Université de Lille – Faculté des Sciences et Technologies

École Doctorale

Sciences de la Matière, du Rayonnement et de l'Environnement

Thèse pour obtenir le titre de

Docteur de l'Université de Lille – Faculté des Sciences et Technologies

Chimie Organique, minérale, industrielle

Lucie DELFORCE

New Methodologies for Characterizing Particles, Complex Oils and Surfactants: Relations Between Chemical Structure, Physicochemical Properties and Applicative Properties

---

Nouvelles méthodologies pour la caractérisation des particules, des huiles complexes et des tensioactifs : relations entre structures chimiques, propriétés physicochimiques et propriétés applicatives

Thèse dirigée par :

Pr Jean-Marie AUBRY, Directeur

Pr Véronique NARDELLO-RATAJ, Co-Directrice

Dr Jesús Fermín ONTIVEROS, Co-Encadrant

Soutenue le 18 Novembre 2022

Devant le jury composé de :

Pr Patrick SAULNIER	Université d'Angers	Rapporteur, Président du jury
Pr Cécile PAGNOUX	Université de Limoges	Rapportrice
Dr François BOUTON	Brenntag	Examinateur
Pr Adlane SAYEDE	Université d'Artois	Membre Invité
Pr Jean-Marie AUBRY	Centrale Lille	Directeur
Pr Véronique NARDELLO-RATAJ	Centrale Lille	Co-Directrice
Dr Jesús Fermín ONTIVEROS	Centrale Lille	Co-Encadrant

Acknowledgements

Résumé

Les objectifs de la thèse sont multiples et s'inscrivent dans le cadre de la mise en place d'une nouvelle plateforme technologique HT-SMART-FORMU dédiée à la formulation. L'accent est mis sur le développement de méthodes expérimentales fiables et d'outils théoriques et prédictifs, afin d'établir des relations entre la composition chimique, les propriétés physicochimiques et les propriétés applicatives.

Les dispersions de particules sont le premier type de systèmes étudiés. Leur est traditionnellement rationnalisée à l'aide de la théorie DLVO, mais des études récentes proposent l'utilisation des Paramètres de Solubilité de Hansen pour décrire leur stabilité en milieu non aqueux. Dans le premier chapitre, une méthode analytique systématique basée sur la mesure du potentiel zêta et le suivi de la stabilité par diffusion de la lumière permet de déduire une complémentarité des deux théories pour décrire la stabilité des dispersions de TiO_2 en solvants non aqueux.

La problématique de la prédiction de l'hydrophobie des huiles par la modélisation d'EACN, qui est un paramètre clé pour concevoir des systèmes dispersés tensioactif/huile/eau (SOW), est abordée dans le deuxième chapitre. Deux modèles mathématiques, conçus à l'aide de méthodes de machine-learning, sont proposés pour la prédiction rapide de l'EACN des huiles, à savoir les réseaux de neurones (NN) et les machines graphiques (GM). Alors que le modèle GM est implémenté à partir des codes SMILES, le modèle NN est alimenté par des descripteurs σ-moments calculés avec le logiciel COSMOtherm. La fiabilité des prédictions des deux modèles est discutée sur la base d'un ensemble de test de 10 molécules complexes.

Dans les chapitres 3 et 4, l'étendue des propriétés applicatives d'un tensioactif non ionique à base de glycérol ($C_{12}Gly_2$) est étudiée. Le chapitre 3 se porte sur son agrégation en solution aqueuse et la formation de cristaux liquides (CL) à faible concentration, en comparaison avec les alcools gras polyéthoxylés et les alkyl polyglucosides. L'influence de ses propriétés physico-chimiques, en particulier la viscoélasticité de dilatation de l'interface air/eau, est mise en relation avec la faible moussabilité et la stabilité de mousse durable observées.

Enfin, les propriétés du $C_{12}Gly_2$ en systèmes SOW sont étudiées dans le chapitre 4. La déviation hydrophile-lipophile normalisé (HLD_N), un outil théorique puissant, est considéré comme un moyen de rationaliser les caractéristiques des émulsions et des microémulsions. Ainsi, une quantification approfondie de l'amphiphilie des tensioactifs, de leur sensibilité à la température et de leur tolérance au sel est présentée. L'utilisation du $C_{12}Gly_2$ comme émulsifiant H/E et E/H est ensuite étudiée : la granulométrie et la stabilité des émulsions obtenues en faisant varier l'huile concordent avec les valeurs du HLD_N. Un minimum est observé à $HLD_N = 0$, puis la granulométrie et la stabilité augmentent pour des valeurs de HLD_N négatives et positives.

Mots-clés : Particules ; Huiles ; Tensioactifs ; Méthodes Prédictives ; Mousse ; HLD

Abstract

The multiple aims of this thesis fall within the implementation of a new HT-SMART-FORMU technological platform dedicated to formulation science. The focus is set on the development of reliable experimental methods, theoretical and predictive tools, so as to establish relations between chemical composition, physicochemical properties and applicative properties.

The first type of systems investigated in this work consists of solid dispersions. Their stability has long been approached by the DLVO theory, but more recent studies suggest the use of Hansen Solubility Parameters to describe their stability in non-aqueous media. In the first chapter, a systematic analytical method based on zeta potential measurement and light scattering stability monitoring allows deducing a complementarity of both theories to describe the stability of TiO_2 dispersions in non-aqueous solvents.

The problematic of oil hydrophobicity prediction through EACN values modelling, which is a key parameter to design surfactant/oil/water (SOW) dispersed systems, is addressed in the second chapter. Two mathematical models, designed using machine-learning methods, are proposed for the rapid prediction of the EACN of oils, namely Neural Networks (NN) and Graph Machines (GM). While the GM model is implemented from the SMILES codes, the NN model is fed with σ-moments descriptors computed with the COSMOtherm software. The prediction reliability of both models is discussed based on a complex 10-molecule test set.

In chapters 3 and 4, the scope of applicative properties of a nonionic glycerol-based surfactant are investigated. Firstly, chapter 3 focuses on its aggregation behaviour in aqueous solutions and the formation of liquid crystals (LC) at low concentration, in comparison with the benchmark polyethoxylated fatty alcohols and alkyl polyglucosides. The influence of its physicochemical properties, in particular the air / water interface dilational viscoelasticity, is put in relation with the observed poor foamability and long-lasting foam stability.

Finally, $C_{12}Gly_2$ properties in SOW systems are investigated in chapter 4. The Normalized Hydrophilic-Lipophilic Deviation (HLD_N), a powerful theoretical tool, is regarded as a way to rationalize the characteristics of both emulsions and microemulsions. In this way, a thorough quantification of surfactants amphiphilicity, temperature sensitivity and salt-tolerance are presented. The potential use of $C_{12}Gly_2$ as O/W and W/O emulsifier is then investigated: the granulometry and stability of emulsions obtained by varying the oil concurs with HLD_N values. A minimum is observed at $HLD_N = 0$ and increases for negative and positive HLD_N values.

Keywords: Particles; Oils; Surfactants; Predictive methods; Foam; HLD

Table of contents

General Introduction

From our morning routine with the cosmetics we use and the clothe we wear, to the processed food we eat at lunch, or even the components of the car we drive to go to work, formulated products fulfil practical functions in our daily lives. Other common examples include medicines, perfumes, home care products, detergents, paints and coatings, inks, glues and construction materials. Components of such products are carefully chosen to confer precise applicative properties to the final composition, within a strict regulatory framework and in a costly manner. However, most of the time, the formulated product consists of a fine dispersion of non-miscible phases: macroscopically homogeneous, but microscopically heterogeneous, and thermodynamically unstable. This is the case for paintings where pigments tend to sediment, cosmetic creams or even mayonnaise where oil tends to separate from water over time. Solid particles, non-miscible liquids or gas dispersed in a continuous liquid phase result in the formation of solid dispersions, emulsions, and foams respectively. This multiphasic nature and the formation of an interface between both phases is thermodynamically unfavoured, implying preparation and stability problematics for designing effective products.

Nowadays, the fast evolution of legislations and market demands induces the replacement of some petroleum-based ingredients with new performant, bio-based or harmless ones. One typical illustration is the "made without" trend in cosmetics, causing the replacement of a number of conservatives, e.g. parabens, and active materials like sulphate surfactants in shampoos or alumina salts in deodorants. In order to accompany and ease this transition, formulators should rely on optimized systematic physicochemical approaches. Understanding the intermolecular interactions and how this relates to the material properties is the key to unlocking future innovations and designing functional products. The works presented in this thesis fall within the implementation of the new HT-SMART-FORMU technological platform, dedicated to bringing innovative solutions for formulation science. The focus is set on the development of reliable experimental methods, theoretical and predictive tools, so as to establish relations between chemical composition, physicochemical properties and applicative properties.
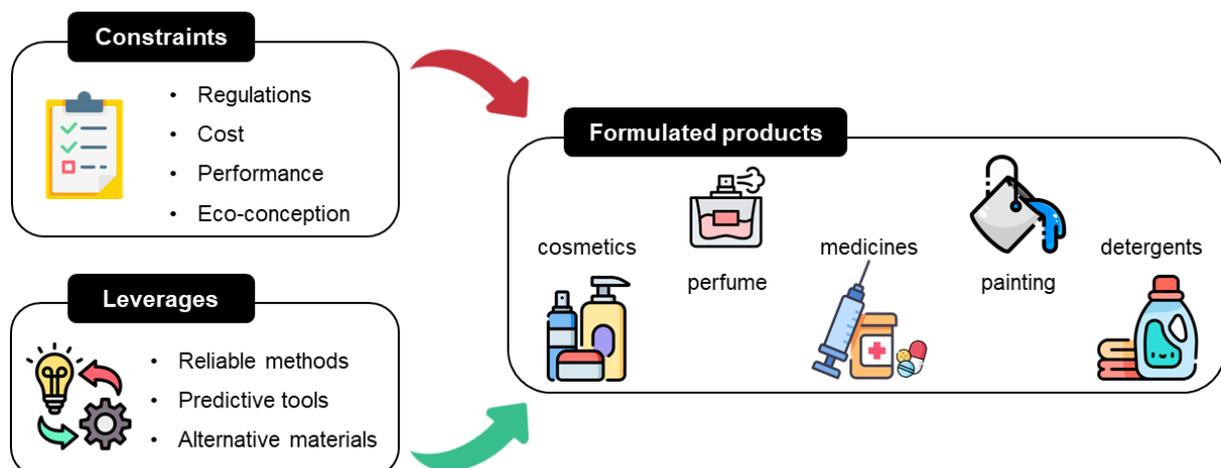
The first type of systems investigated in this work consists of solid dispersions. Their stability has long been approached by the DLVO theory, according to which particles interact with one another through the solvent, acting as a continuous dielectric media. In this framework, the influence of ionic species, electrolytes, acids and bases has been widely studied. However, more recently, a rationalization of particle dispersions stability was proposed based on the Hansen Solubility Parameters approach, considering molecular interactions between solvent molecules and particle surface. Both theories are based on very different hypothesis and one target of this work is thus to determine what theory applies when, and how to discriminate their scope of application. Indeed, those are practical tools that would allow predicting the stability of solid dispersions in various solvent types. In the first chapter, this study is applied to TiO_2 nanoparticles, widely used in the construction sector, cosmetics but also catalytic applications.

When the dispersed phase is a liquid, the system tends to destabilize due to poor affinity between both the dispersed and the continuous phase. To tackle this, the interface is usually stabilized by means of surface-active agents, also known as surfactants. The non-miscibility of a liquid with water can be interpreted in terms of hydrophobicity, literally "lack of affinity for water", that needs to be reliably quantified. Indeed, a variety of liquid compounds can be regrouped under the general term "oils". This class of components comprises vegetable oils, but also petro-sourced hydrocarbons, terpenes and siloxanes, to name a few. Perfume molecules are also included in this category as their aqueous solubility is usually very low.

Historically, the research on enhanced crude oil recovery in the 1970s initiated investigations on oil hydrophobicity and resulted in the EACN classification scale. In a nutshell, it refers to the equivalent length of the linear alkane that would behave similarly to the oil under study in a Surfactant / Oil / Water (SOW) system. Such EACN value being tedious to obtain experimentally, methods based on mathematical data treatment can be implemented for predictive purposes. Based on an experimental dataset, two EACN modelling methods are investigated in the second chapter: neural networks which are non-linear mathematical models, and graph machines that estimate a property directly from topological information. In this way, the hydrophobicity of any molecule could be estimated *in silico*, using either of those models. EACN values of oils can help predict their behaviour in mixture with water and surfactant for the formation of emulsions and microemulsions.

Not only oils are influent on the properties of SOW systems, but the nature of the surfactant and its amphiphilic properties are also tuneable for reaching desired features. As briefly evoked in previous paragraphs, the stabilization of interfaces is one major challenge in formulation science. Surfactants of many types are designed to that aim. Their ability to adsorb at

interfaces is due to their amphiphilic structure, consisting of both a hydrophobic chain of at least 8 carbon atoms, and a polar hydrophilic head. Depending on its nature, they can be classified in four main categories: anionic, cationic, amphoteric and nonionic. This last one represented a global market estimated to be over USD 38 billon in 2021 and expected to increase in the following years. The home care industry dominated the nonionic surfactant application market in 2020, accounting for over 50% of the overall product demand.[1] In volume, the polyethoxylated fatty alcohols are the most widely used type of nonionic surfactants. However, the market demand for bio-sourced molecules and increased performance currently drives the research and innovation in the surfactant market.

In chapters 3 and 4, the scope of applicative properties of a promising nonionic glycerol-based surfactant are investigated. Firstly, chapter 3 focuses on the aggregation behaviour of this 1-O-dodecyl diglyceryl ether ($C_{12}Gly_2$) surfactant in aqueous solutions, in comparison with the benchmark polyethoxylated fatty alcohols and alkyl polyglucosides. The influence of its physicochemical properties, in particular the air / water interface viscoelasticity, is put in relation with its foaming properties. Finally, $C_{12}Gly_2$ properties in SOW systems are further investigated in chapter 4. The Normalized Hydrophilic-Lipophilic Deviation (HLD_N), a powerful theoretical tool, is regarded as a way to rationalize the characteristics of both emulsions and microemulsions. In this way, a thorough quantification of surfactants amphiphilicity, temperature sensitivity and salt-tolerance are presented. The potential use of $C_{12}Gly_2$ as O/W and W/O emulsifier is investigated in this last chapter, by characterizing emulsions obtained with various types of oils.

Chapter 1

❖

Stabilizing interactions in particle dispersions
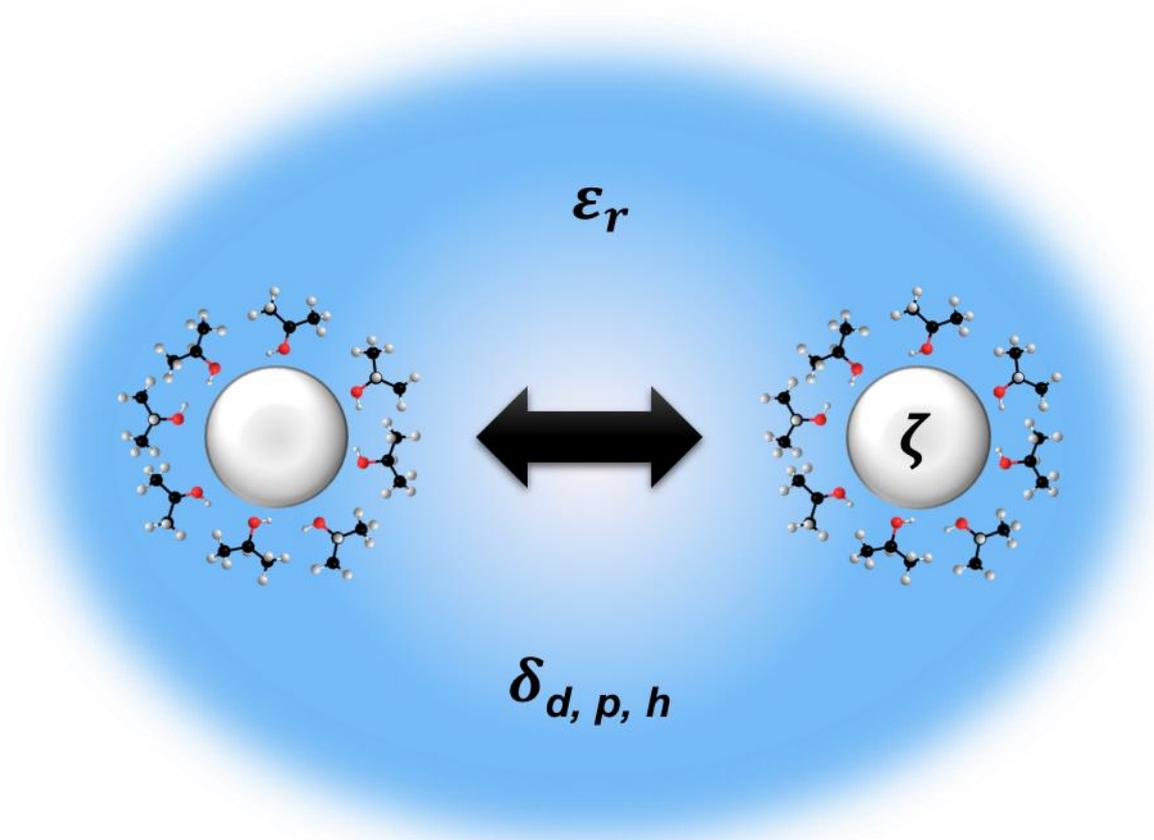
1. Introduction

Nanoscience occupies an important place in contemporary research and industry. Due to their size less than 100 nm,[2] nanoparticles (NPs) have peculiar properties relevant for applications in coatings, cosmetics, pharmaceutics, energy and agriculture, to name a few.[3,4] Titanium dioxide nanoparticles (TiO$_2$ NPs) are among the most widely used NPs, in particular as UV filters in sunscreen cosmetics[5] and plastics to avoid solar degradation,[6] but also as photocatalysts in wastewater treatment,[7] in self-cleaning transparent coatings,[8] in solar cells or as silicon heat-stabilizers[9] and many other applications that require their dispersion in fluid or solid matrixes. The efficiency of UV filters for skin protection is highly dependent on the distribution of this filter in the sunscreen film.[10] When TiO$_2$ NPs are used as catalysts for wastewater treatment, photocatalytic degradation of pollutants increases with NPs dispersion as it is necessary that a large fraction of the catalytic area is accessible.[11] Therefore, it is of major importance that NPs be and remain homogeneously dispersed in the matrix to achieve optimal properties and stability.

Nanoparticle dispersions can be studied by various experimental methods such as DLS (Dynamic Light Scattering) for size measurement[12–14] as well as gravitational[15–18] or centrifugal[14,19,20] sedimentations coupled with multiple light scattering methods which can provide more detailed information regarding the destabilization mechanisms of dispersions. These techniques are based on the measurement of transmitted and backscattered light over time along the whole height of a cell containing the dispersion. Both methods allow to determine hydrodynamic radius using sedimentation rate. However, agglomeration and flocculation of NPs, due to interparticle attraction, are more efficiently observed by gravitational sedimentation. Actually, when a sample is centrifuged, sedimentation is greatly accelerated whereas attractive interactions remain identical to that under gravitational field and, therefore, do not have time to induce agglomeration of the particles.

In this chapter, the evolution over time of aqueous and organic liquid dispersions of uncoated TiO$_2$ nanoparticles was studied using an optical device. The experiments are carried out under gravitational field in order to observe both the agglomeration and flocculation phenomena that occur when formulations are stored under ordinary conditions. In aqueous media, interparticle interactions are well described by the DLVO theory, named after the researchers Derjaguin, Landau, Verwey and Overbeek. Electrostatic repulsions between charged particle surfaces and van der Waals attractions coming from the solid core of particles contribute to the overall particle interactions[21] and become dominant over gravity as particles get smaller.[22,23] However, it is recognized in the literature that organic solvents behave in a more complex manner than water as NP dispersing media because of their diversity in structure and polarity.

Hansen Solubility Parameters (HSP) approach derives from the Hildebrand solubility

parameter which is the square root of cohesive energy density. Hansen's assumption is that this cohesive energy density is due to three types of interactions between molecules: polar (δ_p), dispersive (δ_d) and hydrogen bonding (δ_h) interactions.[24] Originally, Hansen's solubility parameters were developed to study and anticipate the solubilization of molecular and macromolecular compounds in organic solvents. Hansen himself proposed to apply his method to characterize the surface of dispersed particles by arguing that organic liquids which adsorb most strongly to the surfaces of particles are those in which dispersions are most stable.[24,25] It is also argued that for small nanoparticles, energy of solvation can be negative[26] and thus, solvation would be thermodynamically favoured, placing NPs in the frame of Hansen Parameters.[27] However, as the physicochemical phenomena involved in the dispersion of particles are definitely different from those involved in the dissolution of organic compounds, Süß et al. proposed to use the term "Hansen's Dispersion Parameters" (HDP) instead of HSP when Hansen's approach is used to study the dispersibility and stability of particles.[14] HSP has been shown to be a versatile tool for rationalizing and predicting the stability of various types of NP dispersions such as carbon black,[14] carbon nanofibers,[28] fullerene,[29] graphene[30] and carbon nanotubes[31] but also titanium carbides[32] and inorganic nanoparticles of ZnO, Al_2O_3, ZrO_2,[33] hydroxyapatite and TiO_2.[19] This chapter is based on Hansen Parameters as the three complementary parameters allow a practical 3D representation and a clear visualization of effective and non-effective solvents. However, it can be expected that interparticle electrostatic interactions, not considered in Hansen's approach, also play a significant role, especially in organic solvents with notable dielectric constant.

Herein we discuss the respective contributions of DLVO and non-DLVO interactions in the stability of TiO_2 P25 nanoparticle dispersions, with a special emphasis on the relevance of the HSP concept to rationalize non-DLVO interactions in organic solvents. Zeta potential measurements in organic and aqueous media are carried out to identify the solvents in which stability can be explained by the DLVO theory from those for which the stability results from more specific NP-solvent interaction. These latter solvents are used to determine the Hansen sphere of TiO_2 P25 with a Turbiscan as a stability analyser.

