Universidade de Aveiro
2019

Departamento de Química

João António Pereira Iglésias

Difusividades de compostos bioativos em líquidos comprimidos: Simulação de dinâmica molecular

Diffusivities of bioactive compounds in compressed liquids: Molecular dynamics simulation

Universidade de Aveiro Departamento de Química
2019

João António Pereira Iglésias

Difusividades de compostos bioativos em líquidos comprimidos: Simulação de dinâmica molecular

Diffusivities of bioactive compounds in liquids comprimidos: Molecular dynamics simulation

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Engenharia Química, realizada sob a orientação científica do Doutor José Richard Baptista Gomes, Equiparado a Investigador Principal do Departamento de Química da Universidade de Aveiro e coorientação do Professor Doutor Carlos Manuel Santos da Silva, Professor Associado do Departamento de Química da Universidade de Aveiro.

o júri

presidente

Professora Doutora Maria Inês Purcell de Portugal Branco
Professora Auxiliar do Departamento de Química da Universidade de Aveiro

Doutor António Augusto Areosa Martins
Investigador de Pós-Doutoramento no Laboratório de Engeharia de Processos, Ambiente, Biotecnlogia e Energia da Faculdade de Engenharia da Universidade do Porto.

Doutor José Richard Baptista Gomes
Equiparado a Investigador Principal do Departamento de Química da Universidade de Aveiro

agradecimentos

Gostaria de exprimir algumas palavras de agradecimento a todo o grupo EgiChem pelo acolhimento e apoio durante a realização desta dissertação. Um obrigado destacado ao Mestre Bruno Zêzere por me acompanhar nesta aventura que é a dinâmica molecular.

Um obrigado especial ao meu orientador Doutor José Richard Baptista Gomes pela introdução no mundo da química computacional e pelo seu permanente bom humor. E ao meu co-orientador Professor Doutor Carlos Manuel Santos da Silva por tudo que me ensinou ao longo do curso e pela ajuda na escolha e realização da dissertação.

Agradeço também a todos os que partilharam a sua amizade comigo e me apoiaram ao longo do tempo e acima de tudo à minha mãe e irmãos pela paciência de me aturar desde os primórdios da minha existência.

palavras-chave

Coeficiente de difusão, dinâmica molecular, propriedades de transporte, quercetina, etanol, acetato de etilo.

resumo

A quercetina, um flavonoide que pode ser extraído de várias fontes vegetais, apresenta bioatividade interessante, devido às suas boas propriedades antioxidantes ou anti-carcinogénicas. Uma das técnicas de extração deste flavonoide é por extração sólido-líquido, usando, por exemplo, solventes *verdes* como etanol (EtOH) ou acetato de etilo (EtOAc), utilizados na indústria alimentar.

A difusividade, D_{12}, é uma propriedade importante nas extrações sólido-líquido pois, muito frequentemente, estas operações unitárias encontram-se limitadas pela cinética de transferência de massa, sendo assim relevante conhecer o coeficiente de difusão para o projeto e otimização destes processos.

As difusividades da quercetina em acetato de etilo e em etanol foram medidas pelo método cromatográfico de abertura de pico (CPB), na gama de temperaturas de 303,15 a 333,15 K e de pressões de 1 a 150 bar. No caso do etanol, os valores de D_{12} já tinham sido medidos anteriormente, no mesmo laboratório, por outro investigador. Os valores experimentais de D_{12} da quercetina em etanol encontram-se entre $3{,}985\times10^{-6}$ e $7{,}826\times10^{-6}$ $cm^2\ s^{-1}$, e no caso da quercetina em acetato de etilo entre $1{,}018\times10^{-5}$ e $1{,}628\times10^{-5}$ $cm^2\cdot s^{-1}$.

Os resultados experimentais obtidos seguem as dependências esperadas com a temperatura e pressão, nomeadamente, derivadas positivas e negativas, sendo a variação com a temperatura muito mais expressiva.

Paralelamente, foram realizadas simulações de dinâmica molecular (MD) clássica utilizando o software GROMACS para estimar as difusividades e averiguar a possibilidade de utilizar esta técnica computacional para calcular valores de D_{12} noutras condições de pressão e de temperatura. Com este fim em vista, foram testados diferentes conjuntos de parâmetros em simulações no *ensemble* NVT, tais como o raio de corte das interações de curto alcance, número de moléculas de solvente e duração da simulação, para analisar a sua influência na exatidão das estimativas. A otimização dos parâmetros usados nas simulações de MD conduziu a valores de D_{12} em boa concordância com os dados experimentais no caso do etanol a 1 bar, com erro relativo inferior a 6.54 %. Foi ainda demonstrado que a pressões mais elevadas é possível obter valores fiáveis de D_{12}, introduzindo um fator multiplicativo nas cargas dos átomos do etanol. No caso do acetato de etilo, o erro a 30 °C e 1 bar foi −22.51 %. Como os valores do coeficiente de auto-difusão do acetato de etilo, estimados por MD, diferem bastante dos valores experimentais, sugere-se neste trabalho a otimização dos parâmetros do campo de forças utilizado para modelar este solvente.

A concordância entre as difusividades da quercetina em etanol medidas e estimadas por MD clássica demonstra que é realmente possível obter valores fiáveis de D_{12} por esta técnica computacional. Sugerem-se estudos adicionais focados em diferentes grupos funcionais e estruturas de flavonoides, através de análises estruturais usando funções de distribuição radial e de distribuição espacial.

keywords

Diffusion coefficient, molecular dynamics, transport properties, quercetin, ethanol, ethyl acetate.

abstract

Quercetin, a flavonoid that can be extracted from various plant sources, exhibit interesting bioactivity due to relevant antioxidant or anti-carcinogenic properties. One way of extracting this flavonoid is by solid-liquid extraction using, for example, green solvents like ethanol (EtOH) or ethyl acetate (EtOAc), which are well accepted in the food industry.

Diffusivity, D_{12}, is an important property in solid-liquid extraction, since this separation is frequently limited by mass transfer kinetics, which requires the knowledge of D_{12} for the accurate design and optimization of that unitary operation.

The diffusivities, D_{12}, of quercetin in ethyl acetate and ethanol were measured by the chromatographic peak broadening (CPB) method in the temperature range 30-60 °C and pressure range 1-150 bar. The diffusivities in ethanol were measured in the same laboratory by another researcher. The D_{12} values ranged from 3.985×10^{-6} to 7.826×10^{-6} cm^2 s^{-1}, in the case of ethanol, and 1.018×10^{-5} to 1.628×10^{-5} cm^2 s^{-1} for ethyl acetate.

The obtained D_{12} data followed the expected trends with temperature and pressure, namely, positive and negative derivatives, being the influence of temperature much more significant.

In parallel, classical molecular dynamics (MD) simulations were performed using the GROMACS software package to estimate the diffusion coefficient in order to assess the possibility of using this computational technique to generate diffusivities for distinct pressure and temperature conditions. Different parameters sets were adopted to carry out simulations in NVT ensemble, such as the cut-off radius for short-range interactions, number of solvent and solute molecules, and simulation duration, with the objective to verify their influence on the quality of D_{12} estimates. The optimization of the parameters used in the MD simulations led to D_{12} values in good agreement with the experimental data for ethanol at 1 bar, with relative deviations less than 6.54 %. It was also shown that it is possible to obtain reliable results at higher pressures after introducing a multiplicative factor on the atoms charges of ethanol. In the case of ethyl acetate, the error at 30 °C and 1 bar was −22.51 %. Since the MD self-diffusivities of ethyl acetate also differ significantly from the experimental data, it is suggested in this work to optimize the force field parameters used to model this solvent.

The agreement found between experimental and MD quercetin diffusivities in ethanol demonstrates that it is possible to obtain reliable D_{12} values by classical MD simulations. Further studies are suggested on the influence of different functional groups and structure of other flavonoids on D_{12}, with a structural analysis using the radial distribution and spatial distribution function.

Contents

Figure Index

Table Index

Nomenclature

C	Tracer's concentration	(mol m^{-3})
$\bar{C}\ (z,t)$	Calculated average radial concentration of the solute	(mol m^{-3})
$C^{\exp}\ (z,t)$	Solute concentration at exit column	(mol m^{-3})
C_{n}	Number of carbons of the molecule	(Dimensionless)
C_{t}	Total concentration	(mol m^{-3})
D	Dispersion coefficient	(m^2 s^{-1})
D_{i}	Diffusion coefficient of particle i	(m^2 s^{-1})
D_{12}	Tracer diffusion coefficient of solute 2 through solvent 1	(m^2 s^{-1})
De	Dean number, $\mathrm{De} = \mathrm{Re}/\sqrt{\xi}$	(Dimensionless)
f	Constant in coloumb potential $f = 1/\ 4\pi\varepsilon_0$	(s^4 A^2 m^{-2} kg^{-1})
F_{i}	Force on atom i	(kg m s^{-2})
F_{ij}	Force applied on atom i by atom j	(kg m s^{-2})
H	Height of theoretical plate	(m)
$J_{\mathrm{i,z}}$	Unidimensional molar flux of substance i	(m^2 mol s^{-1})
k_{ij}	Force constant of stretching vibrations	(kg s^{-2} mol^{-1})
$\mathrm{k}_{\mathrm{ijk}}^{\theta}$	Force constant of angle vibrations	(kg m^2 s^{-2} mol^{-1})
K	Characteristic constant	(Dimensionless)
L	Length of the column	(m)
m	Total mass of solute injected	(mol)
m_{i}	Particle mass	(g)
n_{d}	Refractive index	(Dimensionless)
$N_{\mathrm{simulations}}$	Number of simulations averaged	(Dimensionless)
N_{solute}	Number of solute molecules	(Dimensionless)
N_{solvent}	Number of solvent molecules	(Dimensionless)
P	Pressure	(kg m^{-1} m^2)
P_0	Atmospheric pressure	(kg m^{-1} m^2)
q	Electric charge	(C)
r	Radial coordinate	(m)
$\boldsymbol{r}_{\mathrm{i}}$	Position vector of particle i	(m)
r_{ij}	Distance between particle i and j	(m)
r_{ij}^0	Equilibrium distance between particle i and j	(m)
R	Inner column radius	(m)

R_c	Tube coil radius	(m)
Re	Reynolds number $(\mathrm{Re} = u\,\rho\,R\,/\,\mu)$	(Dimensionless)
S	Zeroth moment	(mol dm $^{-3}$ s)
Sc	Schmidt number $\mathrm{Sc} = \mu_1/\rho_1\,\mathbf{D}_{12}$	(Dimensionless)
t	Time	(s)
t_i	Time at 10% peak height in the fitting method	(s)
t_r	Retention time	(s)
$\bar{t}$	Average retention time	(s)
T	Absolute temperature	(K)
T_c	Critical temperature	(K)
$\bar{u}$	Average linear velocity	(m s^{-1})
u_{opt}	Optimum linear velocity	(m s^{-1})
V	Potential function	(kg m^2 s^{-2} mol^{-1})
$V_{ij}(r_{ij})$	Potential of non-bonded interactions of particle i and j	(kg m^2 s^{-2} mol^{-1})
$V_{\mathrm{non-bonded}}$	Total non-bonded interactions potentials	(kg m^2 s^{-2} mol^{-1})
V_{ij}^{LJ}	Lennard-Jones potential	(kg m^2 s^{-2} mol^{-1})
$V_{ij}^{Coulomb}$	Coulomb potential	(kg m^2 s^{-2} mol^{-1})
$V_{ij}^{stretching}$	Bond stretching potential	(kg m^2 s^{-2} mol^{-1})
V_{ij}^{angle}	Bond angle potential	(kg m^2 s^{-2} mol^{-1})
V_{ijkl}^{RB}	Ryckeart-Bellemans dihedral angle potential	(kg m^2 s^{-2} mol^{-1})
$V_{ijkl}^{Fourier}$	Fourier series dihedral angle potential	(kg m^2 s^{-2} mol^{-1})
$W_{0.607}$	Peak half-width measured at 60.7% of total peak height	(s)
x_i	Molar composition of the component *i*	(Dimensionless)
z	Axial coordinate	(m)

Greek letters

$\delta(z)$	Dirac's function	(Dimensionless)
ε	Root mean square error	(Dimensionless)
ε_0	Vacuum permittivity	(s^4 A^2 m^{-2} kg^{-1})
ε_i	LJ well depth parameter of particle i	(kg m^2 s^{-2} mol^{-1})
ε_{ij}	Combined LJ well depth parameter of particles i and j	(kg m^2 s^{-2} mol^{-1})
ε_r	Relative permittivity	(Dimensionless)
ζ	Curvature ratio	(Dimensionless)

λ	Wavelength	(nm)
θ_{ijk}	Bond angle	(Dimensionless)
$\theta^{0}_{\mathrm{ijk}}$	Equilibrium bond angle	(Dimensionless)
ϕ_{ijkl}	Dihedral angle	(Dimensionless)
μ	Viscosity	(kg m^{-1} s^{-1})
μ_1	Viscosity of the solvent	(kg m^{-1} s^{-1})
ρ	Density	(kg m^{-3})
ρ_0	Density at atmospheric pressure	(kg m^{-3})
ρ_1	Density of the solvent	(kg m^{-3})
σ_{i}	Lennard-Jones diameter of particle i	(m)
σ_{ij}	Combined Lennard-Jones diameter of particle i and j	(m)

Subscripts

1	Solvent
2	Solute
11	Self
12	Binary
c	Critical property
Fourier	Fourier dihedral
i	Component or particle i
ijk	Angle composed by particles i, j and k
ijkl	Angle of the two planes composed by particles i, j, k and l
j	Component or particle j
LJ	Lennard-Jones
RB	Ryckaert-Bellemans dihedral

Superscripts

0	Equilibrium
exp	Experimental
MD	Molecular dynamics
OPLS	Optimized potentials for liquid simulations
LOPLS	Long-Optimized potentials for liquid simulations

Abbreviations

AARD	Average absolute relative deviation
AMBER	Assisted model building with energy refinement
BPR	Back pressure regulator
CPB	Chromatographic peak broadening method
GAFF	Generalized AMBER force field
GROMACS	Groningen machine for chemical simulations software package
MD	Molecular dynamics
LINCS	Linear constraint solver algorithm
LJ	Lennard-Jones
LOPLS	Long optimized potentials for liquid simulations
MD	Molecular dynamics
OPLS-AA	Optimized potentials for liquid simulations – all atom
PME	Particle mesh Ewald
RB	Ryckaert-Bellemans
RDF	Radial distribution function or pair correlation function
SDF	Spatial distribution function
UV	Ultraviolet
Vis	Visible

1. Introduction

Flavonoids are a class of polyphenol metabolites that are broadly present in plants, frequently in glycosylated or esterified forms, which have been the focus of researchers throughout the years because of their potential to act as antioxidants or as anti-tumorals, among other interesting properties [1–3]. An example of a flavonol is the quercetin molecule (Figure 1), which is found in many medicinal plants, fruits and vegetables, such as green tea and black tea leaves, onions, cranberries, apples, red grapes and celery. This compound is found both in its free form and in the form of glycosides [4].

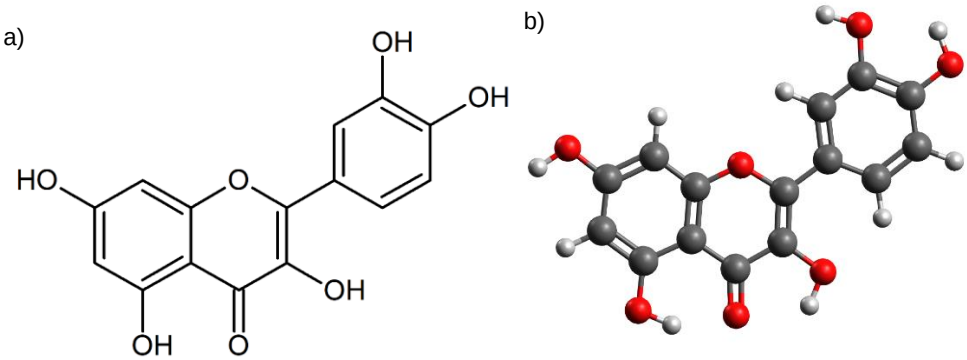


Figure 1 –a) Skeletal structural formula of the quercetin molecule made in ACD/ChemSketch [5] software. b) Molecular structure of quercetin rendered in the software Avogadro [6,7].

Quercetin has interesting biological activities, such as, anti-inflammatory, antihistamine, anti-edematous, anti-oxidant, anticancer effects, stabilizes cell membranes and inhibits the aging process of skin, cornea and myocardium [4,8,9]. It can be extracted by several methods, including solid-liquid extraction from crushed plant materials [4]. For its sustainable extraction, it is also important that the solvents used are environmentally friendly. Therefore, so-called green solvents, as ethanol or ethyl acetate are preferred [10]. Indeed, ethanol and ethyl acetate are food grade solvents, considered usable in the food industry as solvents or flavoring agent by the Join FAO/WHO Expert Committee on Food Additives (JECFA) [11–13].

The extraction methods are often limited by mass transfer phenomena, making the diffusion coefficient or diffusivity an essential parameter for the design and optimization of separation processes [14]. In this dissertation, the diffusivity of quercetin in ethyl acetate was measured experimentally, using the chromatographic peak broadening (CPB) method based on the work of Taylor and Aris [15–17], described in Chapter 2. The diffusivities of quercetin in ethanol were measured prior to the start of this dissertation by another researcher using the same method and equipment. Additionally, computer simulations were

performed to estimate such diffusion constants, that were compared with the experimental data, and to obtain structural information for an atomistic understanding of the interactions between quercetin and two different solvents.

There are three main families of approaches used in the field of computational chemistry, namely, quantum mechanics, molecular mechanics and statistical mechanics methods. Quantum mechanics methods (*ie*. *ab initio* methods) attempt to solve the Schrödinger equation to describe the properties of molecular systems with high accuracy. This family of computational chemistry approaches include the Hartree-Fock and post-Hartree-Fock wavefunction methods and the density functional theory methods, which are essential to compute properties of systems requiring the explicit consideration of electrons. The statistical mechanics methods use probability theory to average the distribution of molecular motions and states in a molecular system. This family includes the method of Monte Carlo simulation, which is applied to calculate static macroscopic properties because it considers only configurational space without the processing forces acting on the constituents of the system [18]. The molecular mechanics approaches are based on classical mechanics within force fields to compute the potential energy of systems as a function of the nuclear coordinates. The energies of the systems computed either with quantum or molecular mechanics approaches can be employed to obtain forces acting on the constituents of a molecular system. From the knowledge of the forces, it is possible to determine the acceleration of each atom in the system *e.g.* from Newton's second law, $\vec{F} = m\vec{a}$, where $\vec{F}$ is the force exerted on the particle, m is its mass and $\vec{a}$ is its acceleration. The integration of the equations of motion generates a trajectory describing the variation with time of the positions, velocities and accelerations of the constituents of a molecular system. This is the basis of a sub-family of computational tools that have found vast application in computational chemistry, the so-called molecular dynamics (MD) simulations. These computational techniques allow for the prediction of many macroscopic physical properties, either static properties (such as the potential energy or radial distribution function of a system) or dynamic properties (such as viscosity or diffusivity) [19]. Because of the enormous computational requisites needed by the quantum mechanics methods, *ab initio* MD simulations are limited to systems with tens to a few hundreds of atoms and short simulation times (up to a few picoseconds). The classical MD simulations are more straightforward and can cope with systems containing thousands of atoms and can reach simulation times in the microsecond scale.

Many software packages were used to make this dissertation possible: the "GROningen MAchine for Chemical Simulations" (GROMACS) software package [20], for MD simulations

and property calculations; the Avogadro molecular builder [6,7], for the construction of molecules; the “Visual Molecular Dynamics” (VMD) software [21], for the 3D visualization and rendering of molecular dynamics.

This dissertation is divided into six chapters. In Chapter 2 “Diffusion coefficients. Measurement methods.” are discussed the fundaments of the measurement of diffusion coefficients; In Chapter 3 “Molecular Dynamics” there is an introduction to the basics of molecular dynamics including the estimation of diffusion coefficients through that computational technique. In Chapter 4 “Materials and methods” a description of used correlations for density and viscosity, the experimental method and equipment, and the molecular dynamics procedure is presented. In Chapter 5 “Results” the molecular dynamics estimated data and measured data are discussed and compared. Lastly in Chapter 6 the main conclusions of this work are shown along future work suggestions.

2. Diffusion coefficients. Measurement methods.

There are two main mechanisms of transport in mass transfer: convection and diffusion. Convection corresponds to the macroscopic movement of a fluid, while diffusion, results from the microscopic movement of molecules, *i.e.* a random and spontaneous displacement linked to their thermal movement [22].

Diffusion can be described mathematically by Fick's first law (Eq. 2.1), suggested as an analogy to Fourier's first law of heat conduction, and derived in 1855 by Fick with three key observations [22]:

- Mass transfer occurs as consequence of a concentration gradient; in binary mixtures, the components move to regions of inferior concentration;
- The rate of mass transfer is proportional to the area normal to the direction of the mass transfer, being expressed as a flux;
- Once uniformity is achieved, the net mass transfer is null.

$$J_2 = -D_{12}\, C_{\mathrm{t}}\, \nabla x_2 \qquad \text{(Eq. 2.1)}$$

with J_2 as the flux of a substance 2, D_{12} as the diffusion coefficient of component 2 through 1, C_{t} being the total concentration and $\nabla x_2$the gradient of molar fraction of component 2.

There are several experimental techniques to measure binary diffusion coefficients in liquids and supercritical fluids [23]: i) solid dissolution technique (SD); ii) photon correlation spectroscopy (PCS); iii) nuclear magnetic resonance (NMR); iv) radioactive tracer response; and v) chromatographic peak broadening method (CPB) or Taylor dispersion method [24]. There are also two additional variants to CPB, *i.e.*, the chromatographic impulse response method (CIR) and the modified Taylor-Aris method [24]. However, these two were not used in studies included in this dissertation, which employed exclusively the CPB method.

