

On the Prediction of Polyphenol Properties

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de rector magnificus prof. ir. K.Ch.A.M. Luyben
voorzitter van het college voor promoties,
in het openbaar te verdedigen op Dinsdag 30^{de} 2015 om 12:30 uur

door

David MÉNDEZ SEVILLANO

Chemical Engineer
geboren te San Andrés del Rabanedo, Spanje

Dit proefschrift is goedgekeurd door de promotor:

Prof. dr. ir. L.A.M. van der Wielen

Copromotor: Dr. Ir. M. Ottens

Samenstelling promotiecommissie:

Rector Magnificus

voorzitter

Prof. dr. ir. L.A.M. van der Wielen

Technische Universiteit Delft, promotor

Dr. Ir. M. Ottens

Technische Universiteit Delft, copromotor

Independent members:

Prof.dr.ir. C.G.P.H. Schroen

Wageningen University of Research

Prof.dr.ir. M.C. Kroon

Technische Universiteit Eindhoven

Prof.dr. G.J. Witkamp

Technische Universiteit Delft

Dr. A. Krijgsman

Unilever

Prof.dr.ir. H.J. Noorman

Technische Universiteit Delft, reservelid

The research described in this thesis was performed at the Department of Biotechnology, Faculty of Applied Sciences, Delft University of Technology, the Netherlands. The research was financially supported by the Institute for Sustainable Process Technology (ISPT) in the project FO-10-03.

Cover: Diego García

“It is change, continuing change, inevitable change, that is the dominant factor in society today. No sensible decision can be made any longer without taking into account not only the world as it is, but the world as it will be.... This, in turn, means that our statesmen, our businessmen, our everyman must take on a science fictional way of thinking.”

-Isaac Asimov-

Table of Contents

Chapter 1: Introduction	7
Chapter 2: Model comparison for the prediction of Green Tea Catechins Solubility in ethanol/water mixtures	27
Chapter 3: MPP-UNIFAC, A Predictive Activity Coefficient Model for Polyphenols	57
Chapter 4: Resin Selection for the Separation of Caffeine from Green Tea Catechins	81
Chapter 5: Mechanism of Isoflavone Adsorption from Okara extracts onto Food-Grade Resins	101
Chapter 6: A Thermodynamic Lattice Model Describing Adsorption of Phytochemicals onto Macroporous Polymeric Resins	127
Chapter 7: Outlook	149
Summary/Samenvatting	151
Curriculum Vitae	159
Publications and Conferences	161
Acknowledgements/Dankwoord	163

Chapter 1: Introduction

1.1. Background and Motivation

Food materials contain many useful compounds that either get discarded during processing or not entirely exploited. The generated waste is around 38% of the original material for fruits or 57% in the case of meat in the United States¹ and up to 97% of this waste is usually discarded or used as landfill². Currently there are several initiatives to use this waste as starting material for other processes³ like, for example, the production of bio-hydrogen.⁴

This food waste still contains many high-value compounds such as proteins, vitamins or polyphenols. Some of these compounds show favourable health properties which classifies them as nutraceuticals. A very interesting family of nutraceuticals are the flavonoids because of their antioxidant properties. Flavonoids are present in small concentrations in the food matrix. The key to their purification is to bring them into solution and treat them via different unit operations until the required purity and composition is met.

The design of the purification process, or processes in general, may be aided by mechanistic mathematical models⁵. Process modelling in Chemical Engineering, Biotechnology or Food Processing is a combination of mass balances, energy balances, mass transfer and equilibrium models. Some of these models are required in a specific step of the process (like isotherm models in chromatography) and some of them are required throughout the process (like solubility giving the maximum concentration).

The use of these models decreases dramatically the time and money put into a process design by doing process optimization *in-silico*, which allows to optimize the

yield and productivity by considering all possible scenarios and not only the ones that can be deduced from prior experience.

In this work, we apply state of the art models to predict and describe the behaviour of flavonoids in different steps of the process validating them against experimentally determined properties. Furthermore, when the experiments show a different behaviour than the one predicted by the models, we develop new models to capture this behaviour and statistically analyse its validity. We limit the study to equilibrium models since good predictions can be made for kinetic phenomena based on existing models and correlations, but prediction of equilibrium phenomena is less developed for these molecules.

Below an overview of the molecules this thesis deals with, as well as the models describing equilibrium properties will be introduced.

1.2. Flavonoids

1.2.1. Definition

Flavonoids are a large group of phytochemicals that are formed by two phenyl rings joined by a 3-carbon chain (Figure 1). In most cases, that chain is bound through an oxygen to the first phenyl ring (ring A in Figure 1).

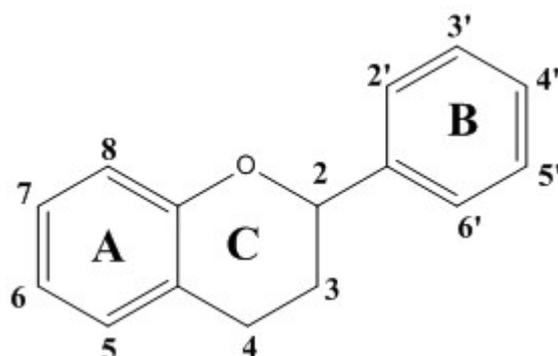


Figure 1: Typical backbone of a flavonoid molecule

Flavonoids are derived from phenylalanine in plant cells⁶ which is used for the production of p-coumarate, a direct precursor of the flavonoids. Each of the flavonoid-producing plants has a different pathway that decorates the flavonoid with different functionalized groups that can vary from a hydroxyl group to a sugar moiety.

A large number of flavonoids have been discovered in a large variety of plants such as tea, soy, berries, grapes, etc. These flavonoids have been associated to the beneficial properties that the ingestion of these plants can give and many studies are being conducted to obtain a better understanding of the influence of said flavonoids on human health⁷⁻¹¹. Besides *in-vivo* studies on human population many studies *in vitro*¹² and in different types of animals¹³ are being published as well with new evidence on the beneficial properties of these compounds.

1.2.2. Types

As previously stated, different functional groups define the different flavonoids. However, in some cases the differences are in the backbone (i.e. a ketone group in C4 of Figure 1) and based on those changes, the flavonoids are separated into different subgroups:

1.2.2.1. Major flavonoids. They are the biggest subgroup. Their main molecular difference from other groups is that the ring B (Figure 1) is bound to the carbon 2 and this carbon is bound to the ring A through an ether bond. This subgroup consists of several families (Figure 2) that differ mainly on the ring C. Well-known polyphenols can be found amongst these families such as kaempferol, quercetin (flavonols), luteolin, apigenin (flavones), hesperidin, naringin (flavanones), the catechins (flavanols), delphinidin and pelargonidin (as part of the anthocyanins).

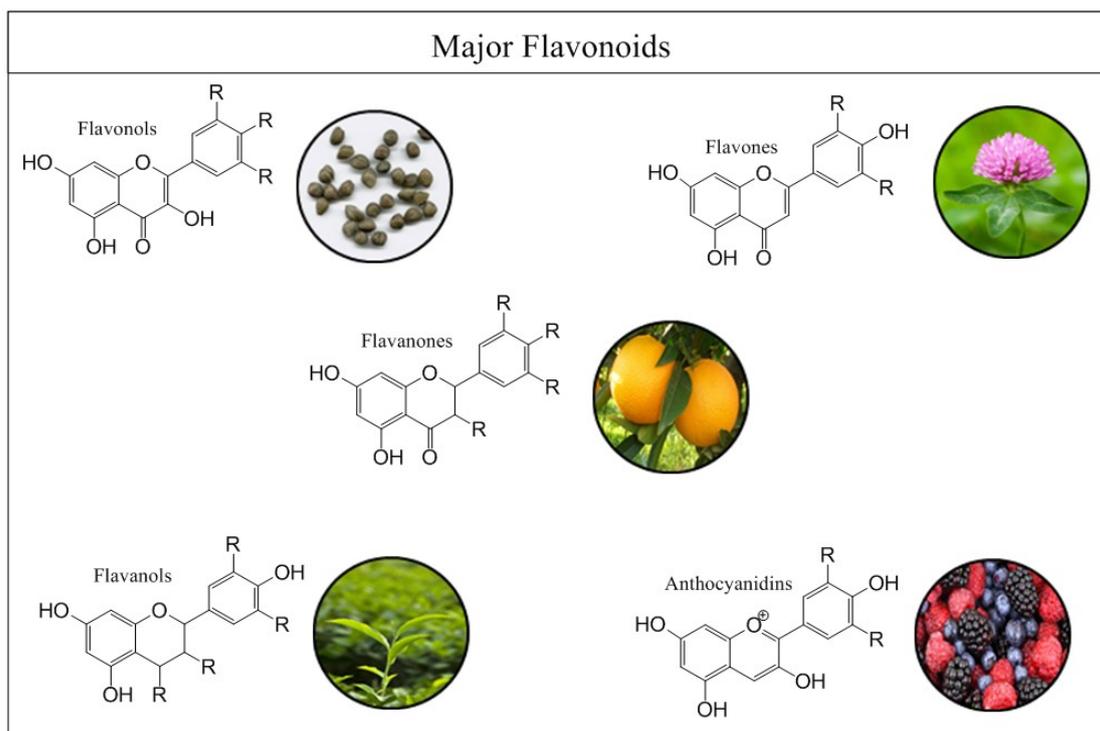


Figure 2: Different types of major flavonoids

1.2.2.2. Isoflavonoids: The polyphenols in this subgroup differ from others in the binding between rings B and C (Figure 1). In this case, the ring B is bound to ring C in carbon 3, instead of carbon 2 like the major flavonoids. Isoflavones are part of the so-called “phytoestrogens”,¹⁴ plant chemicals that can cause estrogenic effects due to their similarity with 17- β -estradiol (also in Figure 3). This subgroup consists of three families being isoflavones the most important of them with polyphenols like genistein or daidzein. Pumilanol (isoflavanols) or glyasperin F (isoflavanones) are other flavonoids that belong to this subgroup.

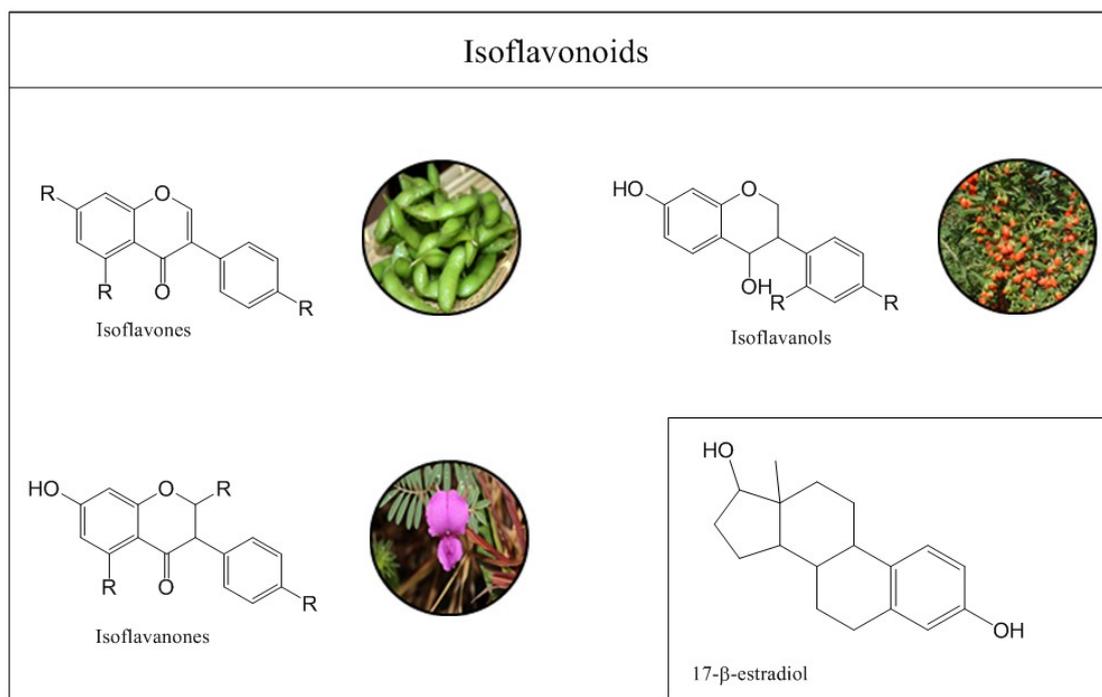


Figure 3: Different types of Isoflavonoids and an estrogen molecule.

1.2.2.3. Neoflavonoids. Neoflavonoids are a rare type of polyphenols (Figure 4).

Their main chemical feature is that the ring B (Figure 1) is bound to the ring C in carbon 4, instead of carbon 2 (like the major flavonoids) or carbon 3 (like isoflavonoids). Dalbergin or nivetin are examples of neoflavonoids.

1.2.2.4. Stilbenoids. This last subgroup of flavonoids shares the least resemblance with the flavan molecule (Figure 1). In stilbenoids (Figure 4) only ring A and B remain and they are linked by an unsaturated 2 carbon chain which enables geometric isomerism within the subgroup. The best and most known example of stilbenoids is trans-resveratrol present in wine.

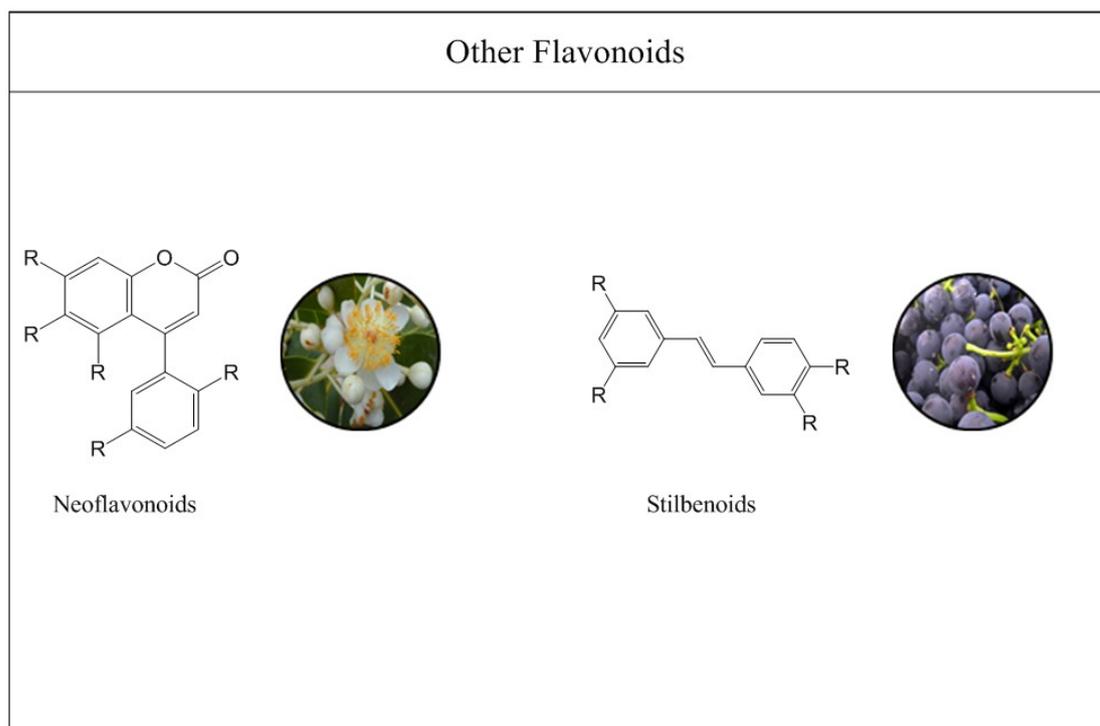


Figure 4: Different types of major flavonoids

1.2.3. Current claims

The influence of flavonoids on health can be classified by the different effects that flavonoids show.¹⁵

1.2.3.1. Effects on the cardiovascular system: McCullough et al. shows a clear lower risk of deaths related to cardiovascular disease with flavonoid intake.¹⁶ The studied flavonoids studied belong to the major flavonoids group, namely anthocyanins, flavanols, flavones, flavanols. Cassidy et al.¹⁷ shows a reduction in the cases of hypertension with anthocyanin and flavone (apigenin) intake.

1.2.3.2. Effects on the nervous system: Several studies are taking place to identify flavonoids that inhibit an enzyme responsible for some of the symptoms of Alzheimer's disease¹⁸⁻²⁰ and others are focussing on the their use for improving brain activity.²¹

1.2.3.3. Effects on cancer: Some examples such as quercetin (a flavonol) showed to lower the risk of lung cancer.²² Epigallocatechin gallate (a flavanol present in green

tea) showed positive effects as an antitumor-promotor in different skin tumors.⁷

Genistein inhibits growth in breast cancer cells.⁹

1.3. Correlation of Properties

As shown before, flavonoids are a group of very interesting molecules available in plants but in order to have a concentrated or even pure compound, a whole process needs to be designed and optimized. Therefore, reliable models are needed to be able to simulate the process *in silico* and optimize it. In most cases, these models are fitted to experimental data and then interpolated or extrapolated (which brings a higher uncertainty) to calculate how the different compounds will behave. Models can be correlative (fitted to data) or predictive (based on molecular properties), being the first ones more reliable than the second ones in most cases. An overview of correlative models for properties is given below.

1.3.1. Properties of a mixture of compounds in a (or more) solvent/s

In most steps of a process, the stream to be treated will consist of the target compound (or compounds) to concentrate or purify and a mixture of other compounds in a solvent. A way of characterizing the behaviour of a compound in a mixture is by its activity coefficient. Activity coefficients²³ are the ratio between the activity of a compound and its molar fraction which means that they account for the nonideal behaviour of the compound. Activity coefficient is a concept that applies mainly to compounds in a liquid phase and it is analogous to the fugacity coefficient in gas phases.

Several models²³ have been developed for the modelling of experimentally calculated activity coefficients like Margules, van Laar, NRTL (Non-Random Two Liquids) and UNIQUAC (UNIversal QUAsi-Chemical model), being Margules the simplest and

UNIQUAC the most complex one²⁴. All of them fit the experimental data (mostly Vapour-Liquid Equilibria data) by the regression of interaction parameters and, in some cases, size related parameters. Those parameters can be used later for the calculation of activity coefficients in different conditions (mainly molar fraction and sometimes temperature) than the ones they were regressed from.

These models have rarely been applied to flavonoid mixtures (with some exceptions²⁵). The main reason for that is the difficulty of acquiring reliable data. The most common method to calculate activity coefficients is by measure the concentrations of each compound during phase equilibria (being VLE the one that is used the most).

The low vapour pressure of the flavonoids and their instability with moderate to high temperatures^{26, 27} force liquid-liquid and solid-liquid to be the equilibria from which the calculation of activity coefficients is done.

1.3.2. Partition coefficients between phases.

As previously stated, vapour-liquid equilibria is likely not to be a key part in any process step. Therefore, only liquid-liquid and solid-liquid will be considered.

In the case of liquid-liquid equilibria, the activity of a compound in a phase is equal to the activity of the same compound in the other phase. Therefore, an accurate prediction of the partition coefficient can be made with a working activity coefficient model.

For solid-liquid equilibria, two very important properties have to be taken into account:

1.3.2.1. Solubility: Solubility of a solute in a solvent is defined as the maximum concentration that the solute can reach in that solvent. At said concentration, an equilibrium is reached between the solute in solution and a newly formed pure solid phase of the solute.

A common way to model solubility in literature is to use the relation between the activity of the solute in the solvent and its melting properties²³. This relation can be used to obtain experimental activity coefficients by fitting such relation to the experiments.

1.3.2.2. Isotherms: Isotherm is the relation between the concentration in the liquid and the concentration in the solid phase (a resin) when adsorption occurs on the surface of that solid phase at a constant temperature. The most widely known isotherm equation is the one developed by Langmuir²⁸, with which two very important parameters can be regressed, namely the affinity constant and the maximum loading of the resin. Such model has been applied to flavonoids thoroughly in literature²⁹⁻³³. Other models, such as Freundlich are equally used but lack of the mechanistic approach of the Langmuir model and therefore, the fitted parameters do not have physical meaning.

1.4. Prediction of properties

In most cases, flavonoids are present in small concentrations and interacting with other components from the plant matrix. This interactions are different from flavonoid to flavonoid and the process design for one of them might not be suitable for a different polyphenol. As well, because of their small concentration, the required experiments to fit the correlative models might be prohibitive due to costs of material or time. Therefore, it is as well very important the development of models to predict properties without the need of previous experiments to act as a toolbox for the design of future processes

An overview of the available predictive models in literature is given below.

1.4.1. Properties of a pure compound

Properties of a pure compound refer to those physicochemical or thermodynamic characteristics of a substance without taking into account its interactions with molecules of other compounds. Most prediction models are based on group contribution (GC), which is a method that divides molecular structures into small groups (-OH, -CH₃, etc.) and correlates the different properties to the number and type of groups that the molecules are made of. The properties of interest can be divided into physicochemical properties (critical temperature, pressure and volume (T_c , P_c , V_c), boiling and melting temperature (T_b , T_m)) or thermodynamic properties (Gibbs energy and enthalpy of formation (ΔG_f , ΔH_f), enthalpy of fusion and vaporization (ΔH^{fus} , ΔH^{vap}), heat capacity (C_p) and Hansen solubility parameters (δ_d , δ_p , δ_{hb})). The most renowned GC model is the one of Joback et al.³⁴. Other models tried to improve the prediction by adding neighbouring effects,³⁵ conformational effects³⁶ or position effects.³⁷ An overview of some of the models on the basis on the properties they estimate can be seen in Table 1.

Table 1: GC models for the prediction of pure component properties

Property to estimate	Model
T_c , P_c , V_c , T_b , ΔG_f , ΔH_f , ΔH^{vap}	Joback et al. ³⁴ , Marrero and Gani ³⁵
T_m	Joback et al. ³⁴ , Wang et al. ³⁷ , Jain et al. ^{36, 38} , Marrero and Gani ³⁵
ΔH^{fus}	Joback et al. ³⁴ , Jain et al. ^{36, 38} , Marrero and Gani ³⁵
ΔS^{fus}	Jain et al. ^{36, 38}
C_p ,	Joback et al. ³⁴
δ_d , δ_p , δ_{hb} ,	Stefanis et al. ³⁹

1.4.2. Properties of a mixture of compounds in a (or more) solvent/s

As explained in Section 1.3.1, the main property to be considered here would be the activity coefficient. The most used predictive activity coefficient model is the UNIFAC model⁴⁰ (UNIQuac Functional Activity Coefficient model). This model is based on the UNIQUAC model with the main difference of instead of considering the system as a mixture of compounds, the UNIFAC model considers it a mixture of groups bonded into compounds. This difference allows for the regression of the interaction parameters between groups from a data set and the application of the same parameters to another dataset with different compounds (as long as the groups are the same). Using the vast Dortmund database, the group of Prof. Jürgen Gmehling⁴⁰⁻⁴³ was able to regress a big set of interaction parameters for a long list of groups.

Another predictive model that deserves mentioning is COSMO-RS (Conductor-like Screening Model for Real Solvents) which uses quantum chemistry for the calculation of the chemical potential of a compound in a mixture. While the performance of both models is similar⁴⁴, COSMO-RS seems to have a wider applicability than UNIFAC at the cost of a much higher complexity.

1.4.3. Partition coefficients between phases.

The prediction of partition coefficients between phases relies mainly in having a working activity coefficient model. Chemical potentials of substances must be equal in all phases in equilibrium and that can be used to predict the behaviour of substances.

1.4.3.1. Solubility: As previously stated in Section 1.3.2.1, there is a relation between the melting properties of a compound and its activity in a solvent when reaching maximum concentration²³. Previously, this relation was described as a way to

calculate experimental activity coefficients, but if a working activity coefficient model is provided, the relation can be used for the prediction of solubility values.

1.4.3.2. logP: A very useful property is the partition coefficient of a compound between two liquid phases, namely 1-octanol and water. The logarithm of this partition coefficient is called logP and is used as a measure for the hydrophobicity of said compound. Since logP is a partition coefficient between two liquid phases, it can be predicted by using an activity coefficient method, however there are simpler GC methods⁴⁵ that can give an good estimate saving computational time.

1.4.3.3. logD: LogP is defined as the partition coefficient of the non-charged species but sometimes, the compound (or a relevant fraction of the total amount of molecules) will be charged and therefore, logP will not give an accurate description of the behaviour . For those cases, logD is the property to use since it gives the partition coefficient as a function of the pH of the solution. LogD GC models tend to be more computationally intensive since it needs the prediction of the pKa values of the compound as an input but nonetheless there are several models in literature tackling this problem and giving good estimates for logD^{46, 47} .

1.5. Application of models

An overview of which models should be used in each step of the process is given bellow.

1.5.1. Extraction from the raw material

Flavonoids are mostly located in the vacuoles⁴⁸ of the plant cells. As most extraction processes, the solvent and temperature of use will determine the profile of compounds that later on, have to be dealt with in the purification steps. To maximize yield of this step, logP is a key property to consider. The yield of the extraction will be maximized

when using a solvent (or mixture of solvents) that have a similar logP to the compound to be extracted. Another aspect to take into account is the logP of the contaminants since a lower yield in the extraction but a higher selectivity might help downstream for the final purification.

1.5.2. Concentration and purification

After the compound has been extracted from the plant matrix, it is advisable to concentrate it before purifying it. The main difference between the two lies in the target to achieve in the unit operation. While the main focus of a concentration step is to increase the concentration of the compound while keeping a high yield, in the purification step, selectivity is the target.

As a concentration step, liquid-liquid extraction offers a wide range of benefits and for its design, several of the aforementioned models are needed. The key property to look at is the partition coefficient between both liquid phases. The prediction of the partition coefficient is very important for the final solvent selection since a low affinity for the extraction solvent is just as unwanted as a very high affinity (due to the need of a back extraction).

As a purification step, it is recommended to opt for solid liquid extraction (chromatography). Different modes of chromatography can be used (IEC, HIC, MMC, etc.) depending on different factors such as pK_a or logP values. As well, a great variety of resins can be used to maximize selectivity. For an a priori selection of resins and conditions there are two main properties that are needed: isotherms and diffusion coefficients (which can be predicted using models from literature⁴⁹⁻⁵¹). Both act as an input for the steric mass action model that describes the adsorption behaviour of components in the column during the process. It is crucial that the

isotherm parameters are validated for the conditions of operation (pH, temperature, solvent (s), salt concentrations, etc.).

1.5.3. Formulation

The compound/s of interest are in solution with an acceptable amount of impurities and therefore, crystallization is the last necessary step. In this last stage of the process the main property that will play a role is solubility. Two main conditions are used for the design of the step, namely temperature and use of anti-solvent. Choosing either of them relies heavily on the value for solubility at those conditions.

This proposed process and the critical properties needed in each step is summarized in Figure 5.

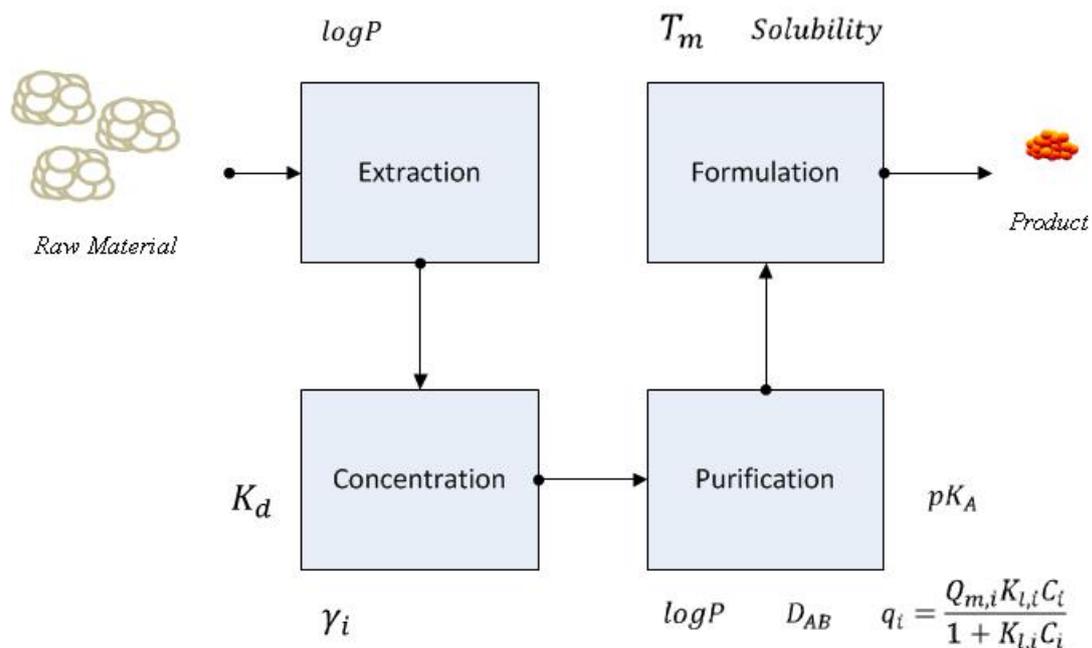


Figure 5. Sketch of a process and critical properties in each step.

1.6. Outline of the Thesis

The state of the art before the beginning of this work was gathered together with a brief guide on the application of the explained models to different steps of the process (Chapter 1). It is worth to mention that in literature almost no models had

been applied to flavonoid-like molecules, making the application of models in this area uncharted territory.

The main focus of the work was on green tea catechins and soy isoflavones. From the beginning, their molecular similarities made clear that group contribution models was the obvious way to reduce complexity by having a universal validated model that could be applied to all flavonoids of interest. Experimental solubility values were determined at different temperatures and different percentages of ethanol in water (Chapter 2). Different predictive models were compared and validated but the activity coefficient model Mod-UNIFAC showed a bad prediction of the experimental solubility values.

Being activity coefficients a key property in the design and optimization of a process, it was very important to have a validated and predictive model for activity coefficients. Therefore an extensive bibliographic work was done and solubility values of flavonoids were gathered in a database (410 datapoints). That database was used for the regression of new interaction parameters for the Mod-UNIFAC model with a new definition of groups more applied to polyphenols. Those parameters were regressed and validated in what is called Modified PolyPhenols-UNIFAC or MPP-UNIFAC (Chapter 3).

With a good understanding on the thermodynamics in solution, the next step of the work needed to be about the equilibrium with a resin. A set of food-grade resins was tested for their suitability to separate catechins from caffeine using green tea as a starting material (Chapter 4). Using regressed parameters from preliminary experiments, a model-based Design of Experiments (DoE) was developed to determine the conditions at which experiments should be performed to minimize the error and guarantee the validity of the model at different ratios of the compounds of

interest. As well, since multicomponent Langmuir isotherm model was used, the regressed parameters have a physical meaning and can be used as criteria for resin selection.

While green tea is a mainly aqueous solution, okara, the solid residue from the soy milk production needs to be pretreated before adsorption. This pretreatment requires the extraction of the flavonoids from the fibrous matrix and can be done by the use of different solvents. As seen throughout this work, several properties of the flavonoids do not behave ideally for mixtures of ethanol and water: solubility has a maximum for 0.5-0.6 ethanol molar fraction (Chapter 2 and a great number of cases in Chapter 3), isotherms are flatter for such ethanol fractions than for pure water or pure ethanol (Chapter 5). Therefore, it was decided to perform the extraction of okara with water and use that extract as the inlet for the adsorption. Adsorption isotherms were measured (Chapter 5) showing an almost linear behaviour, typical of small concentrations. In some cases it was observed a non-linearity that was further studied and explained as adsorption of isoflavones on top of already adsorbed proteins. This phenomenon occurred only with not very hydrophobic isoflavones on resins that were very hydrophobic. Moreover, a linear relation was discovered between the logarithm of the slope of the isotherm and the $\log D$ value of the compound, giving a very clear outlook into possible predictability of the isotherms for isoflavones.

To apply this knowledge to higher concentrations, where the isotherms become non-linear, a more thermodynamically sound method needed to be applied. Therefore a lattice model was developed (Chapter 6) in which the surface of the resin was divided into sites and the molar fraction in the adsorbed phase was defined as fraction of occupied sites. Based on that definition and a few more thermodynamic relations, experimental activity coefficients in the adsorbed phases could be determined.

Subsequently, a simple activity coefficient model was applied for the fit and one interaction parameter was regressed per compound. In this case, as it happened in the previous chapter, a linear relation was discovered between the logarithm of equilibrium constant and the logD value. Moreover, the regressed interaction parameters showed a linear correlation with the logD values of the compounds in equilibrium.

This way we have shown the dependence of isotherm equilibrium parameters on the logD value, setting the ground for further study on this relation and the development of fully predictive adsorption models for polyphenol-like molecules onto food-grade resins.

Nomenclature

Latin symbols

C_p	Specific Heat Capacity at Constant Pressure ($J K^{-1} mol^{-1}$)
HIC	Hydrophobic Interaction Chromatography
IEC	Ion Exchange Chromatography
MMC	Mixed Mode Chromatography
P	Pressure (bar)
T	Temperature (K)
V	Molar volume ($cm^3 mol^{-1}$)

Greek symbols

ΔG	Gibbs Energy ($J mol^{-1}$)
ΔH	Enthalpy ($J mol^{-1}$)
ΔS	Entropy ($J mol^{-1}$)
η_L	Liquid Viscosity
$\delta_d, \delta_p, \delta_{hb}, \delta_t$	Hansen Solubility Parameters (dispersion, polar, hydrogen bonding and total respectively) ($MPa^{1/2}$)

Superscripts

fus	Fusion
vap	Vaporization

Subscripts

b	Boiling
c	Critical
f	Formation
m	Melting

References

- (1) Barnard, N. D., Trends in food availability, 1909–2007. *The American Journal of Clinical Nutrition* 2010.
- (2) Levis, J. W.; Barlaz, M. A.; Themelis, N. J.; Ulloa, P., Assessment of the state of food waste treatment in the United States and Canada. *Waste Manag.* 2010, 30, 8–9, 1486.
- (3) Rosales, E.; Rodríguez Couto, S.; Sanromán, A., New uses of food waste: application to laccase production by *Trametes hirsuta*. *Biotechnol. Lett.* 2002, 24, 9, 701.
- (4) Kim, S.-H.; Han, S.-K.; Shin, H.-S., Feasibility of biohydrogen production by anaerobic co-digestion of food waste and sewage sludge. *Int. J. Hydrogen Energy* 2004, 29, 15, 1607.
- (5) Seader, J. D.; Henley, E. J., *Separation process principles*. Wiley: 1998.
- (6) Ververidis, F.; Trantas, E.; Douglas, C.; Vollmer, G.; Kretzschmar, G.; Panopoulos, N., Biotechnology of flavonoids and other phenylpropanoid-derived natural products. Part I: Chemical diversity, impacts on plant biology and human health. *Biotechnol. J.* 2007, 2, 10, 1214.
- (7) Agarwal, R.; Katiyar, S. K.; Zaidi, S. I. A.; Mukhtar, H., Inhibition of skin tumor promoter-caused induction of epidermal ornithine decarboxylase in SENCAR mice by polyphenolic fraction isolated from green tea and its individual epicatechin derivatives. *Cancer Res.* 1992, 52, 13, 3582.
- (8) Johnson, R.; Bryant, S.; Huntley, A. L., Green tea and green tea catechin extracts: An overview of the clinical evidence. *Maturitas* 2012, 73, 4, 280.
- (9) Messina, M. J.; Persky, V.; Setchell, K. D. R.; Barnes, S., Soy intake and cancer risk: A review of the in vitro and in vivo data. *Nutr. Cancer* 1994, 21, 2, 113.
- (10) Cabrera, C.; Artacho, R.; Giménez, R., Beneficial Effects of Green Tea—A Review. *J. Am. Coll. Nutr.* 2006, 25, 2, 79.
- (11) Hertog, M. G. L.; Feskens, E. J. M.; Hollman, P. C. H.; Katan, M. B.; Kromhout, D., Dietary antioxidant flavonoids and risk of coronary heart disease: The Zutphen Elderly Study. *Lancet* 1993, 342, 8878, 1007.
- (12) Alming, M.; Aura, A. M.; Bohn, T.; Dufour, C.; El, S. N.; Gomes, A.; Karakaya, S.; Martínez-Cuesta, M. C.; McDougall, G. J.; Requena, T.; Santos, C. N., In vitro models for studying secondary plant metabolite digestion and bioaccessibility. *Comprehensive Reviews in Food Science and Food Safety* 2014, 13, 4, 413.
- (13) Guo, Q.; Zhao, B.; Li, M.; Shen, S.; Wenjuan, X., Studies on protective mechanisms of four components of green tea polyphenols against lipid peroxidation in synaptosomes. *Biochimica et Biophysica Acta - Lipids and Lipid Metabolism* 1996, 1304, 3, 210.
- (14) Setchell, K. D. R.; Cassidy, A., Dietary Isoflavones: Biological Effects and Relevance to Human Health. *The Journal of Nutrition* 1999, 129, 3, 758.
- (15) Kozłowska, A.; Szostak-Wegierek, D., Flavonoids--food sources and health benefits. *Roczniki Państwowego Zakładu Higieny* 2014, 65, 2, 79.
- (16) McCullough, M. L.; Peterson, J. J.; Patel, R.; Jacques, P. F.; Shah, R.; Dwyer, J. T., Flavonoid intake and cardiovascular disease mortality in a prospective cohort of US adults. *The American Journal of Clinical Nutrition* 2012.