2. Bibliographic background
2.1. TiO₂ particles generalities
2.1.1. Global market and industrial production

With a global production volume of almost 6 million tons,[34] the size of the titanium dioxide market was estimated to be USD 20.9 billion in 2021.[35] The driver market for this type of particle is the construction sector. Moreover, the nano-scaled titanium dioxide particles, that are among the most widely used nanoparticles, were estimated to be a USD 3.4 billion market in 2015.[6]

Titania naturally occurs in ilmenite and rutile ores. Ilmenite is a mixture of iron and titanium oxide $FeTiO_3$ whereas rutile is mostly TiO_2 with up to 10% iron oxides in a rutile crystal structure. Those two types of ores undergo different processes to yield pure TiO_2 particles.[34] In the first one, known as the sulphate process and first commercialized in 1931, ilmenite is lixiviated by H_2SO_4 so as to separate soluble iron and titanium complexes. Fe^{3+} ions are reacted with solid Fe to form Fe^{2+}, forming a precipitate with sulphate ions. After filtration of $FeSO_4$ crystals, $TiOSO_4$ hydrolysis regenerates titanium hydrated oxides. Finally, particles are calcined so as to remove water and traces of sulfuric acid and yield TiO_2 in both anatase and rutile crystalline phases.[34,36] In the second one, known as the chloride process, developed and commercialized by DuPont in the 1950s, rutile ore is heated with coke while chlorine is added under gaseous form to yield $TiCl_4$ and other chlorinated metals which are separated by distillation. Indeed, $TiCl_4$ has a lower boiling point than other metal chlorides. Oxidation by O_2 addition then regenerates TiO_2 in the rutile crystalline phase.[34,36] In both cases, the resulting particles are generally milled to produce the desired size. A third type of crystal structure, brookite, can be obtained, but is of little industrial interest. Brookite turns into rutile when heated above 500 °C.[37] The three crystalline phases are represented in Figure 1.1.

| **a.** Tetragonal | **b.** Tetragonal | **c.** Orthorhombic |
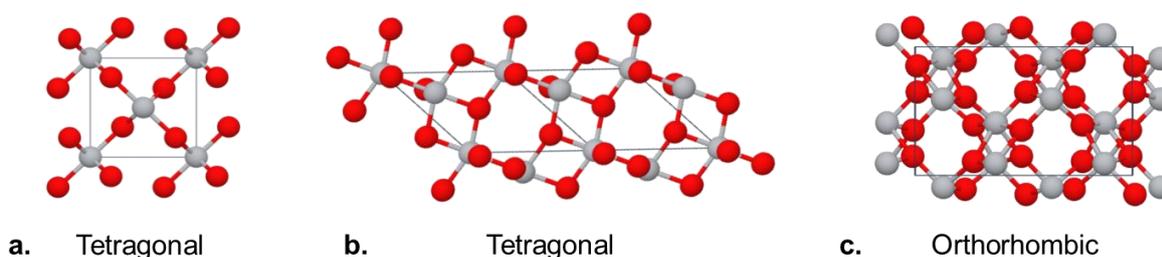
Figure 1.1. Crystalline phases of titanium dioxide a. rutile, space group P4₂/mnm, b. anatase, space group I4₁/amd and c. brookite, space group Pbca.

The first class of application is the construction sector with the use of TiO_2 in paintings, coatings, plastics and glass, representing about 84% of the global TiO_2 market.[35] As a white pigment and UV-filter, TiO_2 is also widely used in the cosmetics industry, in particular in sunscreen products. Finally, its catalytic properties make it a widely studied catalyst and catalyst-support material.

2.1.2. Applications in the construction sector

The refractive index of TiO_2 of 2.53 (anatase) to 2.75 (rutile)[38] is one of the highest among widely used particles and makes it the most efficient white pigment used in industry. Light diffusion, which is responsible for the opacity of a pigment, is directly linked to the difference in refractive index of the matrix and the particle: the greater the difference, the greater light diffuses and the whiter the pigment appears. Another major parameter is the particle size. TiO_2 white pigments are usually 0.2 to 0.3 µm as Mie diffusion phenomenon is optimal: light is diffused by particles independently of the wavelength (see section 2.3.1).[39] Therefore, particles diffuse equally every colour, making the pigment look white and bringing opacity to coatings.[40] Most commercially available TiO_2 pigments surface are modified and coated with alumina and/or silica to control its wettability, dispersibility and performance in various matrixes. The coating acts as a spacer between particles, preventing their agglomeration, and can also provide stabilizing electrostatic repulsion.

In specific paintings for the construction industry, TiO_2 particles are also studied as components of intumescent paintings designed for fire-retardancy. It was shown that the insulation of intumescent coatings is improved in the presence of TiO_2, in particular rutile phase, that enhances thermal stability and mechanical integrity of the char protective layer and prevents its oxidation.[41,42]

Titanium dioxide is also known as an efficient additive to building materials such as cement pastes, mortars and concretes. The main interest in the construction industry is the introduction of self-cleaning, air-depolluting and antimicrobial properties to those materials.[43–46] Indeed, these properties are due to the photocatalytic properties of TiO_2 and in that case, non-modified TiO_2 is used. The first patented applications of TiO_2 in building materials appeared in the 1990s and concerned mainly the self-cleaning properties and NOx removal from air.[47,48] The photoactivation of TiO_2 creates superhydrophilicity by favouring the hydration of the surface. As a result, surface wettability is improved and dirt is removed more easily from outdoor TiO_2-containing materials.[49,50] The photoactivation of TiO_2 also contributes to discoloration of pigments and degradation of organic pollutants.[51] This technology is not only used in cement-based materials but also in clear glass and coatings as the use of nano-scaled TiO_2 particles yields transparent materials.[8]

2.1.3. Use in cosmetics products

Titanium dioxide particles also find their applications in some cosmetic products, mostly as UV filters in sunscreens or as pigments in foundations.[5] TiO_2, due to its wide band gap of 3.23 eV for the anatase form and 3.06 eV for rutile,[52] absorbs light in the UV region. This feature is particularly interesting for sunscreen products. TiO_2 is an inorganic UV-filter approved by the Food and Drug Administration (FDA) in the USA and by the Scientific Committee for Consumer

Safety (SCCS) in Europe. In order to avoid the white film aspect of sunscreen products, nano-scaled TiO_2 is mostly used in this type of products, with a size below 100 nm,[5] whereas bigger particle size around 200 to 300 nm is preferred for make-up foundations in order to increase its covering power.

However, there exist some health hazards concerns associated to the use of TiO_2 nanoparticles. It was shown that when used in topical applications, the risk with nano-scaled TiO_2 particles is low due to the absence of percutaneous absorption.[5] In 2010, health issues associated to inhalation of TiO_2 nanoparticles caused it to be classified possible carcinogenic (group 2B) when inhaled in large doses by the International Agency for Research on Carcinogens (IARC).[53–55]

Their use in spray cosmetics is thus forbidden in Europe as it is a direct route to bloodstream absorption through the pulmonary barrier, and limited to a concentration of 25% in other formulations.[56] In Europe, only TiO_2 particles with the following characteristics are allowed in cosmetic products: purity superior to 99%, rutile phase or no more than 5% anatase, median size value of 30 nm at least based on number size distribution, and particles must be coated to avoid forming reactive radicals. It is estimated that the amount of generated radicals in sunscreen formulations is small enough to be contained by the skin's antioxidant species.[5] It was also shown that the efficiency of UV filters for skin protection is highly dependent on the distribution of this filter in the sunscreen film, that is why TiO_2 nanoparticles should be efficiently dispersed.[10]

2.1.4. Catalytic applications

At both the academic and the industrial levels, TiO_2 particles have been studied for their catalytic properties, either as main catalyst or as support material for heterogeneous catalysis due to strong metal support interaction, chemical stability, and acid-base property.[57] The use of TiO_2 improves the performance and catalytic activities of many reactions such as dehydrogenation,[58,59] hydrodesulphurization,[60] water-gas shift,[61,62] thermal catalytic decomposition,[63] selective reduction of NO by NH_3,[64] various mild oxidation reactions such as ethanol to acetaldehyde.[65] This general tendency shows in the bibliometrics of Figure 1.2, as the main keywords associated to titanium dioxide in the literature are related to catalysis and reactivity.

This variety in TiO_2 applications can be attributed to a variety of interesting properties compared to other types of metal oxides. TiO_2 semi-conductor behaviour addresses it for catalysis applications. As stated in the previous section, due to a wide band gap of 3.23 eV for anatase and 3.06 eV for rutile,[52] it absorbs light in the UV region. The UV light absorbed by TiO_2 only accounts for about 5% of solar energy. Modifying the band gap and shifting the absorption to the visible region would allow gathering about 43% of solar energy. This can be

achieved by either associating TiO_2 with other catalytic material, *i.e.*, as support material, or by doping.[66,67]
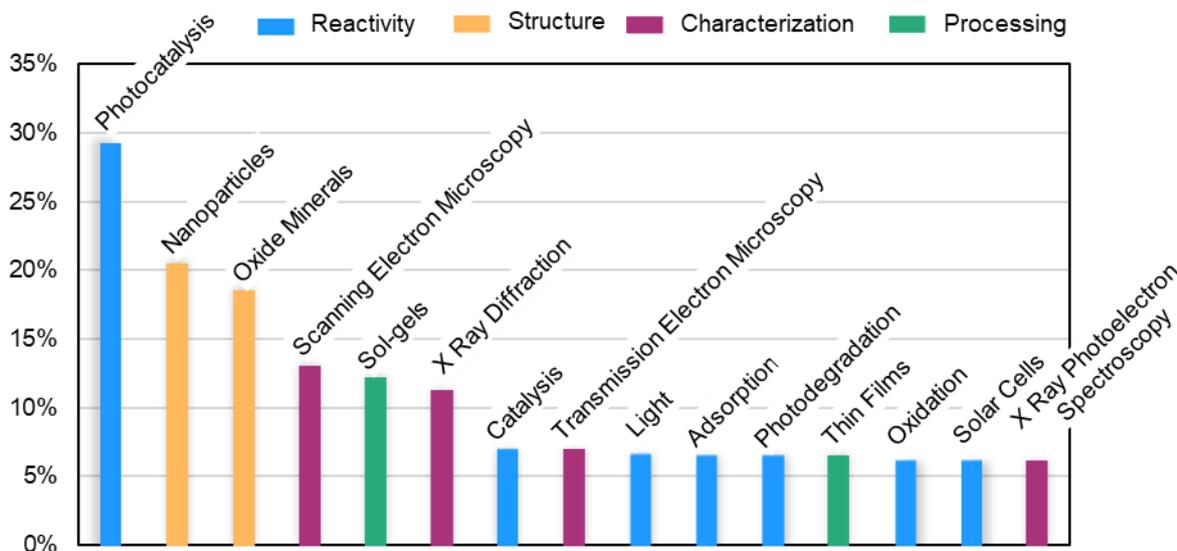
Figure 1.2. Main keywords occurrence associated to "titanium dioxide" in the Scopus database (134,681 documents) regrouped in clusters indicated by colours.

In the electrochemistry field, TiO_2-based materials are of interest due to a high conductivity and stability in both acidic and basic media.[57] That is why it can be regarded as a support for heterogeneous catalysts, guaranteeing stability in electrochemical environment and commercial availability. Pure anatase phase mesoporous TiO_2 with large surface area and narrow pore distribution was synthesized by Nolan *et al.* [68] to increase the degree of distribution and homogeneity of immobilized catalyst. It is well known that TiO_2, with small particle size and highly porous structure, greatly improves the photocatalytic performance of composite materials.[69] The synthesis of various nanosized metal particles immobilized on TiO_2 surface was reported.[70–72] It is inferred that the surface of growing particles is immediately complexed by TiO_2, therefore limiting the grain growth.[71]

The improved properties of TiO_2-supported catalysts are attributed to the photoinduced electron-hole pairs, increasing electron transfer and chemical reactivity. The TiO_2 excited electrons from the conduction band cause the reduction of metallic ions at TiO_2 surface. At the same time, holes, which are highly oxidizing species, contribute to oxidation reactions. The metal acts as an electron-trap specie, increasing the oxidizing power of TiO_2 holes. Moreover, the oxidation of surface hydroxyl groups can form highly reactive OH radicals.[73,74]

2.2. Dispersion agglomeration behaviour

For all the above-mentioned applications of TiO_2, such as catalysts synthesis and utilization, it is of primary importance that the particles remain well dispersed. The same way, the coverage of a paint or coating is enhanced with pigment dispersion. In cosmetics, it was shown that UV

skin protection directly depends on the homogeneous dispersion of TiO_2 particles. Due to the diverse existing forces in particle dispersions, these systems can either stay in a well dispersed state or evolve and tend to form agglomerates and sediment as the particle density is often larger than that of the dispersing media. In this section, the main forces involved in stabilization and destabilization of particle dispersions in liquid media are described.

2.2.1. Brownian motion

Particles with a diameter smaller than 1 µm are small enough to be impacted by the thermic agitation of solvent molecules. As a result, particles vibrate in an irregular and random motion. This motion is named after the botanist R. Brown who observed it in 1827 while looking at pollen and described the phenomenon.[75] In a heterogeneously concentrated dispersion, this Brownian motion contributes to the diffusion of particles from highly concentrated areas to less concentrated area of the bulk media. However, this random movement exists regardless of a concentration gradient as it is only due to thermic agitation. For particles larger than about 1 µm, and due to their larger inertia, collisions with solvent molecules are not energetic enough to put a particle in motion.[76]

2.2.2. Interparticle interactions in the DLVO theory

When colliding with each other, particles can either agglomerate or rebound depending on the relative strength of the van der Waals attraction and the electrostatic repulsion. These interactions are well described in the DLVO theory, named after the researchers Derjaguin, Landau, Verwey and Overbeek. It quantitatively accounts for these two types of interactions and was initially developed to rationalize the stability of dispersions in aqueous media. [77,78] The DLVO theory can be extended to any dispersing media provided that its relative permittivity, viscosity and refractive index are accounted for. Total interaction potential between two spheres is the sum of attractive potential V_A and the repulsive potential V_R, which is given by the following expression.[79]

$$V_R = 2 \pi \, \varepsilon' \, a \, \zeta^2 \ln(1 + exp(-r\kappa)) \tag{1}$$

where ε' is the solvent permittivity, a is the particle radius, ζ is the zeta potential, r is the distance between two spheres and κ^{-1} the Debye-Hückel distance defined by [80]:

$$\kappa^{-1} = \sqrt{\frac{\varepsilon' k_B T}{2 N_A e^2 I}} \tag{2}$$

where k_B is the Boltzmann constant, N_A the Avogadro constant, e the electron charge and I the ionic strength. The attractive component V_A, described by Hamaker [81], is given by equation (3).

$$V_A = \frac{-A_{12} \, a}{12 \, r} \tag{3}$$

Hamaker's constant A_{12} of component 2 in medium 1 is calculated with equation (4) from each

component constants A_1 and A_2 expressed by equation (5).[22]

$$A_{12} = \left(\sqrt{A_1} - \sqrt{A_2} \right)^2 \tag{4}$$

$$A_i = \frac{3}{4} k_B T \frac{(\varepsilon'_{r,i} - 1)^2}{(\varepsilon'_{r,i} + 1)^2} + \frac{3 h \nu_e (n_i^2 - 1)^2}{16\sqrt{2}(n_i^2 + 1)^{3/2}} \tag{5}$$

ε'_r is the relative permittivity, h is Planck's constant, ν_e is the main electronic absorption frequency for the dielectric permittivity calculated based on ionization potential IP and n is the refractive index. Hamaker's constant calculations for TiO_2 and all solvents used are available in appendix A2.

Figure 1.3 shows a typical example of the evolution of V_R and V_A over the distance r between two particles. When $V_T = V_R + V_A$ shows a positive maximum there exists a repulsive force preventing particle from agglomerating. When $V_T < 0$, no energy barrier is opposed to the agglomeration of particles and in the absence of other stabilizing forces the dispersion is usually very unstable.[82] It is estimated that if $V_T > 25\ k_B T$, the electrostatic repulsion is so predominant over the Van der Waals attraction that the dispersion is kinetically stable although the thermodynamically stable state corresponds to particles being in contact.[83]
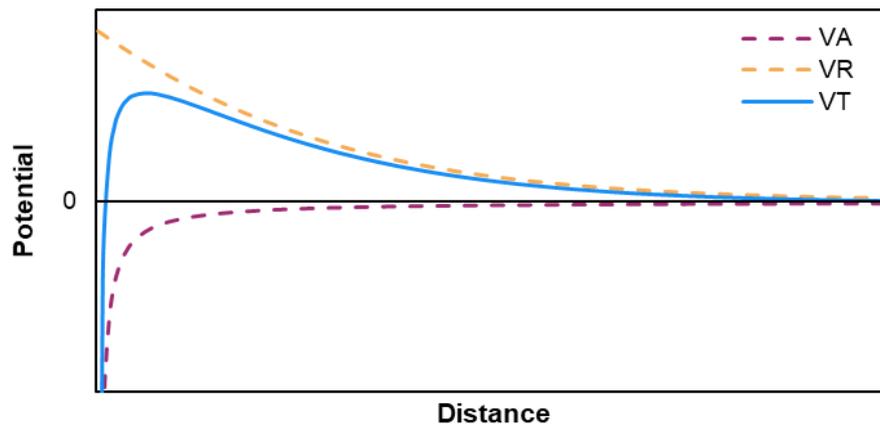
Figure 1.3. Typical evolution of total interaction potential $V_T = V_R + V_A$ over the distance r separating two spherical particles.

2.2.3. Particle and dispersing media interactions

Some non-DLVO interactions can also contribute to stabilizing particle dispersions. They intervene for surfaces with adsorbed layers of water (hydration), solvent (solvation), ions, surfactants, polymers or nanobubbles. The thickness of the adsorbed layer depends on the respective properties of the particles and the medium and can create an additional distance between two particles, thus preventing them from getting in the high interaction potential area as shown in Figure 1.3. Indeed, when the particle size is smaller than 100 nm, the thickness of adsorbed layer is of the same order of magnitude than the van der Waals interaction distance.[84,85] Besides this steric stabilization of surfaces, there can also exist additional

attraction forces. In the case of adsorbed hydrophobic compounds or nanobubbles at the particle surface in a polar medium, hydrophobized surfaces attract strongly with each other [22]. In the case of adsorbed nanobubbles, this attraction is effective at long range (a few hundreds of nanometres) and is attributed to surface tension of the bridging gas bubbles.[86,87]

In the case of particle surface solvation, interactions between solvent molecules and particle surface can be described in many ways. The nature of the particle can induce surface acidity, *e.g.*, for metal oxides, local polarization, or hydrophobic interactions with the solvent, *e.g.*, for carbon-rich particles such as graphene. There exist many solvents scales to classify solvent-solute interactions that can be extended to particle-solvent interactions. Mostly two solvent scales are exploited in the literature: the donor-acceptor numbers in relation with the Lewis acidity and basicity of solvents interacting with the particle, and the Hansen solubility parameters (HSP) associated to the Hildebrand parameter and cohesive energy density of a liquid.

Donor-Acceptor Numbers

This theory was developed by Victor Gutmann in 1976 and is based on the principle that interactions between a solvent and a solute or a particle are mainly Lewis acidity and basicity. The ability of a solvent to accept electrons, *i.e.*, its Acceptor Number (AN) or Lewis acidity is measured by means of ^{31}P NMR and using triethylphosphine oxide Et_3PO as a probe solute. Indeed, the chemical shift of Et_3PO gets higher as the solvent AN increases.[88] The Donor number (DN) of a compound is associated to the enthalpy of reaction between the compound and $SbCl_5$, a typical Lewis acid, dissolved in 1,2-dichloroethane, a non-coordinating solvent.[88] AN values are correlated to other Lewis acidity scales such as Dimroth-Reichardt $E_T(30)$ and Kamlet-Taft α parameter. DN values are correlated to Kamlet-Taft β parameter.[89]

Hansen Solubility Parameters

This approach derives from the Hildebrand solubility parameter δ_H, which is the square root of cohesive energy density, see equation (6). The cohesive energy corresponds to the amount of energy necessary to remove a volume unit of molecules from the bulk, *i.e.*, the heat of vaporization divided by the molar volume. Hansen's assumption is that this cohesive energy density is due to three types of interactions between molecules: polar (δ_p), dispersive (δ_d) and hydrogen bonding (δ_h) interactions as defined in equation (7).[24]

$$\delta_H = \sqrt{\frac{\Delta H_v - RT}{V_m}} \tag{6}$$

with R the molar gas constant, T the temperature, ΔH_v the heat of vaporization and V_m the molar volume of the considered compound.

$$\delta_H = \sqrt{{\delta_p}^2 + {\delta_d}^2 + {\delta_h}^2} \tag{7}$$

Originally, Hansen's solubility parameters were developed to study and anticipate the solubilization of molecular and macromolecular compounds in organic solvents, on the hypothesis that similar solubility parameters for both the solute and the solvent lead to favourable solubilization. Hansen himself proposed to apply his method to characterize the surface of dispersed particles by arguing that organic liquids which adsorb most strongly to the surfaces of particles are those in which dispersions are most stable and that this adsorption depends on δ_p, δ_d and δ_h differences.[24,25] It is also argued that for small nanoparticles, energy of solvation $\Delta G_{solv.}$ can be negative[26] and thus, solvation would be thermodynamically favoured, placing NPs in the frame of Hansen Parameters.[27] However, as the physicochemical phenomena involved in the dispersion of particles are definitely different from those involved in the dissolution of organic compounds, Süß et al. proposed to use the term "Hansen's Dispersion Parameters" instead of HSP when Hansen's approach is used to study the dispersibility and stability of particles.[14] HSP has been shown to be a versatile tool for rationalizing and predicting the stability of various types of NP dispersions such as carbon black,[14] carbon nanofibers,[28] fullerene,[29] graphene[30] and carbon nanotubes[31] but also titanium carbides[32] and inorganic nanoparticles of ZnO, Al_2O_3, ZrO_2,[33] hydroxyapatite and TiO_2.[19] In practice, the particles are dispersed in a series of carefully chosen solvents according to a standard protocol. Each solvent is assigned a score based on the stability of the dispersion, then a Hansen sphere including the most effective solvents is built in the 3D Hansen space as described in Figure 1.4.
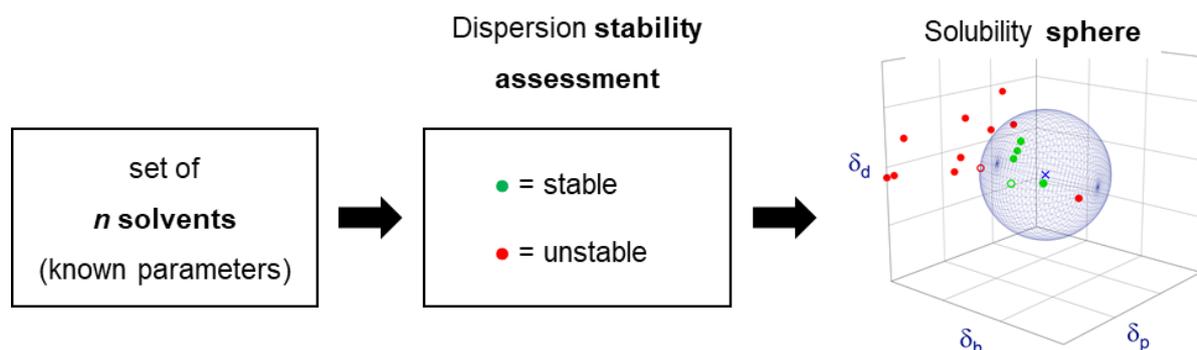
Figure 1.4. Experimental procedure to determine the Hansen solubility sphere of a particle. The centre of the resulting sphere corresponds to the experimental HSP of the particle.