Chromatographic peak broadening method (CPB)

The chromatographic peak broadening (CPB) method is based on work originally performed by Taylor [15,16,25] and then continued by Aris [17], which was solely devoted to study the dispersion of a solute pulse in straight tubes under laminar flow. This method was later applied to measure diffusion coefficients of solutes in gases at low and high pressures [26,27], liquids [28] and, finally, supercritical fluids [29]. This technique allows the determination of D_{12} values in a relatively short time period. Small quantities of solute are utilized, *e.g.* in the range of 3.96 x 10^{-5} – 1.32 x 10^{-3} μmol [30–32], this coefficient corresponds to diffusion at infinite dilution, as it is not possible to relate diffusion and solute concentration [23].

The general behavior of the CPB methods is exemplified in Figure 2 [24] as the pulse of solute goes through an uncoated open tube and typically produces a response of that type.

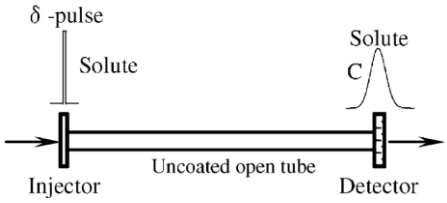


Figure 2 – Output of the CPB method to a pulse input signal. Taken from reference [24].

The concentration profile of a pulse of solute introduced in a cylindrical column of constant diameter filled with solvent (ideally the same as the carrier fluid of the pulse), assuming the physical properties are constant throughout the experiment, is [25,33]:

$$\frac{\partial C}{\partial t} = D_{12}\left[\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial C}{\partial r}\right) + \frac{\partial^2 C}{\partial z^2}\right] - 2\bar{u}\left(1 - \frac{r^2}{R^2}\right)\frac{\partial C}{\partial z} \qquad \text{(Eq. 2.2)}$$

the concentration of solute, C, can be expressed as a function of time, t, with axial coordinate, z, and radial coordinate, r, with $\bar{u}$ as the mean velocity of the solvent stream and R the internal radius of the column. Taylor demonstrated that a pulse of solute, when injected in a capillary narrow tube with circular cross section with a solvent in laminar flow, will broaden due to solvent convection along the axis of the tube and due to radial molecular diffusion [23]. The axial dispersion is negligible since it takes a very long time when compared to the radial dispersion. Therefore, Taylor proposed the following restriction in order to neglect the axial dispersion term from Eq. 2.2 [25]:

$$\frac{L}{\bar{u}} \gg \frac{R^2}{3\times 8^2 D_{12}} \qquad \text{(Eq. 2.3)}$$

with L as the column length. Applying this approximation, the fundamental equation of the process reduces to:

$$\frac{\partial C}{\partial t} = D_{12}\left[\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial C}{\partial r}\right)\right] - 2\bar{u}\left(1 - \frac{r^2}{R^2}\right)\frac{\partial C}{\partial z} \qquad \text{(Eq. 2.4)}$$

The initial and boundary conditions for this equation are the following:

at $t = 0$, $\quad C = \frac{m}{\pi R^2}\delta(z)$ (Eq. 2.5)

at $r = 0$ and $r = R$, $\quad \frac{\partial C}{\partial r} = 0$ (Eq. 2.6)

at $z = \pm\infty$, $\quad C = 0$ (Eq. 2.7)

with m as the injected amount of tracer and $\delta(z)$ as the Dirac's function.

The detector at column outlet measures the average concentration over the cross-sectional area of the tubing, $\bar{C}$, which is calculated by:

$$\bar{C} = \frac{2}{R^2}\int_0^{R} C\,(r,z,t)\,\mathrm{d}r \qquad \text{(Eq. 2.8)}$$

While $\bar{C}$ and C are not the same it has been proven that $\bar{C}$ is a good approximation. [24] Hence, Eqs. 2.4-2.7 can be reduced to [17]:

$$\frac{\partial \bar{C}}{\partial \mathrm{t}} = D\frac{\partial^2 \bar{C}}{\partial z^2} - \bar{u}\frac{\partial \bar{C}}{\partial z} \qquad \text{(Eq. 2.9)}$$

with D as the dispersion coefficient described in Taylor's work [17,34]:

$$D = D_{12} + \frac{R^2\bar{u}^2}{48D_{12}} \qquad \text{(Eq. 2.10)}$$

This simplification implies that the model is only mathematically valid for tubes of infinite length.

With the substitution of the axial variable z by $z'=z-\bar{u}t$ in Eq.2.9, it is obtained:

$$\frac{\partial \bar{C}}{\partial t} = D\frac{\partial^2 \bar{C}}{\partial z'^2} \qquad \text{(Eq. 2.11)}$$

with the boundary conditions of:

at $t = 0$, $\quad \bar{C} = C_0 = \frac{m}{\pi R^2}\delta(z')$ (Eq. 2.12)

at $z' = \pm\infty$, $\quad \bar{C} = 0$ (Eq. 2.13)

The analytical solution of the concentration profile is [34,35]:

$$\frac{\bar{C}}{C_0} = \frac{1}{\sqrt{4\pi Dt}}\exp\left[-\frac{(z-\bar{u}t)^2}{4Dt}\right] \qquad \text{(Eq. 2.14)}$$

This concentration profile can be described in terms of the peak variance in units of length, with σ^2 being the variance and H the theoretical plate height [23]:

$$\sigma^2 = \frac{2DL}{\bar{u}} = \frac{2D_{12}L}{\bar{u}} + \frac{R^2\bar{u}L}{24D_{12}} = HL \qquad \text{(Eq. 2.15)}$$

However, Eq. 2.15 was derived for straight tubes at constant temperature. As the tubes might be considerably long, they need to be coiled in order to fit them into a bath or oven. As reported by Nunge *et al.* [36], there are two major implications to this: if the velocity profile is elongated, it leads to a greater dispersion of the peak measuring apparently lower diffusivity values; or if a centrifugal effect is present, it results in secondary flow perpendicular to the flow direction, increasing mixing effects resulting in narrower peaks, which generates higher apparent diffusivity values [37]. The fluid is forced to the walls of the tubing, and since it cannot accumulate, the secondary flow is formed and the fluid returns to the center of the column, which results in a circular movement as illustrated in Figure 3 [34].

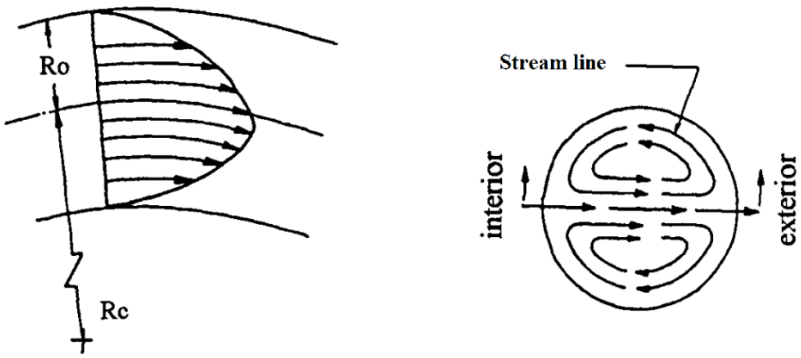


Figure 3 – Laminar flow velocity profile and secondary flow representation in a coiled tube. Taken from reference [34].

The consequences of the coiling lead to deviations to Taylor and Aris' assumptions. Such assessement can be expressed as a function of the Reynolds number, Re, Eq. 2.16, Schmidt number, Sc, Eq. 2.17, and the geometric factor, ζ, which characterizes the curvature of the coil. ζ is calculated as the ratio of the tube coil radius, R_c, and the inner column radius, R_0. Under certain conditions Re and ζ are not independent, and the Dean number, De, Eq. 2.18, describes this relation [34,37]:

$$Re = \frac{\bar{u}\,\rho_1\,R}{\mu_1} \qquad \text{(Eq. 2.16)}$$

$$Sc = \frac{\mu_1}{\rho_1\,D_{12}} \qquad \text{(Eq. 2.17)}$$

$$\mathrm{De} = \mathrm{Re}\zeta^{-0.5} \qquad \text{(Eq. 2.18)}$$

$$\zeta = \frac{R_c}{R_0} \qquad \text{(Eq. 2.19)}$$

The behaviour of the peaks is also described by these dimensionless numbers. The peak broadening effect is proportional to $\mathrm{Re}^2\mathrm{Sc}\zeta^2$, which dominates at lower Re values if $\zeta<10$. On the other hand the narrowing effect is proportional to $\left(\mathrm{De}^2\mathrm{Sc}\right)^2$, which is dominant at higher Re values [36,37]. Thus the secondary flow can be neglected if the following condition is met:

$$\mathrm{DeSc}^{0.5} < 10 \qquad \text{(Eq. 2.20)}$$

This restriction was proposed by Moulijin *et al*. [38], Alizadeh *et al*. [39] and Springston and Novotny [40]. To guarantee an error lower than 1 %, Funazukuri *et al.* [24,41] recommend $\mathrm{DeSc}^{0.5}$ lower than 8 instead of 10 for this condition. To neglect temperature and pressure perturbations that may ocurr outside of the oven (near the detector), it was established the following restriction by van der Laan [42]:

$$\frac{\bar{u}L}{D} > 1000 \qquad \text{(Eq. 2.21)}$$

In order to the concentration profile approximate a Gaussian form, the following condition is necessary as established by Levenspiel and Smith [43]:

$$\frac{D}{\bar{u}L} < 0.01 \qquad \text{(Eq. 2.22)}$$

If all previous conditions are met, the value for D_{12} can be calculated by [31]:

$$D_{12} = \frac{\bar{u}}{4}\left[H \pm \left(H^2 - \frac{R^2}{3}\right)^{0.5}\right] \qquad \text{(Eq. 2.23)}$$

To calculate the experimental theoretical plate height, H, from chromatographic peaks, there are several different methods. A simple and precise method is to measure the half width of the peak at 60.7 % of its height, $W_{0.607}$, and calculate with the following expression [23]:

$$H = \frac{LW_{0.607}^2}{t_r^{\,2}} = \frac{\bar{u}^2 W_{0.607}^2}{L} \qquad \text{(Eq. 2.24)}$$

with t_r as the retention time.

Since Eq. 2.23 is a quadratic equation, it may have two real solutions. Giddings and Seager [44] have shown that the best way to determine the solution is to calculate the velocity that minimizes H in Eq. 2.15, giving rise to Eq. 2.25. Then calculating D_{12} by Eq. 2.23, considering the optimum velocity u_{opt} from Eq. 2.25 if the velocity is below or equal to the optimum, the positive root is the most appropriate solution; otherwise, if the velocity is higher the negative root should be taken. In liquids and dense fluids, the negative root is generally chosen since the optimum velocity is generally very low and overcome by the velocity in the tubing [23].

$$u_{\text{opt}} = \sqrt{48}\,\frac{D_{12}}{R} \qquad \text{(Eq. 2.25)}$$

This method to calculate the diffusion coefficient is called the graphical method [15]. However, there are other methods to process the peaks and obtain valid values of diffusion. Two other methods that are frequently used are the method of moments and the fitting method. The method of moments consists in calculating D_{12} by using the zeroth, first and second moment and in the simplification of Eq. 2.15, combining it with Eq.2. 26 to reach Eq. 2.27, which is valid if $D_{12}\bar{t}/R^2 > 10$, and having a maximum error of ± 1 % associated:

$$\sigma^2 = \frac{2D\bar{t}}{\bar{u}^2} \qquad \text{(Eq. 2.26)}$$

$$D_{12} = \frac{R^2\bar{t}}{24\sigma^2} \qquad \text{(Eq. 2.27)}$$

with $\bar{t}$ as the average retention time and σ^2 the variance. These can be obtained by the zeroth (S), first and second moments [45]:

$$S = \int_0^{\infty} \mathrm{C(t)}\ \mathrm{dt} \qquad \text{(Eq. 2.28)}$$

$$\bar{\mathrm{t}} = \frac{1}{\mathrm{S}} \int_0^{\infty} \mathrm{C(t)}\ \mathrm{dt} \qquad \text{(Eq. 2.29)}$$

$$\sigma^2 = \frac{1}{\mathrm{S}} \int_0^{\infty} (\mathrm{t} - \bar{\mathrm{t}})^2 \mathrm{C(t)}\ \mathrm{dt} \qquad \text{(Eq. 2.30)}$$

Lastly, in the fitting method the average retention time, $\bar{t}$, and the variance, σ^2 are found by nonlinear fitting of the peak, minimizing the root mean square error, ε [46]:

$$\varepsilon = \sqrt{\frac{\int_{t1}^{t2} \left(C^{exp}(t) - (\bar{C}(L,t)\right)^2 dt}{\int_{t1}^{t2} \left(C^{exp}(t)\right)^2 dt}} \qquad \text{(Eq. 2.31)}$$

with $C^{\exp}$ as the experimentally measured concentration and $\bar{C}(L,t)$ the calculated concentration by Eq. 2.14 with $z = L$. Furthermore, t_1 and t_2 are the time values at which the peak is at 10 % of peak height, with $t_1 < t_2$ [33]. This fitting can be considered good if the value of ε is below 1 %, acceptable if between 1 % and 3 %, and rejected if higher than 3 % [46]. Another important parameter used to guarantee the validity of the results is the asymmetry factor, whose value should be below 1.1-1.3, otherwise the peak should be rejected [23].

Between the method of moments and the fitting method, the second one is considered to be more precise [24]. This is due to the fact that the first overestimates errors related to the frontal and tailing portions of the response curve [33].

Another important aspect of the CPB method is that the linearity of the UV-vis detector is important for the accuracy of the measurements performed [47]. The results with best linearity should be selected for peak detection, they were found to correspond to the diffusion coefficients results of larger value [47]. A distinct procedure in this dissertation was followed (see Chapter 4.2 and Appendix B), and consisted of testing several wavelengths and selecting the diffusion coefficient of least error, similar to other authors [24,33].

The CPB method is precise and relatively fast to determine diffusivities at infinite dillution. However, it suffers from constraints, such as: i) it's not possible to measure too close to the critical point in the case of supercritical fluids as the mixture between the fluid and the solute may not attain supercritical state in all of the column resulting in abnormal peaks [24]; ii) polar solutes or compounds of high molecular weight will cause an undesired prolongation of the peak called tailing, which is caused by the adsorption of these compounds leading to a significant error in the results [24]; iii) it must be possible to inject the solute and solubilize it in the solvent inside the column, so it should not lead to a dramatically viscous mixture with the solvent. Nevertheless, it is possible to overpass the viscosity limitation upon the utilization of an additional solvent other than the one already in the column, keeping in mind that the peak that will appear in such case will be the combination of both solvents, in which

case the CIR method or the Modified Taylor-Aris method must be employed. The main difference between these two techniques and the first one is that the column has an internal polymeric coating, as is the case for CIR, or an initial portion of it, like a combination of CIR, and then CPB, as is the case of the Modified Taylor-Aris method, which allows for the chromatrographic separation of the organic solvent from the solute [24,41] and the second being suitable for volatile compounds [48]. The biggest disadvantage of these chromatographic methods is that only diffusion coefficients at infinite dillution are measured.

3. Molecular dynamics

Classical molecular dynamics (MD) simulations are a class of computational techniques that allow for the prediction of several thermo-physical properties, under equilibrium and non-equilibrium conditions, for a wide variety of systems [49–52]. The data calculated via microscopic molecular dynamics simulations are important for the prediction of properties and the development of reliable macroscopic models.

In this dissertation the molecular dynamics software package Groningen Machine for Chemical Simulations (GROMACS) was used [20].

3.1. Classical molecular dynamics simulations

Classical MD simulations are based on solving Newton's classical equations of motion for systems of N interacting particles [19,53]:

$$m_i \frac{\partial^2 \vec{r}_i}{\partial t^2} = \vec{F}_i,\ i = 1 \dots N. \qquad \text{(Eq.3.1)}$$

where $\vec{r}_\mathrm{i}$ is the vector of position of particle i, m_i the mass of particle i, and $\vec{F}_\mathrm{i}$ the force on particle i. The forces acting on the particles are the negative derivatives of a potential function $V\ (\vec{r}_1, \vec{r}_2, \dots, \vec{r}_\mathrm{N})$:

$$\vec{F}_i = -\frac{\partial V}{\partial \vec{r}_i} \qquad \text{(Eq.3.2)}$$

The equations are solved simultaneously in small time steps. Assigned the simulation conditions and guaranteed the temperature and pressure stay at the desired values (NPT ensemble with the utilization of a thermostat and of a barostat), the simulation runs for a selected time and the system coordinates are output at a regular and selected frequency. The coordinates as a function of time are the trajectory of the system. The system will eventually reach an equilibrium state after some time. By averaging over a trajectory, many macroscopic properties of the system can be calculated. The forces involved are calculated by considering the potential energy of the interactions between the particles, both intramolecular and intermolecular, according to appropriate force fields, which are functions of interatomic potentials or energy functions with force field specific parameter sets.

3.1.1. Force fields

In molecular dynamics, the forces between particles are calculated based on potential functions, with their own set of parameters, which are called force fields. The potential energy is calculated taking into consideration two components, the non-bonded interactions and the bonded interactions.

Non-bonded interactions are pair additive which means that the potential is the result of the sum of each pair interaction potential. Often, no polarization, charge transfer effects [19] or three-body (and higher order) interactions [53] are taken into account. These interactions are described by [19]:

$$V_{non-bonded} = \sum_{i<j}^{N} V_{ij}(\vec{r}_{ij}) \qquad \text{(Eq. 3.3)}$$

The non-bonded interactions are characterized by three terms, a repulsion term, a dispersion term and a Coulomb term for electrostatic charges. The repulsion and dispersion terms are combined, for example, in the Lennard-Jones potential (6-12 potential) or the Buckingham potential (exp-6 potential) [19].

The Lennard-Jones (LJ) potential (see Figure 4), first proposed by Sir John Lennard-Jones [54], is the one used by the force fields adopted in this dissertation:

$$V_{ij}^{LJ} = 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \qquad \text{(Eq. 3.4)}$$

in which $-\varepsilon_{\mathrm{ij}}$ is the depth of the potential well for particles i and j, that occurs for $r_{\mathrm{ij}} = 2^{\frac{1}{6}}\,\sigma_{\mathrm{ij}}$, σ_{ij} the diameter of collision for low energy collisions for particles i and j, r_{ij} the distance between particles i and j [34].

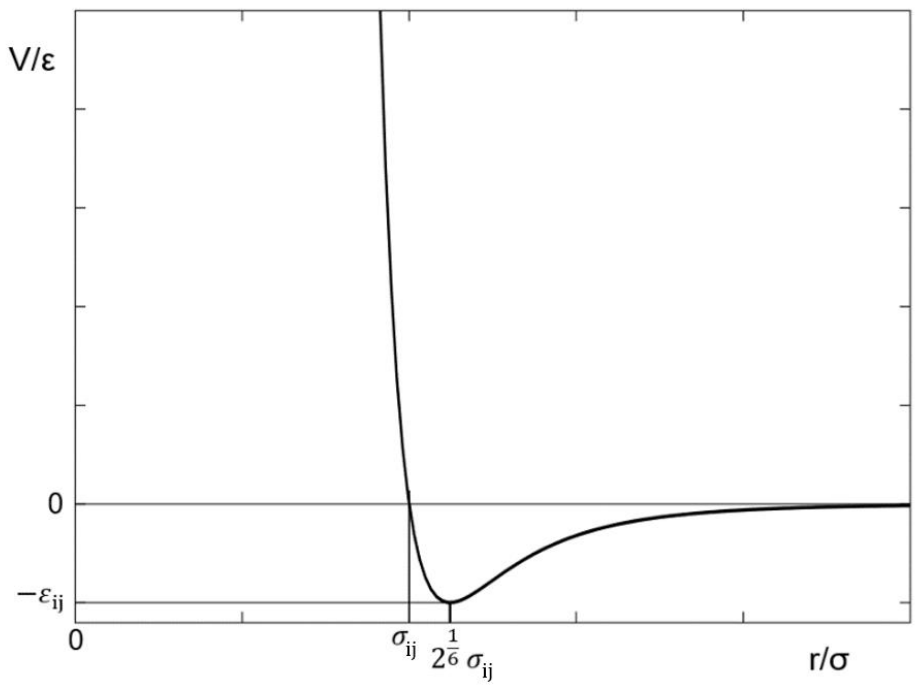


Figure 4 – Lennard-Jones interaction potential. Adapted from ref. [55].

These parameters are introduced in GROMACS in the form of *atom types* and can be obtained from combining the parameter in two ways [19]:

- Through geometric averages:

$$\sigma_{ij} = \left(\sigma_{ii}\sigma_{jj}\right)^{1/2} \quad \text{(Eq.3.5)}$$

$$\varepsilon_{\mathrm{ij}} = \left(\varepsilon_{\mathrm{ii}}\varepsilon_{\mathrm{jj}}\right)^{1/2} \quad \text{(Eq.3.6)}$$

- Through Lorentz-Berthelot rules. An arithmetic average for σ_{ij}, while a geometric average for $\varepsilon_{\mathrm{ij}}$:

$$\sigma_{\mathrm{ij}} = \frac{1}{2}\left(\sigma_{\mathrm{ii}} + \sigma_{\mathrm{jj}}\right) \quad \text{(Eq.3.5)}$$

$$\varepsilon_{\mathrm{ij}} = \left(\varepsilon_{\mathrm{ii}}\varepsilon_{\mathrm{jj}}\right)^{1/2} \quad \text{(Eq.3.6)}$$

This potential can be cut-off at a distance where the interactions are considered negligible in order to reduce the amount of necessary calculations. In the Verlet cut-off scheme, the potential is shifted by a constant so that it is zero at the cut-off distance [19] without breaking continuity.

The Coulomb potential for the electrostatic charges between two particles with charges q_{i} and q_{j} is given by [19,53]:

$$V_{\mathrm{ij}}^{\mathrm{Coulomb}} = \frac{1}{4\pi\varepsilon_0}\frac{q_{\mathrm{i}}q_{\mathrm{j}}}{\varepsilon_{\mathrm{r}}r_{\mathrm{ij}}} \quad \text{(Eq. 3.7)}$$

with ε_0 as the absolute permittivity of free space, ε_{r} the relative permittivity (or dielectric constant).

