PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE

ESCUELA DE INGENIERIA

USING MOLECULAR DYNAMICS SIMULATIONS TO PREDICT AQUEOUS SOLUBILITIES OF AROMATIC COMPOUNDS

RAIMUNDO GILLET BOUCHON

Thesis submitted to the Office of Research and Graduate Studies in partial fulfillment of the requirements for the Degree of Master of Science in Engineering

Advisors:

JOSÉ R. PÉREZ CORREA

LORETO M. VALENZUELA ROEDIGER

Santiago de Chile, July, 2017

PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
ESCUELA DE INGENIERIA

USING MOLECULAR DYNAMICS SIMULATIONS TO PREDICT AQUEOUS SOLUBILITIES OF AROMATIC COMPOUNDS

RAIMUNDO GILLET BOUCHON

Members of the Committee:

JOSÉ RICARDO PÉREZ CORREA

LORETO MARGARITA VALENZUELA ROEDIGER

ROBERTO CANALES MUÑOZ

ANGÉLICA MARIA FIERRO HUERTA

IGNACIO LIRA CANGUILHEM

Thesis submitted to the Office of Research and Graduate Studies in partial fulfillment of the requirements for the Degree of Master of Science in Engineering

Santiago de Chile, July, 2017

Dedicated to you, but you weren't listening

ACKNOWLEDGEMENTS

I am grateful firstly to my family for the permanent support and encouragement they have provided me.

I am also deeply thankful to my uncle Pedro Bouchon, whose advices were a permanent guide through this process.

To my guide professors, Loreto Valenzuela and Ricardo Pérez, who introduced me to the world of investigation as an undergraduate student, and kept guiding me to these days.

To Angelica Fierro, since the majority of this work was carried out at her laboratory. Furthermore, she taught me the principles of molecular dynamics and was always open for thoughtful conversations.

I thank José Cuevas for his advice at an early stage of this investigation, and Andrés Mejía for thoughtful conversations.

And finally, enormous thanks to the investigation groups of Loreto Valenzuela and Angélica Fierro for their permanent concern and advice.

GENERAL INDEX

TABLE INDEX

Page

FIGURE INDEX

Page

RESUMEN

La separación y purificación de polifenoles es un activo tópico de investigación debido al alto valor bioactivo de estos compuestos. En la optimización de dichos procesos, la forma de la curva de solubilidad a distintas condiciones es más relevante que los valores exactos de la solubilidad. Obtener datos de solubilidad a partir de experimentos y en distintas condiciones es costoso y demoroso. Alternativamente, las simulaciones de dinámica molecular (MD) requieren muy pocos datos experimentales como *input*, a diferencia de otras aproximaciones teóricas. Sin embargo, en la actualidad la predicción de solubilidad a través de esta herramienta se encuentra limitada por la complejidad asociada a la estimación de las propiedades de sublimación asociadas al fenómeno de solubilidad, las cuales aún no se comprenden a cabalidad. Se aplicaron simulaciones de MD para el estudio de la solubilidad, sin considerar las propiedades de sublimación, con el fin de obtener propiedades termodinámicas que puedan entregar información relevante sobre la tendencia de la solubilidad a distintas temperaturas. Se calcularon energías libres de solvatación a distintas temperaturas utilizando simulaciones de MD para tres compuestos aromáticos. Los resultados de las simulaciones fueron capaces de predecir con precisión el efecto de la temperatura en la solubilidad acuosa para los compuestos estudiados.

Palabras Claves: Polifenoles; Extracción sólido-líquido; Energía libre de solvatación; Simulaciones de dinámica molecular.

ABSTRACT

The separation and purification of polyphenols is an active research topic due to the high bioactive value of these compounds. In the optimization of such processes, the trend of solubility at different conditions is of major interest rather than the exact solubility values. Obtaining solubility data from experiments at different conditions is lengthy and costly. Alternatively, molecular dynamics (MD) requires little or no experimental data, differently from other theoretical approaches. Nevertheless, solubility prediction with this tool is limited by the estimation of sublimation properties related to the solubility phenomena, which are not thoroughly understood. We applied MD for the study of solubility without considering sublimation properties, with the aim of obtaining thermodynamic properties which can deliver relevant information of the trend of solubility at different temperatures. Free energies of solvation at different temperatures were calculated from MD for three aromatic compounds. Simulation results provided accurate predictions of the effect of temperature on aqueous solubility for the studied compounds.

Keywords: Polyphenols; Solid-liquid extraction; Solubility; Free energy of solvation; Molecular dynamics simulations.

1. INTRODUCTION

This work developed from the interest in polyphenols, a wide group of natural compounds which have attracted much attention in the scientific community due to their high bioactivity. One of the main interests in this field is the extraction of polyphenols from their natural sources, in order to purify and use them in alimentary, nutraceutical, medical and cosmetic, among other applications. Such task is challenging, principally because of the complexity of obtaining solubility data from experiment, which is needed in the design of the extraction process. Therefore, this work was focused in the prediction of solubility data of polyphenol-related molecules from Molecular Dynamics (MD), which would serve as an alternative to the experimental determination of this property. As a result of this investigation, an article was sent to the Food Chemistry journal (https://www.journals.elsevier.com/food-chemistry).

In the present chapter, five themes are treated as a contextualization: the relevance of polyphenols in academia and industry; the importance of solubility data in the design of solid-liquid extraction processes; the current approaches for solubility prediction; an introduction to MD; and the application of MD for solubility prediction. At the end of the chapter, the hypothesis and objectives of this work are exposed. In chapter 2, the methods employed to fulfill these objectives are exposed. In chapter 3, results are shown and discussed. Finally, in chapter 4, the conclusions, along with the contribution of this work and further research projections are exposed.

1.1. The relevance of polyphenols

Polyphenols are a group of compounds widely distributed in the plant kingdom, and are characterized by the presence of one or several phenolic groups in their structure. Initially, the term “polyphenol” was attributed to complex molecules presenting several phenolic hydroxyl groups and aromatic rings, but it has been extended considerably over the years to include much simpler phenolic structures (Ferrazzano et al. 2011). These compounds may be classified according to the number of phenol rings that they contain and to the functional groups that bind them. Thus, structural distinctions can be made between phenolic acids, flavonoids, stilbenes, and lignans

(Manach et al. 2004). A thorough description of this classification is given elsewhere (Manach et al. 2004). The main sources of polyphenols are seasoning foods, fruits, seeds, vegetables, non-alcoholic beverages, cereals, cocoa products, alcoholic beverages and oils (Pérez-Jiménez et al. 2010). In plants, they are generally involved in defense against ultraviolet radiation or aggression by pathogens (Manach et al. 2004).

Polyphenols are known for producing a wide variety of biological effects in the human body as a result of their antioxidant, anti-inflammatory, vasodilating, and prebiotic properties (Leopoldini, Russo, and Toscano 2011). For example, it has been discussed that polyphenols might play a role in preventing atherosclerosis, as well as protecting DNA from oxidative damage, which would prevent the development of some cancers (Halliwell 2009). Also, polyphenols can induce the generation of nitric oxide from blood vessels, which induces the expression of cardiovascular-protective genes (Stoclet et al. 2004). Many other biological effects of polyphenols in the human body are described elsewhere (Landete 2012).

In search for simplicity, three low molecular weight aromatic compounds were selected as case studies for this work: benzoic acid, catechin and toluene. Their molecular structures are presented in figure 1.1.

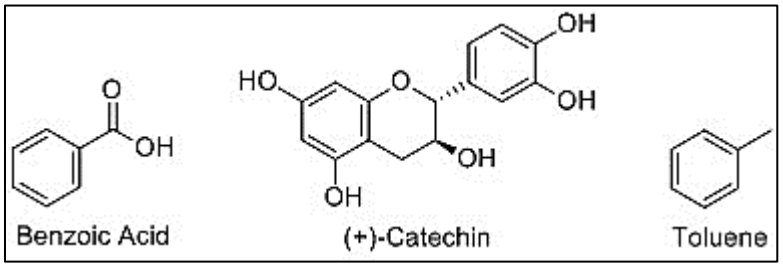

Figure 1.1: Molecular structure of the selected compounds for this study.

Catechin is a polyphenol from the flavonoid family. Toluene is not a polyphenol, and it was selected as a negative control. Benzoic acid is an intermediate case, since it is a precursor of the phenolic acids subset of the polyphenol family and is produced as a result of the metabolism of some polyphenols (Shelnutt et al. 2000), but it is not

considered part of the family due to the lack of phenol functionalities in its molecular structure. The three compounds share in common being widely studied in literature, and their simple structures, which would simplify the development of an unexplored solubility prediction method.

1.2. The importance of solubility data in the design of solid-liquid extraction processes

The current approach for optimizing extraction processes is the Response Surface methodology (Bogdanovic et al. 2016; Heleno et al. 2016; Derrien et al. 2017). This methodology consists in performing extraction experiments at multiple conditions that are produced by varying parameters such as temperature, extraction time or solvent composition. Then, the extraction yield is measured in each case, and as a result a curve or surface informing the extraction yield at different conditions can be assembled. However, the disadvantage of the Response Surface method is that it requires several experiments, involving high costs and requiring considerable time. Another alternative is using solubility data for the optimization of the extraction process. Solubility plays a crucial role in the design of solid-liquid extraction processes, since it delivers information about the yield of the extraction process. Therefore, the knowledge of the behavior of solubility at different temperatures, pressures or solvent conditions is of primal interest for the optimization of the process. In such optimizations, the trend of solubility at different conditions is much more important than the specific values. Commonly, solubility data is obtained from experiments. Unfortunately, experimental determination of solubility is difficult, time consuming and expensive, even more when it is required at many different conditions. Predicting reliable solubility values of bioactive compounds at an early stage of the process design or the product formulation, with minimum or no experiments, would simplify, accelerate and reduce the costs of the development of new natural products.

1.3. Physical-chemical models and current prediction tools for solubility

Solid-liquid solubility is commonly defined as the concentration of the solute in solution when an equilibrium state is reached between the solid phase (solute) and the liquid phase (solution). At equilibrium, the rate of molecules joining the solid phase, commonly assumed to be in crystalline form, equals the rate of molecules leaving the solid phase (Schnieders et al. 2012). In thermodynamics, this phenomena is modeled as the complex interplay between several energy contributions such us (1) the disarray of the pure solid crystal lattice required to bring the solute into solution, (2) the interactions between the solute and the solvent in solution and (3) the creation of a void in the solvent to host the solute molecule (Lipinski et al. 1997). Frequently, Equations of State (EOS) are used for describing thermodynamic properties of pure substances and mixtures, such as solubility. Even though EOS solubility models have been adapted to model complex mixtures by the inclusion of solution and mixing effects, these models contain several semi-empirical parameters that must be calibrated with extensive experimental data (Held et al. 2014; Ji et al. 2007). Quantitative Structure-Property Relationships (QSPR), which consist in developing regression models between structural parameters of the solute molecule and experimental data of a given property such as solubility, is another useful modelling approach, but they also rely on extensive experimental data (Duchowicz et al. 2013; Salahinejad, Le, and Winkler 2013). Additionally, physical-chemical analytical models have been proposed for the prediction of activity coefficients of small carbohydrates (Ben Gaïda, Dussap, and Gros 2006), but in these models, two parameters (i.e., hydration number and equilibrium constant) must be estimated from experimental solubility data.