- (17) Cassidy, A.; O'Reilly, É. J.; Kay, C.; Sampson, L.; Franz, M.; Forman, J.; Curhan, G.; Rimm, E. B., Habitual intake of flavonoid subclasses and incident hypertension in adults. *The American Journal of Clinical Nutrition* 2011, 93, 2, 338.
- (18) Shen, Y.; Zhang, J.; Sheng, R.; Dong, X.; He, Q.; Yang, B.; Hu, Y., Synthesis and biological evaluation of novel flavonoid derivatives as dual binding acetylcholinesterase inhibitors. *J. Enzyme Inhib. Med. Chem.* 2009, 24, 2, 372.
- (19) Sheng, R.; Lin, X.; Zhang, J.; Chol, K. S.; Huang, W.; Yang, B.; He, Q.; Hu, Y., Design, synthesis and evaluation of flavonoid derivatives as potent AChE inhibitors. *Bioorg. Med. Chem.* 2009, 17, 18, 6692.
- (20) Khan, M. T. H.; Orhan, I.; Şenol, F. S.; Kartal, M.; Şener, B.; Dvorská, M.; Šmejkal, K.; Šlapetová, T., Cholinesterase inhibitory activities of some flavonoid derivatives and chosen xanthone and their molecular docking studies. *Chem-Biol. Interact.* 2009, 181, 3, 383.
- (21) Brickman, A. M.; Khan, U. A.; Provenzano, F. A.; Yeung, L.-K.; Suzuki, W.; Schroeter, H.; Wall, M.; Sloan, R. P.; Small, S. A., Enhancing dentate gyrus function with dietary flavanols improves cognition in older adults. *Nat. Neurosci.* 2014, advance online publication.
- (22) Le Marchand, L.; Murphy, S. P.; Hankin, J. H.; Wilkens, L. R.; Kolonel, L. N., Intake of Flavonoids and Lung Cancer. *J. Natl. Cancer Inst.* 2000, 92, 2, 154.
- (23) Prausnitz, J. M., *Molecular Thermodynamics of Fluid-Phase Equilibria*. Prentice-Hall: Englewood Cliffs, New Jersey, 1969.
- (24) Poling, B.; Prausnitz, J.; Connell, J. O., *The Properties of Gases and Liquids*. McGraw-Hill Education: 2000.
- (25) Queimada, A. n. J.; Mota, F. t. L.; Pinho, S. P.; Macedo, E. n. A., Solubilities of Biologically Active Phenolic Compounds: Measurements and Modeling. *The Journal of Physical Chemistry B* 2009, 113, 11, 3469.
- (26) Li, N.; Taylor, L. S.; Mauer, L. J., Degradation Kinetics of Catechins in Green Tea Powder: Effects of Temperature and Relative Humidity. *J. Agric. Food Chem.* 2011, 59, 11, 6082.
- (27) Su, Y. L.; Leung, L. K.; Huang, Y.; Chen, Z. Y., Stability of tea theaflavins and catechins. *Food Chem.* 2003, 83, 2, 189.
- (28) Langmuir, I., The constitution and fundamental properties of solids and liquids. Part I. Solids. *J. Am. Chem. Soc.* 1916, 38, 11, 2221.
- (29) Dong, Z. B.; Liang, Y. R.; Fan, F. Y.; Ye, J. H.; Zheng, X. Q.; Lu, J. L., Adsorption behavior of the catechins and caffeine onto polyvinylpyrrolidone. *J. Agric. Food Chem.* 2011, 59, 8, 4238.
- (30) Tang, R.-C.; Tang, H.; Yang, C., Adsorption Isotherms and Mordant Dyeing Properties of Tea Polyphenols on Wool, Silk, and Nylon. *Ind. Eng. Chem. Res.* 2010, 49, 19, 8894.
- (31) Lu, Q.; Sorial, G. A., A comparative study of multicomponent adsorption of phenolic compounds on GAC and ACFs. *J. Hazard. Mater.* 2009, 167, 1-3, 89.
- (32) Bretag, J.; Kammerer, D. R.; Jensen, U.; Carle, R., Adsorption of rutin onto a food-grade styrene-divinylbenzene copolymer in a model system. *Food Chem.* 2009, 114, 1, 151.
- (33) van Vliet, B. M.; Weber Jr, W. J.; Hozumi, H., Modeling and prediction of specific compound adsorption by activated carbon and synthetic adsorbents. *Water Res.* 1980, 14, 12, 1719.

- (34) Joback, K. G.; Reid, R. C., Estimation of Pure-Component Properties from Group-Contributions. *Chem. Eng. Commun.* 1987, 57, 1-6, 233.
- (35) Marrero, J.; Gani, R., Group-contribution based estimation of pure component properties. *Fluid Phase Equilib.* 2001, 183-184, 183.
- (36) Jain, A.; Yang, G.; Yalkowsky, S. H., Estimation of Melting Points of Organic Compounds. *Ind. Eng. Chem. Res.* 2004, 43, 23, 7618.
- (37) Wang, Q.; Ma, P.; Neng, S., Position Group Contribution Method for Estimation of Melting Point of Organic Compounds. *Chin. J. Chem. Eng.* 2009, 17, 3, 468.
- (38) Jain, A.; Yalkowsky, S. H., Estimation of melting points of organic compounds-II. *J. Pharm. Sci.* 2006, 95, 12, 2562.
- (39) Stefanis, E.; Panayiotou, C., Prediction of Hansen Solubility Parameters with a New Group-Contribution Method. *Int. J. Thermophys.* 2008, 29, 2, 568.
- (40) Fredenslund, A.; Gmehling, J.; Michelsen, M. L.; Rasmussen, P.; Prausnitz, J. M., Computerized design of multicomponent distillation columns using the UNIFAC group contribution method for calculation of activity coefficients. *Industrial and Engineering Chemistry Process Design and Development* 1977, 16, 4, 450.
- (41) Gmehling, J.; Li, J.; Schiller, M., A modified UNIFAC model. 2. Present parameter matrix and results for different thermodynamic properties. *Ind. Eng. Chem. Res.* 1993, 32, 1, 178.
- (42) Gmehling, J.; Lohmann, J.; Jakob, A.; Li, J.; Joh, R., A Modified UNIFAC (Dortmund) Model. 3. Revision and Extension. *Ind. Eng. Chem. Res.* 1998, 37, 12, 4876.
- (43) Gmehling, J.; Wittig, R.; Lohmann, J.; Joh, R., A modified UNIFAC (Dortmund) model. 4. Revision and extension. *Industrial and Engineering Chemistry Research* 2002, 41, 6, 1678.
- (44) Mokrushina, L.; Buggert, M.; Smirnova, I.; Arlt, W.; Schomäcker, R., COSMO-RS and UNIFAC in prediction of micelle/water partition coefficients. *Industrial and Engineering Chemistry Research* 2007, 46, 20, 6501.
- (45) Klopman, G.; Li, J.-Y.; Wang, S.; Dimayuga, M., Computer Automated log P Calculations Based on an Extended Group Contribution Approach. *J. Chem. Inf. Comput. Sci.* 1994, 34, 4, 752.
- (46) Xing, L.; Glen, R. C., Novel Methods for the Prediction of logP, pKa, and logD. *J. Chem. Inf. Comput. Sci.* 2002, 42, 4, 796.
- (47) Viswanadhan, V. N.; Ghose, A. K.; Revankar, G. R.; Robins, R. K., Atomic physicochemical parameters for three dimensional structure directed quantitative structure-activity relationships. 4. Additional parameters for hydrophobic and dispersive interactions and their application for an automated superposition of certain naturally occurring nucleoside antibiotics. *J. Chem. Inf. Comput. Sci.* 1989, 29, 3, 163.
- (48) Hemingway, R. W.; Laks, P. E.; Branham, S. J., *Plant Polyphenols: Synthesis, Properties, Significance*. Springer US: 1992.
- (49) Wilke, C. R.; Chang, P., Correlation of diffusion coefficients in dilute solutions. *AIChE J.* 1955, 1, 2, 264.
- (50) Young, M. E.; Carroad, P. A.; Bell, R. L., Estimation of diffusion coefficients of proteins. *Biotechnol. Bioeng.* 1980, 22, 5, 947.
- (51) Sofer, G. K.; Hagel, L., *Handbook of Process Chromatography: A Guide to Optimization, Scale Up, and Validation*. Academic Press: 1997.

Chapter 2: Model comparison for the prediction of Green Tea Catechins

Solubility in ethanol/water mixtures

Abstract:

This chapter evaluates some of the most used solubility models and their applicability to antioxidants, namely, the family of catechins. From the list of possible catechins, the main four were selected for this study: epicatechin, epigallocatechin, epicatechin gallate, and epigallocatechin gallate. Both the enthalpy of fusion and melting temperature were measured experimentally for these four molecules and compared with the values obtained using predictive methods. Solubility measurements for mixtures of water and ethanol were performed with high throughput experimentation at the temperatures of 293 and 303 K. The experiments show a strong nonideal behavior in both pure and mixed solvents. A comparison of different models was performed, identifying the advantages and disadvantages of each model and their performance in this particular case. Finally, the NRTL-SAC model was found to be the most accurate model for predicting solubility, with an RMSE of 0.550.

Published as: Model comparison for the prediction of the solubility of green tea catechins in ethanol/water mixtures Méndez Sevillano, D.; Van Der Wielen, L. A. M.; Trifunovic, O.; Ottens, M., *Industrial and Engineering Chemistry Research* **2013**, 52, (17), 6039-6048.

2.1. Introduction

The term nutraceutical has been used in the past decade to define a group of drug-like molecules present in our diet with health-related properties. These molecules are usually present in small amounts and immersed in complex multicomponent mixtures. Well-known components such as peptides, antioxidants, vitamins, etc. are examples of nutraceuticals.

Of the diverse set of nutraceuticals, antioxidants have received increasing attention in the past few years because of their claimed health properties. Catechins are one example of antioxidants that can be found in food products such as cocoa and tea (especially green tea). Green tea has the highest concentration of these antioxidants, which is around 30-42% of the dry weight (dw). The family of green tea catechins consists of 21 different molecules but the four epimers in Figure 1 account for 93% (dw) of them with epigallocatechin gallate (EGCg) having the highest concentration, 46% (dw).¹

The special interest in these molecules is because of the wide range of health benefits they could provide². The numerous hydroxyl groups in these molecules can scavenge free radicals, which prevents damages on a cellular level. Because of this scavenging effect, catechins have been identified with benefits that range from antiarteriosclerotic³ to antitumorigenic.⁴

Because antioxidants are such potentially valuable molecules, physicochemical and thermodynamic properties are needed for the design of concentration and purification processes.^{5, 6} Solubility is one of the most important properties for such design. Especially for molecules that contain hydrophobic and hydrophilic groups such as antioxidants, data and proper models are scarce. In these cases, solubility in mixed solvents can have a major impact on design.

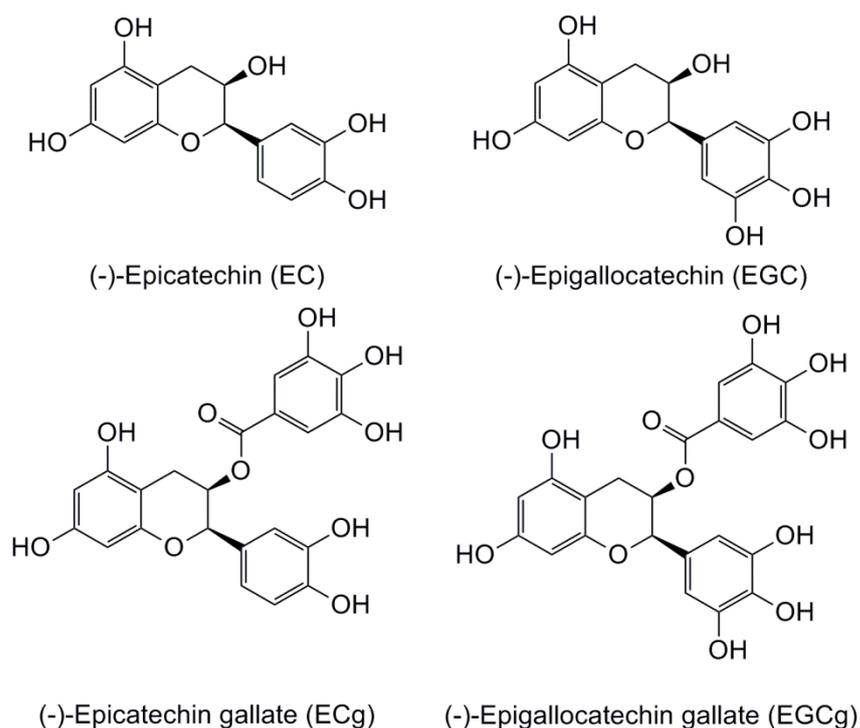


Figure 1. Main compounds in the family of catechins.

The aim of this study was to investigate the behavior of a class of nutraceuticals, namely, the catechins, in pure and mixed solvents and to test the performance of relevant models. These components originate from food-industry products and are meant for use in the food industry, so their possible solvents are limited to food-grade solvents: water and ethanol.

2.2.Theoretical Background

2.2.1. State of the art for Solubility Modelling

2.2.1.1. Solubility Calculation Using Melting Properties. At multiphase equilibrium, the fugacity of each component is the same in all phases. Using well-known basic thermodynamic equations and assuming a unique pure solid phase, one can calculate activity coefficients using Equation 1.⁷

$$x_i \gamma_i^l = \exp \left\{ \frac{\Delta H_i^{sl}}{RT_{m_i}} \left(\frac{T - T_{m_i}}{T} \right) + \frac{\Delta C_{p_i}^{sl}}{R} \left[\ln \frac{T}{T_{m_i}} - \left(\frac{T - T_{m_i}}{T} \right) \right] \right\} \quad [1]$$

This equation has been further simplified throughout literature by neglecting the contribution of $\Delta C_{p,i}^{sl}$, but doing so can lead to errors for compounds with high melting temperatures. Another simplification that has proven to decrease these errors for substituted aromatic compounds⁸, is to assume $\Delta C_{p,i}^{sl} = \Delta H_i^{sl}/T_{m,i}$ which leads to Equation 2.

$$x_i \gamma_i^l = \exp \left[\frac{\Delta H_i^{sl}}{RT_{m,i}} \left(\ln \frac{T}{T_{m,i}} \right) \right] \quad [2]$$

However, this simplification proved to be less accurate for non-substituted aromatic molecules (i.e. benzene, naphthalene, etc.)

Activity coefficients are usually modelled by G^E models like UNIFAC (universal functional activity coefficient)⁹⁻¹³. This model has proven to be successful for small biomolecules, but it has rarely been applied in cosolvency predictions. The UNIFAC model is considered to be predictive, although not entirely, because it does not require any experimentation for the calculation of properties but uses parameters that have already been regressed from a vast database of equilibria. The non-random two-liquid-segment activity coefficient (NRTL-SAC)^{14, 15} model is another model that has gained importance in the field. This model does require experimental data for the components although it can be used fairly accurately to predict properties based on the activity coefficients given by the model.

In addition, other G^E models and other activity coefficient models are of importance in the field such as the conductor-like screening model for realistic solvents (COSMO-RS)¹⁶ and the conductor-like screening model–segment activity coefficient (COSMO-SAC).¹⁷ These models show good performance when the quantum properties can be calculated, but they require these properties for proper performance. Another method of calculating these activity coefficients is by means of an equation of state (EoS) such as perturbed-chain polar statistical associating fluid theory (PCP-

SAFT),¹⁸ but again, quantum and/or electrostatic properties are needed for proper performance.

2.2.1.2. Solubility Calculations from Excess Solubility Models. Excess solubility of a component in a mixture can be defined as the difference between the true value of solubility and the value resulting by presuming that solubility is an ideal property. It can be calculated with Equation 3.

$$\ln x_{i,m}^E = \ln x_{i,m} - \sum_j x'_j \ln x_{i,j} \quad [3]$$

Assuming that the equilibrium occurs at infinite-dilution, based on the typically low solubility values of drug-like molecules, Equation 3 can be written as Equation 4.

$$\ln x_{i,m}^E = \ln \gamma_{i,m}^\infty - \sum_j x'_j \ln \gamma_{i,j}^\infty \quad [4]$$

An activity coefficient model can be used for the calculation of the excess solubility. This type of approach requires the solubilities in pure solvents if a value of the true solubility is needed.

Based on the described simplification, several authors have developed excess solubility models including Ellegaard et al.¹⁹ and Gude et al.²⁰ These two models express the interaction between solute and solvent(s) by means of interaction parameters in different manners, and the model of Ellegaard et al. includes an activity coefficient contribution that is not present in Gude et al.

2.2.1.3. Solubility Calculation from Correlations. Several correlations have been developed in the literature for cosolvency equilibria. The most famous is the log-linear model.²¹ Throughout the literature, several authors have added different polynomial terms to this model for a better fit.^{22, 23} One of these models worth mentioning is that developed by Jouyban et al.,²⁴ who used more than 100 different equilibria of drug-like compounds in binary or ternary solvents at different temperatures to fit their model. (Appendix section)..

2.2.2. Models Used in This Work.

Despite the vast number of models available, only four of them were chosen for comparison in this work. This selection was mainly based on the importance, simplicity (fewer parameters to regress) and convenience for programming of the models. These criteria align with the aim of the work, which was to provide a useful approach to a broader chemical engineering audience.

Several group-contribution (GC) models were used for the prediction of properties in this study. For most of the models, the property was calculated by the addition of the different contributions of the groups plus a zero term. The divisions of the groups were similar in all of the models, but some presented a second or even a third order of complexity by accounting for proximity or conformation effects. UNIFAC,⁹⁻¹³ on the other hand, calculates activity coefficients with a complex set of equations using volume, surface and interaction group parameters (Appendix section). Melting properties,²⁵⁻²⁸ Hansen solubility parameters²⁹ and activity coefficients (Modified version of the UNIFAC model, Mod-UNIFAC)⁹⁻¹³ have been calculated with GC models in this study.

Non-GC models were used as well. They describe the property of a component by parameters that account for either the entire molecule or segments of it as in the NRTL-SAC^{14, 15} which is widely used for solubility predictions in both pure and mixed solvents.³⁰⁻³²

A short description of each model is provided next.

Wang et al.²⁸ developed a GC model that takes neighbouring molecules into account for the prediction of the temperature of fusion. The model considers neighbouring molecules for the selection of the groups and some conformation effects

(i.e., para, meta, and ortho conformations). A drawback of this specificity is the lack of groups for uncommon conformations.

Jain et al.²⁵ developed a GC model that takes into account the chemical composition and three-dimensional configuration of each group for the prediction of enthalpies of melting. The entropy of melting is calculated from the chemical and conformational characteristics of the molecule (e.g., number of symmetry axes). The melting temperature is directly calculated from these two values. For the prediction of the melting enthalpy, every group is considered for the type of bond that is formed with the nearby groups (i.e., sp³ or sp²). An extension of the method was later published²⁶ with an extension of the number of groups and further contributions (apart from bonds formed) such as aromaticity.

Marrero and Gani²⁷ developed a GC model for the prediction of different thermodynamic properties such as boiling, critical, and melting properties. The method has three levels of definition: The first accounts for the groups that form the molecule. The second and third are for effects due to neighbouring molecules on the small (i.e., isobutyl group) and large (i.e., fused aromatic rings) scales, respectively.

The NRTL-SAC^{14, 15} model is an activity coefficient model based on the NRTL model. In the original model, equilibrium is described by interaction parameters. In the NRTL-SAC model, interactions are divided into four types or segments (hydrophilic, polar repulsive, polar attractive, and hydrophobic), and the interaction parameters are already built in the model. The only parameter that is needed is the contribution of each of the forces in each molecule. As these segments are conceptual and not physicochemical groups, the contribution of each of the segments cannot be extrapolated to other molecules. Equations can be found in the Appendix, section .

Mod-UNIFAC¹⁰⁻¹³ is the most used GC activity coefficient model. This model is a modification of the original UNIFAC model, which was based on the universal quasichemical (UNIQUAC) model. Three contributions are taken into account: size, surface, and interaction. Groups are sorted into families with similar characteristics, and whereas surface and volume are characteristics of a group, interactions are between families, and the values are the same for each group that is member of the family. Equations can be found in the Appendix, section .

Stefanis et al.²⁹ developed a model for the prediction of Hansen parameters. This model, similar to that of Jain et al.²⁵, presents two levels of characterization: definition and effects due to neighbouring groups.

Jouyban²⁴ developed a purely empirical model for the prediction of excess solubilities for solubility in two or three solvents. It requires the input of the Hansen parameters of the solute and solvents and shows good performance for a moderate range of temperatures, even though it is optimized for 25 °C. Equations can be found in the Appendix, section .

Gude et al.²⁰ used an activity coefficient model for the excess solubility in mixed solvents. This model requires the input of only two parameters, the solvent–solvent interaction parameter and the solvent–solvent–solute parameter. Equations can be found in the Appendix, section.

2.3. Materials and Methods

2.3.1. Materials.

Epicatechin (EC), Epigallocatechin (EGC), Epicatechin gallate (ECg) and Epigallocatechin gallate (EGCg) standards were purchased from Nacalai USA Inc. (San Diego CA). Caffeine was purchased from Merck. Ethanol HPLC grade and

milli-Q water were used for the mixed solvents. Acetonitrile from Sigma was used for the HPLC mobile phase and the green tea stabilizer solution (20% Acetonitrile in water).

2.3.2. Solubility Experiments.

High-Throughput (HT) technology was used for the solubility experiments. A trend toward the implementation of HT techniques is being established for the measurements of the properties of drugs³³ because of the advantages of such techniques. Key advantages include small well volumes, easy handling and easy automation. A broader set of experiments is possible with reduced use of sample and time.

The experiments were carried out in 800 µl flat-bottom 96 well-plates (UNIPLATE Whatman INC, NJ, USA). The desired amount of compound was accurately placed in the vials by pipetting a volume of a stock solution of known concentration in 50% ethanol (v/v) in water. The plate was left to dry in an oven at 30 °C until the solvent had fully evaporated.

A volume of solvent (or mixture of solvents) was then placed in the vial, and the well-plate was left stirring for 24 h. The temperature and stirring speed (1000 rpm) were controlled (Thermomixer Comfort, Eppendorf, Hamburg, Germany) and evaporation was prevented by sealing the plate with an aluminium seal (adhesive foil for microplates. VWR Scientific, Radnor, PA, USA). Turbidity was measured at a visible wavelength (610 nm, Infinite Pro series UV plate reader, Tecan Ltd., Mannedorf, Switzerland) to detect nondissolved particles. The presence of these particles shows that the point of saturation has been reached. A sample of the solution was filtered and the volume of that filtered solution was diluted in green tea stabilizer

solution [250 mg/L Ascorbic Acid, 250 mg/L sodium ethylenediaminetetraacetic acid (Na-EDTA), 10% (v/v) acetonitrile in water] and read in the HPLC.

2.3.3. Differential Scanning Calorimetry (DSC).

Melting properties were measured with a differential scanning calorimeter (DSC-7, Perkin-Elmer, Wellesley, MA). Amounts from 2 to 5 mg of sample were heated at a rate of 10°C/m while measurements of heat flow were made.

2.3.4. HPLC Analysis.

Analysis was performed by HPLC (DIONEX IC-3000) on a C18 column (Luna 5 µm Phenyl-Hexyl, Phenomenex, Aschaffenburg, Germany) using acetic acid in acetonitrile (2% v/v) and acetic acid in 20 mg/L EDTA in water (2% v/v) as mobile phases and read at a wavelength of 278 nm with a photodiode array detector (DIONEX PDA-3000). In the resultant chromatogram (figure 2), the eight catechins plus important compounds from green tea such as gallic acid, theobromine or caffeine can be quantitatively determined.

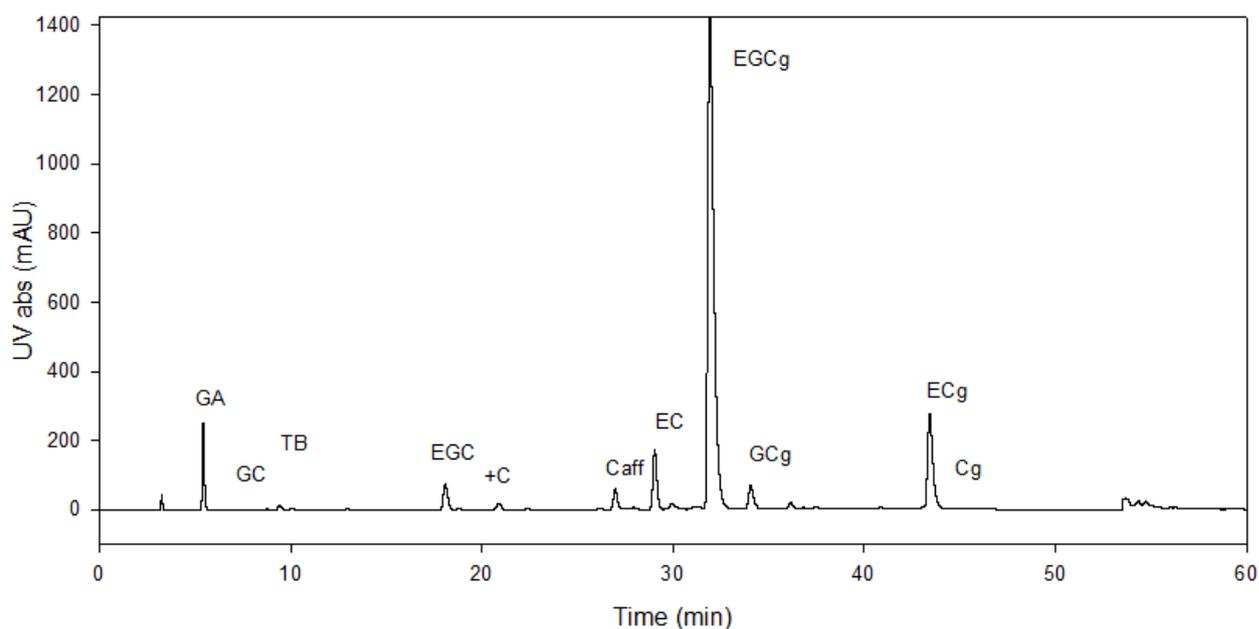


Figure 2. HPLC chromatogram of green tea catechins (gradient elution with water/acetonitrile, 30 °C)

2.3.5. Regression Analysis.

Because part of the aim of this work was to compare the performance of the most relevant models from the literature, the fit of the models was measured using the root-mean-square error (RMSE).

$$RMSE = \sqrt{\frac{\sum_{min}(\ln x_s^{cal} - \ln x_s^{exp})^2}{N}} \quad [5]$$

Experimental values are plotted against calculated ones for each model, as a visual help to show the performance of the models. Values of R^2 can be found in each plot.

The complex models detailed in the Appendix were programmed using the computer software MATLAB (Version 7.11.0.584 MathWorks, Inc., Natick, MA). This program was used as well for the calculation of all regressions and comparisons.

2.4. Results and Discussion

2.4.1. Solubility Profiles.

In Figure 3, solubility (in terms of molar fraction) is plotted against the solute-free ethanol molar fraction for the four epimers and at two different temperatures. For a temperature of 293 K, maxima are observed for all epimers for solute-free ethanol molar fractions from 0.6 to 0.8. For 303 K, on the other hand, solubility follows a linear trend, and the maxima shift to pure ethanol in most cases.

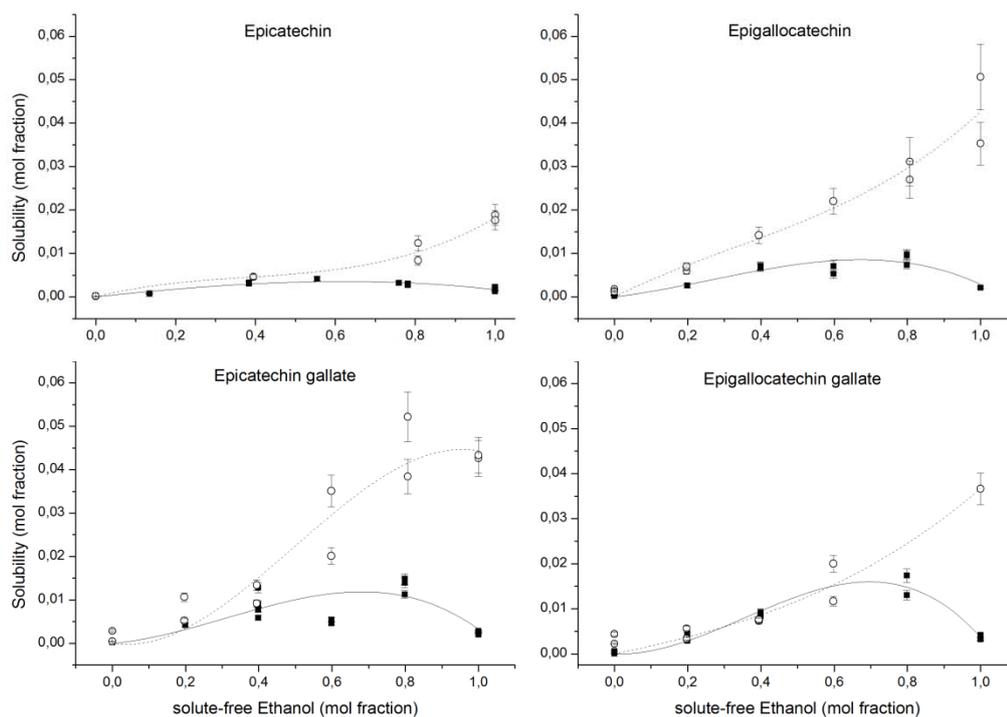


Figure 3. Solubility profiles of different solute-free concentrations of ethanol for the four main catechins at temperatures of 293 K (black squares) and 303 K (white circles). Trend lines (solid for 293 K and dashed for 303 K) are plotted to guide the eye.

For a better understanding of the nonideal behavior of the catechins in the mixed solvents, the ideal solubilities in the mixed of the solvents (dependent on the experimental solubility in each of the pure solvents) has to be subtracted. As a result, the excess solubility (according to Equation 3) is plotted against the solute-free molar fraction of ethanol. (Figure 4)

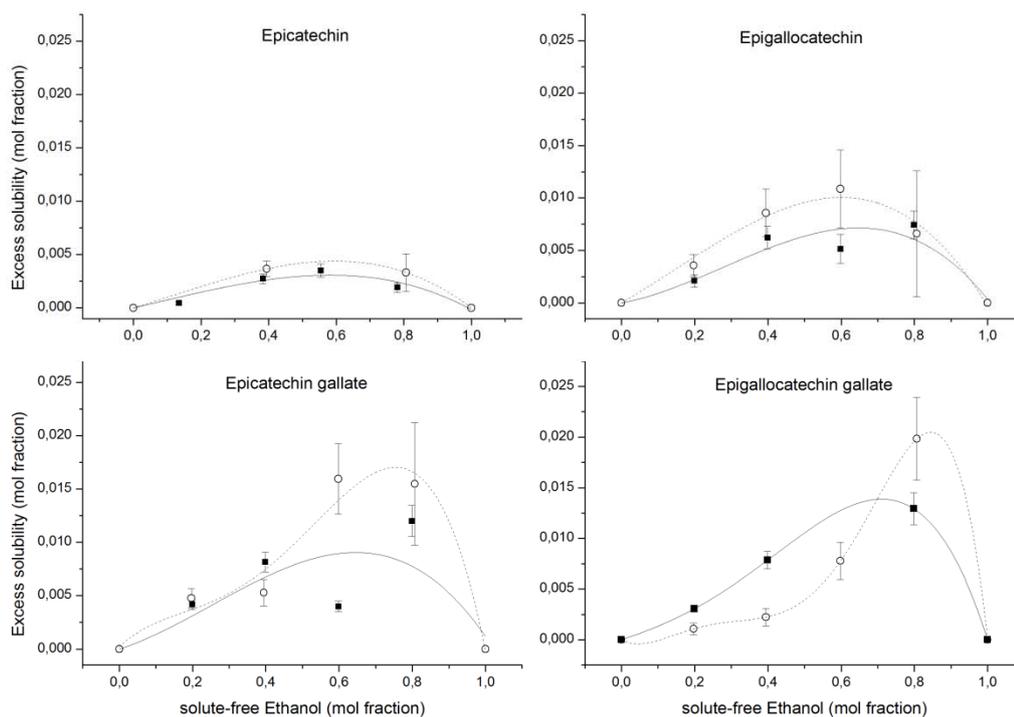


Figure 4. Excess solubility profiles of different solute-free concentrations of ethanol for the four main catechins at temperatures of 293 K (black squares) and 303 K (white circles). Trend lines (solid for 293 K and dashed for 303 K) are plotted to guide the eye.

As seen in Figure 4, the solubilities of all catechins appear to exhibit a convex curve despite the convexity or concavity of their behavior in Figure 3. This is due to the shape of the ideal solubility curve from Equation 4. Because of its concavity, all four molecules show a convex-shaped excess solubility. For both sets of solubility values, the behavior of the epimers is similar to that reported for other amphiphilic molecules like lamotrigine, clonazepam or diazepam³⁴ or even amino acids²⁰. In addition, the nonideal behavior in the measured range seems not to be strongly influenced by the temperature. As seen in the previously mentioned experiments the maximum

solubility occurs when the polarities of the solute and mixture of solvents are closest to each other.

2.4.2. Melting Properties.

The melting points were calculated from DCS experiments. Heat flow was plotted versus temperature, showing a peak for the temperature equal to the melting temperature. The enthalpy of fusion was calculated from the area beneath the peak. In all four cases, a reaction (probably oxidation) occurred after or during the melting of the compound.

Predictive GC models for the melting properties, such as those developed by Jain^{25,26}, Marrero and Gani²⁷ and Wang et al.,²⁸ are present in the literature and their performance was tested as well.

Table 1. Comparison of Melting Temperatures (T_m) of Epicatechin, Epigallocatechin, Epicatechin Gallate and Epigallocatechin Gallate

T_m (K)	Experimental values			Predicted values		
	This work	Shi et al. ³⁵	Park et al. ³⁶	Wang et al. ²⁷	Marrero et al. ²⁷	Jain et al. ²⁵
EC	526*	509	525	409	514	507
EGC	493*	-	-	429	533	553
ECg	542*	509	-	438	571	637
EGCg	503*	-	429	451	585	676

*Value used for calculations of the different models further on.

In the case of melting temperature, as can be seen in Table 1, every catechin showed a melting point within the range of 490-550 K. The experimental values suggest that the introduction of the gallate group stabilizes the molecule because gallated molecules show a higher melting point than their nongallated counterparts ($T_{m,ECg} > T_{m,EC}$ and

$T_{m,EGCg} > T_{m,EGC}$). On the other hand, the effect of the extra hydroxyl group in EGC and EGCg seems to have the opposite effect.

In the case of the predicted values, no model was entirely precise. The four models investigated in this work predict the increase of the melting point with molecular weight but fail to predict the decrease that the epigallomers (EGC and EGCg) show in comparison with their corresponding epimers (EC and ECg, respectively). Of all of the models, that of Marrero and Gani²⁷ shows more congruence with the experimental results.

As previously stated, the enthalpy of fusion is not directly measured but, rather, is calculated from the area beneath the melting peak. For ECg and EGCg, baseline separation was almost achieved between the two phenomena, whereas for EC and EGC, this was not the case. The value of the enthalpy of fusion is not sufficiently precise, because of the oxidation of part of the material before melting.

As Table 2 shows, the experimental value for EGCg is consistent with the literature, whereas as previously discussed, EC and EGC present lower values than expected. Literature values are available for EC,³⁵ and because the reported temperature is consistent with the experimental one, the reported value for ΔH_m was used from here on.

Table 2. Comparison of Melting Enthalpy Values (ΔH_m kJ/mol) of Epicatechin, Epigallocatechin, Epicatechin Gallate and Epigallocatechin Gallate.

ΔH_m (kJ/mol)	Experimental values		Predicted values	
	This work	Shi et al. ³⁵	Marrero et al. ²⁷	Jain et al. ²⁵
EC	11.198	56.682*	47.089	40.310
EGC	22.830	-	53.568*	44.020
ECg	63.290*	-	77.041	60.050
EGCg	58.039*	63.215	83.520	63.760

* Value used for calculations of the different models further on.

Table 2 also shows that, whereas the best performance for the gallate epimers (ECg and EGCg) was obtained with the Jain model, for the previously discussed EC value, the Marrero model was more precise. Consequently, the value for the ΔH_m of EGC was 53.568 J/mol.

The gallate epimers show the same trend in enthalpy of fusion as in melting temperature and that trend is not predicted with any of the models. This difference in the prediction of the models can be explained by an underestimation of the contribution parameter of the group aC-COO- to the enthalpy of fusion in the model of Marrero²⁷.

2.4.3. Performance of the Different Models.

2.4.3.1. G^E models. The performance of two models is compared. Regressed parameters of the latter can be found in table 3:

Table 3. NRTL-SAC Regressed Parameters ($r_{i,j}$) for Different Temperatures

Compound	X	Y ⁻	Y ⁺	Z
<i>T = 293.15 K</i>				
EC	0.243	0.767	0.000	0.744
EGC	0.268	0.799	0.000	0.726
ECg	0.260	1.316	0.000	1.006
EGCg	0.323	0.946	0.000	1.307
<i>T = 303.15 K</i>				
EC	0.363	1.013	0.000	0.000
EGC	0.253	1.031	0.000	0.000
ECg	0.272	1.542	0.000	0.000
EGCg	0.137	0.861	0.000	0.000

2.4.3.2. Excess Solubility Models. The clear drawback of this type of model is that excess solubility is not a useful property without the values of the solubilities in the pure solvents, which have to be determined experimentally. Jouyban model uses Hansen Parameters than can be predicted by the use of the Stefanis et al.²⁹ model (Table 4)

Table 4. Hansen Parameters Predicted with Stefanis et al.²⁹

Compound	δ_p	δ_d	δ_{hb}	δ_t
	MPa ^{1/2}	MPa ^{1/2}	MPa ^{1/2}	MPa ^{1/2}
EC	22.76	6.78	41.55	47.86
EGC	23.18	8.42	48.94	54.80
ECg	25.48	12.25	54.74	61.61
EGCg	25.89	13.88	62.13	68.72

In addition to the previously explained drawback of excess solubility models, the model of Gude et al.²⁰ does not include a temperature dependence. To account for this deficiency, different parameters were regressed for the different temperatures. The value for $A_{w,ethanol}$ used was 1.55, as recommended in the original publication and the regressed values of C_{ijl} are listed in Table 5.

Table 5. Regressed Ternary Interaction Parameters for the Model Developed by Gude et al.²⁰ at Different Temperatures

C_{ijl}	EC	EGC	ECg	EGCg
293 K	7.144	5.750	8.301	4.590
303 K	10.309	5.030	9.735	2.895

Two trends can be observed in Table 5: Whereas the values of C_{ijl} seem to decrease with temperature for the epigallomers (EGC and EGCg), the other epimers (EC and ECg) behave oppositely. These two different behaviours of epigallomers and nonepigallomers were observed, as well, for melting properties.

2.4.3.3. Discussion of the Performance of the Different Models. A summary of the RMS error values is displayed in Table 6. For every model, the number of regressed parameters is also included. In the case of excess solubility models, as previously explained, the values of solubility in pure solvents are needed for the model to produce meaningful calculations, but because they are determined experimentally and not regressed, they are not shown.

Table 6. RMS Error Values for the Different Models

	RMS	n
Mod-UNIFAC	1.961	0
NRTL-SAC	0.550	16
Jouyban et al.	0.762	0
Gude et al.	0.625	8

n: number of parameters regressed.

In the case of activity coefficient models, the performance of Mod-UNIFAC is quite poor for low solubility values ($\ln x_s \leq -6$). This behavior can be caused by a wrong performance of the model when the solvent consists mainly of water, because those are the values for which the solubility is low. The NRTL-SAC model shows an exceptional error value, comparable with those in the literature¹⁵ (0.027 to 1.007). For excess solubility calculations, both models perform remarkably well. The fact that solubility parameters were predicted with one model and applied to a different one does not affect the performance. On the other hand, the model gives a good fit when only one parameter per equilibrium was regressed.

2.5. Conclusions

A set of experimental solubility values for the four main catechins in water, ethanol, and their mixtures has been presented. Experimental values of melting properties for most compounds were reported whereas for others, the values were calculated from predictive models.

The performance of different GC methods for predicting properties such as melting temperature, melting enthalpy or solubility parameters was compared with experimental values, showing agreement between them.