A plain Coulomb interaction should be used without cut-off as the decay of the potential function towards 0 is much slower than the previous LJ terms and would lead to a sudden and large change in the force at the cut-off distance. In a similar way to LJ potential, the function may also be shifted to zero at a certain cut-off distance, *e.g.,* this is done when Ewald summation or particle-mesh Ewald (PME) is used to calculate the long-range interactions [19,56].

Bonded interactions include types of interactions such as bond stretching (2-body) interactions, bond angle (3-body) interactions and dihedral angle (4-body) interactions.

The bond stretching for two covalently bonded atoms (see Figure 5) may be represented in different ways depending on the force field, such as a harmonic potential [19,53]:

$$V_{\mathrm{ij}}^{\mathrm{stretching}} = \frac{1}{2}k_{\mathrm{ij}}^{\mathrm{b}}\left(r_{\mathrm{ij}} - r_{\mathrm{ij}}^{0}\right)^2 \quad \text{(Eq. 3.8)}$$

where k_{ij} is the force constant of stretching vibrations and r_{ij}^{0} is the equilibrium distance between i and j.

The bond angle for three atoms i, j and k (see Figure 6) may be represented by a harmonic potential acting on the angle θ_{ijk} [19,53]:

$$V_{\mathrm{ij}}^{\mathrm{angle}} = \frac{1}{2} k_{\mathrm{ijk}}^{\theta} \left(\theta_{\mathrm{ijk}} - \theta_{\mathrm{ijk}}^{0}\right)^{2} \qquad \text{(Eq. 3.9)}$$

where $k_{\mathrm{ijk}}^{\theta}$ is the force constant for angle vibrations and $\theta_{\mathrm{ijk}}^{0}$ the equilibrium bond angle.

The dihedral angles or torsion angles can be divided into two types, proper dihedrals and improper dihedrals (see Figure 7). Proper dihedral angles are the angle ϕ_{ijkl} between two planes ijk and jkl, with $\phi_{\mathrm{ijkl}} = 0^{\mathrm{o}}$ corresponding to the *cis* configuration (i and l on the same side), or with $\phi_{\mathrm{ijkl}} = 180^{\mathrm{o}}$ corresponding to the *trans* configuration (i and l on opposite sides). Improper dihedral angles are special dihedrals meant to keep planar some parts of molecules (*e.g.* amide bond in amino acids) or maintain the chirality of three atoms centered around one atom (*e.g.* tetrahedral angle).

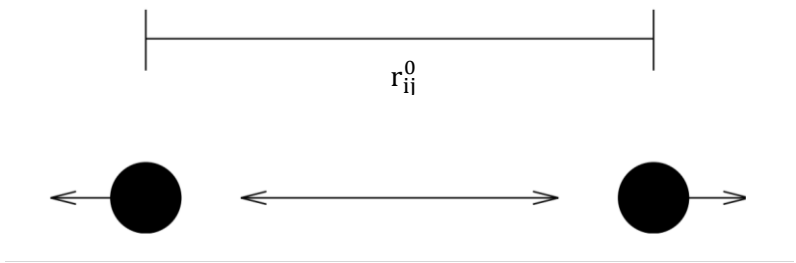


Figure 5 – Bond stretching representation. Adapted from [19].

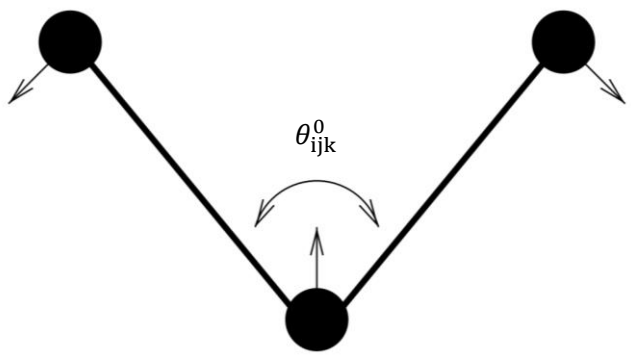


Figure 6 – Bond angle representation. Taken from [19].

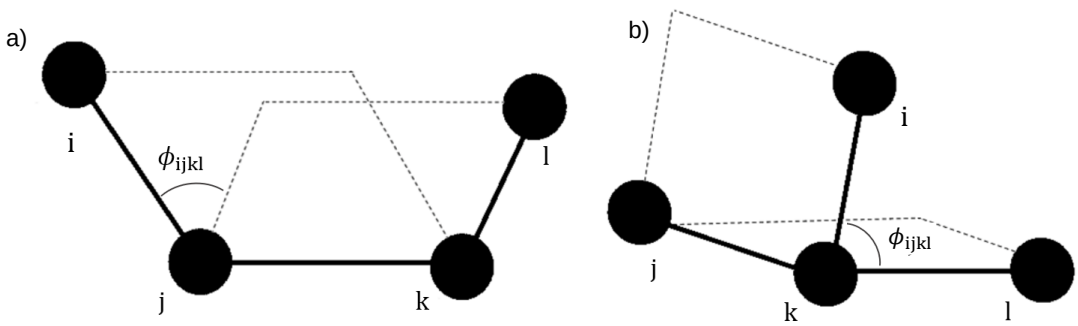


Figure 7 – Torsion/dihedral angle representation: a) proper dihedral; b) improper dihedral. Adapted from [19].

There are various functions for dihedral angles, a potential often used is the Ryckaert-Bellemans (RB) function, a function based on expansion in powers of $\cos(\phi)$ [19]:

$$V_{ijkl}^{RB} = \sum_{\mathrm{n}=0}^{5} C_{\mathrm{n}} \left(cos\ (\psi_{\mathrm{ijkl}})\right)^{\mathrm{n}} \quad \text{(Eq.3. 10)}$$

$$V_{ijkl}^{RB} = \sum_{\mathrm{n}=0}^{5} (-1)^{\mathrm{n}} C_{\mathrm{n}} \left(cos\ (\phi_{\mathrm{ijkl}})\right)^{\mathrm{n}} \quad \text{(Eq.3. 11)}$$

where $\phi_{\mathrm{ijkl}} = \psi_{\mathrm{ijkl}} + 180$, as trans orientation ("polymer convention") and $\phi_{\mathrm{ijkl}} = 0$ as cis orientation ("biochemical convention"). The conversion between conventions for this function can be done by multiplying the constant by $(-1)^{\mathrm{n}}$.

Another possible function is the Fourier series of $\cos(\phi)$:

$$V_{ijkl}^{Fourier} = \frac{1}{2}[F_1(1 + cos(\phi)) + F_2(1 - cos(2\phi)) + \\ + F_3(1 + cos(3\phi)) + F_4(1 - cos(4\phi))] \quad \text{(Eq.3. 12)}$$

The Fourier series parameters can be converted to RB parameters in the following way [19]:

$$C_0 = F_2 + \frac{1}{2}(F_1 + F_3) \quad \text{(Eq.3. 13)}$$

$$C_1 = \frac{1}{2}(-F_1 + 3F_3) \quad \text{(Eq.3. 14)}$$

$$C_2 = -F_2 + 4F_4 \quad \text{(Eq.3. 15)}$$

$$C_3 = -2F_3 \quad \text{(Eq.3. 16)}$$

$$C_4 = -4F_4 \quad \text{(Eq.3. 17)}$$

$$C_5 = 0 \quad \text{(Eq.3. 18)}$$

This happens in GROMACS for the Optimized Potentials for Liquid Simulations (OPLS) force field as it uses RB's code to compute Fourier dihedrals.

These bond distances and angles can be maintained by imposing constraint on the equations of motion with algorithms, like the Linear Constraint Solver (LINCS) [57] algorithm. LINCS resets bonds to their correct lengths after an unconstrained update. If the algorithm cannot fulfill the constraints and the molecule rotates more than a predefined angle it will not crash, it will instead generate a warning and only stop the simulation after a predefined number of warnings [19].

Commonly used force fields are the Assisted Model Building with Energy Refinement (AMBER) [58], or the Optimized Potentials for Liquid Simulations – All Atom (OPLS-AA) [59], but there are many others for varied purposes [19,58,60]. The GROMACS documentation [19] recommends GROMOS-96 [61,62] for united-atom setups and OPLS-AA for all-atom setups.

3.1.2. Periodic boundaries

In MD simulations, the molecular systems are finite and, thus, there are edge effects that may be quite undesirable. To minimize these edge effects *periodic boundary conditions* (PBC) may be applied. The simulated system instead of having boundaries is surrounded by translated copies of itself, Figure 8, resulting in a system without boundaries. Hence, the edge effects of an isolated system are replaced with the edge effects of periodic conditions. These edge effects are reduced for periodic systems like crystalline systems. However, in non-periodic systems like liquids or solutions the error can be caused by periodicity itself. These errors are expected to be less severe than errors from an unnatural boundary with vacuum, and should be evaluated by comparing various system sizes [19].

There are several shapes for the unit cells of periodic systems. For simplicity there is the cube, but shapes like the rhombic dodecahedron or the truncated octahedron are closer to a sphere than a cube, which might be useful for simulating liquids [63] or 0D systems (*e.g.* a protein in water). Periodic boundaries in GROMACS are combined with minimum image convection, which means that only the nearest image of each particle is considered for short-range non-bonded interactions. For long-range electrostatics this may not be enough so lattice sum methods such as Ewald sum or PME may be employed [19].

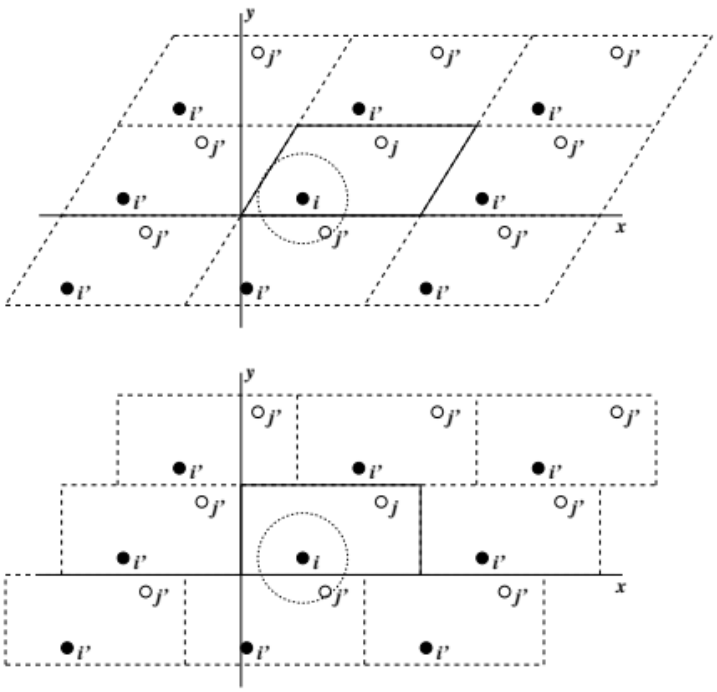


Figure 8 – Periodic boundary conditions in two dimensions. Taken from [19].

3.1.3. Thermodynamic ensemble

Another important concept for MD is the thermodynamic ensemble. To determine macroscopic properties, ensemble averages are always done over an adequate representative statistical ensemble of molecular systems [19]. In a thermodynamic ensemble, a number of variables are fixed while the other thermodynamic quantities are calculated by ensemble averaging [63].

Commonly used statistical ensembles are: the microcanonical or constant-NVE (number of particles, volume and energy) ensemble, the canonical or constant-NVT (number of particles, volume and temperature) ensemble, the isothermal-isobaric constant-NPT (number of particles, pressure and temperature) ensemble [19], [63].

As illustrated in Figure 9, the NVE ensemble consists of an isolated system of fixed volume, the NVT ensemble consists of a closed system of constant volume that exchanges energy with the exterior, like it is surrounded by a thermostatic bath to control the temperature, and the NPT ensemble can be regarded as a closed system of variable volume and controlled temperature.

In molecular dynamics the ensemble average is replaced by a trajectory average. Newton's equations of motion generate a succession of states in accordance to the NVE ensemble as Newton's equations of motion obey the laws of conservation of energy. Often, it is desirable to have constant temperature (NVT) and pressure (NPT), much like laboratorial experiments, in order to do so a modification of the equations of motion may be derived, which could involve a stochastic (random) or deterministic element or it can have no relation with normal dynamics [63,64].

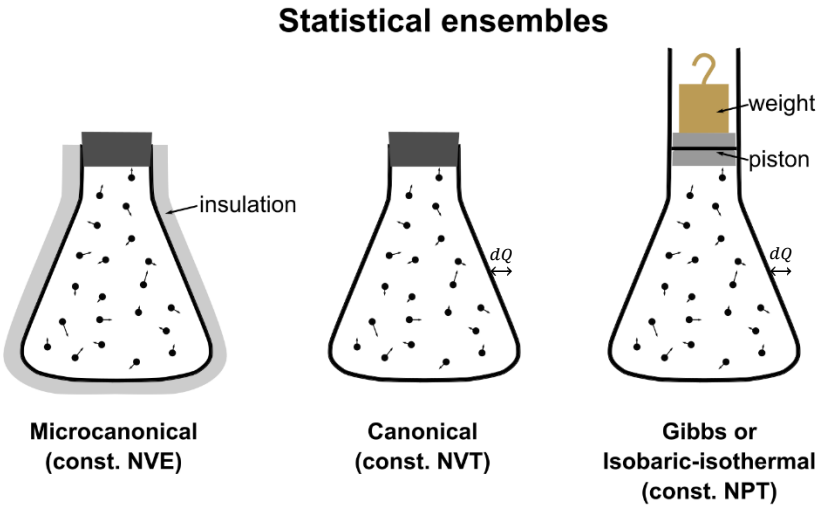


Figure 9 – Representation of three statistical ensembles. Adapted from [65].

3.1.4. Limitations of classical molecular dynamics

Classical MD simulations have limitations as approximations are necessary to feasibly simulate the movement of atoms. When attempting to run simulations and processing their data, one must be aware of such limitations and access the accuracy of the simulation. Summarized, they are [19]:

- Usage of classical mechanics to describe the motion of atoms. Newton's equations of motion are based on classical mechanics which is appropriate for most atoms at regular temperatures. However, under certain conditions, some particles may behave differently, *e.g.* hydrogen atoms, as they are quite light, or protons' motion, as they may have quantum mechanical characteristics at times that cannot be described properly by classical dynamics [19]. Liquid helium may also be wrongly described by classical mechanics. While these can be circumstantial limitations, the approximation that is implied regarding high frequency vibrations of covalent bonds is of greater importance. The definition of a harmonic oscillator is considerably different from a quantum oscillator when the resonance frequency ν approximates or exceeds $k_B T$. At room temperature the wavenumber $\sigma = 1/\lambda = \nu/c$ at which $h\nu = k_B T$ is approximately 200 cm^{-1}, as a result all frequencies higher than 100 cm^{-1} may be represented inaccurately in classical simulations. This translates into almost every bond and bond-angle vibration. Two solutions for this are, using an harmonic oscillator for bonds while making corrections to the total internal energy and specific heat [19], or to treat bonds and bond angles as constraints in the equations of motion, this being because a quantum oscillator in its ground state resembles a constrained bond closer than a classical oscillator. GROMACS can use constraints for bonds and can convert bond angles into bond constraints as the algorithm is also more versatile and allows for larger time steps.
- Electronic motions are not taken into consideration. Molecular dynamics uses a force field that is a function of positions of atoms only, while electrons are taken into account in a very approximate way as point charges located at, most often, the atomic positions. This prohibits the representation of electron transfer processes, electronically exited states and chemical reactions.
- The force field is pair-additive, all non-bonded forces result from the sum of non-bonded pair interactions. Non pair-additive interactions, such as atomic polarizability, are not represented. The contributions from these kinds of phenomena are averaged in the force field parameters. This however implies that the interactions may not be valid for isolated pairs or for situations that differ too much from the systems or models used for the parameterization of the force field.

- Long-range interactions can be cut off. In GROMACS not all the interactions are calculated, and they are always cut-off after a defined radius. This is usually fine for Lennard-Jones and sometimes Coulomb interactions considering that the cut-off radius is large enough, but in the presence of charged particles this can lead to large errors in the energies of the system. The usage of a long-range electrostatic algorithm such as PME is advised.
- Boundary conditions in MD are unnatural, *i.e.* the systems in molecular dynamics are not infinite, they are small and have unwanted boundaries. To simulate a bulk system this must be avoided. To solve this issue, periodic boundary conditions are used to avoid real phase boundaries. For large systems the error may be small but, for small systems, with high internal spatial correlation, periodic boundaries may increase internal correlation, leading to unnatural systems, which results in large errors. Thus, system size is still considerably relevant, making it important to test the influence of the system size on the simulation and on the desired properties.

3.2. Diffusion coefficients in molecular dynamics

Mass, energy or momentum transfers through a system can be described by phenomenological relations with the form of $flux = -coefficient \times gradient$. Newton's law of viscosity, Fourier's law of heat conduction, Ohm's law of electrical conduction and Fick's law of diffusion are examples of such relations. They are usually seen applied to situations of nonequilibrium but they also apply to the microscopic fluctuations that occur in a system at equilibrium [66]. So, the transport coefficients can be extracted from equilibrium molecular dynamics. The diffusion coefficient is often calculated by two types of relations: Green-Kubo or Einstein relations. They should give the same result and there is little advantage of one over the other [66]. Nevertheless, the Green-Kubo relation usually needs trajectories from long simulations to originate reliable values [67]. In this dissertation the Einstein relation was used.

Considering the one-dimensional diffusion as described by Fick's law [66]:

$$\mathrm{J} = -D_{\mathrm{i}} \frac{\partial N}{\partial z} \qquad \text{(Eq. 3.19)}$$

where N is the number of atoms per unit of volume located at position z at time t, D_{i} is the diffusion coefficient and J is the flux. From the material balance on a differential element of fluid the equation of continuity is obtained [66]:

$$\frac{\partial N}{\partial t} + \frac{\partial J}{\partial z} = 0 \qquad \text{(Eq. 3.20)}$$

Combining Eqs. 3.19 and 3.20, the following relation is obtained:

$$\frac{\partial N}{\partial t} = D_{\mathrm{i}} \frac{\partial^2 N}{\partial z^2} \qquad \text{(Eq. 3.21)}$$

By establishing initial conditions, the equation can be solved for temporal and spatial evolution of $N(z,t)$. For instance, in the case of N_0 atoms concentrated at the origin $z = 0$ and at time $t = 0$, the solution is [66]:

$$N(z,t) = \frac{N_0}{2\sqrt{\pi D_{\mathrm{i}} t}} e^{\left(\frac{-z^2}{4D_{\mathrm{i}} t}\right)} \qquad \text{(Eq. 3.22)}$$

So, for any time $t > 0$, the atoms are spatially distributed in a Gaussian shape around the origin and, as time evolves, the atoms diffuse and the Gaussian distribussion collapses. At any time $t > 0$, the second moment of the distribution is the mean-square displacement of atoms [66].

$$\langle [z(t) - z(0)]^2 \rangle = \frac{1}{N_0} \int z^2 N(z,t)\, dz \qquad \text{(Eq. 3.23)}$$

By combining Eq. 3.22 with Eq. 3.23 and integrating, the mean-square displacement is related to the diffusion coefficient [66]:

$$\langle [z(t) - z(0)]^2 \rangle = 2D_{\mathrm{i}} t \qquad \text{(Eq. 3.24)}$$

This result applies when the elapsed time t is large when compared to the average time between collisions of atoms. Considering $\boldsymbol{r}_{\mathrm{i}}$ as the position of the particles, the n-dimensional analog of Eq. 3.24 is [67]:

$$D_{\mathrm{i}} = \lim_{t \to \infty} \frac{\langle |\boldsymbol{r}_{\mathrm{i}}(t) - \boldsymbol{r}_{\mathrm{i}}(0)|^2 \rangle}{2nt} \qquad \text{(Eq. 3.25)}$$

The angled brackets of Eq. 3.24 and Eq. 3.25 indicate that the mean-square displacement is averaged from over all time origins and particles for which diffusion is being calculated from the simulation [66]. Given that at the conditions studied, the diffusion coefficient is considered a constant, the equation implies that the mean-square displacement grows linearly at large time differences.

This relation can be applied for both binary diffusion coefficients and self-diffusion coefficients [63,66].

4. Materials and methods

This chapter is divided into three main sections:

- The first section describes the correlations used to estimate the density and viscosity of the solvents. These values were necessary in several calculations, and were also used for comparison with the values obtained through molecular dynamics (MD).
- The experimental section (materials, equipment and procedures) focuses the measurement of the tracer diffusion coefficients of quercetin in ethanol and in ethyl acetate.
- The MD simulation section presents the procedures for the calculation of tracer diffusion coefficients.

The experimental data and the MD trajectories were determined for a common set of conditions. This is required for validating the computational recipes that, eventually, will be used to extract information at temperature and pressure conditions outside those considered in the experiments.

4.1. Density and viscosity of ethanol and ethyl acetate

The pure liquid ethanol density was calculated using the Tait [68,69] equation:

$$\frac{\rho-\rho_0}{\rho} = 0.2000\times\log_{10}\left(\frac{B+P}{B+P_0}\right) \quad \text{(Eq. 4.1)}$$

$$B = 520.23\text{-}1240\times\frac{T}{T_C}+827\times\left(\frac{T}{T_C}\right)^2 - F \quad \text{(Eq. 4.2)}$$

where ρ and ρ_0 are densities at the corresponding pressures P and P_0, and F calculated by Eq. 4.3, where C_n is the number of carbons of the molecule). The density at atmospheric pressure (ρ_0) is calculated according to the Eykman method as suggested by Cano-Gómez *et al.* [70].

$$F = 0.015\times C_n\times(1+11.5\times C_n) \quad \text{(Eq. 4.3)}$$

$$\rho_0 = \frac{n_D^2-1}{n_D+0.4}\frac{1}{K} \quad \text{(Eq. 4.4)}$$

where n_D and K are the refractive index and a characteristic constant, respectively, given by:

$$K = 0.72719 - 0.39294\exp(C_n^{-0.89255} \times 0.47272) \quad \text{(Eq. 4.5)}$$

$$n_D = a_0+a_1 C_n^{a_2}+a_3 C_n+\frac{a_4}{C_n^{a_5}}+\left(a_6+a_7 C_n^{0.5}+a_8 C_n^{0.75}\right) \times T(°C) \quad \text{(Eq. 4.6)}$$

In the previous equation the values of the constants are a_0=1.87961; $a_1 = 0.55029$; $a_2 = -0.11935$; $a_3 = -0.00161$; $a_4 = 0.01344$; $a_5 = 13.54426$; $a_6 = -0.00043235$; $a_7 = 0.00000954$; $a_8 = 0.0000022$.