1.4. The Molecular Dynamics Approach

Molecular dynamics (MD) is an alternative modelling method that relies much less on experimental data. This approach requires a detailed description of the atomic interactions to solve exactly a theoretical model of the molecular behavior. Thus,

macroscopic properties of the system (such as solubility) can be determined based on the dynamic interaction of particles. Since this method relies on first principles, the prediction of such properties does not necessarily require experimental measurements. To represent the molecular behavior of a system in the MD approach, the kinetic energy can be derived from the individual momenta of the particles and the potential energy is represented by semi-empirical analytical functions. These functions, known as force fields, represent the potential energy as a sum of interaction terms between bonded and non-bonded atoms. For example, the total potential energy (U_{total}) depends on all atomic positions (r_{ij}), and in the CHARMM (Chemistry at HARvard Macromolecular Mechanics) force field (Vanommeslaeghe et al. 2009) is represented by:

$$U_{total} = \sum_{bond\ i} k_i^b (b_i - b_{0i})^2 + \sum_{angle\ i} k_i^\theta (\theta_i - \theta_{0i})^2 + \sum_{dihedral\ i} k_i^\phi [1 + \cos(n_i \phi_i - \delta_i)] + \sum_{\substack{improper \\ dihedral\ i}} k_i^\varphi (\varphi_i - \varphi_{0i})^2 + \sum_i \sum_{j>i} \varepsilon_{ij} \left[\left(\frac{R_{ij}^{min}}{r_{ij}} \right)^{12} - \left(\frac{R_{ij}^{min}}{r_{ij}} \right)^{6} \right] + \sum_i \sum_{j>i} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \quad (1.1)$$

The first four terms represent the bonded interactions of molecules by the harmonic potential for bonds and angles, a truncated Fourier series for dihedrals and the Urey-Bradley term for improper dihedrals. In these terms, the bond length (b_i), bond angle (θ_i), dihedral angle (ϕ_i) and phase (δ_i), and the improper torsion angle (φ_i) are determined from the atomic positions (r_{ij}). The remaining terms represent the non-bonded interactions of molecules by van der Waals' (vdW) dispersion forces and electrostatic interactions. Dispersion forces are approximated by a Lennard-Jones 6-12 potential, which includes a repulsive short-range term (r_{ij}^{-12}) and an attractive long-range term (r_{ij}^{-6}). Electrostatic interactions between pair of atoms with charges q_i and q_j are modeled using the Coulomb potential. The parameters of the potential terms of the force field (i.e., $k_i, b_{0i}, \theta_{0i}, \varphi_{0i}, R_{ij}^{min}, \varepsilon_{ij}, q_i$) are determined from experiments or *ab initio* calculations. These parameters

for common functional groups or molecules, such as water, are usually found in the literature.

The dynamic behavior of the system is simulated in a computer program by applying the Newtonian equations of motion to each particle:

$$m_i \ddot{\boldsymbol{r}}_i = -\frac{dU_{total}}{d\boldsymbol{r}_i}(\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_N), \quad i = 1,2,\ldots,N \tag{1.2}$$

where m_i is the mass of atom i, $\boldsymbol{r}_i$ is its position (the bold font means that it is a vector). These equations are solved iteratively to simulate a desired time span. Further details regarding the MD approach can be found elsewhere (Sadus 2002).

1.5. Molecular dynamics applied for the study of solubility

Currently, predicting solid-liquid solubility from MD simulations is an active research topic. The accurate prediction of solubility has traditionally been limited by the difficulties of modelling the solid crystal lattice energy, since it is very unlikely for a molecule to have only one way of packing itself. Instead, advances in crystal structure prediction in the last two decades suggest that many molecules show polymorphism (i.e., the molecule can pack itself in several forms). Therefore, while it is relatively easy to predict the crystal structure of some organic compounds, the crystal structure of other compounds will remain essentially unpredictable (Price 2014). Consequently, for some molecules, crystal structure prediction is only a complement to the experimental screening aimed at finding all solid forms. Recently, Schnieders et al. (2012) developed a technique to improve the prediction efficiency of the thermodynamic stability and solubility of organic crystals using MD simulations. However, the technique is still limited to a relatively small number of compounds. Additionally, Salahinejad et al. (2013) studied different approaches for aqueous solubility prediction, comprising that of Schnieders et al. (2012), and found that the inclusion of parameters that accounted for crystal lattice interactions did not significantly improve the accuracy of QSPR solubility models. Hence, we focus on studying solubility from the solvation energy of a solute.

Starting from the thermodynamic definition of solid-liquid equilibrium and considering an infinitely diluted system, the following expression can be derived for the calculation of the solubility of a compound (x) from its free energy of solvation (Prausnitz, Lichtenthaler, & Azevedo, 1998; Frenkel & Smit, 2002; Paluch & Maginn, 2013) (see Appendix A):

$$x = \frac{\exp\left(\frac{v^S(T,p)\left[p-\exp\left(-\frac{\Delta H^S(T)}{R}\left(\frac{1}{T}\right)+\frac{\Delta S^S(T)}{R}\right)\right]}{RT}+\frac{\Delta H^S(T)}{R}\left(\frac{1}{T}\right)-\frac{\Delta S^S(T)}{R}\right) v^L(T,p)}{RT} \exp\left(-\frac{\Delta G_{solv}}{RT}\right) \quad (1.3)$$

where $v^S(T,p)$, $\Delta H^S(T)$ and $\Delta S^S(T)$ are the pure solute molar volume, enthalpy and entropy of sublimation, respectively, at temperature T and pressure p. $v^L(T,p)$ is the molar volume of the solvent at the selected conditions, which can be retrieved from a simple molecular simulation or from literature. R is the universal gas constant and ΔG_{solv} is the free energy of solvation of the solute in the solution at the selected conditions. This last term can be calculated from molecular simulations. Chebil et al. (2010) used equation 1.3 to calculate the isothermal relative solubility of a given solute in two different solvents. If equation 1.3 is applied to a solute in different solvents at the same temperature, many of its terms can be grouped in a parameter that we call C, which depends on the temperature and the solute, and therefore can be considered constant for this case:

$$x = C \exp\left(-\frac{\Delta G_{solv}}{RT}\right) \quad (1.4)$$

To calculate the solubility of a solute in a given solvent (x_2) relative to its solubility in another solvent (x_1) at the same temperature, as a function of the free energies of solvation in each solvent (ΔG^2_{solv} and ΔG^1_{solv}, respectively) and the temperature, Chebil et al. derived the following expression:

$$x_2 = x_1 \exp\left(-\frac{\Delta G^2_{solv}-\Delta G^1_{solv}}{RT}\right) \quad (1.5)$$

Thus, they demonstrated that the solubility of a compound in a given solvent could be estimated accurately from the experimental measurement of solubility in

another solvent at the same temperature. Normally this procedure is referred to as relative solubility estimation.

Similarly, in this work we set forth a study of solubility without accounting for crystal structure prediction. In the optimization of separation processes, the trend of solubility at different conditions is of major relevance rather than the exact solubility values. Therefore, we were interested in retrieving information of the trend of solubility at different temperatures from MD simulations that do not require experimental information as input. We focused on studying the free energy of solvation of polyphenol-related compounds at different temperatures in order to obtain a simple computational model to study the solubility of polyphenols. Thus, benzoic acid as a precursor of polyphenols and catechin as a polyphenol itself were used to describe specific solute-solvent interactions, while toluene was used as a negative control of our study.

1.6. Hypothesis and objectives

1.6.1. Hypothesis

Thermodynamic properties calculated from MD simulations can deliver information about the trend of aqueous solubility at different temperatures.

1.6.2. Objectives

The main objective of this work was developing a simple method to determine thermodynamic properties from MD, which can deliver information of the trend of aqueous solubility at different temperatures.

In particular, the specific aims of this thesis were:

- To perform a theoretical analysis of solubility and determine physical-chemical properties that can be related to it.
- To elaborate representative molecular models of the systems of interest.
- To determine a set of MD parameters that will allow performing stable simulations, and subsequently, to perform MD simulations using such parameters and the molecular models elaborated in the third objective.

- To calculate from MD simulations the properties determined in the first objective.

2. MATERIALS AND METHODS

The methodology followed to accomplish the objectives of this work is summarized in figure 2.1.

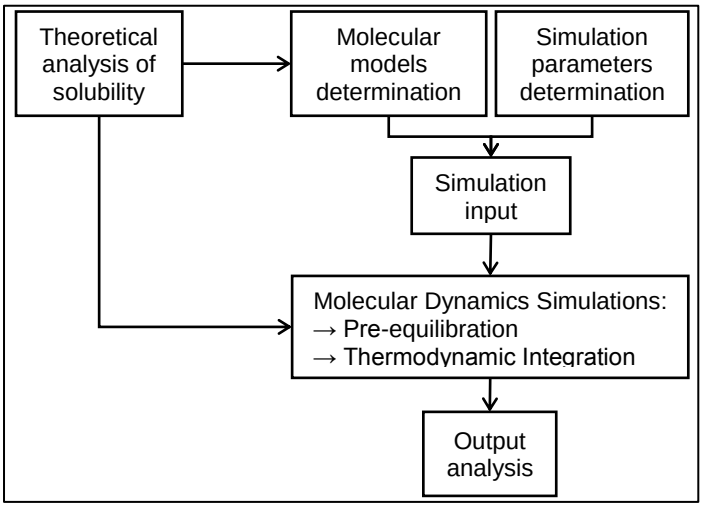

Figure 2.1: Overview of the methodology used in this work.

A theoretical analysis of solubility was employed in order to determine which MD protocol would be pertinent for calculating such property. This information was also employed to determine relevant molecular models for the simulations. These models, along with a set of simulation parameters, were used as the input for the MD simulations. Once the MD simulations were completed, an additional output analysis was performed for the calculation of several thermodynamic properties.

As was mentioned before, to illustrate the methodology, the following simple, low molecular weight aromatic compounds were considered: benzoic acid, (+)-catechin and toluene.

For the three compounds, isobaric aqueous solubility data at different temperatures was retrieved from literature (Cuevas-Valenzuela et al. 2015; Miller and Hawthorne 2000) to subsequently compare it to the simulation results. Particularly, Cuevas-Valenzuela et al. (2015) measured benzoic acid and (+)-catechin solubilities at different temperatures and atmospheric pressure (101.325 kPa), while Miller & Hawthorne (2000) measured toluene

solubilities at different temperatures and 50 bar (5 MPa). In this work, simulations were set up in order to reproduce the experimental conditions, particularly the pressure and temperature.

2.1. Theoretical analysis of solubility

The free energy of solubility ($\Delta G_{solubility}$) in a system is defined as the sum of two energy terms, shown in equation 2.1 (Schnieders et al. 2012). The free energy of sublimation term (ΔG_{sub}) accounts for the disruption of the pure solid crystal lattice, and the solvation free energy term (ΔG_{solv}) accounts for both the energy required to create a void in the solvent into which a solute molecule is hosted, and the solute-solvent interaction.

$$\Delta G_{solubility} = \Delta G_{sub} + \Delta G_{solv} \quad (2.1)$$

As was mentioned before, only the calculation of the free energy of solvation was considered in this work. For the calculation of the free energy of solvation, an infinite dilution approximation was considered, since the aqueous solubilities of the studied compounds were low (which is also the case of the aqueous solubilities of polyphenols in general). The infinite dilution approximation states that if the mole fraction of the solute in solution is sufficiently small, solute-solute interactions can be neglected and the system can be represented in MD as only one solute surrounded by solvent (Paluch and Maginn 2013). In MD simulations, ΔG_{solv} can be estimated based on the thermodynamic cycle shown in figure 2.2.

$$\begin{array}{ccc} solute_{vacuum} & \xrightarrow{\Delta G_{solv}} & solute_{water} \\ \downarrow{\scriptstyle \Delta G_1} & & \downarrow{\scriptstyle \Delta G_2} \\ nothing_{vacuum} & \xrightarrow{\Delta G=0} & nothing_{water} \end{array}$$

Figure 2.2: Thermodynamic cycle used for the calculation of free energies of solvation, where $\Delta G_{solv} = \Delta G_1 - \Delta G_2$. ΔG_1 and ΔG_2 are solute annihilation free energies in vacuum and in water, respectively. Alternatively from calculating the annihilation free energy in vacuum, ΔG_1, the free energy of solvation can be determined only from ΔG_2 if

during this annihilation, the perturbed intramolecular interactions are neglected. Therefore, ΔG_{solv} can be estimated by decoupling the solute-solvent interaction (solute annihilation in water) while the intramolecular interactions are neglected during the whole process.