A study of the most significant models used for drug-like solubility predictions and modelling was carried out and four models were chosen for their use in this work. The performance of those models was compared with the experimental values and with the rest of the models.

Models for the prediction of excess solubility showed, on average, a better fit than those based on activity coefficient calculations, as is clearly shown by the model of Jouyban et al.²⁴ This model performed with a good RSM error without the need for any regression but the molecular configuration. Nonetheless, the NRTL-SAC model is the one that fits the experimental data best, and since the parameters have been regressed, it can be used for other calculations in which activity coefficients of these components are needed. However, because the list of polyphenols is rather vast but the number of groups that comprises them is not that large, it would be interesting to develop a UNIFAC-like activity coefficient model with parameters regressed from solubility data because such a model will be extremely useful for the design of industrial processes for the separation, concentration, and purification of these nutraceutical compounds.

Appendix I: G^E Models:

I.A. Mod-UNIFAC Model: Modified version of the UNIFAC model⁹⁻¹³. The activity coefficient is the sum of a combinatorial part and a residual part (Equation A1):

$$\ln(\gamma_i) = \{\ln(\gamma_i^C)\} + \{\ln(\gamma_i^R)\} = \left\{1 - V_i' + \ln(V_i') - 5q_i \left[1 - \frac{V_i}{F_i} + \ln\left(\frac{V_i}{F_i}\right)\right]\right\} + \left\{\sum_k v_k^{(i)} \left[\ln(\Gamma_k) - \ln\left(\Gamma_k^{(i)}\right)\right]\right\} \quad [\text{A1}]$$

where the parameters for the combinatorial part are calculated according to the equations A2-A6.

$$V_i' = \frac{r_i^{3/4}}{\sum_j x_j r_j^{3/4}} \quad [\text{A2}]$$

$$V_i = \frac{r_i}{\sum_j x_j r_j} \quad [\text{A3}]$$

$$F_i = \frac{q_i}{\sum_j x_j q_j} \quad [\text{A4}]$$

$$r_i = \sum v_k^{(i)} R_k \quad [\text{A5}]$$

$$q_i = \sum v_k^{(i)} Q_k \quad [\text{A6}]$$

and the parameters for the residual contribution are calculated according to the equations A7-A10.

$$\ln(\Gamma_k) = Q_k \left[1 - \ln\left(\sum_m \theta_m \Psi_{nm}\right) - \sum_m \frac{\theta_m \Psi_{nm}}{\sum_n \theta_n \Psi_{nm}}\right] \quad [\text{A7}]$$

$$\theta_m = \frac{Q_m X_m}{\sum_n Q_n X_n} \quad [\text{A8}]$$

$$X_m = \frac{\sum_j v_m^{(j)} x_j}{\sum_j \sum_m v_n^{(j)} x_j} \quad [\text{A9}]$$

$$\Psi_{nm} = \exp\left(-\frac{a_{nm} + b_{nm}T + c_{nm}T^2}{T}\right) \quad [\text{A10}]$$

IB – NRTL-SAC model: The NRTL-SAC model is the combination of the NRTL model with segment theory.^{14, 15} The molecule is divided into four types of segments: hydrophobic (X), polar attractive (Y+), polar repulsive (Y-), and hydrophilic (Z). The

interactions between these segments are already defined in tables. Only four parameters characterize each molecule, namely, the contributions of each of these segments. As in the UNIFAC model, there is a combinatorial part and a residual part

$$\ln(\gamma_i) = \{\ln(\gamma_i^C)\} + \{\ln(\gamma_i^R)\} \quad [\text{A11}]$$

The combinatorial part is calculated from the Flory-Huggins equation for the combinatorial entropy of mixing.

$$\ln(\gamma_i^C) = \left\{ \ln\left(\frac{\phi_i}{x_i}\right) + 1 - r_i \sum_k \frac{r_k}{\phi_k} \right\} \quad [\text{A12}]$$

$$r_i = \sum_l r_{i,l} \quad [\text{A13}]$$

$$\phi_i = \frac{r_i x_i}{\sum_j r_j x_j} \quad [\text{A14}]$$

$$\ln(\gamma_i^R) = \sum_k r_{i,l} [\ln(\Gamma_i^{lc}) - \ln(\Gamma_{i,l}^{lc})] \quad [\text{A15}]$$

$$\ln(\Gamma_i^{lc}) = \frac{\sum_j x_j G_{ji} \tau_{ji}}{\sum_j x_j G_{ji}} + \sum_l \frac{x_l G_{il}}{\sum_k x_k G_{kl}} \left(\tau_{il} - \frac{\sum_j x_j G_{jl} \tau_{jl}}{\sum_k x_k G_{kl}} \right) \quad [\text{A16}]$$

$$\ln(\Gamma_{i,l}^{lc}) = \frac{\sum_j x_{j,l} G_{ji} \tau_{ji}}{\sum_j x_{j,l} G_{ji}} + \sum_l \frac{x_{l,l} G_{il}}{\sum_k x_{k,l} G_{kl}} \left(\tau_{il} - \frac{\sum_j x_{j,l} G_{jl} \tau_{jl}}{\sum_k x_{k,l} G_{kl}} \right) \quad [\text{A17}]$$

$$x_j = \frac{\sum_l x_{j,l} r_{j,l}}{\sum_l \sum_i x_{i,l} r_{i,l}} \quad [\text{A18}]$$

$$x_{j,l} = \frac{r_{j,l}}{\sum_l r_{i,l}} \quad [\text{A19}]$$

$$G_{ji} = e^{-\alpha_{ji} \tau_{ji}} \quad [\text{A20}]$$

Capital letters in a subscripts represent components, whereas lowercase letters represent segments.

Appendix II: Excess Solubility Models:

IIA – Model Developed by Jouyban et al.²⁴: This is an empirical model based on the Jouyban-Acre model²⁴ in which the compound-dependent constants are related to the Hansen parameters²⁹.

$$\log X_{m,T} = f_c \log X_{c,T} + f_w \log X_{w,T} + \left(\frac{f_c f_w}{T}\right) \left[A_0 \delta_{ds} (\delta_{dc} - \delta_{dw})^2 + A_1 \delta_{ps} (\delta_{pc} - \delta_{pw})^2 + A_2 \delta_{hs} (\delta_{hc} - \delta_{hw})^2 \right] + \left(\frac{f_c f_w (f_c - f_w)}{T}\right) \left[A_3 \delta_{ds} (\delta_{dc} - \delta_{dw})^2 + A_4 \delta_{ps} (\delta_{pc} - \delta_{pw})^2 + A_5 \delta_{hs} (\delta_{hc} - \delta_{hw})^2 \right] + \left(\frac{f_c f_w (f_c - f_w)^2}{T}\right) \left[A_6 \delta_{ds} (\delta_{dc} - \delta_{dw})^2 + A_7 \delta_{ps} (\delta_{pc} - \delta_{pw})^2 + A_8 \delta_{hs} (\delta_{hc} - \delta_{hw})^2 \right] \quad [\text{A21}]$$

The model has nine characteristic constants (Table A1) that can be fitted for different conditions. The values reported in Table A1 are appropriate for water + cosolvent mixtures.

Table A1. Specific parameters needed for Jouyban model in the case of water-cosolvent solubility.

A ₀	A ₁	A ₂	A ₃	A ₄	A ₅	A ₆	A ₇	A ₈
0	0.606	0.013	-8.696	0.37 6	0.013	9.277	-0.461	0.017

IIB – Model developed by Gude et al.²⁰: This model characterizes the system by two interaction parameters: one for cosolvent-solvent interactions and a=one for cosolvent-solvent-solute interactions²⁰.

$$\ln x_i^E = \ln r' - \sum_j x'_j \ln r_j + r_i \left(\frac{1}{r'} - \sum_j \frac{x'_j}{r_j} \right) + \sum_j \sum_l [A_{jl} x'_j x'_l (1 + C_{jli})] \quad [\text{A22}]$$

where subscripts j and l relate to solvents and subscript i relates to the solute.

Acknowledgements

This work was supported by the ISPT (Institute for Sustainable Process Technology) under the project FO-10-03 Separation of vitality ingredients. Thanks, as well, to the industrial partner UNILEVER for financial support and valuable comments. Special thanks to Ir. N. Geerlofs for his help with the DSC experiments.

Nomenclature

- A_0 - A_8 model constants for Jouyban et al.²⁴ model in Equation A21.
- A_{jl} solvent-cosolvent interaction parameter in Equation A22.
- a_{nm} UNIFAC group interaction parameter between groups n and m (K) in Equation A10.
- b_{nm} UNIFAC group interaction parameter between groups n and m in Equation A10.
- c_{nm} UNIFAC group interaction parameter between groups n and m (K^{-1}) in Equation A10.
- C_{ijl} ternary interaction parameter from Equation A22.
- dw: dry weight.
- EC: epicatechin.
- ECg: epicatechin gallate.
- EGC: epigallocatechin.
- EGCg: epigallocatechin gallate.
- f volume fraction in Equation A21.
- F_i auxiliary property for component i in Equations A1 and A4.
- G Gibbs energy.
- G_{ij} interaction energy between segments i and j in Equations A16, A17 and A20.
- q_i relative van der Waals surface area of component i in Equations A1 and A4.

Q_k	relative van der Waals surface area of group k in Equations A6 and A8.
$r_{i,I}$	contribution of segment i to molecule I in Equations A12 to A20.
r_I	total contribution of segments of the molecule I in Equations A12 to A20.
r_i	relative van der Waals volume of the component i in Equations A2, A3 and A5.
R	ideal gas constant (8.314 J/(mol K)).
R^2	Coefficient of determination.
R_k	relative van der Waals volume of the group k in Equation A5.
RMS	Root mean square.
T	temperature.
V_i	auxiliary property for component i in Equation A1 and A3.
V'_i	empirically modified V_i -value in Equation A1 and A2.
x	mole fraction in the liquid phase.
x_i	molar fraction of segment i in the mixture in Equations A16 and A18.
$x_{i,J}$	molar fraction of segment i in pure component J in Equations A17 and A19
X	hydrophobic interaction for the NRTL-SAC model.
X_m	group mole fraction of group m in the liquid phase in Equations A8 and A9.
Y	polar attractive interaction for the NRTL-SAC model.
Y^+	polar repulsive interaction for the NRTL-SAC model.
Z	hydrophilic interaction for the NRTL-SAC model.

Greek letters

α_{ij}	nonrandomness parameter between segment i and j in Equation A20.
γ	activity coefficient.
Γ_i^{lc}	group activity coefficient of group i in the mixture in Equations from A15 to A17.
$\Gamma_{i,I}^{lc}$	group activity coefficient of group i in pure I in Equations from A15 to A17.

- $\Gamma_k^{(i)}$ group activity coefficient of group k in the pure substance i in Equation A1.
- δ solubility parameter.
- ΔC_p^{sl} difference between the liquid and solid molar heat capacities.
- ΔH^{sl} fusion enthalpy (J/mol).
- $\nu_k^{(i)}$ number of structural groups of type k in molecule i in Equations A1, A5, A6 and A9.
- τ_{ij} binary interaction parameter between segment i and j in Equations A12 to A20.
- ϕ_I fraction of the total contribution of segments of compound I in Equations A12 and A14
- Ψ_{nm} UNIFAC group interaction parameter between groups n and m in Equation A7 and A10.

Superscripts

- ' solute-free.
- ∞ infinite dilution.
- C combinatorial.
- E excess.
- l liquid phase.
- R residual.

Subscripts

- c cosolvent (ethanol in most cases).
- d dispersion, for solubility parameters.
- hb hydrogen-bonding, for solubility parameters.
- i component.
- m melting.

p	polar, for solubility parameters.
s	solute.
t	total, for solubility parameters.
w	water.

References

- (1) Graham, H. N., Green tea composition, consumption, and polyphenol chemistry. *Prev. Med.* **1992**, 21, 3, 334.
- (2) Ding, Z. H.; Chen, Y.; Zhou, M.; Fang, Y. Z., Inhibitory effect of green tea polyphenol and morin on the oxidative modification of low density lipoprotein. *Chinese Journal of Pharmacology and Toxicology* **1992**, 6, 4, 263.
- (3) Khan, N.; Mukhtar, H., Tea polyphenols for health promotion. *Life Sci.* **2007**, 81, 7, 519.
- (4) Agarwal, R.; Katiyar, S. K.; Zaidi, S. I. A.; Mukhtar, H., Inhibition of skin tumor promoter-caused induction of epidermal ornithine decarboxylase in SENCAR mice by polyphenolic fraction isolated from green tea and its individual epicatechin derivatives. *Cancer Res.* **1992**, 52, 13, 3582.
- (5) Van Der Wielen, L. A. M.; Rudolph, E. S. J., On the generalization of thermodynamic properties for selection of bioseparation processes. *J. Chem. Technol. Biotechnol.* **1999**, 74, 3, 275.
- (6) Ahamed, T.; Ottens, M.; Nfor, B. K.; Van Dedem, G. W. K.; Van Der Wielen, L. A. M., A generalized approach to thermodynamic properties of biomolecules for use in bioseparation process design. *Fluid Phase Equilib.* **2006**, 241, 1-2, 268.
- (7) Prigogine, I. a. R. D., *Chemical Thermodynamics*. Longman: London, 1954.
- (8) Neau, S. H.; Flynn, G. L., Solid and Liquid Heat Capacities of n-Alkyl Para-aminobenzoates Near the Melting Point. *Pharm. Res.* **1990**, 7, 11, 1157.
- (9) Gmehling, J.; Wittig, R.; Lohmann, J.; Joh, R., A Modified UNIFAC (Dortmund) Model. 4. Revision and Extension. *Ind. Eng. Chem. Res.* **2002**, 41, 6, 1678.
- (10) Gmehling, J.; Lohmann, J.; Jakob, A.; Li, J.; Joh, R., A Modified UNIFAC (Dortmund) Model. 3. Revision and Extension. *Ind. Eng. Chem. Res.* **1998**, 37, 12, 4876.
- (11) Gmehling, J.; Li, J.; Schiller, M., A modified UNIFAC model. 2. Present parameter matrix and results for different thermodynamic properties. *Ind. Eng. Chem. Res.* **1993**, 32, 1, 178.
- (12) Fredenslund, A.; Gmehling, J.; Michelsen, M. L.; Rasmussen, P.; Prausnitz, J. M., Computerized design of multicomponent distillation columns using the UNIFAC group contribution method for calculation of activity coefficients. *Industrial and Engineering Chemistry Process Design and Development* **1977**, 16, 4, 450.
- (13) Abbas, R.; Gmehling, J., Vapour-liquid equilibria, azeotropic data, excess enthalpies, activity coefficients at infinite dilution and solid-liquid equilibria for binary alcohol-ketone systems. *Fluid Phase Equilib.* **2008**, 267, 2, 119.

- (14) Chen, C.-C.; Crafts, P. A., Correlation and Prediction of Drug Molecule Solubility in Mixed Solvent Systems with the Nonrandom Two-Liquid Segment Activity Coefficient (NRTL–SAC) Model. *Ind. Eng. Chem. Res.* **2006**, *45*, 13, 4816.
- (15) Chen, C. C.; Song, Y., Solubility modeling with a nonrandom two-liquid segment activity coefficient model. *Industrial and Engineering Chemistry Research* **2004**, *43*, 26, 8354.
- (16) Klamt, A.; Eckert, F.; Hornig, M.; Beck, M. E.; Bürger, T., Prediction of aqueous solubility of drugs and pesticides with COSMO-RS. *J. Comput. Chem.* **2002**, *23*, 2, 275.
- (17) Hsieh, C.-M.; Wang, S.; Lin, S.-T.; Sandler, S. I., A Predictive Model for the Solubility and Octanol–Water Partition Coefficient of Pharmaceuticals. *J. Chem. Eng. Data* **2010**, *56*, 4, 936.
- (18) Cassens, J.; Ruether, F.; Leonhard, K.; Sadowski, G., Solubility calculation of pharmaceutical compounds – A priori parameter estimation using quantum-chemistry. *Fluid Phase Equilib.* **2010**, 299, 1, 161.
- (19) Ellegaard, M. D.; Abildskov, J.; O’Connell, J. P., Molecular Thermodynamic Modeling of Mixed Solvent Solubility. *Ind. Eng. Chem. Res.* **2010**, *49*, 22, 11620.
- (20) Gude, M. T.; Van Der Wielen, L. A. M.; Luyben, K. C. A. M., Phase behavior of α -amino acids in multicomponent aqueous alkanol solutions. *Fluid Phase Equilib.* **1996**, 116, 1-2, 110.
- (21) Yalkowsky, S. H.; Flynn, G. L.; Amidon, G. L., Solubility of nonelectrolytes in polar solvents. *J. Pharm. Sci.* **1972**, *61*, 6, 983.
- (22) Valvani, S. C.; Yalkowsky, S. H.; Roseman, T. J., Solubility and partitioning IV: Aqueous solubility and octanol-water partition coefficients of liquid nonelectrolytes. *J. Pharm. Sci.* **1981**, *70*, 5, 502.
- (23) Li, A., Predicting cosolvency. 3. Evaluation of the extended log-linear model. *Industrial and Engineering Chemistry Research* **2001**, *40*, 22, 5029.
- (24) Jouyban, A.; Shayanfar, A.; Panahi-Azar, V.; Soleymani, J.; Yousefi, B. H.; Acree, W. E.; York, P., Solubility prediction of drugs in mixed solvents using partial solubility parameters. *J. Pharm. Sci.* **2011**, *100*, 10, 4368.
- (25) Jain, A.; Yang, G.; Yalkowsky, S. H., Estimation of Melting Points of Organic Compounds. *Ind. Eng. Chem. Res.* **2004**, *43*, 23, 7618.
- (26) Jain, A.; Yalkowsky, S. H., Estimation of melting points of organic compounds-II. *J. Pharm. Sci.* **2006**, *95*, 12, 2562.
- (27) Marrero, J.; Gani, R., Group-contribution based estimation of pure component properties. *Fluid Phase Equilib.* **2001**, 183-184, 183.
- (28) Wang, Q.; Ma, P.; Neng, S., Position Group Contribution Method for Estimation of Melting Point of Organic Compounds. *Chin. J. Chem. Eng.* **2009**, *17*, 3, 468.
- (29) Stefanis, E.; Panayiotou, C., Prediction of Hansen Solubility Parameters with a New Group-Contribution Method. *Int. J. Thermophys.* **2008**, *29*, 2, 568.
- (30) Mota, F. L.; Queimada, A. J.; Andreatta, A. E.; Pinho, S. P.; Macedo, E. A., Calculation of drug-like molecules solubility using predictive activity coefficient models. *Fluid Phase Equilib.* **2012**, 322–323, 0, 48.
- (31) Sheikholeslamzadeh, E.; Rohani, S., Solubility Prediction of Pharmaceutical and Chemical Compounds in Pure and Mixed Solvents Using Predictive Models. *Ind. Eng. Chem. Res.* **2011**, *51*, 1, 464.

- (32) Bouillot, B.; Teychené, S.; Biscans, B., An evaluation of thermodynamic models for the prediction of drug and drug-like molecule solubility in organic solvents. *Fluid Phase Equilib.* **2011**, 309, 1, 36.
- (33) Alsenz, J.; Kansy, M., High throughput solubility measurement in drug discovery and development. *Adv. Drug Delivery. Rev.* **2007**, 59, 7, 546.
- (34) Shayanfar, A.; Fakhree, M. A. A.; Acree Jr, W. E.; Jouyban, A., Solubility of lamotrigine, diazepam, and clonazepam in ethanol + water mixtures at 298.15 K. *J. Chem. Eng. Data* **2009**, 54, 3, 1107.
- (35) Shi, J.; Yu, J.; Pohorly, J. E.; Kakuda, Y., Polyphenolics in Grape Seeds - Biochemistry and Functionality. *J. Med. Food* **2003**, 6, 4, 291.
- (36) Park, K. A.; Lee, H. J.; Hong, I. K., Solubility prediction of bioantioxidants for functional solvent by group contribution method. *Journal of Industrial and Engineering Chemistry* **2010**, 16, 3, 490.

Chapter 3: MPP-UNIFAC, A Predictive Activity Coefficient Model for Polyphenols

Abstract

Modified UNIFAC (Mod. UNIFAC) is a proven model for the prediction of activity coefficients of molecules in non-ideal mixtures. However, Mod. UNIFAC is often not accurate when hydrogen bonding or strong hydrophobic interactions are present. An interesting group of molecules called polyphenols presents both types of interactions and therefore, Mod. UNIFAC predictions are not always accurate. A polyphenol-specific UNIFAC-based method (MPP-UNIFAC) has been developed in this study for its application on polyphenol-like molecules. Interaction parameters were regressed from 411 solubility data points from polyphenols in literature showing a good fit. The average error of the fit is 0.094 log units, better than the average value for Mod. UNIFAC (0.5 log units). Moreover, the model was validated against three datasets that were not used for the regression giving more accurate predictions than the Mod. UNIFAC it was developed from.

Published as: MPP-UNIFAC, a predictive activity coefficient model for polyphenols.

Méndez Sevillano, D.; van der Wielen, L. A. M.; Hooshyar, N.; Ottens, M., *Fluid Phase Equilib.* **2014**, 384, 82-88.

3.1.Introduction

Polyphenol is a vast concept that groups different types of molecules. These molecules consist of two or more phenyl rings with a number of hydroxyl groups attached to them. Besides these hydroxyl groups, they can have any type of group attached to it. The type of polyphenols that this work focuses on is flavonoids.

Flavonoids consist of two phenolic rings connected by a three-carbon and one-oxygen chain that forms a third ring fused to the first one. The name of this type of polyphenols comes from the similarities between their backbone to the molecule flavan (Figure 1).

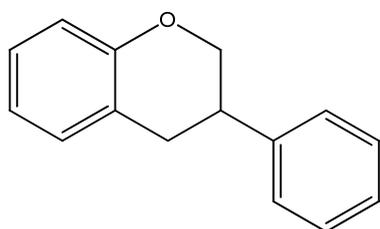


Figure 1: Molecular structure of flavan.

The non-aromatic ring can change by either the substitution of one of the alkane groups for a ketone group, the presence of double bonds or even the substitution of the entire ring by a saturated 2-carbon chain (i.e. stilbenoids). All flavonoids have a number of hydroxyl groups attached (either to the aromatic rings or the non-aromatic one). Other attached groups can be glucose molecules, methyl ethers, gallic esters, etc.

Flavonoids can be classified in different groups based on the conformation of the three rings that form the backbone of the molecule ¹. That classification can be found in Figure 2.

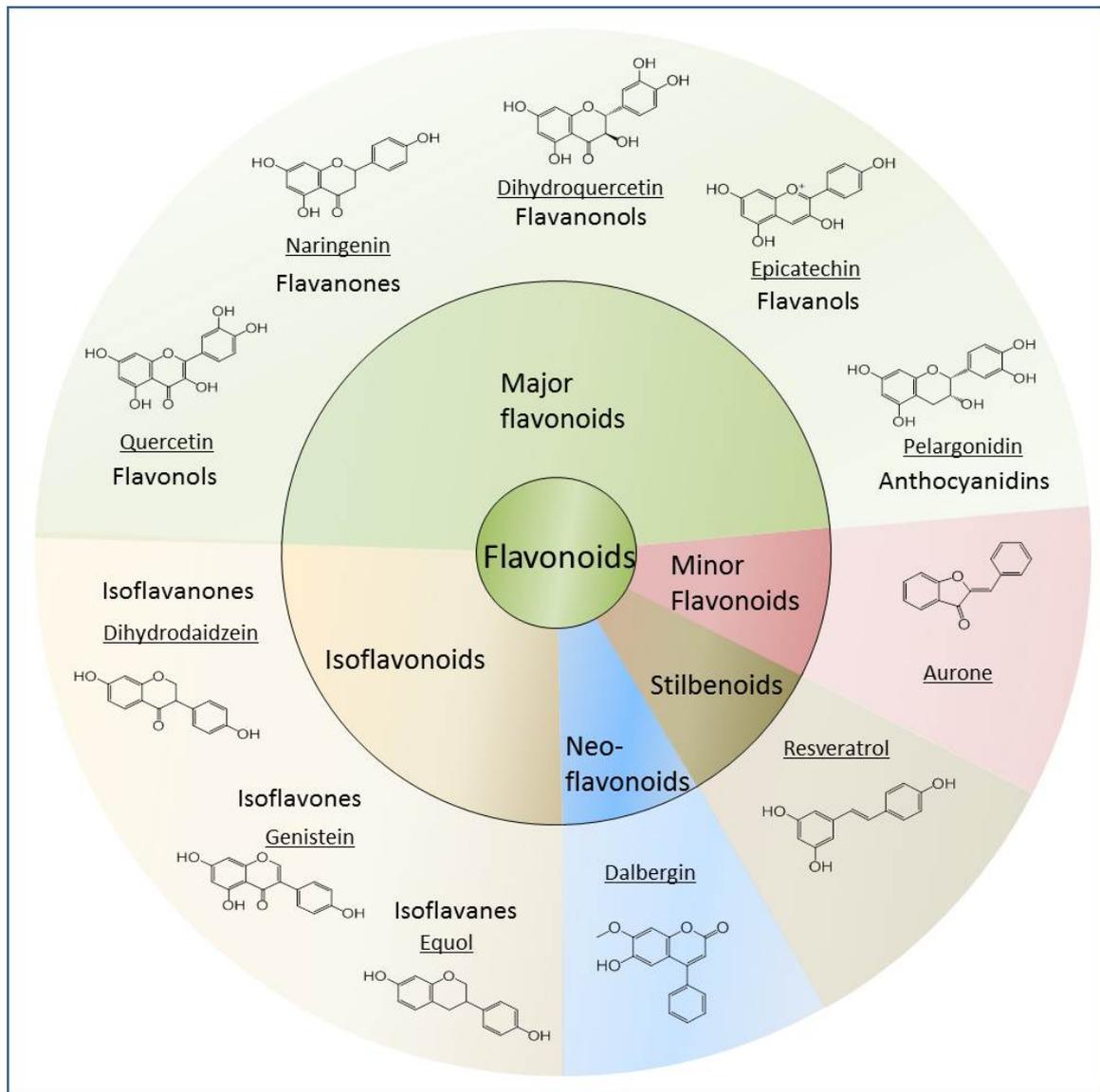


Figure 2. Classification of flavonoids.

Part of the importance of these molecules are the health benefits that some of them show. Flavanones such as naringenin or hesperidin show lowering of high cholesterol levels,² some studies relate isoflavones with reduction of hot flashes³ or colon cancer⁴ and catechins from green tea show how to reduce obesity and certain types of cancer.⁵ Such polyphenols are usually present in food sources in very small concentrations. Therefore, their separation and purification entails complex processes such as liquid-liquid extraction, solid-liquid extraction, crystallization, etc. In order to design and model

these processes, specific physicochemical and thermodynamic properties of the polyphenols are needed which would require costly and lengthy experiments.

A large number of properties can be predicted a priori using activity coefficients. Activity coefficients have proven to be a powerful tool for the prediction of the behavior of a wide range of components in different kinds of mixtures. These activity coefficients can be either modelled with models such as NRTL-SAC⁶ or predicted with models such as UNIFAC⁷.

UNIFAC (UNIQUAC Functional-group Activity Coefficients) is a group contribution model for activity coefficients fully based on the UNIQUAC model⁸. The difference between these two models is based on the interaction parameters. In the case of UNIQUAC, they are regressed from experimental data while in the case of UNIFAC, they are calculated with the contribution of the different molecular subgroups. Even though the interaction parameters between those groups have been regressed from experimental data, they can be applied to molecules not considered in the original dataset. The strength of UNIFAC lies on its prediction capability without the need of previous experimental effort and therefore, it is thoroughly used in industry and found in the open literature. Later on, the Modified UNIFAC model⁹ (and its later revisions and extensions¹⁰⁻¹³) was developed to improve the performance of the original model.

Mod. UNIFAC has proven its usefulness in the prediction of properties for small and medium sized molecules^{14,15}. However, it has shown rather inaccurate predictions when the studied molecules are either too large or present strong van der Waals forces, such as hydrogen bonding or weak ones, such as π - π interactions¹⁶. Méndez Sevillano et al.¹⁷ showed how the prediction of Mod. UNIFAC was most inaccurate for smaller activity coefficient values, typically related to mixtures in which attractive forces being present, such as hydrogen bonds. These challenges have been overcome in the past by the

extension and/or redefinition of new and specific groups in cases like ionic liquids¹⁸ or polymers¹⁹.

The aim of this work is the development and validation of a mod-UNIFAC-based activity coefficient model, MPP (Modified Poly Phenol) UNIFAC, specific for polyphenols. A new set of 9 groups and 16 subgroups has been defined to its application to polyphenols and interaction parameters have been regressed from literature solubility data.

Furthermore, the model has been validated by comparing its prediction to experimental datasets not used for the regression.

3.2. Theoretical framework

Activity coefficients in literature are calculated from different types of equilibrium properties (VLE, LLE, SLE, etc.). Due to the high melting point of polyphenols, they are mostly found in a solid state and solubility is the most commonly published equilibrium data. From solubility data, activity coefficients can be calculated based on Equation 1. This equation is derived from basic well-known thermodynamic relations assuming a unique pure solid phase.

$$x_i \gamma_i^l = \exp \left\{ \frac{\Delta H_i^{fus}}{RT_{m_i}} \left(\frac{T - T_{m_i}}{T} \right) + \frac{\Delta C_{p_i}^{fus}}{R} \left[\ln \frac{T}{T_{m_i}} - \left(\frac{T - T_{m_i}}{T} \right) \right] \right\} \quad [1]$$

This equation can be further simplified considering $\ln(T/T_{m,i})$ equal to $(T - T_{m,i})/T$ ²⁰ leading to Equation 2.

$$x_i \gamma_i^l = \exp \left[\frac{\Delta H_i^{fus}}{RT_{m_i}} \left(\ln \frac{T}{T_{m_i}} \right) \right] \quad [2]$$

Mod-UNIFAC calculates activity coefficients taking into account two contributions: the combinatorial and the residual. The first one accounts for the difference in size and shape of the groups while the second one accounts for the possible interactions between them.

$$\ln(\gamma_i) = \ln(\gamma_i^C) + \ln(\gamma_i^R) \quad [3]$$

Each of the contributions can be calculated with Equations 4 and 5.

$$\ln(\gamma_i^C) = 1 - V_i' - \ln(V_i') + 5q_i \left[1 - \frac{V_i}{F_i} + \ln\left(\frac{V_i}{F_i}\right) \right] \quad [4]$$

$$\ln(\gamma_i^R) = \sum_k v_k^{(i)} \left[\ln(\Gamma_k) - \ln(\Gamma_k^{(i)}) \right] \quad [5]$$

as a function of the following variables and the matrix $v_k^{(i)}$, a descriptor of the different groups that form the molecules present in the equilibrium.

$$V_i' = \frac{r_i^{3/4}}{\sum_j x_j r_j^{3/4}} \quad [6]$$

$$V_i = \frac{r_i}{\sum_j x_j r_j} \quad [7]$$

$$F_i = \frac{q_i}{\sum_j x_j q_j} \quad [8]$$

$$r_i = \sum v_k^{(i)} R_k \quad [9]$$

$$q_i = \sum v_k^{(i)} Q_k \quad [10]$$

$$\ln(\Gamma_k) = Q_k \left[1 - \ln(\sum_m \theta_m \Psi_{km}) - \sum_m \frac{\theta_m \Psi_{km}}{\sum_n \theta_n \Psi_{nm}} \right] \quad [11]$$

where:

$$\theta_m = \frac{Q_m X_m}{\sum_n Q_n X_n} \quad [12]$$

$$X_m = \frac{\sum_j v_m^{(j)} x_j}{\sum_j \sum_m v_n^{(j)} x_j} \quad [13]$$

$$\Psi_{nm} = \exp\left(-\frac{a_{nm} + b_{nm}T + c_{nm}T^2}{T}\right) \quad [14]$$

The calculation of $\Gamma_k^{(i)}$ is similar to Γ_k but, while the second represents the group activity of the group k in the mixture, the first accounts for the group activity of the group k in the molecule i .

This model needs certain inputs for a correct performance:

-*Conditions*: temperature and mole fractions of all components need to be supplied at different parts of the model (Equations 6-8 and 13-14).

-*Combinatorial parameters*: they are present in the first part of the model and they account for non-ideal behaviours caused by differences in size (R_k) and surface area (Q_k) of the group (Equations 9-12).

-*Residual parameters*: The variable Ψ_{nm} accounts for interaction between groups (n and m in this case) which is a polynomial equation of the temperature (Eq. 14). Each of the constants of that polynomial equation is a parameter specific for each interaction between groups.

3.3. Materials and methods

3.3.1. Selection of dataset

As previously mentioned, the goal of the set of parameters regressed in this work is to offer a more accurate prediction of flavonoids behavior in mixtures. Furthermore, UNIFAC is a non-electrolyte activity coefficient, so fully charged groups are not modelled by it. The developed model LIFAC²¹ deals with this issue by defining charged groups as different ones. However, the data gathered is not enough for determining the interaction parameters needed for two extra needed groups (COO^- and O^+). In order to follow that restriction, a group from the aforementioned classification (Figure 2) has to be left out: the anthocyanins. As well, phenolic acids are not used for this work due to their protonation in presence of water.

All data consider for this study can be found in Table 1

Table 1: Literature data used for the regression

Molecules	Solvents	Temperature Range (K)	Number of data points	Reference
Rutin	Ethanol, water	293.2-313.2	33	²²
Luteolin	Ethanol, water	293.2-313.2	33	²³
Apigenin	Ethanol, water	293.2-313.2	33	²⁴
Hesperetin	Methanol, ethanol, 1-butanol, acetone, water, ethyl acetate	288.2-323.2	56	^{25, 26}
Naringin	Water	279.2-348.2	7	²⁷
Epicatechin	Water, ethanol	293.2-303.2	10	¹⁷
Epigallocatechin	Water, ethanol	293.2-303.2	12	¹⁷
Epicatechin gallate	Water, ethanol	293.2-303.2	12	¹⁷
Epigallocatechin gallate	Water, ethanol	293.2-303.2	11	¹⁷
Quercetin	Water	298.6-355.1	4	²⁸
Daidzein	Water	293	1	²⁹
Puerarin	Methanol, acetic acid	293.2-333.2	35	³⁰
Resveratrol	Methanol, ethanol, 1-propanol, 2-propanol, butanol, pentanol, hexanol, ethyl acetate, acetone and water	278-318	163	^{31, 32}

The solubility of a total of 13 phenolic compounds is considered in a total of 11 solvents (pure and mixed). The studied temperature range is from 278 to 355 K and there is a total of 411 data points.

Three zones are distinguishable within the dataset, the first one comprises polyphenols with a relatively high solubility in the solvents they are dissolved in, the second consists of moderately hydrophobic polyphenols in equilibrium with high molar fractions of water and the third one is made by highly hydrophobic polyphenols in equilibrium with water.

3.3.2. *Melting properties*

Experimental activity coefficients can be calculated from the solubility mole fractions by Equation 2 and the melting properties of the solute. Melting points of polyphenols can be found in literature / databases often but melting enthalpies are rarely published for this type of compounds. Therefore, Marrero and Gain model³³ was used for the prediction of the melting properties that were not bibliographically available.

In order to determine the accuracy of the model to the dataset of this work, a comparison was made between the predicted melting temperatures and the published ones. Figure 3 shows a pareto plot for this comparison.

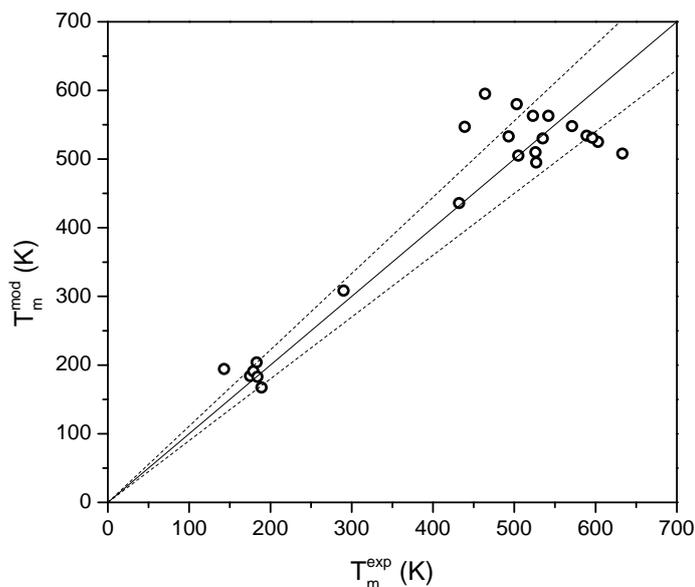


Figure 3. Pareto plot of the prediction of the melting temperature for the studied molecules. Straight line represents 0% error in the prediction, dashed lines represent 10% error.

Average error of the prediction for the melting point is 10.14%, which proves the accuracy of the model for this case. Since in most cases the enthalpy of fusion for these compounds is not known, is expected error is the one reported in the original manuscript 15.7%. Values of the experimental and predicted melting temperatures and enthalpies can be found in the Appendix.

3.3.3. Group definition

Due to the limited amount of data available in literature, there cannot be a large set of groups since that would increase the number of parameters regressed and, as well, decrease their reliability. New groups have to be introduced since Mod-UNIFAC cannot describe the groups cyclic ether or ketone ($-O_{cyc}-$ or $-CO_{cyc}-$ in this work).

The grouping of subgroups was done according the lines on Mod-UNIFAC with the exception of the groups $CH=C_{cyc}$ and $C=C_{cyc}$. Those two subgroups were combined with

the rest of the alkanes in group 1 since having them in a different group did not improve the fit considerably and added 18 more parameters more to regress.

3.3.4. Calculation of size and volume-related parameters

As previously explained in Section 2 of this chapter, Mod-UNIFAC uses three types of parameters: size related, volume related and interaction parameters. In the original UNIFAC the size and volume related parameters (R_k and Q_k) were calculated from the Van der Waals radii of the different groups following the instructions in Bondi³⁴ while the interaction parameters were regressed. One of the changes in Mod-UNIFAC was the regression of R_k and Q_k instead of its calculation. This was possible because of the vast library of equilibria that was used for the regression but due to the limited amount of available data in this case it has been decided to calculate R_k and Q_k from Bondi's method³⁴ as in the original UNIFAC. The calculated values can be found in Table 2.

Table 2: R_k and Q_k parameters.

Group	Subgroup	R_k	Q_k
1	CH ₃	0.9011	0.8480
	CH ₂	0.6744	0.5400
	CH	0.4469	0.2280
	C	0.2195	0.0000
	CH=C _{cyc}	0.8886*	0.6760*
	C=C _{cyc}	0.6605*	0.2440*
2	H ₂ O	0.9200	1.4000
3	OH	1.0000	1.2000
4	aCH	0.5313	0.4000
	aC	0.3652	0.1200
5	aC-OH	0.8952	0.6800

6	-O-	0.3428*	0.2960*
7	-CO-	0.7713*	0.6400*
8	-COO-	1.0020	0.8800
	-COOH	1.3013	1.2240
9	glucose ring	5.5773*	5.0360*

Cyc stands for cyclical and *a* for aromatic. * Parameters calculated in **this work**.

