QUANTITATIVE STRUCTURAL ACTIVITY RELATIONSHIP OF FLAVONOIDS: STUDIES OF ANTIOXIDANT PROPERTIES AND HUMAN INTESTINAL PERMEABILITY

by

HAMIN HWANG

(Under the Direction of William Kerr)

ABSTRACT

Flavonoids are common antioxidants and are found in many plant species. However, only a few studies are based on quantification of structural information and bioavailability. Therefore, both QSAR (quantitative structural-activity relationship) and QSPR (quantitative structural-permeability relationship) study on flavonoids were conducted. In the QSAR study, flavonoid structures were obtained from the Cambridge Molecular Structure Database, minimized, and studied using Grid and Volsurf. The generated data was compared to experimental TEAC (Trolox equivalent antioxidant capacity) data from a previous study. The QSAR study found that antioxidant potential of flavonoids increases with hydrophobicity, smaller molecular weight, lack of rugosity, and increasing number of hydroxy groups. The PCA (principal components analysis) 7 component model explained 88.7% and PLS 6 component model gave an $R^2$ value of 0.8626, showing high correlation between antioxidant potential and 23 flavonoid molecules. In the next portion of the study, a QSPR study was performed using structures from the Cambridge Software Molecular Database. After energy minimization, QSPR software, known as Volsurf, was used to generate the permeability data. This data was correlated and compared to the experimental Caco-2 data from another study. In the QSPR
study, the computational study matched with 94.1% of data determining permeability of 17 structures. In addition, smaller molecular weight and hydrophobic flavonoids showed much higher permeability than larger and hydrophilic flavonoid molecules.

INDEX WORDS: Flavonoids, QSAR, Structure relationship, Computational study, Volsurf, Caco-2 cell monolayer, Digestion, Absorption, Antioxidant.
QUANTITATIVE STRUCTURAL ACTIVITY RELATIONSHIP OF FLAVONOIDS: STUDIES OF ANTIOXIDANT PROPERTIES AND HUMAN INTESTINAL PERMEABILITY

By

HAMIN HWANG

B.S., The University of Georgia, 2006

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2008
QUANTITATIVE STRUCTURAL ACTIVITY RELATIONSHIP OF FLAVONOIDS: STUDIES OF ANTIOXIDANT PROPERTIES AND HUMAN INTESTINAL PERMEABILITY

by

HAMIN HWANG

Major Professor: William Kerr
Committee: Ron Pegg
              Robert Woods

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2008
DEDICATION

This dissertation is dedicated to my family--Ju Sang Hwang (father), Myoung Sook Yoon (mother), Ha Kyung Hwang (brother), and Ye Kang (grandmother), who recently passed away at the age of 92.

Special appreciation goes to Polly Cleveland, from whom I’ve learned much computer knowledge.
ACKNOWLEDGEMENTS

I would like to express my gratitude to the following people:

My major professor, Dr. William Kerr, has accepted me for whom I am, shown me the necessary steps to further advance myself, being my friend and family. Thanks to my committee members, Dr. Ron Pegg and Dr. Robert Woods for their guidance during this research and the pursuit of higher education. I show my special thanks to Dr. Robert Shewfelt and Dr. Yao-wen Huang for introducing food science to my life. I am thankful to Dr. Romeo Toledo for his tough yet mind-opening engineering exams. Special thanks to Carl Ruiz for his companionship, twisted humor, and religious guidance. My lab mates Mark Corey, Laura Brindle, George Cavender, Katherine Acosta, and Jinhee Yi for their help all these years. Dr. Ramsey Bakali, Dr. Joe Mouldin, Dr. Gene Pesti, Dr. Hardy Edwards, Dr. Sammy Aggry for their lifelong lessons in life and teachings that will carry my heart throughout my life.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>ACKNOWLEDGEMENTS</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>v</td>
<td></td>
</tr>
</tbody>
</table>

## CHAPTERS

1. INTRODUCTION ................................................................. 1

2. LITERATURE REVIEW .......................................................... 9

3. QUANTITATIVE STRUCTURAL-ACTIVITY RELATIONSHIP OF FLAVONOIDS AND ANTIOXIDANT POTENTIAL .................................................. 34

4. QUANTITATIVE STRUCTURE-PERMEABILITY RELATIONSHIP OF FLAVONOIDS USING CACO-2 CELLS: INDEPTH STUDY OF HUMAN INTESTINAL PERMEATION OF LAVONOIDS .............................................. 52

5. SUMMARY AND CONCLUSIONS .................................................. 76
CHAPTER 1
INTRODUCTION

In the face of the new world and scientific knowledge that drives forth such machine, few forces can reshape how people think, believe, and prosper. Even more so than ever, consumers and mankind thrive to discover new ideas, understand mechanisms. In the paper to follow, I present my dedication and works that hopefully change the spectacle of sands or the path that light bends when it hits that little dust particle—all with a hope that it is a small step into the light of knowledge.

Flavonoids are phenolic compounds that are available in plants, and they represent more than half of the 8000 known phenolics found in nature (Harborne, Baxter et al. 1999)(Harborne, Baxter et al. 1999). They are derived from amino acids, tyrosine and phenylalanine and play significant roles in the diet. It is common knowledge that diets rich in vegetables and fruits are beneficial to human health. Therefore, it is reasonable to state that consumption of fruits and vegetables, containing flavonoids, are related to positive health benefits. How much of such benefits are specifically due to flavonoids is indeterminant, but flavonoids are believed to be strong antioxidants and are shown to help prevent many diseases. Antioxidants quench reactive oxygen and nitrogen species such as the superoxide anion radicals, hydroxyl radicals, and peroxyl radicals. These reactive oxygen species reduce oxygen (Williams and Jeffrey 2000)(Williams and Jeffrey 2000) and can produce harm in the biological system. Such harm includes protein denaturation, membrane activity, DNA alterations, and
lipid peroxidation (Kinsella, Frankel et al. 1993). Therefore, antioxidants like flavonoids can provide health benefits such as prevention of atherosclerosis (Hertog, Hollman et al. 1992; Keys 1995), anti-inflammation (Middleton, Kandaswami et al. 2000), anti-aging, prevention of coronary heart disease, prevention of diabetes mellitus (Slater 1984; Cheng, Lin et al. 2003), Alzheimer’s disease (Smith, Rottkamp et al. 2000), and anti-microbial effects (Harborne and Williams 2000).

Due to the many benefits flavonoids offer to man, quantification methodologies are an important part of the science and further understanding of flavonoids. Currently, popular methods include the Trolox equivalent antioxidant capacity (TEAC) assay (Miller, Rice-Evans et al. 1993; Rice-Evans and Miller 1994; Pellegrini, Proteggente et al. 1999), DPPH· method (Brand-Williams, Cuvelier et al. 1995; Sánchez-Moreno, Larrauri et al. 1998), ferric reducing ability of plasma (FRAP) assay (Benzie and Strain 1999), oxygen radical absorbance capacity (ORAC) assay (Glazer 1990), total radical trapping parameter (TRAP) (Wayner, Burton et al. 1985), dichlorofluorescin diacetate (DCFH-DA) based assay (Valkonen and Kuusi 1997; Amado, Jaramillo et al. 2007), cyclic voltammetry method (Kohen, Beit-Yannai et al. 1999), total oxyradical scavenging capacity (TOSC) assay (Winston, Regoli et al. 1998), photochemiluminescence (PCL) assay (Popov, Lewin et al. 2000), and other methods (Kohen, Beit-Yannai et al. 1999).
1987; Popov and Lewin 1994; Popov and Lewin 1996)(Popov, Lewin et al. 1987; Popov and Lewin 1994; Popov and Lewin 1996). Among the many, the TEAC assay was selected due to its ability to study both hydrophilic and lipophilic compounds as well as successful recent studies performed (Dastmalchi, Damien Dorman et al. 2007; Srinivasan, Chandrasekar et al. 2007)(Dastmalchi, Damien Dorman et al. 2007; Srinivasan, Chandrasekar et al. 2007).