Cano-Gómez *et al.* [70] suggested using the Mamedov equation (Eq. 4.7) for determining ethanol viscosity at high pressures:

$$\frac{\mu}{\mu_0} = \left(\frac{\rho}{\rho_0}\right)^{\mathrm{A}} \qquad \text{(Eq. 4.7)}$$

$$A = 10.4 + 0.0006\mathrm{C_n^{3.5}} - \frac{5}{\mathrm{C_n}} \qquad \text{(Eq. 4.8)}$$

$$\log_{10} \mu_0 = \mathrm{A} + \frac{\mathrm{B}}{\mathrm{T}} + \mathrm{C} \times \mathrm{T} + \mathrm{D} \times \mathrm{T}^2 \qquad \text{(Eq. 4.9)}$$

where $A = 0.72719$, $B = -0.39294$, $C =$ -0.89255 and $D = 0.47272$.

These correlations have an average absolute percentage deviation of 0.11 % for density, for pressures up to 279 MPa and temperatures between 173.15 to 373.15 K. And for viscosity within 2.14 % at atmospheric pressure and 3.38 % for pressures up to 423 MPa at temperatures between 293 to 423 K.

The pure liquid ethyl acetate density was estimated using the Tait equation (Eq. 4.1) [68,71] as well. However, the parameters ρ_0 and B are calculated by:

$$B = 494\text{-}1110 \times \frac{T}{T_\mathrm{C}} + 672 \times \left(\frac{T}{T_\mathrm{C}}\right)^2 - (C_\mathrm{n} - 6) \qquad \text{(Eq. 4.10)}$$

$$\rho_0 = a \times b^{-\left(\frac{1-T}{T_c}\right)^{\mathrm{n}}} \qquad \text{(Eq. 4.11)}$$

The equation to estimate the density at atmospheric pressure is a modified form of the Racket equation [72]. The temperature is in K, ρ_0 in g cm^{-3} and $a = 0.30654$ and $b = 0.25856$.

To estimate the viscosity of ethyl acetate the correlation developed by Cano-Gómez *et al*. [70], Eq. 4.7, was used. However, in this case the parameters of Eq. 4.9 are $A = -3.6861$, $B = 552.28$, $C = 0.0080018$ and $D = -0.000010439$.

These correlations have an average absolute percentage deviation of 0.5 % for density, for pressures up to 152 MPa and temperatures between 253.15 to 313.15 K.

4.2. Measurement of diffusion coefficients

This section pertains to the experimental part of the dissertation and in it are presented the materials, equipment and procedures used for the measurement of tracer diffusion coefficients.

4.2.1 Chemicals

The chemicals used in the experiments were quercetin, CAS number 117-39-5, purity ≥ 95 wt.%, purchased from Sigma-Aldrich and ethyl acetate, CAS number 141-78-6, purity ≥ 99.5 wt.%, purchased from VWR Chemicals. All chemicals were used directly without further purification.

4.2.2 Systems and experimental conditions

In this work the tracer diffusion coefficients of quercetin were measured in liquid ethyl acetate the studied experimental conditions and respective solvent properties are presented in Table 1:

Table 1 – Systems, experimental conditions for diffusion coefficient measurement and solvent properties.

Experiment	System		P (bar)	T (°C)	ρ_{calc} (kg m^{-3})	μ_{calc} (cP)
1	Quercetin	Ethyl Acetate	1	30	887.7	0.3994
2				40	875.6	0.3590
3				50	863.4	0.3247
4				60	850.8	0.2952
5	Quercetin	Ethyl Acetate	50	30	892.0	0.4223
6				40	881.0	0.3802
7				50	869.0	0.3444
8				60	856.0	0.3138
9	Quercetin	Ethyl Acetate	100	30	897.0	0.4439
10				40	885.0	0.4003
11				50	873.8	0.3632
12				60	861.0	0.3315
13	Quercetin	Ethyl Acetate	150	30	901.0	0.4644
14				40	890.0	0.4193
15				50	878.0	0.3810
16				60	867.0	0.3484

4.2.3 Equipment and experimental procedure

The equipment used to employ the CPB method is presented in Figure 10 [73]. It consists of two reservoirs, (1) and (5), and two syringe pumps, a Teledyne ISCO model 260D with 266.06 cm^3 capacity for CO_2 (2) and a Teledyne ISCO model 100DM 102.97 cm^3 capacity for liquids (4). The CO_2 pump (2) is coupled with a Julabo F12 thermostatic bath (3), to avoid temperature oscillations, which would cause flow rate fluctuations when working in supercritical conditions. Followed by a stainless-steel tubing (7) placed inside a LSIS-B2V/IC 22 oven (Venticell, MMM Group) (9) to pre-heat the solvents, then the tubing is connected to an open capillary tubing (8) (Polyether ether ketone (PEEK) tubing, with dimensions R = 0.261 mm, L = 10.243 m and Rc = 0.150 m) followed by an UV-vis detector (UV Detector 2500, Knauer) (10) set to a system-specific wavelength. After reaching steady-state (constant pressure, temperature and baseline, generally 1 to 2 hours after

startup) a small volume of solute is injected (0.1 μL) as a pulse using a C74H-1674 injector (6) from Valco Instruments Co. Inc.. At the outlet, a Jasco BP-2080 back pressure regulator (BPR) (12) is included in order to control pressure inside the system.

To ensure reliable values in pressurized liquids, such as ethanol and ethyl acetate, the following procedure should be carefully followed. First and foremost, the oven, UV-visible detector and the BPR must be turned on, setting the desired operating conditions of temperature and pressure in all tubing. The UV-vis detector set on the desired wavelength. Then, the syringe pump (4) connected to the liquid reservoir (5) should be turned on and refilled by opening the pump inlet valve (I) and then starting the refill. Once full, the valve should be closed and ensured that the outlet check valve (II) is locked as well. Then, the pump should also be pressurized up to the desired pressure. Once the pump and tubing have reached the desired conditions and the pump stopped, the outlet check valve (II) is opened and the desired flow rate is defined in the syringe pump. In case of changing the conditions, the whole system will need to re-stabilize for 1-2 hours, to ensure steady-state operation.

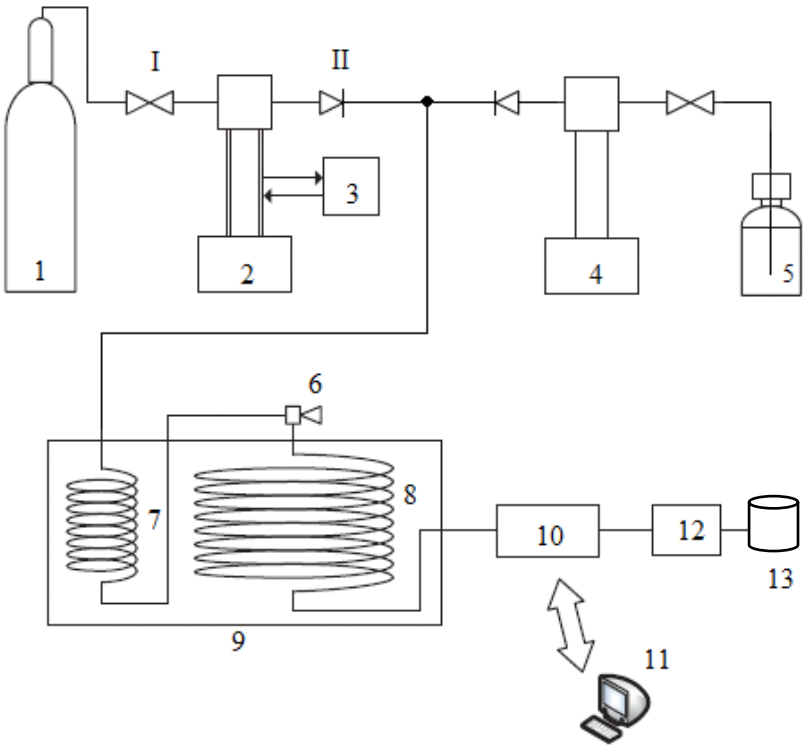


Figure 10 – Scheme of the experimental apparatus used to measure tracer diffusion coefficients in liquid or supercritical fluids: (1) CO_2 cylinder, (2) CO_2 syringe pump, (3) thermostatic bath, (4) liquid syringe pump, (5) liquid reservoir, (6) injector, (7) pre-heating column, (8) diffusion column, (9) oven, (10) UV-vis detector, (11) data acquisition software, (12) back pressure regulator – BPR, (13) liquid waste container, (I) on/off valves, and (II) check valves on both pumps. Adapted from [73].

Lastly, all that is required is to analyze and recover the absorbance data measured by the UV-visible detector and proceed to data treatment using at least one of the methods previously mentioned (graphical, moments, or fitting methods). Three to six measurements were taken to average over and calculate the uncertainty for each experimental point presented in Chapter 5.3.

Before proceeding to measurements using different conditions, the optimal most linear wavelength must be found, and this requires the following procedure. Several pulses of solute are injected in a range of 205 to 410 nm in ethanol and 250 to 400 nm in ethyl acetate to determine the wavelength that ensures the smallest error. The range was selected by UV-visible spectrophotometry to determine the region of maximum absorbance of the solute.

The UV-vis spectra of the quercetin solutions were also frequently checked in a UV-visible spectrometer to verify if the quercetin had oxidized [74,75], which is undesirable for the determination of diffusivities.

4.3. Molecular dynamics simulations

This section pertains to the molecular dynamics part of the dissertation and in it are presented the systems for which calculations were done and the procedures used.

4.3.1 Systems and conditions studied

In this work quercetin in liquid ethanol and quercetin in liquid ethyl acetate was simulated at different conditions in order to calculate tracer diffusion coefficients. Several test simulations were done to find the optimal simulation parameters. The conditions of the simulations and respective objective are presented in Table 2.

The experimental work of quercetin in ethanol used for comparison was done by another researcher using the same equipment and procedure.

Table 2 – Systems, conditions for molecular dynamics simulations and their respective objective.

Simulation	System	P (bar)	T (°C)	Objective
MD1	Ethanol	1	25	Self-diffusion of ethanol using MD calculated density
MD2	Ethanol	1	25	Self-diffusion of ethanol using calculated density
MD3	Ethyl Acetate	1	25	Self-diffusion of ethyl acetate OPLS-AA
MD4	Ethyl Acetate	1	25	Self-diffusion of ethyl acetate LOPLS-AA
MD5	Quercetin/Ethanol	1	30	Parameter test: Simulation duration
MD6	Quercetin/Ethanol	1	30	Parameter test: Cut-off radius of 1.3 nm
MD7	Quercetin/Ethanol	1	30	Parameter test: Cut-off radius of 1.4 nm
MD8	Quercetin/Ethanol	1	30	Parameter test: Cut-off radius of 1.5 nm
MD9	Quercetin/Ethanol	1	30	Parameter test: Solvent molecules 500
MD10	Quercetin/Ethanol	1	30	Parameter test: Solvent molecules 1000
MD11	Quercetin/Ethanol	1	30	Parameter test: Solvent molecules 2500
MD12	Quercetin/Ethanol	1	30	Parameter test: Solvent molecules 4000
MD13	Quercetin/Ethanol	1	30	Parameter test: Solute molecules 3
MD14	Quercetin/Ethanol	1	30	Parameter test: Solute molecules 12
MD15	Quercetin/Ethanol	1	30	Calculate tracer diffusion coefficient*
MD16	Quercetin/Ethanol	1	40	Calculate tracer diffusion coefficient*
MD17	Quercetin/Ethanol	1	50	Calculate tracer diffusion coefficient*
MD18	Quercetin/Ethanol	1	60	Calculate tracer diffusion coefficient*
MD19	Quercetin/Ethanol	50	30	Calculate tracer diffusion coefficient*
MD20	Quercetin/Ethanol	150	60	Calculate tracer diffusion coefficient (Standard atom charge x1.0)*
MD21	Quercetin/Ethanol	150	60	Calculate tracer diffusion coefficient (Scaled atom charge x1.01)*
MD22	Quercetin/Ethanol	150	60	Calculate tracer diffusion coefficient (Scaled atom charge x1.02)*
MD24	Quercetin/Ethanol	150	60	Calculate tracer diffusion coefficient (Scaled atom charge x1.10)*
MD25	Quercetin/Ethyl Acetate	1	30	Calculate tracer diffusion coefficient*

***** For these simulations the procedure and simulation parameters were already established.

4.3.2 Calculation procedure

For the MD simulations, the GROMACS software package, version 2019.3 [19,20], was used. Consulting other authors on how to obtain diffusion values from molecular dynamics simulations [49,76,77], it was decided that there would be three major steps to optimize a computational recipe to calculate accurate and consistent D_{12} values:

1. Testing the force field parameters for the solvents to verify their accuracy by comparing the values of a chosen set of appropriate predicted thermo-physical properties with the literature data. The ones chosen in this dissertation were the density, ρ, and self-diffusion coefficient, D_{11}, two properties heavily related to the binary diffusion coefficient, D_{12}, of a solute in any given solvent.
2. Testing the influence on D_{12} of the different MD simulation parameters during the equilibration and production phases of the simulation in NVT ensemble. In the dissertation, the following parameters were considered: short-range interactions cut-off radius (simply called cut-off), duration of the simulation, number of molecules of the solvent and number of molecules of the solute. This step is necessary to find the optimal parameters to ensure that i) the simulation box is large enough to represent the desired system; ii) the simulation time is long enough to achieve equilibrium and to originate accurate values; and iii) the cut-off is large enough to guarantee an accurate representation of the interactions and, consequently, to minimize errors intrinsic to these parameters, while maintaining the simulation feasible with the available computational resources in terms of processing power/time and file size.
3. Comparison of the tracer diffusion coefficients D_{12} obtained through MD simulations for the different conditions of pressure and temperature with available experimental values.

Throughout all the simulations the cell temperature was kept constant using the Nosé-Hoover [78,79] temperature coupling algorithm for canonical ensemble computations (i.e., a thermostat) while maintaining realistic dynamics [64]. They were carried out using a leap-frog algorithm [80] to integrate Newton's equations of motion, designated *md* in the GROMACS package. The LINCS algorithm was used in these simulations for all bonds length constraints. The non-bonded short-range interactions were assigned a cut-off of 1.4 nm and the neighbor list was updated every 10 time steps, as tested during step two. For the long-range electrostatic interactions, the particle-mesh Ewald summation (PME) was used [56]. Having the random configurations generated in any of the simulations, an energy minimization was performed using the steepest descent algorithm to relax the molecular systems. Cubic boxes and standard periodic boundary conditions were used.

- For the first step, and for each solvent, the NPT ensemble (constant number of molecules, system pressure P, and temperature T) was used in the determination of the density of each system. The simulation runs for 10 ns, with a 2 fs step size, using the Parrinello-Rahman [81,82] pressure coupling algorithm for true NPT ensemble. The first 200 ps were discarded, after which the density of the system reached an equilibrium and was used to obtain an average density of the system at that temperature and pressure, ρ_{MD}. Then, an NVT ensemble simulation for self-diffusion coefficients D_{11} of 12 ns, with a step size of 1 fs and a box of 2500 molecules of solvent was accomplished. Unlike the third step below, due to the much larger number of molecules contributing to the average of the diffusion value, a shorter duration is required to obtain consistent diffusion values. The conditions of these simulations were 25 °C and 1 bar.

- For the second step several simulations at the same conditions were made in NVT ensemble, each one varying a specific and single parameter of the simulation, in a similar fashion to Vaz *et al.* targeting the estimation of D_{12} of ketones in supercritical CO_2 [49]. The parameters were the following: duration of the simulation up to a total of 75 ns with a step size of 1 fs; cut-off radius for non-bonded short-range interactions between 1.3 and 1.5; frequency of written frames, for the values between every 1 frame to every 2000 frames, to reduce file size without compromising the results; and system size, *i.e.* the number of solvent molecules inside the box (500, 1000, 2500 and 4000 molecules). For each simulation, it was analyzed the influence of those parameters on the quality of the D_{12} values of quercetin in the solvent. The number of solute molecules was tested as well; however, since quercetin molecule is relatively big this implies increasing the box size considerably to ensure infinite dilution for numbers larger than 3 molecules.

- For the third step, the simulations were conducted in the NVT ensemble. The established number of molecules was 2500 molecules of solvent and 3 molecules of quercetin, to guarantee that each quercetin molecule didn't interact at close distance for long periods of time with another of the same type and jeopardize the calculation of the tracer diffusion coefficient. This results in a concentration of 0.12 mol% (equivalently 0.78 wt.% in ethanol and 0.41 wt.% in ethyl acetate). The cubic cells had their volume fixed in order to match the system density at the desired conditions. The densities used were the average from preceding NPT simulations of 10 ns duration, enough for obtaining equilibrated density values. For comparison the densities of the MD systems, with quercetin of concentration considered small enough for infinite dilution, were compared with those for pure solvents [69,71], see Chapter 4.1. Then, the NVT ensemble simulation was carried out with an integration time step of 1 fs and initial velocities generated according to the Maxwell

distribution. The simulation was divided into an equilibration phase of 15 × 10^6 time steps followed by a production phase from which a trajectory of 60 × 10^6 time steps were taken from to calculate and average diffusivities, giving the simulation a total of 75 × 10^6 time steps or 75 ns simulation time.

The force field parameters used for ethanol and ethyl acetate were the default OPLS-AA force field parameters [83–85], included in the GROMACS code. The only exception were the LOPLS-AA parameters of the ethyl acetate ester group, which were taken from Pluhackova *et al.* [86]. The force field parameters used are presented in Tables 3 and 4. The labels used to identify each atom are shown in Figure 11 over the skeletal structure of both molecules.

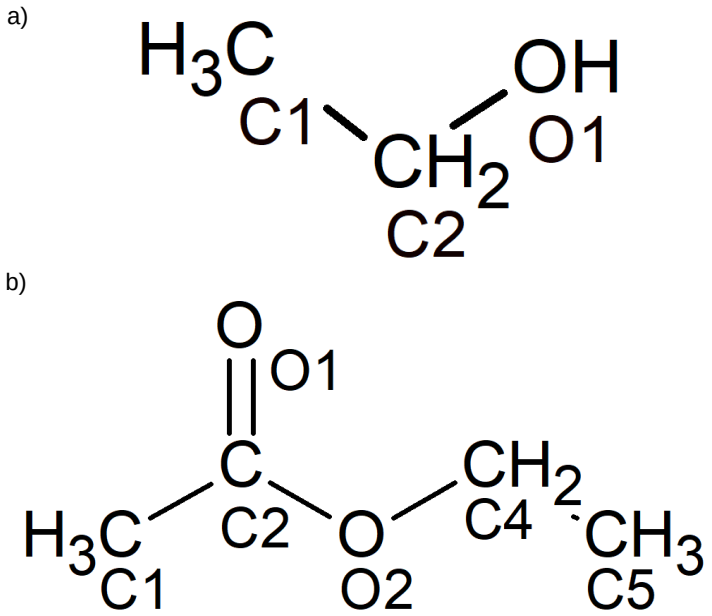


Figure 11 – a) Ethanol and b) ethyl acetate skeletal structure labelled, made using ACD/ChemSketch [5].

Table 3 – OPLS-AA atom types used for ethanol.

Atom type	σ_{ii} (10^{-1} nm)	ε_{ii} (10^{-1} kJ mol^{-1})	q_i (e*[1])	Atoms*[2]	Atom type description
opls_135	3.50000	2.76144	-0.180	C1	Alkane CH_3
opls_140	2.50000	1.25520	0.060	C1 H atoms, C2 H atoms	Alkane H
opls_157	3.50000	2.76144	0.145	C2	CH_3 & CH_2: alcohols
opls_154	3.12000	7.11280	-0.683	O1	O: mono alcohols
opls_155	0.00000	0.00000	0.418	O1 H atom	H(O): mono alcohols

*[1] 1 e=1.602176634×10^{-19} C; *[2] See atom labels in Figure 11.

Table 4 – OPLS-AA atom types used for ethyl acetate.

Atom type	σ_{ii} (10^{-1} nm)	ε_{ii} (10^{-1} kJ mol^{-1})	q_i(e*[1])	Atoms*[2]	Atom type description
opls_135	3.50000	2.76144	-0.180	C1, C5	Alkane CH_3
opls_136	3.50000	2.76144	0.190	C4	Alkane CH_2
opls_140	2.50000	1.25520	0.060	C1 H atoms, C5 H atoms	Alkane H
opls_777	2.42000	6.27600	0.030	C4 H atoms	α-methoxy H
opls_4651*[3]	3.1875	4.39320	0.750	C2	Ester carbonyl C
opls_4662*[3]	3.1080	7.02912	-0.550	O1	Ester carbonyl O
opls_4671*[3]	2.55000	7.11280	-0.450	O2	Alkoxy O

*[1] 1 e=1.602176634×10^{-19} C; *[2] See atom labels in Figure 11; *[3] Taken from Pluhackova *et al.* [86]

The parameters used for quercetin were also from the OPLS-AA for fields but with the Ryckaert-Bellemans parameters for four dihedral angles, namely, O7-C8-C11-C12, O7-C8-C11-C16, C9-C8-C11-C12 and C9-C8-C11-C16, which were taken from the LigParGen OPLS/CM1A Parameter Generator for Organic Ligands by the Jorgensen group [87–89]. The force field parameters for quercetin are presented in Table 5, with the atomic labelling given in Figure 12. The charges for quercetin were calculated with the CHelpG scheme using an optimized geometry for the quercetin molecule in gas phase. The latter calculations considered the B3LYP/6-311+G(d) approach as included in the Gaussian 03 code [90].

