. In this case, activities of specific phytochemical-derivatives can be predicted by QSAR model, and rational strategies for developing functional foods and bio-materials can be established.

Since the antioxidant mechanisms are based on electron or atom transfer, quantum chemical descriptors were used as variables for QSAR models. As a QSAR modelling method, machine learning method was chosen because machine learning method could interpret nonlinear relationship between various explanatory variables and a dependent variable. SVM method is a useful method for determining the effectiveness of each compound; however, it is not an appropriate method for evaluating and predicting activity of each compound. In this study, ANFIS was chosen as a QSAR modelling method for achieving good prediction efficiencies, because quantitative variables could be processed simply by FIS structure and the model could be optimised

by ANN of ANFIS.

Therefore, the aims of this study were to establish and validate QSAR models for antioxidant activities of phytochemicals with high prediction accuracies. This study is composed of three parts. In the first part of the study, ANFIS applied-QSAR models for antioxidant activities of anthocyanins were developed. In the second part of the study, ANFIS applied-QSAR models for antioxidant activities of carotenoids were developed. In the third part of the study, ANFIS applied-QSAR models for antioxidant activities of various phenolic compounds, which did not have common skeleton structures, were developed.

Chapter 2

Prediction of Radical Scavenging Activities of Anthocyanins Applying Adaptive Neuro-Fuzzy Inference System (ANFIS) with Quantum Chemical Descriptors

(Study 1)

(Published in International Journal of Molecular Sciences,

2014; 15: 14715-14727)

2.1. Introduction

As constituents of flavonoid group, anthocyanins and their aglycones, anthocyanidins, are colour pigments originated from plants. Anthocyanins basically have a flavylium ion skeleton attached with different side groups including hydrogen atom, hydroxyl group and methoxy group. Anthocyanins are classified by the position and type of the side groups (Figure 2.1 and Table 2.1).

Radical scavenging activity is one of the characteristics of anthocyanins and other flavonoids. These compounds can reduce reactive oxygen species (ROS), resulting in relieving oxidative stress (Kong et al., 2003; Tsuda et al., 1996). The radical scavenging activities can be explained by two suggested mechanisms: One is hydrogen atom transfer (HAT) and the other is single electron transfer (SET) (Prior et al., 2005). HAT-based radical scavenging activity is explained by hydrogen donation as following equation:

$$\mathbf{X\cdot + AH \rightarrow XH + A\cdot} \qquad (2.1)$$

where X is a free radical, such as ROS, and AH is a molecule of antioxidant. Free radicals are scavenged by hydrogen atoms donated from an antioxidant. SET-based radical scavenging activity can be explained by the scheme below:

$$\mathbf{X\cdot + AH \rightarrow X^{-} + AH^{\cdot +}} \qquad (2.2)$$

An electron from an antioxidant transfers to a radical, and consequently the electron pairs up with unpaired electron of the radical. 2,2'-Azino-bis(3-ethylbenzothiazoline-6-sulphonic acid) (ABTS), a stable radical, is mainly reduced by SET mechanism (Prior et al., 2005; Scott et al., 1993), while 1,1-

diphenyl-2-picryl hydrazyl (DPPH) radical is mainly reduced by HAT mechanism (Bondet et al., 1997). Until now, it has been considered that quantum chemical descriptors of ionisation potential and bond dissociation energy are related with SET and HAT-based radical scavenging activities, respectively (Amić et al., 2007).

The relationship between molecular descriptors (e.g., number of hydroxyl group, presence of double bond on flavonoid C ring and bond dissociation energy of hydrogen) of flavonoids and radical scavenging activities has been studied (Amić et al., 2007). These prediction models can be applied to estimate and design novel radical scavenging bio-materials. However, the relationship between molecular descriptors of anthocyanins and their radical scavenging activities has been poorly understood. Unlike other flavonoids, the structures of anthocyanins vary depending on pH (Kong et al., 2003; Mazza and Brouillard, 1987). At low pH (< 2), flavylium cation (FC) is dominant with red and purple colour. At higher pH, FC decreases, while quinoidal base (QB), carbinol pseudo-base (CP) and chalcone (Ch) forms of anthocyanins increase (Mazza and Brouillard, 1987). The change of the structure leads to the change of molecular descriptors of an anthocyanin molecule. For better understanding of the relationship between molecular descriptors of anthocyanins and their radical scavenging activities, molecular descriptors of various anthocyanin forms should be considered.

To establish quantitative structure-activity relationship (QSAR) model, linear combinations of variables have been generally used in previous studies

(Amić and Lučić, 2010; Amić et al., 2007). The linear combination equation could be constructed easily through regression analysis. However, the interaction effect between variables is ignored, and nonlinear relationship between variables cannot be explained easily by linear combination. These kinds of defects may affect reliability of prediction model. Therefore, another prediction model is necessary for explaining the relationship between variables. Adaptive neuro-fuzzy inference system (ANFIS) is an artificial neural network (ANN)—applied fuzzy inference system (FIS). By using ANFIS, multi-variable related ambiguous relationship can be quantified by defuzzification process of FIS, and error is adjusted by backpropagation algorithm with hidden layer of ANN for reliable prediction (Buyukbingol et al., 2007). Therefore, ANFIS is an effective technique for solving problems which cannot be easily solved by linear regression analysis. The aim of this study was to establish a QSAR model for predicting radical scavenging activities of anthocyanins by ANFIS using quantum chemical descriptors

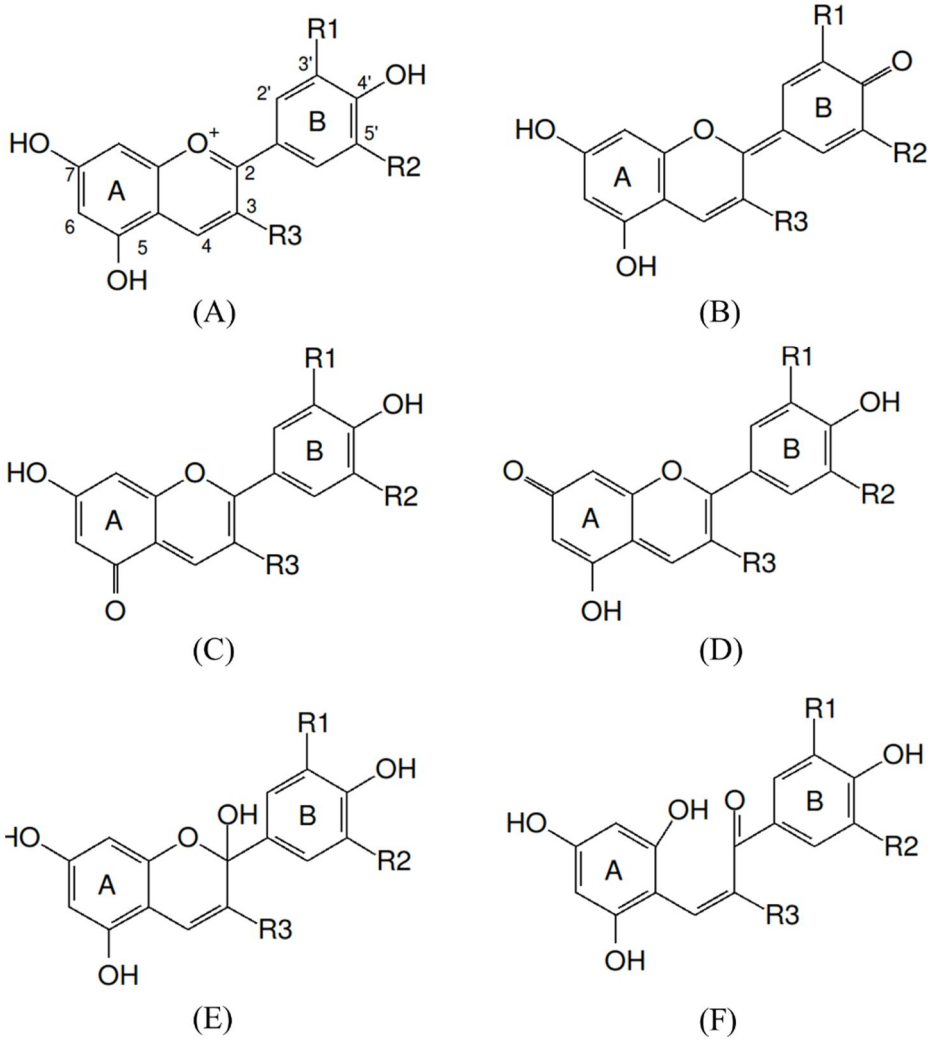

Figure 2.1. (A) Flavylium cation; (B−D) quinoidal bases; (E) carbinol pseudo-base; and (F) chalcone structure of anthocyanidin.

2.2. Materials and methods

2.2.1. DPPH radical scavenging activity

DPPH radical scavenging activity data of 5 anthocyanidins and 16 anthocyanins were adapted from a previous study, which are listed in Table 2.1 (Kähkönen and Heinonen, 2003).

2.2.2. Quantum chemical descriptors

2.2.2.1. Molecular structure preparation

The structures of 21 anthocyanidins and anthocyanins were prepared using Gabedit 2.4.3 (Allouche, 2011).

Table 2.1. Evaluated anthocyanidins and anthocyanins.

Family	Compounds	[a]*n*OH	R1	R2	[b]R3	[b]R4
Anthocyanidin	Cyanidin	5	OH	H	OH	OH
	Delphinidin	6	OH	OH	OH	OH
	Malvidin	4	OCH_3	OCH_3	OH	OH
	Pelargonidin	4	H	H	OH	OH
	Peonidin	4	OCH_3	H	OH	OH
Anthocyanin	Cyanidin-3-coumaroyl-sambubioside-5-galactoside	3	OH	H	coumaroyl -sam	gal
	Cyanidin-3-sambubioside-5-galactoside	3	OH	H	sam	gal
	Cyanidin-3-arabinoside	4	OH	H	ara	OH
	Cyanidin-3-galactoside	4	OH	H	gal	OH
	Cyanidin-3-glucoside	4	OH	H	glc	OH
	Cyanidin-3-rutinoside	4	OH	H	rut	OH
	Cyanidin-3,5-diglucoside	3	OH	H	glc	glc

Table 2.1. (continued)

Family	Compounds	[a]*n*OH	R1	R2	[b]R3	[b]R4
	Delphinidin-3-glucoside	5	OH	OH	glc	OH
	Delphinidin-3-rutinoside	5	OH	OH	rut	OH
	Malvidin-3-galactoside	3	OCH_3	OCH_3	gal	OH
	Malvidin-3-glucoside	3	OCH_3	OCH_3	glc	OH
Anthocyanin	Malvidin-3,5-diglucoside	2	OCH_3	OCH_3	glc	glc
	Pelargonidin-3-glucoside	3	H	H	glc	OH
	Peonidin-3-galactoside	3	OCH_3	H	gal	OH
	Peonidin-3-glucoside	3	OCH_3	H	glc	OH
	Petunidin-3-glucoside	4	OH	OCH_3	glc	OH

[a]*n*OH: number of hydroxyl groups on flavonoid core; [b]sam: sambubioside; gal: galactoside; ara: arabinoside; glc: glucoside; and rut: rutinoside.

2.2.2.2. Calculation of quantum chemical descriptors

Geometrical optimisation of molecular structures was performed by semi-empirical methods of PM6 (Stewart, 2007) and PM7 (Stewart, 2013) using MOPAC2012(Stewart, 2014). Quantum chemical descriptors of ionisation potential (I), electron affinity (A), chemical hardness (η), softness (S), electronegativity (χ), chemical potential (μ), electrophilicity (ω) and hydrogen atom dissociation energy (BDE) were calculated to establish QSAR models.

I was calculated as the following equation:

$$\boldsymbol{I = E(N - 1) - E(N)} \quad (2.3)$$

where $E(N - 1)$ is energy of an anthocyanin radical generated after electron abstraction and $E(N)$ is energy of an anthocyanin molecule. A was calculated as the following equation:

$$\boldsymbol{A = E(N) - E(N + 1)} \quad (2.4)$$

where $E(N + 1)$ is energy of an anthocyanin radical generated after addition of an electron.

η was calculated from the following equation:

$$\boldsymbol{\eta = (I - A)/2} \quad (2.5)$$

S was calculated from the following equation:

$$\boldsymbol{S = 1/2\eta} \quad (2.6)$$

χ and μ were calculated from the following equation:

$$\boldsymbol{\chi = -\mu = (I + A)/2} \quad (2.7)$$

ω was calculated from the following equation:

$$\omega = \mu^2/2\eta \quad (2.8)$$

BDE was calculated from the following equation:

$$\mathbf{BDE} = \boldsymbol{E}(\mathbf{A{-}O\cdot}) + \boldsymbol{E}(\mathbf{H}) - \boldsymbol{E}(\mathbf{A{-}OH}) \quad (2.9)$$

where E(A–O·) is energy of a hydrogen-abstracted anthocyanin phenoxyl radical, E(H) is the energy of a hydrogen atom and E(A–OH) is energy of the anthocyanin molecule.

2.2.3. QSAR model development

2.2.3.1. Correlation analysis

Pearson's correlation analysis was conducted to examine the relationship between the quantum chemical descriptors and the DPPH radical scavenging activities using GNU R 3.0.1 (http://cran.r-project.org/). Quantum chemical descriptors which were highly correlated with radical scavenging activities were chosen as independent variables for QSAR models.

2.2.3.2. ANFIS

ANFIS models were made using a fuzzy logic toolbox of Matlab 8.2 (Mathworks, Natick, MA, USA). Previous studies reported that the number of phenolic OH groups of flavonoid core (*n*OH) was correlated with radical scavenging activities of flavonoids (Amić and Lučić, 2010) and anthocyanins (Chang et al., 2008). Therefore, *n*OH as well as quantum chemical descriptors was used for establishing QSAR model development.

For development of ANFIS models, *A* and χ of FC, *I* of QB7 and *n*OH were used as independent variables. DPPH radical scavenging activities of anthocyanins were used as a dependent variable. Two triangular-shaped membership functions for each independent variable, 16 if-then rules and 16 linear type output membership functions were developed (Figure 2.2). The ANFIS models were optimised by backpropagation method with 100 learning epochs. To validate the constructed ANFIS models, bootstrap validation procedure was repeated 1000 times. Mean absolute error (MAE) was calculated and compared by bootstrapping. MAE was calculated from the following equation:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y'_i - y_i| \qquad (2.10)$$

where y'_i is predicted DPPH radical scavenging activity and y_i is experimental one. Also, Q square between predicted radical scavenging activities and experimental ones was calculated and validated by bootstrap validation.

2.3. Results and discussion

2.3.1. Correlation analysis

The result of correlation analysis between the calculated quantum chemical descriptors (*I*, *A*, *η*, *S*, χ, *μ* and *ω*) of anthocyanin structures and the DPPH radical scavenging activities is presented in Table 2.2. Since 4 out of 21 tested anthocyanins had a glycosidic bond on 5-position, QB5 forms could not be

generated. Thus, QB5 structures were excluded for further analysis. The χ and μ of FC, and *I* of QB7 calculated by both of PM6 and PM7 methods were significantly correlated with DPPH radical scavenging activities ($p < 0.01$). *A* calculated by PM6 and PM7 methods were also significant at $p < 0.01$ and $p < 0.05$, respectively. Some of the other quantum chemical descriptors of FC, QB7 and CP, calculated by PM6 and PM7 methods, were significant ($p < 0.05$). However, none of the quantum chemical descriptors of QB4' nor Ch was significantly correlated with radical scavenging activities. This result suggests that FC and QB7 forms are main contributors for radical scavenging characteristics of anthocyanins rather than CP and Ch. Since the radical scavenging experiment by Kähkönen and Heinonen (2003), of which data was adapted to this study, was not conducted on a highly acidic condition nor a highly basic condition, QB and FC seemed to exist abundantly at the experimental condition. The structural change affects quantum chemical descriptors and redox properties of molecules (Amić and Lučić, 2010; Karelson et al., 1996). Borkowski et al. (2005) also reported that radical scavenging activities of anthocyanins were influenced by pH. In the case of QB, QB7 was suggested as a primary contributor for radical scavenging activities, because a higher correlation was observed between quantum chemical descriptors of QB7 and radical scavenging activities compared with QB4'.

Table 2.2. Correlation coefficient between 1,1-diphenyl-2-picryl hydrazyl (DPPH) radical scavenging activities and quantum chemical descriptors of flavylium cation, quinoidal base, carbinol pseudo-base and chalcone.

Descriptors	PM6					PM7				
	FC	QB4'	QB7	CP	Ch	FC	QB4'	QB7	CP	Ch
I	0.554**	0.343	0.674**	0.233	0.121	0.504*	0.355	0.708**	0.269	0.251
A	0.673**	−0.166	−0.256	−0.473*	−0.052	0.538*	−0.206	−0.037	−0.289	−0.199
η	0.124	0.288	0.523*	0.530*	0.121	0.263	0.310	0.418	0.378	0.298
S	−0.153	−0.171	−0.503*	−0.535*	−0.138	−0.238	−0.132	−0.398	−0.411	−0.284
ω	0.328	−0.127	−0.305	−0.470*	−0.054	0.202	−0.112	−0.085	−0.326	−0.212
χ	0.701**	0.194	0.282	−0.344	0.009	0.567**	0.087	0.551**	−0.149	0.020
μ	−0.701**	−0.194	−0.282	0.344	−0.009	−0.567**	−0.087	−0.551**	0.149	−0.020

FC: flavylium cation; QB7 and QB4': quinoidal bases (Figure 2.1. B and D, respectively); CP: carbinol pseudo-base; Ch: chalcone; I: ionisation potential; A: electron affinity; η: chemical hardness; S: chemical softness; ω: electrophilicity; χ: electronegativity; and μ: chemical potential. *, ** significance ($p < 0.05$ or $p < 0.01$).