When new tools and technology are introduced, keen scientists take advantage of the tools and application that can be performed. In this day and age, there have been dramatical advances in computer technology and related fields. For example, robots and computer programs are present in many things that surround life, society, and work. To an extent, they are being used from sending a man into the moon or used for garbage disposal units. With that in mind, there have been an increasing number of computational studies to perform scientific experiments. In this case, application of computational study was used to study the quantitative structure–activity relationships (QSAR) of flavonoids. In this study, linkage between antioxidant potential to the molecular structures will be examined. This was examined using Volsurf (Volsurf version 3.0 software by Molecular Discovery Ltd)(Volsurf version 3.0 software by Molecular Discovery Ltd) program.

In the next portion of the study, a similar computational study is performed to examine the properties of molecules that can predict and relate human intestinal absorption of flavonoids. Human digestion studies are important due to the new scientific evidence they can provide. Much like yester-age of vitamins, antioxidants, like flavonoids, are considered the new generation nutrients, providing something extra on top of basic nutrition. Therefore, RDIs or
AlIs would be interesting to observe in the future. However, for RDI or AI to exist, absorption studies must be performed and more so, quantitative studies are absolutely necessary.

Recently, a few different methods were employed to predict or measure the absorption of flavonoids. These include computational studies (Ekins, Durst et al. 2001; Ponce, Perez et al. 2004)(Ekins, Durst et al. 2001; Ponce, Perez et al. 2004), in vitro studies (Kuo 1998; Murota, Shimizu et al. 2000; Masataka Oitate 2001)(Kuo 1998; Murota, Shimizu et al. 2000; Masataka Oitate 2001), and in vivo studies (Hollman, Vries et al. 1995; Hollman, Trijp et al. 1997)(Hollman, Vries et al. 1995; Hollman, Trijp et al. 1997). Among the many however, Caco-2 monolayer cells are widely used and accepted in the scientific community as a human intestinal absorption model (Artursson, Palm et al. 2001)(Artursson, Palm et al. 2001). Therefore, combination of Caco-2 absorption methods and computational methods can be very powerful tool in predicting and determining molecular characteristics that are linked with absorption data. The reason behind such combinatorial research lies in few factors. As more data is obtained from in vivo and in vitro experiments, these data can be compiled into a large database. This can be used to predict different types of compounds. As a matter of fact, the Volsurf (Volsurf version 3.0 software by Molecular Discovery Ltd)(Volsurf version 3.0 software by Molecular Discovery Ltd) program has a Caco-2 database of 751 chemical compounds. Moreover, the time consumption, economics of experimental equipments, reagents, and differences in lab methods and materials can be detrimental to Caco-2 studies, which do not seem to be an issue in computational studies. Furthermore, in vivo and in vitro experiments only show quantifying data for Caco-2 permeable samples, leaving all impermeable samples as zeros in the permeation
measurements. On the other hand, computation studies can provide a magnitude for such impermeable samples, further providing data that can explain the degree of impermeability.

By using state-of-the-art equipment and methodologies, combining traditional chemistry and experiments to newest computation studies, an area never explored can be researched, and further deliver light into the darkness that exists in the world. For such reason, and being one of the earlier pioneers of the food science to combine and pursue computational study, incredible new findings of flavonoids, antioxidant potential, and human intestinal absorption have been found and reported throughout this manuscript.
References


Volsurf version 3.0 software by Molecular Discovery Ltd.


CHAPTER 2

LITERATURE REVIEW

2.1 Basic Flavonoid Structure

Flavonoids are derived from amino acids, tyrosine and phenylalanine, and has three aromatic ring structure. As it loses a hydrogen atom to quench a free radical, resonance occurs and stabilizes the flavonoid molecule after a hydrogen atom is donated. Additionally, previous studies (Vrielynck, Cornard et al. 1993)(Vrielynck, Cornard et al. 1993) noted that energy differences between the planar and twisted structure is minimal. However, flavones should exist as a planar structure due to their lower energy and more favorable condition to packing or rugosity. The presence of hydroxyl groups at the C-3 and C-4 position on the B ring of the flavonoids increases the antioxidant activity. Moreover, the basic structure of flavonoids is important as well. This can be seen in the example of kaempferol (Silva, Santos et al. 2002)(Silva, Santos et al. 2002), where the number of hydroxyl groups and their location on the aromatic ring determines antioxidant properties. Quantitative Structure-Activity Relationship study of 42 different flavonoids was performed. Using a beta carotene linoleate system oxidation to test antioxidant and antiradical activities of flavonoids (Burda and Oleszek 2001)(Burda and Oleszek 2001) found that the presence of hydroxyl groups at the C-3 position is associated with good antioxidant activity. Additionally, they also found that antiradical activity is primarily connected with free hydroxyl groups at the C-4 position.
A study conducted by (Lien, Ren et al. 1999)(Lien, Ren et al. 1999) shows the presence of a 3-OH, 4-oxo, and 0-dihydroxy in the B ring is needed in addition to a 2,3 double bond for the highest antioxidant potential. Additionally, higher antioxidant potentials were correlated with increasing numbers of hydroxyl groups. Besides structural placement of hydroxy groups, flavonoids show varying degrees of antioxidant behavior depending on the food system where being used (Chen, Chan et al. 1996)(Chen, Chan et al. 1996). According to an earlier study, ascorbic acid is stable under acidic conditions; however, when exposed to neutral or alkaline solution conditions, oxidation can rapidly occur. Thus, flavonoids can be used to preserve ascorbic acid, while providing antioxidants for food items (Thompson, Williams et al. 1976)(Thompson, Williams et al. 1976).

2.2 Existing Computational Studies

*Ab initio* molecular orbital calculation has been performed to fully optimize structure geometry. In this study, an explanation of catechin and taxifolin has been conducted and they have used Log P values to study lipophilicity of the compounds. Moreover, researchers used acidity values from a computation study and *in vivo* data (Teixeira, Siquet et al. 2005)(Teixeira, Siquet et al. 2005) to examine relationship between them. In another study of melatonin and related indoles, semiempirical AM1 (Austin Model 1) and DFT (Density Functional Theory) were used to measure changes in Gibbs free energy to test molecules for their antioxidant potential. In their experiment design, both vacuum and aqueous settings were employed to find out that no significant difference existed between the two settings. Experimenters concluded that this is due to a lack of charged species used in the computational methodology. The obtained data
was compared to other established data and found some qualitative similarity (Turjanski, Rosenstein et al. 1998)(Turjanski, Rosenstein et al. 1998). Similar studies conducted using DFT was conducted. B3LYP/6-31+G(d) density functional theoretical level was used to study QSAR. They used heat of formation as their primary measurement to observe antioxidant potential. They were able to find a close correlation between experimental data and calculated data (Vafiadis and Bakalbassis 2003)(Vafiadis and Bakalbassis 2003). Another research found that the dissociation enthalpy parameter of O-H bonds provides the best antiradical activity of chalcones. This study was conducted using the B3P86 theory level of quantum calculations (Kozlowski, Trouillas et al. 2007)(Kozlowski, Trouillas et al. 2007). Additionally, another study examined the length of the C-O bond from phenol molecules and discovered the correlation between proton affinities and electron transfer enthalpies (Klein and Lukes 2006)(Klein and Lukes 2006).

Predictions of thermodynamic properties such as enthalpy of formation, bond dissociation energy, and ionization potential of flavonoids were studied and compared to experimental values obtained from x-ray crystallography and calorimetric techniques. The researchers found good correlations between thermodynamic properties and experiment data (Mendoza-Wilson, Lardizabal-Gutierrez et al. 2007)(Mendoza-Wilson, Lardizabal-Gutierrez et al. 2007). The free energy relationship can be used to predict antioxidant potential of target compounds given the proportionality between Gibbs free energy of activation and the overall Gibbs free energy of reaction. Thus, this concept will allow numeric measurement of the target antioxidant (Rhodes, Tran et al. 2004)(Rhodes, Tran et al. 2004). In 2004, scientists used the Dragon program package (Todeschini, Consonni et al. 2002)(Todeschini, Consonni et al. 2002)
and PLS method to study flavonoids and to correlate antioxidant potential. In another study (Burda and Oleszek 2001), 36 flavonoids and their 2D descriptors were calculated. Then, the descriptors and antioxidant values were compared to experimentally measured data. More recent studies show acceptable model between predicted antioxidant data and experimental data (Farkas, Jakus et al. 2004). This study shows that antioxidant activity and structure can be assessed using computational study and the PLS statistical analysis.