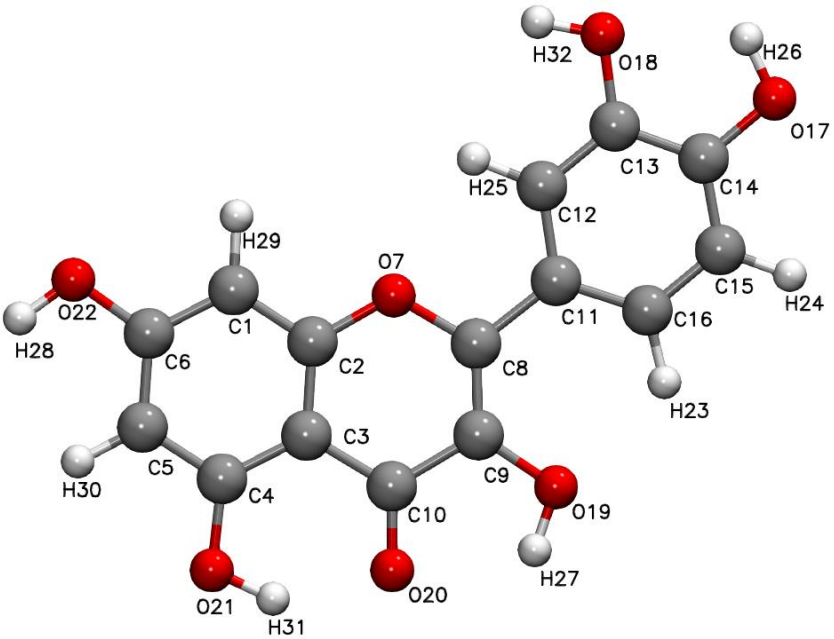


Figure 12 – Quercetin molecule labelled with atom names for identification, rendered in the VMD software [21].

Table 5 – OPLS-AA atom types used for Quercetin.

Atom type	σ_{ii} (10^{-1} nm)	ε_{ii} (10^{-1} kJ mol^{-1})	q_i (e*[1])	Atoms*[2]	Atom type description
opls_145	3.55000	2.92880	−0.505	C1	Benzene C
			0.456	C2	
			−0.472	C3	
			−0.567	C5	
			−0.326	C12	
			−0.258	C15	
			−0.145	C16	
opls_145B	3.55000	2.92880	0.163	C8	Biphenyl C1
			0.117	C11	
opls_146	2.42000	1.25520	0.140	H23	Benzene H
			0.168	H24	
			0.155	H25	
			0.217	H29	
			0.204	H30	
opls_166	3.55000	2.92800	0.573	C4	C(OH) phenol
			0.527	C6	
			0.036	C9	
			0.210	C13	
			0.341	C14	
opls_167	3.07000	7.11280	−0.559	O17	O phenol
			−0.606	O18	
			−0.506	O19	
			−0.572	O21	
			−0.561	O22	
opls_168	0.00000	0.00000	0.438	H26	H phenol
			0.413	H27	
			0.411	H28	
			0.436	H31	
			0.447	H32	
opls_280	3.75000	4.39320	0.520	C10	AA C: ketone
opls_281	2.96000	8.78640	−0.582	O20	AA O: ketone
opls_571	2.90000	5.85760	−0.312	O7	Oxazole O

Missing dihedral	C0 (kJ mol^{-1})	C1 (kJ mol^{-1})	C2 (kJ mol^{-1})	C3 (kJ mol^{-1})	C4 (kJ mol^{-1})	C5 (kJ mol^{-1})
	9.079	0.000	−9.079	0.000	0.000	0.000

*[1] 1 e=1.602176634×10^{-19} C; *[2] See atom labels in Figure 12.

As mentioned in Chapter 3.2, the diffusion coefficients were calculated using the long time limit of the mean square displacement (MSD) through the Einstein relation for three dimensions, since the molecule is in isotropic media and all directions may be taken into account [66,67]:

$$D_{\mathrm{i}} = \lim_{\mathrm{t}\to\infty} \frac{\langle |\boldsymbol{r}_{\mathbf{i}}(t) - \boldsymbol{r}_{\mathbf{i}}(0)|^2 \rangle}{6t} \qquad \text{(Eq. 3.25)}$$

where t is the elapsed time from the time origin t_0 (or observed time, in the sense that the displacement is the difference in the position of an interval of t time units), and $\boldsymbol{r}_i$ is the position of a particle. The average is carried out over all possible time origins and over all molecules. The diffusion coefficient is calculated from the slope of the linear MSD *versus* t plot obtained from the MD trajectories with the *gmx msd* tool included in the GROMACS code. The uncertainty associated with this calculation is the uncertainty of the slope of this linear regression, if more simulations were done the uncertainty would be the greater between the uncertainty of the slope and the standard deviation between the calculations from a number of different simulations, at the same conditions.

Finally, the structure of the rings was analyzed by the means of an angle distribution (*gmx angle* tool included in GROMACS), while the interactions between ethanol molecules and quercetin were analyzed from the radial distribution functions (*gmx rdf* tool included in GROMACS) between the ethanol oxygen atom and each of the quercetin's oxygen atoms and from the spatial distribution functions (*sdf* tool included in the TRAVIS (Trajectory Analyzer and Visualizer) code [91]) for spatial representation of ethanol carbon and oxygen atom density around quercetin.

5. Results

This chapter presents the benchmarking of the solvents force field parameters upon calculation of relevant solvent properties (Section 5.1), the optimization of simulation parameters (Section 5.2), the results for tracer diffusion coefficients of quercetin in ethanol and quercetin in ethyl acetate (Section 5.3) and, lastly, structural and distribution function analyses of the quercetin in ethanol systems (Section 5.4).

5.1. Solvent properties from molecular dynamics

The experimental values for the self-diffusion coefficient, D_{11}^{exp}, were taken from Kato *et al.* [92] for ethanol, and from Uminski *et al.* [93] for ethyl acetate, and are presented in Table 6 along with the values of density calculated with the Tait equation [68,69]:

Table 6 – Self-diffusion coefficients [92,93] and density values calculated using the Tait equation [68,69].

Solvent	D_{11}^{exp} (10^{-10} m^2 s^{-1})	ρ_{calc} (kg m^{-3})
Ethanol	10.70±0.03 [92]	785.9
Ethyl Acetate	27.70 [93]	893.6

As a first step, to verify the validity of the solvents force field parameters, the density, ρ_{MD}, and the self-diffusion coefficient, D_{11}^{MD}, were calculated, respectively, from a NPT ensemble simulation of 10 ns duration and step size of 2 fs, and from an NVT ensemble simulation of 12 ns duration, 6 ns of equilibration phase and production phase, and step size of 1 fs. The simulations contained 2500 molecules of the solvent, at a temperature of 25 °C and a pressure (or volume equivalent to the pressure) of 1 bar. An NVT ensemble simulation for pure ethanol was performed with the density from a previous NPT simulation, ρ_{MD}, giving an estimated self-diffusion coefficient, D_{11}^{MD}, with 2.80 % of relative error towards D_{11}^{exp}. However, using the values of density calculated using the Tait equation, ρ_{calc}, the error increased to 28.97 %, suggesting that a 1.11 % difference in density may incur in a large error in diffusion, as consequence it was decided to use ρ_{MD} for the remaining simulations. The results of these simulations are present in Table 7.

Table 7 – Self-diffusion coefficients and density of ethanol calculated from MD simulations. Relative error in % towards D_{11}^{exp} in parenthesis.

	Ethanol
$D_{11}^{\mathrm{MD},\rho_{\mathrm{calc}}}$ (10^{-10} m^2 s^{-1})	13.786±0.137 (28.97 %)
$D_{11}^{\mathrm{MD},\rho_{\mathrm{MD}}}$ (10^{-10} m^2 s^{-1})	10.953±0.858 (2.80 %)
ρ_{MD} (kg m^{-3})	794.6 (1.11 %)

The MD simulations for ethyl acetate, presented in Table 8 were run using ρ_{MD} for system density. The first simulation was run using the OPLS-AA force field parameters [83–85] included in GROMACS giving a value of 16.573 x 10^{-10} m^2 s^{-1} with an error of −40.17 %. To solve this issue, new parameters were searched for and the Pluhackova *et al.* LOPLS-AA parameters [86] were used. While the results improved the D_{11} to a value of 23.432 × 10^{-10} m^2 s^{-1}, the relative error remained high, *i.e.* −15.52 %, suggesting that the parameters may not be adequate for diffusion coefficient determination.

Table 8 – Self-diffusion coefficients of ethyl acetate calculated by MD simulation using ρ_{MD} for the system density, with GROMACS' OPLS-AA force field parameters [83–85] and the LOPLS-AA force field parameters [86]. Relative error in % towards experimental values in parenthesis.

Ethyl Acetate	
$D_{11}^{\mathrm{MD,OPLS}}$ (10^{-10} m^2 s^{-1})	16.573±0.246 (−40.17 %)
$D_{11}^{\mathrm{MD,LOPLS}}$ (10^{-10} m^2 s^{-1})	23.432±0.237 (−15.52 %)
$\rho_{\mathrm{MD}}^{\mathrm{OPLS}}$ (kg m^{-3})	918.6 (2.79 %)
$\rho_{\mathrm{MD}}^{\mathrm{LOPLS}}$ (kg m^{-3})	894.2 (0.07 %)

5.2. Optimization of simulation parameters

As a second step, MD simulations in NVT ensemble containing varied number of molecules of quercetin and ethanol were run to verify the influence of the parameters on the value of the tracer diffusion coefficient of quercetin in ethanol, D_{12}^{MD}, at 30 °C and 1 bar. The experimental value of the tracer diffusion coefficient of quercetin in ethanol, D_{12}^{exp}, at these conditions is 4.415±0.026 × 10^{-10} m^2 s^{-1} and the calculated density, ρ_{calc}, is 782.0 kg m^{-3}. Taking the parameter of 1.4 nm cut-off distance from Vaz *et al.* [49] as a starting point, the short-range interaction cut-off distance, the duration of the simulation (both equilibration phase and production phase), number of molecules of solvent (*i.e.* system size) and, additionally, the frequency to write trajectory frames were varied and tested.

For the short-range interaction cut-off distance, the values were varied in the range 1.3-1.5 nm in increments of 0.1 nm, and the average and standard deviations between simulations of the same cut-off are presented in Table 9. The systems had 2500 molecules of ethanol and 1 of quercetin, kept at 30 °C and with the density for that temperature at 1 bar from a 10 ns NPT simulation, ρ_{MD}, of 792.2 kg m^{-3}. The equilibration phase was of 15 ns and the production phases were of 60 ns.

Table 9 – Short range cut-off distance variation averaged results for a system of 2500 molecules of ethanol and 1 quercetin at 30 °C, 1 bar, equilibration phase of 15 ns and 60 ns production phase. Relative error in % towards the experimental value in parenthesis.

Cut-off (nm)	D_{12}^{MD} (10^{-10} m^2 s^{-1})
1.3	5.305±0.245 (20.15 %)
1.4	4.655±0.411 (5.43 %)
1.5	4.780±0.135 (8.26 %)

As the value of the cut-off increased, the error decreased considerably from 1.3 to 1.4 nm of cut-off radius. However, from 1.4 to 1.5 nm, the improvement was not significant, making it smaller than the uncertainties currently associated with the estimated values. Therefore 1.4 nm is enough to obtain reliable values for all remaining simulations while being computationally more advantageous.

In Table 10 are reported the results from the tests varying the number of molecules of solvent (*i.e.* system size). The simulations considered a cut-off of 1.4 nm, 15 ns duration for equilibration phase and 60 ns for production phase, and ranged between 500 to 4000 the molecules of solvent, and 1 to 12 molecules of solute. 500 molecules of ethanol makes for a very small system, presenting a relative error of -12.67 %, which decreased as the number of molecules was increased, reaching values of 0.63 % and 12.09 % with 2500 molecules of ethanol and 1 molecule of quercetin, 5.43 % with 2500 molecules of ethanol and 3 molecules of quercetin, and 4.57 % with 4000 molecules of ethanol and 1 molecule of quercetin. Inserting more molecules of quercetin in the simulation showed improvement on the diffusion coefficient averaging than just a single molecule. In order to improve the accuracy of the D_{12} taken from a single simulation, more molecules of the solute should be placed in the box. However, this can lead to the molecule encountering and moving close to another of itself and compromising infinite dilution conditions as the molecule is large, which can be seen with 12 molecules of quercetin with which resulted in a relative error of -8.70 %. As such, 2500 molecules of ethanol and 3 molecules of quercetin were the fixed values for the remaining simulations.

Table 10 – Study of the influence of the system size and number of solute molecules upon D_{12}^{MD}, at 30 °C, 1 bar, duration of 15 ns equilibration phase and 60 ns of production phase, cut-off distance of 1.4 nm. Relative error in % towards the experimental value in parenthesis.

$\mathrm{N}_{solvent}$	N_{solute}	D_{12}^{MD} (10^{-10} m^2 s^{-1})
500	1	3.856±0.167 (−12.67 %)
1000	1	3.903±0.337 (−11.60 %)
2500	1	4.949±0.361 (12.09 %)
		4.443±0.863 (0.63 %)
2500	3	4.502±0.014 (1.96 %)
2500	12	4.031±0.308 (−8.70 %)
4000	1	4.617±1.414 (4.57 %)

The duration of the simulations for both the equilibration and production phases were tested as well, employing a cut-off value of 1.4 nm, for systems composed of 2500 molecules of ethanol and 3 of quercetin, kept at 30 ºC and 1 bar. The averages and standard deviations are presented in Table 11.

Table 11 – Influence of phase duration upon D_{12}^{MD} for a system of of 2500 molecules of ethanol and 3 of quercetin at 30 °C, 1 bar, equilibration phase of 6 ns, cut-off distance of 1.4 nm. Relative error in % towards the experimental value in parenthesis.

Production phase (ns)	Equilibration phase (ns)	D_{12}^{MD} (10^{-10} m^2 s^{-1})
30	5	4.853±0.170 (9.91 %)
	35	4.712±0.077 (6.72 %)
30	15	4.541±0.008 (2.85 %)
	45	4.482±0.066 (1.51 %)
60	3	4.937±0.089 (11.82 %)
60	5	4.759±0.087 (7.79 %)
60	15	4.502±0.014 (1.96 %)
55	3	5.082±0.014 (15.10 %)
55	5	4.897±0.069 (10.91 %)
55	10	4.617±0.160 (4.57 %)
55	15	4.617±0.032 (4.57 %)
55	20	4.418±0.033 (0.06 %)
45	15	4.761±0.011 (7.83 %)
45	20	4.518±0.038 (2.33 %)

The relative error decreases with the increase of the equilibration phase, indicating the simulation did not reach equilibrium until 20 ns, suggesting that simulation times greater equal this value for the equilibration phase. As for the production phase a duration of 30 ns still accounts for some variability in the results. For these reasons the equilibration phase was fixed in 20 ns and the production phase 55 ns. A much larger simulation should be done to find the optimal parameters.

Finally, due to the large file sizes of the simulations, the frame write frequency was varied in order to reduce the file size without losing the accuracy of the results. Picking random simulations performed during the dissertation, the values for different writing frequencies are compared in Table 12. Between frequencies of every 1 frame to 2000 frames the values seemed nearly unaffected and the trajectory file size for systems of 2500 molecules of ethanol reduced from 3688.0 GB to 1.8 GB, respectively, allowing for the use of whichever is preferred. However, it must be considered that smaller values are needed for smaller time intervals to be taken into account.

Table 12 – Diffusion values for different simulations at different frame writing frequencies and trajectory file size. Relative error in % towards the experimental value in parenthesis.

Simulation conditions	Writing frequency	D_{12}^{MD} (10^{-10} m^2 s^{-1})	File size (GB)
60 ºC, 1 bar 1.4 nm cut-off radius	100	8.184±0.208 (4.57 %)	61.7
	1000	8.184±0.208 (4.57 %)	6.2
30 ºC, 1 bar 1.5 nm cut-off radius	40	5.305±0.245 (20.15 %)	97.8
	80	5.305±0.245 (20.15 %)	48.9
	120	5.305±0.245 (20.02 %)	32.6
	200	5.299±0.204 (20.15 %)	19.6
	1000	5.305±0.244 (20.15 %)	3.9
30 ºC, 1 bar 1.5 nm cut-off radius	1	4.780±0.135 (8.26 %)	3688.0
	20	4.780±0.135 (8.26 %)	184.4
	40	4.780±0.135 (8.26 %)	92.2
	80	4.780±0.135 (8.26 %)	46.1
	100	4.780±0.135 (8.26 %)	36.9
	1000	4.780±0.133 (8.26 %)	3.7
	2000	4.781±0.132 (8.28 %)	1.8

Ideally all the prior tests would have been repeated, and done for the quercetin-ethyl acetate systems, but due to time constraints the results obtained in the tests for quercetin-ethanol systems were used instead. The parameters that yielded the best values for tracer diffusion coefficients in these tests were: a cut-off distance of 1.4 nm, equilibration phase of 20 ns, production phase of 55 ns, 2500 solvent molecules and 3 quercetin molecules.

5.3. Tracer diffusion coefficients

Quercetin in Ethyl Acetate

The measured results for tracer diffusion coefficients, $D_{12}^{\exp}$, of quercetin in ethyl acetate at a wavelength of 270 nm (study of wavelength in Appendix B), are presented in Table 13 along with the calculated values of density and viscosity through the Tait and Mamedov equations [68–70]. The values ranged from 10.18 × 10^{-10} to 16.28 x 10^{-10} m^2 s^{-1}, for the temperature range of 30–60 °C and pressure range of 1–150 bar, and are presented graphically as function of pressure in Figure 13. These values were averaged over at least three to six different measurements from which the associated uncertainties were calculated. The $D_{12}^{\exp}$ values are in the same order of magnitude of those for other compounds in compressed liquids, such as eucalyptol in ethanol [94], astaxanthin and squalene in ethyl acetate [95], benzyl acetate, 2-phenylethyl acetate and 3-phenylpropyl acetate in ethanol [96].

Table 13 – Experimental diffusivity results for quercetin in ethyl acetate. and calculated values of density and viscosity.

P (bar)	T (°C)	ρ_{calc}*1 (kg m^{-3})	μ_{calc}*2 (cP)	$D_{12}^{\exp}$ (10^{-10} m^2 s^{-1})	P (bar)	T (°C)	ρ_{calc}*1 (kg m^{-3})	μ_{calc}*2 (cP)	$D_{12}^{\exp}$ (10^{-10} m^2 s^{-1})
1	30	887.7	0.3994	11.11±0.03	100	30	897.0	0.4439	10.61±0.02
	40	875.6	0.3590	12.70±0.07		40	885.0	0.4003	12.13±0.04
	50	863.4	0.3247	14.56±0.05		50	873.8	0.3632	13.76±0.08
	60	850.8	0.2952	16.28±0.05		60	861.0	0.3315	15.31±0.02
50	30	892.0	0.4223	10.90±0.04	150	30	901.0	0.4644	10.18±0.03
	40	881.0	0.3802	12.42±0.08		40	890.0	0.4193	11.64±0.09
	50	869.0	0.3444	13.89±0.08		50	878.0	0.3810	13.24±0.04
	60	856.0	0.3138	15.80±0.06		60	867.0	0.3484	14.75±0.08

*1 Calculated by Tait equation [68,69]; *2 Calculated by Mamedov equation [70].

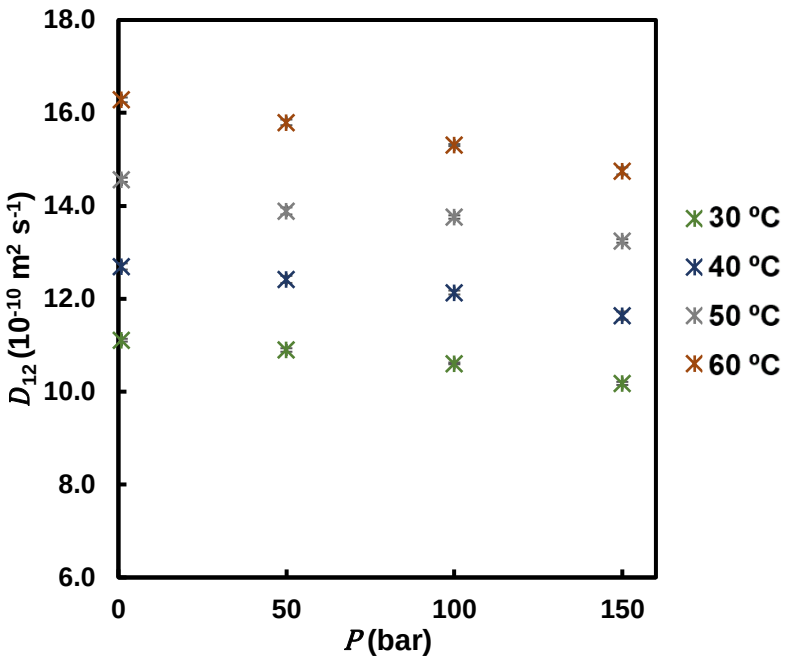


Figure 13 – Tracer diffusion coefficient, D_{12} (10^{-10} m^2 s^{-1}), measured for quercetin in ethyl acetate, as function of pressure at distinct temperatures.

The MD simulation results for quercetin in ethyl acetate at 30 °C and 1 bar was 8.608 x 10^{-10} m^2 s^{-1}, which corresponds to a relative error of about -22.51 % towards the experimental value. This was expected since the self-diffusion of ethyl acetate had a similar relative error. As such no more simulations for ethyl acetate were performed. It is probable that the force field parameters might not be optimized for self-diffusion coefficient determination and, consequently, binary diffusion coefficients. As such, it is suggested the development of new parameters as future work [19].