For further investigation, BDE of FC and QB7 were calculated and compared. Some previous studies reported that negative relationship was observed between the lowest BDE (BDE_{min}) and radical scavenging activities (Amić and Lučić, 2010; Amić et al., 2007). Physicochemically, a hydrogen abstraction reaction is more favoured at the position with a lower BDE, because a hydroxyl group with a lower BDE needs less energy to dissociate hydrogen. In addition, the dissociated hydrogen pairs up with a free radical to scavenge it. For this reason, BDE_{min} represents the radical scavenging activity by HAT mechanism. Average BDE of FC and QB7 structures were lowest at the hydroxyl group on 4'-position (Table 2.3). This tendency was also observed in a previous study by density function theory method (Kozlowski et al., 2007; Marı et al., 2007; Trouillas et al., 2006). It has been reported that BDE_{min} of flavonoids were negatively correlated with radical scavenging activities by Amić and Lučić (2010). In this study, however, BDE_{min} of FC and QB7 calculated by PM6 and PM7 were not significantly correlated with DPPH radical scavenging activities ($p > 0.05$). This result was probably caused by conformational difference between anthocyanins, of which conformation changes depending on pH, and flavonoids. Thus, BDE was excluded for establishing QSAR model.

Table 2.3. Hydrogen atom dissociation energy of flavylium cations and quinoidal bases calculated by semi-empirical methods.

Compounds	Method	[a]BDE of FL						BDE of QB7				
		3	5	7	3'	4'	5'	3	5	3'	4'	5'
Cyanidin	PM6	**80.98**	84.8	90.2	82.06	83.34	-	**67.08**	75.5	76.04	72.44	-
	PM7	81.77	**81.77**	91.41	83.43	84.78	-	**68.98**	75.72	75.54	75.16	-
Delphinidin	PM6	78.15	84.24	90.53	83.16	**75.61**	78.55	**65.65**	74.8	79.17	68.74	74.47
	PM7	82.55	85.75	92.2	85.58	**77.6**	82.08	**68.2**	75.42	78.74	71.29	75.13
Malvidin	PM6	77.33	82.05	87.19	-	**73.9**	-	**65.49**	74.83	-	68.04	-
	PM7	81.19	85.22	90.98	-	**75.21**	-	**71.31**	78.69	-	73.36	-
Pelargonidin	PM6	**81.01**	84.79	90.39	-	89.53	-	**68.45**	75.5	-	74.43	-
	PM7	**81.86**	85.71	91.51	-	90.6	-	**69.68**	75.89	-	76.87	-
Peonidin	PM6	80.86	84.58	90.45	-	**77.37**	-	**68.38**	76.99	-	69.2	-
	PM7	**81.86**	85.71	91.51	-	90.6	-	**71.53**	78.3	-	73.51	-
Cyanidin-3-coumaroyl-sambubioside-5-galactoside	PM6	-	-	88.15	**80.05**	82.17	-	-	-	**74.12**	75.16	-
	PM7	-	-	86.57	**77.29**	78.88	-	-	-	**76.74**	77.52	-

Table 2.3. (continued)

Compounds	Method	[a]BDE of FL						BDE of QB7				
		3	5	7	3'	4'	5'	3	5	3'	4'	5'
Cyanidin-3-sambubioside-5-galactoside	PM6	-	-	89.15	**80.07**	82.15	-	-	-	**74.12**	75.23	-
	PM7	-	-	98.43	**88.66**	90.97	-	-	-	**76.74**	78.07	-
Cyanidin-3,5-diglucoside	PM6	-	-	91.23	**80.23**	83.08	-	-	-	**74.85**	75.92	-
	PM7	-	85.22	-	**80.39**	82.39	-	-	-	**77.19**	78.28	-
Cyanidin-3-arabinoside	PM6	-	86.49	91.25	**80.53**	82.61	-	-	77.32	76.27	**73.99**	-
	PM7	-	88.14	92.69	**82.63**	85.54	-	-	77.87	**77.09**	77.48	-
Cyanidin-3-galactoside	PM6	-	93.17	93.58	**81.03**	83.71	-	-	76.74	**74.76**	75.55	-
	PM7	-	96.62	96.76	**84.86**	85.8	-	-	81.26	80.34	**79.3**	-
Cyanidin-3-glucoside	PM6	-	86.87	92.58	**81.78**	84.4	-	-	76.07	**74.91**	75.63	-
	PM7	-	87.79	82.92	**82.92**	85.51	-	-	**77.07**	77.52	77.99	-
Cyanidin-3-rutinoside	PM6	-	88.51	94.22	**80.41**	83.25	-	-	76.6	**74.72**	76.12	-
	PM7	-	90.49	94.22	**81.73**	84.65	-	-	80.23	**77.36**	78.9	-
Delphinidin-3-glucoside	PM6	-	86.92	91.58	82.36	**75.31**	77.35	-	76.9	78.53	**69.28**	74.42
	PM7	-	89.94	92.99	84.51	**77.89**	80.94	-	80.64	80.53	**77.06**	80.23
Delphinidin-3-rutinoside	PM6	-	88.23	91.05	82.52	75.93	**75.74**	-	75.22	77.2	**70.95**	72.55
	PM7	-	84.54	92.68	84.59	77.79	**76.34**	-	**72.62**	80.15	73.79	73.62

Table 2.3. (continued)

Compounds	Method	[a]BDE of FL						BDE of QB7				
		3	5	7	3'	4'	5'	3	5	3'	4'	5'
Malvidin-3,5-diglucoside	PM6	-	-	86.32	-	**73.67**	-	-	-	-	**71.07**	-
	PM7	-	-	91.84	-	**76.05**	-	-	-	-	**72.45**	-
Malvidin-3-galactoside	PM6	-	87.49	89.43	-	**73.81**	-	-	74.52	-	**69.57**	-
	PM7	-	88.43	91.56	-	**75.73**	-	-	78.65	-	**72.45**	-
Malvidin-3-glucoside	PM6	-	85.75	89.51	-	**73.89**	-	-	76.84	-	**68.59**	-
	PM7	-	88.63	94.26	-	**78.17**	-	-	73.52	-	**71.67**	-
Pelargonidin-3-glucoside	PM6	-	87.13	91.13	-	**86.79**	-	-	**71.66**	-	75.48	-
	PM7	-	**86.20**	93.35	-	89.08	-	-	**72.61**	-	78.55	-
Peonidin-3-galactoside	PM6	-	88.79	98.46	-	**76.39**	-	-	79.56	-	**68.31**	-
	PM7	-	88.94	92.01		**78.78**	-	-	77.10	-	**72.78**	
Peonidin-3-glucoside	PM6	-	88.80	90.97	-	**76.38**	-	-	74.66	-	**67.51**	-
	PM7	-	88.57	91.51	-	**78.74**	-	-	78.95	-	**72.74**	-
Petunidin-3-glucoside	PM6	-	86.34	90.85	82.33	**74.17**	-	-	76.79	79.31	**68.40**	-
	PM7	-	88.62	94.78	87.69	**78.51**	-	-	73.73	80.33	**71.51**	-

[a]BDE: hydrogen atom bond dissociation energy (kcal/mol); FL: flavylium cation; and QB7: quinoidal base (Figure 2.1D). The lowest BDE is presented in bold.

The possible mechanism for DPPH radical scavenging activities could be suggested by these results. Since BDE_{min} was not significantly correlated with radical scavenging activities, the hydrogen atom transfer directly from anthocyanin to DPPH radical might not be a main mechanism for radical scavenging. Osman (2011) suggested that DPPH radical and polyphenol compounds form a complex as an intermediate. For an anthocyanin, as a polyphenol, an intermediate forming reaction could occur. Furthermore, the hydrazyl moiety of DPPH radical has nucleophilic characteristic and FC has high A and χ. The intermediate forming step was suggested as a reaction barrier, and this could explain why the DPPH radical scavenging activities were highly correlated with A and χ of FC. On the other hand, from the positive correlation with I of QB7 and radical scavenging activity, electron transfer from QB7 to other molecules possibly lowers radical scavenging activities. Possible mechanism is a single electron transfer from QB7 to FC. It is easy for a QB7 molecule with low I to transfer an electron to other molecules. Single electron addition to FC neutralizes net charge of the molecule to zero, and this will lower A and χ of the molecule. The required energy for intermediate forming reaction between anthocyanin and DPPH radical tends to increase, and consequently total radical scavenging activity decreases. Matsufuji et al. (2007) reported that DPPH radical scavenging activities of anthocyanins were higher in acidic condition (pH 3) than in neutral condition (pH 7), and this tendency could also be explained by our suggested mechanism. Since electron transfer from QB7 to FC is suggested as

decreasing total radical scavenging activities, low occurrence of QB7 and high occurrence of FC in acidic condition contribute to high DPPH radical scavenging activity.

2.3.2. Prediction of radical scavenging activities of anthocyanins

Through correlation analysis, four quantum chemical descriptors of A, χ and μ of FC, and I of QB7 were most highly correlated with radical scavenging activities. Since both of χ and μ were drawn by the same equation (2.7), the correlation coefficient (R square) between these two variables was exactly 1. Hence, because using both of the descriptors was unnecessary, χ, one of the two descriptors, was used for establishing the QSAR model. Previous studies also reported that χ and μ were reliable descriptors correlated with biological activities (Karelson et al., 1996; Pasha et al., 2005). In addition, it has been reported that nOH was significantly correlated with radical scavenging activities of flavonoids and anthocyanins (Amić and Lučić, 2010; Amić et al., 2007). Therefore, nOH was chosen as an independent variable. Consequently, A and χ of FC, I of QB7 and nOH were used as independent variables for establishing QSAR model. The calculated A and χ of FC, I of QB7, nOH and experimental DPPH radical scavenging activities were used as training data set for ANFIS model training. Structure of established ANFIS models is illustrated in Figure 2.2.

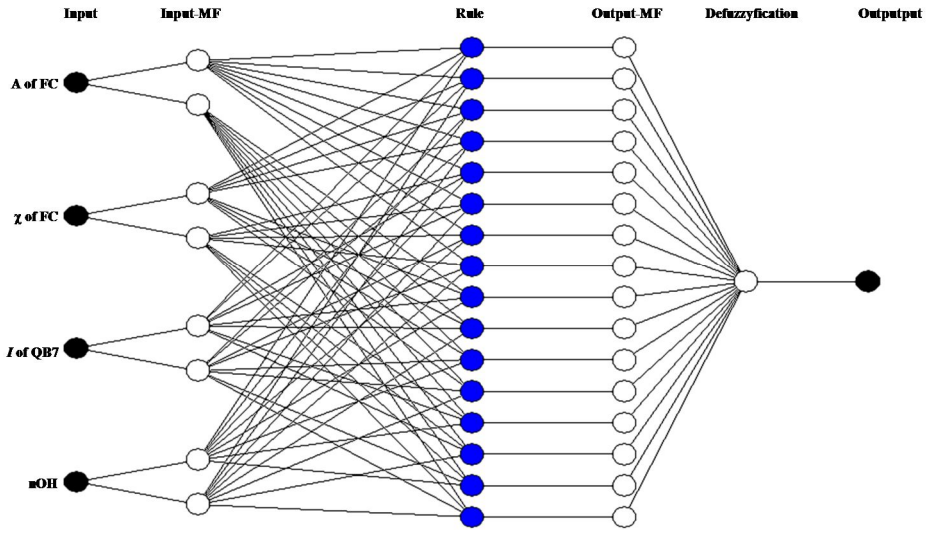

Figure 2.2. Developed adaptive neuro-fuzzy inference system (ANFIS) structure.

The experimental and predicted radical scavenging activities are presented in Table 2.4 and Figure 2.3. Since the number of naturally present and commercially available anthocyanins was limited, the test set with a relatively small number of compounds were only available to analyse. For small data set, cross validation with dividing data set into training set and test set might not be valid. Previous studies suggested a bootstrap validation method for small sample validation rather than cross validation (Braga-Neto and Dougherty, 2004; Kiralj and Ferreira, 2009). Thus, bootstrap validation was performed in this study. The MAE and Q square resulted from bootstrap validation were presented in Table 2.4. The ANFIS models with quantum chemical descriptors calculated by both of PM6 and PM7 semi-empirical methods had high prediction efficiencies ($p < 0.01$).

Table 2.4. Experimental and predicted DPPH radical scavenging activities.

Compounds	[a]Experimental radical scavenging activity	Predicted radical scavenging activity	
		PM6	PM7
Cyanidin	33	35.3	33.2
Delphinidin	42	42.6	41.0
Malvidin	24	26.6	26.1
Pelargonidin	31	28.3	27.6
Peonidin	33	31.4	29.3
Cyanidin-3-coumaroylsambubiose-5-galactoside	26	22.4	25.2
Cyanidin-3,5-diglucoside	21	22.6	20.9
Cyanidin-3-arabinoside	26	28.5	27.3
Cyanidin-3-sambubiose-5-galactoside	22	22.1	21.5
Cyanidin-3-galactoside	25	30.3	30.2
Cyanidin-3-glucoside	32	31.1	28.8
Cyanidin-3-rutinoside	25	27.0	27.5
Delphinidin-3-glucoside	42	35.4	38.1
Delphinidin-3-rutinoside	32	33.4	30.4
Malvidin-3,5-diglucoside	14	16.6	14.9

Table 2.4. (continued)

Compounds	[a]Experimental radical scavenging activity	Predicted radical scavenging activity	
		PM6	PM7
Malvidin-3-galactoside	22	22.9	21.3
Malvidin-3-glucoside	26	22.6	21.2
Pelargonidin-3-glucoside	20	21.0	21.2
Peonidin-3-galactoside	20	22.2	21.1
Peonidin-3-glucoside	26	23.6	23.5
Petunidin-3-glucoside	23	28.0	25.7
[b]Mean absolute error		**2.43 ± 0.35**	**2.06 ± 0.32**
[b]Q squre		**0.82 ± 0.08**	**0.86 ± 0.08**

[a] Experimental radical scavenging activity data was adapted from a previous study by Kähkönen and Heinonen (2003), percentage of scavenged DPPH radical; [b] Mean absolute error and Q-squre resulted from bootstrap validation, presented as mean ± standard deviation

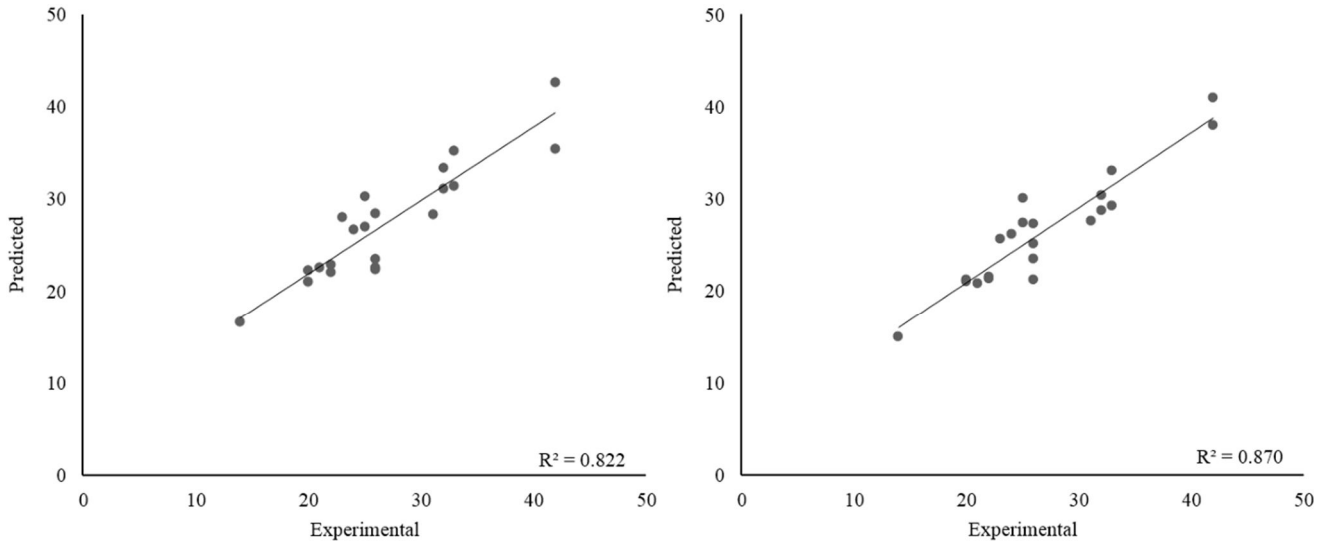

Figure 2.3. Scatter plot between the experimental DPPH radical scavenging activities of anthocyanins by Kähkönen and Heinonen (2003) and the predicted DPPH radical scavenging activities based on the quantum chemical descriptors calculated by (A) PM6 and (B) PM7 methods.

The error of the model continually decreased on every learning epoch (data not shown). After 100 learning epochs, the MAE of the models with quantum chemical descriptors calculated by PM6 and PM7 were 2.44 and 2.06, respectively. Average DPPH radical scavenging activity of the tested compounds was 26.6. Thus, the prediction efficiencies of the established models were over 90%. Previous studies applied linear regression analysis for establishing QSAR models (Amić et al., 2007; Chang et al., 2008; Karelson et al., 1996; Pasha et al., 2005). However, interaction between the variables was ignored in linear equations; therefore, a linear regression technique may not be suitable for multivariable analysis. Because four different independent variables were used in this study, linear regression analysis was not suitable for correcting interaction between the variables. ANN has been applied on scientific studies because ANN can solve problems and predict problems practically as an adaptive intelligent system (Amiryousefi et al., 2011; Buyukbingol et al., 2007; Ghoush et al., 2008). FIS can also solve non-linearly correlated variables by managing uncertainty of functions (Jang et al., 1997; Lin et al., 2009). This explains the reason why ANFIS, an ANN applied problem-solving technique, was applied in this study. As a result, the established model showed a high prediction efficiency and the ANFIS model fitted well with experimental data.