2.3 Measurement of Antioxidant Potential

It is extremely important to quantitatively measure the antioxidant potential of flavonoids in both foods and model systems. Therefore, numerous methods have been developed, each having their own advantages and disadvantages. One of the more widely used methods is called the Trolox equivalent antioxidant capacity (TEAC) assay (Miller, Rice-Evans et al. 1993; Rice-Evans and Miller 1994; Pellegrini, Proteggente et al. 1999)(Miller, Rice-Evans et al. 1993; Rice-Evans and Miller 1994; Pellegrini, Proteggente et al. 1999). In this assay, inhibition of free radicals by ABTS or 2,2'-azinobis (3-ethylbenzothiazoline 6-sulfonate) is used and absorbance is measured. Then, absorbance is compared to the standards or Trolox equivalent. The advantage of this method is that it can be used to measure antioxidant potential for both lipophilic and hydrophilic compounds.

The TEAC assay can be used to measure the antioxidant activity of a compound. However, the TEAC does not always relate to the antioxidant potential because it measures how much radical is scavenged over time, where both reaction compounds and reacted
compounds play a role. Meanwhile, antioxidant activity measures the rate of radical scavenged and it is associated with only the initial reaction compound (Arts, Sebastiaan Dallinga et al. 2003). The oxygen radical absorbance capacity (ORAC) assay (Glazer 1990) is another popular assay to determine antioxidant activities in samples. In this assay, an fluorescenin and a peroxyl radical generator called 2,2'-azobis (2-amidinopropane) dihydrochloride are used to measure antioxidant potential. The strength of this assay lies in the combination of both inhibition time and percentage into a quantitative data.

Following ORAC, the DPPH- method (Brand-Williams, Cuvelier et al. 1995; Sánchez-Moreno, Larrauri et al. 1998) is a well recognized method to determine antioxidant potential, especially for plant samples. In this assay, reaction of the stable 1,1-diphenyl-2-picrylhydrazyl radical and donor sample and color change can be measured. When an antioxidant donates a hydrogen atom, a reduced form of DPPH loses color, which can be measured by using spectrophotometer at 520nm.

Others methods of measuring antioxidant potential include Ferric reducing ability of plasma (FRAP) assay (Benzie and Strain 1999), total radical trapping parameter (TRAP) (Wayner, Burton et al. 1985), dichlorofluorescin-diacetate (DCFH-DA) based assay (Valkonen and Kuusi 1997; Amado, Jaramillo et al. 2007), cyclic voltammetry method (Kohen, Beit-Yannai et al. 1999), total oxyradical scavenging
capacity (TOSC) assay (Winston, Regoli et al. 1998)(Winston, Regoli et al. 1998), and photochemiluminescence (PCL) assay (Popov, Lewin et al. 1987; Popov and Lewin 1994; Popov and Lewin 1996)(Popov, Lewin et al. 1987; Popov and Lewin 1994; Popov and Lewin 1996). Each of these methods has its own pros and cons that is specific to the sample, methods, cost, and easy of performing, accuracy, precision, and time issues.

2.4 Absorption Studies using Caco-2 Cells

For many compounds, including both drugs and nutrients, human absorption information is highly sought after. This absorption data can determine the proper dosage and recommendation intake levels for humans and other animals. Presently, antioxidants are considered as healthy components in foods, and the consumer-driven market is increasing at a fast pace. This is the driving force behind the formulation of new health foods containing flavonoids and other antioxidants. Therefore, quantification plays another level in food formulations and in the supplement sector of the world. Among many different methods to model human gastrointestinal absorption, Caco-2 cells are widely accepted and used worldwide (Murota, Shimizu et al. 2000; Murota, Shimizu et al. 2002)(Murota, Shimizu et al. 2000; Murota, Shimizu et al. 2002). The materials for Caco-2 studies include Caco-2 cells, buffer, and plates as shown in Figure 2.41. If the antioxidant of interest can pass through the medium, it will slowly move through the sample block. Likewise, if the sample cannot pass through the medium, no movement through the sample block will occur. The measurement usually takes 6 days to conclude and distance over time of sample traveled is recorded. It is at this point in the
experiment that quantification can be used, as greater the distance traveled associates with better intestinal absorption.

Currently, there have been many different experiments using Caco-2 cell studies to mimic or model human digestion (Umeda, Yano et al.; Delgado-Andrade, Seiquer et al. 2008; Kobayashi and Konishi 2008; Laparra, Tako et al. 2008; Lv, Wang et al. 2008; Waltenberger, Avula et al. 2008; Weerachayaphorn and Pajor 2008; Zhang, Yu et al. 2008)(Umeda, Yano et al.; Delgado-Andrade, Seiquer et al. 2008; Kobayashi and Konishi 2008; Laparra, Tako et al. 2008; Lv, Wang et al. 2008; Waltenberger, Avula et al. 2008; Weerachayaphorn and Pajor 2008; Zhang, Yu et al. 2008). These studies include permeation of special nutrients, toxins, and drugs. However, like all methods, Caco-2 studies have a few flaws. Not only does it take a long time to perform the experiment, many different types of Caco-2 cell cultures exist, temperature can affect permeability, it is impossible to use with gas and colorless samples or unstable compounds, and it requires complicated steps for proper data collection. However, the biggest flaw is the exclusion of quantity measurement for impermeable samples. In a Caco-2 study, impermeable samples are all given a value of 0 cm, making it difficult to determine the degree of impermeability.

In recent years, many companies and researchers already started to collect permeability data for vast number of samples. This is exactly why computational studies are so well accepted by the scientific community and is given prestigious welcome to provide further understanding of the complex working knowledge of intestinal permeation (Ungell 2004)(Ungell 2004). The basic concept lies in building a large database of real life experiment
Caco-2 values, obtaining 3D molecular structures, and a tool or software that can transform 3D molecular structures into many descriptors for statistical analysis. Therefore, using a large database of different molecules, Caco-2 data can be calculated from the molecular structure and related descriptors (Volsurf 2000-2004). In order to study molecules and structures, it is critical to understand basic molecular structure and how energies affect conformers as they exist in nature. In the next section, energy minimization will be discussed in detail.

2.5 Energy minimization

Molecules are dynamic in nature as they are not in a fixed state. Often, they exist in the lowest energy state possible. In order to capture such molecular movement and energies, different models have been developed. They include classical force fields such as AMBER, CHARMM, CVFF, GROMACS, GROMOS, ENZYMIX, ECEPP/2, and QCFF/PI as well as second generation or modified force fields like CFF, MMFF, MM2, MM3, and MM4. Each of these force fields has its own advantages and disadvantages. Therefore, careful selection of force fields to minimize molecules is a vital step in accuracy of a computational study.

The Merck Molecular Force Field (MMFF) has been developed and updated by a chemical company called the Merck & Co., Inc. The MMFF is largely based on the MM3 force field, created by Norman Allinger (Allinger, Yuh et al. 1989). It is especially valuable to analyze hydrocarbons and smaller molecular structures containing carbon atoms; however, it is not optimized for larger molecules and non-protein molecules. Because of these limitations, MMFF has been derived and is commonly used for wider range of
molecules. MMFF94 is designed to combine the advantages of MM3, OPLS, AMBER, and CHARMM to develop a force field that is not only great for smaller molecules, but can also be used for larger molecules (Halgren 1996). Due to a lack of high quality experiment data, MMFF94 has been computationally derived, tested, and validated by numerous experiments. Its parameters include structural optimization of 500 structures at HF/6-31G*, 475 structures at MP2/6-31G*, 380 structures at MP4SDQ/TZP level at MP2/6-31G*, 1450 structures at MP2/TZP, and has been expanded using 2800 additional structures from the Cambridge Structural Database. The parameters employed are algorithms that are used in the computational studies. They are both basis set of mathematical tools and theories developed for specific functions. Therefore, MMFF94 can be effectively used for many different molecules, including alcohols and phenols groups present in the flavonoids. In summary, the main focuses of MMF94 are conformational and intermolecular-interaction energies as well as inclusion of molecular geometries, torsional barriers, and intermolecular-interaction geometries, making itself a great tool for energy minimization (Halgren 1996; Halgren 1996; Halgren 1996; Halgren 1996; Halgren 1996; Halgren 1996; Halgren 1996; Halgren 1996).