3.3.5. Optimization

The minimizing function is shown in Equation 15.

$$F = \sum_1^N (\ln \gamma_i^{exp} - \ln \gamma_i^{exp})^2 \quad [15]$$

where N is the number of data points. The error of the fit can be calculated directly by Equation 16

$$Error \text{ (log units)} = \frac{1}{N} \left(\sum_1^N (\ln \gamma_i^{exp} - \ln \gamma_i^{exp})^2 \right)^{1/2} \quad [16]$$

The optimization routine was implemented and performed in MATLAB (Version 7.11.0.584 MathWorks, Inc.) using lsqcurvefit with randomly distributed starting values. Regressed parameters were consistent after several runs

3.4. Results

3.4.1. Regression of parameters

The parameters to be regressed are the interaction parameters, namely a_{mn} , b_{nm} and c_{mn} in Equation 14. Due to the limited set of data and the small range of temperatures studied in this work, it has been decided to keep all b_{nm} and c_{mn} parameters at zero and, therefore, presume the interaction between groups to be temperature independent.

A matrix is built with the 70 possible interaction parameters. This number is achieved by considering self-interaction parameters to be zero as in the original and modified UNIFAC. The total number of parameters is (number of groups)² - number of groups

which would yield a total number of 72. However, there are no equilibria within the dataset that represents interaction between the groups 8 and 9 so those two parameters ($a_{8,9}$ and $a_{9,8}$) are left out of the study. Moreover, the interactions of groups 2,3 and 7 with group 8 ($a_{2,8}$, $a_{3,8}$ and $a_{7,8}$ and their symmetric counterparts) have been found not to be present in enough equilibria to fully support the regressed values and therefore, should be recalculated once more data is available. An overview of the recurrence of the different interactions within the dataset can be found in Table 3

The optimization yielded the parameters shown in Table 3. Those parameters have the same order of magnitude as the reported values for the Mod. UNIFAC model ¹⁰.

Table 3: Interaction parameters and number of interactions between groups.

Groups interacting		Number of interactions	a_{nm}	a_{mn}
n	m			
1	2	719	-346.4	162
1	3	936	65.13	-446.7
1	4	2378	1399	-457.4
1	5	1189	1784	-475.4
1	6	783	714.7	532
1	7	729	523.1	-859.8
1	8	100	858.8	-423.7
1	9	232	-466.9	-808
2	3	178	290.8	-436.6
2	4	516	338.5	-133.0
2	5	258	-217.1	-9.355
2	6	151	-1126	-617.8

2	7	158	-87.84	-520.3
2	8	19	1885 (!)	-755.2 (!)
2	9	37	-694.6	-395.8
3	4	600	552.5	-414.1
3	5	300	-529.7	557.8
3	6	206	-366.5	-363.8
3	7	165	-302.4	1058
3	8	19	905.2 (!)	624.9 (!)
3	9	68	-837.7	420.2
4	5	600	1779	1666
4	6	412	1251	866.3
4	7	330	1133	-965.4
4	8	38	245.8	-206.1
4	9	136	-1631	-597.0
5	6	248	1089	1032
5	7	258	-201.2	1225
5	8	27	-415.9	1025
5	9	75	-950.1	526.4
6	7	203	-594.5	-6586
6	8	27	112.9	742.0
6	9	75	213.4	0
7	8	4	137.3 (!)	151.3 (!)
7	9	75	-172.1	-415.4

Parameters with (!) need to be further validated with a bigger set of data

Figure 4 shows a pareto plot with a comparison between experimental activity coefficients and the ones given by the model with the parameters in Table 3. Figure 4 shows the three zones discussed in Section 3.1 of this chapter as well. The fit is good and the error of the fit is 0.0943 log units, which is lower than the average error for the UNIFAC model (0.5 log units³⁵).

Moreover, as previously discussed in literature,³⁶ solubility data is highly dependent on the crystal conformation and in some cases important differences exist between solubility measurements of the same components by different sources³⁷ which emphasizes even more the accuracy of the model.

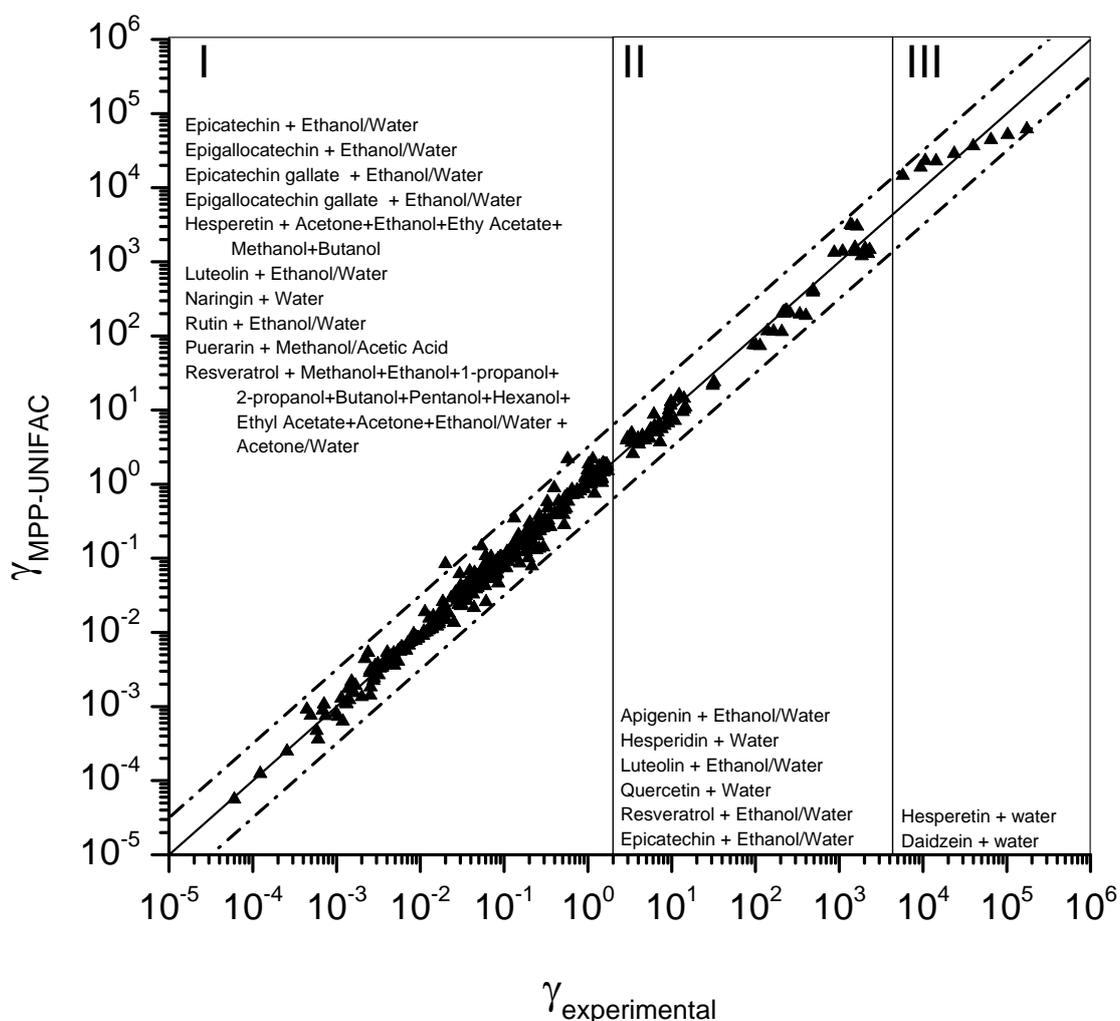


Figure 4: Pareto plot of the performance of the model. The straight line represents the

diagonal on which model and experiments have the same value. Dashed lines represent typical Mod-UNIFAC error of 0.5 log units. The experimental space is divided into three areas and the equilibria present in each is stated in the figure.

3.4.2. Validation

For the validation of the model described in this work, three datasets not previously used for the regression were chosen: genistein in water³⁸ and salicylic acid in ethanol and xylene³⁹. As previously stated in Section 3.1 of this chapter, phenolic acids were not included in the study due to their protonation in the presence of water. However, that protonation does not occur in the presence of ethanol or xylene and therefore, they are eligible for the study.

Figures 5 and 6 show the performance of the model in comparison to experimental values and values predicted with Mod. UNIFAC method. In the case of salicylic acid, the prediction of this work is accurate in both solvents while the one done with Mod. UNIFAC underestimates the solubility in xylene by at least one order of magnitude while MPP-UNIFAC delivered an accurate prediction.

In the case of the solubility of genistein, the molecule cannot be properly defined with the existing groups from Mod. UNIFAC. The closest approximation uses the group -CH₂-CO- defining the molecule with a surplus of two hydrogen atoms. Even though the molecule definition is not the correct one, the prediction delivered by Mod. UNIFAC is accurate. The prediction delivered by MPP-UNIFAC, having a correct representation of groups, is closer to the experimental data.

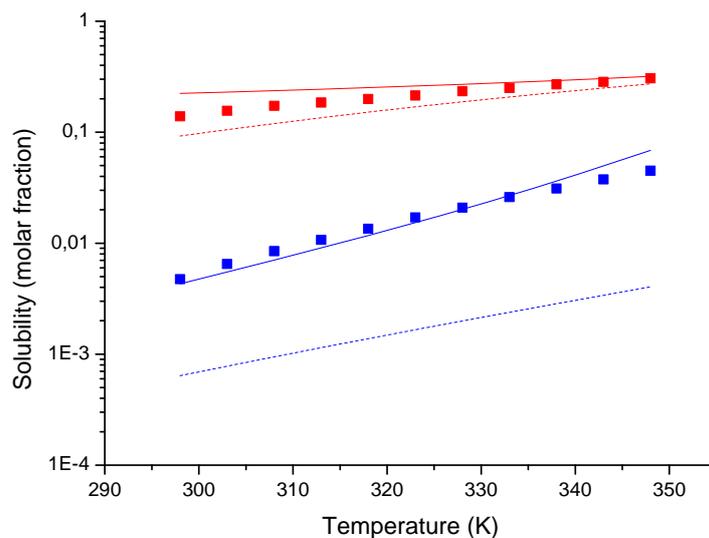


Figure 5: Comparison between experimental solubility of salicylic acid in ethanol (red squares) and in xylene (blue squares), prediction made by MPP-UNIFAC (full lines) and prediction made by mod-UNIFAC (dashed lines).

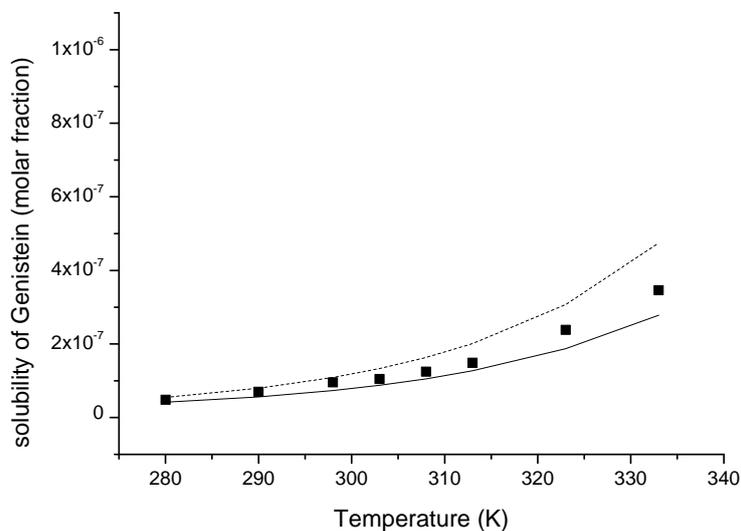


Figure 6: Comparison between experimental solubility of genistein in water (squares), prediction made by MPP-UNIFAC (full lines) and prediction made by mod-UNIFAC (dashed lines).

3.5. Conclusions

An extensive bibliographic search was done to gather solubility data from polyphenols in water or organic solvents. These data were used to regress new interaction parameters applicable for the case of polyphenols in a new model called MPP-UNIFAC. The model proved to fit accurately the data used for the regression with a lower error than the average error that Mod. UNIFAC shows on average (0.5 log units). This method was validated against three experimental datasets from literature showing better predictions in all three cases than Mod. UNIFAC.

Nomenclature

a	aromatic when applied to the definition of groups.
a_{nm}	UNIFAC group interaction parameter between groups n and m (K).
b_{nm}	UNIFAC group interaction parameter between groups n and m .
c_{nm}	UNIFAC group interaction parameter between groups n and m (K^{-1}).
cyc	cyclical when applied to the definition of groups.
F_i	auxiliary property for component i .
q_i	relative van del Waals surface of the component i .
Q_k	relative van der Waals surface area of the group k .
r_i	relative van del Waals volume of the component i .
R	ideal gas constant (J/mol K).
R_k	relative van der Waals volume of the group k .
T	temperature (K).
$T_{m,i}$	temperature of melting of component i (K).
V_i	auxiliary property for component i .
V_i^*	empirically modified V_i -value.

x_i mole fraction in liquid phase.

X_k group mole fraction of group m in the liquid phase in Equation 12 and 13.

Greek letters

γ_i activity coefficient (-).

$\Delta C_{p,i}^{\text{fus,l}}$ difference between the heat capacity of the solid phase and of the liquid phase at equilibrium for component i .

$\Delta H_i^{\text{fus,s}}$ enthalpy of solid-liquid transition of component i (J/mol).

Γ^k group activity coefficient of group k in the mixture.

$\Gamma_{(i)}^k$ group activity coefficient of group k in the pure substance i .

θ_m area fraction for group m in equation 11 and 12.

$\nu_k^{(i)}$ number of structural groups of type k in molecule i .

Ψ_{nm} UNIFAC group interaction parameter between groups n and m in 11 and 14.

Acknowledgements

This work has been carried out within the framework of the Institute for Sustainable Process Technology (ISPT) in the project FO-10-03. The authors want to thank Peter Verheijen for his help in the regression of the parameters.

Appendix

Table A1. Experimental and predicted melting points of the different polyphenols used in this work.

Name	Experimental melting temperature (K)	References	Predicted melting enthalpy (kJ/mol) ³³
Rutin	464	⁴⁰	115.355
Luteolin	603	⁴¹	49.820
Apigenin	633	⁴²	44.811
Hesperetin	505	⁴³	41.828
Naringin	439	⁴⁴	83.199
Epicatechin	526	¹⁷	54.944
Epigallocatechin	493	¹⁷	59.479
Epicatechin gallate	542	¹⁷	84.255
Epigallocatechin gallate	503	¹⁷	89.773
Quercetin	589	⁴⁵	56.731
Hesperidin	535	⁴⁶	63.855
Daidzein	596	⁴⁷	49.057
Puerarin	523	⁴⁸	77.559
Resveratrol	527	⁴⁹	54.952
Genistein	571	⁴⁷	55.536
Salicylic acid	432	⁵⁰	26.592

References

- (1) Ververidis, F.; Trantas, E.; Douglas, C.; Vollmer, G.; Kretzschmar, G.; Panopoulos, N., Biotechnology of flavonoids and other phenylpropanoid-derived natural products. Part I: Chemical diversity, impacts on plant biology and human health. *Biotechnol. J.* **2007**, *2*, 10, 1214.
- (2) Jeon, S.-M.; Kim, H. K.; Kim, H.-J.; Do, G.-M.; Jeong, T.-S.; Park, Y. B.; Choi, M.-S., Hypocholesterolemic and antioxidative effects of naringenin and its two metabolites in high-cholesterol fed rats. *Transl. Res.* **2007**, *149*, 1, 15.
- (3) Williamson-Hughes, P. S.; Flickinger, B. D.; Messina, M. J.; Empie, M. W., Isoflavone supplements containing predominantly genistein reduce hot flash symptoms: A critical review of published studies. *Menopause* **2006**, *13*, 5, 831.
- (4) Verheus, M.; Van Gils, C. H.; Keinan-Boker, L.; Grace, P. B.; Bingham, S. A.; Peeters, P. H. M., Plasma phytoestrogens and subsequent breast cancer risk. *J. Clin. Oncol.* **2007**, *25*, 6, 648.
- (5) Cabrera, C.; Artacho, R.; Giménez, R., Beneficial Effects of Green Tea—A Review. *J. Am. Coll. Nutr.* **2006**, *25*, 2, 79.
- (6) Chen, C.-C.; Crafts, P. A., Correlation and Prediction of Drug Molecule Solubility in Mixed Solvent Systems with the Nonrandom Two-Liquid Segment Activity Coefficient (NRTL–SAC) Model. *Ind. Eng. Chem. Res.* **2006**, *45*, 13, 4816.
- (7) Fredenslund, A.; Gmehling, J.; Michelsen, M. L.; Rasmussen, P.; Prausnitz, J. M., Computerized design of multicomponent distillation columns using the UNIFAC group contribution method for calculation of activity coefficients. *Industrial and Engineering Chemistry Process Design and Development* **1977**, *16*, 4, 450.
- (8) Abrams, D. S.; Prausnitz, J. M., Statistical thermodynamics of liquid mixtures: A new expression for the excess Gibbs energy of partly or completely miscible systems. *AIChE J.* **1975**, *21*, 1, 116.
- (9) Weidlich, U.; Gmehling, J., MODIFIED UNIFAC MODEL. 1. PREDICTION OF VLE, $h^{**}E$, AND γ infinity. *Industrial and Engineering Chemistry Research* **1987**, *26*, 7, 1372.
- (10) Gmehling, J.; Li, J.; Schiller, M., A modified UNIFAC model. 2. Present parameter matrix and results for different thermodynamic properties. *Ind. Eng. Chem. Res.* **1993**, *32*, 1, 178.
- (11) Gmehling, J.; Lohmann, J.; Jakob, A.; Li, J.; Joh, R., A Modified UNIFAC (Dortmund) Model. 3. Revision and Extension. *Ind. Eng. Chem. Res.* **1998**, *37*, 12, 4876.
- (12) Gmehling, J.; Wittig, R.; Lohmann, J.; Joh, R., A Modified UNIFAC (Dortmund) Model. 4. Revision and Extension. *Ind. Eng. Chem. Res.* **2002**, *41*, 6, 1678.
- (13) Jakob, A.; Grensemann, H.; Lohmann, J.; Gmehling, J., Further Development of Modified UNIFAC (Dortmund): Revision and Extension 5. *Ind. Eng. Chem. Res.* **2006**, *45*, 23, 7924.
- (14) Joh, R.; Kreutz, J.; Gmehling, J., Measurement and Prediction of Ternary Solid–Liquid Equilibria. *J. Chem. Eng. Data* **1997**, *42*, 5, 886.
- (15) Domańska, U.; Zawadzki, M.; Królikowski, M.; González, J. A., Phase equilibria and excess molar enthalpies study of the binary systems (pyrrole + hydrocarbon, or an alcohol) and modeling. *Fluid Phase Equilib.* **2014**, *361*, 0, 116.
- (16) Hahnenkamp, I.; Graubner, G.; Gmehling, J., Measurement and prediction of solubilities of active pharmaceutical ingredients. *Int. J. Pharm.* **2010**, *388*, 1–2, 73.

- (17) Méndez Sevillano, D.; Van Der Wielen, L. A. M.; Trifunovic, O.; Ottens, M., Model comparison for the prediction of the solubility of green tea catechins in ethanol/water mixtures. *Industrial and Engineering Chemistry Research* **2013**, *52*, 17, 6039.
- (18) Hector, T.; Gmehling, J., Present status of the modified UNIFAC model for the prediction of phase equilibria and excess enthalpies for systems with ionic liquids. *Fluid Phase Equilib.* **2014**, *371*, 0, 82.
- (19) Zhong, C.; Sato, Y.; Masuoka, H.; Chen, X., Improvement of predictive accuracy of the UNIFAC model for vapor-liquid equilibria of polymer solutions. *Fluid Phase Equilib.* **1996**, *123*, 1-2, 97.
- (20) Neau, S. H.; Flynn, G. L., Solid and Liquid Heat Capacities of n-Alkyl Para-aminobenzoates Near the Melting Point. *Pharm. Res.* **1990**, *7*, 11, 1157.
- (21) Mohs, A.; Gmehling, J., A revised LIQUAC and LIFAC model (LIQUAC*/LIFAC*) for the prediction of properties of electrolyte containing solutions. *Fluid Phase Equilib.* **2013**, *337*, 0, 311.
- (22) Peng, B.; Li, R.; Yan, W., Solubility of Rutin in Ethanol + Water at (273.15 to 323.15) K. *J. Chem. Eng. Data* **2009**, *54*, 4, 1378.
- (23) Peng, B.; Yan, W., Solubility of luteolin in ethanol + water mixed solvents at different temperatures. *J. Chem. Eng. Data* **2010**, *55*, 1, 583.
- (24) Xiao, M.; Yan, W.; Zhang, Z., Solubilities of Apigenin in Ethanol + Water at Different Temperatures. *J. Chem. Eng. Data* **2010**, *55*, 9, 3346.
- (25) Liu, L.; Chen, J., Solubility of Hesperetin in Various Solvents from (288.2 to 323.2) K. *J. Chem. Eng. Data* **2008**, *53*, 7, 1649.
- (26) Ferreira, O.; Pinho, S. P., Solubility of Flavonoids in Pure Solvents. *Ind. Eng. Chem. Res.* **2012**, *51*, 18, 6586.
- (27) Pulley, G. N., Solubility of Naringin in Water. *Industrial & Engineering Chemistry Analytical Edition* **1936**, *8*, 5, 360.
- (28) Srinivas, K.; King, J. W.; Howard, L. R.; Monrad, J. K., Solubility and solution thermodynamic properties of quercetin and quercetin dihydrate in subcritical water. *J. Food Eng.* **2010**, *100*, 2, 208.
- (29) Zhao, C.; Wang, Y.; Su, Y.; Zhang, H.; Ding, L.; Yan, X.; Zhao, D.; Shao, N.; Ye, X.; Cheng, Y., Inclusion complexes of isoflavones with two commercially available dendrimers: Solubility, stability, structures, release behaviors, cytotoxicity, and anti-oxidant activities. *Int. J. Pharm.* **2011**, *421*, 2, 301.
- (30) Wei, D.; Zhang, X., Solubility of puerarin in the binary system of methanol and acetic acid solvent mixtures. *Fluid Phase Equilib.* **2013**, *339*, 0, 67.
- (31) Sun, X.; Peng, B.; Yan, W., Measurement and correlation of solubility of trans-resveratrol in 11 solvents at T=(278.2, 288.2, 298.2, 308.2, and 318.2) K. *The Journal of Chemical Thermodynamics* **2008**, *40*, 4, 735.
- (32) Sun, X.; Shao, Y.; Yan, W., Measurement and Correlation of Solubilities of trans-Resveratrol in Ethanol + Water and Acetone + Water Mixed Solvents at Different Temperatures. *J. Chem. Eng. Data* **2008**, *53*, 11, 2562.
- (33) Marrero, J.; Gani, R., Group-contribution based estimation of pure component properties. *Fluid Phase Equilib.* **2001**, 183-184, 183.
- (34) Bondi, A., Van der waals volumes and radii. *J. Phys. Chem.* **1964**, *68*, 3, 441.
- (35) Kan, A. T.; Tomson, M. B., UNIFAC prediction of aqueous and nonaqueous solubilities of chemicals with environmental interest. *Environ. Sci. Technol.* **1996**, *30*, 4, 1369.

- (36) Klamt, A.; Eckert, F.; Hornig, M.; Beck, M. E.; Bürger, T., Prediction of aqueous solubility of drugs and pesticides with COSMO-RS. *J. Comput. Chem.* **2002**, *23*, 2, 275.
- (37) Srinivas, K.; King, J. W.; Howard, L. R.; Monrad, J. K., Solubility of Gallic Acid, Catechin, and Protocatechuic Acid in Subcritical Water from (298.75 to 415.85) K. *J. Chem. Eng. Data* **2010**, *55*, 9, 3101.
- (38) Wu, J.-G.; Ge, J.; Zhang, Y.-P.; Yu, Y.; Zhang, X.-Y., Solubility of Genistein in Water, Methanol, Ethanol, Propan-2-ol, 1-Butanol, and Ethyl Acetate from (280 to 333) K. *J. Chem. Eng. Data* **2010**, *55*, 11, 5286.
- (39) Shalmashi, A.; Eliassi, A., Solubility of Salicylic Acid in Water, Ethanol, Carbon Tetrachloride, Ethyl Acetate, and Xylene. *J. Chem. Eng. Data* **2007**, *53*, 1, 199.
- (40) Bandyukova, V. A.; Sergeeva, N. V., Rutin in some cultivated plants. *Chem. Nat. Compd.* **1976**, *10*, 4, 535.
- (41) Kaneta, M.; Sugiyama, N., The Constituents of *Arthraxon hispidus* Makino, *Miscanthus tinctorius* Hackel, *Miscanthus sinensis* Anderss, and *Phragmites communis* Trinius. *Bull. Chem. Soc. Jpn.* **1972**, *45*, 2, 528.
- (42) Li, B.; Robinson, D. H.; Birt, D. F., Evaluation of properties of apigenin and [³H]apigenin and analytic method development. *J. Pharm. Sci.* **1997**, *86*, 6, 721.
- (43) Bikiaris, D.; Papageorgiou, G. Z.; Stergiou, A.; Pavlidou, E.; Karavas, E.; Kanaze, F.; Georgarakis, M., Physicochemical studies on solid dispersions of poorly water-soluble drugs: Evaluation of capabilities and limitations of thermal analysis techniques. *Thermochim. Acta* **2005**, *439*, 1–2, 58.
- (44) Castillo, J.; Benavente, O.; Del Río, J. A., Naringin and neohesperidin levels during development of leaves, flower buds, and fruits of citrus aurantium. *Plant Physiol.* **1992**, *99*, 1, 67.
- (45) Ice, C. H.; Wender, S. H., Quercetin and its Glycosides in Leaves of *Vaccinium myrtillus*. *J. Am. Chem. Soc.* **1953**, *75*, 1, 50.
- (46) Scarborough, H., Observations on the nature of vitamin P and the vitamin P potency of certain foodstuffs. *The Biochemical journal* **1945**, *39*, 3, 271.
- (47) Coward, L.; Barnes, N. C.; Setchell, K. D. R.; Barnes, S., Genistein, daidzein, and their .beta.-glycoside conjugates: antitumor isoflavones in soybean foods from American and Asian diets. *J. Agric. Food Chem.* **1993**, *41*, 11, 1961.
- (48) Li, Y.; Yang, D. J.; Zhou, W.; Chen, S. B.; Chen, S. L., Recrystallization of puerarin using the supercritical fluid antisolvent process. *J. Cryst. Growth* **2012**, *340*, 1, 142.
- (49) Roberti, M.; Pizzirani, D.; Simoni, D.; Rondanin, R.; Baruchello, R.; Bonora, C.; Buscemi, F.; Grimaudo, S.; Tolomeo, M., Synthesis and Biological Evaluation of Resveratrol and Analogues as Apoptosis-Inducing Agents. *J. Med. Chem.* **2003**, *46*, 16, 3546.
- (50) Nordström, F. L.; Rasmuson, Å. C., Solubility and Melting Properties of Salicylic Acid. *J. Chem. Eng. Data* **2006**, *51*, 5, 1668.

Chapter 4: Resin Selection for the Separation of Caffeine from Green Tea Catechins

Abstract

This work focuses on the selection of a resin from a defined set of macroporous polymeric resins for the decaffeination of catechins from green tea. High-throughput experimentation and design of experiments is used for retrieving thermodynamic data about the interaction of the different components with the resins. A multicomponent Langmuir isotherm model is used to describe the adsorption and parameters are regressed with high accuracy. These parameters are subsequently used for the definition of criteria to calculate a weighted resin score. The optimal resin is Diaion 20HP with a score of 90.50%, mainly due to its good selectivity for caffeine over catechin (3.021). Based on this good, but not too high, selectivity, a counter-current adsorption process is suggested, like Simulated Moving Bed chromatography, in order to maximize the productivity of the decaffeination of the tea at minimum economic costs.

Published as: Resin Selection for the Separation of Caffeine from Green Tea Catechins.

Sevillano, D. M.; van der Wielen, L. A. M.; Hooshyar, N.; Ottens, M., *Food and Bioproducts Processing* **2014**, 92, (2), 192-198.

4.1. Introduction

Tea is a highly popular beverage, especially in countries like China or India. The beverage is an infusion of the leaves of the plant *Camelia sinensis*. Different types of teas are made by different processing ¹, with green and black tea the most widespread types. The main difference between them is the level of oxidation of the leaves of the plant before the brewing. This oxidation is kept to a minimum in green tea in order to maintain the high concentration of antioxidants in the plant.

An interesting group of antioxidants that can be found in green tea is called catechins (Figure 1). Catechins belong to a group of polyphenols called flavanols and can be found in green tea in percentages from 30 to 42% (dw) ². Even though they are mostly known for their presence in tea, catechins can be found in small quantities in such diverse food sources as coconuts, cocoa, peach or vinegar.

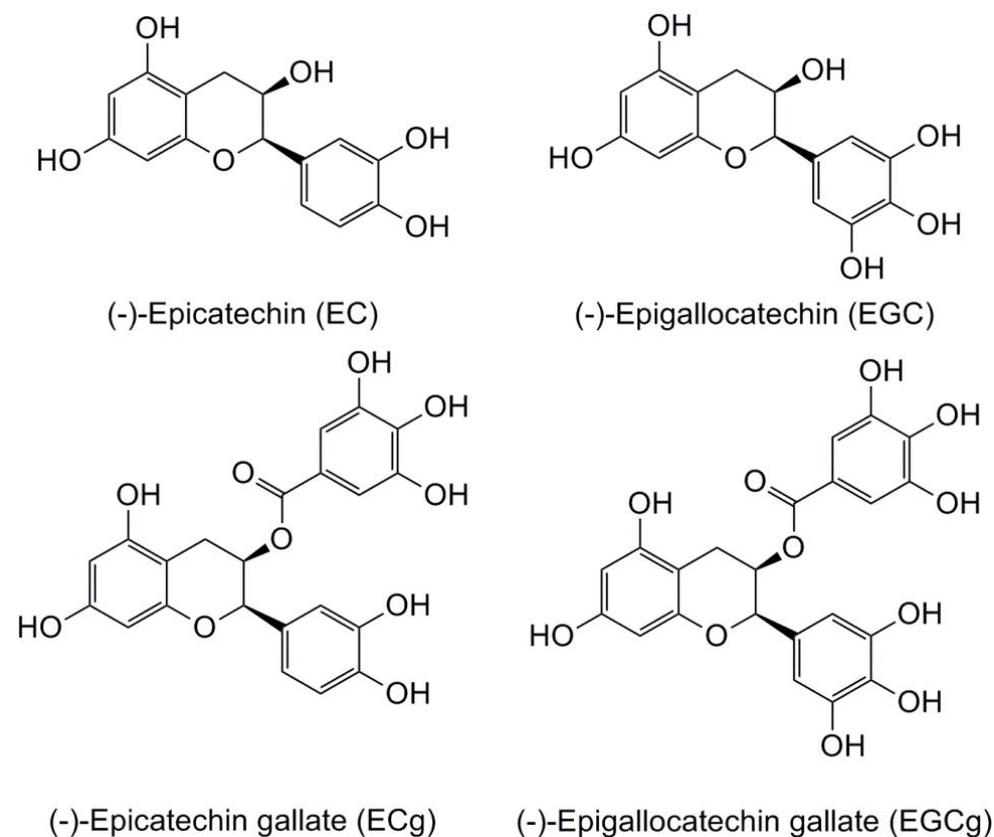


Figure 1. Main components of the family of catechins

Catechins have been related to several health benefits throughout literature. In a recent review ³ the state of the art of clinical evidence on green tea catechins is shown and there seems to be enough evidence to conclude the positive effects of green tea in breast and ovarian cancer. As well, it helps lowering the cholesterol level and it stimulates daily energy expenditure. Liu et al. (2005) ⁴ reported that both black and green tea are able to inhibit the entrance of HIV1 into cells.

Besides polyphenols, green tea contains other components such as proteins, aminoacids, sugars and caffeine. Caffeine is mainly known for its presence in coffee beans and it acts as a stimulant drug. Caffeine has proven to cause several conditions related to the central nervous system such as insomnia ⁵, irritability ⁶ and heart palpitations ⁷ when consumed in large quantities. Due to these effects in the human body it is desirable to decaffeinate both tea beverage and tea extracts.

Decaffeination is a well-known industrial process most used by the coffee industry. The most known methodologies involve supercritical CO₂ extraction (supercritical fluid extraction or scFE) ⁸ or extraction with organic solvents (liquid-liquid extraction or LLE) like chloroform, dichloromethane ⁹, etc. While for the coffee case, the only interest is to reduce the amount of caffeine in either the coffee or the coffee bean, the challenge for tea is different since catechins and caffeine behave in a similar manner. Because of this feature, during the extraction of caffeine, catechins are extracted as well, which decreases the yield of the process. Furthermore scFE and LLE require either the use of costly supercritical equipment for the operation or organic solvents (harmful in some cases).. Solid phase extraction offers some advantages with respect to the discussed methods above. These advantages are mainly the non-toxicity and the ease of recovery (when compared to LLE) and a more economic process (in comparison with scFE). Some work has been done toward the use of macroporous resins for the adsorption of polyphenols and

decaffeination, but a thorough study is lacking for the commercially available food-grade resins.

This work focuses on fast resin selection based on the information given by interaction parameters regressed from a small set of experiments. Adsorption behavior of catechins and caffeine are measured on different commercially available food-grade macroporous resins. Isotherms are shown for the different main components and modelled with a multicomponent isotherm model. The regressed parameters are later used for the calculation of selectivity and capacity and based on that, a resin selection is performed.

4.2. Materials and methods

4.2.1. Materials.

Freeze-dried green tea extract was obtained from UNILEVER. Epicatechin, Epigallocatechin, Epicatechin gallate and Epigallocatechin gallate (EC, ECG, ECg and EGCg respectively) were obtained from Nacalai USA, Inc. with a purity of 98% or higher. Caffeine was purchased from Merck. Methanol, formic acid and acetonitrile (purchased from Sigma-Aldrich) and milli-Q water are used as solvents.

4.2.2. Characterization of the green tea.

A good characterization of the main components is important in order to quantify the different influences in the adsorption. However, in high throughput experimentation analytics tend to be the bottleneck of the experimental process and it is crucial to analyse the key components of the system accurately and in the shortest time possible.

EC, EGC, ECg, EGCg and caffeine were analysed with by UHPLC (Ultimate 3000, Thermo Scientific, USA) based on the protocol described by Zhao et al. (2011)¹⁰ in a C18 column (Acquity UPLC HSS column, 1.8 μ m, 2.1mm x100 mm Waters, Milford, USA). Mobile phase A (0.1% formic acid in milli-Q water) and mobile phase B (0.10% formic

acid in acetonitrile) were run through the column with a constant flow of 0.7 ml/min with the following profile of solvent B: 0 min – 5%, 7 min – 16.5%, 7 min – 45%, 7.5 min – 45%, 7.5 min – 5% and 8.5 min – 5%. Most of the peaks showed baseline separation at a wavelength of 270 nm. An example is shown in Figure 2.

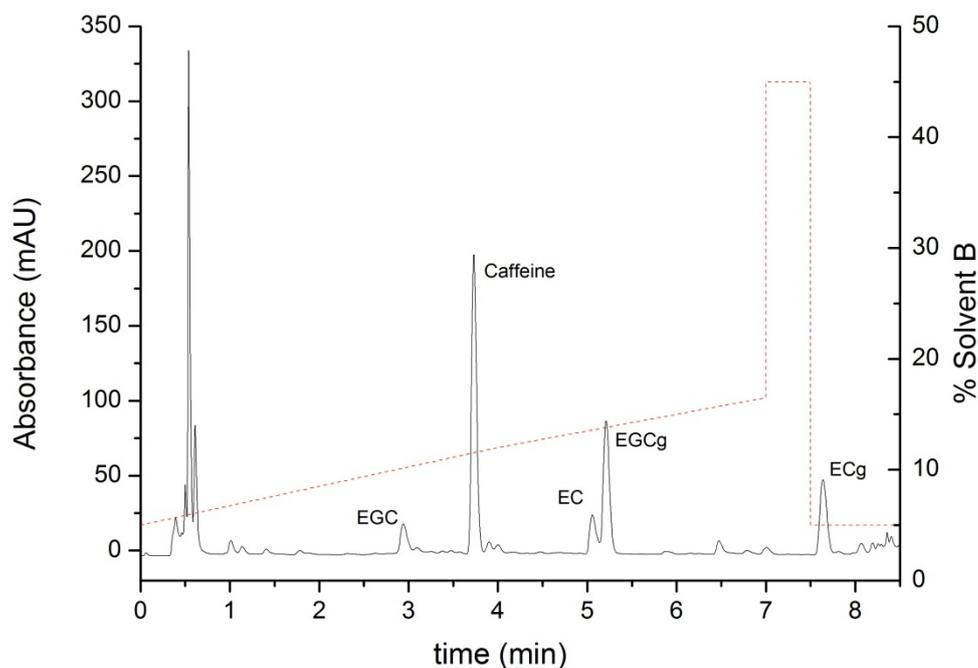


Figure 2: Chromatogram of a green tea sample (solid black line, left axis) and profile of solvent B during the analysis (red dash line, right axis)

In Figure 2 it can be seen that the peaks of EC and EGCg are not fully resolved. Because of this, a deconvolution of the peaks was performed by fitting them to two Gaussian curves.

The concentration of the target components of green tea is as follows (in % dry weight) EC: 1.71, ECG: 7.23, ECg: 3.81, EGCg: 13.29 and Caffeine: 7.29.

4.2.3. Resins.

Seven different food-grade macroporous resins were used. From the AMBERLITE™ XAD™ series, the resins selected were 761, FPX66, 4, 16, 1180N and 7HP while from DIAION™, the resin selected was HP20. FPX66 was purchased from Dow Chemicals while the rest were purchased from Sigma-Aldrich.

4.2.4. Isotherm characterization.

In order to design a process, both kinetics and thermodynamic equilibrium are important. Since diffusion coefficients can be predicted quite precisely with several models in literature¹¹⁻¹⁴, the main experimental challenge lies in measuring adsorption equilibria, or isotherms, as they cannot be easily predicted a priori. Because of the inherent multicomponent nature of the mixture, the multicomponent Langmuir equation was used¹⁵ (Equation 1) for modelling and prediction of the adsorption behavior of the different species.

$$q_i = \frac{Q_{m,i} K_i C_{eq,i}}{1 + \sum K_j C_{eq,j}} \quad [1]$$

$C_{eq,i}$ represents the concentration in the liquid bulk of species i at equilibrium conditions (mg/L), q_i is the load of species i (mg/g_{resin}), $Q_{m,i}$ represents the maximum load (mg/g_{resin}) and K_i the equilibrium constant (L/mg). This equilibrium constant is a measure of the affinity of the component for the resin.