The other previously described solvent scales (Donor-acceptor numbers, $E_T(30)$ and Kamlet-Taft) could be used to quantify solvent-particle interactions, but in this work we chose to use the Hansen Parameters as the three complementary parameters allow a practical 3D representation and a clear visualization of effective and non-effective solvents.

2.3. Dispersion stability and detection methods

A dispersion is considered stable when it is prone to avoiding both agglomeration and sedimentation phenomena. Assessing the stability of a dispersion can be achieved by many techniques. Monitoring the particle size and the sedimentation rate allows a global understanding of a dispersion evolution and stability over time.

In liquid medium, dispersed particles occur in various states and configuration: single primary particles, agglomerates and aggregates as shown in Figure 1.5. Depending on the strength of van der Waals attraction and of the dispersion process energy, agglomerates can be broken down to primary particles in some cases, however aggregates cannot be broken down to smaller entities by physical processes as they result of strong covalent bonds.
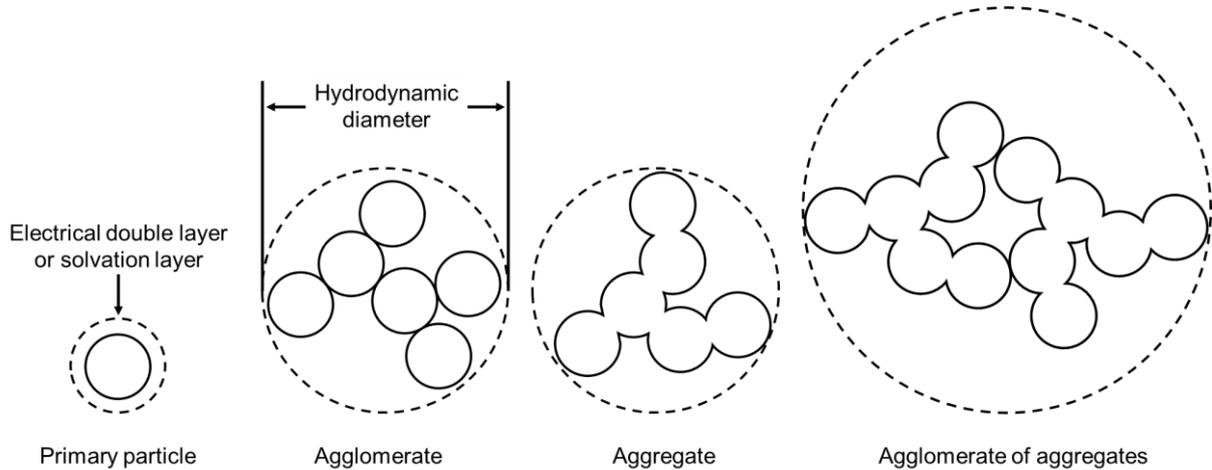
Figure 1.5. States and configurations of dispersed particles in liquid media. Redrawn from [12].

2.3.1. Light scattering behaviour of particle dispersions

When an electromagnetic wave meets a particle, light is scattered in the three directions of space. Depending on the size and shape of the particle, partial light extinction of the incident light is observed. Two main theories characterize this extinction behaviour and are exploited in particle size measurement devices and dispersion stability monitoring devices.

<u>Rayleigh light scattering theory</u>

This model corresponds to light scattering by small objects with a diameter range such as $d < \lambda/20$. In that case, the scattered intensity $I_S(\theta)$ compared to the incident light intensity I_0 is given by Rayleigh's law in equation (8)

$$\frac{I_S(\theta)}{I_0} = \frac{8\pi^4 d^6}{r^2 \lambda^4} \left(\frac{m^2 - 1}{m^2 + 2}\right)^2 (1 + \cos^2 \theta) \tag{8}$$

where d is the particle diameter, r is the distance from the observation point, m is the ratio of the particle refractive index compared to that of the continuous phase and θ is the angle of observation.[90]

It is worth noting that $I_S(\theta)$ is proportional to λ^{-4}. This dependence results in apparent blue reflects as λ_{blue} is the smallest wavelength of the visible light spectrum, and thus the most scattered one. Also, the intensity increases with the particle size, up to the point where light scattering changes behaviour and is defined by the intermediate Rayleigh-Gans-Debye theory. In this intermediary expression, equation (8) is corrected by a form factor $F(\theta)$, accounting for the asymmetric angular distribution of light scattering.[91] When particles get bigger than about $\lambda/10$, the scattering behaviour changes again and is expressed by the Mie light scattering theory.

Mie light scattering theory

The Mie theory, developed by Gustav Mie in 1908, describes light scattering by spherical objects bigger than about $\lambda/10$. In that case, light is scattered mostly in the direction of the incident light but also in other angles, depending on the particle size. Size measurement devices based on this theory are typically laser granulometers, that require working with dilute dispersions to avoid superposition of scattered signals.[40,91]

2.3.2. Size measurements

Dynamic Light Scattering (DLS)

Various techniques exist to measure particle size. Depending on the sample size range, some techniques are more appropriate than others. One of the most famous ones is Differential Light Scattering (DLS), based on optical observation of the sample. Due to thermic agitation, particles vibrate with a size-depending amplitude. In usually used apparatus, an incident laser beam passes through the dispersion and meets vibrating particles that scatter light, as described in section 2.3.1. The scattered light intensity varies as particles vibrate and move inside the sample. The variation speed is measured and converted to particle size distribution. This technique is particularly indicated for sizes comprised between 0.3 nm and 10 µm. Over this limit, Brownian movement is insufficient to be detected. Moreover, quickly sedimenting particles can induce a measurement error as the observed particle movement is a sum of both Brownian motion and sedimentation. Finally, DLS is only applicable to partially clear dispersion as it requires measuring the transmitted light, which is inexistent for an opaque sample. A solution is to dilute a concentrated dispersion, but it implies changing the measurement conditions compared to the original state of the dispersion.[92]

Static Multiple Light Scattering (SMLS)

For concentrated dispersion, the Static Multiple Light Scattering (SMLS) method is indicated as it requires no dilution and is adapted to opaque samples.[93] In SMLS apparatus, typically Turbiscan®, a monochromatic light beam propagates in a dispersion and is scattered by dispersed particles. Two sensors are placed so as to collect the resulting backscattered and

transmitted light. The transmitted *TR* light signal can be expressed by the lambert-Beer law, see equation (9), and the backscattered *BS* signal is approximated by equation (10):[93]

$$TR = TR_0 \exp\left(\frac{-3r_i \Phi Q_e(d)}{d}\right) \tag{9}$$

$$BS = \alpha \sqrt{\frac{3\Phi(1 - g(d))Q_e(d)}{2d}} + \beta \tag{10}$$

with TR_0 the transmitted signal of the continuous phase, r_i the internal radius of the measurement cell, d the particle mean size, Φ the volume fraction of dispersed phase, Q_e the extinction efficiency, α and β the gain and offset of the experimental setup and g the asymmetry factor that quantifies the anisotropy of the light scattered by particles.

This size measurement method is indicated for particle sizes ranging from 10 nm to 1 mm and can also be applied to droplets in an emulsion for instance. This technique allows working with opaque samples, but only yields an average size and no size distribution, which remains less accurate than DLS size measurements.

2.3.3. Sedimentation rate

The kinetics of destabilization can be relevant for a number of applications, *e.g.*, for the stability evaluation of consumer products. To that end, some commercially available devices allow monitoring the sedimentation rate of particle dispersions, under gravitational field or accelerated one. This latter type of device is generally preferred for accelerated stability assessment. The LUMiSizer® products from LUM GmbH are based on this technology: samples are placed in tubes held around a rotative axis. Several light sources are placed perpendicularly to the tubes over the sample height, and detectors are placed both next to the emitting source and in front of it, gathering the transmitted and backscattered light while the sample is being centrifuged at a force between 6×*g* and 2300×*g*. The evolution of these signals over time and sample height give information regarding the size of dispersed objects, directly related to their sedimentation rate.[94]

The second type of device, typically the Turbiscan® products from Formulaction, are based on gravitational sedimentation monitoring. Samples are monitored without centrifugation. A mobile light beam is placed perpendicularly to the sample. Detectors are placed next to the emitting source and facing it so as to measure transmitted and backscattered light.[93] The evolution of samples is, thus, much slower but allows detecting the agglomeration phenomena, which cannot be accelerated by centrifugation as they result from interparticle interactions and are not impacted by gravitational forces. It is the instrument used in this work so as to analyze samples in real-life conditions and detect every type of destabilization.

3. Experimental section

3.1. Chemicals

The nanoparticles AEROXIDE ® TiO$_2$ P25 (titanium dioxide, purity ≥ 99.5 %) were obtained from Acros Organics (Thermo Fisher Scientific Inc., Geel, Belgium). These uncoated TiO$_2$ NPs had a specific surface area of 35 – 65 m²/g [9] and an average primary particle diameter of 21 nm.[95] Crystal structure was mainly anatase (85 %) and rutile (15 %).[96]

Organic solvents used as dispersion media were supplied by the companies Sigma-Aldrich Chemie GmbH (St. Louis, USA), VWR International GmbH (Radnor, Pennsylvania), Acros Organics (Thermo Fisher Scientific Inc., Geel, Belgium), Alfa Aesar (Thermo Fisher Scientific Inc., Heysham, UK), Honeywell (Honeywell International Inc., Morristown, USA), Verbièse (Laboratoire Verbièse, Merville, France) and TCI (Tokyo Chemical Industry Co. Ltd., Tokyo, Japan).They were all of the highest purity available and used as such. Trifluoroacetic acid (99 %) was supplied by Alfa Aesar and tetrabutylammonium hydroxide (1 M in methanol) was supplied by Sigma-Aldrich. Ultrapure water was obtained using a Thermo Scientific Barnstead MicroPure Ultrapure water system with a resistivity of 18.2 MΩ.cm.

3.2. Protocol for dispersing TiO$_2$ P25 nanoparticles

Aqueous dispersions

pH-controlled solutions were prepared by mixing NaCl 10^{-3} M and either NaOH or HCl 10^{-3} M solutions in order to maintain ionic strength at a constant value of 10^{-3} M. 20 mg of TiO$_2$ was placed in a borosilicated glass cell (from Formulaction Company, 27.5 mm diameter), then 20 mL NaCl 10^{-3} M and HCl or NaOH 10^{-3} M were added. The cell was placed in a 23.0 °C thermoregulated bath and sonicated for 12 min using an ultrasonic probe Sonotrode S26d2 (2 mm diameter) immersed by 5 mm in the liquid and operated by the ultrasonic processor UP200St (both from Hielscher). The sonotrode pulse was fixed at 50 % and the amplitude at 20 %. Thermoregulation was carried out by means of a Huber Ministat 125 circulating water bath. pH was measured by immersing the electrode directly in the cell and a sample was taken to measure particle size and zeta-potential ζ using a Zetasizer Nano ZS from Malvern Panalytical. The cell was wiped to remove water drops from the outside wall and was scanned by a thermoregulated Turbiscan LAB from Formulaction.

Dispersions in organic solvents

In organic solvents, 20 mL of solvent ("pure" in the first series of experiments and containing 10^{-3} M trifluoroacetic acid or tetrabutylammonium hydroxide afterwards) were sampled with a graduated pipette and added to a borosilicated glass cell containing 20 mg of TiO$_2$. The particles were then dispersed and analysed according to the protocol described above. Particle size was measured according to the DLS method using a Zetasizer Nano ZS from Malvern

Panalytical. Mean size is given as Z-average size, defined as the harmonic intensity average particle diameter.

3.3. Turbiscan measurement

Dispersions were scanned every 30 seconds for 2 to 72 hours at 25.0 °C. Measured signals are transmitted light (*TR*) and backscattered light (*BS*) represented as a function of the sample height. Variations in *TR* and *BS* can be seen by deducting the first scan to all the following scans yielding *ΔTR* and *ΔBS*. Typical resulting spectra are shown in Figure 1.6.
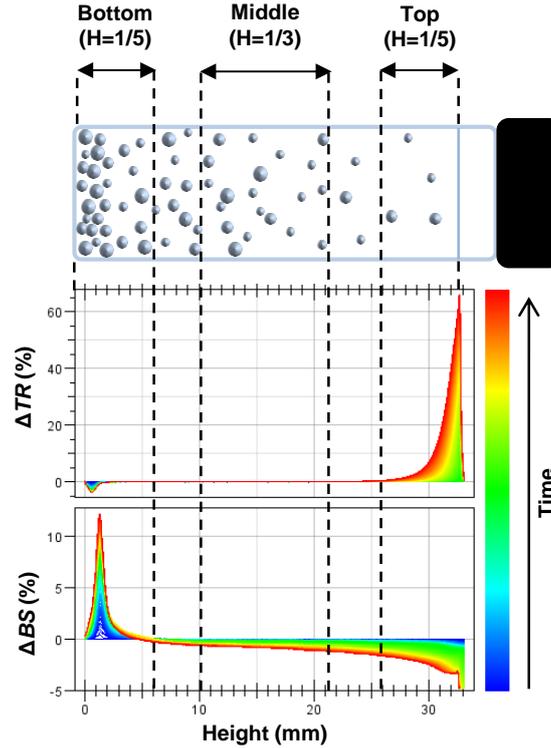
Figure 1.6. Change of transmitted light (*ΔTR)* and backscattered light (*ΔBS*) from Turbiscan data for a partially sedimented dispersion. *TR* increases at the top and *BS* increases at the bottom as particles sediment over time.

Experimental data were processed using the TurbiSoft Lab software (2.3.1.125 FAnalyser) and *TSI* (Turbiscan Stability Index) was computed according to the following equation.[97]

$$TSI(t) = \frac{1}{N_h} \sum_{t_i=1}^{t_{max}} \sum_{z_i=z_{min}}^{z_{max}} |BS,TR(t_i,z_i) - BS,TR(t_{i-1},z_i)| \tag{11}$$

with t_{max} the measurement time at which the *TSI* is calculated, z_{min} and z_{max} the lower and upper selected height limits respectively, $N_h = (z_{max} - z_{min})/\Delta h$ the number of height positions in the selected zone of the scan and *BS,TR* the considered signal (backscattering *BS* if *TR* < 0.2% or transmission *TR* otherwise). Consequently, *TSI* = 0 for t = 0 and increases as the sample gets destabilized. High *TSI* values are characteristic of unstable dispersions. In organic media, a Relative Turbiscan Stability Index (*RTSI*) was defined according to equation (12) to compare solvents with each other, so as to consider their viscosity η [cP] and

their density ρ_i [g.cm^{-3}] which modify the sedimentation rate according to Stokes' law.

$$RTSI = TSI \times \frac{\eta}{(\rho_{particle} - \rho_{solvent})}$$

(12)

Moreover, the variations of *TR* (or *BS* when the sample is opaque) allow the calculation of mean particle size *via* Stokes sedimentation. Using the width evolution at a threshold of the *ΔTR* clarification peak at the top of the cell (see Figure 1.7), the migration rate of particles corrected by the viscosity and density of the solvent is linked to particle size by equation (13).[98] A threshold of $\Delta TR = \Delta TR_{max}/10$ was chosen to measure d_{Stokes} in all samples. When variation of the peak width was not linear, due to sedimentation of different size populations in the sample, the initial slope was used.

$$v = \frac{d^2 \times g \times (\rho_p - \rho_f)}{18\,\eta} \times \frac{1 - \Phi}{1 + \dfrac{4.6\,\Phi}{(1 - \Phi)^3}}$$

(13)

with *v* the sedimentation rate, *Φ* the solid volume fraction, *d* the diameter, *g* the gravitational constant, ρ_p and ρ_f the density of particles and fluid respectively and *η* the viscosity.
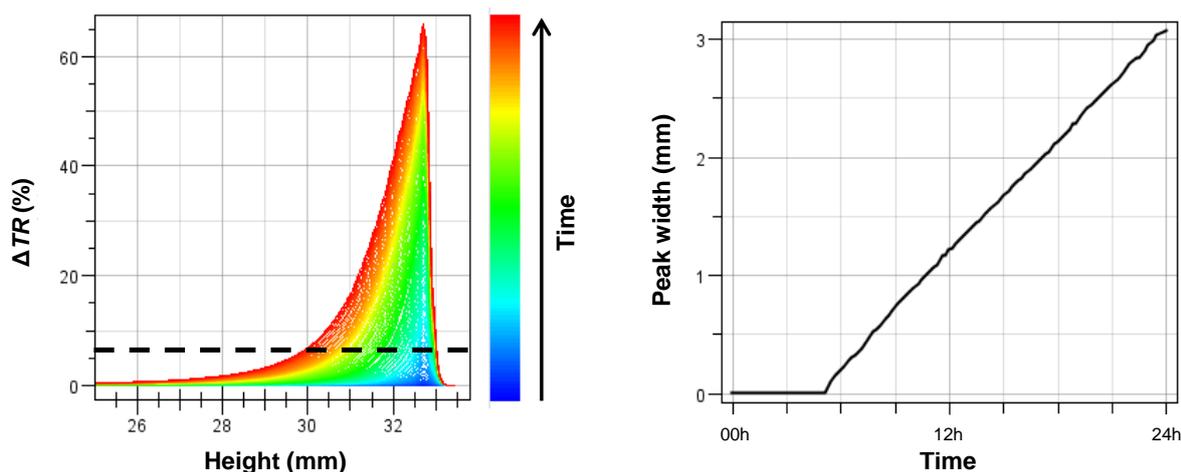
Figure 1.7. Δ*TR* clarification peak and threshold at Δ*TR* = Δ*TR*$_{max}$/10 (left). Peak width evolution at the 1/10 threshold (right) : the slope corresponds to the sedimentation rate *v*.

3.4. Zeta potential measurement

Immediately after dispersing the NPs, 1 mL of dispersion was introduced in a folded capillary zeta cell (aqueous dispersions) or a dip cell (dispersions in organic solvents) and the zeta potential of TiO$_2$ suspensions was measured using a Zetasizer Nano ZS from Malvern Panalytical. In organic solvents, measuring ζ is more challenging than in water. Indeed, the usual folded capillary cell made out of polycarbonate (see Figure 1.8a) cannot be used for two reasons. Firstly, this polymer can be dissolved by some organic solvents. Being made of polyether-ether-ketone (PEEK) polymer and a glass cuvette with excellent chemical compatibility, the dip cell (Figure 1.8b) is more addressed for this type of measurements. Secondly, low conductivity media require a higher applied field to observe electrophoretic

mobility. The dip cell electrodes are positioned closely together on the PEEK probe, producing greater electric fields with lower voltage and thus avoiding local heating and electrolysis effects.[99,100]
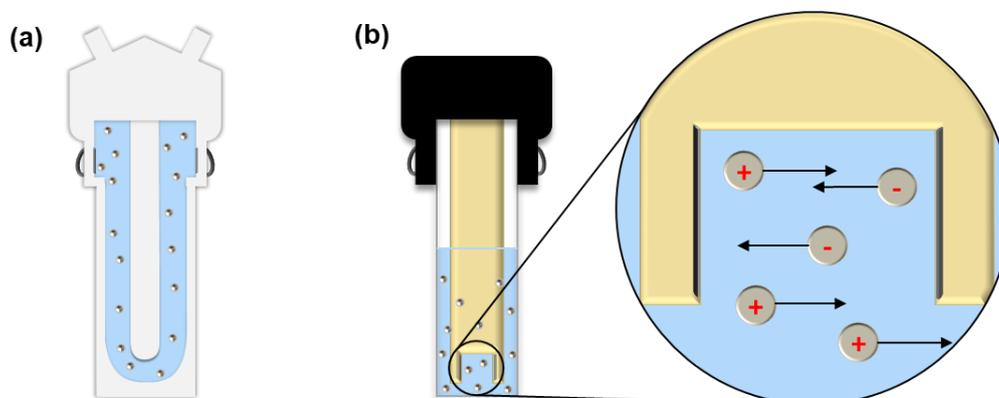
Figure 1.8. Zeta potential (a) capillary polycarbonate cell and (b) dip cell made of PEEK probe and glass cell and particles moving under electrophoretic motion between the electrodes.

Zeta potential is the potential at the surface between a rigidly adsorbed layer and the freely flowing solvent molecules.[101] The functional groups at the particle surface interact with the dispersing medium. Dipole-dipole interactions induce the creation of electric charges that affect the electrostatic forces of particles with one-another. The measurement is based on laser Doppler electrophoresis. The zetasizer measures the frequency difference between the incident and the backscattered beams on the chamber detectors. This Doppler effect is induced by the particle movement under the difference of electric potential. Zeta potential ζ is then calculated using Hückel's equation:[102]

$$u_{Hückel} = \frac{2\,\varepsilon'_0\,\varepsilon'_r\,\zeta}{3\,\eta} \tag{14}$$

with u the particle mobility, ε'_r the relative permittivity of the solvent, ε'_0 the electrical permittivity of vacuum and η the viscosity.