2.6 Grid

The Grid program (Goodford 1985) is used to find interaction sites between a target molecule and one or more probes. Probes include water, hydrophobic, amphipatic, sp2 carboxy oxygen atom, neutral flat NH, sp2 N with a lone pair, sp3 amine NH3 cation, and sp2 phenolate oxygen. The program can be used to study many different arrays of
molecules and more than one molecule can be processed in a fairly short time period. It is popularly used in the protein and ligand binding simulations due to its efficiency in finding energetically favorable regions (Leach 2001)(Leach 2001). Moreover, it is widely employed for its simple systematic search, often called a grid search. The grid search is comprised of many steps. In the beginning, identification of all rotatable bonds occurs as bond angles and lengths are held fixed to a stationary position without movement. Then, each and all rotatable bonds are rotated 360 degrees in very small fixed increments that generates numerous conformations of the structure. This step continues until all rotatable bonds and associated structures are generated and minimized (Baroni, Costantino et al. 1993)(Baroni, Costantino et al. 1993). Due to such bond rotation, if a molecule has many rotatable bonds, the number of conformations generated can increase drastically, a phenomenon known as the combinatorial explosion (Leach 2001)(Leach 2001). However, flavonoids do not contain many rotatable bonds and the Grid can handle basic flavonoids structure quickly, especially after the MMFF94 energy minimization.

2.7 QSAR: Converting 3D information to Volsurf descriptors

The Volsurf software package is used to quantify the 3D structure of molecules into many 2D quantitative descriptors. These descriptors can be used to calculate and build multivariate models dealing with biological responses (Cruciani, Crivori et al. 2000)(Cruciani, Crivori et al. 2000). In details, there are several steps in which Volsurf turns 3D structure into 2D descriptors. First, 3D molecular field maps are converted to simple and easy to understand quantification. Normally, water and hydrophobic probes are used; however, other probes or grid maps produced by semi-empirical or different molecular mechanics could be used. In
many other computational software, molecular surface is always partitioned into small section of tesserae or polyhedral for further rendering and graphic effects. Conversely, Volsurf has a unique method where it starts with many tesserae containing the same information and builds a large framework that incorporates volume and or surface of relative information for a specific molecule. In 2D image, pixels with different color and pattern describe the image. In the 3D molecular field maps, information is contained in many small regular boxes called voxels. Voxels is related to attractive and repulsive forces between the two molecules and it is defined by volume, surface, and interacting energy. Volsurf uses these images to compute volume and surfaces. In the beginning, Voxels are given 1 if within the energy range, and 0 is out of the energy. Then they are separated and grouped according to their energy. However, when energy level changes, the size and shape changes accordingly, ultimately changing the information contained within. Therefore, Volsurf used many energy levels to calculate volume and surfaces of desired molecule. To summarize, Volsurf converts 3D information into simple, easy to understand surfaces and volumes where molecular recognition is achieved by image analysis software. Also, image compression is done by adding external chemical information. Many might argue that Volsurf does not translate 100% of 3D information into its description; however, practical examples exist where most of the relevant information has been extracted (Mannhold, Cruciani et al. 1999)(Mannhold, Cruciani et al. 1999).

2.8 Chemical interpretation

Flavonoids and biological interaction is a complicated to calculate, where many variables must be included to yield an accurate model. Variables include shape, electrostatic forces,
hydrogen bonds, and hydrophobicity. In this step, the GRID force field was selected to transform flavonoids into quantitative descriptors. Simply, volume and surface of the interaction contours are calculated. In more detail, water and hydrophobic probe interactions with target molecule provides 3D molecular description desired. Then, Volsurf descriptors, such as size, shape, hydrophilic, and hydrophobic regions are calculated by Volsurf. The software takes interaction between probes and flavonoids and makes a large list of descriptors. Then, these descriptors are used in conjunction with PCA and PLS to develop a mathematical model.

2.9 Volsurf descriptors: Size and Shape

The size and shape of molecules play a major role in chemistry due to its importance in assessing similarities, regularities, and correlation of drug design optimization. In Volsurf, four important size and shape categories are used to generate further data—molecular volume, molecular surface, ratio volume/surface, and molecular globularity. The molecular volume represents all molecular volume excluding water. Molecular surface represents accessible surface by water probe at +0.20kcal/mol during interaction. Ratio volume/surface refers to the degree of wrinkled surface, known as rugosity. Finally, molecular globularity refers to “sphereness” of a molecule is important measurement of molecular flexibility.

2.10 Volsurf descriptors: Hydrophilic regions

Hydrophilic regions are where molecular envelope attracts water molecule. In Volsurf, hydrophilic descriptors are defined in the molecular fields of -0.2 to -1.0 kcal/mol, including polarisability. This means that interaction between the probe and the molecule is assessed in a
few different energy levels. When considering capacity factors, ratio of the hydrophilic surface and total molecular surface, eight different energy levels are used to calculate, ensuring flexible, yet accurate data.

2.11 Volsurf descriptors: Hydrophobic regions

Hydrophobicity quantifies degree of resistance to polar solvents. In Volsurf, Grid uses dry probe, or 3D lipophilic regions to determine such. Likewise, Volsurf uses 0.0 to -2.0 kcal/mol energy to distinct such variable and eight different energy levels are used to ensure flexible and accurate data.

2.12 Volsurf descriptors: Interaction Energy Moments and others

Interaction energy moments are known as integy moments in the Volsurf. It is used to describe concentration of either hydrophilic or hydrophobic areas of the molecule. Additionally, Volsurf uses many other descriptors, such as local interaction of energy minima, energy minima distance, hydrophilic-hydrophobic balance, amphiphilic moments, critical packing parameters, hydrogen bonding, and polarisability to assess the molecule to generate accurate 2D descriptors. They are tools that allows conversion of structural information into easy to understand 2D numeric descriptors.

2.13 Volsurf statistics

According to Volsurf manual (Volsurf 2000-2004)(Volsurf 2000-2004), there are two major statistical tool that Volsurf uses to convert complicated descriptors into presentable mathematical models—PCA and PLS. PCA or principal components analysis summarizes
complex information into two parts of the X-matrix that can be easily be understood. These two parts include loading and scoring matrices.

Equation 1.  \[ X = TP' + E \]

According to the Eq. 1, loading matrices or P is composed of vectors called principal components and it is the linear combination of the original X-variables. Likewise, score matrices or T contains information that is a projection of principal components. Finally, E stands for unexpected variance that is not covered by either loading or score matrices. PCA is used to group those that are similar from those that are different. The first principal component explains the greatest variance in the data, and consequent principal component explains the next greatest variance in the data excluding the variance covered in the previous principal components. Therefore, first few principal components explain most of the data; therefore, lesser the principal components necessary to represent the data, more accurate the data. The actual calculation of the PC depends on matrix techniques explained by (Chatfield and Collins 1980)(Chatfield and Collins 1980).

PLS or partial least square is a statistical tool that is based on similar concept as PCA, but it relates to the Y variable based on X variables as shown in Equation 2. It can explain Y variables based on linear combination of X or independent variables. Just like PCA, first few latent variables or components explains the most of the variation in the data.

Equation 2.  \[ Y = f(X) + E \]
However, due to predicting of $Y$ by $X$ variable, error can be substantial if $X$ matrix contains more variables than objects. Due to this reason, 3:1 ratio of molecules to variable is necessary to validate multiple linear regression (Volsurf 2000-2004)(Volsurf 2000-2004).

Extensive use of components in both PCA and PLS is associated with lesser accurate data as each added components yield higher $R$-squared value, providing a false sense of explanation of data. In fact, there is a certain point where adding more components does not increase explaining power, but just adds noise to the model. Therefore, $\text{SPRESS}$, $\text{SDEP}$, and bootstrapping methods can be used to determine number of components necessary for a proper model, which is explained in the detail (Leach 2001)(Leach 2001).
Figure 2.41. Shows Caco-2 monolayer cells in a typical setup.
References


Umeda, D., S. Yano, et al. "Involvement of 67-kDa laminin receptor-mediated myosin phosphatase activation in antiproliferative effect of epigallocatechin-3-O-gallate at a physiological concentration on Caco-2 colon cancer cells." Biochemical and Biophysical Research Communications In Press, Uncorrected Proof.


Volsurf version 3.0 software by Molecular Discovery Ltd.