To perform the calculations, solute-solvent interactions are progressively switched off while intramolecular interactions are neglected (see figure 2.2 caption). Hence, the difference in energy between the default system and the system with switched off interactions corresponds to the solvation free energy (this is true only if intramolecular interactions are neglected). A useful method to measure the energy path of a system when certain interactions are switched off during a MD simulation is Thermodynamic Integration (TI) (Beveridge and Dicapua 1989). Therefore, the simulation protocol was defined as a calculation of the solvation free energy for all the selected compounds at each of its respective temperatures by means of TI in MD simulations. In addition, a pre-equilibration of each system was performed before each TI calculation, in order to reach equilibrium at the desired pressure and temperature conditions.

2.2. Molecular models

To simulate the infinitely diluted systems for the three selected compounds, molecular models for the solvent (water) and the three different solutes were determined. In MD simulations, a molecular model consists in the structure of the desired molecule along with all the parameters required for the selected force field. All force field parameters, except for the partial charges of each solute, were taken from the CHARMM force field (Vanommeslaeghe et al. 2009). This force field has been extensively used to model pharmaceutical and natural compounds. The TIP3P (Three-site Transferrable Intermolecular Potential) model (Jorgensen et al. 1983) was selected to represent the water molecules, and atomistic models for benzoic acid, (+)-catechin and toluene were generated in the Spartan'10 v1.1.0 software (Wavefunction 2010). The partial charges for the solutes were calculated

quantum-mechanically at the Hartree-Fock 6-31G* basis set (Hariharan and Pople 1973); these calculations were performed using the Spartan'10 v1.1.0 software. Finally, to build the infinitely diluted systems, each solute was submerged in a pre-equilibrated solvent box of 30 Å (1^{-10} m) of side, which included approximately 800 water molecules. This was done using the Solvate 1.5 plugin (Caddigan et al. 2003), which is incorporated in VMD (Visual Molecular Dynamics) 1.9.2 package (Humphrey, Dalke, and Schulten 1996).

2.3. Simulation Parameters

The following simulation parameters and methods were employed for both the pre-equilibration and the TI calculation. The simulations were performed using the NAMD (NAnoscale Molecular Dynamics) v2.12 simulation package (Phillips et al. 2005). Most of the simulation parameters were taken from a NAMD free energy tutorial (Hénin, Gumbart, and Chipot 2014) . The equations of motion were integrated using a time step of 2 fs (10^{-15} s). Temperature and pressure were kept constant using Langevin dynamics and the Langevin piston method, respectively (Feller et al. 1995). VdW interactions were truncated using a cutoff radius for computational efficiency. To decide the cutoff distance, preliminary MD simulations were conducted for each of the solutes with five different cutoff distances ranging from 4 to 20 Å. It was observed that the calculated vdW interactions for cutoff distances of 12, 16 and 20 Å were approximately equivalent; therefore, the shortest of those distances (12 Å) was selected for computational efficiency (see Appendix B). The electrostatic interactions were calculated using the Particle Mesh Ewald summation (Essmann et al. 1995) with a grid size equal to the size of each simulation box.

2.4. Molecular dynamics simulations

The pre-equilibrations performed for each system at each simulation temperature consisted in:

- An energy minimization of 0.05 ns using the conjugate gradient algorithm (Watowich et al. 1988), to guarantee that each atom took feasible starting positions.
- An equilibration for 1 ns, to assure the convergence of each system to an equilibrium state.

Subsequently, each equilibrated system was subjected to a TI calculation. TI calculations were performed using the TI method (Bhandarkar et al. 2014) included in the NAMD v2.12 package. In this method, the energy terms in the force field are multiplied by a scalar parameter ($0 < \lambda < 1$) in order to switch off the desired interactions. In our case, the non-bonded interactions between the solute and the solvent molecules (vdW and electrostatic potential terms in equation 1.1) were switched off. Equation 2.2 (Bhandarkar et al. 2014) shows the modified vdW and electrostatic non-bonded potentials for the TI method in NAMD.

$$U_{n-b}(r_{ij},\lambda) = \varepsilon_{ij}\lambda_{LJ}\left[\left(\frac{R_{ij}^{min^2}}{r_{ij}^2+\delta(1-\lambda_{LJ})}\right)^6 - \left(\frac{R_{ij}^{min^2}}{r_{ij}^2+\delta(1-\lambda_{LJ})}\right)^3\right] + \lambda_{elec}\frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \qquad (2.2)$$

In addition to λ, an additional parameter known as the soft-core parameter, δ, is added to the Lennard-Jones potential. The modified Lennard-Jones potential shown in equation 2.2 was designed to circumvent difficulties in the integration of the equations of motion caused by appearing or vanishing atoms that collide with solvent molecules. These difficulties lead to unusual increments in the energy known as "end-point catastrophes" (Beutler et al. 1994). The value of the soft-core parameter must be selected in order to soften the non-bonded interparticle interactions.

To implement the TI method, a number of molecular simulations are performed with different values of lambda between 0 and 1. From each of the simulations, the difference in energy is retrieved and finally a curve representing the variation of energy at each λ value can be assembled. This curve is integrated numerically to obtain a total free energy difference of the studied perturbation in the system, as shown in equation 2.3:

$$\Delta G_{solv} = \int_0^1 \langle \frac{\delta H(\lambda)}{\delta \lambda} \rangle_\lambda \, d\lambda \quad (2.3)$$

where ΔG_{solv} is the calculated solvation free energy, H is the Hamiltonian of the system, a function which includes the whole potential energy function (shown in equation 1.1) and the kinetic energy function. In this case, the Hamiltonian is a function of λ.

It is important to note that the total free energy difference calculated by employing only two lambda values (i.e., 0 and 1), might differ significantly from a free energy calculation using numerous intermediate lambda values. Normally the energy path between two states of the system is nonlinear; hence, if an insufficient sampling of λ values is used, the retrieved energy vs λ curve might take leaps (a thorough explanation of this point is given in the Thermodynamic Integration section of the NAMD users' guide (Bhandarkar et al. 2014)). Therefore, in order to get an exact calculation of the free energy difference, a smooth curve is required by sampling several intermediate λ values.

In this work, 50 different λ values uniformly distributed between 0 and 1 were used. This protocol simulated a gradual switching off of the solute-solvent interactions in each of the systems. During this process, intramolecular interactions were neglected to avoid additional vacuum calculations that are needed to complete the thermodynamic cycle (ΔG_1 in figure 2.2). This was accomplished by using the *alchdecouple* option in NAMD. For the soft-core parameter, values lower than 1 nm are usually recommended in the literature; specifically the value used in our study (0.5 nm) is common (Zhang, Tuguldur, and van der Spoel 2015) and also used in a NAMD tutorial for free energy calculations (Hénin, Gumbart, and Chipot 2014). Each λ step was simulated through 0.1 ns (100,000 fs) and the energy was measured each 20 fs, summing 5,000 energy measurements for each λ step. The last 20,000 fs (1,000 energy measurements) of each λ step were collected and the energy was averaged. Along with the average energies, standard deviations for the

same energy measurements were calculated in each lambda step, and were latter used to calculate the error from propagation of uncertainty.

This procedure was implemented for all compounds at each temperature, and repeated 10 times to assure the consistency and repeatability of the results.

2.5. Output analysis

The curves of energy versus lambda retrieved from simulations for each compound, at each temperature and for each of the 10 replicas, were integrated numerically using a python script based on the *namd_ti.pl* perl script that can be found in the NAMD website (http://www.ks.uiuc.edu/Research/namd/utilities/). The *namd_ti.pl* script uses cubic spline interpolation for the numerical integration. In addition to this feature, the coded script computed the standard deviations of the free energies. In order to obtain the standard deviations, the general formula for propagation of uncertainty described in equation 2.4 (Taylor, 1997) was applied to the cubic spline interpolation polynomial. If q is any function of several independent variables $x, \ldots, z$, then the uncertainty of q, δq, can be propagated by the following expression:

$$\delta q = \sqrt{\left(\frac{\partial q}{\partial x}\delta x\right)^2 + \cdots + \left(\frac{\partial q}{\partial z}\delta z\right)^2} \qquad (2.4)$$

Finally, the calculated solvation free energies averaged from the 10 replicas were compared with experimental solubility data.

3. RESULTS AND DISCUSSION

The calculated free energies of solvation in each replica were decomposed into their vdW and electrostatic contributions. The retrieved energy versus λ curves were smooth and no leaps were observed; this was achieved thanks to a high λ sampling and to a proper selection of δ. Although the results for the three compounds at each temperature and for each replica are not shown, a representative result is included in Appendix C.

3.1. Solvation free energy at different temperatures

The calculated free energy of solvation results are presented in table 3.1. Only three replicas are shown; the remaining results are presented in Appendix D.

As can be seen in table 3.1, the calculated results present very low variations between different replicas. Particularly, errors in the electrostatic contribution remain almost invariable between different replicas, and it is relevant to mention their low values, which contribute to statistically significant differences in the trends at different temperatures. Compared to the errors in electrostatic contributions, the errors in vdW contributions present slightly more variability between different replicas, however, the differences in vdW contributions at different temperatures are not statistically significant. These results demonstrate that the selected parameters, simulation times and the integration and uncertainty propagation procedures were adequate for the generation of replicable results. Therefore, such parameters and procedures might be useful for future calculations.

The free energy, averaged from the 10 performed replicas, is shown in figure 3.1. The electrostatic contribution to the free energy of solvation shows a linear, decreasing, statistically significant trend at different temperatures for the three studied compounds. The slope of the linear trend is more pronounced for compounds with a higher amount of polar functional groups (i.e. catechin > benzoic acid > toluene).

Table 3.1: Average electrostatic and vdW contributions to the free energies of solvation calculated for benzoic acid (A), (+)-catechin (B) and toluene (C) at each simulation temperature for three replicas of the procedure. The results for the remaining replicas are presented in Appendix D. Errors correspond to the calculated standard deviations.