Three concordant zeta measurements were achieved for each sample, and values reported in this work are the average zeta values. Uncertainty of measurement is estimated to be the standard deviation. In organic solvents, zeta deviation is generally larger than in aqueous media. A few examples are shown in Figure 1.9. Zeta potentials measured in solvents without the addition of electrolytes should be considered specific to this study as it may differ according to experimental conditions, namely the solvent purity, supplier or batch. Zeta potential of TiO_2 in non-aqueous solvents is very sensitive to the presence and nature of impurities.[103,104]
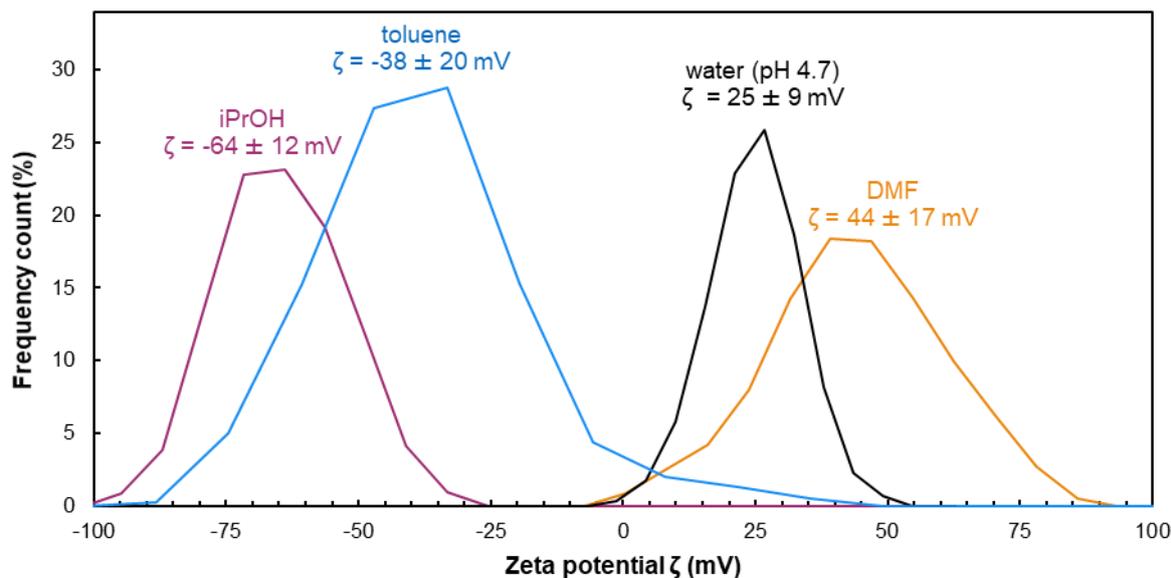
Figure 1.9. Zeta potential raw measurements of 1 g/L TiO_2 P25 dispersions in isopropanol (pink), toluene (blue), dimethylformamide (yellow) and aqueous solution with pH = 4.7 (black). Measurements are achieved with a dip cell in organic solvents and a capillary cell in aqueous media.

3.5. Hansen sphere calculation

Two HSP spheres of TiO_2 P25 were determined based on either one of two distinct stability criteria, namely d_{Stokes} and RTSI. Solvents were rated according to one of these parameters: score 1 for "good" dispersing media (stable dispersion) and score 2 for "bad" dispersing media (unstable dispersion). A threshold was defined for d_{Stokes} and RTSI as follows.

$$\text{Score 1: } d_{Stokes} < 0.31 \text{ μm or } RTSI_{2h}^{top} < 1.0$$

$$\text{Score 2: } d_{Stokes} > 0.31 \text{ μm or } RTSI_{2h}^{top} > 1.0$$

Using HSPiP software and the scores attributed to each solvent, the spheres are computed so as to include scores 1 and exclude scores 2. The fit indictor reflects the quality of the sphere computation: it decreases if scores 1 are excluded and scores 2 are included in the sphere. The center of the sphere, represented by the three coordinates δ_d, δ_p and δ_h, corresponds to the Hansen Solubility Parameters of TiO_2.

4. TiO₂ dispersions stability in aqueous media

Zeta potential ζ is the key parameter to rationalize the stability of aqueous TiO₂ dispersions.[12,105–107] Figure 1.10 shows the evolution of ζ as a function of pH in aqueous solution at constant ionic strength. When dispersed in solution, NPs collide with each other due to the Brownian movement. They can either agglomerate or rebound depending on the relative strength of the van der Waals attraction and the electrostatic repulsion. The DLVO theory which quantitatively accounts for these two types of interactions, was initially developed to rationalize the stability of dispersions in aqueous media.[77,78] Surface charge effects of TiO₂ dispersions in water have been extensively studied.[12,105–109] The pH of the aqueous solution plays an important role as it influences ion exchanges between NPs and water, modifying the surface charge and accordingly, the zeta potential. Indeed, the presence of acidic Ti^{IV} sites on the surface causes water dissociation by adsorption, creating -OH functional groups.[108] NP surface is then modified *via* reaction with H⁺ or HO⁻ ions according to equilibria depicted in Figure 1.11.[108,110]
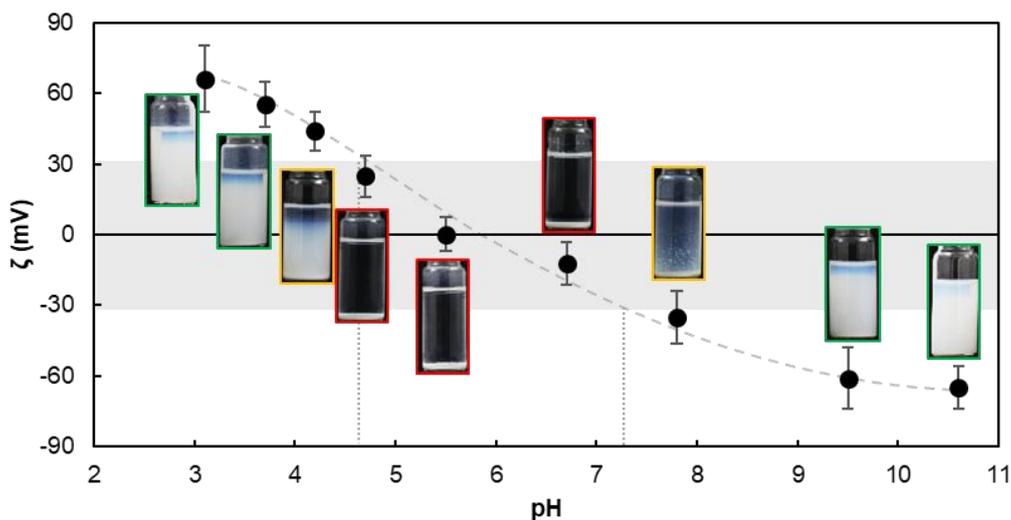
Figure 1.10. Evolution of 1 g/L TiO₂ P25 nanoparticle zeta potential with pH in water at 25.0 °C. pH is adjusted with HCl and NaOH. Ionic strength is kept constant at 10⁻³ M with NaCl. Instability area is marked in grey, corresponding to pH = 4.6 to 7.2. Pictures of dispersions are taken after 24 hours.
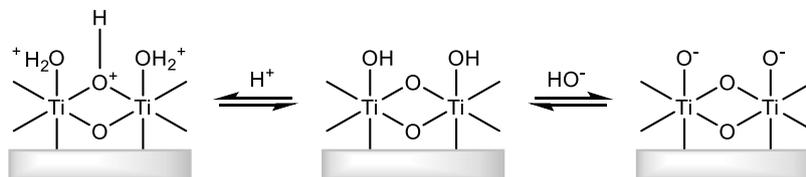
Figure 1.11. Simplified illustration of electric charges formation on TiO₂ surface by acido-basic reactions in water.

Bidentate bridge OH between two Ti atoms (Ti-OH⁺-Ti) and monodentate terminal H₂O adsorbed on 5-fold Ti sites have pK_a values of 2.9 and 12.7 respectively.[111] The reported values for the isoelectric point of TiO₂ P25 are comprised between 5.8 and 6.6.[108,111–113] As shown in Figure 1.10, an isoelectric point of 5.9 was found in accordance with the literature. Electrolyte concentration is also known to strongly impact the zeta potential as more counter-ions can

screen surface charges when its concentration increases. Those differences between measured zeta potential and surface potential are minimized at low electrolyte concentration and with monovalent ions.[77,78,114] In this work, pH was adjusted by NaOH or HCl addition and the ionic strength was then adjusted to a constant value of 10^{-3} M by NaCl addition. pH ranging from 4.6 to 7.2 causes ζ to be comprised within the -30 mV to 30 mV interval, leading to quickly destabilized dispersions. Such destabilizations can be finely analysed using a Turbiscan® LAB through the detection of the transmission TR and backscattered BS light signals. Turbiscan profiles at various pH values are presented in Appendix A3. The differences in stability depending on the pH can be seen in Figure 1.12, showing the evolution of the transmitted TR and backscattered BS lights of two aqueous TiO_2 nanoparticle dispersions at pH = 6.7 and pH = 7.8.
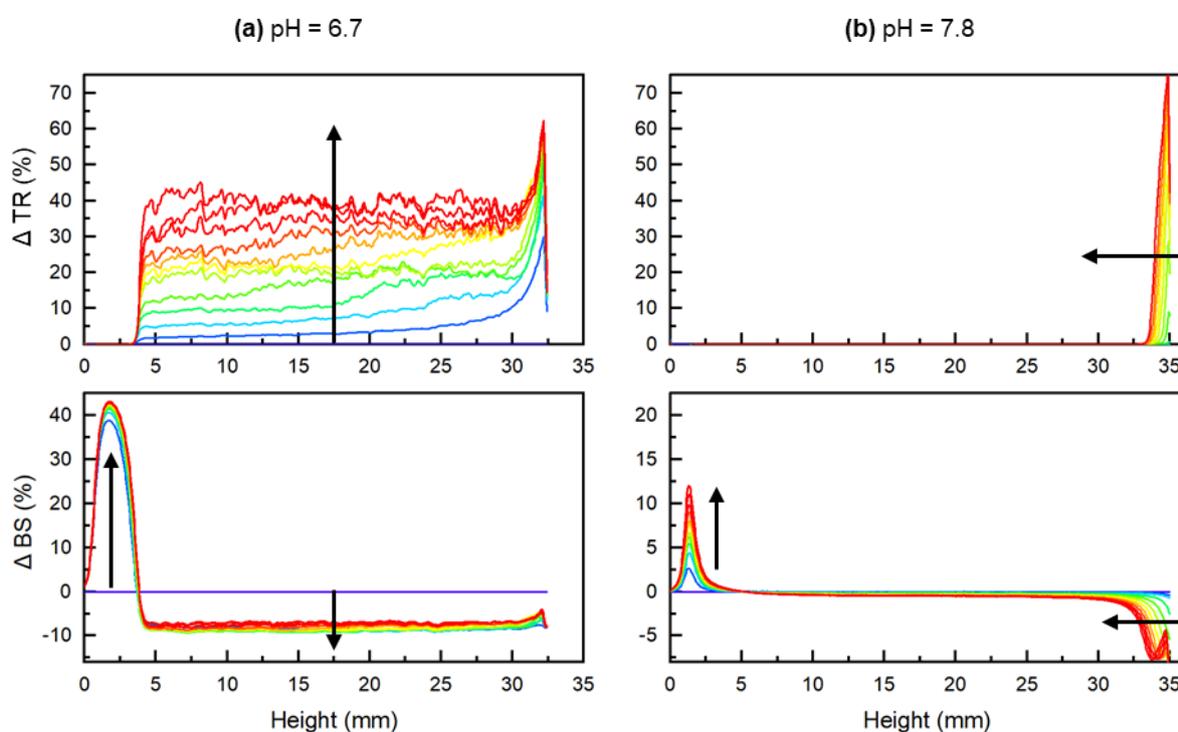
Figure 1.12. Variations over 2 hours of transmitted TR and backscattered BS light of aqueous TiO_2 P25 1 g/L dispersions (T = 25.0 °C) at pH = 6.7 and ζ = - 12 mV (a) and at pH = 7.8 and ζ = - 35 mV (b).

When DLVO repulsion is strong enough to avoid particle agglomeration, the typical profile is the one in Figure 1.12b. TR signal shows no variation except at the top, where slow sedimentation occurs. With time, the sedimentation front (also visible in BS light) would eventually reach the bottom of the cell. BS light increases as particles accumulate at the bottom. Those samples are easily re-dispersed with a simple re-agitation. On the contrary, when DLVO repulsion is weak (i.e., zeta potential is less than 30 mV), as in Figure 1.12a, destabilization is fast. BS decreases and TR increases at the top and in the middle of the cell due to particle agglomeration. Indeed, the particle concentration decreases as they agglomerate. TR even turns into an irregular signal when agglomerates tend to be individually distinguishable. Sedimented particles accumulate faster if DLVO repulsion is low: the

increasing size of agglomerates, in turn, accelerates their sedimentation rate.

All those variations over time can be accounted for by the *TSI* value. Indeed, using *TR* and *BS* signals, the average *TSI* value (see equation (11)) can be calculated on any portion of the cell and at any time of the analysis. It is worth noting that phenomena differ depending on the height: at the top, *TR* increases faster than below as clarification occurs. Figure 1.13 shows the evolution of *TSI* on three portions of the cell in the case of the TiO_2 dispersions at pH = 3.1. Destabilisation can be detected faster when looking at the top of the cell, with a faster increase in *TSI* value than in the middle or the bottom of it.
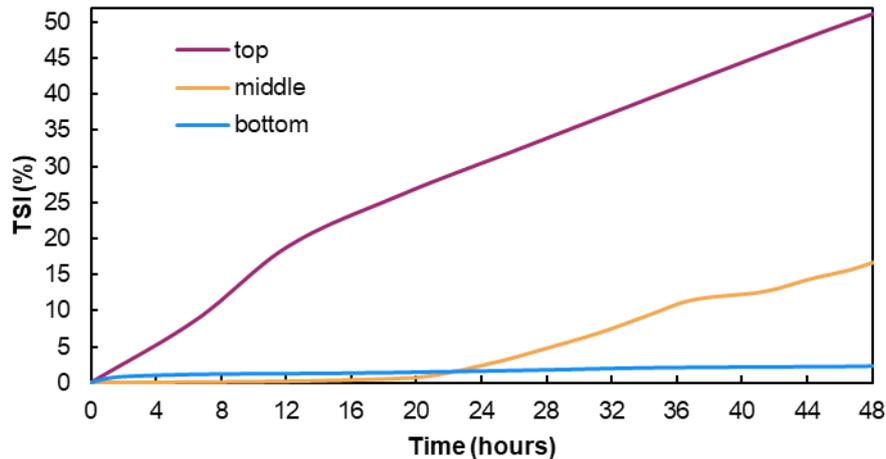
Figure 1.13. Turbiscan Stability Index (*TSI*) evolution in a 1 g/L TiO_2 P25 aqueous dispersion (pH = 3.1, ionic strength = 10^{-3} M) at 25.0 °C over two days.

TSI represented in Figure 1.14a has been calculated at the top of the cell in order to detect the very early changes even for the most stable samples. After 2 hours, TSI_{2h}^{top} values match the expected dispersions behaviour in accordance with the DLVO theory. When $|\zeta| < 30\ mV$, TSI_{2h}^{top} is significantly higher than outside these boundaries. Destabilization is faster within this interval whereas solutions with high $|\zeta|$ remain stable. In Figure 1.14b, the Stokes diameters, calculated based on the sedimentation rates of particles (see experimental section 3.3), have
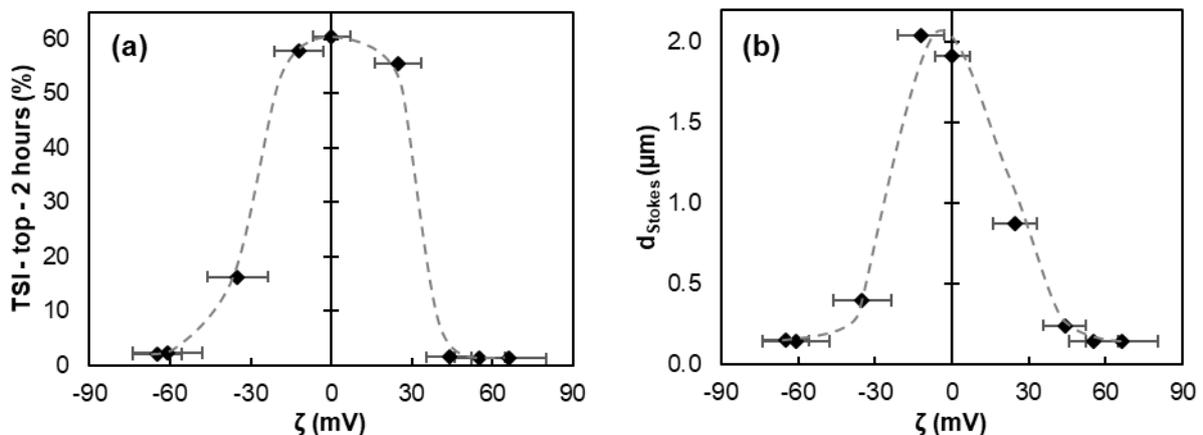
Figure 1.14. Evolution at 25.0 °C of (a) *TSI* at the top of the cell after 2 hours and (b) Stokes diameters determined using the Turbiscan with zeta potential ζ of 1 g/L TiO_2 P25 aqueous dispersions. Ionic strength is kept constant at 10^{-3} M by NaCl addition, pH is adjusted using NaOH and HCl solutions.

been reported as a function of ζ. They reach a maximum around the isoelectric point and are noticeably smaller as |ζ| increases. All these findings clearly show that Turbiscan TSI_{2h}^{top} and Stokes diameters appear as reliable values to quantify dispersion stability. In particular, the maximum d_{Stokes} being attained for zeta potentials close to 0 mV is in accordance with the increase of coagulation rate when electrostatic repulsion is low. In that case, there exists little or no energetic barrier to prevent particles from colliding and coagulation occurs rapidly.[82]

5. TiO$_2$ dispersions in non-aqueous solvents
5.1. DLVO interactions in methanol, a non-aqueous solvent

DLVO theory quantitatively accounts for interactions between particles and can be applied to organic dispersing media.[16,115,116] The main differences between aqueous and organic dispersing media concern dielectric constants and electrolyte concentrations. Both those factors impact the electrostatic repulsion potential. The stability of TiO$_2$ dispersion was investigated in a common polar and protic solvent, namely methanol. Figure 1.15 shows the variation of the zeta potential at different pH values measured using a glass electrode standardized in aqueous buffer. The pH range accessible, based on the water scale, is -1.8 to 17.2 and depends on the dissociation constants of methanol.[117] The lower limit is given by the transfer activity coefficient of H$^+$ ions in methanol $p\gamma_{H_2O \rightarrow MeOH}^{t}(H^+) = -1.8$ and the upper limit is given by the autoprotolysis constant of methanol $pK_{MeOH} = 17.2$.[117–119]
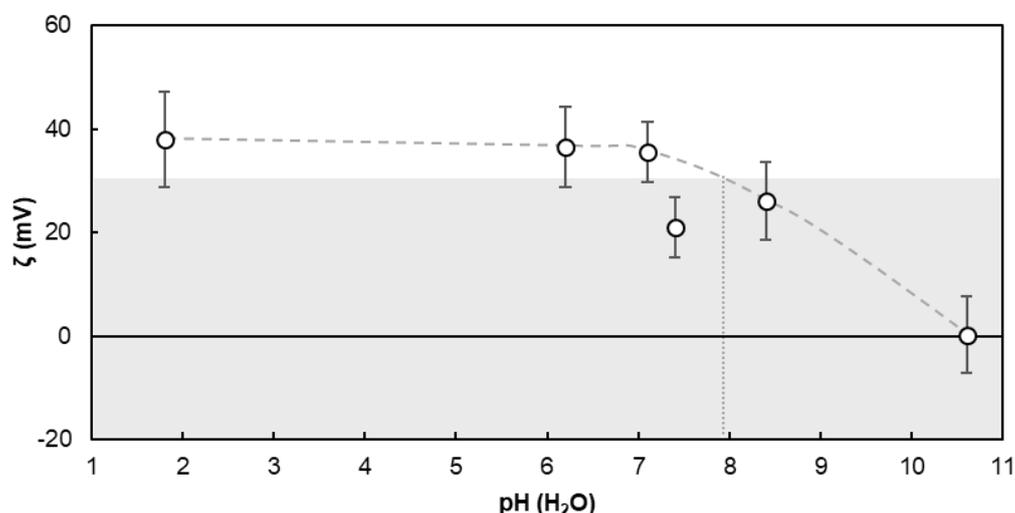
Figure 1.15. Evolution of 1 g/L TiO$_2$ P25 nanoparticle zeta potential with pH in methanol referred to the aqueous pH scale at 25.0 °C. pH is adjusted with HCl and NaOH. Ionic strength is kept constant at 10^{-3} M with NaCl. Instability area is marked in grey.

First of all, in contrast to aqueous media, it appears that whatever the pH value, ζ remains positive. A quite similar observation was made by Kosmulski et al. for TiO$_2$ in water/methanol mixtures: zeta potential became closer to 0 mV with increasing amounts of methanol.[120] On the other hand, it slightly increases with the addition of HCl, whereas NaOH addition brings ζ closer to 0 mV but not in the negative values. Dissociative adsorption of methanol on a

hydrated TiO_2 surface is described by the following equations.[121]

$$MeOH \rightarrow MeO^-_{adsorbed} + H^+ \tag{15}$$

$$H^+ + O^{2-}_{surface} \rightarrow OH^-_{adsorbed} \tag{16}$$

$$MeOH + O^{2-}_{surface} \rightarrow MeO^-_{adsorbed} + OH^-_{adsorbed} \tag{17}$$
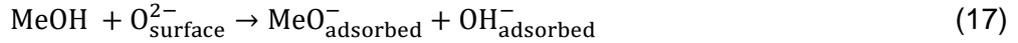
As a result, surface hydroxyl groups are partially replaced by methoxyl groups, as shown in Figure 1.16, which are not able to donate protons. The creation of negative surface charges in the presence of NaOH is thus reduced.
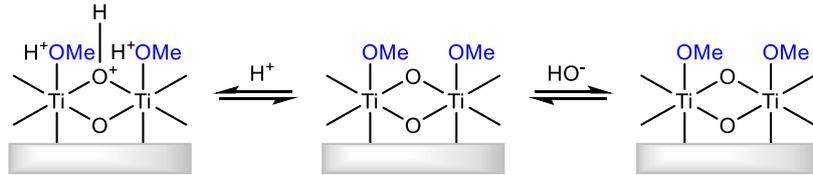
Figure 1.16. Simplified illustration of electric charges formation on TiO_2 surface by acido-basic reactions in the presence of methanol.