The load of a component is the amount of that component that has been adsorbed onto the resin. Equation 2 is used for its calculation.

$$q_i = \frac{C_{ini,i} - C_{eq,i}}{m_{resin}} V \quad [2]$$

C_{ini} and C_{eq} represent the concentration in the bulk at initial and equilibrium conditions, respectively (mg/L), m_{resin} represents the mass of resin in contact with the bulk (g) and V the volume of solution (L).

Due to the high surface area, complete saturation of the first layer of adsorption is likely not to happen. It is expected then, that the regression of $Q_{m,i}$ will yield parameters with a high uncertainty and consequently, influence the precision of the obtained K_i . Because of that, it was decided to calculate a theoretical maximum capacity and regress only the affinity parameter (K_i).

4.2.5. Maximum capacity determination

In order to theoretically calculate the maximum capacity of a layer of a component on a resin it is necessary to compare the molecular area with the surface area of the resin.

While the web-tool Chemicalize¹⁶ was used for calculating the maximum projection area, the total surface area of the resin has to be determined experimentally. This property is usually given by the manufacturer but as seen in e.g. Ku and Lee (2000)¹⁷ these values are not always accurate so an experimental determination was done by the BET method (see Section 4.2.8).

Equation 3 was used for the calculation of the theoretical capacities.

$$Q_{m,i} = \frac{S_A M_{w,i}}{A_{m,i}} \quad [3]$$

where S_A represents the surface area (m^2/g -resin), $M_{w,i}$ the molecular weight of component i (mg/mol) and $A_{m,i}$ is the molecular area of component i (m^2/mol).

In the calculation of this parameter, several assumptions are made: adsorption is non-specific and in a monolayer. Moreover, molecules are adsorbed occupying the largest area possible and all available space of the resin surface.

4.2.6. Batch uptake experiments

Experiments were done in Multiscreen Filter Plates (Millipore USA) with the aid of the robotic liquid handling system Tecan FreedomEvo 200 (Tecan Group Ltd. Switzerland) in pipetting and washing steps. Resin was applied into the filter plate with the aid of Titan

96 Well Resin Loader (Radleys, UK) and methanol was pipetted in each well for preventing drying of the resin. After two washing steps with methanol and ethanol respectively, an equilibration step with water was performed and then different concentrations of stock solution and pure components solution and mili-Q water are pipetted in the wells so that different concentrations are achieved in each well. Green Tea Stock solution is prepared by dissolving 8% (w/w) of freeze-dried green tea powder in mili-Q water, then centrifuged and filtered through a WHATMAN 0.2 μm filter. Stock solutions of each standard are each 1 g/L pure standard in mili-Q water. Different volumes of green tea stock solution and standard stock solutions are mixed in order to maximize the ratios of catechins and caffeine at which the model is valid. These ratios are shown in Section 4.2.7.

The filter plate is stirred at 50 rpm for 60 min (initial batch uptake experiments showed an equilibration time of approximately 30 min) and its content is centrifuged onto a deep well plate (VWR International, USA) for posterior UPLC analysis.

4.2.7. Design of experiments (DoE)

Prior the experiments described in this work, a set of batch uptake experiments was done with 8% green tea stock solution (as described in Section 4.2.6) in order to calculate a preliminary set of Langmuir parameters.

Once the parameters were regressed, they were used in the same equation (Equation 1) to simulate the same experiment with different ratios of the components in solution. The results of the simulation were used for regressing new parameters and calculate their standard deviation. The ratios of the 12 simulations that yielded the minimum standard deviation were used for the actual experiments and the volumes used for the experiments are shown in Table 1

Table 1. Volumes of stock solutions in each experiment (in μl)

	Green Tea	EC	EGC	ECg	Caffeine
Well 1	240	240	60	0	60
Well 2	160	160	160	0	120
Well 3	240	60	120	0	180
Well 4	140	106	70	142	142
Well 5	185	46	138	46	185
Well 6	50	150	0	200	200
Well 7	86	343	0	0	171
Well 8	200	0	267	0	133
Well 9	55	218	163	0	164
Well 10	185	46	92	92	185
Well 11	0	450	150	0	0
Well 12	109	54	54	220	163

4.2.8. BET method

A small sample of each of the resins is dried in vacuum for 16 h at 373 K. Once the sample is dried, an adsorption experiment is performed using N_2 at 77K until saturation is reached and then the calculation of the surface area follows from the saturation of the resin.

4.2.9. Resin selection procedure

The goal of this work is to investigate the difference in affinity of catechins and caffeine to industrial macroporous adsorption resins, and choose an optimal resin for a tea decaffeination process. In order to make a quick resin screen, it is important that both the experimentation, modelling and selection time should be kept to the minimum.

Modelling time can be easily reduced by choosing a simple model to describe the isotherms like the Multicomponent Langmuir (Equation 1). The parameters regressed by this model (K_i and $Q_{m,i}$) give us information about the behavior of the tea components with the selected resins.

A correct decision on which criteria will be considered for the selection is just as important as the modelling itself. Luckily, the model provides us with meaningful parameters that can be used for having a numerical insight into the performance of the resin.

Based on this specific system, it would be preferred to have a fractionation of the tea into different fractions: a caffeine-rich one, and one or more catechin-rich fraction. A possible way in which resins can be evaluated is by comparison of the selectivity or the capacity of the resin. Both magnitudes will be accounted for according to Equations 4-7.

$$S_{cat,caff} = \frac{\prod_{cat} \sqrt[4]{K_i}}{K_{caff}} \quad [4]$$

$$S_{caff,cat} = \frac{K_{caff}}{\prod_{cat} \sqrt[4]{K_i}} \quad [5]$$

$$Cap_{cat} = \frac{\sum_{cat} Q_{m,i}}{N_{cat}} \quad [6]$$

$$Cap_{caff} = Q_{m,caff} \quad [7]$$

S means selectivity, Cap represents the capacity, N_{cat} is the total number of catechins, Σ represent the sum and Π the product.

Equations 4 and 5 describe the selectivity factor of the catechins over caffeine and caffeine over catechins, respectively. The selectivity of the catechins is calculated by the geometrical mean of their affinity values. Equation 6 describes the average capacity of the resin for catechins while Equation 7 describes the capacity of the resin for caffeine. In this case the capacity of the catechins has been calculated by the arithmetical mean of the

capacities.

The selectivity from Equations 4 and 5 is mutually exclusive. If the selectivity of catechins over caffeine (Equation 4) is lower than unity, the selectivity of caffeine over catechins (Equation 5) is taken into account and the same otherwise. The capacity follows the same rule, if the selectivity of the resin for catechins is higher than for caffeine, the capacity for catechins (Equation 6) is used and the same case for caffeine.

Since saturation of the resin was not accomplished in any of the experiments, the values of $Q_{m,i}$ are calculated theoretically (as shown in Section 4.2.5) and therefore, taken less into consideration.

Following this reasoning, the weighing factors for each of the influences are 0.8 for selectivity and 0.2 for capacity.

Since Equations 4-7 are not normalized, a normalization has to be done using Equation 8 in order for it to be used in the resin selection.

$$Resin\ Score = \sum weight * \frac{criterion}{maximum\ value\ of\ the\ criterion} \quad [8]$$

An schematic block diagram that explains the entire resin selection methodology can be found in Figure 3.

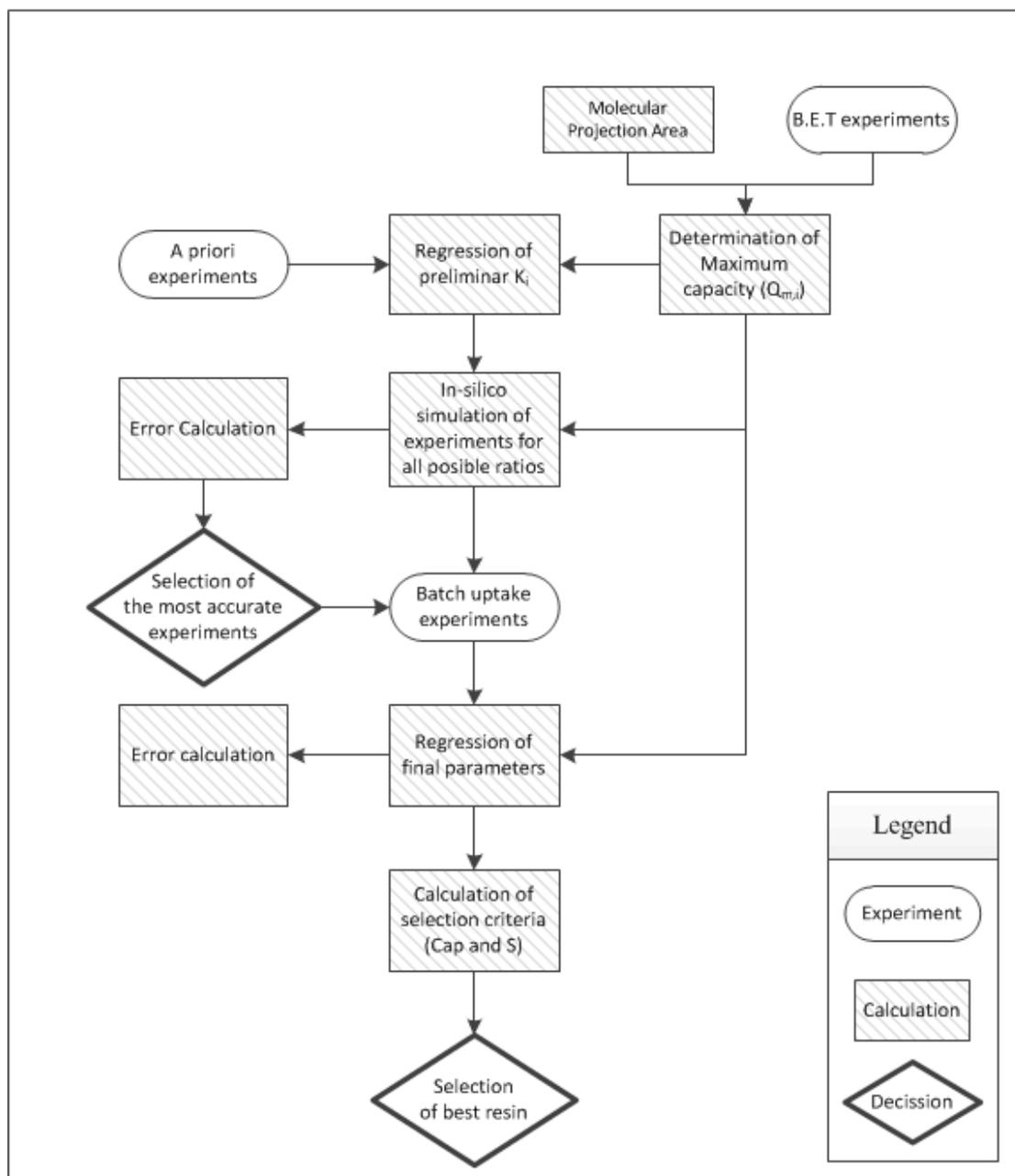


Figure 3: Schematic block diagram of the resin selection methodology used.

4.2.10. Error Calculation

Two types of errors have been taken into account in this work: errors related to measurements and errors related to regressed parameters.

For the calculation of the error due to the measurements the error propagation Equation 9 has been used.¹⁸

$$\sigma_U = \sqrt{\sum_i \left(\frac{\partial U}{\partial i}\right)^2 \sigma_i^2} \quad [9]$$

where σ represents standard deviation, U is the property that is being calculated and i is one of the experimental measurements that this property is calculated from.

The errors related to regressed parameters were taken from the covariance matrix calculated from the Jacobian given by the fitting function.

Computer software MATLAB (Version 7.11.0.584 MathWorks, Inc.) was used for modelling and regression.

4.3. Results and discussion

4.3.1. Determination of maximum capacity.

The values obtained from the BET experiment are shown in Table 2.

Table 2: Surface area of the selected resins

	XAD-4	XAD-7HP	XAD-761	XAD-16	Diaion 20HP	XAD-1180N	FPX66
Experimental surface area (m ² /g)	862	405	232	877	618	597	914
Manufacturer surface area (m ² /g)	725	450	200	800	600	600	700

As it can be seen in Table 2, most of the values are close to the values specified by the manufacturer. However, FPX66 appears to have a much higher surface area compared to the provided information from the manufacturer. Regardless this information, all of them present a very high surface area. From these values, the maximum capacities can be easily calculated as previously explained in Section 4.2.7, following Equation 3. The resultant values can be found in Table 3.

Table 3: Q_m parameter from the multicomponent Langmuir equation for different components and resins

Q_m (mg/g- resin)	XAD-4	XAD- 7HP	XAD- 761	XAD- 16	Diaion 20HP	XAD- 1180N	FPX66
EC	538.5	253.0	144.9	547.9	386.1	373.0	571.0
EGC	537.6	252.6	144.7	547.0	385.4	372.3	570.0
ECg	642.5	301.9	172.9	653.7	460.6	445.0	681.2
EGCg	653.2	306.9	175.8	664.6	468.3	452.4	692.6
Caffeine	460.2	216.2	123.9	468.2	329.9	318.7	487.0

4.3.2 Regression of the affinity parameters.

From the batch uptake experiments described in Section 4.2.4, Equation 1 was used for the modelling of the data. The fitting yielded an affinity parameter per combination of component-resin and a measure of the uncertainty of this parameter. All of them are listed in Table 4

Table 4: K parameter and its standard deviation from the multicomponent Langmuir equation for different components and resins

K (l/g)	XAD-4	XAD-7HP	XAD-761	XAD-16	Diaion 20HP	XAD-1180N	FPX66
EC	0.0212 ± 0.0062	0.0937 ± 0.0239	0.4685 ± 0.0353	0.0324 ± 0.0071	0.0747 ± 0.0220	0.1587 ± 0.0158	0.0600 ± 0.0058
EGC	0.0162 ± 0.0021	0.0706 ± 0.0073	0.3103 ± 0.0960	0.0145 ± 0.0015	0.0446 ± 0.0042	0.1001 ± 0.0032	0.0431 ± 0.0014
ECg	0.0134 ± 0.0114	0.2110 ± 0.1223	1.4282 ± 0.0883	0.1252 ± 0.0169	0.3653 ± 0.0590	0.7057 ± 0.0507	0.1687 ± 0.0117
EGCg	0.0126 ± 0.0018	0.1462 ± 0.0140	0.3673 ± 0.0114	0.0251 ± 0.0018	0.0768 ± 0.0057	0.1488 ± 0.0048	0.0460 ± 0.0015
Caffeine	0.0299 ± 0.0035	0.0779 ± 0.0091	1.4178 ± 0.0405	0.0943 ± 0.0055	0.2971 ± 0.0208	0.4901 ± 0.0148	0.1242 ± 0.0037

A display of the goodness of fit of the model can be seen in Figure 4.

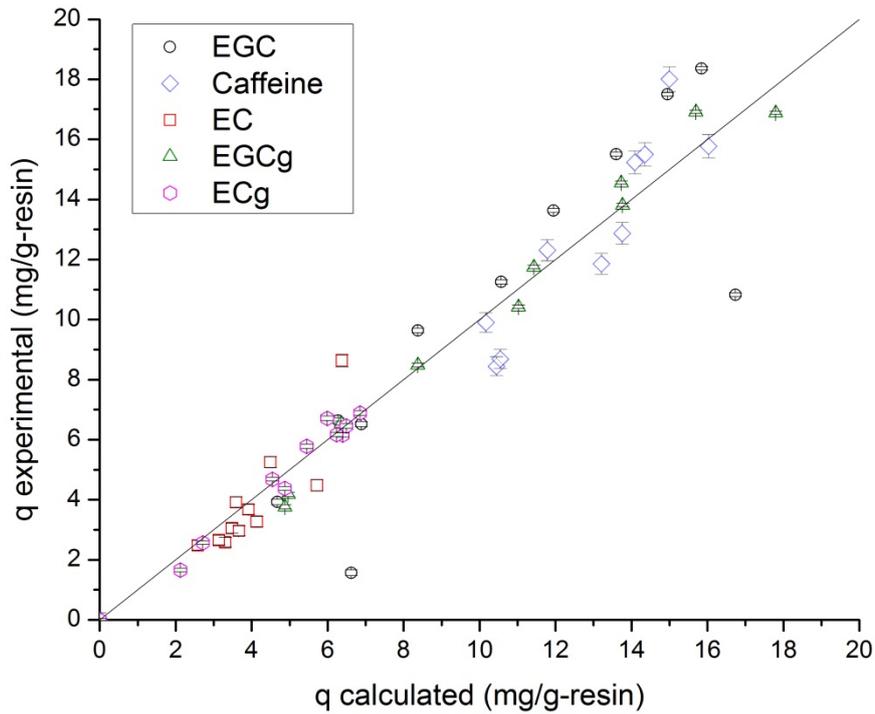


Figure 4: Experimental loading is plotted against calculated one with the Multicomponent Langmuir model for XAD 761, one of the 7 resins. Dashed line, the case of $q_{\text{calculated}}=q_{\text{experimental}}$.

It can be seen that the model fits the data points with great accuracy, proving the validity of the parameters for the experimental range investigated. Error bars display the magnitude of the experimental error calculated with Equation 9.

4.3.3. Criteria, resin scores and comparison.

Following the reasoning outlined in Section 4.2.9 for resin selection, a numerical value is calculated as a measure of the suitability of the resin for each of the criteria. These values are shown in Table 5.

Table 5: Values for the different criteria in each resin.

	XAD-4	XAD-7HP	XAD-761	XAD-16	Diaion 20HP	XAD-1180N	FPX66
$S_{(cat,caff)}$	0.519	1.534	0.371	0.370	0.330	0.412	0.539
$S_{(caf,cat)}$	1.927	0.652	2.698	2.705	3.021	2.425	1.856
C_{cat}	592.9	278.6	159.6	603.3	425.1	410.7	628.7
C_{caff}	460.2	216.2	123.9	468.2	329.9	318.7	487.0

The resin score (according to Equation 8) can be seen in Figure 5.

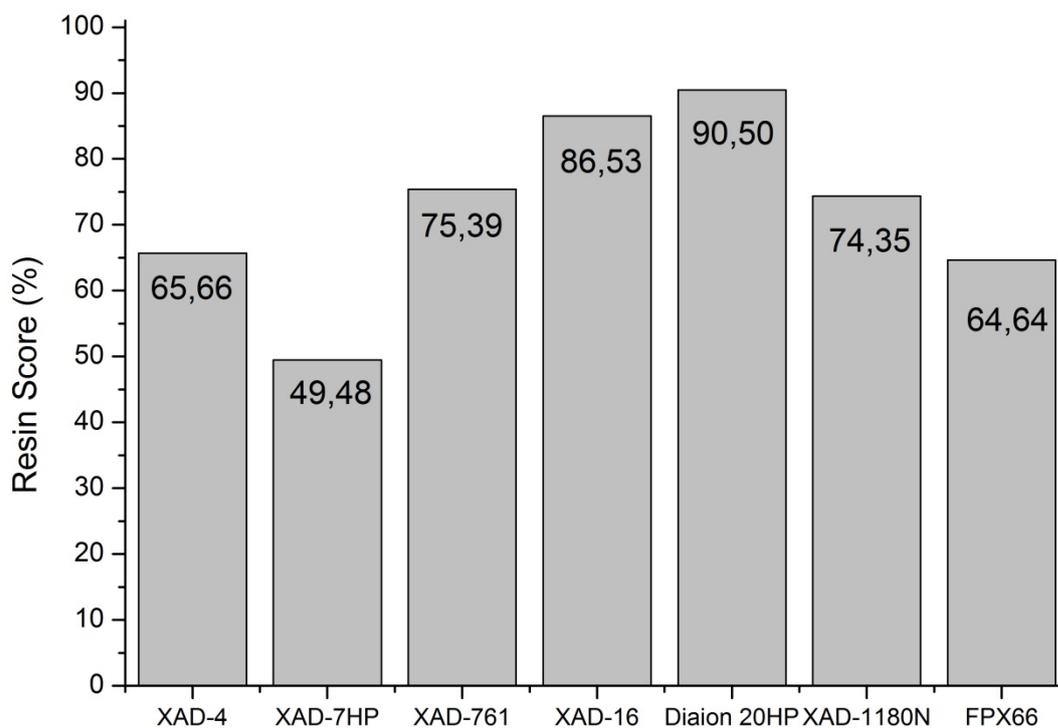


Figure 5: Weighted resin score for each of the resins (highest indicates best).

The highest ranked resin, as seen in Figure 5, is Diaion 20HP with a weighted resin score of 90.50 %. All resins, but one (XAD-7HP), show a higher affinity for caffeine than for catechins. In the case of XAD-7HP the selectivity (and hence, resin score) is not that high in comparison with the rest.

The aim of this work is to present a new methodology for a quick resin selection. The methodology is exemplified with the separation of catechins and caffeine in the green tea. Having an optimal resin selected, brings us to the next step of large scale process development. Different modes of operation can be used, i.e . the Batch Chromatography mode or counter-current Simulated Moving Bed chromatography mode.¹⁹ The latter can be used to improve the economics of the separations process, by reducing eluent use and resin inventory at the same productivity.

Simulated moving bed has proven to be applicable to similar systems like separation of oligosaccharides,²⁰ amino acids,²¹ or even proteins²² and it would be a suitable technique for purification and decaffeination of catechins from green tea at large scale. However, the detailed design and economical evaluation of the large scale process lies beyond the scope of the present study and is material for a future publication.

4.4. Conclusion

This study shows a methodology for selection of resins based on their thermodynamic properties. Optimized batch uptake experiments designed by DoE were performed using high-throughput equipment in order to minimize both the experimental and analytical time. Isotherms were measured and modelled using the multicomponent Langmuir equation with an good fit.

Accurate parameters were regressed for all four catechins and caffeine. The parameters were studied and compared using different criteria. These criteria were analysed and

assigned a weight factor as a function of its importance. From the addition of the different normalized and weighted criteria, a resin score was calculated. The resin Diaion 20HP was chosen based on their similarly high resin score (90.50%) in comparison to other commercially available macroporous resins.

The criterion that influenced the score of this resin the most was its selectivity of catechins over caffeine.

Acknowledgements

This work was supported by the ISPT (Institute for Sustainable Process Technology) under the project FO-10-03 Separation of vitality ingredients. Thanks, as well, to the industrial partner UNILEVER for financial support and valuable comments.

References

- (1) Cabrera, C.; Artacho, R.; Giménez, R., Beneficial Effects of Green Tea—A Review. *J. Am. Coll. Nutr.* **2006**, *25*, 2, 79.
- (2) Graham, H. N., Green tea composition, consumption, and polyphenol chemistry. *Prev. Med.* **1992**, *21*, 3, 334.
- (3) Johnson, R.; Bryant, S.; Huntley, A. L., Green tea and green tea catechin extracts: An overview of the clinical evidence. *Maturitas* **2012**, *73*, 4, 280.
- (4) Liu, S.; Lu, H.; Zhao, Q.; He, Y.; Niu, J.; Debnath, A. K.; Wu, S.; Jiang, S., Theaflavin derivatives in black tea and catechin derivatives in green tea inhibit HIV-1 entry by targeting gp41. *Biochimica et Biophysica Acta (BBA) - General Subjects* **2005**, *1723*, 1–3, 270.
- (5) Karacan, I.; Thornby, J. I.; Anch, A. M.; Booth, G. H.; Williams, R. L.; Salis, P. J., Dose-related sleep disturbances induced by coffee and caffeine. *Clin. Pharmacol. Ther.* **1976**, *20*, 6, 682.
- (6) Darragh, A.; Kenny, M.; Lambe, R. F.; O'Kelly, D. A., Adverse effects of caffeine. *Ir. J. Med. Sci.* **1981**, *150*, 2, 47.
- (7) Greden, J. F., Anxiety or caffeinism: a diagnostic dilemma. *Am. J. Psychiatry* **1974**, *131*, 10, 108.
- (8) Sun, Q. L.; Hua, S.; Ye, J. H.; Lu, J. L.; Zheng, X. Q.; Liang, Y. R., Decaffeination of green tea by supercritical carbon dioxide. *Journal of Medicinal Plants Research* **2010**, *4*, 12, 1161.
- (9) Kazi, T. A method for the production of decaffeinated tea. EP0050482 (B1), 1984-07-11, 1984.

- (10) Zhao, Y.; Chen, P.; Lin, L.; Harnly, J. M.; Yu, L.; Li, Z., Tentative identification, quantitation, and principal component analysis of green pu-erh, green, and white teas using UPLC/DAD/MS. *Food Chem.* **2011**, 126, 3, 1269.
- (11) Einstein, A., Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Annalen der Physik* **1905**, 322, 8, 11.
- (12) Wilke, C. R.; Chang, P., Correlation of diffusion coefficients in dilute solutions. *AIChE J.* **1955**, 1, 2, 264.
- (13) Tyn, M. T.; Calus, W. F., Diffusion coefficients in dilute binary liquid mixtures. *J. Chem. Eng. Data* **1975**, 20, 1, 106.
- (14) Nakanishi, K.; Furusawa, Y., Diffusion in mixed solvents. 4. Iodine in polar nonassociated solutions. *J. Chem. Eng. Data* **1978**, 23, 2, 101.
- (15) Markham, E. C.; Benton, A. F., THE ADSORPTION OF GAS MIXTURES BY SILICA. *J. Am. Chem. Soc.* **1931**, 53, 2, 497.
- (16) <http://www.chemicalize.org> (27-08-2014),
- (17) Ku, Y.; Lee, K.-C., Removal of phenols from aqueous solution by XAD-4 resin. *J. Hazard. Mater.* **2000**, 80, 1–3, 59.
- (18) Worthing, A. G. G., Joseph, Treatment of Experimental Data. *J. Chem. Educ.* **1943**, 20, 11, 572.
- (19) Ruthven, D. M.; Ching, C. B., Counter-current and simulated counter-current adsorption separation processes. *Chem. Eng. Sci.* **1989**, 44, 5, 1011.
- (20) Vaňková, K.; Polakovič, M., Design of Fructooligosaccharide Separation Using Simulated Moving-Bed Chromatography. *Chem. Eng. Technol.* **2012**, 35, 1, 161.
- (21) Makart, S.; Bechtold, M.; Panke, S., Separation of amino acids by simulated moving bed under solvent constrained conditions for the integration of continuous chromatography and biotransformation. *Chem. Eng. Sci.* **2008**, 63, 21, 5347.
- (22) Freydell, E. J.; Bultink, Y.; van Hateren, S.; van der Wielen, L.; Eppink, M.; Ottens, M., Size-exclusion simulated moving bed chromatographic protein refolding. *Chem. Eng. Sci.* **2010**, 65, 16, 4701.

Chapter 5: Mechanism of Isoflavone Adsorption from Okara extracts onto Food-Grade Resins

Abstract

Okara is a by-product of the soy-milk industry containing valuable phytochemicals, called isoflavones, amongst other components (i.e. proteins, sugars, fibers, etc.). As a waste product, okara is an interesting source material for obtaining valuable chemicals and knowledge of the behavior of such components in their complex matrix is a key step for the design of a purification process. Six commercially available macroporous polymeric resins are investigated to measure and model the equilibrium properties of the adsorption of isoflavones, proteins and total solids onto these resins. A new model is evaluated in which adsorption of isoflavones onto a protein layer is proposed describing the system isoflavones-resin XAD 4 better than a linear isotherm model. The parameters for both the linear model and the bilayer model are regressed and reported with their accuracy and correlated to the hydrophobicity of each of the isoflavones.

Published as: Mechanism of Isoflavone Adsorption from Okara Extracts onto Food-Grade Resins. Méndez Sevillano, D.; Jankowiak, L.; van Gaalen, T. L. T.; van der Wielen, L. A. M.; Hooshyar, N.; van der Goot, A.-J.; Ottens, M., *Ind. Eng. Chem. Res.* **2014**, 53, (39), 15245-15252.

5.1. Introduction

Okara is the by-product of the soy milk production. Typically, the use of 1 kg of soybeans produces 1.5 kg of this by-product with a very high moisture content around 80%¹. Soybeans contain a net of insoluble fibers, which remains after the soymilk extraction and becomes, as well, a major component of the okara. In addition to insoluble fibers, proteins and lipids²⁻⁴, okara contains smaller molecules like sugars or polyphenols^{1,4} as well. Amongst these small molecules, there is a class of molecules of special interest because of its health related properties: the isoflavones. These molecules have shown to exhibit estrogen-like activity⁵ which could be beneficial for hormone regulation. Isoflavones belong to the flavonoid family of the polyphenols, mainly found in soy and other plants of the *fabaceae* family⁶. There are three types of isoflavones (Daidzin, Genistin and Glycitin) and within each type, there are four possible configurations (aglycones, glucosides, acetyl -esters and malonyl-esters). These 12 types of isoflavones are presented in Figure 1. The isolation of these components from the by-product okara is therefore of interest.

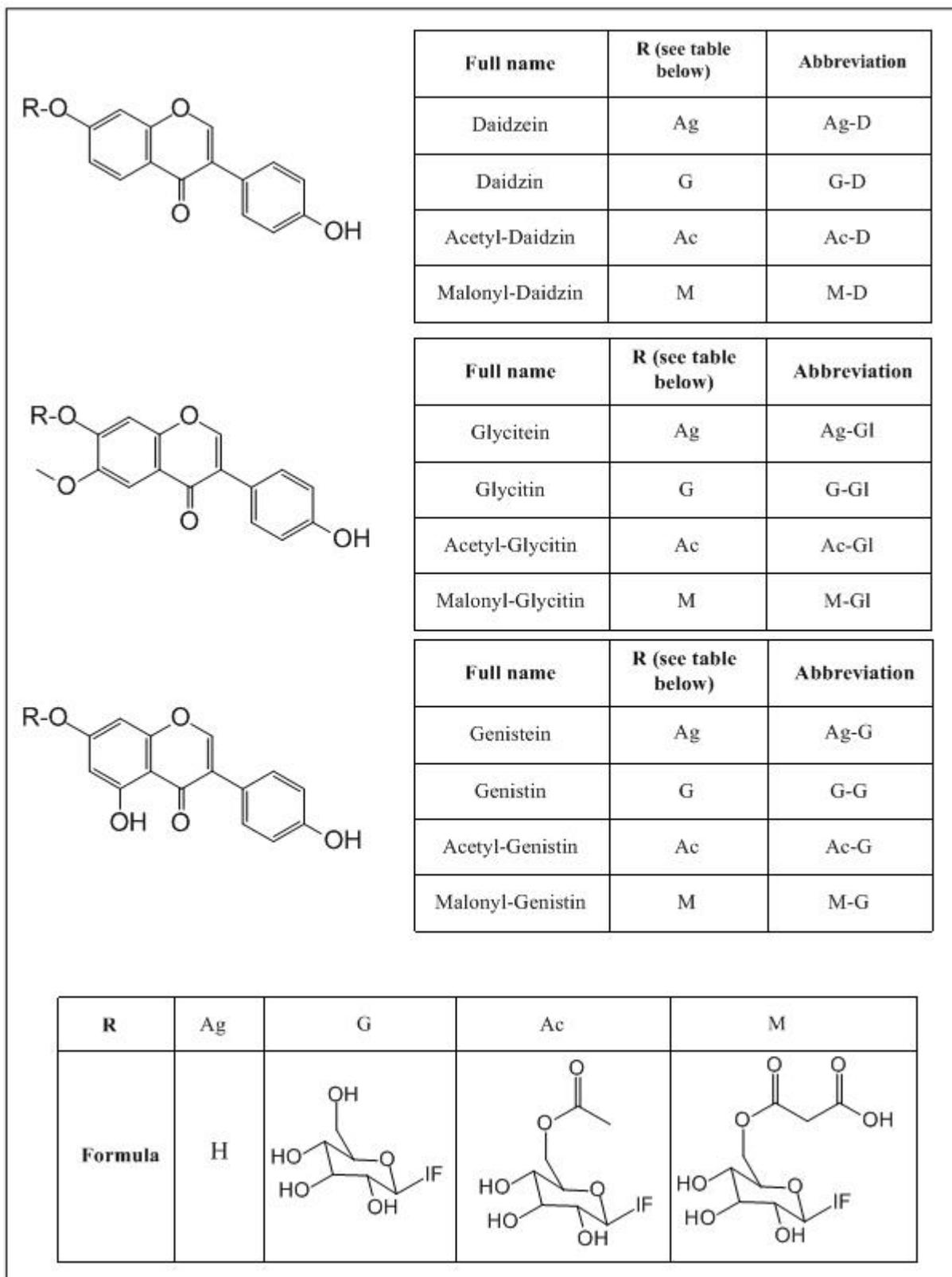


Figure 1: Isoflavones present in soy.

The first step towards purification of isoflavones is to bring them in solution. Previous work has shown that isoflavones are soluble, and therefore can be extracted using

different solvents such as acetonitrile, genapol ® X-080, methanol and ethanol ^{1, 7-10}. Furthermore, latest experiments with water at different pH and increased temperatures have shown promising results ¹¹. A particular challenge of the extraction process with co-solvents is the high water content of okara which will sharply increase the usage of co-solvent if a high co-solvent fraction is needed. As well, the end product should be suitable for human consumption. Therefore, it is preferred to use only food-grade solvents: water and ethanol.

Based on the aromatic backbone of the molecules the next logical separation would be one based on hydrophobicity. Previous work shows the applicability of macroporous polymeric resins to other polyphenols ¹² or to isoflavones from other sources than okara ^{7-9, 13} obtaining purities up to 58%. A resin can be packed in a column facilitating the adsorption and elution without the need of a further solid-liquid separation method. In literature, the performance of resins for certain compounds is investigated through trial-and-error approach and reported. However, optimization of a process using thermodynamic models and a deeper understanding of the thermodynamics of adsorption would result in an optimal process and process understanding. In other cases in literature, the adsorption has been investigated with pure compounds, which is not enough for understanding the different interactions that will take place in the complex system of choice.

This paper focuses on the adsorption step of the purification process of the isoflavones from an okara extract. This study considers the understanding and modelling of the thermodynamic behavior of the different molecules, or groups of molecules, with a total of 6 food-grade resins of choice. The hypothesis of bilayer adsorption is tested in all different scenarios and compared to linear isotherm

behavior. All isotherms have been modelled and the regressed parameters are given with a proper error analysis of their values giving their accuracy.

5.2. Materials and methods

5.2.1. Chemicals.

Isoflavone standards for the UPLC protocol were obtained from Nacalai USA, Inc. with a purity of 98% or higher. Methanol, formic acid and acetonitrile (purchased from Sigma-Aldrich) and mili-Q water were used as solvents. Six different food-grade macroporous resins were used. From the AMBERLITE™ XAD™ series, the resins selected were 761, 4, 16, 1180N and 7HP, while from DIAION™, the resin selected was HP20. All of them were purchased from Sigma-Aldrich.

5.2.2. Okara production.

An ASC50 soymilk system (ProSoya Inc., Ottawa, Canada) was used to produce approximately 3kg okara with a moisture content of 79% by mixing soybeans and water in a ratio of 1:7. The slurry was heated at 105°C for 3 minutes before separation of solids (okara) and liquid (soymilk) by a filter centrifuge¹.

5.2.3. Okara extract production

Clear okara extracts were produced at alkaline pH (10) with subsequent precipitation of the proteins at pH 4.5. 40 g of crude okara was mixed with deionized water to achieve a liquid-to-solid ratio of 20 to 1, including the water present in the okara. The pH was adjusted with 1M NaOH to pH 10. The mixture was placed in a shaking water bath at 50°C for 1h. After extraction of the soluble compounds under the described conditions, the liquid was separated from the fibrous residue by means of an Amicon stirred cell 8400 (Millipore, Massachusetts, USA). A Whatman filter paper no. 1 was placed at the bottom of the Amicon equipment and air pressure was applied. The samples were filtrated until no more extract passed the filter. The extract was

collected in a tube and the pH adjusted with 1M HCl to pH 4.5. The mixture was left standing for 30min for particles to settle before centrifugation at 18.000 x g for 30min with a Sorvall Legend XFR centrifuge (Thermo Scientific, USA). The clear extract was separated from the protein pellet by decanting it through a Buchner funnel. The clarified extract containing the isoflavones was used for adsorption/desorption experiments.

5.2.4. Isoflavones quantification

Isoflavones were analysed with by UHPLC (Ultimate 3000, Thermo Scientific, USA) based on the protocol described in literature ¹⁴ in a C18 column (Acquity UPLC BEH column, 1.7µm, 2.1x50 mm Waters, Milford, USA). An example of a chromatogram is shown in Figure 2.

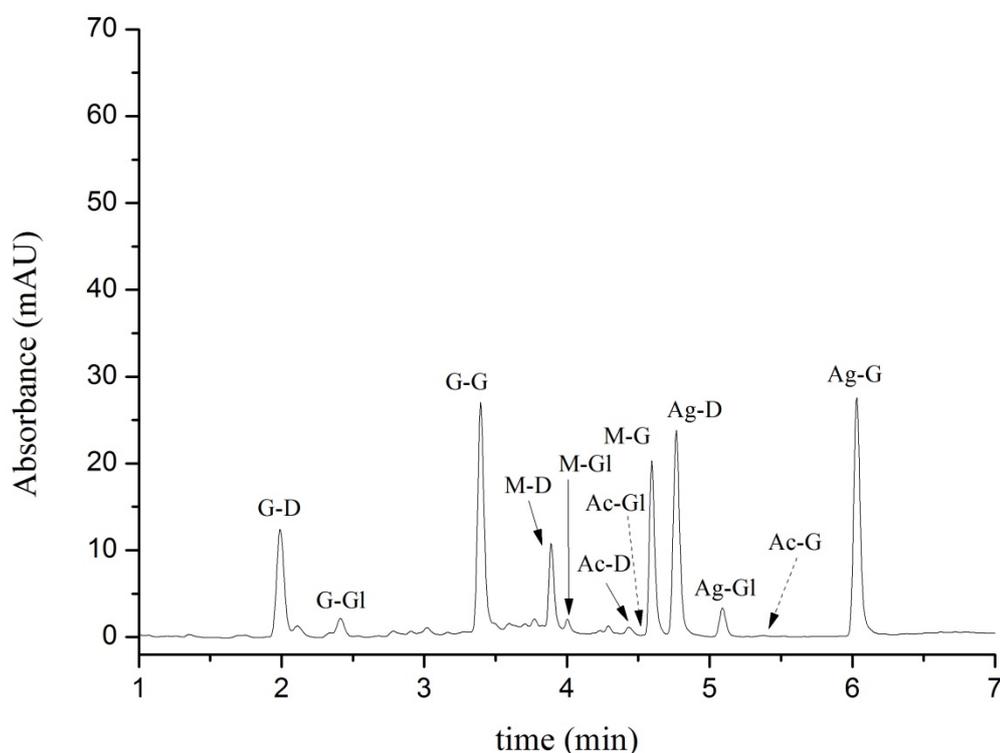


Figure 2: Chromatogram of an extract of okara using abbreviations given in Figure 1. Dashed arrows indicate isoflavones not identified in this study (Ac-Gl and Ac-G).