Chapter 3

Adaptive Neuro-Fuzzy Inference System-Applied QSAR with Quantum Chemical Descriptors for Predicting Radical Scavenging Activities of Carotenoids

(Study 2)

(Published in Plos One, 2015; 10: e0140154)

3.1. Introduction

Carotenoids are natural pigments that have various physiological activities such as conversion to vitamin A and immunological activities (Blount et al., 2003; Faivre et al., 2003; Schwartz et al., 1989). One of the important physiological activities of carotenoids is their radical scavenging activity (El-Agamey et al., 2004; Miller et al., 1996; Müller et al., 2011; Stahl and Sies, 2005). The radical scavenging activity of carotenoids, mainly contributed by conjugated polyene structures of carotenoids (Britton, 1995), is related with direct quenching of singlet oxygen and radical species (Cantrell et al., 2003; Krinsky and Deneke, 1982). On the other hand, quantitative structure-activity relationship (QSAR) studies regarding antioxidant activities and radical scavenging activities of natural products have been conducted to understand radical scavenging mechanisms and to predict the activities (Jhin and Hwang, 2014; Kontogiorgis et al., 2005; Worachartcheewan et al., 2011). The quantum chemical descriptors (such as ionisation energy, electrophilicity, highest occupied molecular orbital (HOMO), lowest unoccupied molecular orbital (LUMO) energies, etc.) have been used as independent variables in QSAR studies, because the quantum chemical descriptors characterise electronic and geometric properties of molecules which affect activities of the molecules (Karelson et al., 1996). There have been QSAR studies on antioxidant activities of carotenoids (Kleinová et al., 2007; Soffers et al., 1999); however, correlation between various quantum chemical descriptors and antioxidant activities of carotenoids have not been reported. Since

quantum chemical descriptors represent physicochemical properties of molecules, QSAR models for antioxidant activities of carotenoids could be developed by using their quantum chemical descriptors.

The relationship between quantum chemical descriptors and radical scavenging activities was studied previously (Jhin and Hwang, 2014; Karelson et al., 1996; Pasha et al., 2008; Sarkar et al., 2011; Vasil'ev et al., 2010). The relationship between quantum chemical descriptors and the radical scavenging activities can be explained theoretically. Radical species could be scavenged by single electron transfer from a radical scavenger (Prior et al., 2005). In a physicochemical manner, it is considered that compounds with low ionisation energy tend to be good radical scavengers because electron transfer from the HOMO of a radical scavenger to the singly occupied molecular orbital (SOMO) of a radical occurs more easily on the compounds with a lower ionisation energy compared to the compounds with a higher ionisation energy (Karelson et al., 1996). In case of nucleophilic radical species, the radical scavenging reaction is attributed to the interaction between SOMO of a radical and the LUMO of a radical scavenger (Park et al., 1986; Poblet et al., 1983). According to Koopmans' theorem (Koopmans, 1934), the energy level of LUMO is a negative electron affinity; therefore, a compound with a higher electron affinity tends to be attacked easily by a nucleophilic radical. The energy level of the SOMO of radical is dropped by interaction with the LUMO of a radical scavenger, followed by single electron transfer from HOMO of radical scavenger to the SOMO of radical to scavenge it. The

low HOMO-LUMO gap allows electron transfer from HOMO more easily. Since the concept of chemical hardness is based on the HOMO-LUMO gap (Parr and Pearson, 1983), compounds with lower chemical hardness tend to have higher radical scavenging activity (Karelson et al., 1996; Kontogiorgis et al., 2005). In addition, the driving force for electron transfer is chemical potential (negative electronegativity), which is the slope of the energy versus number of electron curve (Parr et al., 1978). In this manner, the chemical potential indicates the rate and the direction of chemical reactions including the radical scavenging activity.

Linear regression method is usually applied to develop QSAR models in several studies (Amić and Lučić, 2010; Kleinová et al., 2007; Kontogiorgis et al., 2005; Pasha et al., 2008; Soffers et al., 1999); however, the interaction effect between independent variables and nonlinear relationships could not be interpreted easily by the linear regression method. Adaptive neuro-fuzzy inference system (ANFIS) is an artificial neural network (ANN)-applied fuzzy inference system. ANFIS is an empirical method that defines rules by training and backpropagation processes, and multivariate nonlinear relationship could be solved by fuzzy if-then rules (Jang, 1993). Thus, in this study, ANFIS-applied QSAR models were developed with quantum chemical descriptors for predicting and analysing antioxidant activities of carotenoids.

3.2. Materials and methods

3.2.1. Data set

Total three data sets of radical scavenging activities of carotenoids were adapted from previous studies. The first trolox equivalent antioxidant capacity (TEAC) data set by Müller et al. (2011) was constituted of 17 carotenoids (Figure 3.1), and the second data set by Miller et al. (1996) was constituted of 8 carotenoids. A DPPH radical scavenging activity data set was adapted from the report by Jiménez-Escrig et al. (2000).

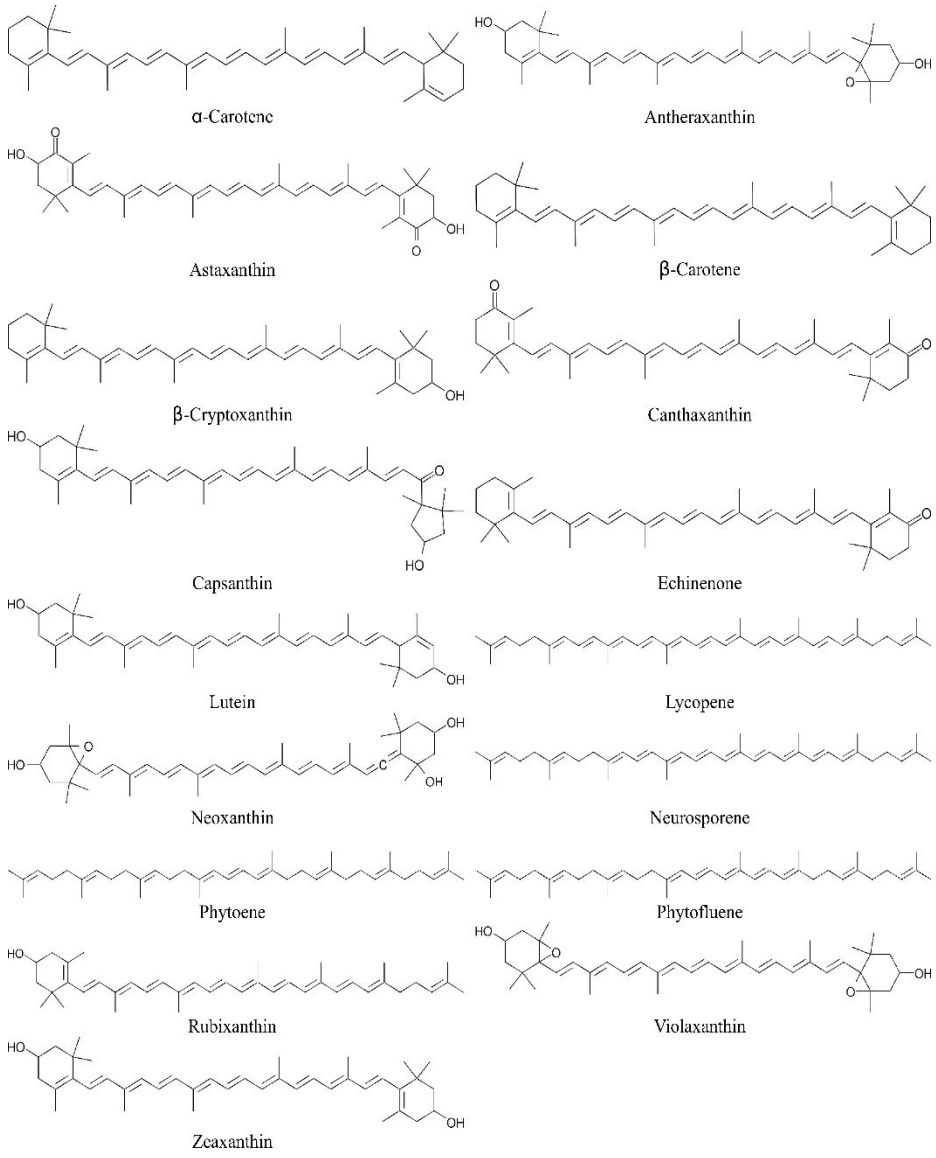

Figure 3.1. Structures of carotenoids presented by Müller et al. (2011).

3.2.2. Molecular structure preparation

The structures of all-trans α-carotene, antheraxanthin, astaxanthin, β-carotene, β-cryptoxanthin, canthaxanthin, capsanthin, echinenone, lutein, lycopene, neoxanthin, neurosporene, phytoene, phytofluene, rubixanthin, violaxanthin and zeaxanthin were obtained from ZINC database (Irwin and Shoichet, 2005) as mol2 format and converted into MOPAC input format by Openbabel (O'Boyle et al., 2011).

3.2.3. Quantum chemical calculations

Semi-empirical PM6 (Stewart, 2007) and PM7 (Stewart, 2013) quantum chemical calculations were done by MOPAC2012 (Stewart, 2014). The heat of formation of structurally optimised carotenoid molecules was retrieved from MOPAC output files. The quantum chemical properties of ionisation energy (***I***), electron affinity (***A***), chemical hardness (***η***), electronegativity (***χ***), chemical potential (**μ**) and electrophilicity (***ω***) were calculated as following equations:

$$\boldsymbol{I = E(N-1) - E(N)} \quad (3.1)$$

$$\boldsymbol{A = E(N) - E(N+1)} \quad (3.2)$$

$$\boldsymbol{\eta = (I - A)/2} \quad (3.3)$$

$$\boldsymbol{\chi = -\mu = (I + A)/2} \quad (3.4)$$

$$\boldsymbol{\omega = \mu^2/2\eta} \quad (3.5)$$

where $\boldsymbol{E(N)}$ is energy of neutral carotenoid molecule, $\boldsymbol{E(N-1)}$ is energy of monovalent carotenoid cation and $\boldsymbol{E(N+1)}$ is energy of monovalent anion.

Likewise of the quantum chemical properties of neutral carotenoid molecules, the ionisation energy (I_{cat}), electron affinity (A_{cat}), chemical hardness (η_{cat}), electronegativity (χ_{cat}), chemical potential (μ_{cat}) and electrophilicity (ω_{cat}) of monovalent cationic carotenoid molecules were calculated. To analyse cross-level effects between neutral and cationic carotenoid molecules, the cross-product terms of quantum chemical property values of neutral carotenoid molecules and those of cationic carotenoid molecules were calculated and abbreviated as ionisation energy (I_{cross}), electron affinity (A_{cross}), chemical hardness (η_{cross}), electronegativity (χ_{cross}), chemical potential (μ_{cross}) and electrophilicity (ω_{cross}).

3.2.4. Quantitative structure-activity relationship

3.2.4.1. Correlation analysis

Prior to developing QSAR models, Pearson's correlation analysis between quantum chemical properties and antioxidant activities of carotenoids was done by using GNU R (http://cran.r-project.org/). Quantum chemical properties highly correlated with antioxidant activities were selected as independent variables for QSAR models.

3.2.4.2. Adaptive neuro-fuzzy inference system

In order to develop QSAR models, ANFIS was applied using Matlab 8.2 (Mathworks, Natick, MA, USA). From the correlation analysis, I, I_{cat} and

$\boldsymbol{\mu_{cross}}$ were chosen as independent variables for QSAR models. Each independent variable was standardised to be adjusted to the same scale by following equation:

$$\boldsymbol{x_i^{std}} = \frac{x_i - \bar{x}}{\sum_{i=1}^{N}(x_i - \bar{x})^2 / N} \tag{3.6}$$

where $\boldsymbol{x_i^{std}}$ is standardised value of the sample, $\boldsymbol{x_i}$ is original value of the sample, $\boldsymbol{\bar{x}}$ is the mean of each variable and $\boldsymbol{N}$ is the number of the samples in the data set.

Two triangular-shaped membership functions were applied for each independent variable, 8 if-then rules and 8 linear type output functions were applied for ANFIS. To train and optimise ANFIS models, back-propagation method was used. The structure of ANFIS is illustrated on Figure 3.2. To validate constructed QSAR models, 1000 times of bootstrap resampling validation procedure were applied. To measure prediction efficiency, mean absolute error (MAE) was calculated as follows:

$$\mathbf{MAE} = \frac{1}{N}\sum_{i=1}^{n} |y'_i - y_i| \tag{3.7}$$

where $\boldsymbol{y'_i}$ is predicted value resulted from QSAR model and $\boldsymbol{y_i}$ is experimental value from the literature.

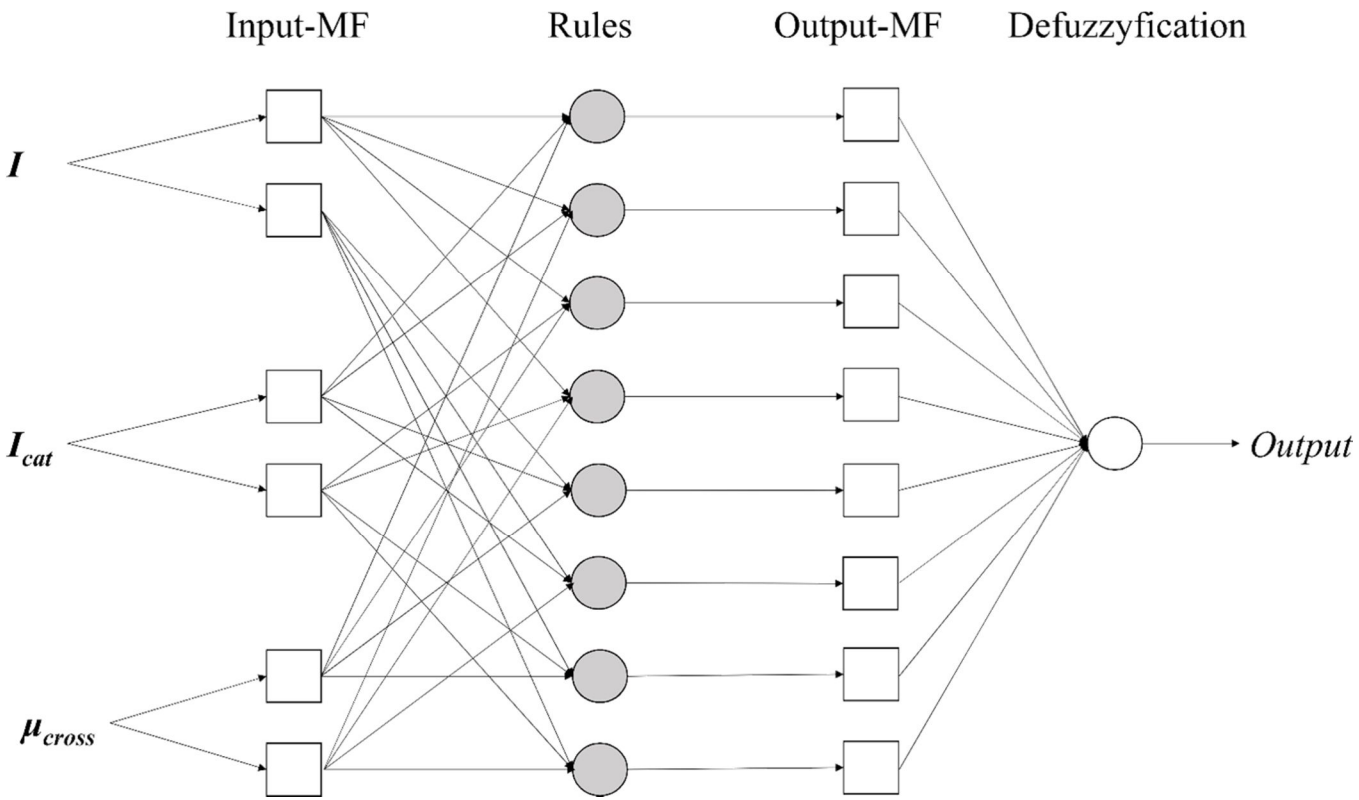

Figure 3.2. Developed adaptive neuro-fuzzy inference system (ANFIS) structure.

3.3. Results and discussion

The molecular structures and the list of carotenoids adapted from the study of Müller et al. (2011) are illustrated in Figure 3.1. Their calculated quantum chemical properties by PM6 method are presented in Table 3.1. The quantum chemical descriptors of monovalent cationic carotenoid molecules were also calculated because both of the neutral and monovalent cationic carotenoid molecules exist at the same time during the radical scavenging reaction (Jeevarajan et al., 1994; Mairanovsky et al., 1975). In addition, the cross-product terms should be calculated and introduced including interaction effects between neutral and cationic carotenoid molecules on QSAR models.

Pearson's correlation coefficients between quantum chemical properties calculated by PM6 and PM7 methods and the TEAC values of carotenoids from Müller et al. (2011) are presented in Table 3.2. The correlations analysed by both of the PM6 and PM7 methods showed identical tendencies. The negative correlation between calculated $\boldsymbol{I}$ values and TEAC values could be explained in physicochemical manner. Due to conjugation of double bonds on the long carbon chain skeleton of carotenoids, carotenoids oxidise easily via electron transfer mechanism (Britton, 1995; El-Agamey et al., 2004; Mortensen et al., 2001; Stahl and Sies, 2005). Since electron transfer is a main mechanism of antioxidant activity of carotenoids, carotenoids with lower $\boldsymbol{I}$ are expected to have higher antioxidant capacity than carotenoids with higher $\boldsymbol{I}$. Not only the $\boldsymbol{I}$ but also the $\boldsymbol{I}_{cat}$ had a significant correlation with the TEAC.