Moreover, in the presence of electrolytes, there exists a charge screening from Na^+ and Cl^- ions. It is assumed that Cl^- ions interact with the surface through $TiOH_2^+Cl^-$ interactions, whereas Na^+ ions interact in a non-specific way with negatively charged groups on TiO_2 surface.[121] When NaOH is added to methanol dispersions, $TiOH_2^+$ disappears from the surface and Cl^- ion adsorption decreases while Na^+ ion adsorption ability remains identical. This may explain why the zeta potential remains positive whereas surface potential may be negative. Figure 1.17 displays the evolution of TSI^{top}_{2h} and Stokes diameters determined with the Turbiscan as a function of the zeta potential ζ in methanol.
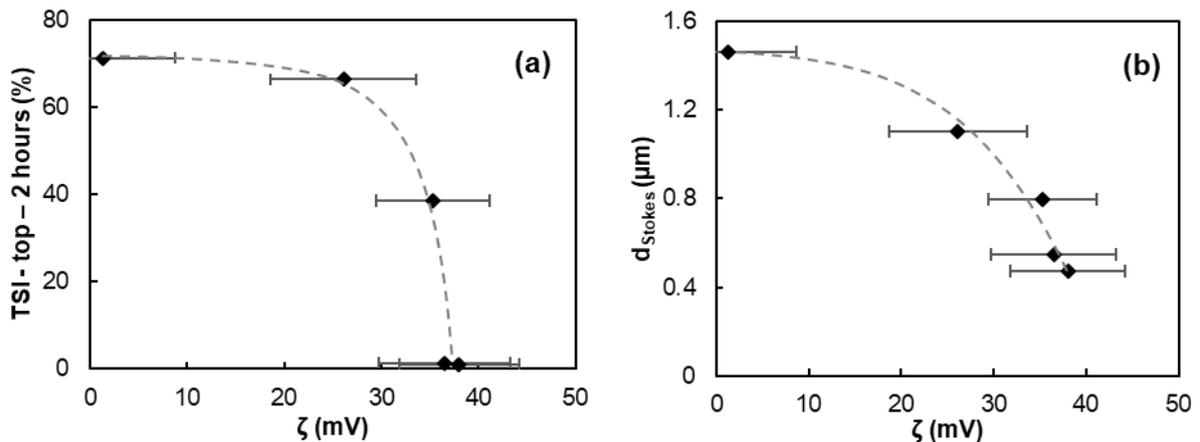
Figure 1.17. Evolution of (a) TSI at the top of the cell after 2 hours and (b) Stokes diameters d_{Stokes} as a function of the zeta potential of 1 g/L TiO_2 P25 particles in methanol. Ionic strength is kept constant at 10^{-3} M with NaCl and pH is adjusted with NaOH or HCl.

As in water, TSI^{top}_{2h} and particle diameters vary depending on the zeta potential suggesting heavily that the stability of TiO_2 dispersions in methanol is mainly influenced by electrostatic repulsions. Actually, it is well established that DLVO theory is also applicable in polar organic

solvents and may explain the stability of dispersions provided that the dielectric constant is high enough. However, for solvents of lower permittivity ε', electrostatic repulsions significantly decrease even when zeta potential is relatively high.

It can be assumed that these types of interactions occur in others non-aqueous solvents, and directly impact the stability of particle dispersions. In the following section, both the DLVO and the HSP contributions to stabilization are considered to rationalize the behaviour of TiO_2 dispersions in non-aqueous solvents.

5.2. Complementarity of HSP and DLVO approaches to rationalize stability
5.2.1. DLVO observations in non-aqueous solvents

The same way that in water and methanol, zeta potential ζ was measured in samples containing 1 g/L TiO_2 P25 nanoparticles dispersed in a series of 17 "pure" solvents, i.e., of the highest purity commercially available and free of any additional compound. The analysis of the transmitted and backscattered signals recorded by the Turbiscan for two hours provides the average diameter of the aggregates (d_{Stokes}) as well as the so-called Relative Turbiscan Stability Index ($RTSI_{2h}^{top}$), which is calculated from the *TSI* by considering the viscosity and density of the solvents, see equation (12). Turbiscan profiles are available in Appendix A4. Solvents are listed in Table 1.1 according to the increasing size of the agglomerates.

For comparison purposes, size measurements obtained by DLS (d_{DLS}) and obtained by sedimentation rate (d_{Stokes}) were compared. Results are presented in Figure 1.18. It appears that d_{DLS} is globally larger than d_{Stokes} for most of our samples. Indeed, DLS measurements are based on the variations of light scattering by the sample, which changes as particles are subject to Brownian motion. However, sedimentation also occurs in the meantime and increases the variation rate of light scattering signal modification, making the particles look
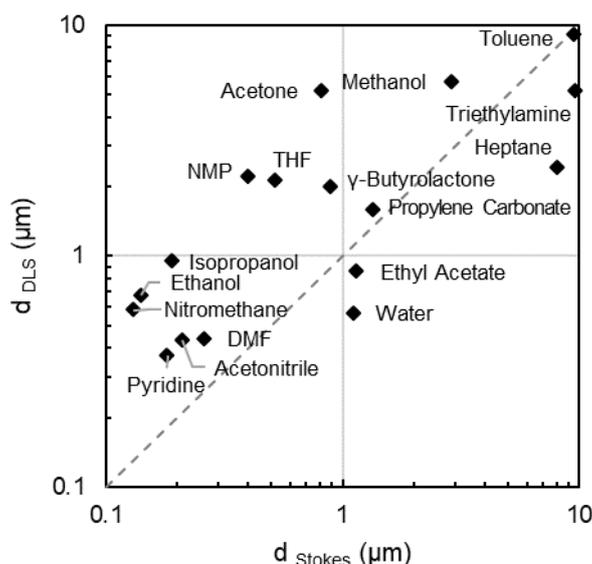
Figure 1.18. Comparison of DLS and Stokes diameters for TiO_2 P25 dispersions in nonaqueous solvents used as received.

larger than they are. Also, DLS measurements are based on the hypothesis that light scattering objects are spherical, but in our case, particles are mostly agglomerates and most probably not spherical. For the following results, only d_{Stokes} was considered as it reflects the destabilization of samples, which is a key information in this study.

Table 1.1. Stability results and physicochemical characteristics of TiO$_2$ dispersions (1 g/L) in various "pure" solvents at 25 °C: viscosity (η), density (ρ), dielectric constant (ε'$_r$), zeta potential (ζ), Relative Turbiscan Stability Index (RTSI), Stokes diameters (d$_{Stokes}$) and total interaction energy $(V_R+V_A)_{max}$ calculated according to equations (1)-(5). Calculation details are given in appendix A2.

Solvent	Structure	η (cP)	ρ (g/cm³)	ε'$_r$	ζ (mV)	$RTSI_{2h}^{top}$	d$_{Stokes}$ (µm)	$\dfrac{(V_R + V_A)_{max}}{k_BT}$
Nitromethane		0.67	1.13	39.0	-31±9	0.3	0.13	2.1
Ethanol		1.22	0.82	24.5	31±1	0.4	0.14	1.1
Pyridine		0.88	0.98	12.4	44±10	0.2	0.18	0.9
Isopropanol		2.1	0.79	17.9	-70±9	0.8	0.19	5.8
Acetonitrile		0.37	0.79	37.5	-34±10	0.9	0.21	2.5
DMF		0.92	0.94	36.7	43±14	0.8	0.26	4.6
NMP		1.67	1.03	33.0	-7±5	0.3	0.40	< 0
THF		0.95	0.98	7.6	19±2	1.4	0.53	< 0
Acetone		0.32	0.79	20.7	-26±2	7.2	0.81	0.2
γ-Butyrolactone		1.75	1.13	41.0	-4±2	13.1	0.89	< 0
Water		0.89	1.00	80.1	-5±4	14.5	1.11	< 0
Ethyl Acetate		0.46	0.9	6.0	44±12	2.2	1.14	< 0
Propylene Carbonate		2.8	1.2	64.9	-46±6	28.4	1.34	11.0
Methanol		0.54	0.79	32.7	11±13	10.8	2.88	< 0
Heptane		0.42	0.68	1.9	-10±3	5.6	8.05	< 0
Toluene		0.55	0.89	2.4	-30±8	9.7	9.47	< 0
Triethylamine		0.36	0.73	2.4	6±6	4.8	9.59	< 0

Most of the observed stabilities can be rationalized by the DLVO theory. It thus appears that the six solvents leading to the finest particles (d$_{Stokes}$ < 0.3 µm in Table 1.1), namely nitromethane, ethanol, pyridine, isopropanol, acetonitrile and dimethylformamide, correspond to particles with a high zeta potential (|ζ| > 30 mV) dispersed in quite polar solvents (ε'$_r$ > 10).

On the contrary, the seven solvents in which the aggregates are the largest ($0.9 < d_{Stokes} < 10$ µm) have either a very low dielectric constant ($\varepsilon'_r < 10$), namely ethyl acetate, heptane, toluene and triethylamine, or a low zeta potential ($|\zeta| < 6$ mV) insufficient for the electrostatic repulsion to dominate the Van der Waals attraction, namely, γ-butyrolactone, water and methanol. Three solvents (N-methyl pyrrolidone, tetrahydrofuran and acetone) have an intermediate behaviour because either their dielectric constant or the zeta potential of the particles is slightly lower than the threshold values defined above.

Likewise, the evolution $RTSI_{2h}^{top}$ (Table 1.1) follows the same trend as the best solvents exhibit low values ($RTSI_{2h}^{top} < 1$) whereas the less efficient solvents have very high values ($RTSI_{2h}^{top} > 10$). However, two solvents fail to fall into this general framework of explanation: N-methyl pyrrolidone (NMP) in which the dispersion is stable although the electrostatic repulsion is negligible ($|\zeta| = 7$ mV), and propylene carbonate which meets both stability criteria ($\varepsilon'_r = 64.9$ and $|\zeta| = 46$ mV) but in which the dispersion is extremely unstable. The presence of these outliers indicates that other phenomena not considered in the DLVO theory are involved in the stability of TiO$_2$ dispersions. To analyse the influence of non-DLVO forces, the alternative approach of Hansen's solubility parameters (HSP) was next investigated.

5.2.2. Limiting DLVO forces to measure HSP

The HSP method is a pragmatic and versatile tool originally developed to facilitate the finding of solvents able to dissolve paint resins. The principle of the method is based on the idea that "like dissolves like", which means that a solvent should effectively dissolve a solute provided it resembles it. This concept was quickly extended to solid/liquid dispersions to help formulators in designing the most suitable media for dispersing pigments. This variant is based on the "like disperses like" principle assuming that particles disperse better in solvents having high affinity for the surface.

In the HSP approach, only three types of interaction between the particle and the surrounding medium are considered, namely hydrogen bonding, dipolar interactions and dispersive interactions due to London forces. The attractive interactions of Van der Waals are considered in both theories HSP and DLVO. On the other hand, electrostatic repulsions are only considered in DLVO theory while hydrogen and dipole bonds are only considered in Hansen's approach. The decisive impact of electrostatic stabilization is clearly established in the case of charged TiO$_2$ particles dispersed in polar solvents, but it is unable to explain the stability observed for some TiO$_2$ dispersions when $|\zeta|$ is low.

Estimating the respective contributions of these complementary interactions would provide valuable information to rationalize the experimental results usually interpreted on the basis of only one of these theories. The difficulty to quantify the stability of a dispersion has been

brought up in the literature and Süß et al.[14] proposed to use analytical centrifugation and quantify the sedimentation rate. This technique allows analysing a large number of samples in a timely manner compared to gravitational sedimentation, but interparticle attraction and agglomeration cannot be detected by this technique. In this work, a Turbiscan device based on gravitational sedimentation is used, so as to detect the two phenomena involved in the destabilization of dispersions under ordinary storage conditions. Namely, on the one hand, the agglomeration of particles resulting from Brownian motion and Van der Waals attraction and, on the other hand, the sedimentation of particles and aggregates under the effect of gravity.

To focus on the influence of non-DLVO forces, the contribution of electrostatic repulsion to stabilization was minimized by decreasing the zeta potential of particles and removing the energy barrier. Thus, whenever $(V_R + V_A)_{max}/k_B T$ in "pure" solvents was positive (Table 1.1), 10^{-3} M acid or base was added and the zeta potential was measured again to verify that it is weak enough. Trifluoroacetic acid (TFA) and tetrabutylammonium hydroxide (TBAH) were chosen to allow counter-ion solubility in organic solvents and avoid ionic adsorption at the particle surface. TFA was expected to increase ζ whereas TBAH was expected to decrease ζ. Of course, when solvents included in Table 1.1 already have a negative $(V_R + V_A)_{max}/k_B T$, the stability of dispersions was interpreted according to HSP without adding TFA or TBAH. Turbiscan profiles of corresponding samples are presented in Appendix A5. Table 1.2 summarizes the experimental stabilities of the non-electrostatically stabilized dispersions which are interpreted on the basis of HSP. The solvents are listed in Table 1.2 according to the increasing size of the aggregates. Each solvent is then assigned a score of 1 (stable) or 2 (unstable) based on the *RTSI* and the diameter of the aggregates:

$$\text{Score 1: } d_{Stokes} < 0.31 \text{ µm or } RTSI_{2h}^{top} < 1.0$$

$$\text{Score 2: } d_{Stokes} > 0.31 \text{ µm or } RTSI_{2h}^{top} > 1.0$$

Two HSP spheres were determined using HSPiP software considering each of the stability criteria, i.e., d_{Stokes} or $RTSI_{2h}^{top}$ calculated based on different methods described hereafter. The first one based on the size of the agglomerates, $d_{Stokes,}$ is shown in Figure 1.19. The "fit" indicator, which can vary between 0 and 1, expresses the effectiveness of the sphere for modelling experimental results. The more misplaced solvents, the more the fit decreases. Here the fit is very good (0.90), since only one effective solvent (isopropanol) is excluded from the sphere and one poor solvent (acetone) is included.

Table 1.2. TiO$_2$ dispersions (1 g/L) in various solvents at 25 °C for which zeta potential (ζ) and total interaction energy *(V$_R$ - V$_A$)$_{max}$* are minimized by addition, when necessary, of 10^{-3} M TFA (a) or TBAH (b): Relative Turbiscan Stability Index ($RTSI_{2h}^{top}$), Stokes diameters (d$_{Stokes}$), zeta potential (ζ), Hansen solvent parameters and Relative Energy Difference (RED) for the Hansen's sphere based on d$_{Stokes}$.

Solvent	Structure	$RTSI_{2h}^{top}$	d$_{Stokes}$ (µm)	$\dfrac{(V_R - V_A)_{max}}{k_B T}$	ζ (mV)	δ$_d$	δ$_p$	δ$_h$	RED
Nitromethane [a]		0.30	0.12	< 0	-19±6	15.8	18.8	6.1	0.99
Acetonitrile [a]		0.20	0.14	< 0	-11±7	15.3	18	6.1	0.95
Isopropanol [a]		0.50	0.18	< 0	15±6	15.8	6.1	16.4	1.11
DMF [a]		0.90	0.25	< 0	12±1	17.4	13.7	11.3	0.64
Ethanol [b]		1.00	0.31	< 0	-14±11	15.8	8.8	19.4	0.99
NMP		0.30	0.4	< 0	-7±5	18.0	12.3	7.2	1.00
THF		1.40	0.53	< 0	19±2	16.8	5.7	8.0	1.36
Acetone [a]		8.10	0.58	< 0	-7±4	15.5	10.4	7.0	0.99
γ-Butyrolactone		13.10	0.89	< 0	-4±2	18.0	16.6	7.4	1.01
Propylene Carbonate [a]		4.50	1.09	< 0	-18±2	20.0	18	4.1	1.61
Water		14.50	1.11	< 0	-5±4	15.5	16	42.3	3.21
Ethyl Acetate		2.20	1.14	< 0	44±12	15.8	5.3	7.2	1.41
Methanol		10.80	2.88	< 0	11±13	14.7	12.3	22.3	1.01
Pyridine [b]		14.60	3.34	< 0	-24±11	19.0	8.8	5.9	1.49
Heptane		5.60	8.05	< 0	-10±3	15.3	0.0	0.0	2.38
Toluene		9.70	9.47	< 0	-30±8	18.0	1.4	2.0	2.21
Triethylamine		4.80	9.59	< 0	6±6	15.5	0.4	1.0	2.27

[a] the solvent contains 10^{-3} M trifluoroacetic acid

[b] the solvent contains 10^{-3} M tetrabutylammonium hydroxide

It is worth noting that with the first sphere (Figure 1.19), the two misplaced solvents are both located near the boundary separating effective and poor solvents. Thus, the poor solvent acetone is erroneously slightly inside the sphere with a RED of 0.99 while isopropanol (effective solvent) is slightly outside with a RED of 1.11. RED is the "Relative Energy Difference" defined by the relationship RED = R$_a$ / R where R$_a$ is the distance between the solvent and the sphere centre and R is the sphere radius. So, when RED < 1, the solvent is inside the sphere and when RED > 1, the solvent is outside the sphere.
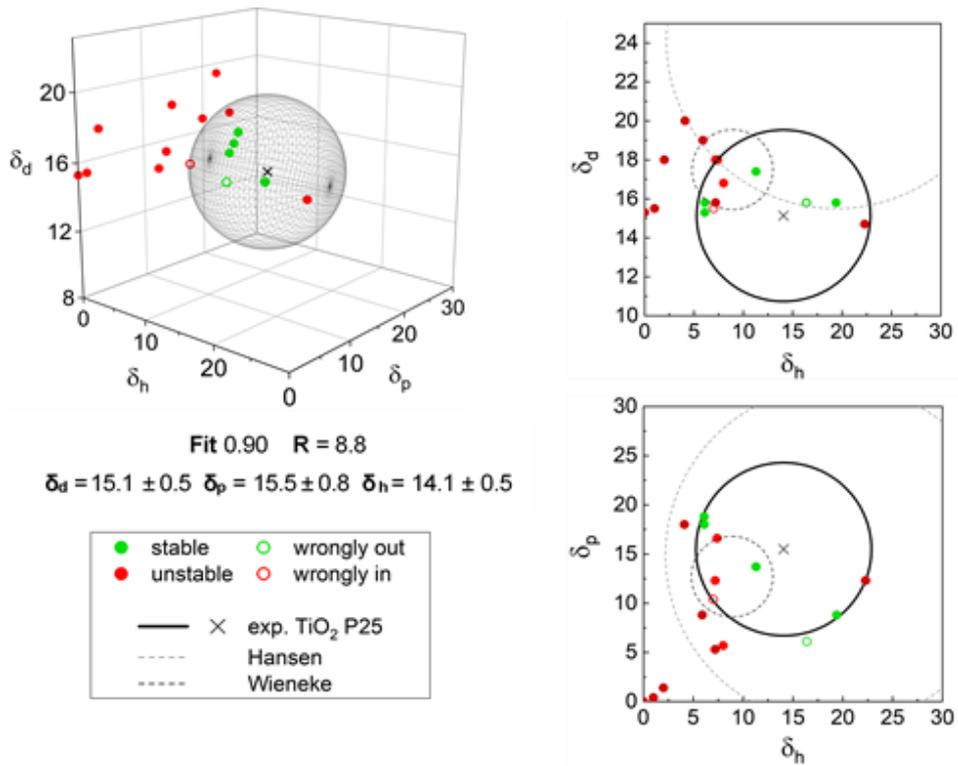
Figure 1.19. Solubility sphere and 2D projections of TiO_2 P25 1 g/L in 17 organic solvents based on Stokes diameters d_{Stokes} at 25 °C.
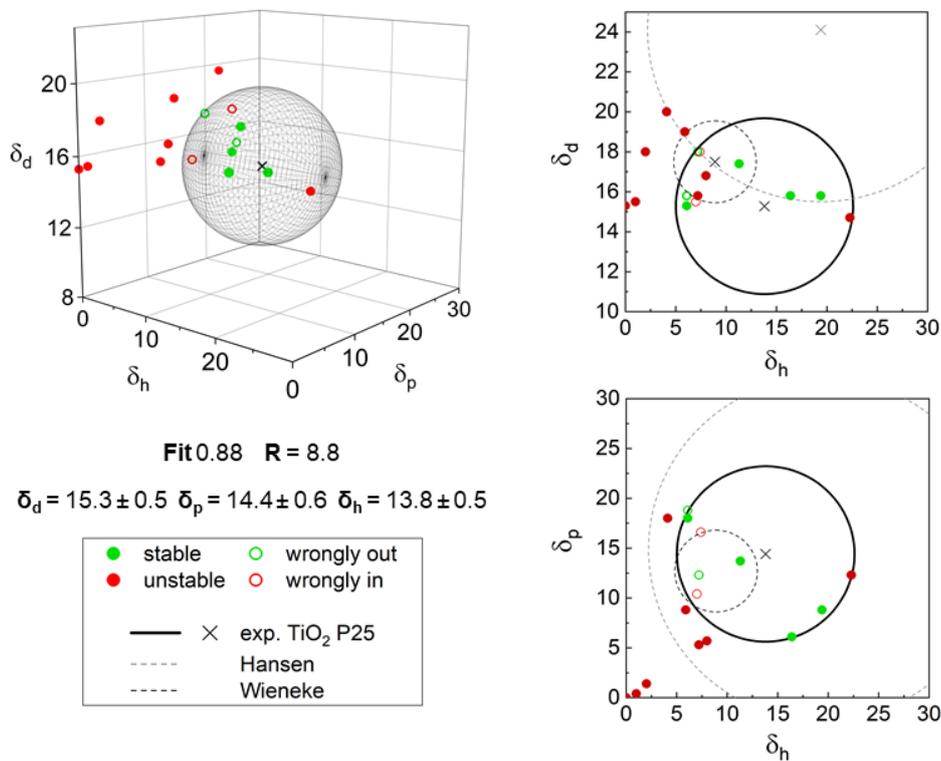
Figure 1.20. Solubility sphere and 2D projections of TiO_2 P25 1 g/L in 17 organic solvents based on $RTSI_{2h}^{top}$ at 25 °C.