Quercetin in ethanol

The experimental results for $D_{12}^{\exp}$ of quercetin in ethanol, measured by another researcher of the EgiChem group, are presented in Table 14 along with the calculated values of density and viscosity through the Tait and Mamedov equations [68–70]. The values ranged from 3.985 × 10^{-10} to 7.823 x 10^{-10} m^2 s^{-1}, for the temperature range of 30–60 °C and pressure range of 1–150 bar, and are presented graphically as function of pressure in Figure 15. The $D_{12}^{\exp}$ values are an order of magnitude close to that of the previously mentioned solutes in compressed liquids and those of quercetin in ethyl acetate.

Table 14 – Experimental diffusivity results for quercetin in ethanol, and calculated values of density and viscosity.

P (bar)	T (°C)	ρ_{calc}*1 (kg m^{-3})	μ_{calc}*2 (cP)	D_{12}^{exp} (10^{-10} m^2 s^{-1})	P (bar)	T (°C)	ρ_{calc}*1 (kg m^{-3})	μ_{calc}*2 (cP)	D_{12}^{exp} (10^{-10} m^2 s^{-1})
1	30	782.0	0.9650	4.42±0.03	100	30	789.9	1.0488	4.11±0.05
	40	773.0	0.8100	5.42±0.03		40	781.7	0.8854	5.02±0.08
	50	764.0	0.6870	6.55±0.02		50	773.5	0.7547	6.10±0.02
	60	756.0	0.5870	7.83±0.06		60	765.2	0.6484	7.27±0.05
50	30	785.8	1.0067	4.26±0.01	150	30	793.7	1.0898	3.98±0.01
	40	777.4	0.8478	5.22±0.02		40	785.7	0.9221	4.90±0.04
	50	769.0	0.7209	6.29±0.07		50	777.7	0.7876	5.92±0.03
	60	760.6	0.6179	7.49±0.08		60	769.6	0.6780	7.09±0.02

*1 Calculated by Tait equation [68,69]; *2 Calculated by Mamedov equation [70].

Figure 14 presents experimental and MD D_{12} results for quercetin in ethanol against temperature, while Figure 15 shows the variation of the experimental and calculated D_{12} values as function of pressure.

By observing both figures, the experimental values show a decrease of D_{12} with increasing pressure, however very small when compared to the increase with temperature increment, possibly because of the low compressibility of ethanol. Similar trends were found by other authors [94,95] for other solutes in pure liquid ethanol and, as can be seen in Figure 13, quercetin in ethyl acetate also shows analogous behavior. Considering the free volume theory [97], this decrease of D_{12} with increasing pressure is explained by the increase in solvent density, which reduces the free volume available for the solute to move [97]. While the increment caused by temperature increase is explained by the increment of internal energy of the system, allowing the solute to overcome the energy barrier necessary to escape from the solvent's force field more easily, and lower density of the solvent (higher free volume) [97].

The values of quercetin in ethyl acetate, in Table 13 are an order of magnitude above quercetin in ethanol, this may be explained by the polar nature of quercetin [98], as ethanol is a polar solvent [99] and capable of forming strong polar bonds with quercetin hydroxyl groups it might reduce the mobility of the quercetin molecules, while ethyl acetate is less polar [99] and has no hydroxyl groups.

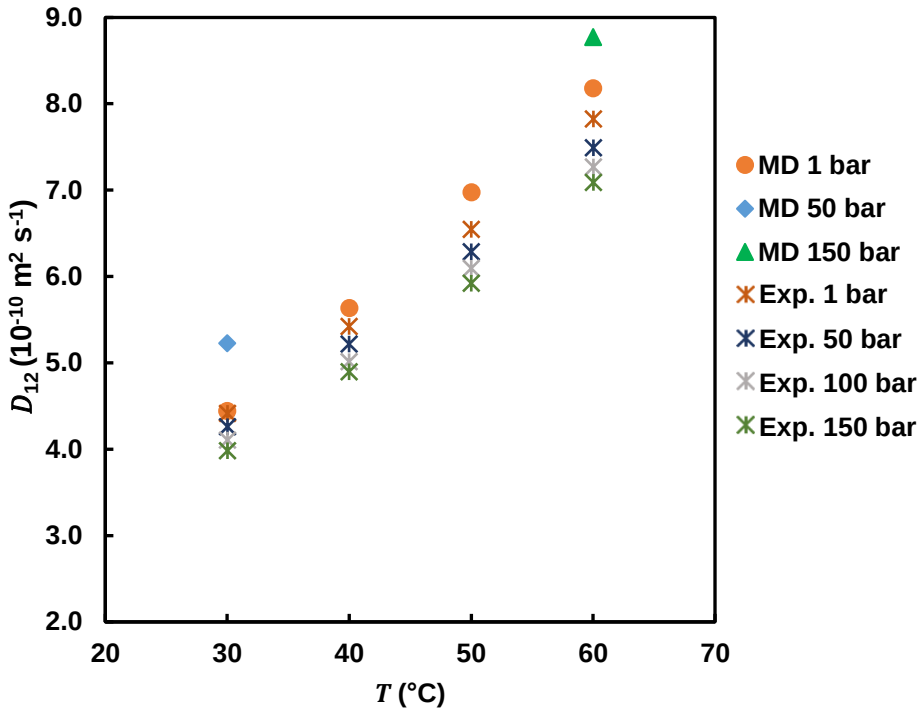


Figure 14 – Tracer diffusion coefficient (experimental and MD simulation) quercetin in ethanol, as function of temperature.

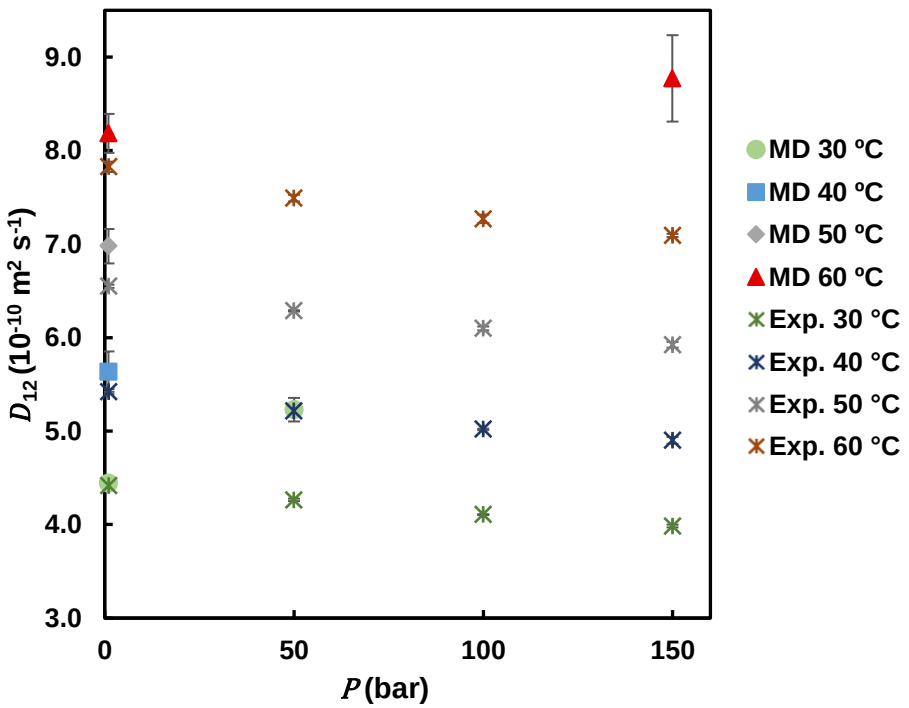


Figure 15 – Tracer diffusion coefficient (experimental and MD simulation) quercetin in ethanol, as function of pressure.

The MD simulation results for quercetin in ethanol at 30 °C and 1 bar are presented in Table 15, and range between 4.418 × 10^{-10} and 8.184 × 10^{-10} m^2 s^{-1} with relative errors below 6.54 %. They follow the same trend for temperature however show large enough uncertainties that may not ensure the same for the pressure trends as the decrease of diffusion with pressure in compressed liquids is rather small. The large uncertainty can be due to the low quercetin count/short duration of the simulation leading to an average with large deviations, requiring larger simulations or repetitions, or due to the fact that a small deviation in density can cause a large error in diffusion, as previously seen with ethanol self-diffusion in Table 7. Density in these NVT ensemble simulations is the way pressure is ensured in the simulation and seeing that in liquids with low compressibility a small change in density can mean a large change in pressure. In order to improve the density values, a longer NPT ensemble simulation could be adequate. To ensure the correct pressure in NVT ensemble, aside from a more precise density, a larger system for the simulation may help as large pressure fluctuations are to be expected but they reduce in proportion to the square root of the number of system particles [19].

Table 15 – System conditions, MD diffusivities of quercetin in ethanol and density calculated *via* separate NPT simulations. Relative deviations in % between MD and experimental values are in parenthesis.

System conditions		Estimated property via MD simulation	
P (bar)	T (°C)	D_{12}^{MD} (10^{-10} m^2 s^{-1})	ρ_{MD} (kg m^{-3})
1	30	4.418±0.033 (0.06 %)	792.2 (1.30 %)
	40	5.636±0.216 (3.94 %)	780.6 (0.98 %)
	50	6.977±0.184 (6.54 %)	769.5 (0.72 %)
	60	8.814±0.208 (4.57 %)	757.4 (0.19 %)
50	30	5.229±0.126 (22.62 %)	795.7 (1.26 %)
150	60	8.772±0.462 (23.69 %)	771.4 (0.23 %)

MD simulation D_{12}^{MD} values are in good accordance with experimental values as can be seen for the conditions of 30, 40, 50 and 60 °C at 1 bar presented in Table 15. The values at 50 bar and 150 bar have large errors of above 20 %, which may be attributed to the fact that force field parameters were not developed to represent ethanol at higher pressures. As suggested by Hölzl *et al*. [100], it is possible to empirically adjust the charges of the atoms for the pressures in question to avoid total re-parameterization, taking into consideration that the electronic structure of the molecules in the solution change slightly upon compression, which may impact the interactions between solute and solvent. To test this hypothesis, the atom charges of ethanol were multiplied by factors of 1.01, 1.02 and 1.1 at 60 °C and 150 bar as shown in Figure 16.

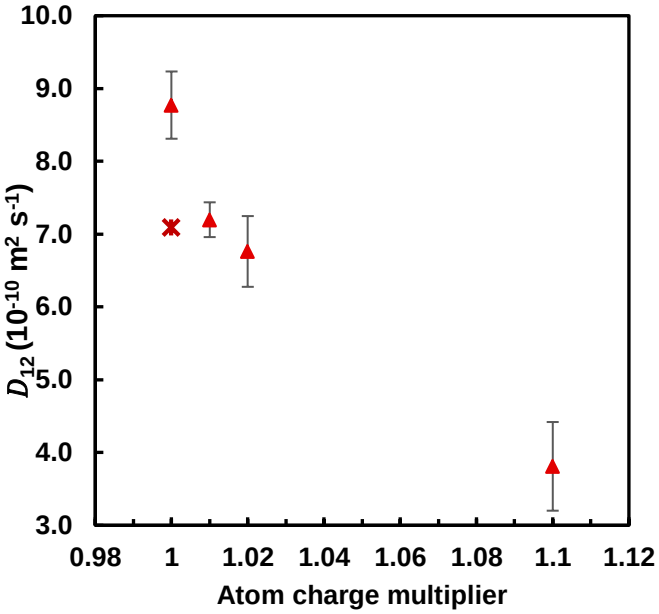


Figure 16 – MD diffusivities of quercetin in ethanol at 60 °C and 150 bar for different values of atom charge multiplier. The experimental value is shown as a red asterisk for comparison.

With standard charges the simulation yielded a diffusion of (8.772 ± 0.462) × 10^{-10} m^2 s^{-1} with a relative error of 23.69 % towards the experimental value. Using scale factors of 1.01, 1.02 and 1.1, the D_{12}^{MD} values were (7.198 ± 0.239) × 10^{-10} m^2 s^{-1} (1.49 %), (6.762 ± 0.486) × 10^{-10} m^2 s^{-1} (−4.65 %) and (3.810 ± 0.486) × 10^{-10} m^2 s^{-1} (−46.28 %), respectively, indicating that D_{12}^{MD} values can be improved significantly by empirically simulating the effects of the compression on the atom charges and the interactions between solute and solvent. This should be tested for other pressures in ethanol and in a proper ethyl acetate force field to verify if the same scale factors can be applied to less polar solvents or they must be tuned for each solvent.

5.4. Structure and distribution function analysis

The interactions between the solute and the solvent were analyzed in this dissertation taking into account angle distributions, radial distribution functions (RDF or pair correlation function, $g(r)$) and spatial distribution functions (SDF). These were computed for quercetin in ethanol at 30°C and 1 bar, using the appropriate tools included in the GROMACS software package [19,20].

First, an angle distribution was calculated for the atoms O7-C8-C11-C12 (a dihedral or torsion angle as labeled in Figure 17), since this angle distribution allows to verify if the single ring on quercetin rotates in ethanol. The angle distribution is a probability function where the angle probability is calculated over the time of the simulation and averaged over

all angles of the same type. The angles vary between -180° through 0° up to 180°, where -180° and 180° means that the first and last particle of the angle are in opposite sides (trans conformation), and 0° that they are on the same side (cis conformation), as illustrated in Figure 17. Only three molecules of quercetin were included in the simulation box, as such this torsion angle distribution corresponds to the angles averaged over the 55 ns of production phase of the simulation for those molecules. It can be observed in Figure 18 that the ring rotates freely, with preference to the cis conformation which is the starting conformation for all quercetin molecules for in the performed simulations. It is expected that the rings of quercetin rotate, as the energy barrier for this rotation decreases in solution, and prefer to oscillate close to either cis or trans conformation of this angle as they are identical in terms of potential energy [101,102].

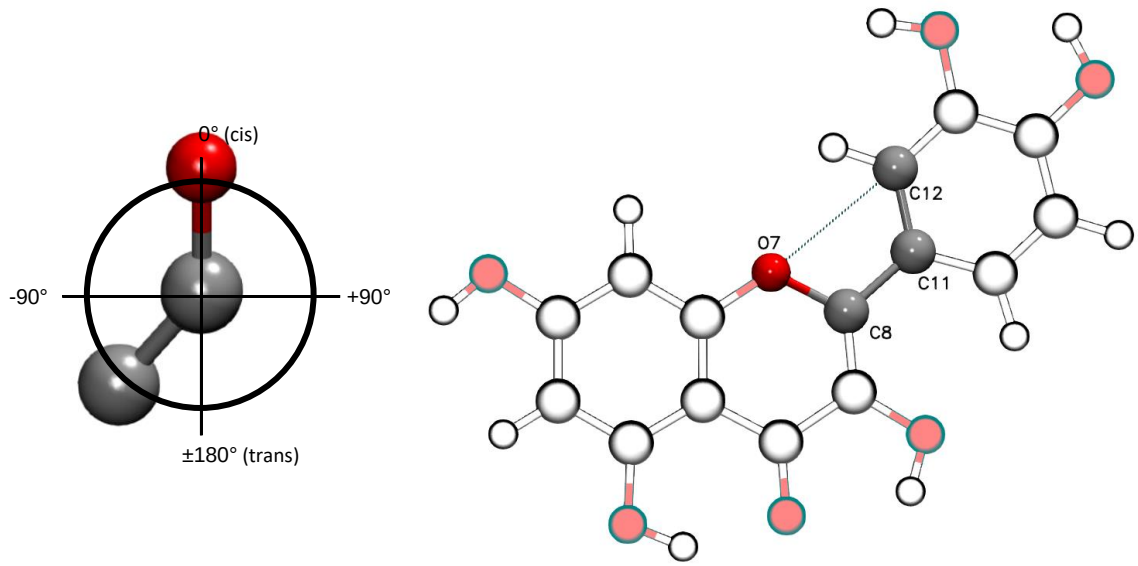


Figure 17 – a) Angle conformation values example diagram and b) Dihedral angle O7-C8-C11-C12 in cis conformation.

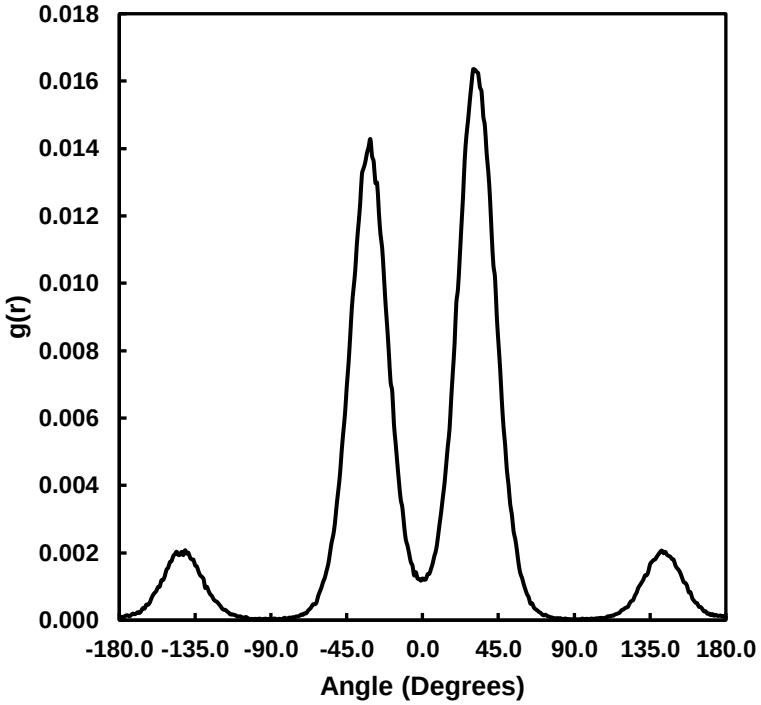


Figure 18 – Torsion angle distribution function of the angle O7-C8-C11-C12 of quercetin in the simulation at 30°C and 1 bar.

The radial distribution function is calculated between two particle types: a central particle A and the surrounding particles B, giving a measure of the density of particles B at a distance r from a central particle A, averaging over all particles A [19,49,63]. The spatial distribution function is in a way a "three-dimensional version" of the RDF, where iso-surfaces of fixed particle density (particles/nm^3) can be observed [91], allowing to extract information of how different particles locate themselves towards a set of reference particles.

It is possible to observe in Figure 19 that the quercetin oxygens have very different probabilities of interacting with ethanol hydroxyl group. The quercetin oxygens O18, O22 and O17 exhibit the most favorable interactions with ethanol hydroxyl groups, followed by the O19 and O20 oxygens, by that order. The O21 and O7 oxygens do not present relevant interaction with this group given that the density of ethanol hydroxyl groups around them is lower than that in the bulk. When looking at the charges in Table 5 the charges of oxygens alone don't seem to always decrease in the same order as the order of preference of the ethanol hydroxyl group. The density of hydroxyl groups surrounding O18, O22 and O17 decreases with the charge of the atoms −0.606 e > −0.561 e > ~−−0.559 e, respectively (see Table 5). The atoms O19 and O20 show weaker interaction closer to that of the bulk despite O20 has the second largest negative charge −0.582 e. Oxygen O21 despite possessing

third largest negative charge −0.572 e has the second lowest density of hydroxyl groups. The oxygens O19, O20 and O21 are located near each other, the hydroxyl groups of oxygens O19 and O21 may have intra-molecular interactions with the ketone group oxygen O20 and may result in this behavior, prompting further study of the involved atoms. The atom O7 has the smallest negative charge of −0.312 e and is part of the middle of a ring of carbons.

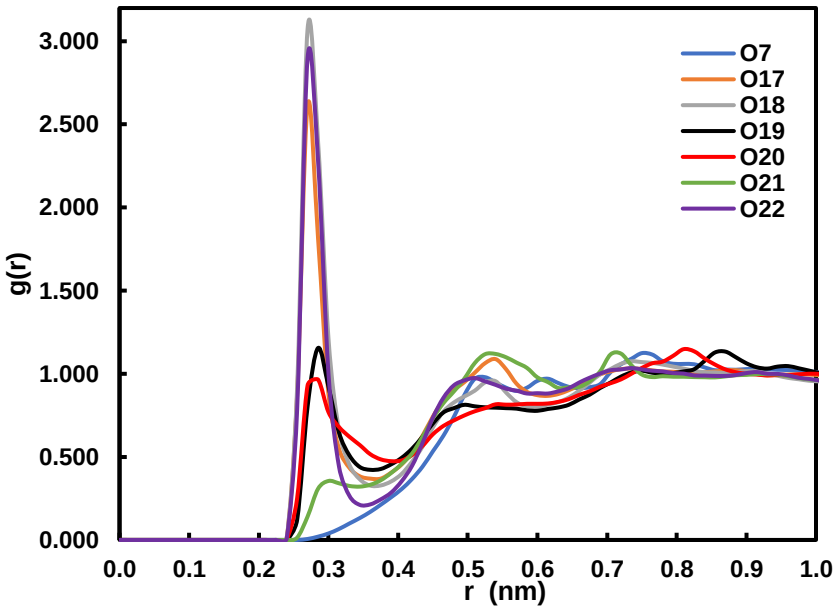
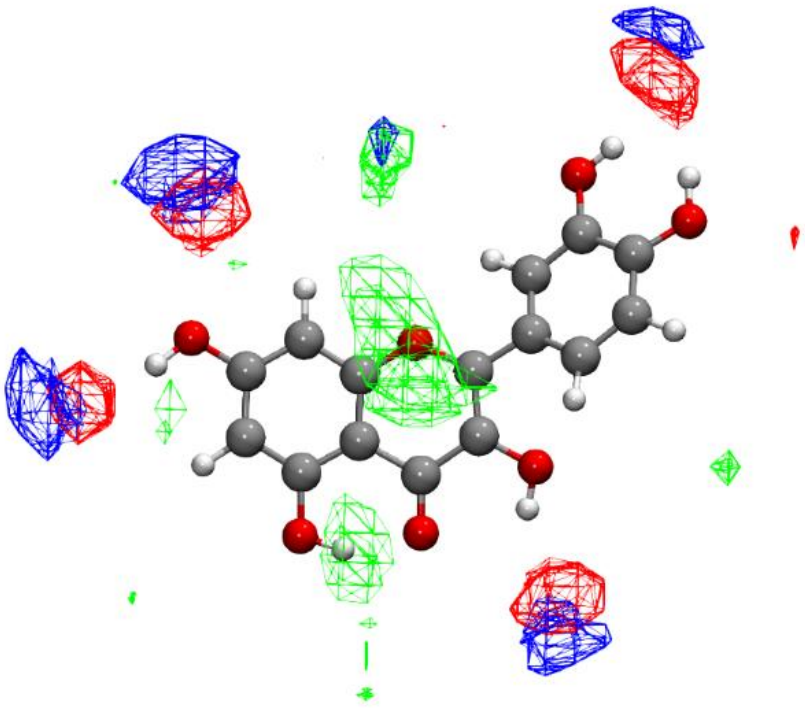


Figure 19 – Radial distribution function (RDF) of all quercetin oxygens towards ethanol OH group oxygen at 30 °C, 1 bar.