Chapter 3

QUANTITATIVE STRUCTURAL-ACTIVITY RELATIONSHIP OF FLAVONOIDS
AND ANTIOXIDANT POTENTIAL
ABSTRACT

Flavonoids are powerful antioxidants that are commonly available in many fruits and vegetables. They are linked with many positive health benefits and thus receiving increased attention in the foods and diets. There have been many studies available, but many are qualitative and lack good quantitative research. In this study, 23 common flavonoids were selected and their structural activity relationships have been compared to existing TEAC studies. The QSAR found that flavonoids antioxidant potential increase with increasing hydrophobicity, smaller molecular weight, lack of rugosity, and presence of hydroxyl groups. The PCA 7 component model explained 88.68% and PLS 6 component model showed R² value of 0.8626, indicating a good correlation between structures and antioxidant potential measured by TEAC.
Introduction

Flavonoids are polyphenolic compounds that are often found in nature (Shils and Goodhart 1956), and they are widely distributed throughout the plant world, especially in fruits and vegetables that are popular in the human diet. Increasing public awareness of flavonoids and their positive health roles fuel the need for further research into antioxidants. Such increase in attention is given to flavonoids due to their health benefits, including antioxidant activity and various cure or prevention of diseases (Packer, Hiramatsu et al. 1999; Teixeira, Siquet et al. 2005) (Packer, Hiramatsu et al. 1999; Teixeira, Siquet et al. 2005). These diseases include cancer, atherosclerosis (Hertog, Hollman et al. 1992; Keys 1995) (Hertog, Hollman et al. 1992; Keys 1995), cardiovascular diseases (Hertog, Feskens et al. 1993) (Hertog, Feskens et al. 1993), and degenerative diseases (Ejaz, Ejaz et al. 2006) (Ejaz, Ejaz et al. 2006). Moreover, flavonoids are shown to have anti-tumor effect (Middleton and Kandaswami 1992) (Middleton and Kandaswami 1992), anti-inflammatory (Middleton, Kandaswami et al. 2000) (Middleton, Kandaswami et al. 2000), and anti-microbial effects (Harborne and Williams 2000) (Harborne and Williams 2000). Moreover, flavonoids are powerful antioxidant that can quench free radicals and super oxide anion radical that are associated with many chronic and degenerative diseases. In order to assess flavonoids for antioxidant potential, many lab techniques exist. However, Quantitative Structural-Activity Relationship is a powerful method to further study the flavonoids.

QSAR can be used to predict ADME or absorption, distribution, metabolism, excretion of drugs, nutrients, and various molecules. It does this by calculating relevant descriptors for
ADME models, optimizing drug properties, and screening compound databases. Volsurf is an important QSAR tool that can translate 3D molecular structure into simple 2D descriptors to study flavonoids. The Volsurf can quickly convert structural information into simple descriptors that can be analyzed by statistical methods. The program uses different probes to calculate various molecular characteristics, such as size, shape, hydrophilic, and hydrophobic regions of the molecule (Volsurf version 3.0 software by Molecular Discovery Ltd)(Volsurf version 3.0 software by Molecular Discovery Ltd). However, the greatest advantage of this software is the quickness of calculation and its ability to calculate multiple molecules in a single run. Volsurf can operate in two major ways. It can be used to generate simple descriptors and probes to determine activities of molecular structures or it can be used with pre-existing models to compare and analyze new molecules to the database. Database includes models like blood brain barrier permeation, termodinamic solubility, Caco-2 permeation, biopharmaceutical classification, protein binding, volume of distribution, hERG, water and DMSO solubility, and CYP3A4 metabolic stability (Crivori, Cruciani et al. 2000; Lombardo, Obach et al. 2002; Oprea, Zamora et al. 2002; Cruciani and Meniconi 2003; Pearlstein, Vaz et al. 2003; Crivori, Zamora et al. 2004)(Crivori, Cruciani et al. 2000; Lombardo, Obach et al. 2002; Oprea, Zamora et al. 2002; Cruciani and Meniconi 2003; Pearlstein, Vaz et al. 2003; Crivori, Zamora et al. 2004).

Among many classes of antioxidants, flavonoids were studied due to their abundant nature in popular fruits such as apples, peaches, pears, plums, and cherries (Chun, Kim et al. 2003; Kim, Jeong et al. 2003)(Chun, Kim et al. 2003; Kim, Jeong et al. 2003). Many antioxidant data available study flavonoids in groups, as analyzed by ORAC. Therefore, additional work in the sub-class or individual flavonoids is needed to further understand physicochemical
characteristics that contribute to antioxidant potential based on TEAC experimental finding. This will not only reveal detailed structural characteristics of flavonoids, but also, show characteristics that dictate the level of antioxidant potential in flavonoid molecules. Furthermore, Volsurf has many great advantages. Compared to traditional wet chemistry methods, QSAR is much quicker, less costly, and can be used for unstable samples. Therefore, QSAR approach was used to study flavonoids and antioxidant potential in this study.

Materials and Methods

3.1 Structures and Energy Minimization

Twenty three molecular structure of different class of flavonoids were selected from the established work of (Rice-Evans and Miller 1998)(Rice-Evans and Miller 1998). Then, these molecules were searched and obtained from the chemacx.com, a structure database that is linked with ChemBio3D Ultra software package. Energy minimization was conducted for all 23 molecules using MMFF94 (Merck Molecular field 94) (Thomas and Halgren 1996)(Thomas and Halgren 1996). The maximum iteration was set at 5000 and minimum RMS gradient set at 0.0010 to ensure good minimized structure for further analysis. This means either molecules reach 5000 steps or RMS gradient of less than 0.0010 should be sufficient to determine that the molecular structure has been minimized to its energy minima. This method of energy minimization is called the Monte Carlo energy minimization. This method is chosen for the
experiment due to its strength dealing with conformational changes. The details and further working result can be refer to (Rubinstein 1981; Adams 1983)(Rubinstein 1981; Adams 1983)

3.2 Probes

After energy minimization, molecules were complied and loaded in the Volsurf program. The criteria used for Volsurf program includes, keep kont option, using OH2, Dry, and O probes. The OH2 probe is water, and it is designed to finds hydrophilic regions around the molecule. Moreover, it has sp³ tetrahedral geometry, or sp² flat trigonal geometry that is designed to donate two hydrogen bonds and accept two. The dry probe is the hydrophobic probe that finds hydrophobic regions when both probe and molecule of interest is immersed in water. The O probe is the carbonyl oxygen atom that accepts two hydrogen bonds in the direction of its loan pairs. It is especially useful for many other atoms that share similar geometry, such as aldehyde, amide, nitro, nitroso, phenolate, sulphonamide, and sulphoxide oxygens (Volsurf version 3.0 software by Molecular Discovery Ltd)(Volsurf version 3.0 software by Molecular Discovery Ltd). It is this step that the Grid program (Goodford 1985; Bobbyer, Goodford et al. 1989)(Goodford 1985; Bobbyer, Goodford et al. 1989) was used to calculate 3D molecular interaction fields. Afterwards, the 3D molecular interaction field is converted or calculated to simple molecular descriptors by the Volsurf program. This is followed by comparing descriptors to the TEAC data from a previous study (Rice-Evans and Miller 1998)(Rice-Evans and Miller 1998) as shown in Table 3.1.
**Result and Discussion**

The first part of the study was intended to look for the relationship between molecular structure and Volsurf descriptors. At this time, no biological activity input was performed. In the analysis, we found that the PCA (Principle Component Analysis) of 7 components is 88.68%. This is the percent of variance explained by the X-matrix or Volsurf descriptors. In the search of the best model, R-squared vs. Q-squared statistics tools were used to determine that the 7 components model showed the best correlation with the flavonoid structures (Leach 2001).

According to Figure 3.3, chrysin is an outlier among the group. This is explained by the fact that it is the only molecule lacking a hydroxy group in the B ring and that it has the least number of hydroxy groups. Additionally, molecules rutin, hesperidin, and narirutin are located far away from the rest of the molecules, representing themselves as outliers to the others. Although the difference is not readily visible from the molecular structure, energy minima revealed something much more valuable where these molecules have much higher total energy at 242.5, 213.0, and 237.8 K cal/mol in comparison to average of 66.7 K cal/mol for all molecules in the data set.