The second sphere presented in Figure 1.20 and calculated from $RTSI_{2h}^{top}$ values is somewhat less satisfactory (fit = 0.88) since two effective solvents are outside the sphere and two poor solvents are inside. In both cases the radius of the sphere is 8.8 but the first sphere including all dispersions with d_{Stokes} < 0.31 μm appears to be the most reliable. In a study on the influence of different dispersion conditions on the size of aggregates, Jiang et al. showed that the ultrasonic probe Sonotrode provided aggregates of 155 nm but was unable to further break the TiO_2 P25 aggregates down to the elementary particle (25 nm).[12] In this work, we consider that particles size up to twice this minimal size provide the most stable dispersions. Beyond this size, a competition takes place between NP-solvent and NP-NP interactions that prevent the formation of smaller particles.

Other authors have characterized TiO_2 particles[24] and nanoparticles[19] using HSP. However, none of them considered electrostatic repulsion as a possible stabilization phenomenon. Comparative results are displayed in Table 1.3. Characteristic parameters of the sphere calculated by Hansen regarding TiO_2 pigments are very different from those determined in this study. However, Hansen investigated the paint pigment TiO_2 Kronos RN 57 at a concentration 20 times greater than that of the present work and no information is available regarding coating and particle diameter.[24] Actually, it is likely that this white pigment has a diameter close to 0.3 μm to maximize its opacifying power and that it is covered with a mixture of oxides to avoid the photocatalytic degradation of the organic paint film. Hence, the spheres are hardly comparable.

Table 1.3. Experimentally determined Hansen parameters (δ_d, δ_p, δ_h), of the center and the radius (R) of the solubility sphere and calculated Hildebrand parameter (δ_H) compared to literature values.

Method	δ_d [MPa$^{1/2}$]	δ_p [MPa$^{1/2}$]	δ_h [MPa$^{1/2}$]	δ_H [MPa$^{1/2}$]	R [MPa$^{1/2}$]
d_{Stokes}	15.1 ± 0.5	15.5 ± 0.8	14.1 ± 0.5	25.8 ± 0.9	8.8
$RTSI_{2h}^{top}$	15.3 ± 0.5	14.4 ± 0.6	13.8 ± 0.5	25.1 ± 0.8	8.8
Hansen [24]	24.1	14.9	19.4	34.3	17.2
Wieneke [19]	17.5	12.7	8.9	23.4	4.1

Differences with Wieneke's results could come from the nature of the solvents used in both cases. Moreover, Wieneke et al.[19] studied smaller TiO_2 NPs with an average particle size of 5.3 nm composed at 95 % of anatase and 5 % of rutile with a specific surface area of 265 m²/g which is much greater than that of TiO_2 P25. However, the sphere obtained by Wieneke et al. is mostly included inside the one calculated in this work and the sphere radius is more than twice smaller. Also, as TiO_2 NPs get smaller, interparticle forces get stronger, making dispersion harder to achieve at the expense of solvent-particle interactions.[23] Another impacting factor is the sample concentration which was of 0.015% for Wieneke's study as the

visual changes in stability are easier to detect with the naked eye in dilute samples. In their case, coagulation was evaluated through the colour change from bluish to white, corresponding to a change in light scattering behaviour from Rayleigh to Mie scattering.[19] Using the Turbiscan, quantitative descriptors allow detecting destabilization before the naked eye could detect anything.

The Turbiscan-based method developed in this work is more reliable than those reported in the literature to determine HSP of TiO_2 nanoparticles. Indeed, the phenomena occurring during the destabilization, namely the coagulation of particles and the sedimentation of aggregates, can be observed separately. Moreover, this work clearly shows that, when particles are charged and dispersed in polar solvents, the electrostatic repulsion must absolutely be considered to rationalize the observed stabilities using DLVO theory. As electrostatic repulsions are not considered in Hansen's theory, only dispersions of weakly charged particles in media of low polarity can correctly be interpreted on the basis of HSP. Therefore, it makes no sense to interpret the stability of DLVO stabilized samples using HSP theory. On the other hand, the analysis of the stability of the dispersions jointly by Hansen's and DLVO theories provides complementary information allowing a more accurate interpretation of the results and highlights the solvents exhibiting a significant affinity for the particles' surface.

6. Conclusions

Hansen Parameters have been used for rationalizing and predicting the stability of titanium dioxide particles dispersions.[14,19,24] However, interparticle electrostatic interactions, not considered in Hansen's approach but considered in the DLVO theory, play a significant role in organic solvents having a notable dielectric constant.[16,115,116] HSP interpretation for NPs dispersions can be achieved provided that electrostatic repulsion is negligible, in particular regarding inorganic NPs that can easily acquire surface charges.

In water, it was shown that the Turbiscan detected high *TSI* (Turbiscan Stability Index) and Stokes diameters when zeta potential was elevated, in accordance with the DLVO theory. In organic solvents, it was intended to interpret the stability using HSP. This model was shown insufficient to describe with fidelity the dispersions stability since no sphere could fit the experimental observations. The relevance of DLVO interpretation came in to play as total repulsive energy $(V_R + V_A)_{max}$ was non-negligible in numerous solvents. When the zeta potential of NPs and the dielectric constant of the solvent are both high, the dispersion benefits from additional stabilization while when the electrostatic repulsion is negligible, only the solvents within a Hansen dispersion sphere give stable dispersions. The two interpretations are therefore complementary to describe the behaviour of TiO_2 dispersions in organic solvents.

Using a zeta dip cell, with higher chemical compatibility than the regular folded capillary cells, it was possible to measure zeta potential in organic solvents and assess the strength of the electrostatic contribution to stabilization. Solvents were scored according to Turbiscan indicators, namely $RTSI_{2h}^{top}$ (Relative Turbiscan Stability Index after 2 hours at the top of the cell) and d_{Stokes} calculated from the sedimentation rate to yield the TiO_2 P25 Hansen Solubility Sphere. A comparison with DLS size measurements showed notable differences as particle sedimentation disturbs DLS measurements. In the case of stability study, sedimentation-based size measurements are more indicated. The spheres obtained based on d_{Stokes} scoring (R = 8.8, δ_d = 15.1, δ_p = 15.5, δ_h = 14.1 MPa$^{1/2}$) and $RTSI_{2h}^{top}$ scoring were compared to that of Hansen[24] and that of Wieneke et al.[19] for TiO_2, both determined based on visual scoring. In this study, both zeta measurement and Turbiscan monitoring of stability were complementary. $RTSI_{2h}^{top}$ and Stokes diameters are quantitative descriptors avoiding an empirical visual rating of dispersions stability. This standardized HSP determination method could be extended to other types of NPs such as metal oxides, but carbon-based NPs dispersions, such as graphene for instance, are expected to behave differently as surface charging mechanisms are different.

Chapter 2

❖

Predicting hydrophobicity of complex oils using Neural Networks and Graph Machines
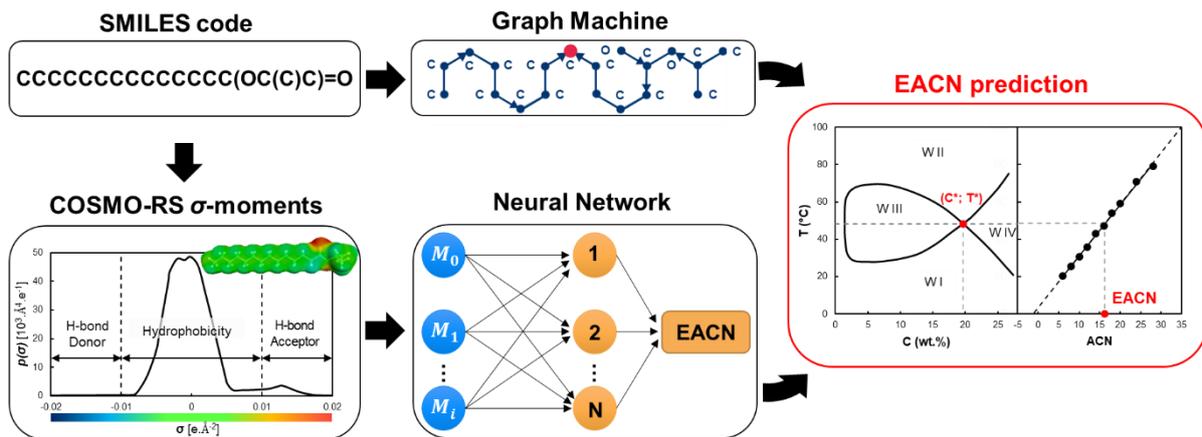
SMILES code

CCCCCCCCCCCCCC(OC(C)C)=O

Graph Machine

COSMO-RS σ-moments

$\rho(\sigma)$ [$10^3 \cdot \mathring{A}^4 \cdot e^{-1}$]

H-bond Donor · Hydrophobicity · H-bond Acceptor

σ [e.$\mathring{A}^{-2}$]

Neural Network

M_0 M_1 ⋮ M_l

1 2 ⋮ N

EACN

EACN prediction

W II

W III

(C^*; T^*)

W IV

W I

T (°C)

C (wt.%)

ACN

EACN

1. Introduction

A large diversity of natural and synthetic liquid compounds are grouped under the generic term "oils" due to their non-miscibility with water. They are key components of Surfactant/Oil/Water (SOW) systems such as swollen micelles, microemulsions or emulsions which are found in numerous industries and various end-use products. For instance, the physico-chemical characteristics of oils are essential in the petroleum industry dealing with crude oil extraction and refined oil properties. Enhanced oil recovery (EOR) has been driving the research on oil solubility in SOW systems since the 1970's. As about 70% of crude oil remains trapped in the reservoir rocks after regular oil extraction, solubilizing the remaining crude oil to extract it represents a tremendous financial interest. Another area with great interest in the study of oil behaviour is the cosmetics industry. Indeed, cosmetic products contain various types of oils acting as emollients or moisturizing agents that can be formulated in emulsions, gels, balms and other forms. Also, in the perfumery industry, most perfume molecules are non-miscible to water. For all these applications, the quantitative evaluation of the hydrophobicity of the oil is extremely important because it allows choosing the most effective SOW system composition and, *in fine*, optimizing its performances in applications. Due to health and safety concerns, a major tendency in end-use products is to substitute ingredients based on their innocuousness and low environmental impact. To that end, replacing an oil by another one having the same characteristics avoids changing others components to yield the same applicative properties in the final product.

Several concepts have thus been developed to characterize the hydrophobicity/polarity of oils in SOW systems. The best known are log P, i.e. the logarithm of the *n*-octanol-water partition coefficient[122] and the "required HLB" (Hydrophilic-Lipophilic-Balance) of Griffin.[123,124] A more relevant concept to quantify the hydrophobicity of an oil in SOW systems was introduced in 1977 by Wade et al. as a dimensionless number: the so-called EACN (Equivalent Alkane Carbon Number).[125,126] It corresponds to the number of carbon atoms of the *n*-alkane which exhibits a phase behaviour similar to that of the oil under consideration.[127] In practice, measuring accurately the EACN value of an oil is tedious. The standard method is based on the elaboration of the so-called "Fish diagrams" which represent the phase behaviour of equilibrated SOW systems where the surfactant is a well-defined polyethyleneglycol monoalkyl ether and the water-to-oil ratio (WOR) is equal to 1.[128]

While reliable and accurate, the experimental determination of EACNs from fish tail diagrams is, however, a lengthy process (from one to several weeks) which is limited by experimentally accessible conditions in terms of temperature (T ≈ 5 to 80 °C). Thus, *in-silico* estimation of the EACN values of oils without any experiments would be considerably time saving.

To date, a few predictive models of EACN values have been reported. EACN value of complex oil mixtures, i.e. crude oil, was predicted by Creton et al. using an evolutionary algorithm coupled to data mining.[129] Bouton et al. built a QSPR model by applying genetic algorithms to structural molecular descriptors of polar hydrocarbon oils.[130] A multilinear regression based on the σ-moments calculated by the COSMO-RS approach[131,132] was applied to polar hydrocarbons[133] and aprotic polar oils[134] by Lukowicz et al. These works showed that depending on the chemical functions of molecules, the relevant descriptors differ and EACN estimations were less satisfactory in the case of polar oils.

On the other hand, a variety of chemical and physicochemical properties such as surface tension[135], viscosity[136], flash point, cetane number of fuels,[137] bioactivity of drugs[138,139] and other thermodynamic properties[140–142] can be predicted accurately with graph machines (GM)[143] and neural networks (NN)[144]. Both these theoretical tools are non-linear models that learn a pathway from input values to a resulting output. For neural networks, that are basically standard multi-layer perceptrons (MLPs), the inputs are either measured or computed from molecular simulations, while for graph machines the inputs are the 2D molecular structures entered as their SMILES (Simplified Molecular Input Line Entry Specification) codes.

In this chapter, two approaches for predicting the EACN of functionalized oils are reported using NN and GM. To that goal, a set of 111 molecules with a reliable experimental EACN was gathered either from literature or from our laboratory database.[127,128,130,133,134,145–147] A GM regression based solely on the readily accessible molecular SMILES codes and a NN regression using as inputs COSMO-RS computed σ-moments are designed for the 111 molecules. After a selection step of the optimal model in each case, predictions are performed on a test set of ten cosmetic or perfumery molecules for which experimental EACN have been determined. The respective reliability of the two models is finally evaluated by predicting the EACN of compounds belonging to nine homologous series.

2. Bibliographic background
2.1. Characterizing oil hydrophobicity

Hydrophobicity is a way to express the non-miscibility of two liquids such as oil and vinegar. Thermodynamically, the miscibility of two liquids is rationalized by the free Gibbs energy of mixture, ΔG_{mix}. A closed system at constant temperature and pressure tends to evolve towards the lowest energy state, i.e. the minimization of ΔG_{mix}. When mixed, two liquids are miscible when $\Delta G_{mix} = \Delta H_{mix} - T\Delta S_{mix} < 0$. The enthalpic term of mixing ΔH_{mix} can be either positive or negative for linear alkanes.[148] However, when introducing a non-polar molecule such as an alkane into bulk water, a cavity is formed, causing water molecules to reorganize to a higher organization state around the alkane molecule. This is accompanied by a decrease in entropy

and $\Delta S_{mix} < 0$, which is unfavored. The importance of the entropic term makes the miscibility unfavored, such as $\Delta G_{mix} > 0$. This is known as the hydrophobic effect.[148,149] In his model, Winsor separated the affinity between oil and water as the hydrogen bonds contribution (H, hydrophilic) and the low energy Van der Waals contribution (L, lipophilic) according to the following equation.[150]

$$A_{OW} = A_{H,OW} + A_{L,OW}$$

However, quantifying hydrophobicity by thermodynamic values would require measuring hardly accessible values. Instead, several ways to approach hydrophobicity quantification were developed. The most widely known and used are the partition coefficient log P and the required HLB of an oil, which are further explained hereafter.

2.1.1. Partition coefficient Log P

The determination of the partition coefficient Log P of a solute is a popular method to assess the hydrophobicity of a chemical compound. More exactly, it assesses the preferential solubilization of the compound either in water or in 1-octanol. Log P is widely used in environmental and medical sciences since it expresses the ability of a non-ionizable substance to partition between aqueous and lipophilic compartments of organisms. However, this value gives no indication regarding the behaviour of an oil at the O/W interfaces in SOW systems since no surfactant is considered in this approach.[122]

2.1.2. Required HLB of Griffin

The required HLB values of an oil, introduced by Griffin,[123,124] correspond to the HLB values of the mixture of surfactants providing the most stable emulsions with the oil under study.[123,151] For one oil, there exist two required HLB values corresponding to O/W (oil in water) or W/O (water in oil) emulsions.[152] In practice, one should prepare emulsions containing the oil under study and a mixture of two surfactants with a span of HLB values. The most stable emulsions are formed with the surfactant mixture of the oil required HLB. This method is sometimes used to classify complex liquids such as essential, vegetable, or animal oils.[153–157] However, the required HLBs thus obtained are imprecise and poorly reproducible because they depend on the mixture of surfactants used in the study and on the emulsification process. Furthermore, the required HLB concept is based on the HLB of surfactants which is itself an approximate empirical parameter only reliable for polyethoxylated nonionic surfactants.

2.1.3. The Equivalent Alkane Carbon Number (EACN) concept

In the 1970's, the research about enhanced oil recovery triggered many key findings. In order to harvest the most quantity of crude oil from reservoir rocks, the mechanisms of surface tension and oil solubilization involved in SOW systems were investigated. Depending on the affinity of the surfactant for either the aqueous or the oil phase, microemulsions are formed

and can be water-rich (Winsor I phase), oil-rich (Winsor II) or equilibrated in oil and water (Winsor III). At high amounts of surfactant, a total Winsor IV microemulsion solubilizing all of the oil and water is formed.

Interfacial tensions in SOW systems with various *n*-alkanes were measured by Cash et al. and Wade et al., giving birth to the EACN concept. Actually, an oil or oil mixture's EACN value is the length of the linear alkane for which the system reaches the lowest surface tension in the same conditions of temperature, salinity and using the same surfactant.[125,126] These works were based on the EACN additivity principle, meaning that the EACN of a two-oil stoichiometric mixture is supposed to be the average value of each oil's EACN.[126] Salager et al. showed in 1979 that the minimal interfacial tension phenomenon, which is the optimal condition for enhanced oil recovery, coincides with the Winsor III microemulsion system and was referred to as the "optimal formulation" in the oil industry.[158] The EACN value of an oil expresses its ability to penetrate the interfacial film of SOW systems and to modify its spontaneous curvature.[159–161] In the case where the surfactant is a polyethoxylated fatty alcohol C_iE_j, some molecules of oil penetrate the interfacial film according to their affinity for C_iE_j molecules. In particular, when the oil has a polar function, its affinity for the film is stronger than apolar oils and its EACN is much lower than N_C, its number of carbon atoms. Indeed, Figure 2.1 illustrates the identical Winsor phase behaviour of octyl octanoate ($N_C = 16$) and *n*-octane ($N_C = 8$), which is the linear alkane having an ACN equal to the EACN of the ester.
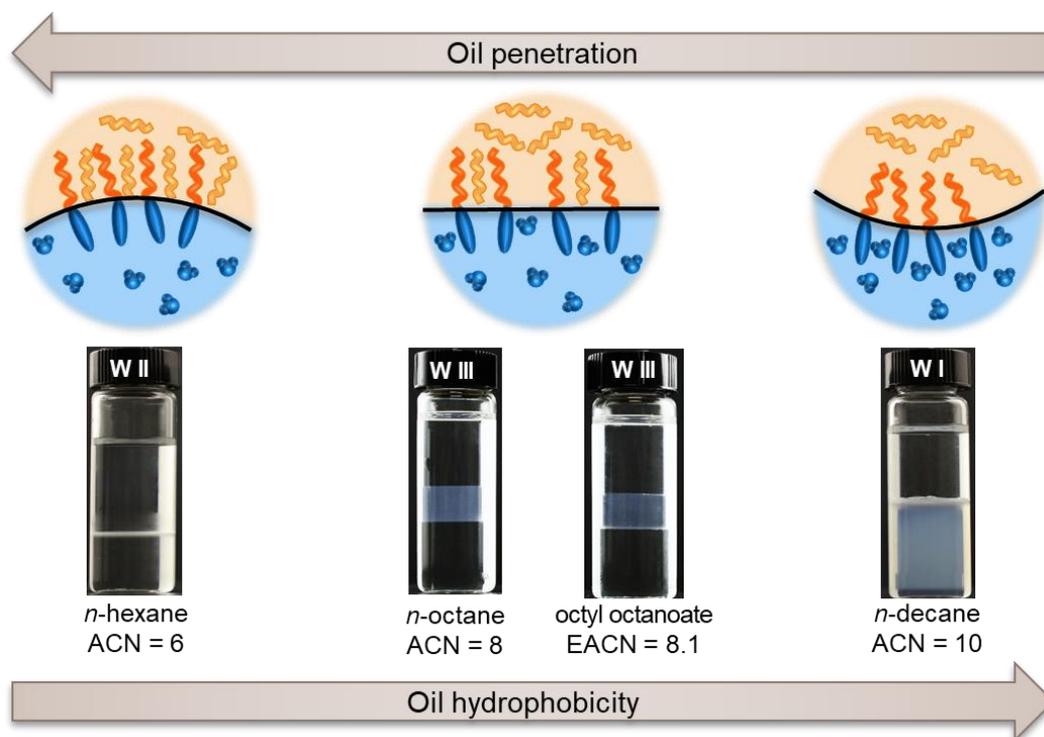
Figure 2.1. Effect of oil penetration on the spontaneous curvature of the interfacial film (top) and $C_{10}E_4$/Oil/Water (water-to-oil ratio = 1 (v/v)) microemulsion systems equilibrated at 25.0 °C yielding Winsor II, Winsor III and Winsor I microemulsions respectively (bottom). Systems with *n*-alkanes contain 3% $C_{10}E_4$ and the one with octyl octanoate contains 7% $C_{10}E_4$.

In the initial work of Cash et al., the EACN of alkylcyclohexanes and alkylbenzenes were characterized.[126] Over time, several categories of molecules were studied to determine their EACN such as esters and triglycerides,[147] aliphatic, aromatic and chlorinated hydrocarbons[127] and terpenes.[146] This concept also has been used to describe the partition of polar oils[162] and perfume molecules[160] in mixtures with less polar oils, that do not behave as ideal mixtures.