Table 3.1. Molecular quantum chemical properties of carotenoids calculated by PM6 semi-empirical quantum chemistry method

	[a]I	A	η	μ	χ	ω	[b]I_{cat}	A_{cat}	η_{cat}	μ_{cat}	χ_{cat}	ω_{cat}
α-Carotene	142.76	49.83	46.46	-96.29	96.29	99.78	190.12	142.76	23.68	-166.44	166.44	584.83
Antheraxanthin	145.03	51.84	46.59	-98.44	98.44	103.98	194.18	145.03	24.57	-169.60	169.60	585.35
Astaxanthin	151.37	60.29	45.54	-105.83	105.83	122.98	203.93	151.37	26.28	-177.65	177.65	600.41
β-Carotene	142.96	49.00	46.98	-95.98	95.98	98.05	185.09	142.96	21.06	-164.02	164.02	638.68
β-Cryptoxanthin	142.91	50.20	46.35	-96.55	96.55	100.57	187.17	142.91	22.13	-165.04	165.04	615.32
Canthaxanthin	148.87	55.19	46.84	-102.03	102.03	111.12	199.23	148.87	25.18	-174.05	174.05	601.49
Capsanthin	146.36	62.30	42.03	-104.33	104.33	129.50	199.33	146.36	26.49	-172.84	172.84	563.97
Echinenone	143.88	53.38	45.25	-98.63	98.63	107.50	193.72	143.88	24.92	-168.80	168.80	571.77
Lutein	144.74	55.52	44.61	-100.13	100.13	112.38	192.37	144.74	23.81	-168.55	168.55	596.61
Lycopene	139.28	53.45	42.92	-96.36	96.36	108.19	183.63	139.28	22.18	-161.46	161.46	587.78
Neoxanthin	145.50	54.32	45.59	-99.91	99.91	109.47	197.76	145.50	26.13	-171.63	171.63	563.63
Neurosporene	142.19	48.31	46.94	-95.25	95.25	96.64	192.75	142.19	25.28	-167.47	167.47	554.72
Phytoene	160.00	19.10	70.45	-89.55	89.55	56.92	248.57	160.00	44.29	-204.28	204.28	471.15
Phytofluene	151.72	33.62	59.05	-92.67	92.67	72.71	222.44	151.72	35.36	-187.08	187.08	494.86
Rubixanthin	140.67	52.91	43.88	-96.79	96.79	106.75	187.20	140.67	23.26	-163.94	163.94	577.59
Violaxanthin	147.79	52.65	47.57	-100.22	100.22	105.56	199.75	147.79	25.98	-173.77	173.77	581.09
Zeaxanthin	144.51	51.03	46.74	-97.77	97.77	102.26	188.52	144.51	22.01	-166.51	166.51	629.94

(kcal/mol).

[a] ***I, A,*** **η, μ, χ** and **ω** are ionisation energy, electron affinity, chemical hardness, chemical potential, electronegativity and electrophilicity of neutral carotenoid, respectively.

[b] I_{cat}, A_{cat}, η_{cat}, μ_{cat}, χ_{cat} and ω_{cat} are ionisation energy, electron affinity, chemical hardness, chemical potential, electronegativity and electrophilicity of cationic carotenoid, respectively.

Table 3.2. Pearson's correlation coefficients between quantum chemical descriptors and radical scavenging activities of carotenoids calculated by PM6 and PM7 methods.

	PM6			PM7		
	Neutral	**Cation**	**Product**	**Neutral**	**Cation**	**Product**
Ionisation energy	-0.785***	-0.653**	-0.683**	-0.732***	-0.632**	-0.650**
Electron affinity	0.211	-0.785***	0.099	0.245	-0.732***	0.150
Chemical hardness	-0.440	-0.572*	-0.493*	-0.434	-0.581*	-0.501*
Chemical potential	0.226	0.692**	0.847***	0.143	0.659**	0.830***
Electronegativity	-0.226	-0.692**	-0.847***	-0.143	-0.659**	-0.830***
Electrophilicity	0.148	0.416	0.226	0.200	0.570*	0.346

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

This result postulates a radical scavenging mechanism of carotenoids that two electrons of a single carotenoid molecule transfer to radical species to scavenge them. Previous studies also reported that carotenoids undergo sequential loss of two electrons in oxidation reaction (Jeevarajan and Kispert, 1996; Jeevarajan et al., 1994; Mairanovsky et al., 1975). For the reason that the $\boldsymbol{I}$ was lower than the $\boldsymbol{I}_{cat}$ (Table 3.1) and the $\boldsymbol{I}$ was more significantly correlated with the TEAC compared to the $\boldsymbol{I}_{cat}$ (Table 3.2), electron transfer process from neutral carotenoid molecule to radical occurs more favourably than that from monovalent carotenoid cation. Also, a significantly high correlation between $\boldsymbol{\mu}_{cross}$ ($-\boldsymbol{\chi}_{cross}$) and TEAC values ($p<0.001$) could be noted, but lower significances were observed in the cases of $\boldsymbol{\mu}$ and $\boldsymbol{\mu}_{cation}$ (Table 3.2). This result postulates that antioxidant activities of carotenoids were mainly contributed by both of the neutral and cationic carotenoid molecules rather than either of them. Since the chemical potential (negative electronegativity) is a thermodynamic property derived by differentiating the energy with

respect to the number of electrons (Parr et al., 1978), it indicates the direction of chemical reactions as a partial free energy. Thus, a positive relationship was observed between the chemical potential and the TEAC. The postulated mechanism of radical scavenging activities of carotenoids with limited information from correlation analysis is related with the sequential donation of two electrons to radical species. Chemical potential, $\boldsymbol{\mu}$, as a driving force for electron transfer (Parr et al., 1978) is responsible for the first electron transfer reaction and $\boldsymbol{\mu}_{cat}$ is responsible for the second electron transfer reaction to radical species. The electron transfer occurs not only between the carotenoids and radical species, but also between the neutral and cationic carotenoids (Edge et al., 1998; Jeevarajan and Kispert, 1996). The significant relationship between $\boldsymbol{\mu}_{cross}$ and TEAC was supposed to be caused by the interaction effect between the neutral and cationic carotenoids. In this physicochemical manner, the quantum chemical descriptor of chemical potential could be applied on QSAR studies. Worachartcheewan et al. (2011) introduced chemical potential to predict radical scavenging activities of curcuminoids. Pasha et al. (2008) developed QSAR model for radical scavenging activities of flavonoids with electronegativity as a dependent variable.

From the result of correlation analysis, $\boldsymbol{I}$, $\boldsymbol{I}_{cat}$ and $\boldsymbol{\mu}_{cross}$ were selected as dependent variables for developing QSAR models, for reason that these properties had a significant correlation with antioxidant activities of carotenoids. Most of previous QSAR studies developed prediction models by

linear regression (Amić and Lučić, 2010; Kontogiorgis et al., 2005; Pasha et al., 2008; Sarkar et al., 2011; Soffers et al., 1999). However, traditional QSAR models by linear regression analysis ignore the interaction effects between dependent variables; thus, this ignorance may have effects on the prediction efficiency of the model. The ANN-applied techniques could be an alternative to conventional linear regression model. ANN-based models are constructed and adjusted by empirical training process. It is useful to solve and analyse the problems which could not be solved easily with traditional linear regression analysis. Especially, ANN-based modelling is useful to analyse multivariate nonlinear relationship (Gemperline et al., 1991). For this reason, a higher prediction efficiency could be achieved by applying empirical training-based ANN modelling technique on predicting bioactivities of molecules (Agatonovic-Kustrin and Beresford, 2000). In this study, ANFIS, an ANN-based system, was used to achieve high prediction efficiency of QSAR models. At first, the ANFIS-applied QSAR models were developed using a data set from Müller et al. (2011). The TEAC values were predicted by the developed QSAR models (Table 3.3 and Figure 3.3). Because carotenoids were limited to a variety of naturally occurring and commercially available ones, data sets with a relatively small sample size were available to be analysed. Bootstrap validation is an appropriate method to validate models with a small sample size (Braga-Neto and Dougherty, 2004; Kiralj and Ferreira, 2009). Thus, in this study, MAE and Q square values of developed models were estimated by the bootstrap method. Both of the QSAR models

with quantum chemical properties calculated by the PM6 and PM7 semi-empirical methods had high prediction efficiencies.

Table 3.3. Experimental and predicted TEAC of carotenoids.

Compounds	Experimental TEAC	Predicted TEAC	
		PM6	PM7
α-Carotene	3.3	2.79	2.83
Antheraxanthin	2.0	1.80	1.71
Astaxanthin	0.8	0.87	0.92
β-Carotene	3.1	3.15	3.21
β-Cryptoxanthin	3.2	2.97	3.01
Canthaxanthin	0.8	1.11	1.24
Capsanthin	2.0	1.63	1.51
Echinenone	2.2	2.10	2.11
Lutein	2.0	2.01	1.86
Lycopene	3.9	3.89	3.89
Neoxanthin	1.1	1.60	1.68
Neurosporene	2.1	2.55	2.51
Phytoene	1.0	1.00	1.00
Phytofluene	1.0	0.99	0.97
Rubixanthin	3.3	3.35	3.25
Violaxanthin	1.6	1.21	1.32
Zeaxanthin	1.9	2.30	2.34
Mean absolute error		0.22 ± 0.04	0.24 ± 0.05
Q square		0.905±0.043	0.895±0.049

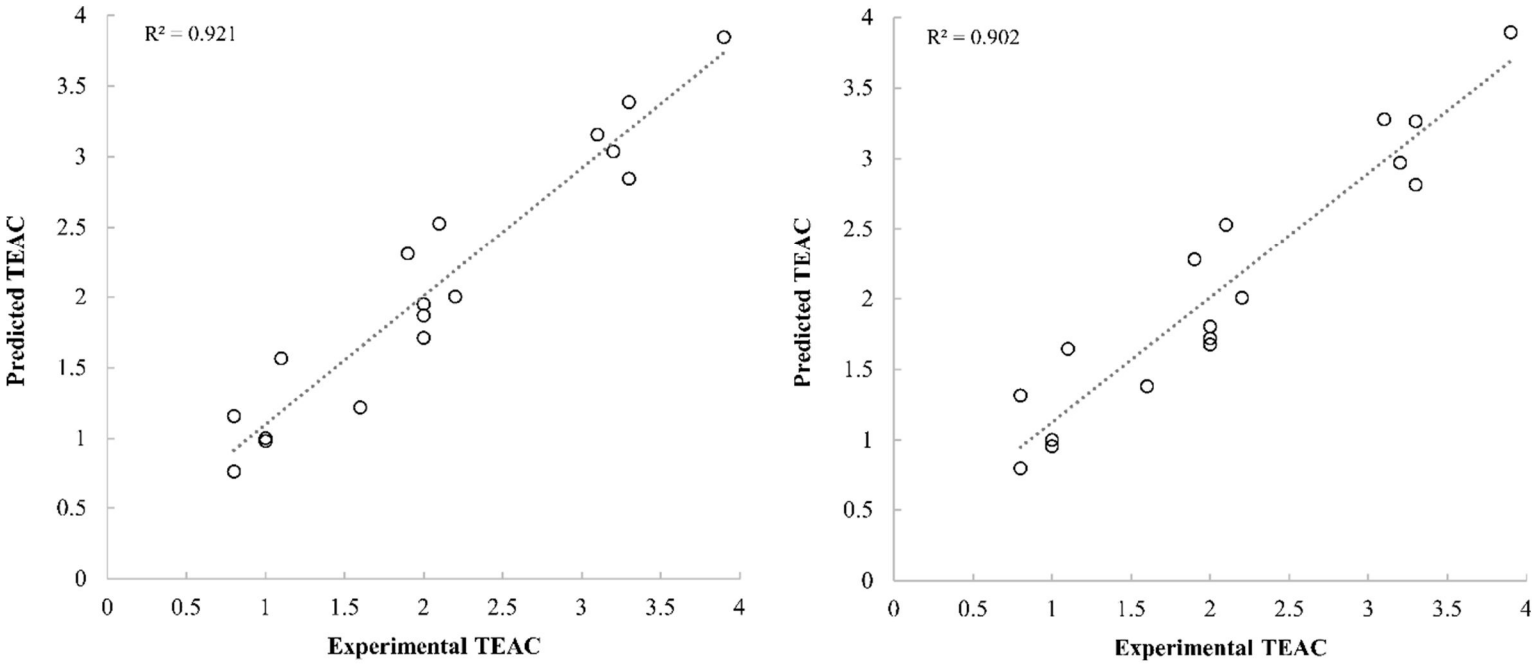

Figure 3.3. Scatter plot between the experimental TEAC of carotenoids reported by Müller et al. (2011) and the predicted TEAC based on the quantum chemical descriptors calculated by (A) PM6 and (B) PM7 methods.

To verify the reliabilities of the selected quantum chemical properties (i.e., $\boldsymbol{I}$, $\boldsymbol{I_{cat}}$ and $\boldsymbol{\mu_{cross}}$) and the design of developed QSAR models, QSAR models were developed with radical scavenging activity data sets from other previous studies. TEAC values of 8 carotenoid species were adapted from Miller et al. (1996) and EC50 data of 6 carotenoid species were adapted from Jiménez-Escrig et al. (2000). The results from correlation analysis of these data sets are presented in Table 3.4. The $\boldsymbol{I}$, $\boldsymbol{I_{cat}}$ and $\boldsymbol{\mu_{cross}}$ of carotenoids calculated by both of PM6 and PM7 were significantly correlated with the TEAC values from Miller et al. at $p < 0.001$. Although statistical significances were low on the correlation between quantum chemical properties of carotenoids and the experimental EC50 due to a small sample size of the data set by Jiménez-Escrig et al. (n=6), positive correlations were observed between the selected quantum chemical properties of carotenoids and the EC50 A previous QSAR study regarding radical scavenging activities of carotenoids by Soffers et al. (1999) reported a positive relationship between ionisation energies and TEAC values using single data set constituted of 9 carotenoids. In addition to the previous study, this positive tendency was also verified using several number of data sets in this study. Kleinová et al. (2007) also constructed a QSAR model for electrochemical redox potentials of carotenoids using polarizability as an independent variable. Since the polarizability was closely correlated with ionisation energy (Beg et al., 2013), the use of polarizability as an independent variable seemed adequate. However, based on electrochemical properties of carotenoids, the results from the previous study could not be

Table 3.4. Pearson's correlation coefficients between quantum chemical descriptors and radical scavenging activities of carotenoids calculated by PM6 and PM7 methods.

	TEAC by Miller et al. (1996)		EC50 by Jiménez-Escrig et al. (2000)	
	PM6	**PM7**	**PM6**	**PM7**
I	-0.923***	-0.917***	0.892*	0.850*
I_{cat}	-0.937***	-0.942***	0.724	0.812*
μ_{cross}	-0.892***	-0.914***	0.830*	0.802

*p < 0.05, **p < 0.01, ***p < 0.001
I, ionisation energy of neutral carotenoid.
I_{cat}, ionisation energy of cationic carotenoid.
μ_{cross}, product of chemical potential of neutral and monovalent cationic carotenoid.

directly interpreted to radical scavenging activities of carotenoids. In the present study, the reliabilities of the selected quantum chemical properties for predicting radical scavenging activities and the prediction efficiency of ANFIS-applied QSAR models were confirmed by applying various data sets (Figure 3.4 and 3.5).

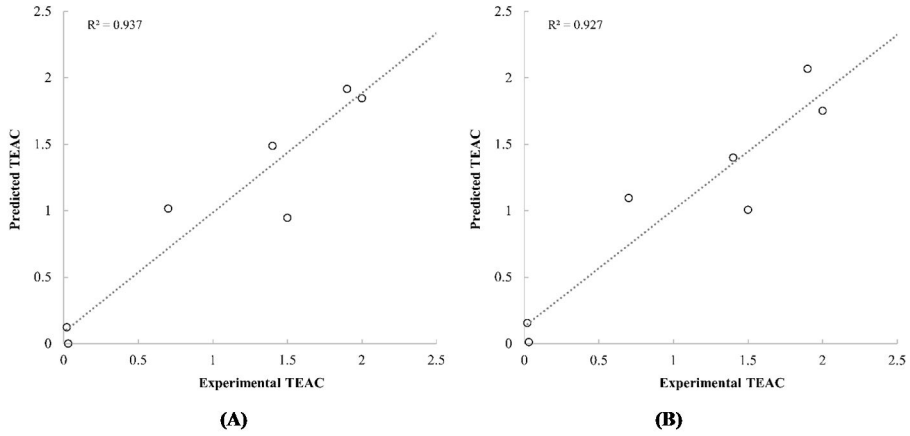

Figure 3.4. Scatter plot between the experimental TEAC of carotenoids reported by Miller et al. (1996) and the predicted TEAC based on the quantum chemical descriptors calculated by (A) PM6 and (B) PM7 methods.