(A)	Electrostatic contribution to ΔG_{solv} (kJ mol^{-1})			vdW contribution to ΔG_{solv} (kJ mol^{-1})		
T [K]	R1	R2	R3	R1	R2	R3
278.4	-36.3 ± 3.5	-33.5 ± 3.4	-33.8 ± 3.3	-313.3 ± 15.6	-310.6 ± 14.8	-320.9 ± 14.0
282.9	-73.1 ± 4.9	-67.6 ± 4.9	-69.1 ± 4.7	-309.1 ± 18.9	-305.9 ± 17.7	-320.2 ± 17.4
294.1	-107.8 ± 6.0	-105.2 ± 5.9	-106.7 ± 5.7	-304.7 ± 21.4	-297.4 ± 20.6	-318.7 ± 20.4
300.4	-144.3 ± 7.0	-141.5 ± 6.9	-142.9 ± 6.7	-296.0 ± 23.5	-299.9 ± 22.8	-322.1 ± 23.0
307.7	-178.9 ± 7.9	-175.2 ± 7.8	-176.0 ± 7.6	-291.9 ± 25.7	-295.1 ± 25.1	-319.7 ± 25.2
317.7	-212.2 ± 8.7	-207.9 ± 8.5	-210.0 ± 8.4	-277.9 ± 28.4	-286.3 ± 27.7	-310.6 ± 27.9
326.7	-245.7 ± 9.5	-242.8 ± 9.3	-245.8 ± 9.4	-271.0 ± 30.7	-269.1 ± 30.5	-302.6 ± 30.2
336.6	-281.1 ± 10.2	-277.3 ± 10.1	-280.5 ± 10.1	-267.2 ± 32.4	-258.5 ± 32.6	-296.5 ± 32.1
340.9	-315.6 ± 10.8	-310.7 ± 10.8	-316.1 ± 10.8	-262.0 ± 34.5	-252.1 ± 34.3	-290.7 ± 34.3

(B)	Electrostatic contribution to ΔG_{solv} (kJ mol^{-1})			vdW contribution to ΔG_{solv} (kJ mol^{-1})		
T [K]	R1	R2	R3	R1	R2	R3
277.6	-99.3 ± 5.4	-100.1 ± 5.3	-101.7 ± 5.5	-626.1 ± 20.8	-632.1 ± 21.8	-648.3 ± 19.8
282.7	-219.3 ± 7.8	-209.4 ± 8.0	-212.8 ± 8.0	-630.2 ± 25.6	-628.2 ± 26.0	-653.1 ± 24.8
297.7	-321.2 ± 10.1	-304.8 ± 10.1	-315.7 ± 9.8	-629.1 ± 29.7	-632.8 ± 30.5	-651.3 ± 28.7
308.4	-431.3 ± 11.9	-413.8 ± 11.7	-430.0 ± 11.8	-627.0 ± 33.6	-639.2 ± 33.7	-642.2 ± 33.6
324.9	-525.2 ± 13.4	-511.7 ± 13.0	-529.2 ± 13.3	-621.8 ± 37.7	-633.9 ± 37.3	-652.1 ± 36.2
331.2	-632.3 ± 14.9	-623.6 ± 14.5	-633.6 ± 14.7	-630.5 ± 40.6	-636.8 ± 40.3	-662.1 ± 39.8

(C)	Electrostatic contribution to ΔG_{solv} (kJ mol^{-1})			vdW contribution to ΔG_{solv} (kJ mol^{-1})		
T [K]	R1	R2	R3	R1	R2	R3
298	-10.1 ± 1.8	-10.5 ± 1.8	-10.3 ± 1.7	-40.0 ± 11.9	-42.2 ± 10.3	-32.9 ± 10.9
323	-19.9 ± 2.7	-21.1 ± 2.6	-19.8 ± 2.5	-28.9 ± 16.5	-28.7 ± 14.7	-22.8 ± 14.9
373	-28.8 ± 3.3	-31.2 ± 3.3	-29.6 ± 3.3	-19.9 ± 19.7	-15.7 ± 19.9	-15.4 ± 18.0
423	-38.0 ± 3.9	-40.3 ± 3.8	-38.2 ± 3.8	-10.2 ± 22.8	-0.6 ± 23.4	-0.3 ± 21.3
473	-45.9 ± 4.4	-48.6 ± 4.3	-45.4 ± 4.3	1.4 ± 26.6	9.1 ± 26.2	9.2 ± 24.0

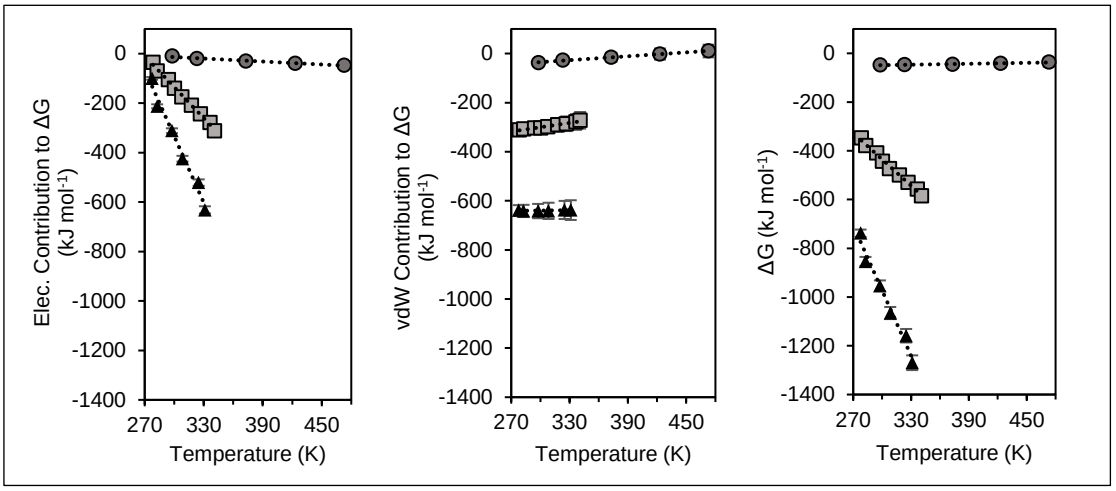

Figure 3.1: Calculated electrostatic (elec.) and vdW contributions, and total free energy of solvation (ΔG) at different temperaturesor benzoic acid (▫), catechin (▲) and toluene (●). Error bars correspond to the standard deviation of each result, calculated by propagation of uncertainty. Dotted lines correspond to adjusted linear trends to each data set.

Electrostatic interactions are due to point charge interactions, thus, when more kinetic energy is added to the system –and the probability of collision between the particles of the system increases–, point interactions are more probable. This explains the observed straight decreasing trends for increasing temperatures (contributing to more stability in each system), independently of the nature of the compounds. VdW contributions for (+)-catechin do not show any variation with temperature, and a slightly increasing variation is observed for benzoic acid and toluene. As a whole, benzoic acid and (+)-catechin present linear, decreasing trends in their total free energy of solvation at different temperatures. Toluene does not present any variation in its total free energy of solvation, due to the cancelation of its increasing vdW and decreasing electrostatic trends. Also, larger free energy values (more negative) are observed for compounds with a higher amount of polar functional groups.

3.2. Comparison between the MD results and experimental solubility data

Since the electrostatic contribution had a straight linear, decreasing behavior at different temperatures for the three compounds, it was compared with experimental solubility values, as depicted in figure 3.2.

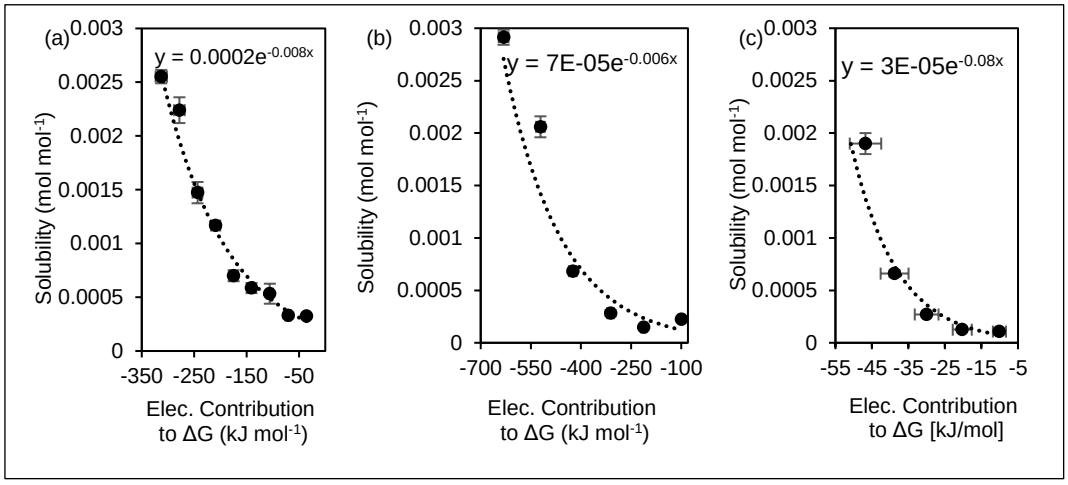

Figure 3.2: Electrostatic (elec.) contribution to the free energies of solvation vs. experimental solubility at different temperatures for benzoic acid (a), (+)-catechin (b) and toluene (c). All the studied compounds show exponential trends (adjusted; dotted lines). Horizontal error bars correspond to the standard deviations of the free energy calculations, obtained by propagation of uncertainty. Vertical error bars correspond to the reported errors in each experimental reference. Exponential trends with statistically significant differences are observed. In the three studied cases, the exact same exponential trend was observed. Therefore, such type of interactions were able to yield important information about the shape of solubility at different temperatures, with independence of the nature of each compound.

In figure 3.3, the calculated free energies of solvation were compared to experimental solubility values.

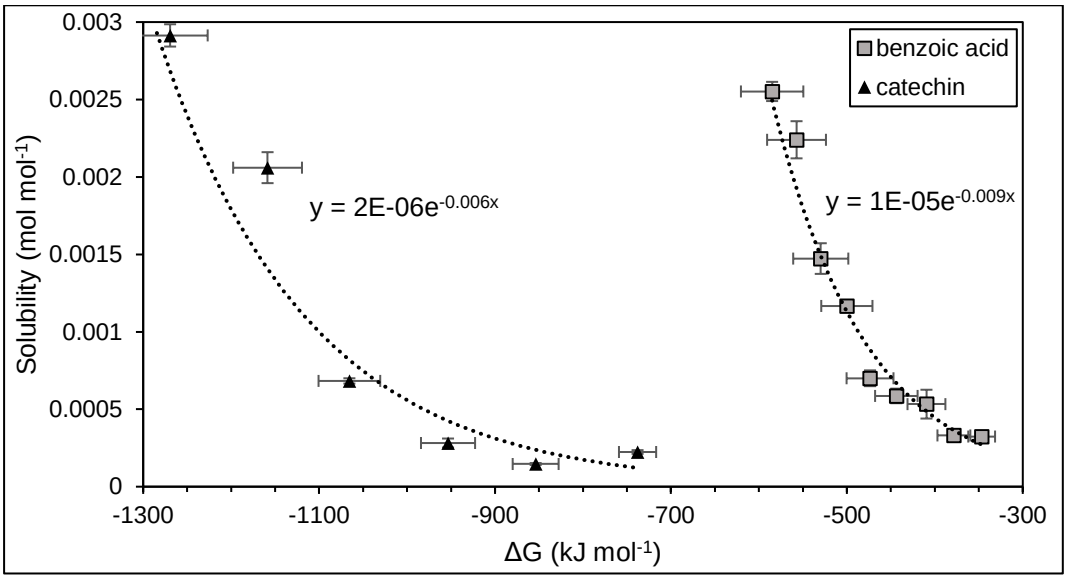

Figure 3.3: Calculated free energies of solvation (ΔG) vs. experimental solubility at different temperatures. Benzoic acid and (+)-catechin show exponential trends (adjusted; dotted lines). Toluene is not shown since no trend was observed. Horizontal error bars correspond to the standard deviations of the free energy calculations, obtained by propagation of uncertainty. Vertical error bars correspond to the reported errors in each experimental reference.