Following PCA analysis, we tried to predict properties of flavonoids using antioxidant activity as obtained from earlier work of Dr. Rice-Evan (Rice-Evans and Miller 1998)(Rice-Evans and Miller 1998) that shows TEAC values of numerous flavonoids, using PLS (Partial Least Square). In the PLS analysis, 6 components model has been utilized with R-square value of 0.8626, explaining a good correlation of TEAC values and the flavonoids structures. In order to
select the ideal number of components, SDEC (standard deviation of error of correlation) vs. SDEP (standard errors of prediction) and R-square vs. Q-square plots were examined (Volsurf version 3.0 software by Molecular Discovery Ltd; Leach 2001). Figure 3.4 shows the PLS coefficients plot that correlates antioxidant activity of the flavonoids to the Volsurf descriptors. Antioxidant activity increases with capacity factors (Cw1 – Cw4) measured at -0.2 -0.5 -1.0 -2.0 kcal/mol energy level. Capacity factors show the ratio of hydrophilic volume to the molecular surface. Moreover, antioxidant properties increase with best volumes (BV21 BV22) or measurement of three best local hydrophilic volumes. Finally, hydrophobic regions (D1 - D8) in the molecule have a direct relationship to increasing antioxidant properties. Log P derived from Volsurf descriptors Vs. experimentally observed water/octanol partition coefficient, are positively related to increasing antioxidant activity. Conversely, antioxidant activity decreases with polar interaction sites, ratio of volume to surface (R), rugosity or wrinkles in the molecules, hydrophilic regions (W1 – W8). Moreover, best volumes (BV11 BV21 BV31 BV12 BV22 BV32) that measures three best local hydrophilic volumes, integy moments (Iw1 - Iw8) that shows hydrated regions are clustered together in one part of the molecule decreases antioxidant potential. Furthermore, capacity factors (Cw6 - Cw8) measured at -4.0 -5.0 -6.0 kcal/mol energy levels shows the ratio of hydrophilic region and molecular surface, hydrophobic integy moments (ID1 - ID8) that shows unequal distribution of hydrophobic regions throughout the molecule, and hydrogen bonding (HB1 - HB8) all contribute to decreasing antioxidant potential. The findings are similar to the findings of (Farkas, Jakus et al. 2004)(Farkas, Jakus et al. 2004) that double bonds between 2 and 3 position.
of the C ring contributes to superior antiradical activities of flavonoids, creating less rugosity as molecules with less folds can donate hydrogen atoms much easier. Moreover, double bonds represent more hydrophilic activity in the ring, increasing antioxidant potential. The works of (Lien, Ren et al. 1999) has shown that antioxidant potential and number of hydroxyl groups has been clearly established. Therefore, strong hydrogen bond will make hydrogen atom donation more difficult, lowering antioxidant potential.
Conclusion

Volsurf computational study of 23 flavonoids was conducted to determine Quantitative Structural Activity Relationship. In this study, Volsurf converted 3D molecular structure into 2D descriptors and analyzed using PCA and PLS statistical analysis. The 7 components PCA explains 88.68% of the data and 6 components PLS has R-square value of 0.8626. Both of these statistical tools show that there is definitely a correlation between the molecular structure and TEAC value. Overall, the study showed that antioxidant properties of flavonoids heavily depended on hydrophobic regions, lack of rugosity, lack of strong hydrogen bonds. Although simple and much time consuming that many other computational study, Volsurf computational study has provided good descriptors and factors that determines both positive and negative effects of each descriptors to the antioxidant potential.
<table>
<thead>
<tr>
<th>Subclass</th>
<th>Molecule ID</th>
<th>Compound</th>
<th>TEAC value</th>
<th>Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flavanol</td>
<td>2</td>
<td>epigallocatechin gallate (EGCG)</td>
<td>4.80</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>epicatechin (EC)</td>
<td>2.50</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>taxifolin</td>
<td>1.90</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>catechin</td>
<td>2.40</td>
<td></td>
</tr>
<tr>
<td>Flavonol</td>
<td>7</td>
<td>quercetin</td>
<td>4.70</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>myricetin</td>
<td>3.10</td>
<td></td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>morin</td>
<td>2.55</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>rutin</td>
<td>2.40</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Compound</td>
<td>Value</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>--------------------------</td>
<td>-------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>kaempferol</td>
<td>1.34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>luteolin</td>
<td>2.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>apigenin</td>
<td>1.45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>chrysin</td>
<td>1.43</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>naringenin</td>
<td>1.53</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>hesperetin</td>
<td>1.37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>hesperidin</td>
<td>1.08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>narirutin</td>
<td>0.76</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>dihydrokaempferol</td>
<td>1.39</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No.</td>
<td>Compound</td>
<td>Value</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>------------</td>
<td>-------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>eriodictyol</td>
<td>1.80</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[Diagram]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>delphinidin</td>
<td>4.44</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[Diagram]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>cyanidin</td>
<td>4.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[Diagram]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>peonidin</td>
<td>2.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[Diagram]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>malvidin</td>
<td>2.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[Diagram]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>pelargonidin</td>
<td>1.30</td>
<td>[Diagram]</td>
<td></td>
</tr>
</tbody>
</table>

**Table 3.1.** Compounds used for the analysis. Compounds are arranged by their subclass of flavonoids. The data set was obtained from (Lien, Ren et al. 1999)(Lien, Ren et al. 1999) and (Rice-Evans and Miller 1998)(Rice-Evans and Miller 1998).
Table 3.2. MMFF94 Energy Minimized structures.

<table>
<thead>
<tr>
<th>Molecule ID</th>
<th>Total Energy (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>34.8</td>
</tr>
<tr>
<td>4</td>
<td>51.2</td>
</tr>
<tr>
<td>5</td>
<td>70.0</td>
</tr>
<tr>
<td>6</td>
<td>43.1</td>
</tr>
<tr>
<td>7</td>
<td>43.4</td>
</tr>
<tr>
<td>8</td>
<td>51.4</td>
</tr>
<tr>
<td>9</td>
<td>54.5</td>
</tr>
<tr>
<td>10</td>
<td>242.5</td>
</tr>
<tr>
<td>11</td>
<td>57.5</td>
</tr>
<tr>
<td>12</td>
<td>10.0</td>
</tr>
<tr>
<td>14</td>
<td>34.7</td>
</tr>
<tr>
<td>15</td>
<td>46.0</td>
</tr>
<tr>
<td>17</td>
<td>35.5</td>
</tr>
<tr>
<td>18</td>
<td>44.9</td>
</tr>
<tr>
<td>19</td>
<td>237.8</td>
</tr>
<tr>
<td>20</td>
<td>213.0</td>
</tr>
<tr>
<td>21</td>
<td>21.3</td>
</tr>
<tr>
<td>22</td>
<td>21.3</td>
</tr>
<tr>
<td>23</td>
<td>31.4</td>
</tr>
<tr>
<td>24</td>
<td>31.3</td>
</tr>
<tr>
<td>26</td>
<td>52.0</td>
</tr>
<tr>
<td>27</td>
<td>63.9</td>
</tr>
<tr>
<td>28</td>
<td>42.9</td>
</tr>
</tbody>
</table>
Figure 3.3. 2D PCA score plot

2D PCA score plot shows relative position of molecules in comparison with others. It is used to find outliers and cluster of related molecules.
Figure 3.4. PLS Coefficient plot

PLS Coefficient plot shows relationship among variables that contribute to increasing or decreasing antioxidant potential of the flavonoids.
Reference


Volsurf version 3.0 software by Molecular Discovery Ltd.
Chapter 4

QUANTITATIVE STRUCTURE-PERMEABILITY RELATIONSHIP OF FLAVONOIDS USING CACO-2 CELLS: INDEPTH STUDY OF HUMAN INTESTINAL PERMEATION OF LAVONOIDS
ABSTRACT

Flavonoids are widely available compounds that are often present in human diet. Many studies of flavonoids exist, but not many studies focus on bioavailability. As interest in flavonoids grow stronger due to health benefits of antioxidant, further bioavailability study is essential. Over the years, Caco-2 cells are used to model human intestinal absorption of many compounds. In recent years, computational studies have been developed in conjunction with experimental data to develop a highly sophisticated and accurate Caco-2 model. In this study, QSPR or Quantitative Structural-Permeability Relationship study of 17 molecules was conducted for Caco-2 study. All of the molecules were obtained from the Cambridge Software Molecular Database. The structures were energy minimized using MMFF94 and converted into 2D descriptors by the Grid and Volsurf Program. Then, descriptors were compared to the 751 existing Caco-2 database. 94.1% of QSAR structures match permeability data of experimentally derived data. Moreover, we were able to conclude hydrophobic and smaller molecular weight flavonoids show much higher permeability.
Introduction