5.2.5. Protein quantification

A Coomassie (Bradford) Protein Assay Kit was used for the quantification of proteins¹⁵. Samples were compared to a BSA calibration line for quantification.

5.2.6. Dry mass determination

Dry mass determination was done by weight difference. 10 ml of sample was taken to dryness in an oven at 40° C. The measurements of the dry mass yielded the solid content in solution. The amount of unquantifiable solids was calculated by subtraction of the known amount of proteins and isoflavones from the solid content.

5.2.7. Batch uptake experiments

Batch uptake experiments were carried out as follows. A known amount of resin (around 0.02 g) was mixed in a 96 deep well plate with 1 ml of okara extract at different dilutions. The resin was previously washed with methanol and equilibrated with mili-Q water. The deep-well plate was left stirring overnight at 600 rpm to achieve equilibrium. Samples of the starting solutions as well as the equilibrium solutions were taken and analysed by both, the UHPLC and the Bradford method. The adsorbed amount was calculated from the measurements according to Equation 1.

$$Q_i = \frac{(C_{i, \text{initial}} - C_{i, \text{equilibrium}}) V}{m_{\text{resin}}} \quad [1]$$

Where Q_i represents the concentration of component i in the solid (mg/g-resin), C_i is the concentration of the component i in the initial (or equilibrium) solution (mg/L), m_{resin} , the mass of resin material (g) and V is the for volume of the solution in the well (L). Elution with 75% (w/w) ethanol/water was done to prove the applicability of Equation 1. Mass balances closed within the experimental error.

5.2.8. Error Calculation

Two types of errors have been taken into account in this work: errors related to measurements, and errors related to regressed parameters.

For the calculation of the error due to the measurements the error propagation

Equation 2 has been used.¹⁶

$$\sigma_U = \sqrt{\left(\frac{\partial U}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial U}{\partial y}\right)^2 \sigma_y^2 + \dots} \quad [2]$$

Where σ represents the standard deviation, U is the property that is being calculated, and x, y, etc. represent the different experimental measurements that this property is calculated from.

The errors related to regressed parameters were taken from the covariance matrix calculated from the Jacobian given by the fitting function (lsqcurvefit).

The computer software MATLAB (Version 7.11.0.584 MathWorks, Inc.) was used for modelling and regression.

5.3. Results and discussion

5.3.1. Loading experiments

Table 1 shows the initial concentrations of the major isoflavones present in the okara extract. In soy, 12 isoflavones have been identified (Fig. 1). However, both Ac-GI and Ac-G were not identified in the okara extract, which is in line with previous research¹⁷⁻¹⁹. Ac-D, Ag-GI, and M-GI were present in such low concentrations that we chose to concentrate our study on the main isoflavones present in the extract as listed in Table 1. The purity of those seven isoflavones together in the mixture was 0.51% and will be named from hereon as total isoflavones.

Table 1: Composition of okara extract

Compound	Concentration in the extract (mg/L)
<i>G-D</i>	6.191
<i>G-GI</i>	1.325
<i>G-G</i>	4.969
<i>M-D</i>	4.225
<i>M-G</i>	5.641
<i>Ag-D</i>	2.492
<i>Ag-G</i>	1.190
<i>Proteins</i>	292.8
<i>Unidentifiables</i>	4900

As only ethanol and water can be used as solvents, it is preferred to use water as a loading solvent, and a mixture of ethanol-water as an eluting agent. This knowledge is based on experience with similar molecules^{20,21} and rules of thumb but not on actual data for isoflavones. Therefore, a screening of the adsorption behavior of the isoflavones dissolved in different solvents was performed prior to actual adsorption experiments. The okara extracts were produced with ethanol in water ratios ranging from 0%-80% (w/w). Higher ethanol concentrations in the mixtures were not possible to achieve due to the moisture that the okara holds. The affinity of isoflavones in different ethanol-water okara extracts onto one of the resins (XAD 761) is shown Figure 3.

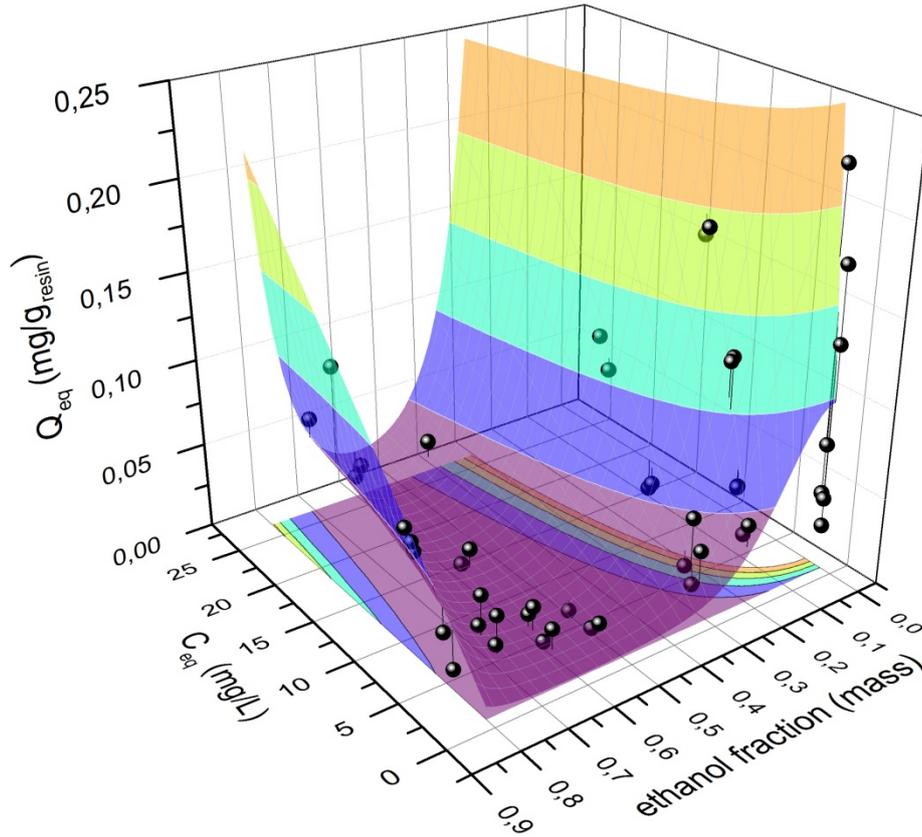


Figure 3: Isotherms for the total isoflavones at different ethanol mass fractions of okara extract with resin XAD 761. The surface is shown to guide the eye.

The slope of the isotherm is a good indication of the affinity of the compound for the resin it is adsorbing onto. Figure 3 shows that the slope of the isotherm was largest for water based solutions and it decreased with the increase of ethanol until a mass fraction of 0.5-0.6 in which the minimum is located. For higher ethanol fractions the slope increases again.

The difference (even with the closest one, the 20% ethanol) is big enough to choose water as the preferred solvent for the adsorption of isoflavones.

The adsorption of the different isoflavones and proteins on the selected resins was studied extensively. For a better understanding, the behavior of the different types of isoflavones is displayed to compare those with the other compounds of their group

(Figure 4). Most of the compounds behave in all the resins similarly. In Figure 4, resin XAD 761 was chosen for clarity.

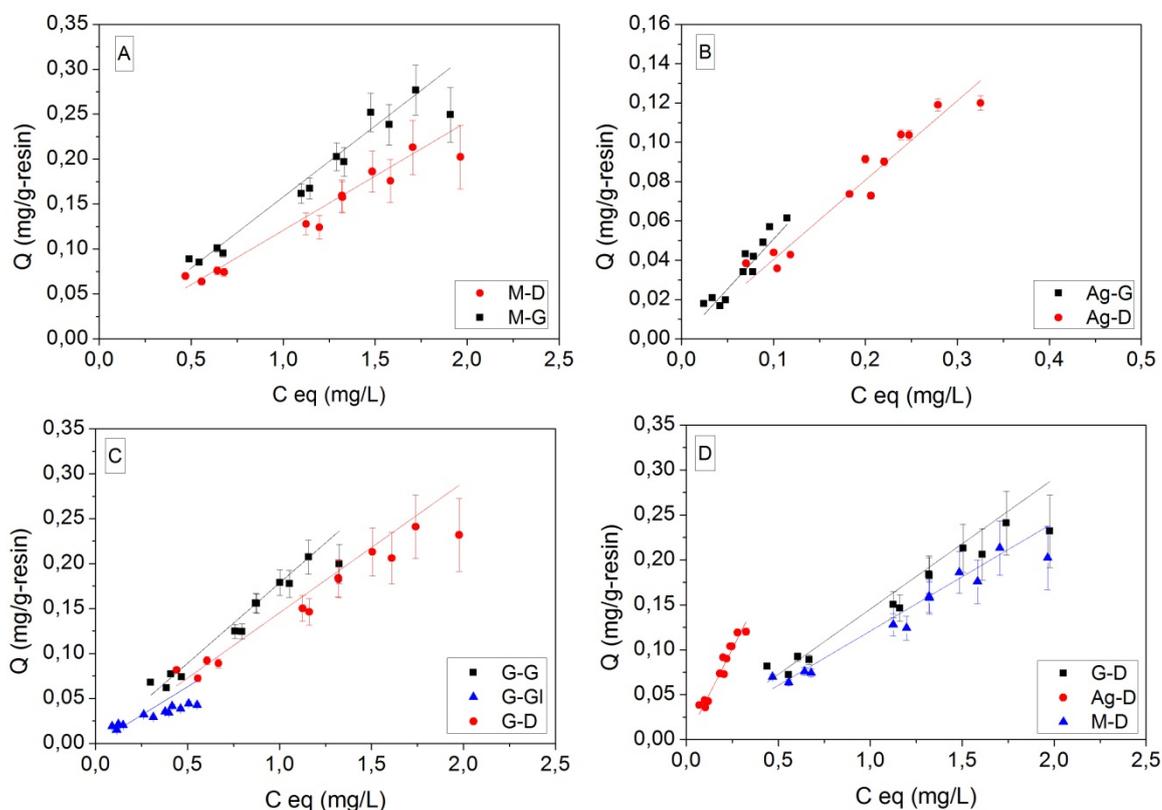


Figure 4: Different isotherm behavior of the isoflavones in adsorption of the water-based extract with resin XAD 761. A: Glucosides, B: Aglycones, C: Malonyl esters and D: daidzin-based isoflavones. Lines represent the linear isotherm the data has been fitted to.

In the three groups (aglycones, glucosides and malonyl esters) the genistin form presents always a higher slope than the daidzin one. G-GI is the only glycitin form taken into account and therefore the only glycitin-based type to be discussed. Its slope is considerably lower than any of the other forms, which indicates a lower affinity for the resin and/or a higher affinity for the solvent. The comparison of adsorption between different groups is shown in Figure 4-D. The malonyl ester has the lowest affinity, possibly because it is charged at the pH of the experiment. The slope of the

aglycones is the highest due to their hydrophobicity, which is higher than the one of the glucosides.

A common way to quantify the hydrophobicity is to determine the logarithm of the partition coefficient between 1-octanol and water, the so-called logP-value. However, logP does not account for charged species, therefore, logD was used in this work.

LogD is the value of logP as a function of pH. The values are shown in Table 2. When combining the information provided in Table 2 and Figure 4 it can be clearly seen that a higher the value of log D coincides with a steeper slope of the isotherm, which both points in the direction of higher hydrophobicity

Table 2: LogD values of most common isoflavones for a pH of 4.5. ²²

	Aglycones	Glucosides	Malonyl esters
Genistin	3,08	0,81	-0.02
Daidzin	2,73	0,46	-0.31
Glycitin	2,57	0,3	-0.56

Besides a different behavior of the isoflavones in a resin, individual isoflavones behave differently in different resins. This difference in behavior can be due to chemical and/or physical properties of the resin such as porosity, surface area, surface chemistry, and pore diameter. Due to the many possible comparisons, only total isoflavones and total proteins behavior are shown in Figures 5 and 6. For a more-in-detail comparison, all isotherms have been modelled and their regressed parameters tabulated in Tables A1-A3 (Appendix) and Table 3.

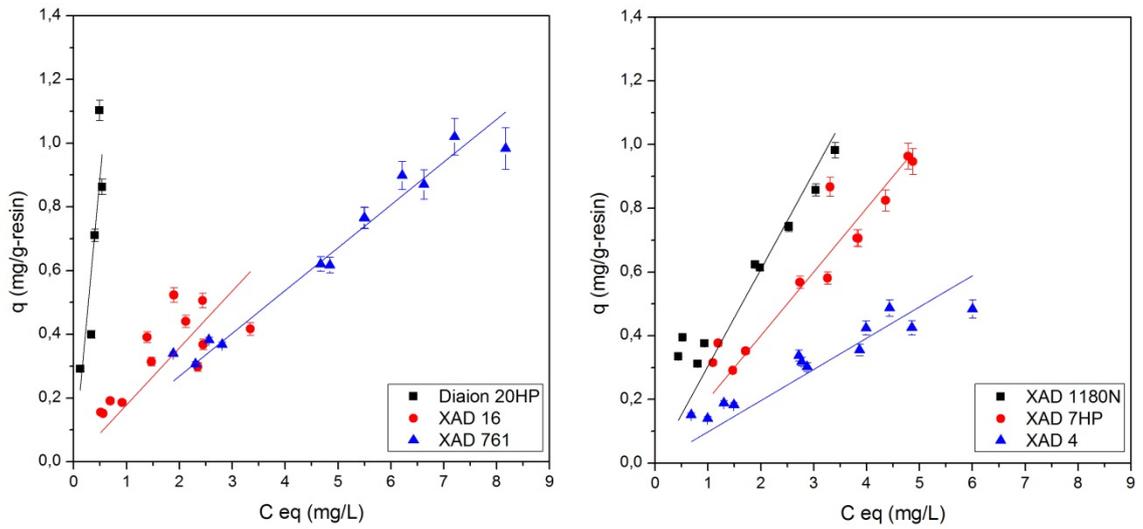


Figure 5: Isotherms of total isoflavones of the water-based okara extraction for the 6 resins studied in this work. Lines represent the linear isotherm the data has been fitted to.

Figure 5 shows that all isotherms present linear or close to linear behavior. Diaion 20HP is the resin with the steepest slope.

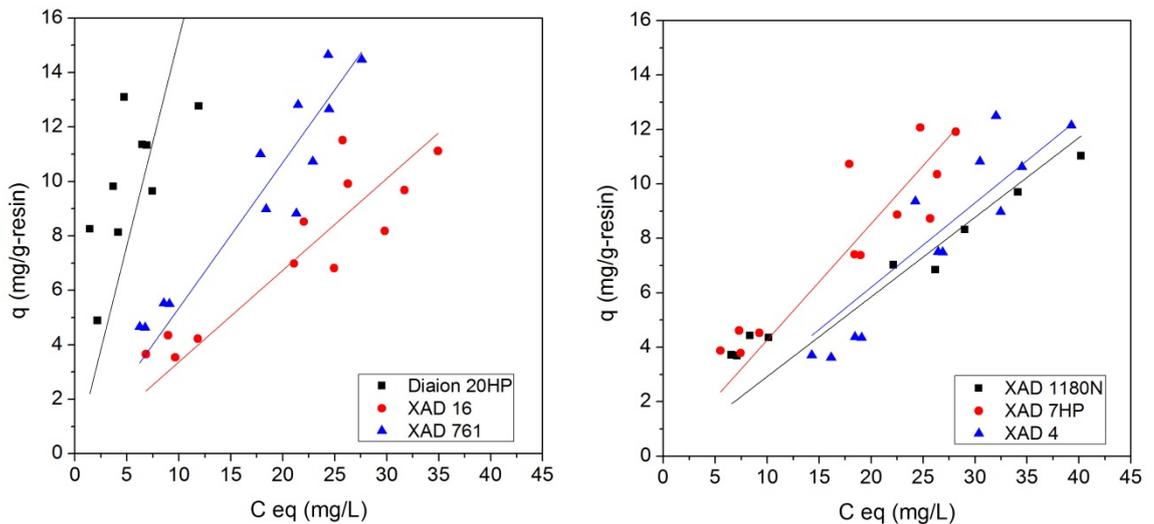


Figure 6: Isotherms of total proteins of the water-based okara extraction for the 6 resins studied in this work. Lines represent the linear isotherm the data has been fitted to.

The results for the total proteins can be seen in Figure 6. All isotherms present a close to linear behavior and, like in Figure 5, Diaion-20HP is the one with the steepest slope.

5.3.2. Modeling

Since the concentrations in this work are rather low in comparison with the large surface area of the resins, and based on the experimental results shown in Figures 4-6, it was decided to use linear isotherms as a model for fitting.

It can be observed in Table A3 (Appendix) that the value of the slope for the total unidentifiables is between 0.009 and 0.023 which is more than 10 times lower than the value for the slope for isoflavones (from 0.14 to 1.8). This shows the high affinity of this type of resins for isoflavones and proteins in comparison to non-phenolic compounds.

Unfortunately, some of the parameters from Table A1-A3 (Appendix) present a large error when fitted to a linear isotherm (specially the parameters from XAD 4) which might point to a model different from linear isotherm model. As mentioned at the beginning of this section, concentrations of the isoflavones are small in comparison with the surface area of the resin, which makes saturation effects an unlikely cause of non-linearity. However, the existence of an interaction between proteins and isoflavones is plausible and their interaction and complexation has been reported in literature²³⁻²⁵. Therefore, the adsorption of isoflavones on top of already adsorbed proteins is not unlikely. Moreover, when plotting the slope of the linear isotherm for each experiment versus the amount of adsorbed protein (Figure 7), a clear trend is observed. As well, a sketch of this interaction is shown in Figure 8.

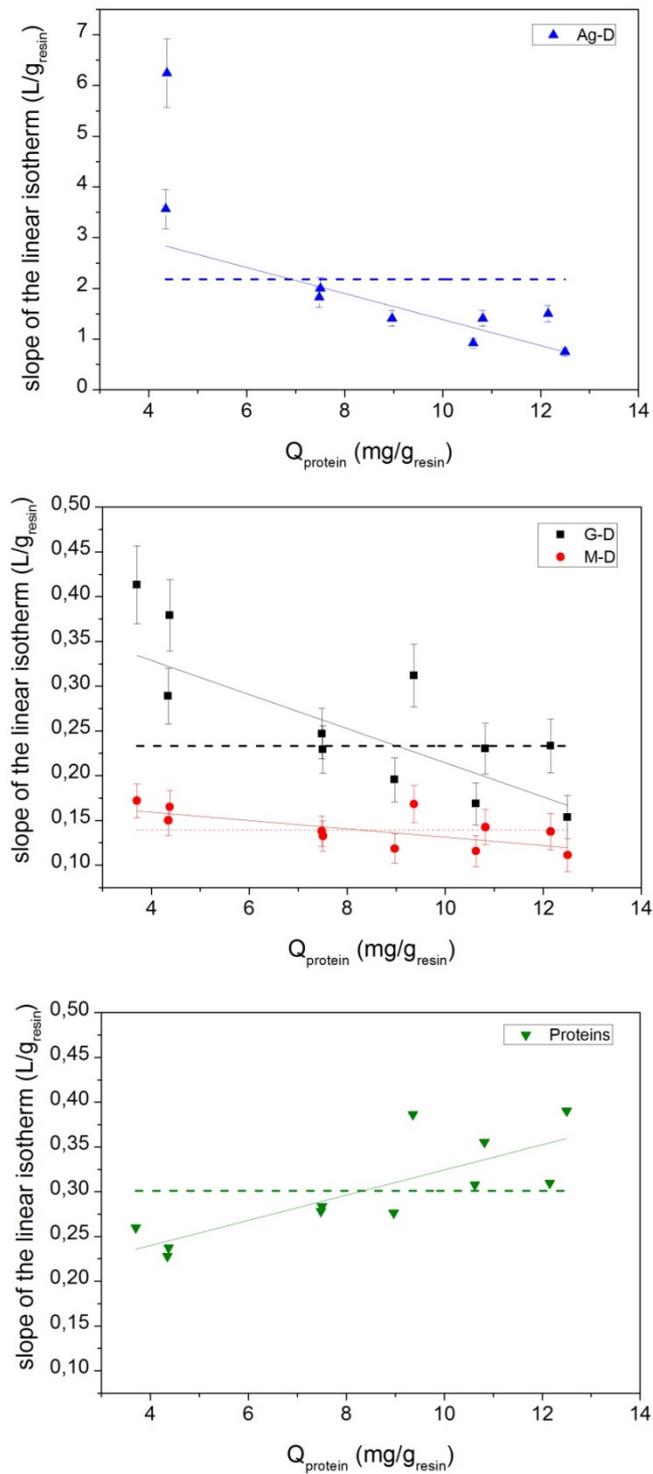


Figure 7: Examples of the relationship of the slope of the isotherm with the amount of adsorbed proteins. Solid lines represent the model described in Equation 8 with the parameters from Table 3 and dashed lines represent fit with parameters in Tables A1-A3 (Appendix).

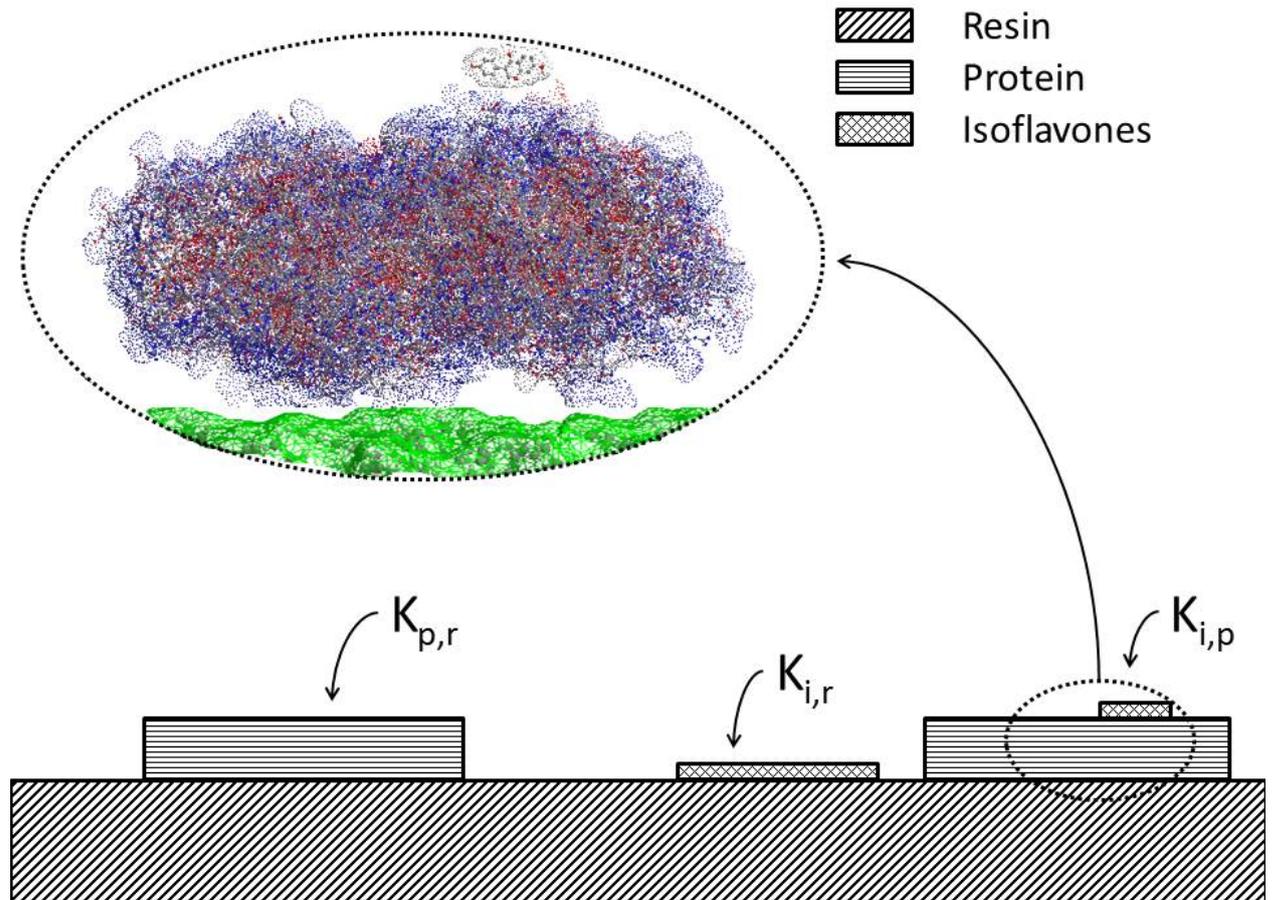


Figure 8: Sketch representing the bilayer model with distinctive affinity parameters (K) given and zoom in of the isoflavone-protein-resin complex showing the scale of the different components.

For all isoflavones, an increase in the amount of adsorbed proteins, decreases their affinity for the resin. In the case of proteins, the effect is opposite, the higher the amount of adsorbed proteins, the higher their affinity is.

A simplified model for this system can be developed in which a linear adsorption is presumed in both systems: isoflavone-resin and isoflavone-protein. Equation 3 represents the well-known Langmuir Equation for isotherms.

$$Q_{eq,i} = \frac{q_{max,i}K_{l,i}C_{eq,i}}{1+K_{l,i}C_{eq,i}} \quad [3]$$

in which $K_{l,i}$ represents the affinity coefficient (L/g_{resin}), $q_{mas,i}$ the maximum capacity (mg/g_{resin}) and $C_{eq,i}$ the concentration (mg/L) of species “i” in the liquid at

equilibrium. When the concentration of the species is really small, Equation 3 can be simplified to Equation 4 by presuming that $K_{l,i}C_{eq,i} \ll 1$.

$$Q_{eq,i} = q_{max,i}K_{l,i}C_{eq,i} \quad [4]$$

In both systems (isoflavone-resin and isoflavone-protein), $q_{max,i}$ is a function of the protein adsorbed. In the case of the isoflavone-resin system, $q_{max,i}$ decreases linearly with the amount of adsorbed protein, since the amount of available resin is lower when protein is adsorbed. In the case of the system isoflavone-protein, $q_{max,i}$ increases linearly with the amount of protein adsorbed onto the resin. Taking this into consideration, Equation 5 is the one used for the system isoflavone-resin while Equation 6 is the one used for the isoflavone-protein system.

$$Q_{eq,i/r} = (q_{max,i} - aQ_{eq,p})K_{l,i}C_{eq,i} \quad [5]$$

$$Q_{eq,i/p} = aQ_{eq,p}K_{l,p}C_{eq,i} \quad [6]$$

Where “ a ” represents a surface parameter for proteins and “ p ” stands for proteins.

Equation 7 can be used for the entire system by combining Equations 5 and 6, and it can be simplified further to Equation 8.

$$Q_{eq,i} = [(K_{l,p}a - K_{l,i}a)Q_{eq,p} + K_{l,i}q_{max,i}]C_{eq,i} \quad [7]$$

$$Q_{eq,i} = (K_{i/p}Q_{eq,p} + K_{i/r})C_{eq,i} \quad [8]$$

In this context $K_{i/p}$ relates to the affinity of the isoflavone for the protein in comparison with the affinity of the isoflavone for the resin. This parameter will be negative when the affinity for the protein is lower than the affinity for the resin. $K_{i/r}$ relates directly to affinity of the isoflavone for the resin. When Equation 8 is applied to proteins, instead of isoflavones, it becomes Equation 9.

$$Q_{eq,p} = \frac{K_{p/r}C_{eq,p}}{1 - K_{p/p}C_{eq,p}} \quad [9]$$

The parameters have been regressed for the system and are displayed in Table 3.

Table 3: $K_{i/r}$ and $K_{i/p}$ parameters from Equation 8 and 9. Bold numbers represent the equilibria for which Equation 8 is preferred to Equation 4.

		G-D	G-GI	G-G	M-D	M-G	Ag-D	Ag-G	Proteins
XAD 4	$K_{i/p}$ [L/ mg]	-0.0225 ± 0.0075	-0.0145 ± 0.0047	-0.0236 ± 0.0085	-0.0060 ± 0.0019	-0.0077 ± 0.0031	-0.256 ± 0.076	-0.38 ± 0.44	0.0142 ± 0.0028
	$K_{i/r}$ [L/ g _{resin}]	0.440 ± 0.071	0.298 ± 0.046	0.484 ± 0.081	0.191 ± 0.016	0.231 ± 0.027	3.95 ± 0.83	5.7 ± 5.3	0.182 ± 0.017
XAD 16	$K_{i/p}$ [L/ mg]	-0.017 ± 0.016	-0.017 ± 0.016	-0.019 ± 0.019	-0.0046 ± 0.0084	-0.012 ± 0.014	-0.041 ± 0.067	-0.0073 ± 0.058	-0.016 ± 0.011
	$K_{i/r}$ [L/ g _{resin}]	0.49 ± 0.13	0.48 ± 0.13	0.54 ± 0.15	0.304 ± 0.066	0.42 ± 0.11	1.16 ± 0.57	0.89 ± 0.49	0.491 ± 0.061
XAD 1180 N	$K_{i/p}$ [L/ mg]	-0.027 ± 0.011	-0.036 ± 0.016	-0.036 ± 0.020	-0.0121 ± 0.0072	-0.029 ± 0.017	-0.080 ± 0.055	-0.131 ± 0.076	-0.050 ± 0.011
	$K_{i/r}$ [L/ g _{resin}]	0.486 ± 0.079	0.56 ± 0.13	0.65 ± 0.15	0.328 ± 0.051	0.56 ± 0.13	1.45 ± 0.45	2.12 ± 0.65	0.711 ± 0.054
XAD 761	$K_{i/p}$ [L/ mg]	-0,0021 ± 0.0017	-0,0073 ± 0.0018	-0,0012 ± 0.0019	-6.7E-4 $\pm 6.5E-4$	-9.7E-4 ± 0.0013	0,0018 ± 0.0043	0,0104 ± 0.0072	-0,0218 ± 0.0075
	$K_{i/r}$ [L/ g _{resin}]	0,157 ± 0.014	0,179 ± 0.019	0,183 \pm 0.016	0,0830 \pm 0.0051	0,126 ± 0.010	0,385 ± 0.043	0,415 ± 0.070	0,779 ± 0.055
XAD 7HP	$K_{i/p}$ [L/ mg]	-0,0053 ± 0.0035	-0,0074 ± 0.0034	-0,0059 ± 0.0063	-3.2E-4 ± 0.0014	-0,0022 ± 0.0046	-0,010 ± 0.021	-0,010 ± 0.036	-0,026 ± 0.012
	$K_{i/r}$ [L/ g _{resin}]	0,218 ± 0.028	0,182 ± 0.031	0,314 ± 0.053	0,126 ± 0.011	0,241 ± 0.037	0,70 ± 0.18	0,93 ± 0.31	0,684 ± 0.074
Diaio n 20HP	$K_{i/p}$ [L/ mg]	0,063 ± 0.027	-0,09 ± 0.15	0,176 ± 0.072	0,0890 ± 0.0075	0,199 ± 0.062	0,23 ± 0.33	N/A	-0,07 ± 0.13
	$K_{i/r}$ [L/ g _{resin}]	0,57 ± 0.23	3,1 ± 1.6	0,66 ± 0.58	0,253 ± 0.059	0,44 ± 0.49	5,6 ± 2.9	N/A	2,8 ± 1.2

N/A: Not enough data points for a meaningful regression

As seen in Table 3, this approach yields lower errors even though in some cases they are still significant.

The assumption of linear isotherms for proteins and isoflavones is strengthened by the calculation of the surface coverage. Okara protein is assumed to be beta-conglycinin² with a molar weight of 150 kDa and a maximal projection area of $3.326 \cdot 10^3 \text{ \AA}^2$ (properties calculated using YASARA 13.3.12). Properties from the resin are taken from Mendez Sevillano et al. (2014)¹². The protein coverage never surpassed 1% of the total surface of the resin, having 0.84% as a maximum value for the experiments with XAD 761 (the resin with the lowest surface area).

Two models have been studied, the one described by Equation 4 (linear model) and the one described by Equation 8 (bilayer model). The bilayer model shows a better fit than the linear model (Figure 7) but it uses one more parameter so their comparison is not as easy as comparing fits. Akaike Information Criteria (AIC)²⁶ can be used to compare different models with different amount of regressed parameters. The model that obtains the minimum AIC value is the more likely to describe the system.

In this work it has been considered that the bilayer model is the one closest to describe the adsorption if its AIC value is 2 units below the AIC of the linear model (as it is accustomed for comparisons with different parameters²⁷). Equilibria in which that occurred have been bolded in Table 3. The bilayer model describes the experiments well with resins XAD 4 and Dianion 20HP and in a lesser extent with the other resins. Those two resins exhibit a higher hydrophobicity as it can be seen from their higher $K_{i/r}$ for aglycones (the most hydrophobic isoflavones). This higher hydrophobicity of these resins might drive the less hydrophobic isoflavones to adsorb onto the adsorbed proteins, thereby making the bilayer model a more accurate description of the actual adsorption.

The regressed parameters for the different compounds are correlated to their logD (Table 2) in Figure 9.

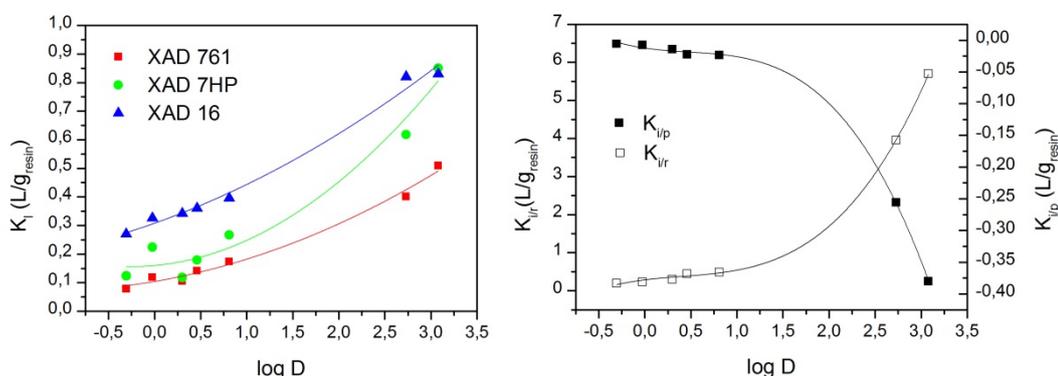


Figure 9: Slope of the linear isotherm of the isoflavones (Table A1) vs. $\log D$ of all of the isoflavones for three resins (left). $K_{i/r}$ and $K_{i/p}$ (Table 3) vs. $\log D$ of all of the isoflavones for XAD 4. Lines are shown to guide the eye.

Figure 9 shows a clear relationship between the equilibrium constants and the $\log D$ of each isoflavone species of the solution which points at hydrophobicity as the main driving force in this system.

5.4. Conclusion

Batch adsorption has been carried out on different resins using the extract of okara (of different ethanol / water mixtures) and several food-grade macroporous polymeric resins. The content of each isoflavone, proteins and dry mass has been analysed for those samples yielding isotherms for each of those compounds/ group of compounds. The isotherms are modelled by regressing parameters that give affinity information crucial for the understanding of the different interactions that take place in this complex mixture and how to use them for the optimization of a process. A model is proposed and developed based on observed data by which isoflavones form a bilayer adsorbing onto proteins as well as on the surface of the resin. Akaike criterion shows

enough evidence for the formation of a bilayer in adsorption. This bilayer model is proved for the equilibria of all isoflavones with resin XAD 4 and the proteins with all the resins but Diaion 20HP. Regressed parameters and their accuracy is reported. Moreover, hydrophobicity is shown to be an important factor to drive the process in both types of adsorption: when polyphenols adsorb onto the resin and when they adsorb onto a protein layer. The regressed thermodynamic parameters ($K_{l,i}$, $K_{i/p}$ and $K_{i/r}$) clearly correlate with the hydrophobicity of the molecules. This finding can be used in future work for the development of a predictive model in which the properties of the solute and of the resin are taken into account.

Appendix

Table A1: Slopes of the different isoflavones for each resin [L/g_{resin}]

	XAD-4	XAD 16	XAD 1180N	XAD 761	XAD 7HP	Diaion 20HP
G-D	0.239 ±0.027	0.360 ±0.036	0.305 ±0.034	0.1416 ±0.0056	0.179 ±0.010	1.08 ±0.12
G-GI	0.162 ±0.017	0.342 ±0.037	0.263 ±0.045	0.1055 ±0.0092	0.118 ±0.010	2.20 ±0.46
G-G	0.268 ±0.029	0.395 ±0.043	0.393 ±0.053	0.1736 ±0.0063	0.267 ±0.018	2.00 ±0.31
M-D	0.1436 ±0.0077	0.270 ±0.020	0.248 ±0.020	0.0782 ±0.0020	0.1237 ±0.0039	0.92 ±0.14
M-G	0.169 ±0.011	0.326 ±0.032	0.356 ±0.045	0.1188 ±0.0040	0.224 ±0.013	1.91 ±0.33
Ag-D	1.18 ±0.22	0.82 ±0.13	0.822 ±0.131	0.401 ±0.015	0.618 ±0.055	7.50 ±0.75
Ag-G	1.18 ±0.26	0.83 ±0.10	1.03 ±0.17	0.509 ±0.027	0.851 ±0.092	8.12 ±1.60

Table A2: Slopes of the different groups of isoflavones for each resin [L/g_{resin}]

	XAD-4	XAD 16	XAD 1180N	XAD 761	XAD 7HP	Diaion 20HP
Total Aglycones	1.23 ±0.21	0.82 ±0.12	0.89 ±0.15	0.430 ±0.176	0.674 ±0.614	7.62 ±0.65
Total Glucosides	0.241 ±0.027	0.372 ±0.039	0.331 ±0.041	0.1508 ±0.0063	0.199 ±0.013	1.44 ±0.16
Total Malonyl esters	0.1573 ±0.0095	0.300 ±0.026	0.300 ±0.030	0.0986 ±0.0029	0.1695 ±0.0071	1.31 ±0.21

Table A3: Slopes of the major compounds of okara extract [L/g_{resin}]

	XAD-4	XAD 16	XAD 1180N	XAD 761	XAD 7HP	Diaion 20HP
Total Isoflavones	0.220 ±0.015	0.369 ±0.036	0.349 ±0.039	0.1409 ±0.0048	0.209 ±0.011	1.58 ±0.20
Total Proteins	0.259 ±0.014	0.404 ±0.026	0.468 ±0.037	0.633 ±0.029	0.537 ±0.034	2.48 ±0.73
Total Unidentifiables	9.45e-3 ±7.3e-4	1.38e-2 ±1.1e-3	1.45e-2 ±1.1e-3	2.27e-2 ±1.8e-3	1.02e-2 ±7.9e-4	2.32e-2 ±1.8e-3

Acknowledgements

This work has been carried out within the framework of the Institute for Sustainable Process Technology (ISPT) in the project FO-10-03.