The free energy measurements for (+)-catechin and benzoic acid from MD simulations were able to deliver information of the trend of experimental solubility data. Moreover, when comparing the simulation results with solubility data, exponential shapes are observed for the two compounds. Toluene (not shown in figure 3.3) does not show any trend due to the invariable free energy of solvation with temperature. In addition, the calculated values showed small errors; hence, the observed trends were statistically significant. The trends for (+)-catechin and benzoic acid were fitted to the following expression:

$$x = C \exp(-B\, \Delta G_{solv}) \qquad (3.1)$$

where C and B are constants. Noticeably, this expression is similar to equation 1.4, with the only difference that B is not constant and varies inversely with temperature (nevertheless, note that the fitted values of B differ only in two orders

of magnitude with $1/RT$, employing SI units and an average of 300 K for T; $\frac{1}{RT} = \frac{1}{8.314*10^{-3}*300} = 0.4$). Moreover, C is not a function of temperature for (+)-catechin and benzoic acid. This might suggest that, for these cases, sublimation properties do not play an essential role in the shape of the solubility vs temperature curves. Moreover, for similar compounds –i.e., polyphenols– the prediction of the shape of solubility at different temperatures might be possible without requiring sublimation properties. On the other hand, for toluene, which differs mainly in polarity from polyphenols, sublimation properties are likely to be strongly dependent of temperature at the selected pressure conditions (5 MPa) for both solubility measurements and MD simulations. In fact, high-pressure conditions were selected on the experimental setup in order to improve the aqueous solubility of toluene.

4. CONCLUSION

Free energies of solvation calculated from MD were able to provide information for the studied compounds of the shape of the aqueous solubility vs temperature curves, without requiring experimental data as input. For the studied polyphenolic compounds (i.e., benzoic acid and (+)-catechin), the shapes of the solubility vs temperature relationship, were highly dependent on the free energy of solvation and nondependent on sublimation properties. This study should be extended to more polyphenolic compounds to assess how general is this behavior. In addition, the differences observed among replicas were small, which confirms that the method is repeatable, and that the parameters chosen here were suitable for this context.

As long as we are aware of, none molecular dynamic studies of solubility prediction at different temperatures have been published before. Also, none studies before have been able to retrieve information of the trend of solubility at different temperatures without requiring experimental data as input. In this work, we were able to fulfill both tasks by demonstrating that: (1) the computed solvation free energy varied linearly with temperature for the studied compounds; (2) the solvation free energy varied exponentially with the polyphenol-related compounds' solubilities; (3) the polyphenol-related compounds' solubilities were more dependent on solvation properties than on sublimation properties. These statements must be proved by extending the proposed methodology to more polyphenol-related compounds. Also, different conditions might be varied. For instance, simulations with co-solvents at different co-solvent concentrations might be performed.

Also, the learning of MD principles and complex techniques such as TI are profound contributions to my formation as an investigator, and might be useful for me to contribute to different areas such as molecular biology, in which MD can be used as a powerful tool.

REFERENCES

Ben Gaïda, L., C. G. Dussap, and J. B. Gros. 2006. "Variable Hydration of Small Carbohydrates for Predicting Equilibrium Properties in Diluted and Concentrated Solutions." *Food Chemistry* 96(3): 387–401.

Beutler, Thomas C. et al. 1994. "Avoiding Singularities and Numerical Instabilities in Free Energy Calculations Based on Molecular Simulations." *Chemical Physics Letters* 222(6): 529–39.

Beveridge, D L, and F M Dicapua. 1989. "MOLECULAR SIMULATION : Applications to Chemical and Biomolecular Systems." *Annual Review of Biophysics and Biophysical Chemistry* 18: 431–92.

Bhandarkar, M et al. 2014. "NAMD User's Guide NAMD Version 2.10b1 NAMD Molecular Dynamics Software Non-Exclusive, Non-Commercial Use License." http://www.ks.uiuc.edu/Research/namd/2.10b1/ug.pdf (June 9, 2017).

Bogdanovic, Aleksandra et al. 2016. "Supercritical Carbon Dioxide Extraction of Trigonella Foenum-Graecum L. Seeds: Process Optimization Using Response Surface Methodology." *The Journal of Supercritical Fluids* 107: 44–50.

Cacace, J.E., and G. Mazza. 2003. "Mass Transfer Process during Extraction of Phenolic Compounds from Milled Berries." *Journal of Food Engineering* 59(4): 379–89.
Caddigan, E, J Cohen, J Gullingsrud, and J Stone. 2003. "VMD User's Guide." http://www.ks.uiuc.edu/Research/vmd/.

Chebil, Latifa et al. 2010. "Solubilities Inferred from the Combination of Experiment and Simulation. Case Study of Quercetin in a Variety of Solvents." *Journal of Physical Chemistry B* 114(38): 12308–13.

Cuevas-Valenzuela, José et al. 2015. "Solubility of (+)-Catechin in Water and Water-Ethanol Mixtures within the Temperature Range 277.6-331.2K: Fundamental Data to Design Polyphenol Extraction Processes." *Fluid Phase Equilibria* 382: 279–85.

Derrien, Maëlle et al. 2017. "Optimization of a Green Process for the Extraction of Lutein and Chlorophyll from Spinach by-Products Using Response Surface Methodology (RSM)." *LWT - Food Science and Technology* 79: 170–77.

Duchowicz, Pablo R., Miguel A. Giraudo, Eduardo A. Castro, and Alicia B. Pomilio. 2013. "Amino Acid Profiles and Quantitative Structure–property Relationship Models as Markers for Merlot and Torrontés Wines." *Food Chemistry* 140(1): 210–16.

Essmann, Ulrich et al. 1995. “A Smooth Particle Mesh Ewald Method.” *The Journal of Chemical Physics* 103(19): 8577–93.

Feller, Scott E., Yuhong Zhang, Richard W. Pastor, and Bernard R. Brooks. 1995. “Constant Pressure Molecular Dynamics Simulation: The Langevin Piston Method.” *The Journal of Chemical Physics* 103(11): 4613–21.

Ferrazzano, Gianmaria et al. 2011. “Plant Polyphenols and Their Anti-Cariogenic Properties: A Review.” *Molecules* 16(12): 1486–1507.

Frenkel, Daan, and Berend Smit. 2002. “Chapter 7. Free Energy Calculations.” In *Understanding Molecular Simulation*, San Diego: Academic Press, 167–200.

Halliwell, Barry. 2009. “Establishing the Significance and Optimal Intake of Dietary Antioxidants: The Biomarker Concept.” *Nutrition Reviews* 57(4): 104–13.

Hariharan, P. C., and J. A. Pople. 1973. “The Influence of Polarization Functions on Molecular Orbital Hydrogenation Energies.” *Theoretica Chimica Acta* 28(3): 213–22.

Held, Christoph et al. 2014. “ePC-SAFT Revised.” *Chemical Engineering Research and Design* 92(12): 2884–97.

Heleno, Sandrina A. et al. 2016. “Optimization of Ultrasound-Assisted Extraction to Obtain Mycosterols from Agaricus Bisporus L. by Response Surface Methodology and Comparison with Conventional Soxhlet Extraction.” *Food Chemistry* 197: 1054–63.

Hénin, Jérôme, James Gumbart, and Christophe Chipot. 2014. “Free Energy Calculations along a Reaction Coordinate: A Tutorial for Adaptive Biasing Force Simulations Adaptive Biasing Force Tutorial.” http://www.ks.uiuc.edu/Training/TutorialsOverview/namd/ABF/tutorial-abf.pdf (June 9, 2017).

Humphrey, William, Andrew Dalke, and Klaus Schulten. 1996. “VMD: Visual Molecular Dynamics.” *Journal of Molecular Graphics* 14(1): 33–38.

Ji, Peijun, Wei Feng, Tianwei Tan, and Danxing Zheng. 2007. “Modeling of Water Activity, Oxygen Solubility and Density of Sugar and Sugar Alcohol Solutions.” *Food Chemistry* 104(2): 551–58.

Jorgensen, William L. et al. 1983. “Comparison of Simple Potential Functions for Simulating Liquid Water.” *The Journal of Chemical Physics* 79(2): 926–35.

Landete, J. M. 2012. “Updated Knowledge about Polyphenols: Functions,

Bioavailability, Metabolism, and Health." *Critical Reviews in Food Science and Nutrition* 52(10): 936–48.

Leopoldini, Monica, Nino Russo, and Marirosa Toscano. 2011. "The Molecular Basis of Working Mechanism of Natural Polyphenolic Antioxidants." *Food Chemistry* 125(2): 288–306.

Lipinski, Christopher A., Franco Lombardo, Beryl W. Dominy, and Paul J. Feeney. 1997. "Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings." *Advanced Drug Delivery Reviews* 23(1–3): 3–25.

Manach, Claudine et al. 2004. "Polyphenols: Food Sources and Bioavailability." *The American journal of clinical nutrition* 79(5): 727–47.

Miller, David J, and Steven B Hawthorne. 2000. "Solubility of Liquid Organics of Environmental Interest in Subcritical (Hot / Liquid) Water from 298 K to 473 K." *Journal of Chemical and Engineering Data* 45: 315–18.

"namd2_ti.pl." http://www.ks.uiuc.edu/Research/namd/utilities/ (June 9, 2017).

Neveu, V. et al. 2010. "Phenol-Explorer: An Online Comprehensive Database on Polyphenol Contents in Foods." *Database* 2010: bap024.

Paluch, Andrew S., and Edward J. Maginn. 2013. "Predicting the Solubility of Solid Phenanthrene: A Combined Molecular Simulation and Group Contribution Approach." *AIChE Journal* 59(7): 2647–61.

Pérez-Jiménez, J, V Neveu, F Vos, and A Scalbert. 2010. "Identification of the 100 Richest Dietary Sources of Polyphenols: An Application of the Phenol-Explorer Database." *European Journal of Clinical Nutrition* 64: S112–20.

Phillips, James C. et al. 2005. "Scalable Molecular Dynamics with NAMD." *Journal of Computational Chemistry* 26(16): 1781–1802.

Prausnitz, JM, RN Lichtenthaler, and EG de Azevedo. 1998. *Molecular Thermodynamics of Fluid-Phase Equilibria*. 3rd ed. ed. Neal R. Amundson. Upper Saddle River: Prentice Hall.

Price, Sarah L. 2014. "Predicting Crystal Structures of Organic Compounds." *Chem. Soc. Rev.* 43(7): 2098–2111.

Rocha, L.G., J.R.G.S. Almeida, R.O. Macêdo, and J.M. Barbosa-Filho. 2005. "A

Review of Natural Products with Antileishmanial Activity." *Phytomedicine* 12(6): 514–35.

Roriz, Custódio Lobo et al. 2017. "Floral Parts of Gomphrena Globosa L. as a Novel Alternative Source of Betacyanins: Optimization of the Extraction Using Response Surface Methodology." *Food Chemistry* 229: 223–34.

Sadus, Richard J. 2002. *Molecular Simulation of Fluids : Theory, Algorithms, and Object-Orientation.* 1st ed. Amsterdam: Elsevier.

Salahinejad, Maryam, Tu C. Le, and David A. Winkler. 2013. "Aqueous Solubility Prediction: Do Crystal Lattice Interactions Help?" *Molecular Pharmaceutics* 10(7): 2757–66.

Schnieders, Michael J. et al. 2012. "The Structure, Thermodynamics, and Solubility of Organic Crystals from Simulation with a Polarizable Force Field." *Journal of Chemical Theory and Computation* 8(5): 1721–36.

Shelnutt, Susan R., Carolyn O. Cimino, Patricia A. Wiggins, and Thomas M. Badger. 2000. "Urinary Pharmacokinetics of the Glucuronide and Sulfate Conjugates of Genistein and Daidzein." *Cancer Epidemiology and Prevention Biomarkers* 9(4).