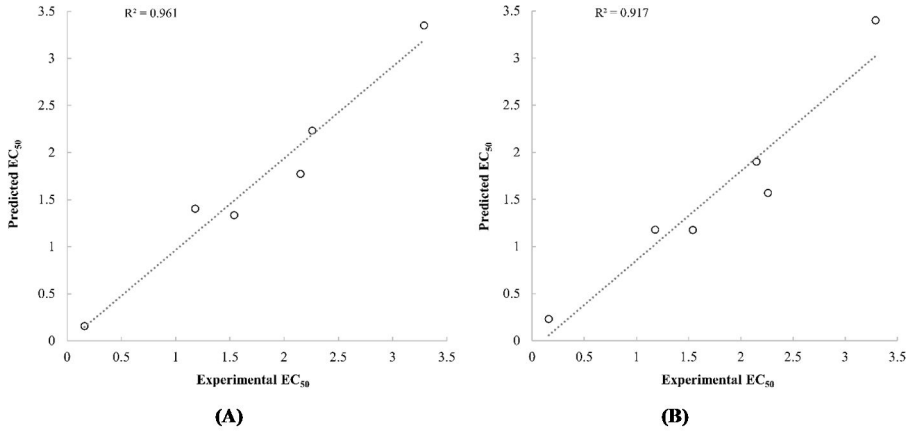

Figure 3.5. Scatter plot between the experimental EC50 of carotenoids reported by Jiménez-Escrig et al. (2000) and the predicted TEAC based on the quantum chemical descriptors calculated by (A) PM6 and (B) PM7 methods.

Chapter 4

Adaptive Neuro-Fuzzy Inference System-Applied QSAR with Bond Dissociation Energy for Predicting Antioxidant Activities of Phenolic Compounds

(Study 3)

4.1. Introduction

Antioxidants are chemical substances that inhibit oxidation promoted by radical species such as oxygen, peroxides, nitric oxide and peroxynitrite (Evans and Halliwell, 2001; Fridovich, 1978). Since dietary antioxidants possibly prevent coronary heart diseases, cancer and Alzheimer's disease, antioxidants play a beneficial role for health (Asplund, 2002; Crichton et al., 2013). Some of phytochemicals, which are chemical compounds existing naturally in plants, have antioxidantive characteristics as a part of plant antioxidant systems to control external and internal oxidative stress (Blokhina et al., 2003). Contributed by phytochemicals, traditional Chinese medicinal plants have antioxidant activity (Cai et al., 2004; Liao et al., 2008). The antioxidant activities of phytochemicals are mainly contributed by phenolic compounds including phenolic acids, flavonoids and phenolic diterpenes (Brewer, 2011). There are two major mechanisms for deactivating radical species by antioxidants. One is hydrogen atom transfer (HAT) and the other is single electron transfer (SET) (Prior et al., 2005). The radical scavenging activity by HAT mechanism is explained by following reaction:

$$\mathbf{X\cdot + AOH \rightarrow XH + AO} \qquad (4.1)$$

where X represents a free radical and AOH represents an antioxidant molecule and its hydroxyl group. As shown on the equation 4.1, single hydrogen atom should be homolytic dissociated from the antioxidant molecule to scavenge radical by HAT. Thus, the bond dissociation energy of O-H bond (BDE) is a crucial variable for estimating antioxidant activity.

The SET mechanism is explained by following equation:

$$\mathbf{X\cdot + AOH \rightarrow X^{-} + AOH^{\cdot +}} \quad (4.2)$$

In the case of SET, the mechanism is governed by ionisation energy which is the energy needed to dissociate an electron from a molecule.

To predict antioxidant activities of phytochemicals, numerous structure-activity relationship (SAR) and quantitative structure-activity relationship (QSAR) studies have been conducted (Amić and Lučić, 2010; Chang et al., 2008; Cruciani et al., 1990; Filipović et al., 2015; Kontogiorgis et al., 2005; Sarkar et al., 2011). Most of previous QSAR studies applied linear regression method for developing QSAR models to predict activities of compounds. While the QSAR studies applying adaptive neuro-fuzzy inference system (ANFIS), which is an artificial neural network (ANN-based fuzzy inference system, achieved high prediction accuracies (Jhin and Hwang, 2014, 2015). Besides, antioxidative characteristics of phenolic compounds, which have common skeleton structures, were analysed in previous QSAR studies. However, to compare the antioxidant activities of compounds, which belong to different classes, an integrated single QSAR model is needed. Therefore, for high prediction accuracy and comparative analysis of various phenolic compounds, ANFIS-applied QSAR models for predicting antioxidant activities of phenolic compounds were developed in this study.

4.2. Materials and methods

4.2.1. Data set

The trolox equivalent antioxidant capacity values of traditional Chinese medicinal plants derived phenolic compounds against 2,2'-azino-bis(3-ethylbenzothiazoline-6-sulphonic acid) ($TEAC_{ABTS}$) and 2,2-diphenyl-1-picrylhydrazyl ($TEAC_{DPPH}$) were adapted from a previous study by Cai et al. (2006). The data set was constituted of total 100 compounds which belong to various classes such as phenolic acids, flavonoids, tannins, stilbenes, coumarins and quinones (Figure 4.1). Carthamin and Chinese gallotannin were excluded from the data set, because their exact structures could not be deduced as these compounds from the previous study had various forms. Since the BDE of hydroxyl group was used as a descriptor for QSAR models, compounds with no hydroxyl group were also excluded from the data set. Therefore, a data set constituted of 88 compounds was used for this study (Table 4.1).

4.2.2. Bond dissociation energy calculation

The chemical structures of compounds were retrieved as SDF format from PubChem database (Wang et al., 2009). Three-dimensional structure of each of the retrieved molecules was generated and converted into MOPAC input format by Open Babel (O'Boyle et al., 2011). Molecular structures were

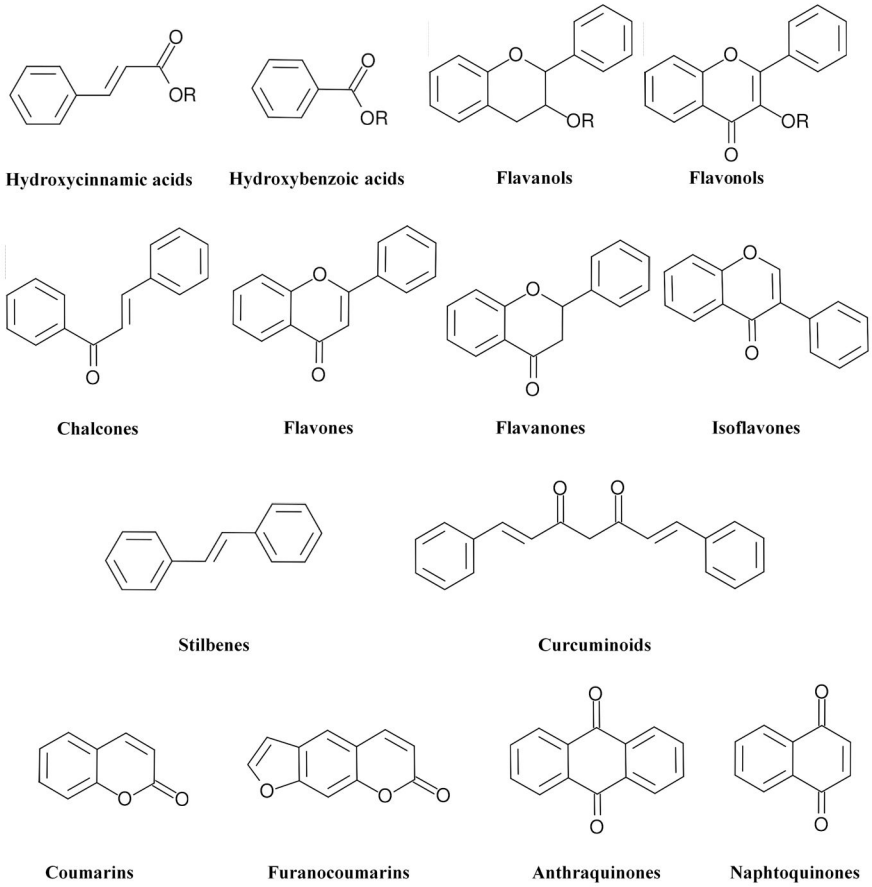

Figure 4.1.The skeleton structures of phenolic compounds existing in traditional Chinese medicinal plants, tested by Cai et al. (2006).

Table 4.1. Antioxidant activities and the lowest bond dissociation energies (BDE_{min}) of phenolic compounds.

Entry	Compounds	[a] $TEAC_{ABTS}$	[a] $TEAC_{DPPH}$	BDE_{min} (kcal/mol)	
				PM6	PM7
1	Caffeic acid	1.31	1.24	73.45	76.19
2	Chlorogenic acid	1.56	1.75	73.21	75.87
3	*o*-Coumaric acid	0.93	0.84	78.79	81.48
4	*m*-Coumaric acid	0.82	0.75	81.39	83.58
5	*p*-Coumaric acid	1.96	1.44	79.98	82.41
6	Ferulic acid	1.92	1.49	72.35	75.09
7	Isoferulic acid	1.53	1.24	74.10	76.54
8	Gallic acid	3.52	3.92	75.14	76.94
9	Protocatechuic acid	1.15	1.29	75.04	77.34
10	2,4-Dihydroxybenzoic acid	1.22	1.27	82.35	83.48
11	*o*-Hydroxybenzoic acid	0.04	0.05	79.71	81.58
12	*m*-Hydroxybenzoic acid	0.03	0.07	81.63	83.80
13	*p*-Hydroxybenzoic acid	0.03	0.06	83.75	85.35
14	Syringic acid	1.39	1.33	74.01	75.18
15	Vanillic acid	0.09	0.06	78.74	80.20
16	(-)-Epigallocatechin gallate	5.95	6.09	74.27	76.13
17	(-)-Epicatechin gallate	5.29	5.26	73.33	75.88
18	(-)-Epigallocatechin	3.71	3.56	74.09	75.80
19	(-)-Epicatechin	3.08	3.18	73.85	76.03
20	(+)-Catechin	3.04	2.95	73.26	75.98
21	Myricetin	1.31	1.38	71.67	74.37

Table 4.1. (continued)

Entry	Compounds	[a] $TEAC_{ABTS}$	[a] $TEAC_{DPPH}$	BDE_{min} (kcal/mol)	
				PM6	PM7
22	Quercetin	4.42	4.60	71.22	74.55
23	Morin	2.68	2.75	77.55	77.06
24	Kaempferol	1.59	1.32	76.36	76.35
25	Galangin	1.12	0.71	76.25	76.58
26	Quercetin-3-glucoside	2.39	2.16	72.39	75.14
27	Quercetin-3-rutinoside	2.02	2.33	70.29	74.00
28	Quercetin-3-rhamnoside	2.18	2.57	71.42	74.39
29	Kaempferol-3-glucoside	0.14	0.15	75.45	82.21
30	Quercetin-3-glucoside-7-rhamnoside	1.56	1.63	71.91	73.60
31	Flavonol	0.71	0.48	76.06	75.79
32	Butein	2.42	2.27	72.76	75.74
33	Phloretin	1.79	1.24	78.35	80.86
34	Sappanchalcone	1.93	1.82	72.41	74.59
35	Luteolin	2.18	2.24	74.68	77.09
36	Baicalein	2.56	2.74	76.96	77.05
37	Apigenin	0.09	0.04	82.02	83.75
38	Chrysin	0.08	0.05	92.31	91.39
39	Luteolin-7-glucoside	1.47	1.39	74.35	76.21
40	Apigenin-8-glucoside	0.22	0.21	81.95	83.64
41	Apigenin-7-glucoside	0.08	0.05	81.26	83.09
42	Baicalein-7-glucuronide	1.55	1.79	66.83	74.63
43	Naringenin	0.22	0.14	80.84	82.79
44	Hesperetin	0.41	0.27	74.02	76.31

Table 4.1. (continued)

Entry	Compounds	[a] $TEAC_{ABTS}$	[a] $TEAC_{DPPH}$	BDE_{min} (kcal/mol)	
				PM6	PM7
45	Naringenin-7-rutinoside	0.10	0.08	76.82	82.89
46	Hesperetin-7-rutinoside	0.10	0.08	73.72	89.64
47	Genistein	0.12	0.10	78.02	80.67
48	Daidzein	0.10	0.03	78.11	80.71
49	Glycitein	0.10	0.02	74.48	76.87
50	Genistein-7-glucoside	0.08	0.03	78.33	80.77
51	Daidzein-7-glucoside	0.07	0.04	78.17	80.60
52	Catechin 3-O-gallate	5.25	5.56	73.49	76.00
53	Procyanidin B-1	6.14	5.94	73.01	75.21
54	Procyanidin B-2-digallate	9.18	8.79	72.40	74.70
55	Procyanidin C-1	8.29	7.93	72.90	74.72
56	Corilagin	7.76	6.98	69.70	71.53
57	Piceatannol	2.53	2.35	71.28	74.07
58	Resveratrol	2.14	1.71	77.34	80.15
59	Piceatannol-3'-glucoside	1.62	1.55	74.15	78.09
60	Resveratrol-3-glucoside	1.35	1.21	77.30	79.92
61	Resveratrol-4'-glucoside	0.56	0.62	83.87	85.06
62	Curcumin	2.24	2.02	71.72	74.84
63	Demethoxycurcumin	1.63	1.48	71.24	74.36
64	Bisdemethoxycurcumin	1.18	1.02	78.14	80.69
65	Esculetin	2.38	2.08	74.68	76.72
66	Scopoletin	0.38	0.21	74.32	76.67
67	Esculetin-6-glucoside	0.16	0.17	80.15	79.24

Table 4.1. (continued)

Entry	Compounds	[a] $TEAC_{ABTS}$	[a] $TEAC_{DPPH}$	BDE_{min} (kcal/mol)	
				PM6	PM7
68	Secoisolariciresinol	0.31	0.18	69.88	74.14
69	Matairesinol	0.25	0.21	69.66	75.34
70	Arctigenin	0.10	0.09	70.27	75.27
71	Magnolol	0.21	0.20	75.31	79.68
72	Purpurin	1.93	2.00	74.27	76.07
73	Pseudopurpurin	1.62	1.70	76.66	79.64
74	Alizarin	1.07	1.05	75.19	76.79
75	Quinizarin	0.55	0.04	78.46	79.33
76	Emodin	0.10	0.03	84.61	85.19
77	Chrysazin	0.07	0.06	82.94	83.76
78	Rhein	0.08	0.06	93.15	94.90
79	Chrysophanol	0.07	0.04	93.13	94.87
80	Physcion	0.07	0.07	94.18	95.56
81	Aloe-emodin	0.08	0.04	92.38	94.81
82	1,5-Dihydroxyanthraquinone	0.08	0.07	83.48	84.36
83	2,6-Dihydroxyanthraquinone	0.07	0.08	83.22	84.55
84	Ruberythric acid	0.07	0.07	86.18	87.69
85	Alizarin-2-glucoside	0.11	0.10	87.47	89.43
86	Juglone	0.11	0.09	83.54	84.70
87	Shikonin	0.12	0.08	77.74	79.66
88	Acetylshikonin	0.11	0.08	77.09	78.81

[a] Trolox equivalent antioxidant capacity values against ABTS ($TEAC_{ABTS}$) and DPPH ($TEAC_{DPPH}$) were adapted from Cai et al. (2006).

geometrically optimised and heat of formations of compounds were calculated by PM6 (Stewart, 2007) and PM7 (Stewart, 2013) semi-empirical quantum chemical calculation methods by MOPAC 2012 (Stewart, 2014). Standard heat of formation value ($\mathbf{\Delta H}_f$) of each molecule was retrieved from MOPAC output files of semi-empirical calculations. Bond dissociation energies of O-H bond (BDE) was calculated from the following equation:

$$\text{BDE} = \mathbf{\Delta H}_f(\text{AO}\cdot) + \mathbf{\Delta H}_f(\text{H}) - \mathbf{\Delta H}_f(\text{AOH}) \quad (4.3)$$

where $\mathbf{\Delta H}_f(\text{AO}\cdot)$, $\mathbf{\Delta H}_f(\text{H})$ and $\mathbf{\Delta H}_f(\text{AOH})$ are heat of formations of a hydrogen abstracted oxy radical of a phenolic compound, a hydrogen atom and a phenolic compound molecule, respectively. The BDE values of enol and phenol groups were calculated for every phenolic compound in the data set, and the lowest BDE value among the BDE of each compound was designated as BDE_{min}.

4.2.3. Quantitative structure-activity relationship

Sum of reciprocals of BDE (X_{BDE}) was calculated from the following equation:

$$X_{BDE} = \sum_{i=1}^{n} \frac{1}{\text{BDE}_i} \quad (4.4)$$

where n is the number of enol and phenol groups. The calculated descriptor, X_{BDE}, was used as a parameter for QSAR models. To investigate the relationship between antioxidant activities and descriptors, Pearson's correlation analysis was conducted by GNU R (http://cran.r-project.org/).

Linear regression and ANFIS-applied QSAR models were developed using Matlab R2016a (Mathworks, Natick, MA, USA). Two Gaussian curve membership functions were applied for each ANFIS-QSAR model. The models were optimised by hybrid training method embedded in Matlab fuzzy logic toolbox. To validate QSAR models, leave-one-out cross-validation (LOOCV) procedure was performed for each developed model. Mean absolute error (MAE) was calculated from the following equation:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y'_i - y_i\right| \quad (4.5)$$

where y'_i is calculated radical scavenging activity value and y_i is observed value.