The SDFs are presented in Figure 20, in which the double ring is fixed to avoid rotational effects, and in Figure 21, where the single ring is fixed to prevent rotational effects. Being a system with a polar solute and a polar solvent it can be observed in these figures that there is a preference for ethanol hydroxyl groups closer to quercetin hydroxyl groups, confirming the presence of polar interactions between these groups. The ethyl carbons of ethanol only show behind the surfaces of the ethanol oxygen atom surface and on the side of quercetin parallel to the aromatic rings where the red surface stands on the outer side, pointing to the occurrence of non-polar interactions with the aromatic rings. In accordance to the RDFs of Figure 19 it can be observed in Figure 20 and Figure 21 that ethanol hydroxyl groups are grouped near O18, O22, O17 and O19 preferentially.

These structural and distribution function studies can be useful for observing the behavior of particles in each system, however more information can be obtained by comparing different and similar systems, in this case other flavonoids.

a)



b)

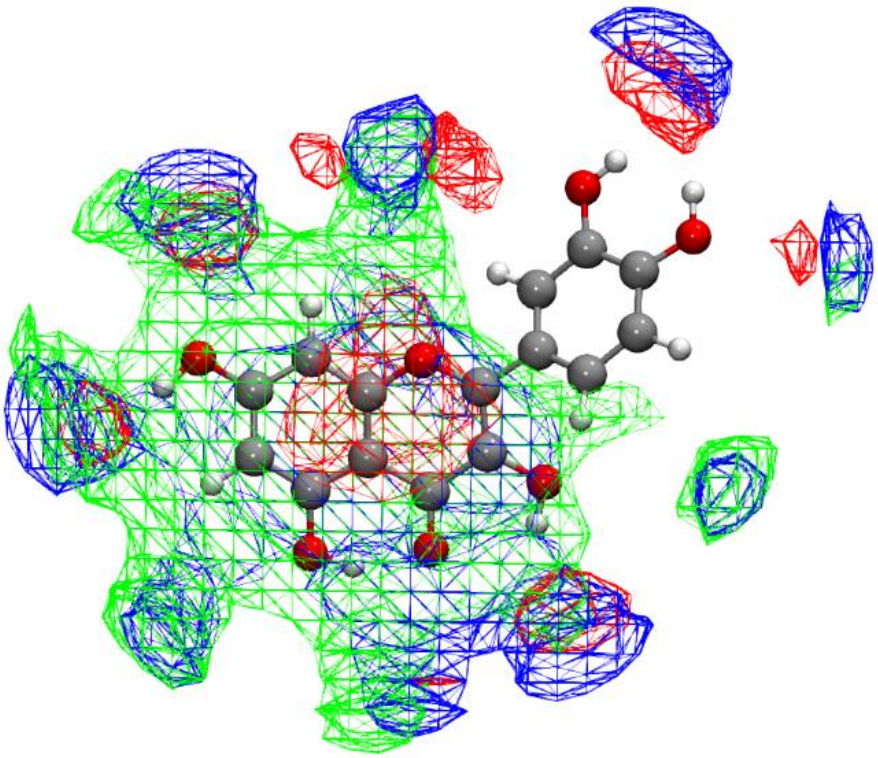
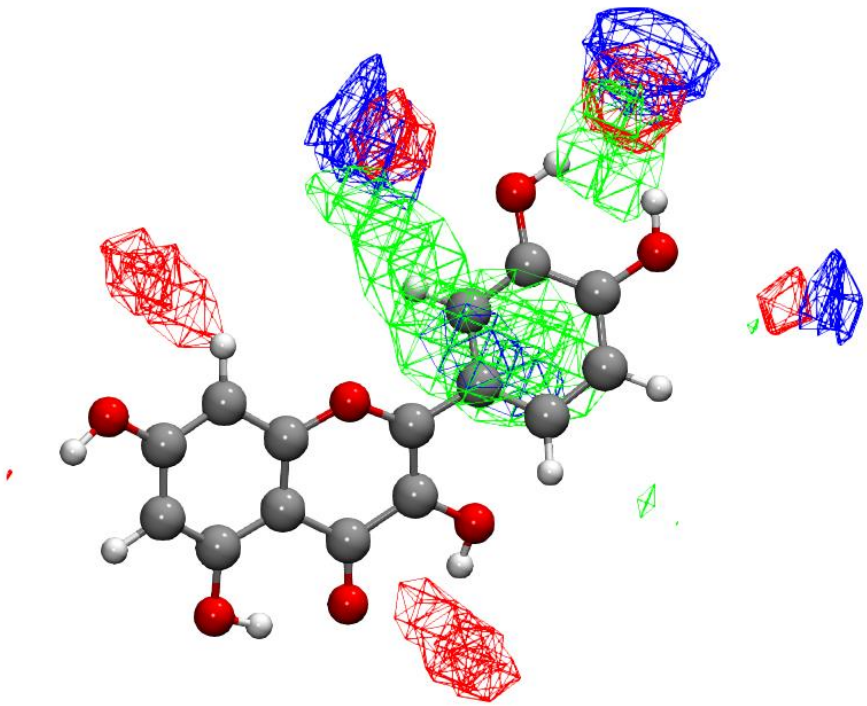


Figure 20 – Spatial distribution function of ethanol hydroxyl oxygen (red surfaces), ethanol CH_2 carbon (blue surfaces), ethanol CH_3 carbon (green surfaces), with the double ring of quercetin fixed as reference, at iso-densities of a) 30 particles/nm^3 and b) 20 particles/nm^3.

a)



b)

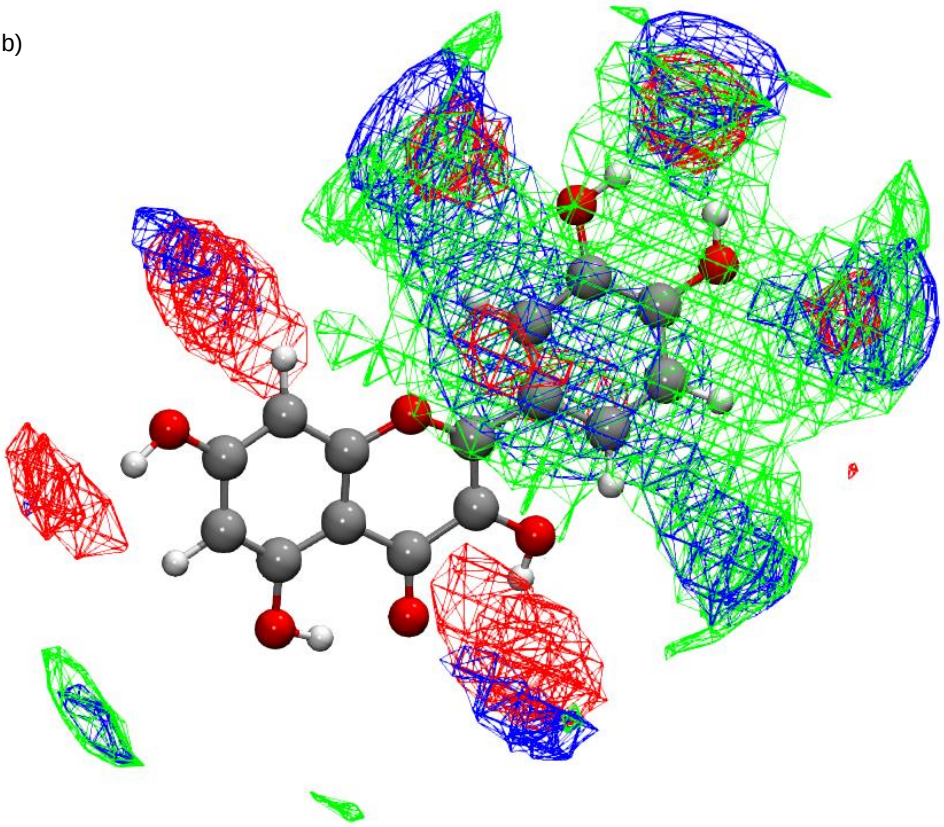


Figure 21 – Spatial distribution function of ethanol hydroxyl oxygen (red surface), ethanol CH_2 carbon (blue surface), ethanol CH_3 carbon (green surface), with the single ring of quercetin fixed as reference, at iso-densities of a) 30 particles/nm^3, b) 20 particles/nm^3..

6. Conclusion and suggestions for future work

Conclusions of the dissertation

In this dissertation, the measurement of tracer diffusion coefficients, D_{12}^{exp}, of quercetin in ethyl acetate was performed using the CPB method at 30−60 °C and 1−150 bar, giving rise to values ranged between 1.018 × 10^{-9} and 1.628 × 10^{-9} m^2 s^{-1}. Their dependency with temperature and pressure was analyzed, and show similar trends and orders of magnitude similar to other systems of solutes in compressed liquids, including quercetin in ethanol.

Molecular dynamics simulations in the NVT ensemble were performed to compute tracer diffusion coefficients, D_{12}^{MD}, of quercetin in ethanol and quercetin in ethyl acetate. A series of preliminary tests were accomplished to validate the force field parameters used for the calculation. Tests including the calculation of the solvents self-diffusion coefficient and an optimization of simulation parameters were performed leading to a cut-off distance of short-range interactions of 1.4 nm, 2500 molecules of solvent and 3 molecules of solute, and duration of the simulation of 20 ns equilibration phase and 55 ns production phase per simulation.

The diffusivity value of quercetin in ethyl acetate, at 30 °C and 1 bar, presents an error of −22.51 %, which is a large deviation consistent with previous MD simulations for the self-diffusivity of ethyl acetate (15.52 %), leading to the conclusion that the parameters used are not suitable for diffusion coefficient calculation in these systems, as they were not parameterized for such end.

For quercetin in ethanol the obtained diffusivities for the temperature range of 30-60 °C at 1 bar are validated with errors below 6.54 %. The temperatures in this interval allow for accurate estimations, however increasing pressure to 50 bar or above results in wrong estimations, meaning that the force field parameters are inappropriate at high pressures. In an attempt to solve this problem atom charges were multiplied to avoid full re-parametrization, resulting in accurate values for a multiplier of 1.01 at 60 °C and 150 bar, showing promising results that require testing for other temperatures and pressures.

Structures of quercetin in ethanol at 30 °C and 1 bar were studied through angle, radial and spatial distribution functions. They show that quercetin is a polar molecule, confirmed by its affinity towards ethanol hydroxyl group, and there was preference between the oxygens of quercetin in the following order O18>O22>O17>>O19>O2, and O21>O7 showing almost close to no specific interaction with the solvent hydroxyl groups.

It is possible to obtain significant information of the molecular behavior from the MD simulations, including accurate predictions of D_{12} and on the relevant interactions between the constituting particles.

Suggestions for future work

For future work it is suggested to finish the simulations for quercetin in ethanol for all planned conditions by testing adjusted atom charges for higher pressures. The densities should be calculated from longer NPT ensemble equilibration simulations to ensure that the values of density are accurate enough to guarantee the target pressure in NVT ensemble, if possible by increasing the number of solvent particles in the system, since this may help with pressure control as it reduces pressure fluctuations in both ensembles with the square root of the number of particles in the system. For ethyl acetate, finding or developing new force field parameters optimized for diffusion coefficients is advised. Further study of the functional groups of oxygens O19, O20 and O21 and their interactions could also be performed. And lastly testing other solvents or other flavonoids (starting with other flavonols) to enrich the structural analysis. Including distinct functional groups could be a good idea to help understand how they affect diffusivity for this family of molecules with interesting biological effects.

References

[1] J. Robak, R.J. Gryglewski, Bioactivity of flavonoids., Pol. J. Pharmacol. 48 (1996) 555–564.

[2] A.N. Panche, A.D. Diwan, S.R. Chandra, Flavonoids: an overview, J. Nutr. Sci. 5 (2016) e47–e47.

[3] T. Wang, Q. Li, K. Bi, Bioactive flavonoids in medicinal plants: Structure, activity and biological fate, Asian J. Pharm. Sci. 13 (2018) 12–23.

[4] S.G. Dmitrienko, V.A. Kudrinskaya, V. V Apyari, Methods of extraction, preconcentration, and determination of quercetin, J. Anal. Chem. 67 (2012) 299–311.

[5] ACD/ChemSketch, version 2018.2.5, (2019). www.acdlabs.com.

[6] Avogadro: an open-source molecular builder and visualization tool, version 1.2.0, (n.d.).

[7] M.D. Hanwell, D.E. Curtis, D.C. Lonie, T. Vandermeersch, E. Zurek, G.R. Hutchison, Avogadro: an advanced semantic chemical editor, visualization, and analysis platform, J. Cheminform. 4 (2012) 17.

[8] A. Wach, K. Pyrzyńska, M. Biesaga, Quercetin content in some food and herbal samples, Food Chem. 100 (2007) 699–704.

[9] P. Rameshthangam, J.P. Chitra, Synergistic anticancer effect of green synthesized nickel nanoparticles and quercetin extracted from Ocimum sanctum leaf extract, J. Mater. Sci. Technol. 34 (2018) 508–522.

[10] C. Capello, U. Fischer, K. Hungerbühler, What is a green solvent? A comprehensive framework for the environmental assessment of solvents, Green Chem. 9 (2007) 927–934.

[11] World Health Organization - areas of work - chemical risks, (n.d.). https://www.who.int/foodsafety/areas_work/chemical-risks/jecfa/en/ (accessed September 6, 2019).

[12] Evaluations of the Joint FAO/WHO Expert Committee on Food Additives (JECFA) - Ethyl Acetate, (n.d.). http://apps.who.int/food-additives-contaminants-jecfa-database/chemical.aspx?chemID=3075 (accessed September 6, 2019).

[13] Evaluations of the Joint FAO/WHO Expert Committee on Food Additives (JECFA) - Ethanol, (n.d.). http://apps.who.int/food-additives-contaminants-jecfa-database/chemical.aspx?chemID=2404 (accessed September 6, 2019).

[14] R. Taylor, R. Krishna, Multicomponent Mass Transfer, Wiley, 1993. https://books.google.pt/books?id=FrpX9dNjWc0C.

[15] G.I. Taylor, Diffusion and mass transport in tubes, Proc. Phys. Soc. Sect. B. 67 (1954) 857–869.

[16] G.I. Taylor, The dispersion of matter in turbulent flow through a pipe, Proc. R. Soc. London. Ser. A. Math. Phys. Sci. 223 (1954) 446–468.

[17] R. Aris, On the dispersion of a solute by diffusion, convection, and exchange between phases, in: Process Syst. Eng., The Royal Society, 1999: pp. 121–135.

[18] D.C. Rapaport, The Art of Molecular Dynamics Simulation, 2nd ed., Cambridge University Press, 2004.

[19] Lindahl, Abraham, Hess, van der Spoel, GROMACS 2019.3 Documentation, (2019).

[20] Lindahl, Abraham, Hess, van der Spoel, GROMACS Source Code, (2019).

[21] W. Humphrey, A. Dalke, K. Schulten, {VMD} -- {V}isual {M}olecular {D}ynamics, J. Mol. Graph. 14 (1996) 33–38.

[22] J.D. Seader, W.D. Seider, D.R. Lewin, L. Boulle, A. Rycrof, Separation Process Principles, Chemical and biochemical separations, Third, Wiley, USA, 2006.

[23] K.K. Liong, P.A. Wells, N.R. Foster, Diffusion in supercritical fluids, J. Supercrit. Fluids. 4 (1991) 91–108.

[24] T. Funazukuri, C.Y. Kong, S. Kagei, Impulse response techniques to measure binary diffusion coefficients under supercritical conditions, J. Chromatogr. A. 1037 (2004) 411–429.

[25] G.I. Taylor, Dispersion of soluble matter in solvent flowing slowly through a tube, Proc. R. Soc. London. Ser. A. Math. Phys. Sci. 219 (1953) 186–203.

[26] J.C. Giddings, S.L. Seager, Rapid Determination of Gaseous Diffusion Coefficients by Means of Gas Chromatography Apparatus, J. Chem. Phys. 33 (1960) 1579–1580.

[27] Z. Balenovic, M.N. Myers, J. Calvin Giddings, Binary diffusion in dense gases to 1360 atm by the chromatographic peak-broadening method, J. Chem. Phys. 52 (1970) 915–922.

[28] A.C. Ouano, Diffusion in Liquid Systems. I. A Simple and Fast Method of Measuring Diffusion Constants, Ind. Eng. Chem. Fundam. 11 (1972) 268–271.

[29] R. Feist, G.M. Schneider, Determination of Binary Diffusion Coefficients of Benzene, Phenol, Naphthalene and Caffeine in Supercritical C02between 308 and 333 K in the Pressure Range 80 to 160 Bar with Supercritical Fluid Chromatography (SFC), Sep. Sci. Technol. 17 (1982) 261–270.

[30] B. Zêzere, J. Cordeiro, J. Leite, A.L. Magalhães, I. Portugal, C.M. Silva, Diffusivities

of metal acetylacetonates in liquid ethanol and comparison with the transport behavior in supercritical systems, J. Supercrit. Fluids. 143 (2019) 259–267.

[31] R. V Vaz, A.L. Magalhães, A.A. Valente, C.M. Silva, Measurement and modeling of tracer diffusivities of α-pinene in supercritical CO2, and analysis of their hydrodynamic and free-volume behaviors, J. Supercrit. Fluids. 107 (2016) 690–698.

[32] J. Cordeiro, A.L. Magalhães, A.A. Valente, C.M. Silva, Experimental and theoretical analysis of the diffusion behavior of chromium(III) acetylacetonate in supercritical CO2, J. Supercrit. Fluids. 118 (2016) 153–162.

[33] C.Y. Kong, T. Funazukuri, S. Kagei, Chromatographic impulse response technique with curve fitting to measure binary diffusion coefficients and retention factors using polymer-coated capillary columns, J. Chromatogr. A. 1035 (2004) 177–193.

[34] C.M. Silva, Coeficientes de Difusão em Misturas Supercríticas, Tese Doutor. Em Eng. Química. (1998).

[35] R. Aris, R - *Chemical Engineering Greetings, in: R.B.T.-P.S.E. Aris (Ed.), Math. Model., Academic Press, 1999: pp. 415–416.

[36] R.J. Nunge, T.-S. Lin, W.N. Gill, Laminar dispersion in curved tubes and channels, J. Fluid Mech. 51 (1972) 363–383.

[37] J.M.H. Levelt Sengers, U.K. Deiters, U. Klask, P. Swidersky, G.M. Schneider, Application of the Taylor dispersion method in supercritical fluids, Int. J. Thermophys. 14 (1993) 893–922.

[38] J.A. Moulijn, R. Spijker, J.F.M. Kolk, Axial dispersion of gases flowing through coiled columns, J. Chromatogr. A. 142 (1977) 155–166.

[39] A. Alizadeh, C.A. Nieto de Castro, W.A. Wakeham, The theory of the Taylor dispersion technique for liquid diffusivity measurements, Int. J. Thermophys. 1 (1980) 243–284.

[40] S.R. Springston, M. Novotny, Coiling-induced secondary flow in capillary supercritical fluid chromatography, Anal. Chem. 58 (1986) 2699–2704.

[41] C.Y. Kong, T. Funazukuri, S. Kagei, G. Wang, F. Lu, T. Sako, Applications of the chromatographic impulse response method in supercritical fluid chromatography, J. Chromatogr. A. 1250 (2012) 141–156.

[42] E.T. van der Laan, Notes on the diffusion-type model for longitudinal mixing in flow, Chem. Eng. Sci. 7 (1958) 187–191.

[43] O. Levenspiel, W.K. Smith, Notes on the diffusion-type model for the longitudinal mixing of fluids in flow, Chem. Eng. Sci. 50 (1995) 3891–3896.

[44] J.C. Giddings, S.L. Seager, Method for the Rapid Determination of Diffusion

Coefficients. Theory and Application, Ind. Eng. Chem. Fundam. 1 (1962) 277–283.

[45] W.A. Wakeham, A. Nagashima, J. V Sengers, International Union of Pure and Applied Chemistry. Commission on Thermodynamics., Measurement of the transport properties of fluids - Experimental Thermodynamics Volume III, Blackwell Scientific Publications ; Distributed in the USA and North America by CRC Press, Oxford [England]; Boston; Boca Raton, Fla., 1991.

[46] T. Funazukuri, C.Y. Kong, S. Kagei, Infinite-dilution binary diffusion coefficients of 2-propanone, 2-butanone, 2-pentanone, and 3-pentanone in CO_2 by the Taylor dispersion technique from 308.15 to 328.15 K in the pressure range from 8 to 35 MPa, Int. J. Thermophys. 21 (2000) 1279–1290.

[47] R. Lin, L.L. Tavlarides, Diffusion coefficients of diesel fuel and surrogate compounds in supercritical carbon dioxide, J. Supercrit. Fluids. 52 (2010) 47–55.

[48] T. Funazukuri, C.Y. Kong, S. Kagei, Measurements of binary diffusion coefficients for some low volatile compounds in supercritical carbon dioxide by input–output response technique with two diffusion columns connected in series, Fluid Phase Equilib. 194–197 (2002) 1169–1178.