Sticher, Otto et al. 2008. "Natural Product Isolation." *Natural Product Reports* 25(3): 517.

Stoclet, Jean-Claude et al. 2004. "Vascular Protection by Dietary Polyphenols." *European Journal of Pharmacology* 500(1–3): 299–313.

Taylor, John R. 1997. "Chapter 3. Propagation of Uncertainties." In *An Introduction to Error Analysis : The Study of Uncertainties in Physical Measurements*, University Science Books, 45–79.

Vanommeslaeghe, K. et al. 2009. "CHARMM General Force Field: A Force Field for Drug-like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields." *Journal of Computational Chemistry* 31(4): 671–90.

Watowich, Stanley J., Eric S. Meyer, Ray Hagstrom, and Robert Josephs. 1988. "A Stable, Rapidly Converging Conjugate Gradient Method for Energy Minimization." *Journal of Computational Chemistry* 9(6): 650–61.

Wavefunction. 2010. "Spartan '10."

Wong, Ka H. et al. 2017. "Optimisation of Pueraria Isoflavonoids by Response Surface

Methodology Using Ultrasonic-Assisted Extraction." *Food Chemistry* 231: 231–37.

Zhang, Jin, Badamkhatan Tuguldur, and David van der Spoel. 2015. "Force Field Benchmark of Organic Liquids. 2. Gibbs Energy of Solvation." *Journal of Chemical Information and Modeling* 55(6): 1192–1201.

APPENDIX

APPENDIX A: MODEL FOR SOLUBILITY

A.1. Definitions

a) Fugacity: Fugacity is defined, for an isothermal change for any component in any system, solid, liquid, or gas, pure or mixed, ideal or not, as follows:

$$\mu_i - \mu_i^0 = RTln\frac{f_i}{f_i^0} \tag{A-1}$$

Where T is the temperature of the system and R is the universal gas constant. Either μ_i^0 or f_i^0 is arbitrary, but both may not be chosen independently; when one is chosen, the other is fixed (Prausnitz, Lichtenthaler, and Azevedo 1998).

b) Chemical Potential (Prausnitz, Lichtenthaler, and Azevedo 1998):

$$\mu_i = \left(\frac{\partial G}{\partial n_i}\right)_{T,P} \tag{A-2}$$

c) Residual Chemical Potential: We can separate the chemical potential in an ideal gas contribution μ_i^{ig} and the residual part μ_i^{res} (Frenkel and Smit 2002).

$$\mu_i = \mu_i^{ig} + \mu_i^{res} \tag{A-3}$$

d) Solubilities of solids in liquids: the solubility of a solid, nonelectrolyte solute (component 2) in a single component solvent (component 1) is described by the equilibrium of the fugacities of component 2 in both phases. Here we are assuming that the components 1 and 2 are miscible only in the liquid phase, but not in the solid phase.

$$f_{2(pure\ solid)} = f_{2(solid\ in\ liquid\ solution)} \tag{A-4}$$

Or

$$f_{2(pure\ solid)} = \gamma_2 x_2 f_2^0 \tag{A-5}$$

where x_2 is the solubility (mole fraction) of the solute in the solvent, γ_2 is the liquid phase activity coefficient, and f_2^0 is the standard-state fugacity to which γ_2 refers (Prausnitz, Lichtenthaler, and Azevedo 1998).

Eq.5 can lead to the following relations (Paluch and Maginn 2013):

$$x_2\gamma_2(T,p,x_2) = \frac{f_2^L(T,p,x_2)}{f_2^0(T,p)} = \frac{f_2^S(T,p,x_2)}{f_2^0(T,p)} \quad \text{(A-6)}$$

Particularly, the leftmost relation is relevant for the purpose of this method. Rearranging it, eq. 7 is obtained:

$$f_2^L(T,p,x_2) = x_2\gamma_2(T,p,x_2)f_2^0(T,p) \quad \text{(A-7)}$$

e) Fugacity of a solid: Prausnitz et al., (1998) derived the following expression for the fugacity of a solid or a liquid:

$$f_i^s(T,p) = p_i^{sat}\exp\left(\frac{v^s[p-p_i^{sat}]}{RT}\right) \quad \text{(A-8)}$$

Where $f_i^s(T,p)$ is the fugacity of the solid (compound i) at pressure p and temperature T, p_i^{sat} is its vapor pressure at T, v^s is the molar volume of the solid at T and p, and R is the universal gas constant. This expression is valid only at conditions remote from critical, where the solid phase can be considered as incompressible. Also, it is required that the saturation pressure of the solid is lower than 1 bar, which normally is the case of compounds which behave as solids or liquids at normal conditions.

f) Vapor pressure: the vapor pressure of a solid or liquid at temperature T can be determined using the Clausius-Clapeyron relation:

$$p_i^{sat} = \exp\left(-\frac{\Delta H_i^s(T)}{R}\left(\frac{1}{T}\right) + \frac{\Delta S_i^s(T)}{R}\right) \quad \text{(A-9)}$$

where $\Delta H_i^s(T)$ and $\Delta S_i^s(T)$ are the sublimation entalpy and entropy, respectively at temperature T and R is the universal gas constant.

A.2. Model

Equation A-3 can be rearranged as follows:

$$\mu_i^{res} = \mu_i - \mu_i^{ig} \tag{A-10}$$

Writing equation A-10 in terms of the quantities relevant for simulation:

$$\mu_2^{res}(T, p, N_1, N_2) = \mu_2(T, p, N_1, N_2) - \mu_2^{ig}(T, \langle V \rangle_{T,p,N_1,N_2-1}, N_2) \tag{A-11}$$

where N1 and N2 are the number of molecules of components 1 (solvent) and 2 (solute), respectively, μ_2 is the chemical potential of component 2 in solution at the same conditions as μ_2^{res} , and μ_2^{ig} is the chemical potential of component 2 in a hypothetical, non-interacting, ideal gas state at the same T, but at a fixed density. The relevant volume is given by the ensemble average volume of the system in the absence of the solute molecule that is being coupled/decoupled to the system, $\langle V \rangle_{T,p,N_1,N_2-1}$.

Taking the ideal gas state as a reference state for equation A-1, equation A-11 can be replaced in it:

$$\mu_2^{res}(T, p, N_1, N_2) = RT \ln \frac{f_2^L(T, p, x_2)}{f_2^{ig}\left(T, \langle V \rangle_{T,p,N_1,N_2-1}, N_2\right)} \tag{A-12}$$

where f_2^{ig} is the ideal gas fugacity of component 2 at the same conditions as μ_2^{ig} in equation A-11. The ideal gas fugacity of component 2 is equivalent to the partial pressure of component 2 in a hypothetical ideal gas state (Prausnitz, Lichtenthaler, and Azevedo 1998), $f_2^{ig} = p_2^{ig}$. Use of the ideal gas equation of state leads to:

$$f_2^{ig}\left(T, \langle V \rangle_{T,p,N_1,N_2-1}, N_2\right) = p_2^{ig}\left(T, \langle V \rangle_{T,p,N_1,N_2-1}, N_2\right) = \frac{N_2 RT}{\langle V \rangle_{T,p,N_1,N_2-1}} \tag{A-13}$$

$$= x_2 \frac{(N_1 + N_2)RT}{\langle V \rangle_{T,p,N_1,N_2-1}}$$

where the definition of the mole fraction of Component 2, $x_2 = N_2/(N_1 + N_2)$ was used in the last term.

Substituting equation A-13 for f_2^{ig} and equation A-7 for f_2^L into equation A-12:

$$\mu_2^{res}(T,p,N_1,N_2) = RTln\frac{x_2\gamma_2(T,p,x_2)f_2^0(T,p)}{x_2\frac{(N_1+N_2)RT}{\langle V\rangle_{T,p,N_1,N_2-1}}} \tag{A-14}$$

Rearranging:

$$\left(\frac{1}{RT}\right)\mu_2^{res}(T,p,N_1,N_2) = \ln\left(\gamma_2(T,p,x_2)\right) + \ln\left(f_2^0(T,p)\right) - \ln\left(\frac{(N_1+N_2)RT}{\langle V\rangle_{T,p,N_1,N_2-1}}\right) \tag{A-15}$$

Note that f_2^0 is a solute-dependent constants, and the other terms are composition-dependent properties of the solution that can be computed from a liquid-phase simulation.

And rearranging gives the following expression for the activity coefficient of component 2:

$$\ln\left(\gamma_2(T,p,x_2)\right) = \left(\frac{1}{RT}\right)\mu_2^{res}(T,p,N_1,N_2) + \ln\left(\frac{(N_1+N_2)RT}{\langle V\rangle_{T,p,N_1,N_2-1}}\right) - \ln\left(f_2^0(T,p)\right) \tag{A-16}$$

Assuming that solute–solute interactions are negligible, and the mole fraction of the solute in solution is sufficiently small so as to be considered infinitely dilute. That is, assume $x_2 \to 0$. This is realized in a simulation by adding a single-solute molecule ($N_2 = 1$) to a system of pure solvent for which $N_1 \gg N_2$. It follows that equation A-16 becomes:

$$\begin{aligned}\ln\left(\gamma_2(T,p,x_2)\right) &\approx \ln\left(\gamma_2^{\infty}(T,p,0)\right) \\ &= \left(\frac{1}{RT}\right)\mu_2^{res,\infty}(T,p,N_1,N_2=1) + \ln\left(\frac{(N_1+1)RT}{\langle V\rangle_{T,p,N_1}}\right) - \ln\left(f_2^0(T,p)\right)\end{aligned} \tag{A-17}$$

Where the superscript ∞ corresponds to properties for an infinitely dilute solute ($x_1 \to 1$ and $x_2 \to 0$). For an infinitely dilute solute, $N_1 \gg N_2$ such that:

$$\frac{(N_1+1)RT}{\langle V\rangle_{T,p,N_1}} \approx \frac{N_1RT}{\langle V\rangle_{T,p,N_1}} = \frac{RT}{v_1(T,p)} \tag{A-18}$$

where v_1 is the intensive molar volume of the solvent. Equation A-17 may be rewritten as:

$$\ln\left(\gamma_2(T,p,x_2)\right) \approx \ln\left(\gamma_2^{\infty}(T,p,0)\right) \tag{A-19}$$
$$= \left(\frac{1}{RT}\right)\mu_2^{res,\infty}(T,p,N_1,N_2=1) + \ln\left(\frac{(N_1+1)RT}{\langle V\rangle_{T,p,N_1}}\right) - \ln\left(f_2^0(T,p)\right)$$

Rearranging:

$$\ln\left(\gamma_2^{\infty}(T,p,0)f_2^0(T,p)\right) = \left(\frac{1}{RT}\right)\mu_2^{res,\infty}(T,p,N_1,N_2=1) + \ln\left(\frac{(N_1+1)RT}{\langle V\rangle_{T,p,N_1}}\right) \tag{A-20}$$

From equation A-6, the following expression can be obtained:

$$\gamma_2(T,p,x_2)f_2^0(T,p) = \frac{f_2^S(T,p,x_2)}{x_2} \tag{A-21}$$

Replacing equation A-21 in equation A-20:

$$\ln\left(\frac{f_2^S(T,p,x_2)}{x_2}\right) = \left(\frac{1}{RT}\right)\mu_2^{res,\infty}(T,p,N_1,N_2=1) + \ln\left(\frac{(N_1+1)RT}{\langle V\rangle_{T,p,N_1}}\right) \tag{A-22}$$