A diet rich in fruits and vegetables has been related to various positive health benefits. These include many chronic conditions that affect numerous people world-wide (Doll 1990; Ames, Shigenaga et al. 1993; Sun, Chu et al. 2002)(Doll 1990; Ames, Shigenaga et al. 1993; Sun, Chu et al. 2002). These protective effects are related to antioxidants (Moreira, Fraga et al. 2004; Kwon, Murakami et al. 2005)(Moreira, Fraga et al. 2004; Kwon, Murakami et al. 2005). Flavonoids are widely present in the plant kingdom and account for approximately half of more than 8000 phenolic compounds found in nature (Harborne, Baxter et al. 1999)(Harborne, Baxter et al. 1999). Additionally, they are associated with health benefits of protection against atherosclerosis (Hertog, Hollman et al. 1992; Keys 1995)(Hertog, Hollman et al. 1992; Keys 1995), inflammatory conditions (Middleton, Kandaswami et al. 2000)(Middleton, Kandaswami et al. 2000), allergenic conditions, artherogenic conditions, microbial infections, thrombotic problems, and provide cardio protective effects beyond basic nutrition (Manach, Mazur et al. 2005)(Manach, Mazur et al. 2005). These medicinal effects are believed to be the antioxidant activity of the flavonoids which quench free radicals, free radical anions, and peroxy radicals. However, proper dosage of flavonoids in the human diets is unknown due to limited studies on bioavailability and bioassesibility. Therefore, digestion and absorption of flavonoids in the human gastrointestinal system is important, and related studies have been conducted using different methods, such as in vivo (Hollman, Vries et al. 1995; Hollman, Trijp et al. 1997)(Hollman, Vries et al. 1995; Hollman, Trijp et al. 1997), in vitro (Kuo 1998; Murota, Shimizu et al. 2000; Masataka Oitate 2001)(Kuo 1998; Murota, Shimizu et al. 2000; Masataka Oitate 2001), and computational studies (Ekins, Durst et al. 2001; Ponce, Perez et al. 2004)(Ekins,
Durst et al. 2001; Ponce, Perez et al. 2004). Diffusion mechanism and permeation study is divided into two main types—passive diffusion and active transport. For the passive diffusion, 2/4/A1 cell, PAMPA, chromatography, and LogP are used. For active transport, Caco-2, MDCK-MDR1, and transformed cells with transporters are used (Ungell 2004)(Ungell 2004). Among these however, Caco-2 cells, (Hidalgo, Raub et al. 1989)(Hidalgo, Raub et al. 1989) or differentiated monolayer cells of human colon adenocarcinoma, are selected due to their wide acceptance as a model for human intestinal absorption. The Caco-2 cell method is active transport method that can determine permeability of samples through the medium. Onto this medium, sample and buffer mix is loaded. The samples that need specific molecules or chemicals, HPLC or other separation methods are necessary to obtain more pure forms. After a certain time period, ranging from few minutes to days, the distance of permeation is measured. Although the distance permeable is different from sample to sample, many experiments use $10^{-6}$ cm/sec or permeation length over time (Tammela, Laitinen et al. 2004; Zuo, Zhang et al. 2006; Serra, Mendes et al. 2008)(Tammela, Laitinen et al. 2004; Zuo, Zhang et al. 2006; Serra, Mendes et al. 2008).

In terms of faster and newer scientific methods, computational studies and models are widely accepted for drug lead optimization and knowledge that can benefit the community. Computational methods, such as quantitative structure-permeability relationships, are important and needed to predict intestinal absorptivity of flavonoids. In QSPR, molecular structures of compounds are turned into descriptors and matched with preexisting database for the purpose of finding structural-permeability relationship. The end result is quantitative permeation data based on structure (Hansch, Leo et al. 2004)(Hansch, Leo et al. 2004). Not
only is this method faster, economically efficient, and has less variability due to the nature of the software, but also it can be used for molecular structures that are difficult to isolate, unstable, and can use different solvents to develop a model that can predict human digestion and absorption model.

In the experiment, Caco-2 cells were chosen because previous studies by (Murota, Shimizu et al. 2000; Murota, Shimizu et al. 2002)(Murota, Shimizu et al. 2000; Murota, Shimizu et al. 2002) have shown that flavonoids are ideal candidates for Caco-2 studies due to their lipophilic nature. Although Caco-2 cells have been used to study bioavailability of many compounds, many flaws do exist. It is costly, time consuming, lack of precision among labs, and requires expertise to handle. Therefore, many recent studies used QSPR have to study permeability. One of the tools for QSPR is Volsurf and it contains 751 Caco-2 permeation data and their structural descriptors.

Volsurf is a tool that can predict properties various molecules and serve important to drug-lead optimization and bioavailability of numerous bioflavonoids. It works by converting 3D molecular information into 2D descriptors, such as hydrophobic regions, hydrophilic regions, strength of hydrogen bonds, log cp values, size, and other characteristics of molecule using various different probes (Volsurf 2000-2004)(Volsurf 2000-2004). The converted descriptors can be matched with database of 751 molecules where the descriptors and Caco-2 permeability has been integrated. In order to generate descriptors, Volsurf uses probes, designed specifically to examine molecules via measuring interaction energies between the probes and molecules of interest. However, one of the most powerful advantages is its accuracy and short
amount of time it takes to convert 3D molecules into 2D descriptors. Furthermore, existing cell culture studies can be used to correlate and build a database that can relate molecules and their properties to selected variable in study (Ungell 2004)(Ungell 2004). Therefore, library of flavonoids was selected to run in the Volsurf program in order to compare Caco-2 properties of experimentally found data to the database exist in the Volsurf. Not only has this method required less time requirements, the models exist in the Volsurf database has been proven to predict various properties of array of different drugs. Moreover, there has been an increase in molecular databases that incorporate structural and experimental Caco-2 data and prediction and accuracy of Caco-2 permeability increases as the database grows (Ungell 2004)(Ungell 2004). Furthermore, experimental Caco-2 studies is expensive, time consuming, lacks quantitative measurement of impermeable molecules, and yields lower quality data due to different materials and methods being used. In a computation study, even the impermeable molecules are assigned a quantitative value, faster to run, and is consistent world-wide. In order to study QSPR, Volsurf (Volsurf version 3.0 software by Molecular Discovery Ltd)(Volsurf version 3.0 software by Molecular Discovery Ltd) is used to predict permeability of flavonoids using its library of 751 related, yet diverse chemical compounds and their experimental Caco-2 permeability data. Additionally, non flavonoid, phenolic compounds were selected to test the Volsurf model using external compounds.
Methods

A Total of 17 structures of 12 flavonoids and 5 non-flavonoids phenolic compounds and their experimental Caco-2 permeabilities were selected from existing experiments (Tammela, Laitinen et al. 2004; Serra, Mendes et al. 2008)(Tammela, Laitinen et al. 2004; Serra, Mendes et al. 2008). These are shown in Figure 4.1. The selected molecules were obtained from the Cambridge Structural Database, a large database that works with ChemBio3D Ultra program (Halgren 1996)(Halgren 1996).

Energy Minimization

All structures were minimized using MMFF94 (Thomas and Halgren 1996)(Thomas and Halgren 1996). The details of this energy minimization called for maximum iteration at 5000 with minimum RMS gradient of 0.0010. This was conducted to ensure that the energies of molecules were closely accurate to its true energy minima. Energy minimization step was carefully conducted with maximum precision as improper energy minimization would not generate accurate data. MMFF94 was selected due to its powerful ability to minimize both small organic and large structures (Halgren 1996)(Halgren 1996).