4.3. Results and discussion

The data set from Cai et al. (2006) contained radical scavenging activities of phenolic compounds by two different radical scavenging activity assays, ABTS assay and DPPH assay. Consistent tendency with high correlation (R=0.995) was observed on hierarchical orders of radical scavenging activities for the 88 phenolic compounds against ABTS and DPPH. High correlation between ABTS and DPPH radical scavenging activities were observed in other studies (Dudonne et al., 2009; Floegel et al., 2011; Thaipong et al., 2006), and this correlation was due to a common radical scavenging mechanism of ABTS and DPPH assays (Karadag et al., 2009). HAT mechanism plays a key role in antioxidant activities of phenolic compounds (Karadag et al., 2009; Wright et al., 2001). Physicochemically,

radical scavenging by HAT mechanism occurs more easily with compounds having lower BDE, because a hydrogen atom can be abstracted easily on a molecule with a lower BDE (Trouillas et al., 2006; Zhang and Wang, 2002). In this manner, BDE was used as a parameter for QSAR and SAR studies. Amić and Lučić (2010) reported a negative correlation between BDE_{min} and radical scavenging activity of flavonoids. In this study, a significant negative correlation was also observed between BDE_{min} and radical scavenging activities of phenolic compounds (Table 4.2). In addition, since the number of hydroxyl groups was positively correlated with antioxidant activities, the number of hydroxyl groups was used as a parameter for QSAR models in previous studies (Anouar, 2014; Chang et al., 2008; Filipović et al., 2015; Jhin and Hwang, 2014; Lien et al., 1999). In the study by Amić and Lučić (2010), as well as BDE_{min}, the number of hydroxyl groups was also used as a structural descriptor for QSAR models. The positive correlation between the number of hydroxyl groups and antioxidant activities postulates that not only single hydroxyl group corresponding to BDE_{min}, but also other hydroxyl

Table 4.2. Pearson's correlation coefficients between bond dissociation energy (BDE)-related descriptors and antioxidant activities.

	[a] BDE_{min}		[b] X_{BDE}	
	PM6	**PM7**	**PM6**	**PM7**
$TEAC_{ABTS}$	-0.469^{***}	-0.524^{***}	0.907^{***}	0.909^{***}
$TEAC_{DPPH}$	-0.470^{***}	-0.522^{***}	0.906^{***}	0.906^{***}

$^{***}p < 0.001$

[a] BDE_{min}: the lowest bond dissociation energy

[b] $X_{BDE} = \sum_{i=1}^{n} \frac{1}{BDE_i}$, where n is the number of phenol and enol groups

groups affect radical scavenging activities of phenolic compounds. For these reasons, in consideration of BDE values of multiple hydroxyl groups and negative correlation between BDE and antioxidant activities, X_{BDE} was calculated and used as a parameter for predicting radical scavenging activities. As mentioned above, *n* in equation 4.4 was defined as the number of enol and phenol groups. This value was different from a previous study by Amić and Lučić (2010), which introduced the number of hydroxyl groups of flavonoidal skeleton and phenol group to QSAR models. Since BDE of enol group was comparable with that of phenol group (Li et al., 2007; Wright, 2002), antioxidant activity might be contributed by HAT of enol group as well as phenol group. Meanwhile, because BDE of alkyl alcohol was higher than those of phenol and enol (Blanksby and Ellison, 2003), the number of alcohols were not included as a parameter. For these reasons, it was rational strategy for developing QSAR models with the number of enol and phenol groups. For instance, (+)-catechin and quercetin had the same number and same position of hydroxyl groups on flavanol and flavonol skeletons, respectively. BDE of hydroxyl group at 3-position of (+)-catechin and quercetin, calculated by PM6 method, were 104.81kcal/mol and 74.98kcal/mol, respectively. Since flavanol does not have a double bond in a heterocyclic ring, the 3-position-hydroxyl group of (+)-catechin was classified as alcohol group, while that of quercetin was classified as enol group. Therefore, the *n* variables for calculating X_{BDE} were designated as 4 for (+)-catechin and 5 for quercetin. Also, radical scavenging activity of quercetin

was greater than that of (+)-catechin (Table 4.1). For another example, in the case of procyanidin B-2 digallate and procyanidin C-1, the numbers of hydroxyl groups were 14 and 15, respectively, and the numbers of only enol and phenol groups were 14 and 12, respectively. The radical scavenging activity of the former was greater than the latter. These results showed that the number of enol and phenol groups was more closely related with radical scavenging activities of phenolic compounds than the number of hydroxyl groups, consequently being more appropriate for a QSAR parameter.

Pearson's correlation analysis was performed between X_{BDE} and radical scavenging activities (Table 4.2). As the absolute R values of X_{BDE} were larger than BDE_{min}, X_{BDE} was more significantly correlated with radical scavenging activities than BDE_{min}. Therefore, X_{BDE} was selected and used as a QSAR parameter for radical scavenging activities of phenolic compounds. The linear regression-applied QSAR models using X_{BDE} calculated by PM6 method were listed as follows:

$$TEAC_{ABTS} = 60.150\ (\pm 3.01)\ X_{BDE} - 0.668\ (\pm 0.140)$$

$n = 88$, $R = 0.907$, $s = 0.512$, $MAE = 0.619$, $Q = 0.901$,

$s_{cv} = 0.531$, $MAE_{cv} = 0.637$, $F(1,86) = 400.516$ (4.6)

$$TEAC_{DPPH} = 59.057\ (\pm 2.970)\ X_{BDE} - 0.695\ (\pm 0.138)$$

$n = 88$, $R = 0.906$, $s = 0.492$, $MAE = 0.623$, $Q = 0.903$,

$s_{cv} = 0.518$, $MAE_{cv} = 0.637$, $F(1,86) = 395.336$ (4.7)

The linear regression-applied QSAR models using X_{BDE} calculated by PM7 method were listed as follows.

$$TEAC_{ABTS} = 61.359\ (\pm 3.041)\ X_{BDE} - 0.663\ (\pm 0.138)$$

n = 88, R = 0.909, s = 0.502, MAE = 0.620, Q = 0.900,

s_{cv} = 0.512, MAE_{cv} = 0.641, F(1,86) = 407.261 (4.8)

$$TEAC_{DPPH} = 60.168\ (\pm 3.023)\ X_{BDE} - 0.687\ (\pm 0.138)$$

n = 88, R = 906, s = 0.491, MAE = 0.622, Q = 0.900,

s_{cv} = 0.509, MAE_{cv} = 0.640, F(1,86) = 396.226 (4.9)

Standard deviations of regression coefficients are presented in brackets. The regression coefficient of the X_{BDE} of each equation was nearly 20 times bigger than its standard error, implying that the calculated structural descriptor was significantly correlated to TEAC values obtained by ABTS and DPPH radical scavenging assays (Topliss and Edwards, 1979). Also the cross-validated regression models had good prediction efficiencies, indicating that the resulted models were not overfitted or biased.

To improve the prediction efficiency, ANFIS-applied QSAR models were developed. Numerous previous QSAR studies developed linear regression-applied prediction models (Amić and Lučić, 2010; Cruciani et al., 1990; Kontogiorgis et al., 2005; Pasha et al., 2008; Sarkar et al., 2011; Soffers et al., 1999). The linear regression is a feasible method for analysing the first-order contribution of each variable to response variable. While ANN-based prediction models are constructed and adjusted by empirical training procedures, high prediction accuracy could be achieved by ANN-applied models (Agatonovic-Kustrin and Beresford, 2000). For this reason, ANN-applied models were used for achieving high prediction accuracy in QSAR

studies (Agatonovic-Kustrin and Beresford, 2000; González-Díaz et al., 2007; Jhin and Hwang, 2014, 2015). ANFIS, an ANN-based model, was applied for achieving high prediction accuracy. The correlation coefficients of ANFIS-applied models were higher than those of linear regression models, while MAE and standard error of fits of ANFIS models were lower than those of linear regression models (Table 4.3). Namely, ANFIS-applied QSAR models had better prediction accuracies than linear regression models. Also cross-validated statistics had similar tendencies to non-cross-validated statistics; therefore, these results demonstrated that the developed ANFIS-applied models were not overfitted or biased.

,

Table 4.3. Statistical results of developed QSAR models.

Activity	QSAR Method	[a] BDE calculation method	[b] Statistical parameters					
			R	**s**	**MAE**	***Q***	$\mathbf{s_{cv}}$	$\mathbf{MAE_{cv}}$
$TEAC_{ABTS}$	Linear regression	PM6	0.907	0.512	0.619	0.901	0.531	0.637
$TEAC_{ABTS}$	Linear regression	PM7	0.909	0.502	0.620	0.903	0.518	0.637
$TEAC_{ABTS}$	ANFIS	PM6	0.919	0.473	0.588	0.911	0.491	0.614
$TEAC_{ABTS}$	ANFIS	PM7	0.919	0.472	0.584	0.913	0.490	0.606
$TEAC_{DPPH}$	Linear regression	PM6	0.906	0.492	0.623	0.900	0.512	0.641
$TEAC_{DPPH}$	Linear regression	PM7	0.906	0.491	0.622	0.900	0.509	0.640
$TEAC_{DPPH}$	ANFIS	PM6	0.918	0.489	0.582	0.910	0.514	0.607
$TEAC_{DPPH}$	ANFIS	PM7	0.919	0.485	0.582	0.911	0.510	0.604

[a] BDE: dissociation energy of O-H bond

[b] *R*: correlation coefficient; s: standard error of fit; MAE: mean absolute error; *Q*: cross-validated *R*; s_{cv}: cross-validated standard error of fit; and MAE_{cv}: cross-validated MAE.

Chapter 5

Summary and Conclusion

Although numerous previous studies performed antioxidant assays for evaluating and comparing antioxidant power of various phytochemicals, rational physicochemical investigations are needed to better understand antioxidant reaction mechanisms and predict antioxidant activities of phytochemicals and their derivatives. In this study, chemical structure-related physicochemical analyses were conducted for antioxidant activities of phytochemicals indigenous to foods. QSAR models for predicting and evaluating antioxidant activities of phytochemicals were developed and the developed models were validated.

Study 1: Semi-empirical quantum chemical calculations of anthocyanins and anthocyanidins were done by PM6 and PM7 methods using MOPAC2012. In this study, structural changes of anthocyanin molecules were considered on semi-empirical quantum chemistry calculation. This study revealed that quantum chemical descriptors of flavylium cation and quinoidal base affect radical scavenging activities of anthocyanins. It suggests that the molecular conformation should be modified depending on surrounding condition before calculating structural descriptors for QSAR analysis. Established QSAR models by ANFIS had good prediction efficiency with a statistical significance. Therefore, applying the ANFIS technique could improve the accuracy of QSAR models. In addition, this is the first study regarding anthocyanins using PM7 method which is the most recently distributed semi-empirical calculation method. This study suggests that PM7, as well as PM6,

is a useful method for calculating quantum chemical descriptors for QSAR analysis.

Study 2: Ionisation energies and chemical potentials of neutral and monovalent cationic carotenoid molecules were demonstrated as descriptors that describe radical scavenging activities of carotenoids. Although some of the previous studies reported a significant relationship between quantum chemical descriptors and radical scavenging activities of phytochemicals, any report that analyses the radical scavenging activities of carotenoids quantitatively has not been found. In addition, in this study, molecular properties of cation molecules were calculated for QSAR models as well as neutral molecules. Although the small sample sizes of data sets might weaken the significance of QSAR models, consistent tendencies, which were observed on various data sets, could demonstrate the reliabilities of the selected quantum chemical descriptors and the significance of QSAR models.

Study 3: In this study, BDE of various phenolic compounds were calculated by semi-empirical methods as structural descriptors for predicting QSAR models. It is apparent that radical scavenging activities of phenolic compounds are mainly contributed by enol groups as well as phenol groups. In most of previous QSAR studies, antioxidant characteristics of phenolic compounds containing a common skeleton structure were analysed. In this study, however, the developed QSAR models, with a new QSAR parameter,

X_{BDE}, could predict and analyse radical scavenging activities of phenolic compounds, even if the compounds did not have common skeleton structure. Therefore, applying the developed QSAR models, radical scavenging activities of phenolic compounds from different subgroups can be compared. Furthermore, the results confirmed that applying ANFIS technique on QSAR models is an appropriate strategy for achieving high prediction accuracy.

To sum up, physicochemical and quantum chemical descriptors of food phytochemicals were appropriate variables for QSAR analysis of antioxidant activities. By applying ANFIS on QSAR analysis, high prediction efficiencies could be obtained. The results from this study confirmed that applying ANFIS technique on QSAR is an appropriate strategy for achieving high prediction accuracy. The ANFIS-applied QSAR models could be used in computer-aided drug design (CADD), cheminformatics and functional food areas by predicting and analysing antioxidant activities of phytochemicals and their derivatives.

The limitation of this study is that only *in vitro*, not *in vivo*, antioxidant activities of various subgroups of phytochemicals were analysed. However, the purposes of utilising antioxidants are not only limited to exert their physiological effects, but also to enhance qualities of food products. For versatile application, therefore, fundamental chemical features of phytochemicals were mainly discussed in this study. Also, *in vitro* antioxidant assays are appropriate methods for measuring the stoichiometric characteristic

of each food phytochemical because the *in vitro* antioxidant activities were measured at chemical equilibrium status. For extrapolating *in vitro* activities of food phytochemicals to *in vivo*, extra descriptors such as hydrophobic constant and seem to be required in addition to this QSAR study.

References

Afiuni-Zadeh, S., and Azimi, G. (2010). A QSAR study for modeling of 8-azaadenine analogues proposed as A1 adenosine receptor antagonists using genetic algorithm coupling adaptive neuro-fuzzy inference system (ANFIS). Anal. Sci. *26*, 897–902.

Agatonovic-Kustrin, S., and Beresford, R. (2000). Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. J. Pharm. Biomed. Anal. *22*, 717–727.

Allouche, A.R. (2011). Gabedit—a graphical user interface for computational chemistry softwares. J. Comput. Chem. *32*, 174–182.

Amić, D., and Lučić, B. (2010). Reliability of bond dissociation enthalpy calculated by the PM6 method and experimental TEAC values in antiradical QSAR of flavonoids. Bioorg. Med. Chem. *18*, 28–35.

Amić, D., Davidović-Amić, D., Beslo, D., Rastija, V., Lucić, B., and Trinajstić, N. (2007). SAR and QSAR of the antioxidant activity of flavonoids. Curr. Med. Chem. *14*, 827–845.

Amiryousefi, M.R., Mohebbi, M., Khodaiyan, F., and Asadi, S. (2011). An empowered adaptive neuro-fuzzy inference system using self-organizing map clustering to predict mass transfer kinetics in deep-fat frying of ostrich meat plates. Comput. Electron. Agric. *76*, 89–95.

Anouar, E.H. (2014). A quantum chemical and statistical study of phenolic

Schiff bases with antioxidant activity against DPPH free radical. Antioxidants *3*, 309–322.

Arora, A., Sairam, R.K., and Srivastava, G.C. (2002). Oxidative stress and antioxidative system in plants. Curr. Sci. *82*, 1227.

Asplund, K. (2002). Antioxidant vitamins in the prevention of cardiovascular disease: a systematic review. J. Intern. Med. *251*, 372–392.

Beg, H., De, S.P., Ash, S., Das, D., and Misra, A. (2013). Polarizability, chemical hardness and ionization potential as descriptors to understand the mechanism of double proton transfer in acetamide dimer. Comput. Theor. Chem. *1005*, 1–8.

Blanksby, S.J., and Ellison, G.B. (2003). Bond dissociation energies of organic molecules. Acc. Chem. Res. *36*, 255–263.

Blokhina, O., Virolainen, E., and Fagerstedt, K.V. (2003). Antioxidants, oxidative damage and oxygen deprivation stress: a review. Ann. Bot. *91*, 179–194.

Blount, J.D., Metcalfe, N.B., Birkhead, T.R., and Surai, P.F. (2003). Carotenoid modulation of immune function and sexual attractiveness in zebra finches. Science *300*, 125–127.

Bondet, V., Brand-Williams, W., and Berset, C. (1997). Kinetics and mechanisms of antioxidant activity using the DPPH. free radical method. LWT-Food Sci. Technol. *30*, 609–615.

Borkowski, T., Szymusiak, H., Gliszczynska-Swiglo, A., Rietjens, I.M., and Tyrakowska, B. (2005). Radical scavenging capacity of wine anthocyanins

is strongly pH-dependent. J. Agric. Food Chem. *53*, 5526–5534.

Braga-Neto, U.M., and Dougherty, E.R. (2004). Is cross-validation valid for small-sample microarray classification?. Bioinformatics *20*, 374–380.

Brewer, M.S. (2011). Natural antioxidants: sources, compounds, mechanisms of action, and potential applications. Compr. Rev. Food Sci. Food Saf. *10*, 221–247.

Britton, G. (1995). Structure and properties of carotenoids in relation to function. FASEB J. *9*, 1551–1558.

Buyukbingol, E., Sisman, A., Akyildiz, M., Alparslan, F.N., and Adejare, A. (2007). Adaptive neuro-fuzzy inference system (ANFIS): A new approach to predictive modeling in QSAR applications: A study of neuro-fuzzy modeling of PCP-based NMDA receptor antagonists. Bioorg. Med. Chem. *15*, 4265–4282.

Byvatov, E., Fechner, U., Sadowski, J., and Schneider, G. (2003). Comparison of support vector machine and artificial neural network systems for drug/nondrug classification. J. Chem. Inf. Comput. Sci. *43*, 1882–1889.

Cadenas, E., and Davies, K.J.A. (2000). Mitochondrial free radical generation, oxidative stress, and aging. Free Radic. Biol. Med. *29*, 222–230.

Cai, Y., Luo, Q., Sun, M., and Corke, H. (2004). Antioxidant activity and phenolic compounds of 112 traditional Chinese medicinal plants associated with anticancer. Life Sci. *74*, 2157–2184.

Cai, Y.Z., Sun, M., Xing, J., Luo, Q., and Corke, H. (2006). Structure–radical scavenging activity relationships of phenolic compounds from traditional

Chinese medicinal plants. Life Sci. *78*, 2872–2888.