And clearing x_2:

$$x_2 = \frac{f_2^s(T,p)\,\nu_1(T,p)}{RT}\exp\left(-\left(\frac{1}{RT}\right)\mu_2{}^{\alpha,res}(T,p,N_1,N_2=1)\right) \tag{A-23}$$

Equation A-8 can be applied for the fugacity of the solid:

$$x_2 = \frac{p_i^{sat}\exp\left(\frac{v^s\left[p-p_i^{sat}\right]}{RT}\right)\nu_1(T,p)}{RT}\exp\left(-\left(\frac{1}{RT}\right)\mu_2{}^{\alpha,res}(T,p,N_1,N_2=1)\right) \tag{A-24}$$

And equation A-9 can be applied for the vapor pressure of the solid:

$$x_2 = \frac{\exp\left(\frac{v^s\left[p-\exp\left(-\frac{\Delta H_2^s(T)}{R}\left(\frac{1}{T}\right)+\frac{\Delta S_2^s(T)}{R}\right)\right]}{RT} - \left(\frac{1}{RT}\right)\mu_2{}^{\alpha,res} - \frac{\Delta H_2^s(T)}{R}\left(\frac{1}{T}\right)+\frac{\Delta S_2^s(T)}{R}\right)\nu_1(T,p)}{RT} \tag{A-25}$$

For the calculation of the chemical potential $\mu_2{}^{\alpha,res}$, equation A-2 can be applied to relate it to a Gibbs free energy term, that can be obtained from simulation, ΔG_{sim}. Considering an infinite dilute system, to calculate this term the interaction of only one solute molecule surrounded by solvent molecules is simulated ($\Delta n_2 = 1$). Therefore, equation A-2 can be rearranged as follows:

$$\mu_2{}^{\alpha,res}(T,p,N_1,N_2=1) = \left(\frac{\partial G}{\partial n_2}\right)_{T,P} = \frac{\Delta G_{sim}}{\Delta n_2} = \frac{\Delta G_{sim}}{1} = \Delta G_{sim} \tag{A-26}$$

Substituting equation A-26 for $\mu_2{}^{\alpha,res}$ in equation A-25, the final expression for the solubility of an infinitely dilute system calculated from molecular simulation is:

$$x_2 = \frac{\exp\left(\frac{v^s\left[p - \exp\left(-\frac{\Delta H_2^s(T)}{R}\left(\frac{1}{T}\right) + \frac{\Delta S_2^s(T)}{R}\right)\right]}{RT} - \left(\frac{1}{RT}\right)\Delta G_{sim} - \frac{\Delta H_2^s(T)}{R}\left(\frac{1}{T}\right) + \frac{\Delta S_2^s(T)}{R}\right) v_1(T,p)}{RT} \quad \text{(A-27)}$$

Where x_2 is the solubility of the solute (compound 2) in the solvent (compound 1) at pressure p and temperature T, v^s, ΔH and ΔS are the molar volume, the enthalpy of sublimation and the enthropy of sublimation, respectively, of the solute at the same conditions, ΔG_{sim} is the free energy of hydration calculated from molecular simulation when the solute molecule vanishes (passes from solid to ideal gas state) from a system that consists in a single solute molecule surrounded by solvent molecules, $v_1(T,p)$ is the molar volume of the solvent at the same conditions and R is the universal gas constant.

APPENDIX B: PRELIMINARY CALCULATIONS TO DETERMINE THE CUTOFF RADIUS

To decide the cutoff distance, preliminary MD simulations were conducted for each of the solutes with five different cutoff distances ranging from 4 to 20 Å. It was observed that the calculated vdW interactions for cutoff distances of 12, 16 and 20 Å were approximately equivalent; therefore, the shortest of those distances (12 Å) was selected for computational efficiency.

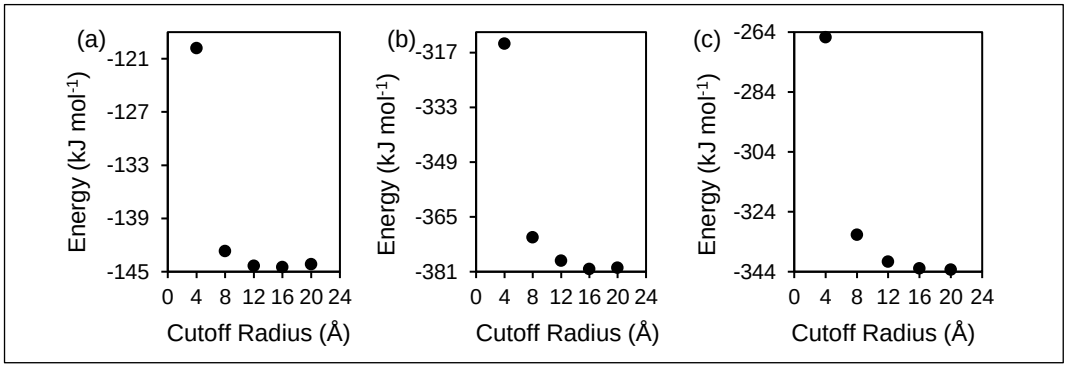

Figure B.1: Cutoff radius determination. Preliminary calculations of vdW potential energies solute-solvent interaction of benzoic acid (a), (+)-catechin (b) and toluene (c) using five different cutoff radius distances.

APPENDIX C: REPRESENTATIVE RESULT OF THE RETRIEVED ENERGY VERSUS LAMBDA CURVES

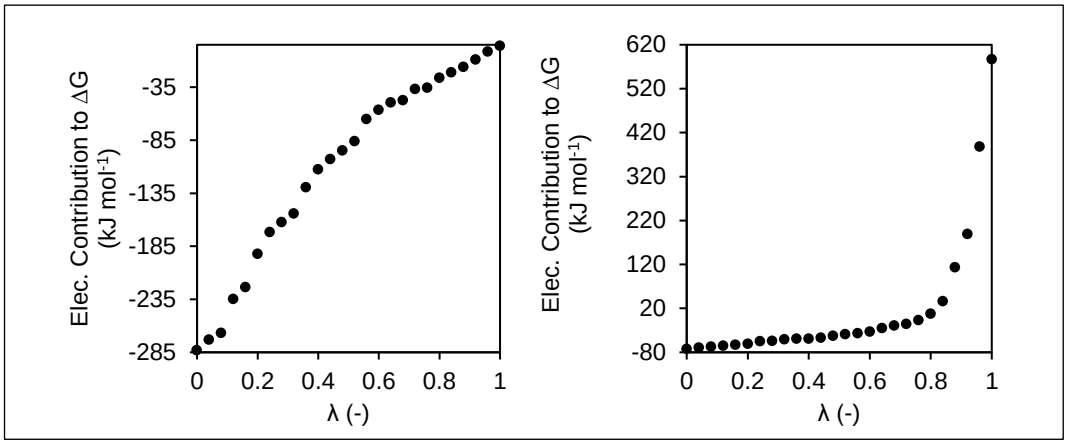

Figure C.1: Representative result of the retrieved smooth energy versus lambda curves. This case corresponds to (+)-catechin at 277.6 K.

APPENDIX D: FREE ENERGY CALCULATION RESULTS

Table D.1: Average electrostatic and vdW contributions to the free energies of solvation from replicas 1-3, denoted as R1, R2 and R3. Results are shown for benzoic acid (A), (+)-catechin (B) and toluene (C) at each simulation temperature. The averages were calculated among 1000 measurements. Errors correspond to the standard deviations calculated for the same samples by propagation of uncertainty.

(A)	Electrostatic contribution to ΔG_{solv} (kJ mol^{-1})			vdW contribution to ΔG_{solv} (kJ mol^{-1})		
T [K]	R1	R2	R3	R1	R2	R3
278.4	-36.3 ± 3.5	-33.5 ± 3.4	-33.8 ± 3.3	-313.3 ± 15.6	-310.6 ± 14.8	-320.9 ± 14.0
282.9	-73.1 ± 4.9	-67.6 ± 4.9	-69.1 ± 4.7	-309.1 ± 18.9	-305.9 ± 17.7	-320.2 ± 17.4
294.1	-107.8 ± 6.0	-105.2 ± 5.9	-106.7 ± 5.7	-304.7 ± 21.4	-297.4 ± 20.6	-318.7 ± 20.4
300.4	-144.3 ± 7.0	-141.5 ± 6.9	-142.9 ± 6.7	-296.0 ± 23.5	-299.9 ± 22.8	-322.1 ± 23.0
307.7	-178.9 ± 7.9	-175.2 ± 7.8	-176.0 ± 7.6	-291.9 ± 25.7	-295.1 ± 25.1	-319.7 ± 25.2
317.7	-212.2 ± 8.7	-207.9 ± 8.5	-210.0 ± 8.4	-277.9 ± 28.4	-286.3 ± 27.7	-310.6 ± 27.9
326.7	-245.7 ± 9.5	-242.8 ± 9.3	-245.8 ± 9.4	-271.0 ± 30.7	-269.1 ± 30.5	-302.6 ± 30.2
336.6	-281.1 ± 10.2	-277.3 ± 10.1	-280.5 ± 10.1	-267.2 ± 32.4	-258.5 ± 32.6	-296.5 ± 32.1
340.9	-315.6 ± 10.8	-310.7 ± 10.8	-316.1 ± 10.8	-262.0 ± 34.5	-252.1 ± 34.3	-290.7 ± 34.3

(B)	Electrostatic contribution to ΔG_{solv} (kJ mol^{-1})			vdW contribution to ΔG_{solv} (kJ mol^{-1})		
T [K]	R1	R2	R3	R1	R2	R3
277.6	-99.3 ± 5.4	-100.1 ± 5.3	-101.7 ± 5.5	-626.1 ± 20.8	-632.1 ± 21.8	-648.3 ± 19.8
282.7	-219.3 ± 7.8	-209.4 ± 8.0	-212.8 ± 8.0	-630.2 ± 25.6	-628.2 ± 26.0	-653.1 ± 24.8
297.7	-321.2 ± 10.1	-304.8 ± 10.1	-315.7 ± 9.8	-629.1 ± 29.7	-632.8 ± 30.5	-651.3 ± 28.7
308.4	-431.3 ± 11.9	-413.8 ± 11.7	-430.0 ± 11.8	-627.0 ± 33.6	-639.2 ± 33.7	-642.2 ± 33.6
324.9	-525.2 ± 13.4	-511.7 ± 13.0	-529.2 ± 13.3	-621.8 ± 37.7	-633.9 ± 37.3	-652.1 ± 36.2
331.2	-632.3 ± 14.9	-623.6 ± 14.5	-633.6 ± 14.7	-630.5 ± 40.6	-636.8 ± 40.3	-662.1 ± 39.8

(C)	Electrostatic contribution to ΔG_{solv} (kJ mol^{-1})			vdW contribution to ΔG_{solv} (kJ mol^{-1})		
T [K]	R1	R2	R3	R1	R2	R3
298	-10.1 ± 1.8	-10.5 ± 1.8	-10.3 ± 1.7	-40.0 ± 11.9	-42.2 ± 10.3	-32.9 ± 10.9
323	-19.9 ± 2.7	-21.1 ± 2.6	-19.8 ± 2.5	-28.9 ± 16.5	-28.7 ± 14.7	-22.8 ± 14.9
373	-28.8 ± 3.3	-31.2 ± 3.3	-29.6 ± 3.3	-19.9 ± 19.7	-15.7 ± 19.9	-15.4 ± 18.0
423	-38.0 ± 3.9	-40.3 ± 3.8	-38.2 ± 3.8	-10.2 ± 22.8	-0.6 ± 23.4	-0.3 ± 21.3
473	-45.9 ± 4.4	-48.6 ± 4.3	-45.4 ± 4.3	1.4 ± 26.6	9.1 ± 26.2	9.2 ± 24.0

Table D.2: Average electrostatic and vdW contributions to the free energies of solvation from replicas 4-6, denoted as R4, R5 and R6. Results are shown for benzoic acid (A), (+)-catechin (B) and toluene (C) at each simulation temperature. The averages were calculated among 1000 measurements. Errors correspond to the standard deviations calculated for the same samples by propagation of uncertainty.