2.1.4. EACN experimental determination

In practice, measuring accurately the EACN value of an oil requires the standard method based on the elaboration of the so-called "Fish diagrams". It represents the phase behaviour of equilibrated SOW systems where S is a well-defined polyethyleneglycol monoalkyl ethers (C_iE_j) and the water-to-oil ratio (WOR) is equal to 1.[128] A typical fish diagram of a system $C_{10}E_4$/Oil/Water is given in Figure 2.2a as an example.
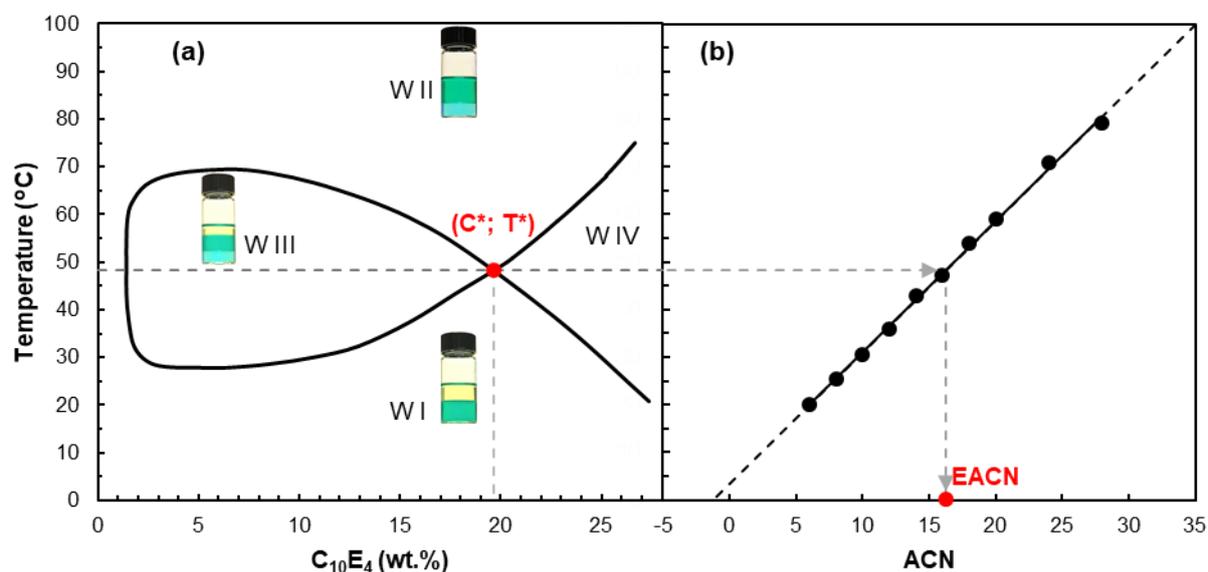
Figure 2.2. Determination of the EACN of an oil from the fish plot of the $C_{10}E_4$/Oil/Water-T system (a). The temperature of the fish-tail point indicated in red is reported to the calibration straight line obtained with a series of $C_{10}E_4$/n-Alkanes/Water-T systems (b).[127]

When varying the temperature and the surfactant concentration, the C_iE_j/oil/water-T systems provide different types of microemulsion behaviours (Winsor I, II, III and IV) depending on the affinity of the surfactant for water and oil. When it is balanced, a Winsor III microemulsion is formed giving a diagram shaped like a fish. The characteristic temperature T* at the intersection of the Winsor III and Winsor IV (i.e. one single phase microemulsion) regions, is then compared to the T* values of a series of n-alkanes (Figure 2.2b) to determine the EACN of the oil which expresses its hydrophobicity.[127,163]

The EACN concept is of interest only if the values assigned to oils do not depend on the nature of the C_iE_j surfactant used for its measurement. This key issue has been checked by Bouton et al. who showed that the EACN values of 26 terpenes and non-linear (branched, unsaturated, cyclic) hydrocarbons were identical within 0.3 unit regardless of the surfactant used namely

C_6E_4, C_8E_4 or $C_{10}E_4$.[130] However, for very polar oils, two major problems decrease the accuracy of EACN measurements. The first one comes from the fact that for oils having an EACN lower than 6, the calibration curve established with *n*-alkanes must be extrapolated to the dotted parts of the regression straight line (see Figure 2.2b). Accordingly, the lower the EACN, the greater the uncertainty over its estimated value. The second problem arises from the monomeric solubility of C_iE_j surfactants in the oil phase which increases the apparent polarity of the oil. As a result, the EACNs measured with short C_iE_j such as C_6E_4 tend to be lower than the EACNs measured with a long C_iE_j whose monomeric concentration in the oil phase is significantly lower. This difficulty was encountered while seeking to model the EACN of diisopropyl ether for which we had previously assigned an EACN equal to 2.2[128] on the basis of the fish diagram determined by Wormuth et al. with the $C_{12}E_6$/Diisopropyl ether/Water system.[164] According to our very first models (GM and NN) the EACN of this oil appeared as an outlier. The EACN of this ether was thus measured again using the same amphiphile (C_6E_4) that was used to measure the EACNs of most other highly polar oils.[133] The new value of EACN thus determined (0.6, see Appendix A6) is, as expected, significantly lower than the previous value and perfectly consistent with the EACNs of other very polar oils as they were determined with the same surfactant (C_6E_4). This revised EACN value has therefore been used to fit our GM and NN models.

2.2. Modelling and predicting physicochemical properties

Quantitative Structure Property/Activity Relationship (QSPR / QSAR) models have been used for over a century to quantitatively correlate molecular properties, represented by descriptors, to macroscopic values. In 1868, Brown and Fraser[165] were the first ones to conceptualize the existence of a quantitative relation between chemical structure and biological activity. Their model attempted to define a mathematical function linking the constitution of poisonous chemicals to their physiological action.[165] The respective contributions of Hansch[166] and Free and Wilson[167] helped developing the currently used QSPR methods. Hansch[166] proposed a model linking biological activity to hydrophobic, electronic and steric properties. Free and Wilson[167] implemented group contribution models for biological activity: each structural group corresponds to an empirical tabulated value. The end value is then the sum of all group contributions.

More recently, the European legislation REACH, aiming at improving the safety in relation with chemicals use, encourages the use of QSAR models as a non-testing approach to avoid handling hazardous chemicals.[168] Nowadays, thanks to much greater calculation capacities, QSPR models also apply to a broad spectrum of physicochemical properties. Molecular descriptors are numerous: the most simple ones are constitutional, 2D ones are topological,

and the most sophisticated are geometric or even quantic but require longer calculation time.[169,170]

Building a QSPR model relies on finding the best relation between a group of descriptors and a target property. Those models require the construction of a reliable database, consisting of entry/output pairs where entries are molecular descriptors and outputs are the target properties.[170] A model is then developed from this database, validated and tested to assess its predictive capacity.[169] Databases are generally divided into the learning base used to build the model on one hand, and the test base to validate its predictive capacity on the other hand. QSPR models can be linear or non-linear models. The most trivial one is the Multiple Linear Regression (MLR) method. However, some properties require a more complex modelling and multiple non-linear regressions with Neural Networks (NN) can be adapted. In this section, we first describe how the COSMO-RS method can be used to obtain molecular descriptors from quantic/thermodynamics calculation. The linear and non-linear modelling methods are then developed. Finally, the principle of Graph Machines (GM) based on topological information is presented.

2.2.1. COSMO-RS

COSMO-RS (COnductor like Screening MOdel for Real Solvents) is a first principle theoretical model based on a combination of quantum chemistry and statistical thermodynamics that serves to estimate, without any prior experience (*ab initio*), a large number of chemical properties, *e.g.* flash point, solubility or pK$_A$ based on thermodynamics laws.[132,171,172]

Due to the presence of polar covalent bonds, molecules carry a surface charge density σ on its so-called "σ-surface", which corresponds to the slightly inflated van der Waals surface. The "σ-profile" p^X(σ) of a molecule X is the curve obtained by smoothing the histogram of surface portions grouped by charge density in the interval [σ − dσ/2, σ + dσ/2].[131] Examples in the case of *β*-ionone and isopropyl myristate are represented in Figure 2.3. Using the COSMOconf software (version 4.3), the lower energy conformations of a molecule in the bulk liquid state are calculated. These conformations are then used as inputs in the COSMOtherm software (version 19.0.4) allowing the calculation of the σ-surface, σ-potential and σ-moments. Klamt[132] has shown that any partition coefficient K can be very well expressed as a Taylor-like development of σ-moments as defined by equation (18). It is estimated that a development up to m equal to six σ-moments is sufficient to satisfactorily express the partition coefficient K according to equation (18).

$$RT \ln K = c_{acc}M_{acc}^X + c_{don}M_{don}^X + \sum_{i=0}^{m} c_i M_i^X \qquad (18)$$

The σ-moments M_i^X are calculated from the σ-profile $p^X(\sigma)$ of the studied compound X according to equations (19)-(21).

$$M_{acc}^X = \int_{+\sigma_{HB}}^{+\infty} p^X(\sigma)(\sigma - \sigma_{HB})d\sigma \tag{19}$$

$$M_{don}^X = \int_{-\infty}^{-\sigma_{HB}} p^X(\sigma)(-\sigma - \sigma_{HB})d\sigma \tag{20}$$

$$M_i^X = \int_{-\infty}^{+\infty} p^X(\sigma)\sigma^i d\sigma \tag{21}$$

The first σ-moments have a simple physical meaning: the zero-order σ-moment M_0^X is the surface area of the molecule, expressed in Å². The first-order one M_1^X is the total polarization charge of this surface, expressed in e (electric charge of the electron). For uncharged molecules, this moment is equal to zero. The second-order σ-moment M_2^X, expressed in e²·Å⁻², is the polarity of the molecule.[173] The third-order M_3^X represents the asymmetry of the σ-profile $p^X(\sigma)$. The other σ-moments up to M_6^X have no particular physical meanings. Finally, M_{acc}^X and M_{don}^X , expressed in e, are the "hydrogen-bonding" σ-moments representing the ability of the molecule to interact with hydrogen-bond acceptors and donors, respectively. Their value is non-zero when the σ-profile outranges the [−σ_HB, +σ_HB] interval, where σ_HB, the hydrogen-bond threshold, is equal to 0.01 e·Å⁻² as shown in Figure 2.3.
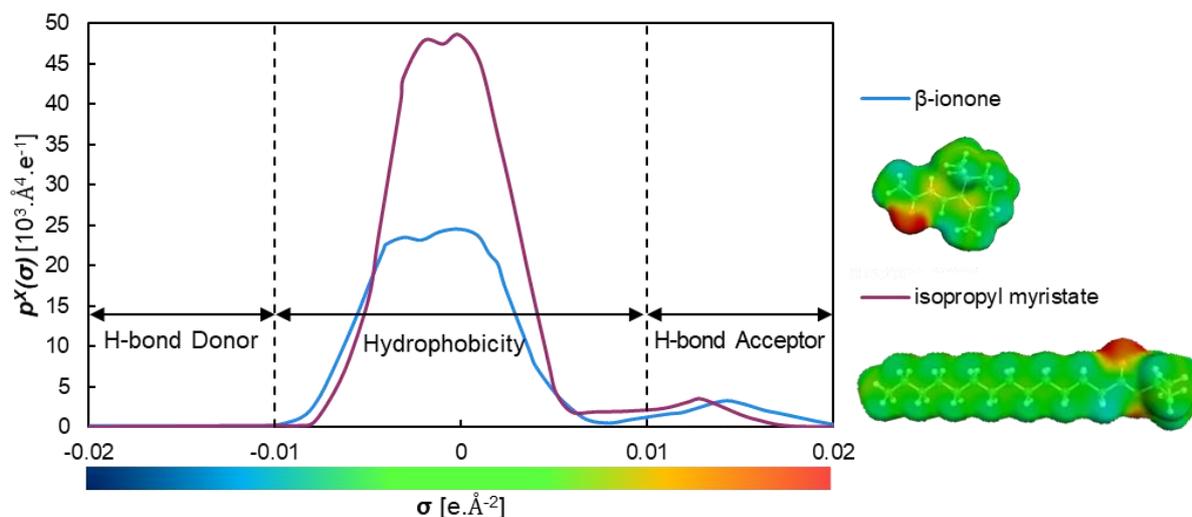
Figure 2.3. σ-profiles and σ-surfaces of β-ionone (in blue) and isopropyl myristate (in purple). The color gradient corresponds to the surface charge density σ.

Neither β-ionone nor isopropyl myristate exhibit Lewis acidity corresponding to the hydrogen bond donor region. However, both of them have a Lewis basicity with non-zero-value σ-profile in the hydrogen bond acceptor region. This is due to the presence of the ester and carbonyl functions inducing locally electron-rich surface areas (in red in both molecules according the colour scale in Figure 2.3). Finally, the central part of the σ-profile shows higher hydrophobicity in the case of isopropyl myristate than for β-ionone, which is in accordance with its longer alkyl moiety.

The computed σ-moments can be used as molecular descriptors and inputted in models such as multilinear regressions or neural networks applied to the prediction of properties such as EACN,[133,134] surface tension,[135] and viscosity.[136]

2.2.2. Multiple Linear Regression

Multiple linear regression (MLR) is the simplest modelling approach as it consists in finding the linear equation from the vector of entry values $x = \{x_k, k = 1 \dots q\}$, balanced by a set of parameters $A = \{a_k, k = 1 \dots q\}$, to an output quantity y. The parameters a_k are called partial regression coefficients: each of them measures the effect of the corresponding variable x_k on the quantity to model, all other variables being kept constant.[170] The general expression for a multiple linear regression is given by equation (22).

$$g(x, A) = \sum_{k=1}^{q} a_k x_k = XA \tag{22}$$

where X is the observation matrix of size (N, q), defined as the matrix which elements in column k correspond to the N measured values of the variable k $(x_{1k}, \dots x_{Nk})$. For each element i $(i = 1 \dots N)$ of the training set, the residual R_i is defined as the difference between the value of the quantity to model y^i and the value estimated by the model $g(x^i, A)$.

$$R_i = y^i - g(x^i, A) \tag{23}$$

The learning step is achieved by minimizing the least squares function $J(A)$, expressed by equation (24), assessing the adjustment of the model g to the learning dataset.

$$J(A) = \sum_{i=1}^{N} (R_i)^2 \tag{24}$$

A multiple linear regression based on the σ-moments calculated by the COSMO-RS approach[131,132] was applied to polar hydrocarbons and aprotic polar oils by Lukowicz et al.[133,134] Their first model was based on a 56-molecule training set containing unsaturated, cyclic, branched and halogenated alkanes and yielded an estimation error of 0.8 EACN units on a 6-molecule test set.[133] The same methodology was then applied to a more diverse 61-molecule training set containing esters, ketones, nitriles and ethers. This second model yielded an estimation error of 1.1 EACN units on a 9-molecule test set.[134] The diversification in the chemicals structures under study made it harder to account for the EACN value using a multilinear regression. In the present work, two points of improvements are proposed. Firstly, the dataset is enriched with new experimental EACN values. Secondly, other types of models, namely neural networks and graph machines, are used.

2.2.3. Neural Networks and Graph Machines

As early as 1943, neural networks were initially developed in an attempt to model the functioning of a nervous system.[174] A formal neuron is a parameterized non-linear function with bounded values, that applies an activation function f to a linear combination of entry values. f is generally a non-linear function: the most commonly used activation functions are the hyperbolic tangent function, the sigmoid function and the identity function.[175,176] The value y computed by the activation function is the neuron's output. The general expression for a formal neuron can be expressed by equation (25), where $\{x_k\}_{k=1\ldots q}$ are the entry values and $\{a_k\}_{k=0\ldots q}$ are the parameters, also called synapses. The a_0 parameter associated to the bias has a value equal to 1.[170]

$$y = f\left(a_0 + \sum_{k=1}^{q} a_k x_k\right) \tag{25}$$

The combination of formal neurons into neural networks allows building more complex models that can be applied to the estimation of many properties of interest. There are various ways of combining neurons. A neural network built by the association of neurons in layers is called a multilayer perceptron (MLP). A schematic MLP representation is shown in Figure 2.4. A parameter is associated to each connection and the total number of parameters depends on the network complexity, i.e. the number of neurons in each layer and the number of layers.
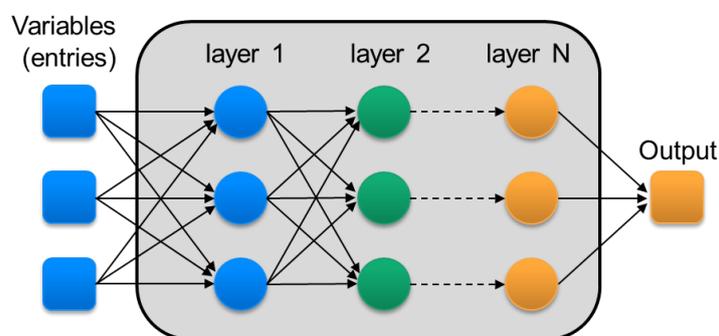
Figure 2.4. Schematic representation of a multilayer perceptron with N layers of hidden neurons.

The most interesting property of NNs is their accuracy in modelling a given property with parsimony compared to linear models in terms of parameters. Indeed, the number of parameters required to reach a given precision increases linearly with the number of variables for NNs, whereas it increases exponentially in the case of linear models. As a result, NNs require less examples in the training dataset.[170]

NN models have been developed and applied to the prediction (estimation) of a variety of chemical properties such as log P,[143] surface tension,[135] or viscosity.[136] However, a drawback associated with the use of NN models is the prior obtention of appropriate molecular descriptors that can, sometimes, result from complex calculations and require specific software

or important computational resources. Some other types of models such as graph machines have the advantage of being more accessible.

Graph machines (GM) on the other hand, are models that estimate a property directly from the topological information provided by the SMILES codes of molecules. In these models, molecules are described as directed acyclic graphs derived from their 2D structures and the parameterized functions that compute the estimate of the property of interest reflect the molecular structures of the compounds.[138,177] As usual in regression or classification models, GM parameters are computed by learning from examples present in an experimental value database.[138]

Graphs are representations of structured data. Instead of translating a structure into descriptors, the structure itself is used as input. To build a graph from a chemical structure, each atom is represented by a node and each bond by an edge. For a structure with n non-hydrogen atoms, the connections between the n nodes of the graph can be described by its adjacency matrix M of size $n \times n$. A six-node graph example and its adjacency matrix of size 6×6 are shown in Figure 2.5. In the matrix, each node connection is represented by a coefficient 1. For instance, node 1 is connected to nodes 2, 3 and 4 so coefficients $m_{1,2}$, $m_{1,3}$ and $m_{1,4}$ are equal to 1. In an oriented graph, the root node is defined as a node from which the arborescence starts. In the example of Figure 2.5, the root node is the number 1 node. In the case of molecular structures, labels can then be attributed to each node to account for the atom nature. Examples of graph machine images are available in the Appendix A8. The structured graph data is then translated into vectoral information, that can be processed by a NN model as described above.[170]

$$M = \begin{pmatrix} 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$
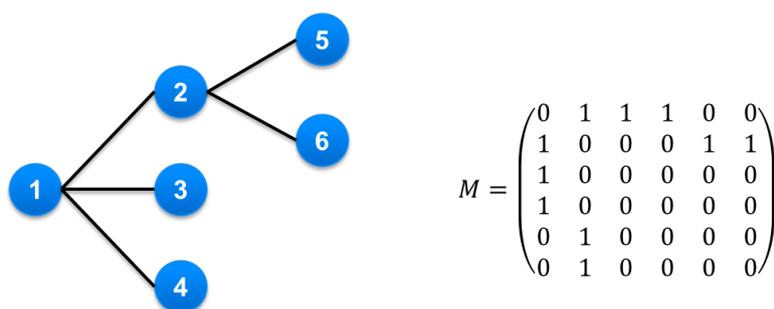
Figure 2.5. Schematic representation of a six-node graph (left) and its adjacency matrix M (right).

Both NN and GM models are built in an iterative way, meaning that a large number of parameters are tested and only the relevant ones are kept. The same way, the number of parameters is incremented until the most satisfying one is obtained. The number of hidden neurons, or layers in the multilayer perceptron, is incremented and the Leave One Out score (LOO score) as well as the learning Root Mean Square Error (RMSE) are calculated at each step. More details on model selection are available in the experimental section 3.4.

3. Experimental section

3.1. Chemicals

Isoamyl laurate (JOLEE 7750, Oleon, 100%), hemisqualane (Neossance™ Hemisqualane, Amyris, > 95%), isohexadecane (2,2,4,4,6,8,8-Heptamethylnonane, Sigma-Aldrich, 98%), pristane (2,6,10,14-tetramethylpentadecane, TCI, > 95%) and isododecane (2,2,4,6,6-pentamethylheptane, TCI, > 98.0%) were analyzed by GC-MS and used as such. 2-methylpentane (Sigma-Aldrich, > 99%), 3-methylpentane (Sigma-Aldrich, > 99%), 2,3-dimethylbutane (Sigma-Aldrich, 98%), isooctane (2,2,4-trimethylpentane, Sigma-Aldrich, > 99%), dipropyl ether (Sigma-Aldrich, > 99%) and diiso-propyl ether (Sigma-Aldrich, > 98.5%) were used as such. Pure tetraethyleneglycol monodecyl ether ($C_{10}E_4$) was synthesized according to a method described elsewhere.[178,179] Its purity was assessed by GC-MS analysis (> 99%) and by comparing its cloud point temperature at 2.6 wt% (20.4 °C *vs* 20.6 °C) with the reference value.[180] Tetraethyleneglycol monohexyl ether (C_6E_4) was synthetized using an analogous method to $C_{10}E_4$ and its cloud point temperature (66.2 °C at 16.4 wt%) was compared to the reference value (66.1 °C at 16.4 wt%).[181]

3.2. Fish diagrams for EACN determination

In order to enrich the EACN values database, 11 new experimental fish diagrams were built, in particular in the case of branched alkanes that were under-represented in the available literature data. The experimental EACN value was determined by establishing the phase behaviour of 50 ± 0.2 wt% water/oil mixtures at different $C_{10}E_4$ or C_6E_4 concentrations as a function of temperature. The Winsor types were determined by visual observation.[182] The most volatile oil samples (2-methylpentane, 3-methylpentane and 2,3-dimethylbutane) were weighed in glass tubes, placed in liquid nitrogen then sealed with a flame. Other samples were prepared in glass tubes closed by screw caps. Samples were first shaken gently several times and left in a thermoregulated bath at T ± 0.1 °C until equilibration. The point (C*; T*) corresponding to the intersection of the Winsor III and the Winsor IV phases was used to determine the oil's EACN: its T* value was reported on the T* vs. ACN (Alkane Carbon Number) reference straight line for linear alkanes using either $C_{10}E_4$ (see Figure 2.2b) or C_6E_4 as the surfactant.[127] The fish diagram of hemisqualane is given as an example in Figure 2.6a. The fish diagram lower concentration limit was determined by extrapolation of Winsor III phase relative volume as described by Burauer et al.[183] Other experimentally determined (C*; T*) points using $C_{10}E_4$ are represented in Figure 2.6b. The experimentally determined partial Fish diagrams for dipropylether and diisopropylether using C_6E_4 are available in Appendix A6 as well as the reference straight line for *n*-alkanes using C_6E_4.
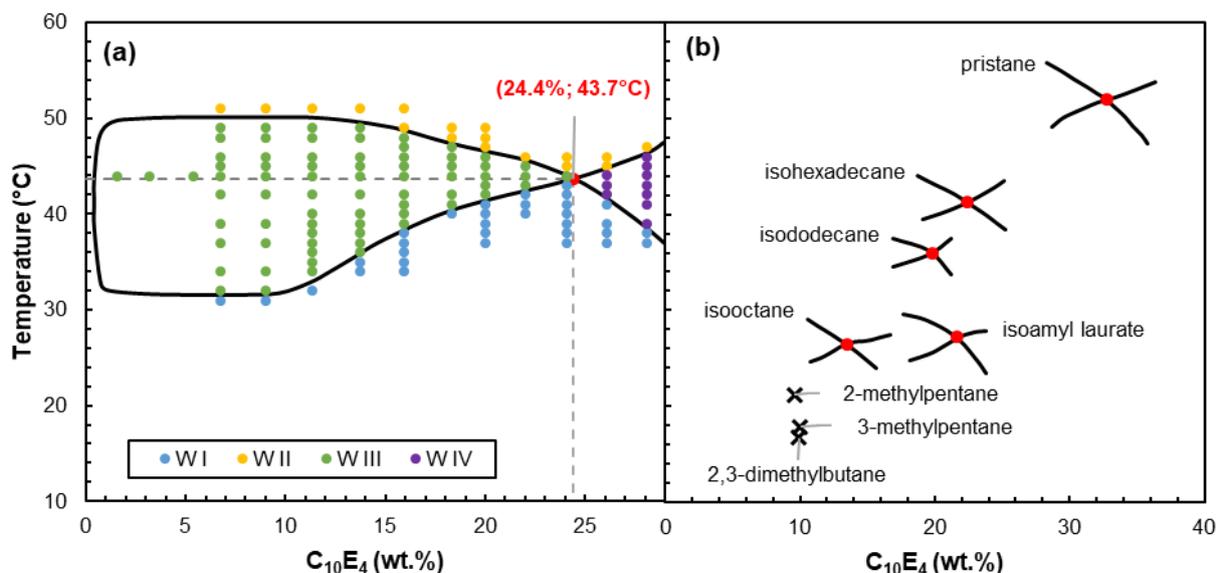
Figure 2.6. (a) Experimental fish plot of $C_{10}E_4$/Hemisqualane/Water-T system at Water/Oil ratio equal to 1 (w/w) and (b) partial fish plot and fish tail points (C*; T*) determined with $C_{10}E_4$ for pristane (2,6,10,14-tetramethylpentadecane), isohexadecane (2,2,4,4,6,6,8-heptamethylnonane), isododecane (2,2,4,6,6-pentamethylheptane), isooctane (2,2,4-trimethylpentane) and isoamyl laurate (3-methylbutyl dodecanoate). Fish tail points (C*;T*) for 2-methylpentane, 3-methylpentane and 2,3-dimethylbutane are represented by cross-marks for clarity.