Cantrell, A., McGarvey, D.J., George Truscott, T., Rancan, F., and Böhm, F. (2003). Singlet oxygen quenching by dietary carotenoids in a model membrane environment. Arch. Biochem. Biophys. *412*, 47–54.

Cao, G., Alessio, H.M., and Cutler, R.G. (1993). Oxygen-radical absorbance capacity assay for antioxidants. Free Radic. Biol. Med. *14*, 303–311.

Chang, H.J., Choi, E.H., and Chun, H.S. (2008). Quantitative structure-activity relationship (QSAR) of antioxidative anthocyanidins and their glycosides. Food Sci. Biotechnol. *17*, 501–507.

Chen, A.F., Chen, D.D., Daiber, A., Faraci, F.M., Li, H., Rembold, C.M., and Laher, I. (2012). Free radical biology of the cardiovascular system. Clin. Sci. *123*, 73–91.

Cooke, M.S., Evans, M.D., Dizdaroglu, M., and Lunec, J. (2003). Oxidative DNA damage: mechanisms, mutation, and disease. FASEB J. *17*, 1195–1214.

Cortes, C., and Vapnik, V. (1995). Support-vector networks. Mach. Learn. *20*, 273–297.

Cramer, R.D., Patterson, D.E., and Bunce, J.D. (1988). Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. J. Am. Chem. Soc. *110*, 5959–5967.

Crichton, G.E., Bryan, J., and Murphy, K.J. (2013). Dietary antioxidants, cognitive function and dementia - a systematic review. Plant Foods Hum. Nutr. *68*, 279–292.

Cruciani, G., Baroni, M., Bonelli, D., Clementi, S., Ebert, C., and Skagerberg, B. (1990). Comparison of chemometric models for QSAR. Quant. Struct. Relatsh. *9*, 101–107.

Dasgupta, A., and Klein, K. (2014). Introduction to free radicals and the body's antioxidant defense. In Antioxidants in Food, Vitamins and Supplements (Elsevier), pp. 1–18.

Del Carlo, M., Sacchetti, G., Di Mattia, C., Compagnone, D., Mastrocola, D., Liberatore, L., and Cichelli, A. (2004). Contribution of the phenolic fraction to the antioxidant activity and oxidative stability of olive oil. J. Agric. Food Chem. *52*, 4072–4079.

Dudonne, S., Vitrac, X., Coutiere, P., Woillez, M., and Mérillon, J.M. (2009). Comparative study of antioxidant properties and total phenolic content of 30 plant extracts of industrial interest using DPPH, ABTS, FRAP, SOD, and ORAC assays. J. Agric. Food Chem. *57*, 1768–1774.

Edge, R., Land, E.J., McGarvey, D., Mulroy, L., and Truscott, T.G. (1998). Relative one-electron reduction potentials of carotenoid radical cations and the interactions of carotenoids with the vitamin E radical cation. J. Am. Chem. Soc. *120*, 4087–4090.

El-Agamey, A., Lowe, G.M., McGarvey, D.J., Mortensen, A., Phillip, D.M., Truscott, T.G., and Young, A.J. (2004). Carotenoid radical chemistry and antioxidant/pro-oxidant properties. Arch. Biochem. Biophys. *430*, 37–48.

Evans, P., and Halliwell, B. (2001). Micronutrients: oxidant/antioxidant status. Br. J. Nutr. *85*, S67–S74.

Faivre, B., Grégoire, A., Préault, M., Cézilly, F., and Sorci, G. (2003). Immune activation rapidly mirrored in a secondary sexual trait. Science *300*, 103–103.

Filipović, M., Marković, Z., Đorović, J., Marković, J.D., Lučić, B., and Amić, D. (2015). QSAR of the free radical scavenging potency of selected hydroxybenzoic acids and simple phenolics. Comptes Rendus Chim. *18*, 492–498.

Floegel, A., Kim, D.O., Chung, S.J., Koo, S.I., and Chun, O.K. (2011). Comparison of ABTS/DPPH assays to measure antioxidant capacity in popular antioxidant-rich US foods. J. Food Compos. Anal. *24*, 1043–1048.

Förstermann, U. (2008). Oxidative stress in vascular disease: causes, defense mechanisms and potential therapies. Nat. Clin. Pract. Cardiovasc. Med. *5*, 338–349.

Foyer, C.H., Lelandais, M., and Kunert, K.J. (1994). Photooxidative stress in plants. Physiol. Plant. *92*, 696–717.

Frankel, E.N. (1996). Antioxidants in lipid foods and their impact on food quality. Food Chem. *57*, 51–55.

Frankel, E.N., Huang, S.W., Aeschbach, R., and Prior, E. (1996). Antioxidant activity of a rosemary extract and its constituents, carnosic acid, carnosol, and rosmarinic acid, in bulk oil and oil-in-water emulsion. J. Agric. Food Chem. *44*, 131–135.

Free, S.M., and Wilson, J.W. (1964). A mathematical contribution to structure-activity studies. J. Med. Chem. *7*, 395–399.

Fridovich, I. (1978). The biology of oxygen radicals. Science *201*, 875–880.

Fujita, T., and Ban, T. (1971). Structure-activity relation. 3. Structure-activity study of phenethylamines as substrates of biosynthetic enzymes of sympathetic transmitters. J. Med. Chem. *14*, 148–152.

Gemperline, P.J., Long, J.R., and Gregoriou, V.G. (1991). Nonlinear multivariate calibration using principal components regression and artificial neural networks. Anal. Chem. *63*, 2313–2323.

Ghoush, M.A., Samhouri, M., Al-Holy, M., and Herald, T. (2008). Formulation and fuzzy modeling of emulsion stability and viscosity of a gum–protein emulsifier in a model mayonnaise system. J. Food Eng. *84*, 348–357.

González-Díaz, H., Bonet, I., Terán, C., De Clercq, E., Bello, R., García, M.M., Santana, L., and Uriarte, E. (2007). ANN-QSAR model for selection of anticancer leads from structurally heterogeneous series of compounds. Eur. J. Med. Chem. *42*, 580–585.

Gray, J.I. (1978). Measurement of lipid oxidation: a review. J. Am. Oil Chem. Soc. *55*, 539–546.

Haenen, G.R., Paquay, J.B., Korthouwer, R.E., and Bast, A. (1997). Peroxynitrite scavenging by flavonoids. Biochem. Biophys. Res. Commun. *236*, 591–593.

Hagerman, A.E., Riedl, K.M., Jones, G.A., Sovik, K.N., Ritchard, N.T., Hartzfeld, P.W., and Riechel, T.L. (1998). High molecular weight plant polyphenolics (tannins) as biological antioxidants. J. Agric. Food Chem. *46*,

1887–1892.

Hájek, P. (1998). Metamathematics of fuzzy logic (Springer Science & Business Media).

Halliwell, B. (2007). Biochemistry of oxidative stress. Biochem. Soc. Trans. *35*, 1147–1150.

Hansch, C., and Fujita, T. (1964). p-σ-π Analysis. A method for the correlation of biological activity and chemical structure. J. Am. Chem. Soc. *86*, 1616–1626.

Irwin, J.J., and Shoichet, B.K. (2005). ZINC-a free database of commercially available compounds for virtual screening. J. Chem. Inf. Model. *45*, 177–182.

Jackson, M.J. (1999). Free radicals in skin and muscle: damaging agents or signals for adaptation?. Proc. Nutr. Soc. *58*, 673–676.

Jang, J.S.R. (1993). ANFIS: adaptive-network-based fuzzy inference system. IEEE T. Syst. Man Cyb. *23*, 665–685.

Jang, J.S.R., Sun, C.T., and Mizutani, E. (1997). Neuro-fuzzy and soft computing: a computational approach to learning and machine intelligence.

Jeevarajan, J.A., and Kispert, L.D. (1996). Electrochemical oxidation of carotenoids containing donor/acceptor substituents. J. Electroanal. Chem. *411*, 57–66.

Jeevarajan, A.S., Khaled, M., and Kispert, L.D. (1994). Simultaneous electrochemical and electron paramagnetic resonance studies of carotenoids: effect of electron donating and accepting substituents. J. Phys. Chem. *98*,

7777–7781.

Jhin, C., and Hwang, K.T. (2014). Prediction of radical scavenging activities of anthocyanins applying adaptive neuro-fuzzy inference system (ANFIS) with quantum chemical descriptors. Int. J. Mol. Sci. *15*, 14715–14727.

Jhin, C., and Hwang, K.T. (2015). Adaptive neuro-fuzzy inference system applied QSAR with quantum chemical descriptors for predicting radical scavenging activities of carotenoids. Plos One *10*, e0140154.

Jiménez-Escrig, A., Jiménez-Jiménez, I., Sánchez-Moreno, C., and Saura-Calixto, F. (2000). Evaluation of free radical scavenging of dietary carotenoids by the stable radical 2,2-diphenyl-1-picrylhydrazyl. J. Sci. Food Agric. *80*, 1686–1690.

Jönsson, K., Szymusiak, H., Tyrakowska, B., Rietjens, I.M., and Narimantas, Č. (1999). Prooxidant toxicity of polyphenolic antioxidants to HL-60 cells: description of quantitative structure-activity relationships. FEBS Lett. *462*, 392–396.

Kähkönen, M.P., and Heinonen, M. (2003). Antioxidant activity of anthocyanins and their aglycons. J. Agric. Food Chem. *51*, 628–633.

Karadag, A., Ozcelik, B., and Saner, S. (2009). Review of methods to determine antioxidant capacities. Food Anal. Methods *2*, 41–60.

Karelson, M., Lobanov, V.S., and Katritzky, A.R. (1996). Quantum-chemical descriptors in QSAR/QSPR studies. Chem. Rev. *96*, 1027–1044.

Kastenholz, M.A., Pastor, M., Cruciani, G., Haaksma, E.E.J., and Fox, T. (2000). GRID/CPCA: a new computational tool to design selective ligands.

J. Med. Chem. *43*, 3033–3044.

Kellogg, G.E., and Abraham, D.J. (1992). KEY, LOCK, and LOCKSMITH: complementary hydropathic map predictions of drug structure from a known receptor-receptor structure from known drugs. J. Mol. Graph. *10*, 212–217.

Kerry, N., and Rice-Evans, C. (1999). Inhibition of peroxynitrite-mediated oxidation of dopamine by flavonoid and phenolic antioxidants and their structural relationships. J. Neurochem. *73*, 247–253.

Kiralj, R., and Ferreira, M.M.C. (2009). Basic validation procedures for regression models in QSAR and QSPR studies: theory and application. J. Braz. Chem. Soc. *20*, 770–787.

Klein, E., and Lukeš, V. (2007). DFT/B3LYP study of the substituent effect on the reaction enthalpies of the individual steps of sequential proton loss electron transfer mechanism of phenols antioxidant action: Correlation with phenolic CO bond length. J. Mol. Struct. (THEOCHEM) *805*, 153–160.

Kleinová, M., Hewitt, M., Brezová, V., Madden, J.C., Cronin, M.T.D., and Valko, M. (2007). Antioxidant properties of carotenoids: QSAR prediction of their redox potentials. Gen. Physiol. Biophys. *26*, 97–103.

Kong, J.M., Chia, L.S., Goh, N.K., Chia, T.F., and Brouillard, R. (2003). Analysis and biological activities of anthocyanins. Phytochemistry *64*, 923–933.

Kontogiorgis, A.C., Pontiki, A.E., and Hadjipavlou-Litina, D. (2005). A review on quantitative structure-activity relationships (QSARs) of natural

and synthetic antioxidants compounds. Mini Rev. Med. Chem. *5*, 563–574.

Koopmans, T. (1934). Über die Zuordnung von Wellenfunktionen und Eigenwerten zu den Einzelnen Elektronen Eines Atoms. Physica *1*, 104–113.

Kozlowski, D., Marsal, P., Steel, M., Mokrini, R., Duroux, J.L., Lazzaroni, R., and Trouillas, P. (2007). Theoretical investigation of the formation of a new series of antioxidant depsides from the radiolysis of flavonoid compounds. Radiat. Res. *168*, 243–252.

Krinsky, N.I., and Deneke, S.M. (1982). Interaction of oxygen and oxy-radicals with carotenoids. J. Natl. Cancer Inst. *69*, 205–210.

Kubinyi, H. (1988). Free Wilson analysis. Theory, applications and its relationship to Hansch analysis. Quant. Struct. Relatsh. *7*, 121–133.

Kubinyi, H. (1997). QSAR and 3D QSAR in drug design Part 1: methodology. Drug Discov. Today *2*, 457–467.

Li, M.J., Liu, L., Fu, Y., and Guo, Q.X. (2007). Accurate bond dissociation enthalpies of popular antioxidants predicted by the ONIOM-G3B3 method. J. Mol. Struct. (THEOCHEM) *815*, 1–9.

Liao, H., Banbury, L.K., and Leach, D.N. (2008). Antioxidant activity of 45 Chinese herbs and the relationship with their TCM characteristics. Evid-Based Compl. Alt. Med. *5*, 429–434.

Liaw, A., and Svetnik, V. (2015). QSAR modeling: prediction of biological activity from chemical structure. Stat. Methods Eval. Saf. Med. Prod. Dev. 66–83.

Lien, E.J., Ren, S., Bui, H.H., and Wang, R. (1999). Quantitative structure-activity relationship analysis of phenolic antioxidants. Free Radic. Biol. Med. *26*, 285–294.

Liew, C.Y., Ma, X.H., and Yap, C.W. (2010). Consensus model for identification of novel PI3K inhibitors in large chemical library. J. Comput. Aided Mol. Des. *24*, 131–141.

Lin, T.C., Liu, H.L., and Kuo, M.J. (2009). Direct adaptive interval type-2 fuzzy control of multivariable nonlinear systems. Eng. Appl. Artif. Intell. *22*, 420–430.

Lipinski, C.A. (2002). Poor aqueous solubility - an industry wide problem in ADME screening. Am. Pharm. Rev. *5*, 82–85.

Logani, M.K., and Davies, R.E. (1980). Lipid oxidation: biologic effects and antioxidants - a review. Lipids *15*, 485–495.

Mairanovsky, V.G., Engovatov, A.A., Ioffe, N.T., and Samokhvalov, G.I. (1975). Electron-donor and electron-acceptor properties of carotenoids: electrochemical study of carotenes. J. Electroanal. Chem. Interfacial Electrochem. *66*, 123–137.

Marı, A., Lardizabal-Gutiérrez, D., Torres-Moye, E., Fuentes-Cobas, L., Balandrán-Quintana, R.R., Camacho-Dávila, A., Quintero-Ramos, A., and Glossman-Mitnik, D. (2007). Optimized structure and thermochemical properties of flavonoids determined by the CHIH (medium)-DFT model chemistry versus experimental techniques. J. Mol. Struct. *871*, 114–130.

Matsufuji, H., Kido, H., Misawa, H., Yaguchi, J., Otsuki, T., Chino, M.,

Takeda, M., and Yamagata, K. (2007). Stability to light, heat, and hydrogen peroxide at different pH values and DPPH radical scavenging activity of acylated anthocyanins from red radish extract. J. Agric. Food Chem. *55*, 3692–3701.

Mazza, G., and Brouillard, R. (1987). Recent developments in the stabilization of anthocyanins in food products. Food Chem. *25*, 207–225.

Michalski, R.S., Carbonell, J.G., and Mitchell, T.M. (2013). Machine learning: An aificial intelligence approach (Springer Science & Business Media).

Miller, N.J., Sampson, J., Candeias, L.P., Bramley, P.M., and Rice-Evans, C.A. (1996). Antioxidant activities of carotenes and xanthophylls. FEBS Lett. *384*, 240–242.

Mortensen, A., Skibsted, L.H., and Truscott, T.G. (2001). The interaction of dietary carotenoids with radical species. Arch. Biochem. Biophys. *385*, 13–19.

Müller, L., Fröhlich, K., and Böhm, V. (2011). Comparative antioxidant activities of carotenoids measured by ferric reducing antioxidant power (FRAP), ABTS bleaching assay (αTEAC), DPPH assay and peroxyl radical scavenging assay. Food Chem. *129*, 139–148.

Musialik, M., Kuzmicz, R., Pawłowski, T.S., and Litwinienko, G. (2009). Acidity of hydroxyl groups: an overlooked influence on antiradical properties of flavonoids. J. Org. Chem. *74*, 2699–2709.

Niculescu, S.P. (2003). Artificial neural networks and genetic algorithms in QSAR. J. Mol. Struct. (THEOCHEM) *622*, 71–83.

O'Boyle, N.M., Banck, M., James, C.A., Morley, C., Vandermeersch, T., and Hutchison, G.R. (2011). Open babel: an open chemical toolbox. J. Cheminformatics *3*, 33.

Osman, A.M. (2011). Multiple pathways of the reaction of 2, 2-diphenyl-1-picrylhydrazyl radical (DPPH) with (+)-catechin: Evidence for the formation of a covalent adduct between DPPH and the oxidized form of the polyphenol. Biochem. Biophys. Res. Commun. *412*, 473–478.

Park, S.U., Chung, S.K., and Newcomb, M. (1986). Acceptor, donor, and captodative stabilization in transition states of 5-hexen-1-yl radical cyclizations. J. Am. Chem. Soc. *108*, 240–244.

Parr, R.G., and Pearson, R.G. (1983). Absolute hardness: companion parameter to absolute electronegativity. J. Am. Chem. Soc. *105*, 7512–7516.

Parr, R.G., Donnelly, R.A., Levy, M., and Palke, W.E. (1978). Electronegativity: the density functional viewpoint. J. Chem. Phys. *68*, 3801–3807.

Pasha, F., Srivastava, H., and Singh, P. (2005). Comparative QSAR study of phenol derivatives with the help of density functional theory. Bioorg. Med. Chem. *13*, 6823–6829.