(A)	Electrostatic contribution to ΔG_{solv} (kJ mol^{-1})			vdW contribution to ΔG_{solv} (kJ mol^{-1})		
T [K]	R4	R5	R6	R4	R5	R6
278.4	-36.8 ± 3.3	-36.2 ± 3.5	-35.5 ± 3.4	-305.7 ± 14.7	-304.3 ± 15.1	-312.2 ± 14.6
282.9	-71.2 ± 4.7	-69.1 ± 4.8	-72.4 ± 4.7	-300.8 ± 18.8	-302.0 ± 17.8	-306.7 ± 18.0
294.1	-103.3 ± 5.8	-103.3 ± 5.9	-109.2 ± 5.9	-297.4 ± 21.1	-299.2 ± 20.3	-300.8 ± 20.8
300.4	-139.2 ± 6.8	-137.5 ± 6.9	-143.6 ± 7.0	-296.8 ± 23.6	-304.0 ± 22.5	-297.8 ± 23.3
307.7	-173.8 ± 7.7	-171.8 ± 7.8	-176.0 ± 7.8	-292.6 ± 25.8	-296.1 ± 25.4	-293.6 ± 25.1
317.7	-207.2 ± 8.5	-206.0 ± 8.6	-210.2 ± 8.6	-279.5 ± 28.1	-291.0 ± 27.5	-288.7 ± 27.4
326.7	-241.4 ± 9.3	-238.7 ± 9.3	-245.4 ± 9.4	-273.5 ± 29.9	-290.3 ± 29.2	-284.8 ± 29.3
336.6	-274.8 ± 10.0	-272.2 ± 10.1	-280.2 ± 10.0	-268.0 ± 31.8	-278.7 ± 31.5	-274.8 ± 31.2
340.9	-309.4 ± 10.6	-307.0 ± 10.8	-314.0 ± 10.6	-264.4 ± 33.3	-270.1 ± 33.3	-263.6 ± 33.1

(B)	Electrostatic contribution to ΔG_{solv} (kJ mol^{-1})			vdW contribution to ΔG_{solv} (kJ mol^{-1})		
T [K]	R4	R5	R6	R4	R5	R6
277.6	-99.0 ± 5.0	-97.5 ± 5.5	-98.9 ± 5.6	-645.1 ± 20.6	-642.3 ± 19.6	-631.0 ± 20.5
282.7	-216.3 ± 7.7	-212.4 ± 8.1	-208.5 ± 8.0	-637.4 ± 25.2	-643.2 ± 24.0	-628.8 ± 25.1
297.7	-318.0 ± 9.7	-311.1 ± 9.7	-306.1 ± 10.1	-637.9 ± 29.4	-639.4 ± 28.5	-632.9 ± 29.2
308.4	-425.4 ± 11.2	-422.1 ± 11.5	-416.5 ± 11.6	-646.7 ± 33.0	-640.5 ± 31.7	-633.1 ± 32.9
324.9	-519.5 ± 12.6	-526.6 ± 13.0	-515.1 ± 13.1	-636.3 ± 36.4	-644.1 ± 35.0	-628.6 ± 37.1
331.2	-632.3 ± 14.1	-633.6 ± 14.4	-629.1 ± 14.6	-644.4 ± 38.9	-646.8 ± 39.1	-626.4 ± 39.9

(C)	Electrostatic contribution to ΔG_{solv} (kJ mol^{-1})			vdW contribution to ΔG_{solv} (kJ mol^{-1})		
T [K]	R4	R5	R6	R4	R5	R6
298	-9.7 ± 1.8	-9.9 ± 1.8	-9.6 ± 1.8	-30.3 ± 11.8	-40.7 ± 10.8	-36.8 ± 10.1
323	-19.1 ± 2.6	-21.3 ± 2.7	-20.5 ± 2.6	-16.0 ± 15.8	-33.6 ± 14.6	-24.2 ± 14.4
373	-28.1 ± 3.2	-31.4 ± 3.4	-30.0 ± 3.3	-2.5 ± 19.6	-22.1 ± 18.3	-14.4 ± 18.1
423	-37.0 ± 3.8	-40.2 ± 3.9	-37.8 ± 3.8	11.3 ± 22.9	-16.4 ± 21.4	0.0 ± 22.2
473	-45.0 ± 4.3	-47.9 ± 4.4	-46.2 ± 4.3	20.8 ± 25.5	-5.6 ± 25.7	18.5 ± 26.1

Table D.3: Average electrostatic and vdW contributions to the free energies of solvation from replicas 7-9, denoted as R7, R8 and R9. Results are shown for benzoic acid (A), (+)-catechin (B) and toluene (C) at each simulation temperature. The averages were calculated among 1000 measurements. Errors correspond to the standard deviations calculated for the same samples by propagation of uncertainty.

(A)	Electrostatic contribution to ΔG_{solv} (kJ mol^{-1})						vdW contribution to ΔG_{solv} (kJ mol^{-1})					
T [K]	R7			R8			R9			R7			R8			R9		
278.4	-36.3	±	3.4	-36.0	±	3.6	-36.6	±	3.4	-306.7	±	15.8	-311.4	±	15.3	-314.2	±	14.3
282.9	-71.1	±	4.9	-71.5	±	4.9	-71.4	±	4.8	-303.3	±	19.2	-312.2	±	18.0	-311.7	±	17.3
294.1	-106.7	±	6.0	-105.5	±	6.0	-106.3	±	6.0	-299.2	±	21.7	-314.8	±	20.4	-310.8	±	19.9
300.4	-141.6	±	7.0	-138.2	±	7.0	-140.4	±	6.9	-298.6	±	23.9	-307.5	±	23.8	-309.6	±	22.3
307.7	-177.0	±	7.9	-174.1	±	7.9	-173.7	±	7.7	-297.9	±	26.4	-301.1	±	26.1	-306.3	±	25.1
317.7	-211.7	±	8.7	-209.4	±	8.7	-208.9	±	8.5	-296.2	±	28.2	-292.8	±	28.1	-301.4	±	27.1
326.7	-244.8	±	9.5	-244.5	±	9.5	-245.6	±	9.3	-293.0	±	30.3	-295.1	±	30.1	-299.5	±	29.1
336.6	-280.3	±	10.2	-278.7	±	10.1	-281.5	±	10.0	-285.8	±	32.3	-286.5	±	32.0	-294.9	±	31.1
340.9	-315.2	±	10.8	-314.1	±	10.8	-315.4	±	10.7	-281.4	±	34.3	-275.4	±	33.9	-285.7	±	32.8

(B)	Electrostatic contribution to ΔG_{solv} (kJ mol^{-1})						vdW contribution to ΔG_{solv} (kJ mol^{-1})					
T [K]	R7			R8			R9			R7			R8			R9		
277.6	-97.3	±	5.3	-100.1	±	5.5	-99.9	±	5.5	-642.2	±	20.2	-644.9	±	20.2	-632.7	±	20.0
282.7	-211.8	±	7.7	-212.0	±	7.9	-203.5	±	7.9	-664.4	±	24.2	-652.1	±	24.4	-639.0	±	24.4
297.7	-310.5	±	9.5	-310.6	±	9.7	-302.1	±	9.9	-666.8	±	27.7	-659.2	±	29.1	-633.9	±	28.8
308.4	-432.9	±	11.3	-430.0	±	11.6	-412.6	±	11.8	-658.0	±	32.5	-658.6	±	33.5	-627.0	±	33.1
324.9	-525.1	±	12.7	-524.1	±	13.0	-505.3	±	13.1	-656.3	±	35.9	-646.4	±	38.6	-625.7	±	36.8
331.2	-640.8	±	14.3	-636.0	±	14.6	-618.8	±	14.7	-655.3	±	39.4	-644.0	±	41.5	-613.0	±	41.0

(C)	Electrostatic contribution to ΔG_{solv} (kJ mol^{-1})						vdW contribution to ΔG_{solv} (kJ mol^{-1})					
T [K]	R7			R8			R9			R7			R8			R9		
298	-10.1	±	1.8	-10.0	±	1.8	-10.2	±	1.7	-39.9	±	12.0	-39.1	±	11.2	-36.8	±	10.2
323	-20.7	±	2.7	-20.2	±	2.6	-20.1	±	2.5	-28.7	±	16.1	-28.4	±	15.1	-28.2	±	14.5
373	-31.2	±	3.3	-30.2	±	3.3	-29.1	±	3.2	-17.3	±	19.8	-20.9	±	18.4	-14.6	±	18.9
423	-39.9	±	3.8	-39.6	±	3.8	-37.3	±	3.7	-6.8	±	22.7	-3.6	±	22.4	-0.9	±	23.4
473	-48.1	±	4.3	-47.9	±	4.4	-45.5	±	4.2	5.7	±	26.0	9.2	±	25.3	14.0	±	26.6

Table D.4: Average electrostatic and vdW contributions to the free energies of solvation from replica 10, denoted as R10. Results are shown for benzoic acid (A), (+)-catechin (B) and toluene (C) at each simulation temperature. The averages were calculated among 1000 measurements. Errors correspond to the standard deviations calculated for the same samples by propagation of uncertainty.

(A)	Electrostatic contribution to ΔG_{solv} (kJ mol^{-1})	vdW contribution to ΔG_{solv} (kJ mol^{-1})
T [K]	R10	R10
278.4	-34.7 ± 3.3	-312.1 ± 14.2
282.9	-69.6 ± 4.7	-305.1 ± 17.4
294.1	-103.2 ± 5.9	-294.9 ± 20.4
300.4	-138.0 ± 6.9	-298.1 ± 22.7
307.7	-171.8 ± 7.7	-293.8 ± 25.0
317.7	-206.6 ± 8.6	-285.4 ± 27.5
326.7	-241.7 ± 9.4	-281.5 ± 29.5
336.6	-275.9 ± 10.1	-278.6 ± 32.0
340.9	-310.7 ± 10.8	-275.9 ± 34.2

(B)	Electrostatic contribution to ΔG_{solv} (kJ mol^{-1})	vdW contribution to ΔG_{solv} (kJ mol^{-1})
T [K]	R10	R10
277.6	-105.0 ± 5.2	-636.2 ± 20.8
282.7	-220.0 ± 8.0	-635.4 ± 25.2
297.7	-316.3 ± 10.0	-634.4 ± 29.7
308.4	-432.9 ± 11.8	-635.7 ± 33.3
324.9	-531.9 ± 13.2	-627.3 ± 37.6
331.2	-635.7 ± 14.6	-618.7 ± 41.0

(C)	Electrostatic contribution to ΔG_{solv} (kJ mol^{-1})	vdW contribution to ΔG_{solv} (kJ mol^{-1})
T [K]	R10	R10
298	-10.3 ± 1.8	-38.2 ± 10.8
323	-20.0 ± 2.6	-26.9 ± 15.8
373	-30.7 ± 3.2	-12.9 ± 19.1
423	-39.6 ± 3.8	5.5 ± 22.7
473	-47.2 ± 4.2	19.8 ± 25.6