[49] R. V Vaz, J.R.B. Gomes, C.M. Silva, Molecular dynamics simulation of diffusion coefficients and structural properties of ketones in supercritical CO_2 at infinite dilution, J. Supercrit. Fluids. 107 (2016) 630–638.

[50] A.R. Leach, Molecular modelling: principles and applications, Pearson education, 2001.

[51] M. Fermeglia, S. Pricl, Equation-of-state parameters for pure polymers by molecular dynamics simulations, AIChE J. 45 (1999) 2619–2627.

[52] H. Feng, W. Gao, Z. Sun, B. Lei, G. Li, L. Chen, Molecular Dynamics Simulation of Diffusion and Structure of Some n-Alkanes in near Critical and Supercritical Carbon Dioxide at Infinite Dilution, J. Phys. Chem. B. 117 (2013) 12525–12534.

[53] M.P. Allen, Introduction to Molecular Dynamics Simulation, Comput. Soft Matter From Synth. Polym. to Proteins. 23 (2004) 1–28.

[54] J.J. E., C. Sydney, On the determination of molecular fields. —II. From the equation of state of a gas, Proc. R. Soc. London. Ser. A, Contain. Pap. a Math. Phys. Character. 106 (1924) 463–477.

[55] O. Lenz, Graph of the 12-6-Lennard-Jones Potential, (2007). https://en.wikipedia.org/wiki/File:12-6-Lennard-Jones-Potential.svg (accessed August 31, 2019).

[56] U. Essmann, L. Perera, M.L. Berkowitz, T. Darden, H. Lee, L.G. Pedersen, A

smooth particle mesh Ewald method, J. Chem. Phys. 103 (1995) 8577–8593.

[57] B. Hess, H. Bekker, H.J.C. Berendsen, J.G.E.M. Fraaije, LINCS: A linear constraint solver for molecular simulations, J. Comput. Chem. 18 (1997) 1463–1472.

[58] J.W. Ponder, D.A. Case, Force fields for protein simulations, Adv. Protein Chem. 66 (2003) 27–85.

[59] W.L. Jorgensen, D.S. Maxwell, J. Tirado-Rives, Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids, J. Am. Chem. Soc. 118 (1996) 11225–11236.

[60] P.E.M. Lopes, O. Guvench, A.D. MacKerell Jr, Current status of protein force fields for molecular dynamics simulations, Methods Mol. Biol. 1215 (2015) 47–71.

[61] W.F. Van Gunsteren, S.R. Billeter, A.A. Eising, P.H. Hunenberger, P. Kruger, A.E. Mark, W.R.P. Scott, I.G. Tironi, The GROMOS96 manual and user guide, Biomos, Zurich, Switz. (1996).

[62] M. Christen, P.H. Hünenberger, D. Bakowies, R. Baron, R. Bürgi, D.P. Geerke, T.N. Heinz, M.A. Kastenholz, V. Kräutler, C. Oostenbrink, C. Peter, D. Trzesniak, W.F. van Gunsteren, The GROMOS software for biomolecular simulation: GROMOS05, J. Comput. Chem. 26 (2005) 1719–1751.

[63] M.P. Allen, D.J. Tildesley, Computer simulation of liquids: Second edition, 2017.

[64] J.E. Basconi, M.R. Shirts, Effects of Temperature Control Algorithms on Transport Properties and Kinetics in Molecular Dynamics Simulations, J. Chem. Theory

[65] Nzjacobmartin, Visual representation of five statistical ensembles., (2017). https://commons.wikimedia.org/wiki/File:Statistical_Ensembles.png (accessed September 5, 2019).

[66] J.M. Haile, Molecular Dynamics Simulation: Elementary Methods, 1st ed., John Wiley & Sons, Inc., New York, NY, USA, 1992.

[67] D. Keffer, The Working Man's Guide to Obtaining Self Diffusion Coefficients from Molecular Dynamics Simulations, Knoxville, 2002.

[68] J.H. Dymond, R. Malhotra, The Tait equation: 100 years on, Int. J. Thermophys. 9 (1988) 941–951.

[69] J.J. Cano-Gómez, G.A. Iglesias-Silva, V. Rico-Ramírez, M. Ramos-Estrada, K.R. Hall, A new correlation for the prediction of refractive index and liquid densities of 1-alcohols, Fluid Phase Equilib. 387 (2015) 117–120.

[70] J.J. Cano-Gómez, G.A. Iglesias-Silva, M. Ramos-Estrada, Correlations for the prediction of the density and viscosity of 1-alcohols at high pressures, Fluid Phase Equilib. 404 (2015) 109–117.

[71] A. Kumagai, H. Iwasaki, Pressure-volume-Temperature Relationships of ethyl acetate and generalized Tait equation for liquids at high pressures, J. Chem. Eng. Data. 24 (1979) 261–263.

[72] C.L. Yaws, X. Lin, L. Bu, D.R. Balundgi, S. Tripathi, Chemical Properties Handbook: Physical, Thermodynamic, Environmental, Transport, Safety, and Health Related Properties for Organic and Inorganic Chemical, McGraw-Hill, Beaumont, Texas, 1999.

[73] J. Leite, A.L. Magalhães, A.A. Valente, C.M. Silva, Measurement and modelling of tracer diffusivities of gallic acid in liquid ethanol and in supercritical CO2 modified with ethanol, J. Supercrit. Fluids. 131 (2018) 130–139.

[74] I.G. Zenkevich, A.Y. Eshchenko, S. V. Makarova, A.G. Vitenberg, Y.G. Dobryakov, V.A. Utsal, Identification of the products of oxidation of quercetin by air oxygen at ambient temperature, Molecules. 12 (2007) 654–672.

[75] M. Mocek, P.J. Richardson, KINETICS AND MECHANISM OF QUERCETIN OXIDATION, J. Inst. Brew. 78 (1972) 459–465.

[76] I.C. Yeh, G. Hummer, System-size dependence of diffusion coefficients and viscosities from molecular dynamics simulations with periodic boundary conditions, J. Phys. Chem. B. 108 (2004) 15873–15879.

[77] G. Guevara-Carrion, S. Ancherbak, A. Mialdun, J. Vrabec, V. Shevtsova, Diffusion of methane in supercritical carbon dioxide across the Widom line, Sci. Rep. 9 (2019).

[78] S. Nosé, A molecular dynamics method for simulations in the canonical ensemble, Mol. Phys. 52 (1984) 255–268.

[79] W.G. Hoover, Canonical dynamics: Equilibrium phase-space distributions, Phys. Rev. A. 31 (1985) 1695–1697.

[80] R.W. Hockney, S.P. Goel, J.W. Eastwood, Quiet high-resolution computer models of a plasma, J. Comput. Phys. 14 (1974) 148–158.

[81] M. Parrinello, A. Rahman, Polymorphic transitions in single crystals: A new molecular dynamics method, J. Appl. Phys. 52 (1981) 7182–7190.

[82] S. Nosé, M.L. Klein, Constant pressure molecular dynamics for molecular systems, Mol. Phys. 50 (1983) 1055–1076.

[83] G. Kaminski, W.L. Jorgensen, Performance of the AMBER94, MMFF94, and OPLS-AA Force Fields for Modeling Organic Liquids, J. Phys. Chem. 100 (1996) 18010–18013.

[84] W.L. Jorgensen, Optimized intermolecular potential functions for liquid alcohols, J.

Phys. Chem. 90 (1986) 1276–1284.

[85] M.L.P. Price, D. Ostrovsky, W.L. Jorgensen, Gas-phase and liquid-state properties of esters, nitriles, and nitro compounds with the OPLS-AA force field, J. Comput. Chem. 22 (2001) 1340–1352.

[86] K. Pluhackova, H. Morhenn, L. Lautner, W. Lohstroh, K.S. Nemkovski, T. Unruh, R.A. Böckmann, Extension of the LOPLS-AA Force Field for Alcohols, Esters, and Monoolein Bilayers and its Validation by Neutron Scattering Experiments, J. Phys. Chem. B. 119 (2015) 15287–15299.

[87] W.L. Jorgensen, J. Tirado-Rives, Potential energy functions for atomic-level simulations of water and organic and biomolecular systems, Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 6665–6670.

[88] L.S. Dodda, J.Z. Vilseck, J. Tirado-Rives, W.L. Jorgensen, 1.14*CM1A-LBCC: Localized Bond-Charge Corrected CM1A Charges for Condensed-Phase Simulations, J. Phys. Chem. B. 121 (2017) 3864–3870.

[89] L.S. Dodda, I. Cabeza de Vaca, J. Tirado-Rives, W.L. Jorgensen, LigParGen web server: an automatic OPLS-AA parameter generator for organic ligands, Nucleic Acids Res. 45 (2017) W331–W336.

[90] J.A. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; R, Gaussian 03, Revision C.01, (n.d.).

[91] M. Brehm, B. Kirchner, TRAVIS - A Free Analyzer and Visualizer for Monte Carlo and Molecular Dynamics Trajectories, J. Chem. Inf. Model. 51 (2011) 2007–2023.

[92] H. Kato, T. Saito, M. Nabeshima, K. Shimada, S. Kinugasa, Assessment of diffusion coefficients of general solvents by PFG-NMR: Investigation of the sources error, J. Magn. Reson. 180 (2006) 266–273.

[93] T. Uminski, J. Dera, G. Kupryszewski, The measurement of self-diffusion coefficient in dielec-tric liquids using the method of open-ended capillaries and radioisotope technique, Acta Phys. Pol. 28 (1965) 17.

[94] B. Zêzere, A.L. Magalhães, I. Portugal, C.M. Silva, Diffusion coefficients of eucalyptol at infinite dilution in compressed liquid ethanol and in supercritical CO2/ethanol mixtures, J. Supercrit. Fluids. 133 (2018) 297–308.

[95] B. Zêzere, J.M. Silva, I. Portugal, J.R.B. Gomes, C.M. Silva, Measurement of astaxanthin and squalene diffusivities in compressed liquid ethyl acetate by Taylor-Aris dispersion method, Sep. Purif. Technol. 234 (2020) 116046.

[96] O. Suárez-Iglesias, I. Medina, C. Pizarro, J.L. Bueno, Diffusion of Benzyl Acetate, 2-Phenylethyl Acetate, 3-Phenylpropyl Acetate, and Dibenzyl Ether in Mixtures of Carbon Dioxide and Ethanol, Ind. Eng. Chem. Res. 46 (2007) 3810–3819.

[97] C.M. Silva, H. Liu, Modelling of Transport Properties of Hard Sphere Fluids and Related Systems, and its Applications BT - Theory and Simulation of Hard-Sphere Fluids and Related Systems, in: Á. Mulero (Ed.), Springer Berlin Heidelberg, Berlin, Heidelberg, 2008: pp. 383–492.

[98] A. Chafer, T. Fornari, A. Berna, R.P. Stateva, Solubility of quercetin in supercritical CO2 + ethanol as a modifier: measurements and thermodynamic modelling, J. Supercrit. Fluids. 32 (2004) 89–96.

[99] C. Reichardt, Solvents and Solvent Effects in Organic Chemistry. Wiley-VCH, Weinheim/Germany, 2003, 2003.

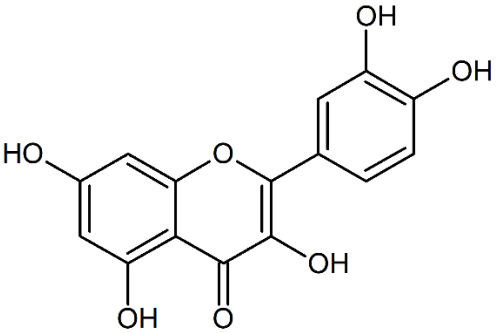
[100] C. Hölzl, P. Kibies, S. Imoto, R. Frach, S. Suladze, R. Winter, D. Marx, D. Horinek, S.M. Kast, Design principles for high-pressure force fields: Aqueous TMAO solutions from ambient to kilobar pressures, J. Chem. Phys. 144 (2016) 144104.

[101] L.A. De Souza, W.M.G. Tavares, A.P.M. Lopes, M.M. Soeiro, W.B. De Almeida, Structural analysis of flavonoids in solution through DFT 1H NMR chemical shift calculations: Epigallocatechin, Kaempferol and Quercetin, Chem. Phys. Lett. 676 (2017) 46–52.

[102] O. Brovarets, D. Hovorun, Conformational transitions of the quercetin molecule via the rotations of its rings: A comprehensive theoretical study, J. Biomol. Struct. Dyn. (2019) 1–19.

Appendix A – Compounds and Software used

Compounds



CAS: 117-39-5

Chemical formula: $C_{15}H_{10}O_7$

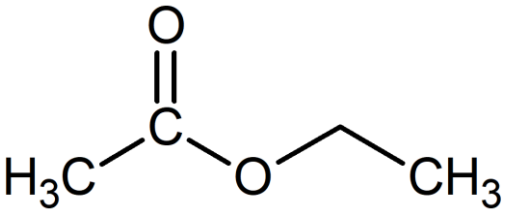
Supplier: Sigma-Aldrich

Molecular weight: 302.24 $g \cdot mol^{-1}$

Purity: ≥ 95 %

Figure A. 3 – Quercetin skeletal molecular formula

Product number: Q4951



CAS: 141-78-6

Chemical formula: $CH_3COOC_2H_5$

Supplier: VWR Chemicals

Molecular weight: 88.11 $g \cdot mol^{-1}$

Purity: ≥ 99.5 %

Figure A. 4 – Ethyl acetate skeletal molecular formula

Product number: 23882

Software

Molecular dynamics simulation software – GROMACS (GROningen MAchine for Chemical Simulations) [19,20]

Molecule editor software – Avogadro [6,7]

Molecular visualization software – VMD (Visual Molecular Dynamics) [21]

2D structure drawing software – ACD/ChemSketch [5]

Trajectory analyzer – GROMACS [19,20] and Travis (Trajectory Analyzer and Visualizer) [91]

Appendix B – Wavelength study for ethyl acetate

The tracer injection response curves were measured between the range of 250 and 400 nm to find the wavelength with minimum experimental noise and error for quercetin in ethyl acetate. After recording several response curves at temperature of 50 °C and pressure of 1 bar, the root mean square errors ε were plotted against wavelength λ in Figure B. 1. This led to the conclusion that 270 nm leads to the least error. On Figure B.2 and Figure B.3 were also plotted the ratio of maximum absorbance over the area of the peak, $maxAbs/AreaPeak$, *versus* the wavelength λ and the diffusion coefficient, D_{12}, *versus* the wavelength λ, to ensure the linearity of the systems and that variations in wavelength didn't cause significant variations in the diffusion coefficient.

The concentration of the solute is also an important aspect of the experimental procedure to guarantee approximate infinite dilution of the solute, so different concentrations were tested in the figures above to confirm that no variation on the diffusion coefficient on the selected concentration of 0.63 mg/mL. Only 0.1 μL of quercetin in ethyl acetate was injected per point.

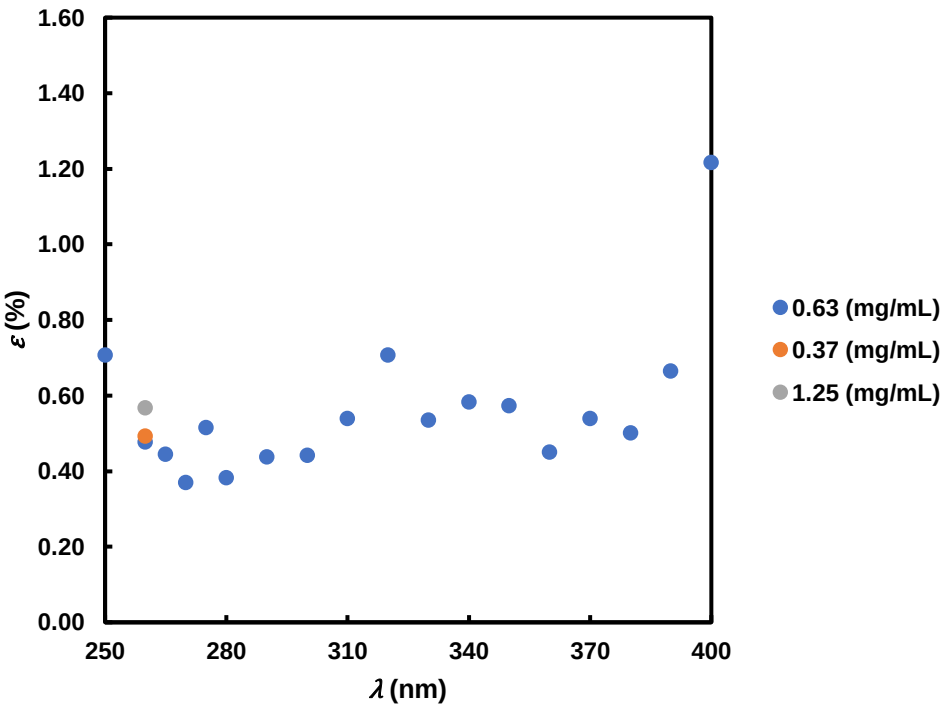


Figure B. 1 – Root mean square error, ε, *versus* wavelength, λ, from response curves of quercetin in ethyl acetate at different concentrations, temperature of 50 °C and pressure 1 bar.

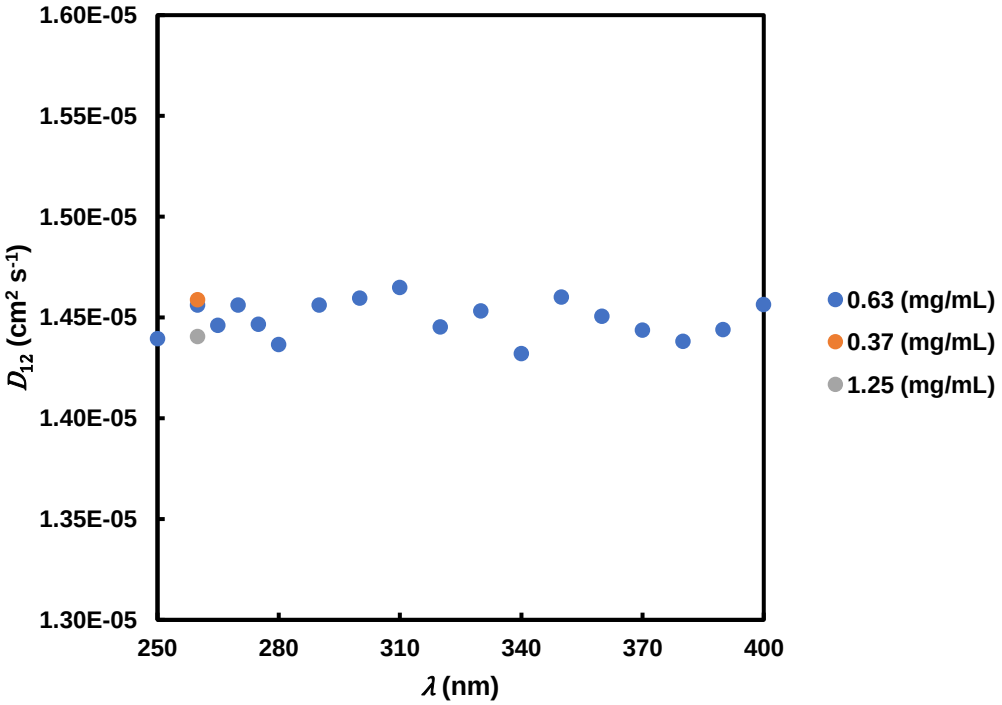


Figure B. 2 – Diffusion coefficient, D_{12}, *versus* wavelength, λ, from response curves of quercetin in ethyl acetate at different concentrations, temperature of 50 °C and pressure 1 bar.

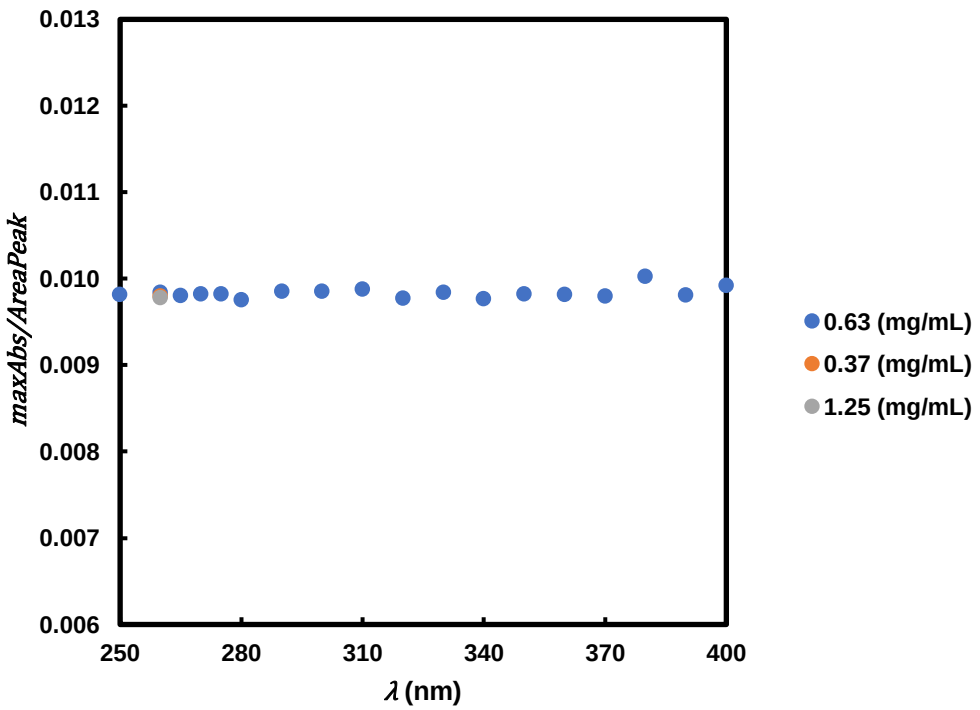


Figure B. 3 – Ratio of maximum absorbance over the area of the peak, $maxAbs/AreaPeak$, *versus* wavelength, λ, from response curves of quercetin in ethyl acetate at different concentrations, temperature of 50 °C and pressure 1 bar.