QSPR Modeling

After MMFF94 energy minimization, all molecules were compiled and submitted to the Volsurf program. During this step, all 3D molecules were turned to 2D descriptors. In this step, OH2, Dry and O probes were used. The probes work to determine important descriptors for molecules. For example, OH2 probe is especially useful in determination of hydrophilic regions,
dry probe looks for hydrophobic regions, and O looks for hydrogen bonds in general. More detail of different probes and their functions are listed in the Volsurf manual (Volsurf version 3.0 software by Molecular Discovery Ltd) and available online http://www.moldiscovery.com/docs/volsurf/intro.html. For the Caco-2 study, 751 different molecules, their related 2D descriptors, and experimentally found Caco-2 values were incorporated into the database. In this experiment, molecular structures of the flavonoids are imposed onto the Volsurf database to determine human intestinal permeability. This is possible by comparing 2D descriptors from flavonoids and comparing to the existing Volsurf library, thus predicting Caco-2 permeation of flavonoid to high degree of accuracy.

**Result and Discussion**

Figure 4.2 shows Caco-2 permeation of all 17 molecules. The three linear lines in the middle of the graph is the dividing line that determines whether a molecule can permeate through the Caco-2 monolayer or not. Two lines around the center line is the confidence interval for the permeation separation line. Molecules further away from the division line to the right shows increasing Caco-2 permeation. Likewise, molecules further away from the left shows decreasing Caco-2 permeation.

The Volsurf model predicted at 94.1% accuracy for all molecules whether they are permeable or impermeable. In detail, all structures were categorized into either permeable or impermeable groups. The relationship between the experimental and the computational data indicate 94.1 % accuracy that QSRP study can determine permeability of flavonoids. Methylgallate and morin are close to the dividing line, indicating increasing chance of error in
true representation of their Caco-2 permeability. However, previous studies show that morin is not permeable, while methylgallate has very slightly permeable (Tammela, Laitinen et al. 2004); therefore, computational study by Volsurf shows its high degree of sensitivity for the Caco-2 permeation. Naringin, diosmin, and hesperidin show high impermeability. This is due to their large molecular size with addition of two gallate groups to basic flavonoid structure. This also correlates to data found in the previous experiment that related increase in molecular size decreases caco-2 permeability (Gonthier, Donovan et al. 2003; Tsang, Auger et al. 2005). In contrast, flavone has the highest permeability due to lacks of hydroxy groups in the molecule, increasing hydrophobicity. In another study, similar result was found (Murota and Terao 2003). They concluded that flavonoids permeability increase as hydrophobicity increases. Paracetamol shows the next highest permeation data, and is largely due to its small molecular weight and size.

In the following portion of the study, we cross examined the Volsurf using phenolic compounds that are non-flavonoids. This was conducted to test the Volsurf and its processing abilities for molecular size, hydroxy groups, and reaction to the hydrophobicity. According to earlier works by (Salucci, Stivala et al. 2002), smaller molecules like gallic acid have good permeability. Further examination shows increasing permeability of gallic acid as hydrocarbon tail length increases. Longer tail increases hydrophobic nature of the molecule, increasing permeability through the Caco-2 cells.
Finally, a model was created using data of Caco-2 permeable molecules. The closest distance between the permeation separation line and the flavonoids were measured. The predicted QSPR values were tested via comparison with experimental Caco-2 data as shown in Figure 4.3. Methylgallate has the highest correlation, followed by diosmetin and hesperetin. On the other hand, paracetamol showed the least correlation. Among all the molecules, flavonoid molecules composed of flavones, naringenin, diosmetin, and hesperatin. All of the flavonoids molecules showed great relationship between experimental and QSPR data. More importantly, Figure 4.3 shows that Volsurf is fairly accurate in representation of Caco-2 permeability for both phenolic compounds and flavonoids.

**Conclusion**

Quantitative structure-permeability relationship using Volsurf and Caco-2 monolayer cells has been conducted. As with the general understanding in the beginning, impermeable molecules were associated with magnitude, that explains how much less likely the permeation through the Caco-2 cells occur. This could be very useful during modification of compounds during drug-lead optimization. Overall, Volsurf direct prediction was able to determine with 94.1% of accuracy whether a molecule will be permeable or not. However, prediction of permeable data proved to be more difficult. But with more data points in the database, the model can become very accurate in the years to come.
### Tables and Figures

<table>
<thead>
<tr>
<th>Molecule ID</th>
<th>Total Energy (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>43.1</td>
</tr>
<tr>
<td>2</td>
<td>51.2</td>
</tr>
<tr>
<td>3</td>
<td>52.4</td>
</tr>
<tr>
<td>4</td>
<td>10.0</td>
</tr>
<tr>
<td>5</td>
<td>12.1</td>
</tr>
<tr>
<td>6</td>
<td>48.2</td>
</tr>
<tr>
<td>7</td>
<td>35.5</td>
</tr>
<tr>
<td>8</td>
<td>224.8</td>
</tr>
<tr>
<td>9</td>
<td>12.9</td>
</tr>
<tr>
<td>10</td>
<td>14.0</td>
</tr>
<tr>
<td>11</td>
<td>43.4</td>
</tr>
<tr>
<td>12</td>
<td>71.5</td>
</tr>
<tr>
<td>13</td>
<td>-11.9</td>
</tr>
<tr>
<td>14</td>
<td>233.0</td>
</tr>
<tr>
<td>15</td>
<td>237.7</td>
</tr>
<tr>
<td>16</td>
<td>39.4</td>
</tr>
<tr>
<td>17</td>
<td>44.9</td>
</tr>
</tbody>
</table>

**Table 4.1.** MMFF94 Energy Minimized structures.
<table>
<thead>
<tr>
<th>Molecule ID</th>
<th>Compound</th>
<th>Experimental Caco-2 value Papp ($10^6$ cm/s)</th>
<th>Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Catechin</td>
<td>0</td>
<td><img src="#" alt="Structure" /></td>
</tr>
<tr>
<td>2</td>
<td>Epicatechin</td>
<td>0</td>
<td><img src="#" alt="Structure" /></td>
</tr>
<tr>
<td>3</td>
<td>Flavone</td>
<td>380</td>
<td><img src="#" alt="Structure" /></td>
</tr>
<tr>
<td>4</td>
<td>Luteolin</td>
<td>0</td>
<td><img src="#" alt="Structure" /></td>
</tr>
<tr>
<td>5</td>
<td>Methylgallate</td>
<td>2.9</td>
<td><img src="#" alt="Structure" /></td>
</tr>
<tr>
<td>6</td>
<td>Morin</td>
<td>0</td>
<td><img src="#" alt="Structure" /></td>
</tr>
<tr>
<td>7</td>
<td>Naringenin</td>
<td>29.4</td>
<td><img src="#" alt="Structure" /></td>
</tr>
<tr>
<td>No.</td>
<td>Compound</td>
<td>Value</td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>--------------</td>
<td>-------</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Naringin</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Octylgallate</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Propylgallate</td>
<td>7.3</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Quercetin</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Ketoprofen</td>
<td>24.6</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Paracetamol</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chemical Name</td>
<td>Value</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>----------------</td>
<td>-------</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Diosmin</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Hesperidin</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Diosmetin</td>
<td>76.2</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Hesperetin</td>
<td>47.1</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 4.1** shows experimentally observed Caco-2 permeation value, molecular structure, and molecular name and related coding for the experiment.
Figure 4.2. shows PLS plot of Caco-2 Permeation data of flavonoids compared to experimentally observed data. Further away from the linear line to the right side shows increasing or higher Caco-2 permeation data, while Further away from the linear line to the left side shows decreasing or lower Caco-2 permeation data. Empty circles represent existing data points from the VolSurf library and filled circles represent 17 molecules of interest in this study.
Figure 4.3. Caco-2 shows permeation relationship between experimental and QSPR data. Paracetamol has the largest gap, while methylgallate has the lowest gap.
References


Volsurf version 3.0 software by Molecular Discovery Ltd.


Chapter 5

SUMMARY AND CONCLUSION

In spite of many studies of flavonoids, good quantitative studies of flavonoids are scarce. With that in mind and that these flavonoids and antioxidants are taking a large fame in the food and nutrition market, these two studies were conducted. In the QSAR study, high antioxidant potential of flavonoids correlates with small molecular structures, number of hydroxy groups on the structure, lack of rugosity, and hydrophobicity. In the QSPR study, the flavonoids that are smaller and more hydrophobic have higher permeability through the Caco-2 cells. In both cases, computational study using Volsurf and Grid show that they are fast, dependable, and can be expanded in the future to generate even more accurate data than ever before.