References

1. Jankowiak, L.; Trifunovic, O.; Boom, R. M.; van der Goot, A. J., The potential of crude okara for isoflavone production. *J. Food Eng.* **2014**, 124, 166.
2. Stanojevic, S. P.; Barac, M. B.; Pesic, M. B.; Vucelic-Radovic, B. V., Composition of Proteins in Okara as a Byproduct in Hydrothermal Processing of Soy Milk. *J. Agric. Food Chem.* **2012**, 60, 9221.
3. Mateos-Aparicio, I.; Redondo-Cuenca, A.; Villanueva-Suárez, M. J., Isolation and characterisation of cell wall polysaccharides from legume by-products: Okara (soymilk residue), pea pod and broad bean pod. *Food Chem.* **2010**, 122, 339.
4. Redondo-Cuenca, A.; Villanueva-Suárez, M. J.; Mateos-Aparicio, I., Soybean seeds and its by-product okara as sources of dietary fibre. Measurement by AOAC and Englyst methods. *Food Chem.* **2008**, 108, 1099.
5. Kuiper, G. G. J. M.; Lemmen, J. G.; Carlsson, B.; Corton, J. C.; Safe, S. H.; van der Saag, P. T.; van der Burg, B.; Gustafsson, J.-Å., Interaction of Estrogenic Chemicals and Phytoestrogens with Estrogen Receptor β . *Endocrinology* **1998**, 139, 4252.
6. Manach, C.; Scalbert, A.; Morand, C.; Rémésy, C.; Jiménez, L., Polyphenols: food sources and bioavailability. *Am. J. Clin. Nutr.* **2004**, 79, 727.
7. Wu, C. Y.; Lai, S. M., Preparative Isolation of Isoflavones from Defatted Soy Flakes. *J. Liq. Chromatogr. Relat. Technol.* **2007**, 30, 1617.
8. Kirakosyan, A.; Kaufman, P. B.; Chang, S. C.; Warber, S.; Bolling, S.; Vardapetyan, H., Regulation of isoflavone production in hydroponically grown *Pueraria montana* (kudzu) by cork pieces, XAD-4, and methyl jasmonate. *Plant Cell Rep.* **2006**, 25, 1387.
9. Chang, L.-H.; Chang, C.-M. J., Continuous hot pressurized fluids extraction of isoflavones and soyasaponins from defatted soybean flakes. *J. Chin. Inst. Chem. Eng.* **2007**, 38, 313.
10. He, J.; Zhao, Z.; Shi, Z.; Zhao, M.; Li, Y.; Chang, W., Analysis of Isoflavone Daidzein in *Puerariae radix* with Micelle-Mediated Extraction and Preconcentration. *J. Agric. Food Chem.* **2005**, 53, 518.

11. Xu, L.; Kumar, A.; Lamb, K.; Layton, L., Production of an isoflavone concentrate from red clover flowers. *J. Food Sci.* **2005**, *70*, S558.
12. Sevillano, D. M.; van der Wielen, L. A. M.; Hooshyar, N.; Ottens, M., Resin Selection for the Separation of Caffeine from Green Tea Catechins. *Food and Bioprod. Process.* **2014**, *92*, 192.
13. Li-Hsun, C.; Ya-Chuan, C.; Chieh-Ming, C., Extracting and purifying isoflavones from defatted soybean flakes using superheated water at elevated pressures. *Food Chem.* **2004**, *84*, 279.
14. Toro-Funes, N.; Odriozola-Serrano, I.; Bosch-Fusté, J.; Latorre-Moratalla, M. L.; Veciana-Nogués, M. T.; Izquierdo-Pulido, M.; Vidal-Carou, M. C., Fast simultaneous determination of free and conjugated isoflavones in soy milk by UHPLC–UV. *Food Chem.* **2012**, *135*, 2832.
15. Bradford, M. M., A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **1976**, *72*, 248.
16. Worthing, A. G. G., Joseph, Treatment of Experimental Data. *J. Chem. Educ.* **1943**, *20*, 572.
17. Jackson, C. J. C.; Dini, J. P.; Lavandier, C.; Rupasinghe, H. P. V.; Faulkner, H.; Poysa, V.; Buzzell, D.; DeGrandis, S., Effects of processing on the content and composition of isoflavones during manufacturing of soy beverage and tofu. *Process Biochem.* **2002**, *37*, 1117.
18. Rinaldi, V. E. A.; Ng, P. K. W.; Bennink, M. R., Effects of Extrusion on Dietary Fiber and Isoflavone Contents of Wheat Extrudates Enriched with Wet Okara. *Cereal Chem.* **2000**, *77*, 237.
19. Wang, H.-J.; Murphy, P. A., Mass Balance Study of Isoflavones during Soybean Processing. *J. Agric. Food Chem.* **1996**, *44*, 2377.
20. Yamamoto, S.; Hakoda, M.; Oda, T.; Hosono, M., Rational method for designing efficient separations by chromatography on polystyrene–divinylbenzene resins eluted with aqueous ethanol. *J. Chromatogr. A* **2007**, *1162*, 50.
21. Kammerer, D. R.; Carle, R.; Stanley, R. A.; Saleh, Z. S., Pilot-Scale Resin Adsorption as a Means To Recover and Fractionate Apple Polyphenols†. *J. Agric. Food Chem.* **2010**, *58*, 6787.
22. <http://www.chemicalize.org> (13-08-2014),

23. Charlton, A. J.; Davis, A. L.; Jones, D. P.; Lewis, J. R.; Davies, A. P.; Haslam, E.; Williamson, M. P., The self-association of the black tea polyphenol theaflavin and its complexation with caffeine. *J. Chem. Soc., Perkin Trans.* **2000**, 317.
24. Charlton, A. J.; Baxter, N. J.; Khan, M. L.; Moir, A. J. G.; Haslam, E.; Davies, A. P.; Williamson, M. P., Polyphenol/Peptide Binding and Precipitation. *J. Agric. Food Chem.* **2002**, 50, 1593.
25. Speroni, F.; Milesi, V.; Añón, M. C., Interactions between isoflavones and soybean proteins: Applications in soybean-protein–isolate production. *LWT--Food Sci. Technol.* **2010**, 43, 1265.
26. Akaike, H., A new look at the statistical model identification. *IEEE Trans. Autom. Control* **1974**, AC-19, 716.
27. Arnold, T. W., Uninformative Parameters and Model Selection Using Akaike's Information Criterion. *J. Wildl. Manage.* **2010**, 74, 1175.

Chapter 6: A Thermodynamic Lattice Model Describing Adsorption of Phytochemicals onto Macroporous Polymeric Resins

Abstract

Adsorption is an important mechanism that is used widely for purification or concentration in several fields. One growing field is that of nutraceuticals, which are products derived from food sources that have health-related properties. This work shows the development of a thermodynamically consistent adsorption model for the four main catechins (nutraceuticals from tea) onto macroporous polymeric resins made from polystyrene divinylbenzene, using a lattice of available adsorption sites. Infrared spectroscopy measurements were performed in order to properly characterize the resin. In addition activity coefficients in the interphase were calculated for the four compounds and modelled with the Margules model. The fit is good (0.26 log units) and the regressed parameters are reported with their accuracy. These parameters can be used for the calculation of the equilibrium constants. All regressed and calculated parameters were correlated to the logD value of the compound and show linearity (R^2 of 0.9989 and 0.9881) indicating that hydrophobicity is a dominant driving force for the adsorption of catechins on polystyrene divinylbenzene resins.

Submitted as: A Thermodynamic Lattice Model Describing Adsorption of Phytochemicals onto Macroporous Polymeric Resins. David Méndez Sevillano, Sumit Sachdeva, Luuk A.M. van der Wielen, Nasim Hooshyar, Louis C.P.M. de Smet and Marcel Ottens to *Langmuir*.

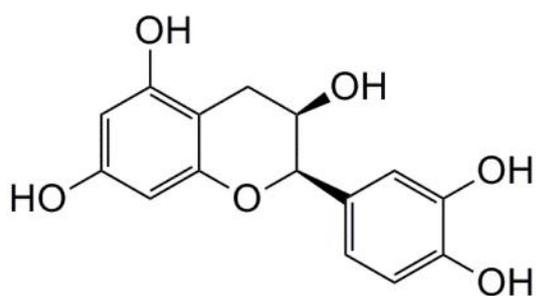
6.1. Introduction

Adsorption is a widely used mechanism for removal, concentration or purification of target compounds from a liquid or gaseous stream. Production of pharmaceuticals¹, bulk chemicals², treatment of waste water³, etc. often rely on adsorbents for their key processing steps. This thorough use of adsorbents has driven research from engineering or empirical models to more mechanistic ones in an attempt to understand and predict the behavior of target substances with the sorbent of choice.

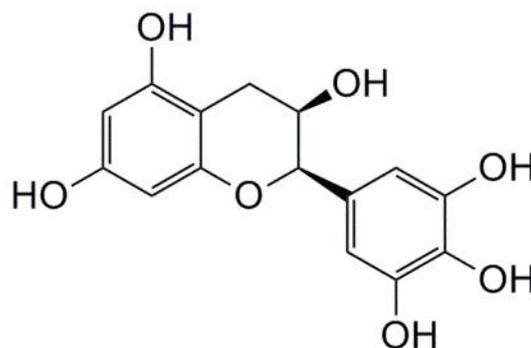
Adsorption is currently being studied for its application in the field of nutraceuticals.

Nutraceuticals are compounds present in our diet with positive health-related effects⁴. These compounds are mostly found in plants and are often present in waste streams of food industry that can be valorise by their recovery. An interesting family of antioxidants are the polyphenols, because of their health-related properties which is the result of a variety of benefits including lowering cholesterol levels⁵, obesity or certain types of cancer⁶.

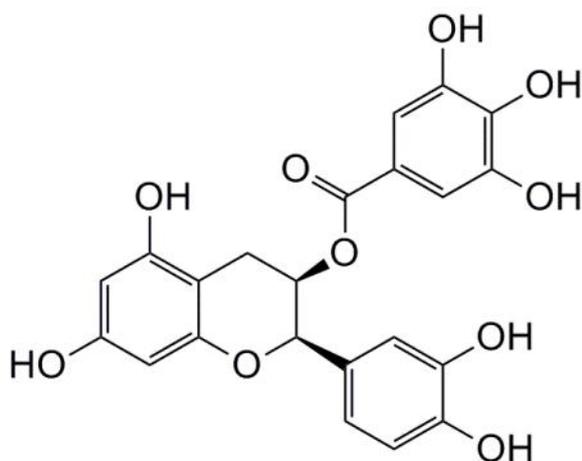
Research on concentration or purification of polyphenols from complex streams has focused on their adsorption onto macroporous polymeric resins⁷⁻¹³. These resins offer some clear advantages such as very high surface area, low cost and food-grade status. Most of these resins are styrene-divinylbenzene polymers and therefore, hydrophobicity is likely to be one of the main driving forces for the adsorption. However, very little mechanistic knowledge has been gathered about this interaction. Moreover, the impossibility of direct measurements at the interphase makes the determination of important thermodynamic properties, such as activity coefficients, challenging to measure, and therefore, model.



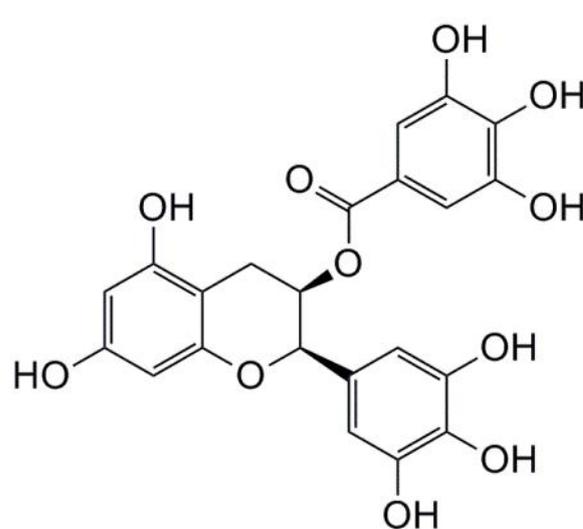
(-)-Epicatechin (EC)



(-)-Epigallocatechin (EGC)



(-)-Epicatechin gallate (ECg)



(-)-Epigallocatechin gallate (EGCg)

Figure 1. Four major catechins present in tea.

The aim of this work is to develop a consistent thermodynamic framework whereby important thermodynamic properties of adsorption can be measured and modelled. This framework is then to be applied to a multicomponent adsorption of polyphenols onto a macroporous polymeric resin. The chosen polyphenols for the experiments are the four main catechins (Figure 1) that can be found in tea leaves or green tea beverage.

6.2. Theoretical framework.

The model proposed in this work is based on surface excess models thoroughly studied in literature¹⁴⁻¹⁶. The system of study is divided in three phases as showed in Figure 2: the bulk phase (or liquid phase), the adsorbed phase and the resin. The adsorbed phase consists of a monolayer of molecules on the surface of the resin. There is no chemical exchange between the resin and the adsorbed phase and therefore, no thermodynamic equilibrium relations can be calculated from their interaction.

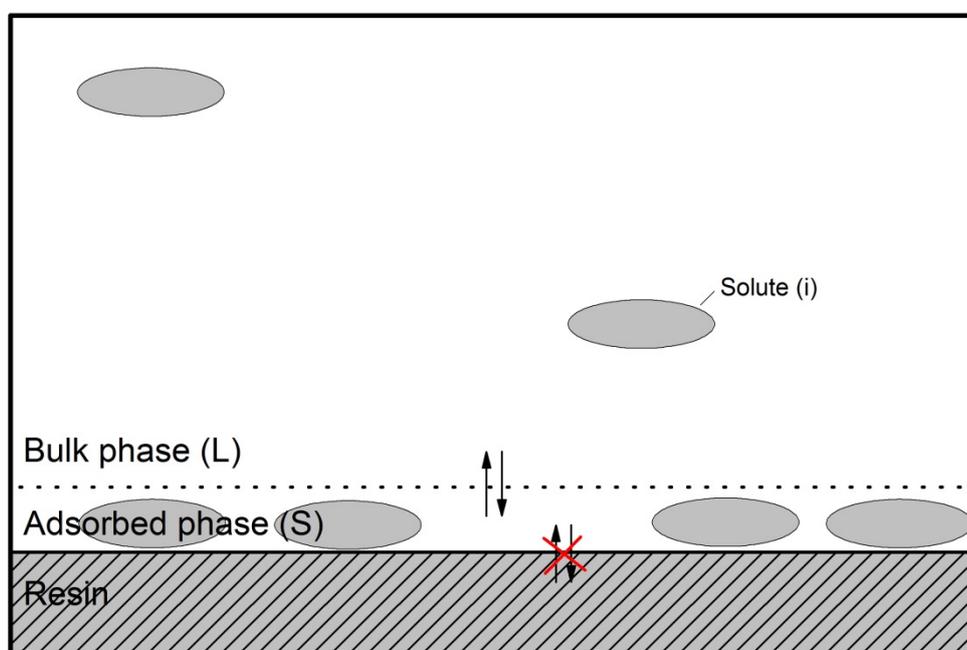


Figure 2: Scheme of the system considered in this study (not to scale). Ellipses represent the solute molecule.

If both the adsorbed phase and the bulk phase are in equilibrium, the chemical potential of each compound (i) is equal to the chemical potential of said compound in the other phase.

$$\mu_i^l = \mu_i^s \quad [1]$$

Where μ represents chemical potential and superscripts l and s represent liquid phase and adsorbed phase.

The change in chemical potential taking only composition effects into account can be calculated with Equation 2.

$$\mu_i = \mu_i^0 + RT \ln(x_i \gamma_i) \quad [2]$$

Where μ_i^0 represents the chemical potential of compound “i” at a standard state (pure i liquid at atmospheric pressure and room temperature), R is the ideal gas constant ($8.314 \text{ JK}^{-1} \text{ mol}^{-1}$), T stands for temperature (in Kelvin), x_i represents molar fraction and γ_i activity coefficient of compound “i”. Equation 3 is the result of substituting Equation 2 on both sides of Equation 1 followed by some operations.

$$\frac{x_i^L \gamma_i^L}{x_i^S \gamma_i^S} = \exp\left(\frac{\mu_i^{S,0} - \mu_i^{L,0}}{RT}\right) = K_i(T) \quad [3]$$

Where K_i is the equilibrium constant. The molar fraction in the bulk phase (x_i^L) can be calculated by measuring directly the concentration of the different components. The activity coefficient in the liquid phase (γ_i^L) can be predicted by models from literature with proven accuracy like NRTL-SAC¹⁷, COSMO-RS¹⁸ or UNIFAC¹⁹. Since this work is focused on polyphenols, the activity coefficient model used in it is MPP-UNIFAC²⁰, an activity coefficient model based on mod-UNIFAC but specific for polyphenol interactions.

Based on a mass balance between initial and equilibrium time, adsorbed moles can be quantified (n_i^0 and n_i^L respectively). The challenge arises for the calculation of the molar fraction in the adsorbed phase since the volume of such phase is unknown. The surface excess model, as explained in literature, applies only for microporous resins as it calculates molar fractions in the adsorbed phase from the volumetric capacity of said micropore. This calculation cannot be used in the case of macroporous resins since the maximum diameter of their pores can go up to $1 \mu\text{m}$ ²¹ making the assumption of constant concentration throughout

the pore implausible. Therefore, molar fraction in the adsorbed phase needs to be calculated in a different manner.

In this context, adsorption is a surface phenomenon and therefore, the logical measure of the ratio of a compound should be the ratio of the surface that is covered by this compound.

Furthermore, since most of these resins are polystyrene-divinylbenzene polymers, π - π bonding is likely to be involved in the adsorption mechanism. Based on this reasoning, a lattice-based model has been developed whereby the surface of the resin consists of a finite number of benzene rings (N_s) and solute molecules (n_i) occupy a defined number of them (L_i) and therefore, molar fractions in the solid can be calculated by Equation 4.

$$x_i^S = \frac{(n_i^0 - n_i^L)L_i}{N_s} \quad [4]$$

After applying Equation 4 to Equation 3, x_i^L , γ_i^L and x_i^S can be calculated and K_i and γ_i^S remain unknown and, in theory, cannot be calculated independently from each other.

However, if Equation 4 is applied to an infinite dilution scenario, it transforms into Equation 5.

$$\frac{x_i^{L,\infty}\gamma_i^{L,\infty}}{x_i^{S,\infty}\gamma_i^{S,\infty}} = K_i(T) \quad [5]$$

The ratio between the molar fractions in the bulk (x_i^L) and in the adsorbed phase (x_i^S) is also known as Henry adsorption constant (H_i) and can be obtained from the experiments. Since K_i is independent of the concentration it can be substituted again into Equation 3 and rearranged to get Equation 6.

$$\frac{\gamma_i^S}{\gamma_i^{S,\infty}} = \gamma_i^{*,S} = \frac{x_i^L \gamma_i^L}{H_i \gamma_i^{L,\infty} x_i^S} \quad [6]$$

Equation 6 allows the determination of the so-called unsymmetric activity coefficient²² ($\gamma_i^{*,S}$) from experiments.

Different activity coefficient models were studied in this work for the modelling of the unsymmetric activity coefficients determined with Equation 6. Most of the activity

coefficient models in literature study the symmetric activity coefficient (γ_i) but they can be used for unsymmetric activity coefficients by calculating γ_i^∞ within the model and dividing said symmetric activity coefficient by the infinite dilution one (γ_i^∞). Symmetric activity coefficients²² follow Equations 7 and 8.

$$x_i \rightarrow 0 \quad \gamma_i = \gamma_i^\infty \quad [7]$$

$$x_i \rightarrow 1 \quad \gamma_i = 1 \quad [8]$$

While the unsymmetric activity coefficient follows Equations 9 and 10.

$$x_i \rightarrow 0 \quad \gamma_i = 1 \quad [9]$$

$$x_i \rightarrow 1 \quad \gamma_i = (\gamma_i^\infty)^{-1} \quad [10]$$

However due to the inherent characteristics of adsorption of biomolecules, the range of molar fractions that are studied very rarely surpass high molar fractions and the calculation of $\gamma_i^{S,\infty}$ requires an extrapolation to $x_i^S=1$ (Equation 10). Therefore, to avoid overparameterized models and the consequent inaccuracy in their parameters, it was decided to use a simple model, the two-suffix Margules²², for modelling surface activity coefficients. For similar reasons, it was assumed that the polyphenol-polyphenol interaction is negligible. This model can be found in Equation 11.

$$\ln \gamma_i^S = \frac{A_i}{RT} (x_w^S)^2 \quad [11]$$

Where A_i is the Margules constant and x_w^S is the molar fraction of water in the adsorbed phase. The infinite dilution activity coefficient can be calculated from Equation 11 and Equation 7, as depicted in Equation 12.

$$\ln \gamma_i^{S,\infty} = \frac{A_i}{RT} \quad [12]$$

Furthermore, the model for unsymmetric activity coefficient is developed from the ratio between γ_i^S and $\gamma_i^{S,\infty}$ and can be seen in Equation 13.

$$\ln \gamma_i^{S,*} = \frac{A_i}{RT} [(x_w^S)^2 - 1] \quad [13]$$

Lastly, the Margules constant (A_i) regressed from the experiments, can be used in both Equations 11 and 13. Thus allowing the calculation of the symmetric activity coefficient and therefore, the calculation of the equilibrium constant K_i using Equation 3.

6.2. Materials and Methods

6.2.1. Materials.

Catechins concentrate SunPhenon 90LB was purchased from Taiyo Kagaku Co. Ltd (Jiangsu, China). Epicatechin, Epigallocatechin, Epicatechin gallate and Epigallocatechin gallate (EC, ECG, ECg and ECGg respectively) were obtained from Nacalai UNA, Inc. with a purity of 98% or higher. Methanol and acetonitrile (purchased from Sigma-Aldrich with a purity higher than 99.9%) and Milli-Q water are used as solvent in different steps.

Ethylenediaminetetraacetic acid (EDTA) and glacial acetic acid (purity higher than 99.85%) were purchased from Sigma-Aldrich (Missouri USA). The resin AMBERLITE™ FPX66 was purchased from Dow Chemical (Midland, Michigan USA).

6.2.2. Batch uptake experiments.

Experiments were done in Multiscreen® Filter Plates (Millipore USA) Resin was applied into the filter plate with the aid of Titan 96 Well Resin Loader (Radleys, UK) and methanol was pipetted in each well for preventing drying of the resin. After two washing steps with methanol and ethanol respectively, an equilibration step with water was performed and then different dilutions of SunPhenon 90LB in Milli-Q water were pipetted in the wells.

The filter plate is stirred at 50 rpm for 60 minutes (initial batch uptake experiments showed an equilibration time of approximately 30 minutes) and its content is centrifuged onto a deep well plate (VWR International, USA) for posterior HPLC analysis.

6.2.3. HPLC quantification

EC, EGC, ECg and EGCg were analysed with by HPLC (DIONEX IC-3000) in a C18 column (Luna 5 μm Phenyl-Hexyl, Phenomenex, ltd.) following protocol detailed in literature²³ using acetic acid in acetonitrile (2% v/v) and in 20 mg/l EDTA in water (2% v/v) as mobile phases and read at a wavelength of 278 nm with a Photodiode Array detector (DIONEX PDA-3000). In the resultant chromatogram (Figure 3), the 4 main catechins plus important compounds from green tea such as gallic acid or alkaloids as theobromine or caffeine can be quantitatively determined.

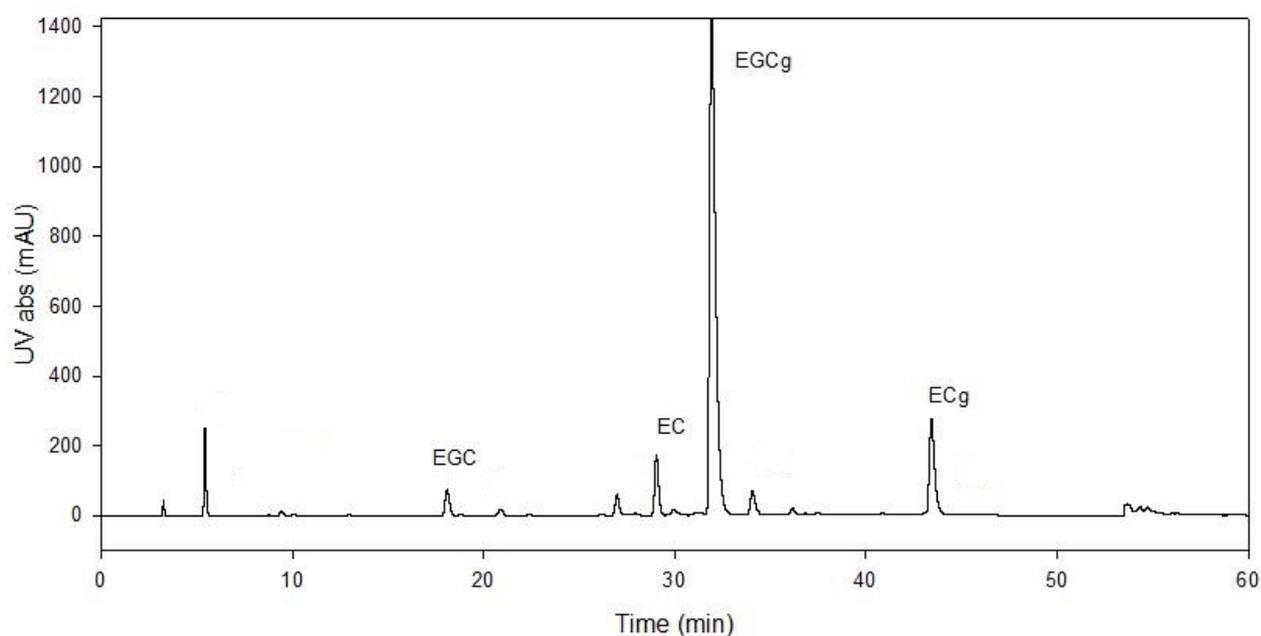


Figure 3. HPLC chromatogram of green tea catechins (gradient elution of water-acetonitrile).

6.2.4. FTIR Spectroscopy

Fourier Transform Infrared Spectroscopy was performed by using a Bruker IFS66 FT-IR Spectrometer with DTGS detector with KBr beam splitter. The spectra were recorded from 4000 cm^{-1} to 400 cm^{-1} wavenumbers with 128 scans and a resolution of 4 cm^{-1} . Samples were dried and pulverized as specified in literature²⁴.

6.2.5. Calculation of Number of sites

The resin used for the experiment is a polystyrene divinylbenzene (SDVB) resin. As Figure 4 shows, the divinylbenzene polymer (B) crosslinks the different oligomers of styrene (A) or in a lesser extent, forms an unresolved crosslink (C). A crosslink between two chains results in a benzene ring missing from one of the chains since the second vinyl group from the divinylbenzene takes the place of the styrene in that chain.

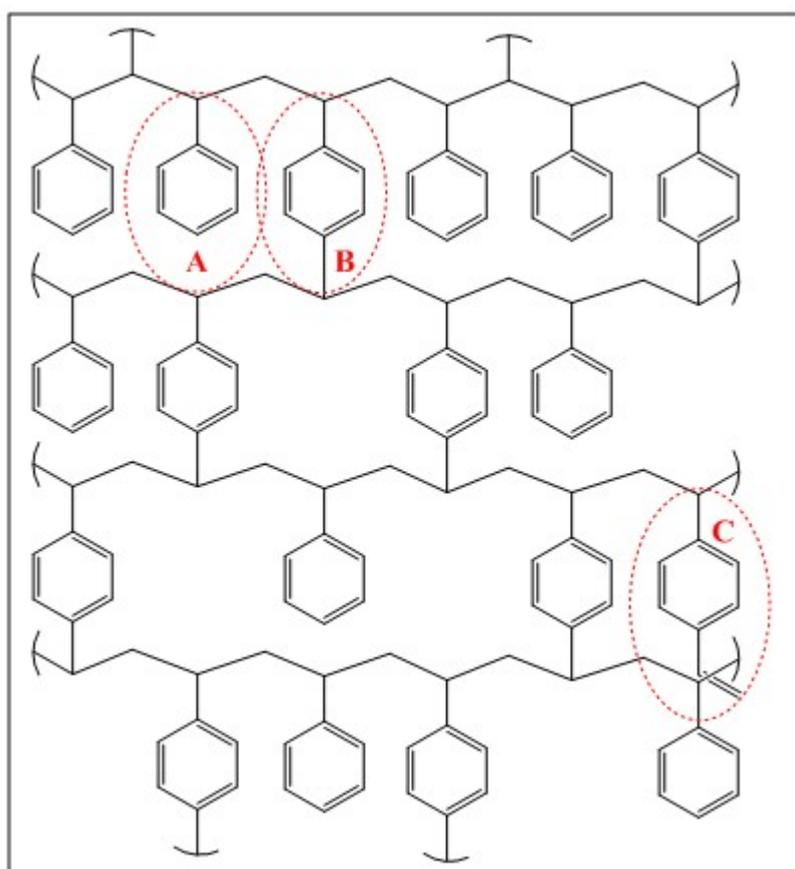


Figure 4: Schematic molecular structure of a cross-linked styrene-divinylbenzene (SDVB), where A and B/C represent the functional groups related to the styrene and divinylbenzene, respectively. While B acts as a cross-linker if both vinyl groups reacted, C represents divinylbenzene with one unreacted vinyl group.

The entire surface area ($S_a m_r$) can be divided in the contribution to the area by the styrene molecules ($S_{a,sty}$) plus the contribution to the area by the divinylbenzene molecules ($S_{a,DVB}$) as seen in Equation 14.

$$S_{a,sty} + S_{a,DVB} = S_a m_r \quad [14]$$

As observed, the area occupied by a molecule of styrene (A in Figure 4) is similar to the maximum projection of a molecule of divinylbenzene (A_{DVB}) while the area occupied by a molecule of divinylbenzene (B in Figure 4) is twice this value. Moreover, the amount of unresolved crosslinks (N_{UC}) does not occupy the same area as a divinylbenzene but as a styrene. This reasoning can be translated into Equation 15.

$$(N_{sty} + N_{UC} + 2(N_{DVB} - N_{UC})) A_{DVB} = S_a m_r \quad [15]$$

From the FTIR measurements the ratio between moles of styrene and unresolved crosslinks and moles of resolved divinylbenzene ($R_{S/DVB}$ in Equation 16) is known.

$$R_{S/DVB} = \frac{N_{sty} + N_{UC}}{N_{DVB} - N_{UC}} \quad [16]$$

When substituting Equation 16 into 15, Equations 17 and 18 are obtained.

$$N_{sty} + N_{UC} = \frac{R_{S/DVB} S_a m_r}{A_{DVB} (R_{S/DVB} + 2)} \quad [17]$$

$$N_{DVB} - N_{UC} = \frac{S_a m_r}{A_{DVB} (R_{S/DVB} + 2)} \quad [18]$$

And since both styrene and divinylbenzene have one aromatic ring and therefore, one adsorption site the amount of sites will be the addition of equation 17 and 18.

$$N_s = N_{sty} + N_{DVB} = \frac{S_a m_r}{A_{DVB}} \left(\frac{R_{S/DVB} + 1}{R_{S/DVB} + 2} \right) \quad [19]$$

Where $R_{S/DVB}$ is the ratio between styrene plus unresolved divinylbenzene and crosslinked divinylbenzene, A_{DVB} is the maximal projection area of the divinylbenzene monomer (Table 2), S_a is the specific surface area ($914 \text{ m}^2/\text{g}$)⁷ and m_r is the mass of resin used for the experiment.

6.2.6. Error Analysis

Two different types of errors were calculated: propagation error and parameter error.

Propagation error applies to experimental data while parameter error is calculated for regressed parameters.

For the calculation of the error propagation²⁵ Equation 20 has been used.

$$\sigma_U = \sqrt{\left(\frac{\partial U}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial U}{\partial y}\right)^2 \sigma_y^2 + \dots} \quad [20]$$

Where σ represents the standard error, U is the property that is being calculated, and x, y, etc. represent the different experimental measurements that this property is calculated from.

The errors related to regressed parameters were calculated by the MATLAB routine “nlparci”.

Optimization: The minimizing function for the fit of the Margules model (Equation 13) is shown in Equation 21.

$$F = \sum \left(\frac{\gamma_i^{s, \text{model}} - \gamma_i^{s, \text{exp}}}{\sigma_{\gamma_i^{s, \text{exp}}}} \right)^2 \quad [21]$$

And since the fit is for an activity coefficient model, the error of the fit can be calculated by Equation 22.

$$\text{Error (log units)} = \frac{1}{N} \sum \left| \ln \frac{\gamma_i^{s, \text{model}}}{\gamma_i^{s, \text{exp}}} \right| \quad [22]$$

Where N represents the total number of experiments.

The optimization routine was implemented and performed in MATLAB (Version 8.0.0.783 MathWorks, Inc.).

6.3. Results and discussion.

6.3.1. Calculation of number of sites

An established method for the calculation of the ratio of copolymers is by FTIR spectroscopy²⁴. Figure 5 shows the FTIR spectra for the resin FPX66. Following the method explained in literature²⁴ a value for $R_{S/DVB}$ equal to 0.558 is obtained. (Table 1)

Table 1: Chemical IR characterization of the resin Amberlite FPX66

Bond	Wavelength (cm^{-1})	Area of the peak	Milli-equivalents/g
Double bond	1630	1.570	0.234
p-benzene	838	7.931	0.769
m-benzene	795	6.245	0.699
Mono-substituted benzene	1493	2.958	0.456

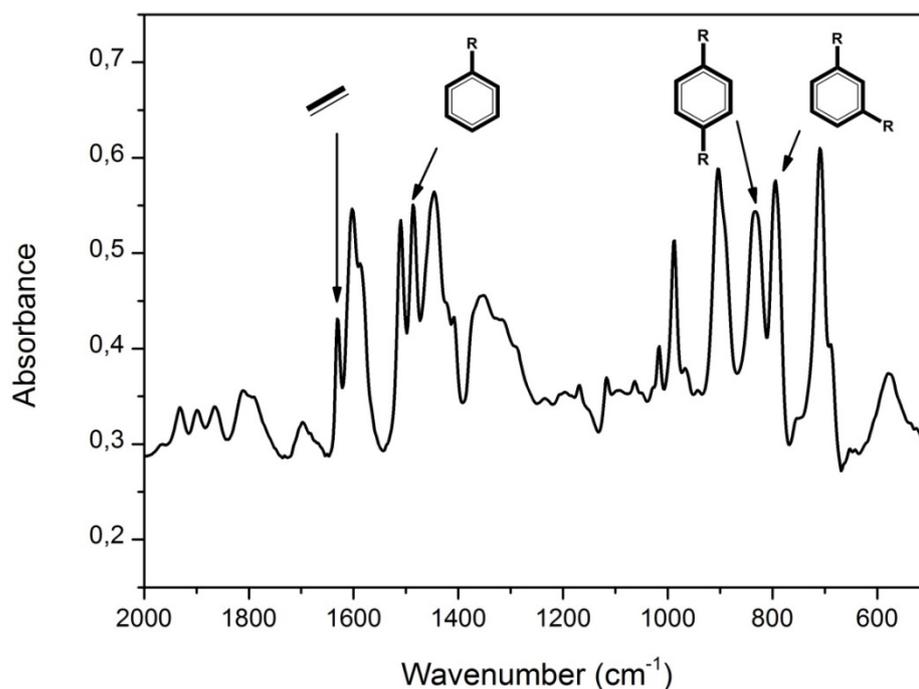


Figure 5: IR Spectrum of the resin FPX66.

6.3.2. Calculation of molecules per site:

Adsorption of the polyphenols onto the sites requires the full occupation of those sites and therefore, a factor (L_i) was added to Equation 4. The values of those factors can be found in Table 2.

Table 2: Comparison of projection areas of the different compounds of this work and a resin site.²⁶

Compound	Maximal projection Area (\AA^2)	Occupied sites per molecule	
		Exact	Considered
EC	77.1	1.41	2
EGC	81.5	1.49	2
ECg	98.5	1.80	2
EGCg	100.4	1.84	2
Water	9.7	0.178	0.2
Resin site	54.6	1.000	1

6.3.3. Activity coefficient in the interphase

By substituting Equation 17 into Equation 4, the molar fractions in the interphase can be estimated. Once known, unsymmetric activity coefficients are directly calculated by using Equation 6 and modelled by the Margules model described in Equation 13. The regressed parameters can be found in Table 2 and the fit can be appreciated in Figure 6. The error of the fit is 0.26 log units which is better than the typical error of widely used models like Mod-UNIFAC (0.5 log units), COSMO-RS (0.6 log units)²⁷ or NRTL-SAC (0.55 log units)²³.