Pasha, F.A., Cho, S.J., Beg, Y., and Tripathi, Y.B. (2008). Quantum chemical QSAR study of flavones and their radical-scavenging activity. Med. Chem. Res. *16*, 408–417.

Poblet, J.M., Canadell, E., and Sordo, T. (1983). Concerning the orientation in free radical additions to olefins. Can. J. Chem. *61*, 2068–2069.

Podsędek, A. (2007). Natural antioxidants and antioxidant capacity of Brassica vegetables: a review. LWT-Food Sci. Technol. *40*, 1–11.

Prior, R.L., Wu, X., and Schaich, K. (2005). Standardized methods for the determination of antioxidant capacity and phenolics in foods and dietary supplements. J. Agric. Food Chem. *53*, 4290–4302.

Priyadarsini, K.I., Maity, D.K., Naik, G.H., Kumar, M.S., Unnikrishnan, M.K., Satav, J.G., and Mohan, H. (2003). Role of phenolic OH and methylene hydrogen on the free radical reactions and antioxidant activity of curcumin. Free Radic. Biol. Med. *35*, 475–484.

Puzyn, T., Suzuki, N., Haranczyk, M., and Rak, J. (2008). Calculation of quantum-mechanical descriptors for QSPR at the DFT level: is it necessary?. J. Chem. Inf. Model. *48*, 1174–1180.

Re, R., Pellegrini, N., Proteggente, A., Pannala, A., Yang, M., and Rice-Evans, C. (1999). Antioxidant activity applying an improved ABTS radical cation decolorization assay. Free Radic. Biol. Med. *26*, 1231–1237.

Rimm, E.B., Ascherio, A., Giovannucci, E., Spiegelman, D., Stampfer, M.J., and Willett, W.C. (1996). Vegetable, fruit, and cereal fiber intake and risk of coronary heart disease among men. Jama *275*, 447–451.

Sanchez, A., Flores-Cotera, L.B., Langley, E., Martin, R., Maldonado, G., and Sanchez, S. (1999). Carotenoides: estructura, función, biosíntesis, regulación y aplicaciones. Rev. Latinoam. Microbiol. *41*, 175–192.

Sarkar, A., Middya, T.R., and Jana, A.D. (2011). A QSAR study of radical scavenging antioxidant activity of a series of flavonoids using DFT based

quantum chemical descriptors - the importance of group frontier electron density. J. Mol. Model. *18*, 2621–2631.

Schwartz, J.L., Sloane, D., and Shklar, G. (1989). Prevention and inhibition of oral cancer in the hamster buccal pouch model associated with carotenoid immune enhancement. Tumor Biol. *10*, 297–309.

Scott, S.L., Chen, W.J., Bakac, A., and Espenson, J.H. (1993). Spectroscopic parameters, electrode potentials, acid ionization constants, and electron exchange rates of the 2, 2'-azinobis (3-ethylbenzothiazoline-6-sulfonate) radicals and ions. J. Phys. Chem. *97*, 6710–6714.

Sharma, O.P., and Bhat, T.K. (2009). DPPH antioxidant assay revisited. Food Chem. *113*, 1202–1205.

Smola, A., and Vapnik, V. (1997). Support vector regression machines. Adv. Neural Inf. Process. Syst. *9*, 155–161.

Snedecor, G.W., and Cochran, W. (1967). Statistical methods. Iowa state (University Press).

Soffers, A.E.M.F., van Haandel, M.J.H., Boersma, M.G., Tyrakowska, B., Laane, C., and Rietjens, I.M.C.M. (1999). Antioxidant activities of carotenoids: quantitative relationships between theoretical calculations and experimental literature data. Free Radic. Res. *30*, 233–240.

Stahl, W., and Sies, H. (2005). Bioactivity and protective effects of natural carotenoids. BBA-Mol. Basis Dis. *1740*, 101–107.

Staikova, M., Wania, F., and Donaldson, D. (2004). Molecular polarizability as a single-parameter predictor of vapour pressures and octanol–air

partitioning coefficients of non-polar compounds: a priori approach and results. Atmos. Environ. *38*, 213–225.

Stewart, J.J. (2007). Optimization of parameters for semiempirical methods V: modification of NDDO approximations and application to 70 elements. J. Mol. Model. *13*, 1173–1213.

Stewart, J.J. (2013). Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters. J. Mol. Model. *19*, 1–32.

Stewart, J.J.P. (2014). MOPAC 2012 (Stewart Computational Chemistry).

Thaipong, K., Boonprakob, U., Crosby, K., Cisneros-Zevallos, L., and Hawkins Byrne, D. (2006). Comparison of ABTS, DPPH, FRAP, and ORAC assays for estimating antioxidant activity from guava fruit extracts. J. Food Compos. Anal. *19*, 669–675.

Topliss, J.G., and Edwards, R.P. (1979). Chance factors in studies of quantitative structure-activity relationships. J. Med. Chem. *22*, 1238–1244.

Torel, J., Cillard, J., and Cillard, P. (1986). Antioxidant activity of flavonoids and reactivity with peroxy radical. Phytochemistry *25*, 383–385.

Trouillas, P., Marsal, P., Siri, D., Lazzaroni, R., and Duroux, J.L. (2006). A DFT study of the reactivity of OH groups in quercetin and taxifolin antioxidants: the specificity of the 3-OH site. Food Chem. *97*, 679–688.

Tsuda, T., Shiga, K., Ohshima, K., Kawakishi, S., and Osawa, T. (1996). Inhibition of lipid peroxidation and the active oxygen radical scavenging effect of anthocyanin pigments isolated from *Phaseolus vulgaris* L.

Biochem. Pharmacol. *52*, 1033–1039.

Tu, J.V. (1996). Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. J. Clin. Epidemiol. *49*, 1225–1231.

Vaca, C.E., Wilhelm, J., and Harms-Ringdahl, M. (1988). Interaction of lipid peroxidation products with DNA. A review. Mutat. Res. Genet. Toxicol. *195*, 137–149.

Vagánek, A., Rimarčík, J., Lukeš, V., and Klein, E. (2012). On the energetics of homolytic and heterolytic O H bond cleavage in flavonoids. Comput. Theor. Chem. *991*, 192–200.

Valko, M., Leibfritz, D., Moncol, J., Cronin, M.T., Mazur, M., and Telser, J. (2007). Free radicals and antioxidants in normal physiological functions and human disease. Int. J. Biochem. Cell Biol. *39*, 44–84.

Vapnik, V. (2013). The nature of statistical learning theory (Springer Science & Business Media).

Vasil'ev, R.F., Kancheva, V.D., Fedorova, G.F., Batovska, D.I., and Trofimov, A.V. (2010). Antioxidant activity of chalcones: The chemiluminescence determination of the reactivity and the quantum chemical calculation of the energies and structures of reagents and intermediates. Kinet. Catal. *51*, 507–515.

Verma, J., Khedkar, V.M., and Coutinho, E.C. (2010). 3D-QSAR in drug design-a review. Curr. Top. Med. Chem. *10*, 95–115.

Wang, Y., Xiao, J., Suzek, T.O., Zhang, J., Wang, J., and Bryant, S.H. (2009).

PubChem: a public information system for analyzing bioactivities of small molecules. Nucleic Acids Res. gkp456.

Wold, S., Sjöström, M., and Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. Chemom. Intell. Lab. Syst. *58*, 109–130.

Worachartcheewan, A., Nantasenamat, C., Isarankura-Na-Ayudhya, C., Prachayasittikul, S., and Prachayasittikul, V. (2011). Predicting the free radical scavenging activity of curcumin derivatives. Chemom. Intell. Lab. Syst. *109*, 207–216.

Wright, J.S. (2002). Predicting the antioxidant activity of curcumin and curcuminoids. J. Mol. Struct. (THEOCHEM) *591*, 207–217.

Wright, J.S., Johnson, E.R., and DiLabio, G.A. (2001). Predicting the activity of phenolic antioxidants: theoretical method, analysis of substituent effects, and application to major families of antioxidants. J. Am. Chem. Soc. *123*, 1173–1183.

Xue, Y., Li, H., Ung, C., Yap, C., and Chen, Y. (2006). Classification of a diverse set of Tetrahymena pyriformis toxicity chemical compounds from molecular descriptors by statistical learning methods. Chem. Res. Toxicol. *19*, 1030–1039.

Yang, G., Yu, J., Wang, Z., Zeng, X., and Ju, X. (2007). QSPR study on the aqueous solubility (− lgSw) and n-octanol/water partition coefficients (lgKow) of polychlorinated dibenzo-p-dioxins (PCDDs). QSAR Comb. Sci. *26*, 352–357.

Yee, L.C., and Wei, Y.C. (2012). Current modeling methods used in

QSAR/QSPR. In Statistical Modelling of Molecular Descriptors in QSAR/QSPR (Wiley-VCH Verlag GmbH & Co. KGaA), pp. 1–31.

Zeng, X., Wang, Z., Ge, Z., and Liu, H. (2007). Quantitative structure–property relationships for predicting subcooled liquid vapor pressure (PL) of 209 polychlorinated diphenyl ethers (PCDEs) by DFT and the position of Cl substitution (PCS) methods. Atmos. Environ. *41*, 3590–3603.

Zhang, H.Y., and Wang, L.F. (2002). Theoretical elucidation on structure–antioxidant activity relationships for indolinonic hydroxylamines. Bioorg. Med. Chem. Lett. *12*, 225–227.

국문초록

식품 유래 파이토케미컬의 항산화활성에 대한 적응형 뉴로-퍼지 추론시스템 적용 QSAR

식품영양학과

서울대학교

진 창 호

항산화제는 산화적 스트레스로부터 기인하는 대상 분자의 손상을 현저하게 지연시키거나, 감소 혹은 방해하는 항산화 활성을 띠는 화합물을 일컫는다. 식품 유래 파이토케미컬은 자유라디칼을 소거함으로써 인체 항산화 체계의 일부 역할을 담당한다. 식품 유래 파이토케미컬은 생리활성 기능 이외에 식품의 품질 및 보존기간에 영향을 줄 수 있다. 파이토케미컬의 우수한 항산화 활성은 과거 수많은 연구를 통해 증명되었다. 하지만 파이토케미컬의 화학구조와 항산화 활성이 밀접하게 연관되어 있음에도 불구하고 구조와 활성 사이의 관계를 체계적으로 분석한 연구는 많이 찾아볼 수 없었다. 따라서 파이토케미컬의 화학적 구조와 항산화활성 사이의 관계를 근거를 기반으로 하여 체계적으로 이해함과 동시에 파이토케미컬의 항산화 활성의 정확한 예측을 위한 연구가 필요하다. 정량적 구조-

활성 관계(quantitative structure-activity relationship; QSAR)는 분자의 화학구조적 변수와 분자의 활성도 사이의 상관관계를 정량적으로 규정한 것 또는 규정한 수학적 함수를 뜻한다. 적절한 화학적 구조 기술자(descriptor)의 선택은 신뢰할 수 있는 QSAR 모델을 구축하기 위해 중요하게 여겨지는 부분이다. 소수성 척도, 구조적 수치 기술자, 양자역학적 기술자, 3차원 구조 기술자 등이 QSAR에서 구조 기술자로 이용될 수 있다. 이중에서, 양자역학적 기술자는 산화-환원 반응 등의 화학반응을 설명할 수 있는 변수이다. 적절한 모델링 기법을 선택하는 것도 QSAR 연구에서 중요하다. 일반적으로 QSAR 모델링 기법으로써 multiple linear regression과 partial least square 방법을 사용한다. 최근에는 좀 변수 사이의 비선형적인 관계를 적합화할 수 있는 더 발전된 방법인 machine learning 방법도 QSAR 연구에 이용하고 있다. 적응형 뉴로-퍼지 추론시스템(adaptive neuro-fuzzy inference system; ANFIS)은 machine learning 방법 중 하나로서 변수 사이의 비선형적 관계를 정확하게 예측할 수 있는 모델링 기법이다.

본 연구의 목적은 식품에 존재하는 주요한 파이토케미컬의 항산화능을 정확히 예측할 수 있는 QSAR 모델을 양자역학적 기술자를 이용하여 구축하는 것이다. 본 연구는 크게 세 부분으로 구성되어 있다. 본 연구의 첫번째와 두번째 부분에서는 각각

안토시아닌(비배당체인 안토시아니딘 포함)과 카로티노이드의 항산화능 예측을 위한 ANFIS 적용 QSAR 모델을 구축하였다. 연구의 세번째 부분에서는 여러 종류의 페놀계 파이토케미컬의 항산화능을 QSAR을 이용하여 분석하고 예측모델을 구축하였다.

첫번째로 수행한 안토시아닌의 ANFIS 적용 QSAR 연구에서는 안토시아닌의 양자화학적 기술자를 계산하기 위하여 용액의 pH에 따라 구조가 달라지는 안토시아닌의 특성을 고려하였다. 안토시아닌의 구조는 pH에 따라 변하기 때문에, 구조-활성 관계 분석을 위해서는 하나의 분자구조가 아니라 pH 변화에 따른 여러 안토시아닌 구조가 모두 고려되어야 한다. 본 연구에서는 안토시아닌의 flavylium cation, quinoidal base, carbinol pseudo-base, chalcone 등의 구조의 양자화학적 기술자를 semi-empirical PM6와 PM7 방법으로 계산하였다. 안토시아닌의 항산화 활성과 구조 사이의 상관분석을 수행한 결과, flavylium cation의 electron affinity와 elenctronegativity, quinoidal base의 ionisation potential이 안토시아닌의 항산화 활성과 유의적인 상관관계가 있었다. 따라서, QSAR 모델을 구축하는 데 있어서 이들 기술자를 독립변수로 이용하였다. 각각의 독립 변수에 대하여 2개의 삼각형 형태의 input fuzzy 함수를 구축하였으며, backpropagation 방법으로 모델을 training하였다. PM6와 PM7 방법으로 계산된

독립변수를 사용하여 구축한 ANFIS-QSAR 예측모델은 Q square 값이 각각 0.819와 0.862으로 높은 예측 정확도를 보였다.

두번째 연구에서는 카로티노이드의 양자화학적 기술자를 PM6와 PM7 방법을 이용하여 계산하였다. 카로티노이드 중성분자뿐만 아니라 카로티노이드 양이온도 화학적 반응계에 존재하기 때문에 카로티노이드 양이온의 양자화학적 기술자도 계산하였다. QSAR 모델 구축에 이용할 독립변수를 선택하기 위하여 상관분석을 수행한 결과, 중성 카로티노이드와 카로티노이드 1가 양이온의 ionisation energy(I, I_{cat}), 그리고 중성 분자와 1가 양이온의 chemical potential 교호작용(μ_{cross})이 항산화 활성과 유의적인 상관관계가 있었다. 따라서 I, I_{cat}, μ_{cross}을 QSAR 모델의 독립변수로 이용하였다. 각 독립변수당 2개의 삼각형 형태의 input fuzzy 함수를 구축하여 ANFIS 모델을 만들었으며, backpropagation 방법에 의해 예측모델을 training하였다. PM6와 PM7 방법으로 계산한 양자화학적 변수를 이용하여 구축한 QSAR 모델의 적합도는 각각 0.921과 0.902로 예측능이 높았다.

세번째 연구에서는 O-H 결합 해리 에너지(BDE)를 semi-empirical PM6와 PM7 방법으로 계산하였다. 과거 선행연구에서 각 분자의 최소 BDE 값(BDE_{min})과 항산화활성 사이의 음의 상관관계 있음이 보고되었다. 본 연구에서는 BDE_{min}에 해당하는

OH뿐만 아니라, 분자 내의 다른 OH그룹도 항산화 활성에 영향을 준다고 가정하였다. 또한, 알코올 그룹의 BDE보다 페놀과 enol의 BDE 값이 낮다는 사실을 확인하였다. 이러한 사항들을 고려하여 QSAR 모델에 이용할 새로운 독립변수로서 페놀과 enol의 BDE 역수의 합을 계산하였다(X_{BDE}). 새롭게 제시된 변수 X_{BDE}는 항산화 활성과 유의적인 상관관계가 있었다. X_{BDE}를 이용하여 선형회귀분석 기반의 QSAR 모델과 ANFIS 기반의 QSAR 모델을 각각 구축하였다. 선형회귀분석법을 이용하여 구축한 QSAR 모델의 R은 0.907이었다. ANFIS를 이용한 QSAR 모델의 R은 0.919로 선형회귀법보다 예측능이 높았다. 본 연구를 통해, 탄소골격이 서로 다른 여러 페놀계 파이토케미컬의 항산화 활성을 통합적으로 예측하고 평가할 수 있는 QSAR 모델이 구축되었으며, 항산화 활성과 유의적인 상관관계가 있는 지표로서 X_{BDE}가 제시되었다. 또한, ANFIS의 적용으로 예측능이 높은 QSAR 모델을 구축할 수 있음을 확인하였다.

본 연구에서는 ANFIS를 적용한 QSAR 모델을 이용하여 식품 유래 파이토케미컬의 항산화 활성을 정확히 예측할 수 있음을 확인하였다. 또한, 파이토케미컬 분자 구조의 양자 역학적인 특성과 항산화능 사이의 관계를 분석함으로써, 각 파이토케미컬의 항산화 기작에 대한 고찰을 하였다. 수행한 연구 결과는 항산화 활성을

띠는 식물 유래의 기능성 식품 및 파이토케미컬 유도체 등의 바이오 소재 개발과 응용을 위하여 이용할 수 있을 것이다.

주요어: 파이토케미컬, 항산화 활성, 라디칼 소거능, 정량적 구조-활성 관계(QSAR), 적응 뉴로-퍼지 추론 시스템(ANFIS).

학 번: 2011-23690