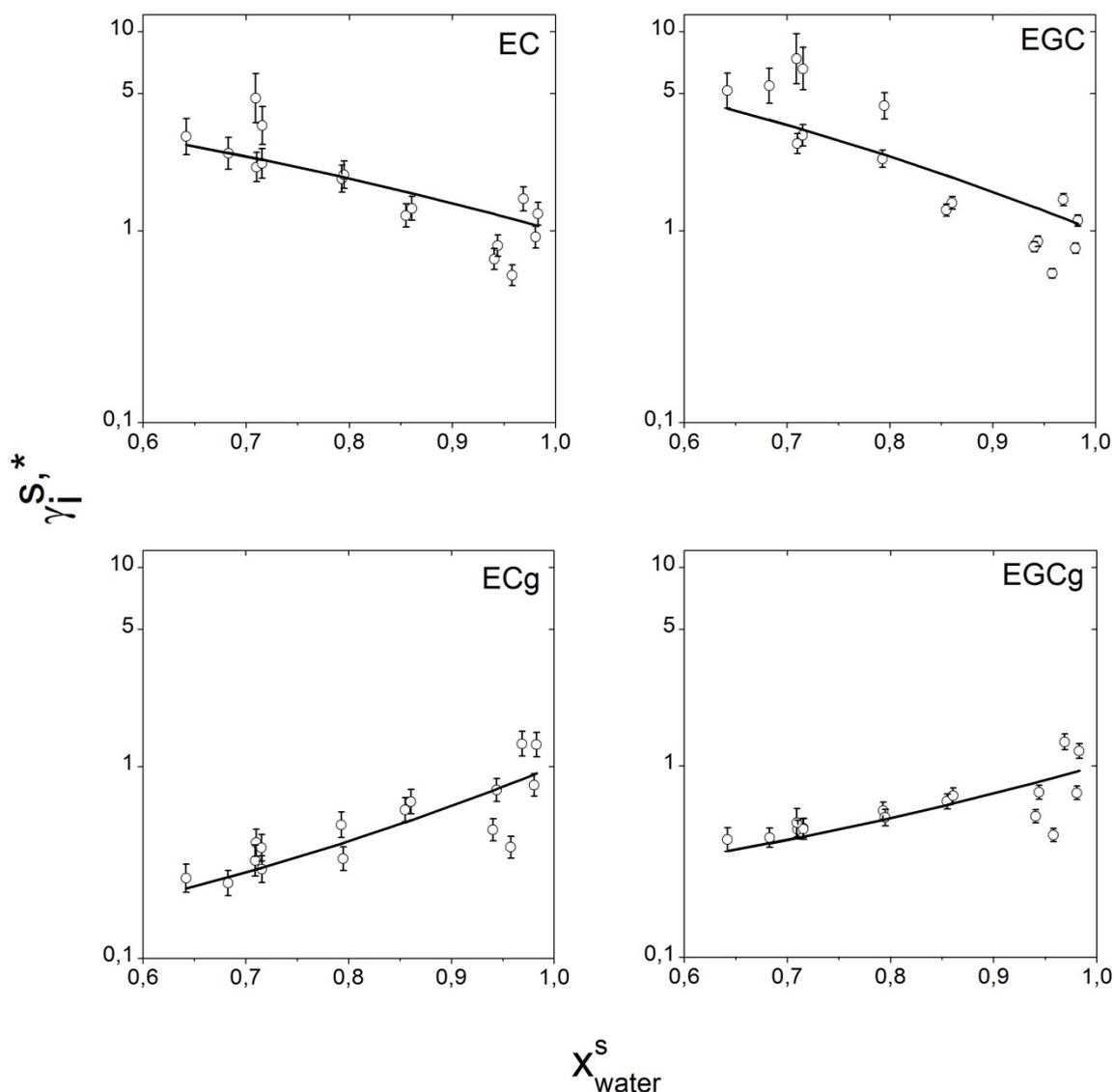


Figure 6: Unsymmetric activity coefficient as a function of the water molar fraction for the four catechins and fit of Equation 13 with the parameters of Table 2.

The fit of the model is good and the parameters regressed are accurate. There seem to be two different trends that coincide with the different types of catechins. The gallated catechins (ECg and EGCg) show increasing values of unsymmetric activity coefficients with increasing water molar fraction while the non-gallated ones (EC and EGC) show decreasing values with increasing water molar fractions. Moreover the extra hydroxyl group that the

epigallocatechins have (EGC and EGCg) increases slightly the values of activity coefficients in comparison to their counterparts (EC and ECg).

6.3.4. Equilibrium constants

Once the parameters for the activity coefficient model (A_i) in this system are regressed, they can be used for the calculation of the equilibrium constant (K_i) using Equation 3. A representation of the equilibrium constants as a function of the molar fraction of the compound is shown in Figure 7 and a weighted average of the values can be found in Table 3.

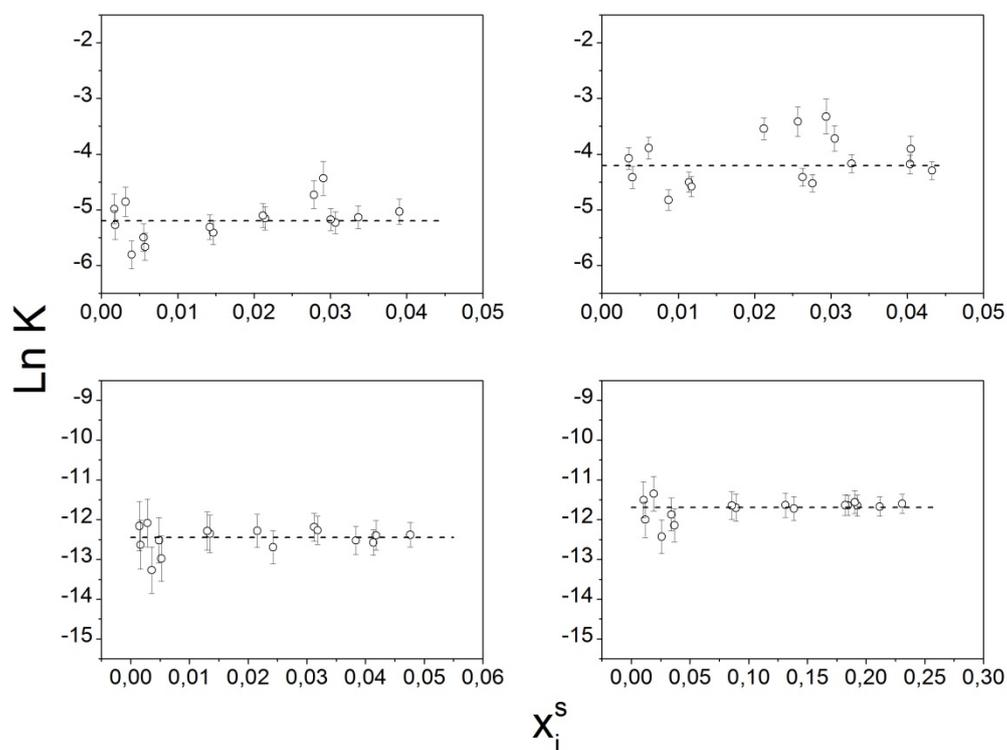


Figure 7: K_i values as a function of molar fraction of the compound. Dashed lines represent the weighted average.

Table 3: Margules coefficient and natural logarithm of the equilibrium constant for the four studied catechins.

	EC	EGC	ECg	EGCg
A_i (J/mol)	-1.73 ± 0.26	-2.45 ± 0.20	2.49 ± 0.63	1.74 ± 0.46
$\ln K_i$	-5.195 ± 0.075	-4.209 ± 0.097	-12.443 ± 0.063	-11.694 ± 0.050

As Figure 7 shows, the value of the equilibrium constants is always lower than one. This is very important since according to the definition of K_i (Equation 3) the lower the value of K_i the is the higher the affinity for the resin is. For a K_i equal to one, there would be an equal distribution between phases.

6.3.5. Predictive power of the model

The main driving force in the adsorption of these compounds is most likely hydrophobicity and therefore, it is to be expected that the regressed and calculated thermodynamic parameters (A_i and K respectively) are related to the hydrophobicity of the molecules involved. A commonly used measure of the hydrophobicity is the so-called logP. LogP indicates the logarithm of the partition coefficient of a neutral substance between two phases: a water phase and a 1-octanol phase. However, depending on the pH, the molecules of interest can be charged which affects the partition coefficient. LogD measures the distribution of a certain compound at a certain pH between water and 1-octanol. Therefore, logD is equal to logP for the pH at which the compound is not charged. Figure 8 shows the correlation between the Margules constant (A_i) and the natural logarithm of the equilibrium constant ($\ln K_i$) and the logD of the four catechins used for this study (Table 4).

Table 4: logD of the four main catechins in tea²⁶.

	EC	EGC	ECg	EGCg
logD (pH=4)	1.81	1.49	3.38	3.08

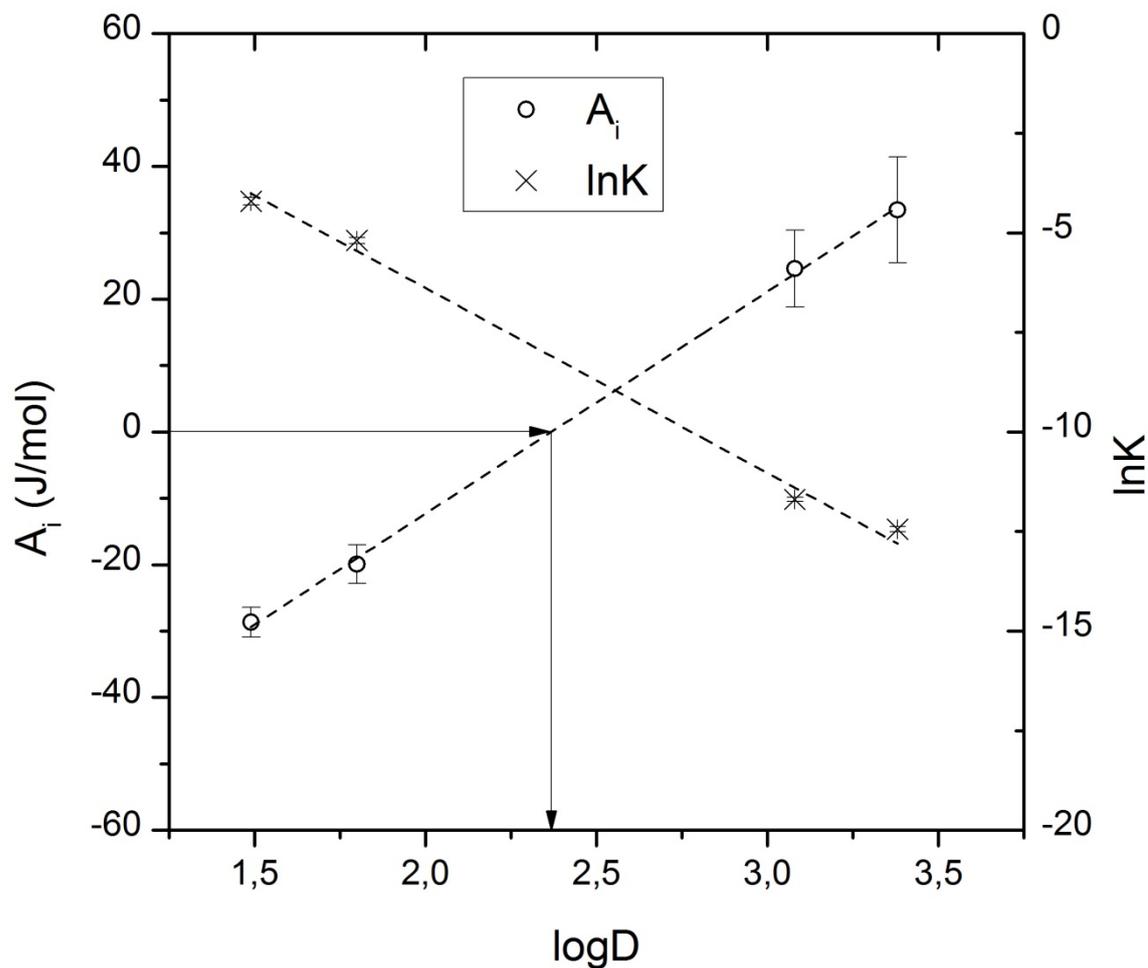


Figure 8: Correlation between A_i and $\log D$ plus $\ln(K_i)$ and $\log D$.

As seen in Figure 8 both parameters are linear with respect to the $\log D$. The R^2 values are 0.9989 in the case of A_i and 0.9881 in the case of $\ln K_i$. These values are so close to one that it can be safely assumed that hydrophobicity is the prominent driving force. Moreover, an interpolation for the case of A_i yields a value of $\log D$ equal to 2.37 for which A_i is equal to zero. In this particular scenario, a null A_i indicates that the activity coefficient is equal to one for any given concentration and therefore behaves ideally. Ideal behaviour in this context means that the distribution of a certain molecule in a phase is governed by randomness since there is no interaction between that molecule and the rest of the molecules around it. Ideal

mixtures tend to occur with very similar compounds (i.e. benzene/toluene) and therefore this $\log D$ for which A_i is zero might coincide with the theoretical $\log D$ of the resin. However, more experiments are needed to further confirm this hypothesis, but falls beyond the scope of this paper and will be content of future work in this area.

6.4. Conclusions

Adsorption experiments were performed with tea catechins in the macroporous polymeric resin Amberlite FPX66. A rigorous thermodynamic study was conducted whereby the resin surface was divided into sites or lattices and the molar fraction in the adsorbed phase was defined as the ratio between occupied and total lattices. Experimental values of unsymmetric activity coefficients in the adsorbed phase were modelled by the Margules model and interaction parameters (A_i) were regressed with good accuracy. A trend is observed whereby the activity coefficients of the gallate versions of the catechins (ECg and EGCg) decrease while the activity coefficient of the non-gallate versions (EC and EGC) increase upon increasing water molar fraction.

Once the molar fractions and activity coefficients of the compounds in both phases were known, the equilibrium constants are calculated. All K_i show values lower than one which is an indication of an efficient adsorption. Moreover, a linear relation was discovered between the $\log D$ of the compounds and both A_i and the natural logarithm of K_i . Due to the high R^2 of the correlation, it can be assumed that hydrophobicity is the only driving force for the adsorption.

Acknowledgements

This work has been carried out within the framework of the Institute for Sustainable Process Technology (ISPT) in the project FO-10-03.

References

- (1) Hanke, A. T.; Ottens, M., Purifying biopharmaceuticals: knowledge-based chromatographic process development. *Trends Biotechnol.* **2014**, *32*, 4, 210.
- (2) López-Garzón, C. S.; van der Wielen, L. A. M.; Straathof, A. J. J., Green upgrading of succinate using dimethyl carbonate for a better integration with fermentative production. *Chem. Eng. J.* **2014**, *235*, 0, 52.
- (3) Bailey, S. E.; Olin, T. J.; Bricka, R. M.; Adrian, D. D., A review of potentially low-cost sorbents for heavy metals. *Water Res.* **1999**, *33*, 11, 2469.
- (4) Dillard, C. J.; German, J. B., Phytochemicals: nutraceuticals and human health. *J. Sci. Food Agric.* **2000**, *80*, 12, 1744.
- (5) Jeon, S.-M.; Kim, H. K.; Kim, H.-J.; Do, G.-M.; Jeong, T.-S.; Park, Y. B.; Choi, M.-S., Hypocholesterolemic and antioxidative effects of naringenin and its two metabolites in high-cholesterol fed rats. *Transl. Res.* **2007**, *149*, 1, 15.
- (6) Cabrera, C.; Artacho, R.; Giménez, R., Beneficial Effects of Green Tea—A Review. *J. Am. Coll. Nutr.* **2006**, *25*, 2, 79.
- (7) Sevillano, D. M.; van der Wielen, L. A. M.; Hooshyar, N.; Ottens, M., Resin Selection for the Separation of Caffeine from Green Tea Catechins. *Food and Bioprocess Processing* **2014**, *92*, 2, 192.
- (8) Huang, J.; Huang, K.; Liu, S.; Luo, Q.; Xu, M., Adsorption properties of tea polyphenols onto three polymeric adsorbents with amide group. *J. Colloid Interface Sci.* **2007**, *315*, 2, 407.
- (9) Yamamoto, S.; Hakoda, M.; Oda, T.; Hosono, M., Rational method for designing efficient separations by chromatography on polystyrene–divinylbenzene resins eluted with aqueous ethanol. *J. Chromatogr. A* **2007**, *1162*, 1, 50.
- (10) Zhao, R.; Yan, Y.; Li, M.; Yan, H., Selective adsorption of tea polyphenols from aqueous solution of the mixture with caffeine on macroporous crosslinked poly(N-vinyl-2-pyrrolidinone). *Reactive and Functional Polymers* **2008**, *68*, 3, 768.
- (11) Bretag, J.; Kammerer, D. R.; Jensen, U.; Carle, R., Adsorption of rutin onto a food-grade styrene-divinylbenzene copolymer in a model system. *Food Chem.* **2009**, *114*, 1, 151.
- (12) Kammerer, J.; Kammerer, D. R.; Jensen, U.; Carle, R., Interaction of apple polyphenols in a multi-compound system upon adsorption onto a food-grade resin. *J. Food Eng.* **2010**, *96*, 4, 544.
- (13) Ku, Y.; Lee, K.-C., Removal of phenols from aqueous solution by XAD-4 resin. *J. Hazard. Mater.* **2000**, *80*, 1–3, 59.
- (14) Larionov, O. G.; Myers, A. L., Thermodynamics of adsorption from nonideal solutions of nonelectrolytes. *Chem. Eng. Sci.* **1971**, *26*, 7, 1025.
- (15) Sircar, S.; Myers, A. L., A thermodynamic consistency test for adsorption from binary liquid mixtures on solids. *AIChE J.* **1971**, *17*, 1, 186.
- (16) Minka, C.; Myers, A. L., Adsorption from ternary liquid mixtures on solids. *AIChE J.* **1973**, *19*, 3, 453.

- (17) Chen, C.-C.; Crafts, P. A., Correlation and Prediction of Drug Molecule Solubility in Mixed Solvent Systems with the Nonrandom Two-Liquid Segment Activity Coefficient (NRTL–SAC) Model. *Ind. Eng. Chem. Res.* **2006**, *45*, 13, 4816.
- (18) Xue, Z.; Mu, T.; Gmehling, J., Comparison of the a Priori COSMO-RS models and group contribution methods: Original UNIFAC, modified UNIFAC(Do), and modified UNIFAC(Do) consortium. *Industrial and Engineering Chemistry Research* **2012**, *51*, 36, 11809.
- (19) Fredenslund, A.; Gmehling, J.; Michelsen, M. L.; Rasmussen, P.; Prausnitz, J. M., Computerized design of multicomponent distillation columns using the UNIFAC group contribution method for calculation of activity coefficients. *Industrial and Engineering Chemistry Process Design and Development* **1977**, *16*, 4, 450.
- (20) Méndez Sevilano, D.; van der Wielen, L. A. M.; Hooshyar, N.; Ottens, M., MPP-UNIFAC, a predictive activity coefficient model for polyphenols. *Fluid Phase Equilib.* **2014**, *384*, 82.
- (21) Huxham, I. M.; Rowatt, B.; Sherrington, D. C.; Tetley, L., Molecular architectural changes in hydrated macroporous styrene-divinylbenzene resin sorbents revealed by transmission electron microscopy using image analysis. *Polymer* **1992**, *33*, 13, 2768.
- (22) Prausnitz, J. M., *Molecular Thermodynamics of Fluid-Phase Equilibria*. Prentice-Hall: Englewood Cliffs, New Jersey, 1969.
- (23) Méndez Sevilano, D.; Van Der Wielen, L. A. M.; Trifunovic, O.; Ottens, M., Model comparison for the prediction of the solubility of green tea catechins in ethanol/water mixtures. *Industrial and Engineering Chemistry Research* **2013**, *52*, 17, 6039.
- (24) Bartholin, M.; Boissier, G.; Dubois, J., Styrene-divinylbenzene copolymers, 3. Revisited IR analysis. *Die Makromolekulare Chemie* **1981**, *182*, 7, 2075.
- (25) Worthing, A. G. G., Joseph, Treatment of Experimental Data. *J. Chem. Educ.* **1943**, *20*, 11, 572.
- (26) <http://www.chemicalize.org> (27-08-2014),
- (27) Kan, A. T.; Tomson, M. B., UNIFAC prediction of aqueous and nonaqueous solubilities of chemicals with environmental interest. *Environ. Sci. Technol.* **1996**, *30*, 4, 1369.

Chapter 7: Outlook

This chapter tries to show what has been learned from the work described in this thesis and what are the logical steps that follow from that knowledge.

Group-contribution (GC) methods are a very powerful tool that even though have been around for more than 100 years, they are still used throughout literature and industry. An important realization in this work was that the effect of neighbouring groups might play a bigger role when GC models are applied to medium sized molecules such as polyphenols. That would also explain why when reducing the dataset to regress parameters from to only polyphenols, GC models give a better prediction (**Chapter 3**) than with the much broader original dataset (**Chapter 2**). Therefore the addition of a term that accounts for distance between groups is key in order to increase the accuracy of GC models like UNIFAC for medium size molecules (or even bigger size biomolecules).

As well, throughout this work, logD has proven to be a really powerful tool to characterize a molecule. Important thermodynamic properties have been directly correlated to it, like the interaction parameter in the solid activity coefficient model (**Chapter 6**) or the equilibrium constant between bulk and resin (**Chapters 5 and 6**). The logical next step is to validate the observed relation between logD and equilibrium parameters such as the equilibrium constant or the Margules parameter and therefore, develop a fully predictive tool able to produce accurate isotherm parameters with the sole input of the structural formula of the molecule.

To conclude, the most important of the different points raised in this chapter. The main aim of this work was to develop tools to predict the behaviour of polyphenols in a process and the most important next step is to use those tools that have been developed. Polyphenol purification processes can be better optimized at a large scale but with the rise in demand of

superfoods in the last decade and the cheapness of the raw material (polyphenols are present in waste streams from the food industry) there is room for the implementation of a knowledge-based design and optimization of the downstream processing. This work gives a collection of tools that can be used for process design and a brief summary of how to use them (**Chapter 1**) and the implementation of these tools to the design should yield on efficient and cheap processes.

Summary

The increasing population demands increasing food supply with the consequent increase in food waste. Moreover, not only the quantity of the demanded food increases but so does the quality. This situation poses an interesting challenge to the food industry that is moving towards the revalorisation of the food waste and, as a part of it, towards the recovery of nutritional compounds that provide health benefits (nutraceuticals) still present in those streams.

However, those streams possess very complex chemical and physical characteristics, namely multiple phases, ionic and hydrophobic interactions, etc. Moreover, the concentration of these nutraceuticals tends to be low, making regular processes costly and inefficient if applied to these cases. Therefore, it is extremely important to have models that describe the behaviour of the different components of the system to be able to design and develop a feasible process to recover these molecules.

The objective of this research, therefore, was to apply or develop and validate models that could predict important properties for the design of a downstream process in each of the cases.

The main focus of the work was on green tea catechins and soy isoflavones. From the beginning, their molecular similarities made clear that Group Contribution (GC) models were the obvious way to reduce complexity. Therefore solubility experiments of the four main catechins (Chapter 2) were performed in order to validate a well know GC activity coefficient model such as Mod. UNIFAC (amongst other models). The validation proved unsuccessful since Mod. UNIFAC did not give good predictions for the interaction between the molecules and water.

Activity coefficients are a key property in design since they account for the non-idealities in the system which made the development of a suitable activity coefficient an

important part of the research. An extensive bibliographic work was done and solubility values of flavonoids (a class of polyphenols) were gathered in a database that was used for the regression of new interaction parameters for the Mod-UNIFAC model with a new definition of groups more applied to polyphenols. Those parameters were regressed and validated in what is called Modified PolyPhenols-UNIFAC or MPP-UNIFAC (Chapter 3).

After having a good understanding of the behaviour of the polyphenols in solution, the next logical step was to model their behaviour in equilibrium with a resin since Solid-Liquid Extraction is one of the most used unit operations. A set of food-grade resins was tested for their suitability to separate catechins from caffeine using green tea as a starting material (Chapter 4). A model-based Design of Experiments was used in order to perform the needed experiments to minimize the uncertainty of the regressed parameters which were further used as criteria for a resin selection.

A similar approach was used for the okara. However since okara is a solid matrix, a pretreatment was needed in order to bring the polyphenols to solution. This pretreatment consisted of a solvent extraction with different ratios of ethanol and water. After testing the behaviour of the different extracts with the resin, it was decided to further study the water extract since it maximized the overall yield. Adsorption isotherms were measured (Chapter 5) showing an almost linear behaviour, typical of small concentrations. In some cases it was observed a non-linearity that was hypothesized as a bilayer adsorption of polyphenols onto adsorbed proteins. This hypothesis was then proved for the interaction of medium-hydrophobic isoflavones and highly hydrophobic resins. Moreover, a linear relation was discovered between the logarithm of the slope of the isotherm and the logD value of the compound, pointing into possible predictability of the isotherms for isoflavones.

This relationship was further studied for higher concentrations in which the isotherms do not behave linearly. A lattice model was developed to calculate the molar fraction of the

compounds in the adsorbed phase in contact with the resin (Chapter 6). This allowed for the experimental calculations of activity coefficients in the adsorbed phase that were modelled with a simple activity coefficient model (Margules). As in previous chapter, a linear relation was discovered between the logarithm of the equilibrium constant and the logD value. Moreover, the regressed interaction parameters showed as well a linear correlation with the logD values of the compounds in equilibrium.

This way a “toolbox” of models is established and validated for the prediction of soy and tea polyphenol properties. These models can be readily used for the design of a separation process or to gain insight and develop an integrated process. Moreover, because of the nature of the models, either group contribution or mechanistic, this toolbox is easily applicable to different polyphenols from different sources.

Samenvatting

De toenemende bevolking eist toenemende voedselvoorziening met de daaruit voortvloeiende toename van de voedselverspilling. Bovendien zijn niet alleen de hoeveelheid van het gevraagde voedsel toeneemt maar ook de kwaliteit. Deze situatie vormt een interessante uitdaging voor de voedselindustrie die tendeeert naar herwaardering van het voedselafval en, als een deel ervan, naar het herstel van nutritionele verbindingen die gezondheidsvoordelen (nutraceuticals) nog in deze stromen leveren.

Echter, deze stromen bezitten zeer complexe chemische en fysische kenmerken, namelijk meerdere fasen, ionische en hydrofobe interacties, etc. Bovendien, de concentratie van deze nutraceuticals neiging laag te zijn, waardoor normale processen kostbaar en inefficiënt indien toegepast op deze gevallen. Daarom is het uiterst belangrijk om modellen die het gedrag van de verschillende componenten van het systeem beschrijven kunnen ontwerpen en ontwikkelen van een haalbare werkwijze om deze moleculen te verhalen.

Het doel van dit onderzoek was dan ook toe te passen of te ontwikkelen en te valideren modellen die belangrijke eigenschappen voor het ontwerp van een stroomafwaarts proces in elk van de gevallen kon voorspellen.

De belangrijkste focus van het werk was op groene thee catechinen en soja-isoflavonen. Vanaf het begin, hun moleculaire gelijkenissen duidelijk gemaakt dat de Groep Bijdrage (GC) modellen was de voor de hand liggende manier om de complexiteit te verminderen. Daarom oplosbaarheid experimenten van de vier catechinen (hoofdstuk 2) werden uitgevoerd om een bekende GC activiteitscoëfficiënt model te valideren zoals Mod. UNIFAC (onder andere modellen). De validatie bleek geen succes sinds Mod. UNIFAC niet goed voorspellingen voor de interactie tussen de moleculen en water te geven.

Activiteit coëfficiënten zijn een belangrijke eigenschap in design omdat zij goed zijn voor de niet-idealiteiten in het systeem dat de ontwikkeling van een geschikte

activiteitscoëfficiënt een belangrijk deel van het onderzoek gedaan. Een uitgebreide bibliografische gewerkt en oplosbaarheid waarden van flavonoïden (een klasse van polyfenolen) werden verzameld in een database die werd gebruikt voor de regressie van nieuwe interactie parameters voor de Mod-UNIFAC model met een nieuwe definitie groepen meer toegepast om polyfenolen. Deze parameters werden regressie en gevalideerd in wat genoemd wordt gewijzigd polyfenolen-UNIFAC of MPP-UNIFAC (hoofdstuk 3).

Na een goed inzicht in het gedrag van de polyfenolen in oplossing de volgende logische stap was hun gedrag evenwichtsmodel met een hars aangezien Solid-vloeistof extractie is een van de meest gebruikte eenheidsprocessen. Een set van food-grade harsen werd getest op hun geschiktheid om catechines van cafeïne te scheiden met behulp van groene thee als uitgangsmateriaal (hoofdstuk 4). Een modelgebaseerde Design of Experiments werd gebruikt om de benodigde experimenten de onzekerheid van de regressie parameters die verder werden gebruikt als criteria voor een hars selectie minimaliseren.

Een soortgelijke werkwijze werd toegepast voor de okara. Maar omdat okara een vaste matrix, is een voorbehandeling om de polyfenolen tot oplossing te brengen nodig. Deze voorbehandeling bestond uit een oplosmidelextractie met verschillende verhoudingen van ethanol en water. Na het testen van het gedrag van de verschillende extracten met de hars werd besloten om verdere studie het water extract gemaximaliseerd omdat de totale opbrengst. Adsorptie-isothermen werden gemeten (hoofdstuk 5) dat een nagenoeg lineair gedrag, typisch kleine concentraties. In sommige gevallen werd waargenomen niet-lineariteit die werd verondersteld als een dubbellaag adsorptie polyfenolen op geadsorbeerde eiwitten. Deze hypothese werd vervolgens bleek voor de interactie van middelgrote hydrofobe isoflavonen en zeer hydrofobe harsen. Bovendien werd een lineaire relatie ontdekt tussen de logaritme van de helling van de isotherm en logD waarde van de verbinding, wijzend naar mogelijke voorspelbaarheid van de isothermen isoflavonen.

Deze relatie werd verder onderzocht hogere concentraties waarin de isothermen niet lineair gedragen. Een rooster model ontwikkeld om de molaire fractie van de verbindingen berekenen de geadsorbeerde fase in contact met de hars (hoofdstuk 6). Dit is toegestaan voor de experimentele berekeningen van de activiteit coëfficiënten in de geadsorbeerde fase die werden gemodelleerd met een eenvoudige activiteit coëfficiënt model (Margules). Zoals in het vorige hoofdstuk, werd een lineair verband ontdekt tussen de logaritme van de evenwichtsconstante en de logD waarde. Bovendien is de regressie interactie parameters vertoonden ook een lineair verband met de logD waarden van de verbindingen in evenwicht.

Op deze manier een "gereedschapskist" van modellen is vastgesteld en gevalideerd voor de voorspelling van soja en thee polyfenolen eigenschappen. Deze modellen kunnen gemakkelijk worden gebruikt voor het ontwerp van een scheidingsproces of inzicht en ontwikkelen van een geïntegreerd proces. Bovendien, vanwege de aard van de modellen, hetzij contributiegroep of mechanistisch Deze toolbox is gemakkelijk toepasbaar op andere polyfenolen uit verschillende bronnen.

Curriculum Vitae

David Méndez Sevillano was born on the 31st of January 1985 in San Andrés del Rabanedo, Spain. After completing Elementary and High School in León (1992-2003) he moved to Valladolid to start his studies in Chemical Engineering at the University of Valladolid (2003-2009) where he focussed on courses related to downstream processes, design and thermodynamics. During the duration of the degree, he did a research project about modelling solubility of gases in ionic liquids. The Master Thesis (2009-2010), entitled “Study of the Resolution of Racemic Ibuprofen with (R)-1- Phenylethylamine in a Supercritical CO₂ Media” was performed in the BME university in Budapest under the supervision of Dr. Edit Székely and Prof. Bela Simándi. As a result of both works, three articles were published.



From November 2010, he started a PhD in the Technical University of Delft in the BST (later BPE) group with Prof. dr. ir. Luuk van der Wielen as promotor and Dr. ir. Marcel Ottens as co-promotor. As a part of his PhD he worked on the development and application of both correlative and predictive models for equilibrium properties of tea and soy polyphenols which is presented in this thesis.

At present he is working as a postdoctoral research fellow at the department of Biotechnology, Technological University of Delft, The Netherlands. His research deals with preparative scale separation and fractionation of polyphenols from complex mixtures.

Publications and conferences

Publications

Méndez Sevillano, D., Van Der Wielen, L. A. M., Trifunovic, O., Ottens, M. (2013) "Model comparison for the prediction of the solubility of green tea catechins in ethanol/water mixtures." Industrial and Engineering Chemistry Research 52(17): 6039-6048.

Méndez Sevillano, D., van der Wielen, L. A. M., Hooshyar, N., Ottens, M., (2014). "MPP-UNIFAC, a predictive activity coefficient model for polyphenols." Fluid Phase Equilibria 384(0): 82-88.

Sevillano, D. M., van der Wielen, L. A. M., Hooshyar, N., Ottens, M., (2014). "Resin Selection for the Separation of Caffeine from Green Tea Catechins." Food and Bioprocess Processing 92(2): 192-198.

Méndez Sevillano, D., Jankowiak, L., van Gaalen, T. L. T., van der Wielen, L. A. M., Hooshyar, N., van der Goot, A. J., Ottens, M., (2014) "Mechanism of Isoflavone Adsorption from Okara Extracts onto Food-Grade Resins." Industrial & Engineering Chemistry Research 53(39): 15245-15252.

Méndez Sevillano D., Sachdeva, S., van der Wielen, L.A.M., Hooshyar N., de Smet, L.C.P.M., Ottens, M., "A Thermodynamic Lattice Model Describing Adsorption of Phytochemicals onto Macroporous Polymeric Resins" Langmuir, *submitted*

Chilamkurthi, S., **Sevillano, D. M.**, Albers, L. H. G., Sahoo, M. R., Verheijen, P. J. T., van der Wielen, L. A. M., den Hollander, J. L., Ottens, M., (2014) "Thermodynamic description of peptide adsorption on mixed-mode resins." Journal of Chromatography A 1341(0): 41-49.

Jankowiak, L., **Méndez Sevillano, D.**, Boom, R., Ottens, M., Zondervan, E., van der Goot, A.J., (2015) "A process synthesis approach for isolation of isoflavones from okara" Industrial and Engineering Chemistry Research, 54(2) 691-699

Conferences

Méndez Sevillano, D., van de Wielen, L.A.M., Trifunovic, O., Ottens, M., (2011) "Separation of Vitality ingredients: Physicochemical and Thermodynamic Properties of Tea and Soy Polyphenols" NPS 11, Arnhem, The Netherlands, (Poster Presentation)

Méndez Sevillano, D., van de Wielen, L.A.M., Trifunovic, O., Ottens, M., (2012) "Physicochemical and Thermodynamic Properties of Tea and Soy Polyphenols" ACS, San Diego, USA, (Poster Presentation)

Méndez Sevillano, D., van de Wielen, L.A.M., Trifunovic, O., Ottens, M., (2012) "Physicochemical and Thermodynamic Properties of Tea and Soy Polyphenols" NBC 14, Ede, The Netherlands, (Poster Presentation)

Méndez Sevillano, D., van de Wielen, L.A.M., Trifunovic, O., Ottens, M., (2012)
“Physicochemical and Thermodynamic Properties of Tea and Soy Polyphenols” ESBES-
ISPPP, Istanbul, Turkey, (Poster Presentation)

Méndez Sevillano, D., van de Wielen, L.A.M., Trifunovic, O., Ottens, M., (2013)
“Physicochemical and Thermodynamic Properties of Tea and Soy Polyphenols” 9th PhD
Seminar on Chromatographic Separation Science, Weggis, Switzerland (Oral Presentation)

Méndez Sevillano, D., van de Wielen, L.A.M., Trifunovic, O., Ottens, M., (2013)
“Thermodynamic approach for the prediction of properties of tea polyphenols” ECCE-
ECAB, The Hague, The Netherlands, (Poster Presentation)

Méndez Sevillano, D., van de Wielen, L.A.M., Trifunovic, O., Ottens, M., (2013)
“Thermodynamic approach for the prediction of properties of tea polyphenols” DPTI Event,
The Hague, The Netherlands, (Poster Presentation)

Méndez Sevillano, D., Jankowiak, L., van Gaalen, T. L. T., van der Wielen, L. A. M.,
Hooshyar, N., van der Goot, A. J., Ottens, M., (2014) “Selection of resins for purification of
antioxidants from waste Streams” 10th PhD Seminar on Chromatographic Separation Science,
Egmond aan Zee, The Netherlands (Oral Presentation)

Méndez Sevillano, D., Jankowiak, L., van Gaalen, T. L. T., van der Wielen, L. A. M.,
Hooshyar, N., van der Goot, A. J., Ottens, M., (2014) “Selection of resins for purification of
antioxidants from waste Streams” NBC 15, Ede, The Netherlands (Poster Presentation)

Méndez Sevillano, D., van de Wielen, L.A.M., Hooshyar, N., Ottens, M., (2014)
“Thermodynamic approach for the prediction of properties of tea polyphenols” ESBES-
IFIBIOP, Lille, France, (Oral Presentation)

Acknowledgements / Dankwoord

Yes, where to begin with? A great deal of thanks to all the people I have crossed paths with and have influenced me into the decision of starting a PhD. This is a very long list but I want to emphasize two names on it: Choni my chemistry teacher in High School and Fidel Mato Chaín who started my passion for Thermodynamics in the University of Valladolid.

And now, I change briefly language to thank my family. Quería agradecer a mi abuela, a mis padres y a mi hermana por todo el apoyo y aceptación que han tenido y porque me han demostrado que la familia siempre está cerca sin importar las distancias. En ook, nu in het Nederlands, hartelijk bedankt, Duijndam familie. Jullie zijn mijn familie geworden in een land zo ver van thuis.

Gracias, muchísimas gracias, a todos los amigos de León, Valladolid y Budapest que me han hecho sentir menos bicho raro con todas sus em... excentricidades. I will not attempt to write down a list of names because it would easily take me at least another chapter of this already bulky enough thesis.

Once thanked the people that helped me arrive to the beginning of this experience, it is time to thank the people that helped me through it. First and foremost, Marcel Ottens. You introduced me to this new world that Biotechnology was (and sometimes still is) to me. With your supervision I have learned through successes and (many) mistakes a great deal of the self-confidence that I have nowadays. Thanks for guiding me when needed and trust me when proposing new ideas. Luuk, thanks for the scientific talks. Especially thanks for keeping me with my feet on the ground when the list of things I wanted to do kept growing exponentially.

Staying in the topic of supervision, many thanks to both Olivera Trifunovic and Nasim Hooshyar for your invaluable support through this project. Both of you knew how to be there when needed and give a very much appreciated feedback. Thanks to Peter Verheijen for putting into words what I not always wanted to hear. From you I have learnt that not everything is black or white and that the size of the grey depends on your confidence interval.

Thanks, many thanks, to all the colleagues I have had the pleasure to work with: María, Adri, Max and Stef, Paz, Jovana, Corjan, Arjan Oudshoorn, Frieder, Camilo, Sushil, Beckley, Sreekanth, Susana, Stefano, Silvia, Carlos, Joana, Victor and the countless other people from the building (PhD's, staff, PDEngs, students) that made Mondays a bit more bearable. To all the party room members: Lodewijk, Kedar, Katelijne, Janneke, Robin, Shima and Marcelo, it's been a pleasure to share room with you guys. To Thomas, my one and only

kid during the PhD, thanks for reminding me of how much I like teaching. I am pretty sure I learned from you just as much as you learned from me. To Pilar, Mark, Victor, Lutz and Eleni thanks for the time and conversations in the lab.

Thanks too, to Lena, Miguel and Andrea, thanks for sharing the trip that this project was, for the talks and phone/skype conversations, the lunches in Utrecht, the long emails, I hope you guys have a great life. To Hugo ¡¡hay esperanza al final del túnel, paisano!!, to Susana, you are the man of this family now, take good care of all the kids. To Arjan, you have become one of the people I find most dear, I hope to keep our friendship wherever our paths take us. To Alex, my work wife, I certainly would not have made it this far without the shared cubic meters of coffee we have drunk together. Thanks, very much thanks for all the shared complaining.

Last but certainly not least, Jelle. Thanks for keeping up with the crazy schedules, the senseless drama and the language chaos that my life entails without giving them for granted and constantly trying to put some common sense in this cazurro. During these four years you have represented the concept of home for me more than any building, city or country. You are the person I laugh the most with and I love you for that.

A last round of thanks to those giants on which shoulders I have been standing thus far and without which, nothing of this would have been possible.