3.3.   EACN database construction

A set of 121 compounds with reliable EACN values was assembled. They were either extracted from literature or determined experimentally. These averaged EACN values are reported in Table 2.1. For training and testing purposes, this set was divided into a training set of 111 compounds and a test set of 10 compounds. The whole set includes *n*-alkanes, esters, ethers, ketones, alkenes, alkynes, cyclic hydrocarbons, aromatics, branched hydrocarbons, nitriles and chloroalkanes, and consequently compounds containing carbon, hydrogen, oxygen, nitrogen and chloride atoms. The perfumery and cosmetic oils from the test set were chosen: (i) for their complex structure, e.g. caryophyllene, a molecule with unsaturations, cycles and branching, and (ii) with EACN experimentally determined in our group to have trusty values. The distribution of the structural features among both data sets are displayed in Figure 2.7.

Table 2.1. Names, SMILES notations, three first σ-moments (different from zero) calculated with COSMO-RS, number of carbon atoms and experimental EACN and average EACN values determined from the Fish-tail-temperature T* reported in the literature for ternary systems C_iE_j/Oil/Water for the 111 molecules of the training set and the 10 molecules of the test set.

Entry	Compound	SMILES code	M_0^{X} [a]	M_2^{X} [b]	M_3^{X} [c]	N_C	EACN$_{exp}$	Ref.
1	Hexane	CCCCCC	159.25	8.17	–0.08	6	6[d]	-
2	Heptane	CCCCCCC	179.32	9.14	–0.03	7	7[d]	-
3	Octane	CCCCCCCC	198.88	10.04	0,00	8	8[d]	-
4	Nonane	CCCCCCCCC	218.16	10.71	0,00	9	9[d]	-
5	Decane	CCCCCCCCCC	238.31	11.78	0.06	10	10[d]	-
6	Undecane	CCCCCCCCCCC	258.33	12.51	0.1	11	11[d]	-
7	Dodecane	CCCCCCCCCCCC	278.75	13.46	0.12	12	12[d]	-

8	Tridecane	CCCCCCCCCCCCC	298.9	14.31	0.22	13	13[d]	-
9	Tetradecane	CCCCCCCCCCCCCC	319.04	15.18	0.26	14	14[d]	-
10	Pentadecane	CCCCCCCCCCCCCCC	339.46	16.11	0.28	15	15[d]	-
11	Hexadecane	CCCCCCCCCCCCCCCC	359.04	17.29	0.38	16	16[d]	-
12	Heptadecane	CCCCCCCCCCCCCCCCC	379.07	18.2	0.43	17	17[d]	-
13	Octadecane	CCCCCCCCCCCCCCCCCC	399.34	18.9	0.42	18	18[d]	-
14	Nonadecane	CCCCCCCCCCCCCCCCCCC	419.21	20.01	0.52	19	19[d]	-
15	Eicosane	CCCCCCCCCCCCCCCCCCCC	440.35	20.71	0.52	20	20[d]	-
16	Cyclohexane	C1CCCCC1	135.77	5.86	−0.12	6	2.1	130,159,184,185
17	Methylcyclohexane	CC1CCCCC1	153.55	7.16	−0.04	7	3.2	130
18	Ethylcyclohexane	CCC1CCCCC1	171.52	7.89	−0.13	8	4.2	127,130
19	Propylcyclohexane	CCCC1CCCCC1	191.55	8.77	−0.09	9	5.9	127,130
20	Butylcyclohexane	CCCCC1CCCCC1	210.84	9.62	−0.05	10	7.3	127,130
21	Decylcyclohexane	CCCCCCCCCC1CCCCC1	332.16	14.95	0.24	16	14.4	127
22	Dodecylcyclohexane	CCCCCCCCCCCC1CCCCC1	371.81	16.81	0.29	18	17.5	127
23	1-Octene	CCCCCCC=C	192.38	18.77	3.59	8	3.9	133
24	1-Decene	CCCCCCCCC=C	234.26	20.46	3.17	10	5.5	133
25	1-Dodecene	CCCCCCCCCCC=C	274.46	22.29	3.31	12	8.1	133
26	1-Octadecene	CCCCCCCCCCCCCCCCC=C	394.66	27.89	3.6	18	14.2	133
27	Dipropylether	CCCOCCC	171.36	26.39	20.59	6	0.4	[e]
28	Dibutylether	CCCCOCCCC	211.38	28.47	21.23	8	3.0	127,134,164
29	Dipentylether	CCCCCOCCCCC	251.19	29.57	20.44	10	4.2	134
30	Dihexylether	CCCCCCOCCCCCC	291.31	31.96	21.34	12	6.2	134
31	Diheptylether	CCCCCCCOCCCCCCC	331.49	32.86	20.36	14	8	134
32-Test	Dioctylether	CCCCCCCCOCCCCCCCC	371.28	35.31	21.42	16	10.3	134
33	1-Chlorodecane	CCCCCCCCCCCl	257.74	29.94	3.76	10	3.5	133,186
34	1-Chlorododecane	CCCCCCCCCCCCCl	297.96	31.52	3.82	12	5.6	133,186
35	1-Chlorotetradecane	CCCCCCCCCCCCCCCl	338.17	33.58	3.92	14	8	133,186
36	1-Chlorohexadecane	CCCCCCCCCCCCCCCCCl	378.22	35.15	3.98	16	9.8	133,186
37	Butylbenzene	CCCCc1ccccc1	197.06	25.15	0.54	10	0.4	133,185
38	Octylbenzene	CCCCCCCCc1ccccc1	280.39	29.6	0.94	14	4	127,130
39	Decylbenzene	CCCCCCCCCCc1ccccc1	320.42	31.55	1.1	16	6	127,130
40	Dodecylbenzene	CCCCCCCCCCCCc1ccccc1	359.87	33.08	1.12	18	7.8	127,159,185
41	1-Octyne	CCCCCCC#C	188.77	33.22	1.53	8	−1.8	133
42	1-Decyne	CCCCCCCCC#C	230.44	35.32	1.68	10	0.1	133
43	1-Dodecyne	CCCCCCCCCCC#C	265.87	36.64	1.52	12	2	133
44	1-Tetradecyne	CCCCCCCCCCCCC#C	310.75	38.72	1.96	14	3.9	133
45	Ethyl decanoate	CCCCCCCCCC(=O)OCC	294.96	53.81	33.56	12	2.1	145,147
46	Ethyl dodecanoate	CCCCCCCCCCCC(=O)OCC	333.62	54.56	32.64	14	3.8	147
47	Ethyl myristate	CCCCCCCCCCCCCC(=O)OCC	375.02	56.48	32.88	16	5.2	147
48	Ethyl palmitate	CCCCCCCCCCCCCCCC(=O)OCC	415.24	59.07	33.25	18	6.8	147
49	Octanenitrile	CCCCCCCC#N	204.22	49.16	22.76	8	−1.7	134
50	Decanenitrile	CCCCCCCCCC#N	243.52	50.82	22.88	10	−0.5	134
51	Dodecanenitrile	CCCCCCCCCCCC#N	283.99	52.55	23.07	12	0.4	134
52	2-Octanone	CCCCCCC(C)=O	201.89	48.96	35.68	8	−3.4	134
53	2-Decanone	CCCCCCCCC(C)=O	242.35	50.66	35.46	10	−2.1	134
54	2-Undecanone	CCCCCCCCCC(C)=O	261.87	51.41	35.83	11	−1.3	134
55	2-Dodecanone	CCCCCCCCCCC(C)=O	282.29	52.17	35.35	12	−0.6	134
56	2-methylpentane	CCCC(C)C	154.78	8.47	−0.17	6	6.4	[e]

57	2,2,4-trimethylpentane	CC(C)CC(C)(C)C	181.75	10.89	–0.13	8	8.3	[e]
58	2,3-dimethylbutane	CC(C)C(C)C	149.31	8.56	–0.25	6	4.8	[e]
59	3-methylpentane	CCC(C)CC	152.61	8.3	–0.28	6	5.2	[e]
60-Test	Isododecane	CC(C)(C)CC(C)CC(C)(C)C	244.43	14.90	–0.11	12	11.7	[e]
61-Test	Hemisqualane	CCC(C)CCCC(C)CCCC(C)C	320.82	16.86	–0.11	15	14.8	[e]
62	Isohexadecane	CC(CC(C)(C)C)CC(C)(C)CC(C)(C)C	303.13	18.48	–0.11	16	13.9	[e]
63	Pristane	CC(C)CCCC(C)CCCC(C)CCCC(C)C	396.57	20.64	0.06	19	17.6	[e]
64	1,2-Dimethylcyclohexane	CC1CCCCC1C	168.83	8.2	–0.05	8	3.3	130
65	1,4-Dimethylcyclohexane	CC1CCC(C)CC1	171.39	8.45	0.03	8	4.6	130
66	Isopropylcyclohexane	CC(C)C1CCCCC1	185.48	9.03	–0.13	9	5.3	130
67	Cyclooctane	C1CCCCCCC1	163.91	7.28	–0.04	8	4.1	133
68	Cyclodecane	C1CCCCCCCCC1	192.46	9.22	–0.38	10	5.6	133
69	cis-Decalin	C1CC[C@@H]2CCCC[C@@H]2C1	188.05	7.8	0.05	10	5.3	133
70	Pinane	CC1CCC2CC1C2(C)C	186.28	10.46	–0.38	10	4.1	146
71	p-Menthane	CC1CCC(CC1)C(C)C	200.29	10.29	–0.15	10	5.8	146
72	Cyclohexene	C1CCC=CC1	132.04	14.28	3.82	6	–1.2	130
73	1-Methyl-1-cyclohexene	CC1=CCCCC1	150.83	14.02	3.11	7	0	187
74	4-Methyl-1-cyclohexene	CC1CCC=CC1	149.89	15.41	3.72	7	0.1	187
75	3-Methyl-1-cyclohexene	CC1CCCC=C1	150.06	15.25	3.73	7	–0.5	187
76	cis-Cyclooctene	C1CCC\C=C/CC1	160.62	14.36	3.25	8	1.5	133
77	1,3-Cyclohexadiene	C1CC=CC=C1	127.77	20.68	3.54	6	–3.1	187
78	1,4-Cyclohexadiene	C1C=CCC=C1	128.11	21.47	4.65	6	–4	187
79	2,5-Norbornadiene	C1C2C=CC1C=C2	133.62	22.95	3.78	7	–3.2	187
80	2,6,10-trimethylundeca-2,6-diene	CC(C)CCC=C(C)CCC=C(C)C	291.67	28.17	4.2	14	10.3	146
81	p-Xylene	Cc1ccc(C)cc1	161.17	23.96	1.22	8	–2.4	133
82	Phenyl-1-butyne	CCC#Cc1ccccc1	195.47	35.72	–0.41	10	–3.3	133
83	p-Cymene	CC(C)c1ccc(C)cc1	195.88	25.88	1.52	10	–0.8	146
84	α-Pinene	CC1=CCC2CC1C2(C)C	186.35	14.89	1.06	10	3.5	146
85	p-Menth-2-ene	CC(C)C1CCC(C)C=C1	197.75	17,00	3.01	10	3.4	146
86	Δ-3-Carene	CC1=CCC2C(C1)C2(C)C	194.74	20.82	3.22	10	2.5	146
87	β-Pinene	CC1(C)C2CCC(=C)C1C2	184.7	18.79	3.18	10	2.2	146
88-Test	Limonene	CC(=C)C1CCC(=CC1)C	197.8	24.85	5.78	10	1.8	146
89	γ-Terpinene	CC(C)C1=CCC(=CC1)C	199.76	22.82	4.25	10	1.7	146
90	α-Terpinene	CC(C)C1=CC=C(C)CC1	199.72	23.33	4.58	10	1.2	146
91	Terpinolene	CC(C)=C1CCC(=CC1)C	198.78	22.97	3.89	10	0.7	146
92	Longifolene	CC1(C)CCCC2(C)C3CCC(C13)C2=C	237.83	19.09	2.2	15	7.0	146
93-Test	Caryophyllene	C\C1=C/CCC(=C)C2CC(C)(C)C2CC1	254.08	25.94	5.64	10	6.0	146
94	Diisopropylether	CC(C)OC(C)C	165.76	31.98	26.8	6	0.6	[e]
95	1,4-dipropoxybutane	CCCOCCCCOCCC	263.26	46.21	37.9	10	1.9	164
96	1,2-dipropoxyethane	CCCOCCOCCC	216.46	43.75	38.01	8	0.4	164
97	1,2-dibutoxyethane	CCCCOCCOCCCC	264.4	50.76	43.18	10	1.7	164
98	Decyl butyrate	CCCCCCCCCCOC(=O)CCC	334.4	53.88	32.77	14	5	147
99	Hexyl octanoate	CCC[C:1]CCCC(=O)OCCCCCC	333.77	53.52	32.68	14	6.2	147
100	Butyl dodecanoate	CCCCCCCCCCCC(=O)OCCCC	374.28	55.45	33.31	16	7.2	147
101	Myristyl propionate	CCCCCCCCCCCCCCOC(=O)CC	394.93	57.2	33.01	17	6.8	147
102-Test	Octyloctanoate	CCCCCCCCOC(=O)CCC[C:1]CCC	371.81	53.73	31.52	16	8.1	147
103-Test	Isopropyl myristate	CCCCCCCCCCCCCC(=O)OC(C)C	392.03	55.25	31.87	17	7.3	127,147,185
104	Isoamyl laurate	CCCCCCCCCCCC(=O)OCCC(C)C	387.57	55.79	32.49	17	8.8	[e]
105	Hexyl dodecanoate	CCCCCCCCCCCC(=O)OCCCCCC	413.93	57.13	33.18	18	9.3	147

106	Ethyl oleate	CCCCCCCC\C=C/CCCCCCCC(=O)OCC	448.42	66.44	35.58	20	7.2	[147]
107	Menthone	CC(C)C1CCC(C)CC1=O	204.53	47.96	39.96	10	−1.5	[145]
108	Eucalyptol	CC12CCC(CC1)C(C)(C)O2	193.69	33.68	30.12	10	−1.6	[145]
109-Test	Rose oxide	CC1CCOC(C1)C=C(C)C	214.57	42.15	32.13	10	−1.7	[145]
110	D-Carvone	CC(=C)[C@H]1CC=C(C)C(=O)C1	201.11	60.13	41.08	10	−3.1	[145]
111	Hexyl methacrylate	CCCCCCOC(=O)C(C)=C	246.73	48.51	27.65	10	0.1	[188]
112	Menthyl acetate	CC(C)C1CCC(C)CC1OC(C)=O	251.45	47.83	31.61	12	−0.1	[134]
113	Citronellyl acetate	CC(CCOC(C)=O)CCC=C(C)C	277.2	60.77	33.79	12	−0.2	[134]
114	Geranyl acetate	CC(=O)OC\C=C(/C)CCC=C(C)C	274.66	65.71	33.06	12	−0.6	[134]
115	α-Damascone	C/C=C/C(=O)C1C(=CCCC1(C)C)C	244.47	55.35	36.19	13	−1.3	[145]
116-Test	Linalyl acetate	CC(C)=CCCC(C)(OC(C)=O)C=C	263.17	58.29	27.24	13	−0.9	[134]
117-Test	β-Ionone	CC(=O)/C=C/C1=C(C)CCCC1(C)C	247.59	61.65	48.43	13	−1.9	[145,159]
118	Methyl dihydrojasmonate	CCCCCC1C(CCC1=O)CC(=O)OC	289.12	88.67	56.94	13	−1.7	[145]
119	Ethylene brassylate	O=C1CCCCCCCCCCCC(=O)OCCO1	311.19	84.83	49.39	15	−1.1	[145]
120	Methyl cedrylether	CO[C@]1(C)CCC23C[C@@H]1C(C)(C)[C@@H]2CC[C@H]3C	263.38	31.76	20.52	16	3.5	[145]
121	Ambrettolid	O=C1CCCC\C=C/CCCCCCCO1	309.44	58.6	35.9	16	1	[145]

[a] M_0^X, expressed in Å^2, is equal to the whole surface area of molecule X; [b] M_2^X, expressed in e^2.Å^{-2}, reflects the polarity of molecule X; [c] M_3^X, expressed in e^3.Å^{-4}, reflects the electrostatic asymmetry of molecule X; [d]EACN value equal to number of carbon atoms by definition; [e]This work.
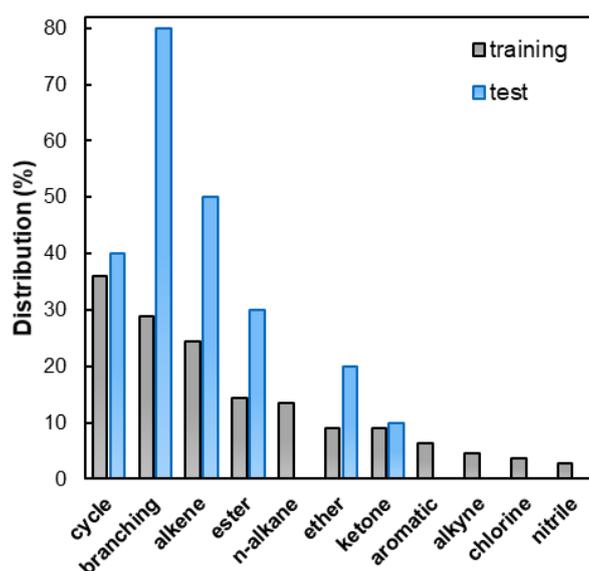
Figure 2.7. Distribution of structural features (in percent) of molecules in the training and test sets of the EACN database.

3.4. Graph Machine (GM) and Neural Network (NN) Models Selection

Basically, NN models are multiple non-linear regressions that estimate an output value of a property of interest from some input descriptors values, hereafter three σ-moments selected from a pull of eight σ-moments, all computed with COSMO-RS according to a procedure described in section 2.2.1. A selection of the σ-moments was performed with Metagen, a software package written in Python (Laboratoire de Chimie Organique, CNRS, ESPCI Paris, PSL Research University). Feature selection by the random probe method showed that for our data, M_0^X, M_2^X and M_3^X are the most relevant for EACN estimation.

Both NN and GM models are built from Multi Layer Perceptrons (MLPs) that contain a single hidden layer of neurons. The complexity of the models is consequently dependent on the number of neurons of that layer, and along with this, on the number of parameters of the models. Since, for a given number of neurons in the MLPs, NN and GM models have a different number of parameters, this later variable will be preferred as a complexity equivalent in the model complexity selection section (4.1).

The selection of a model is a key step in machine learning model design: it consists in finding the model complexity, given the available data for designing it, that will result in the best generalization. To that end, with the 111-molecule set available, trainings are carried out with an increasing number of MLP hidden neurons. The ability of both models to account for the training data is monitored with the root mean square training error (RMSTE) that is computed as follows:

$$RMSTE = \sqrt{\frac{1}{111}\sum_{i=1}^{111}\left(EACN_{exp.}^{i} - EACN_{est.}^{i}\right)^2} \qquad (26)$$

where $EACN_{exp.}^{i}$ is the EACN value determined experimentally for molecule *i*, and $EACN_{est.}^{i}$ is the EACN value estimated by the model for molecule *i* at the end of the training.

The estimation of the generalization error for model selection is usually performed by two methods: the computation of the Leave-One-Out score (LOOs) and the computation of the Virtual Leave-One-Out score (VLOOs). The computation of the LOO score was chosen for the determination of the optimal complexity, since the VLOO score, that is a first order approximation of the LOO score, is less accurate for small size datasets.[136,189] At the end of the LOO process, the leave-one-out score (LOO score) is computed as:

$$LOO\ score = \sqrt{\frac{1}{111}\sum_{i=1}^{111}\left(EACN_{exp.}^{i} - EACN_{pred.}^{i}\right)^2} \qquad (27)$$

where $EACN_{exp.}^{i}$ is the EACN value determined experimentally for molecule *i*, and $EACN_{pred.}^{i}$ is the average EACN prediction value computed for the left-out molecule *i* with 50 models having different initialization parameters. The above equation is the same as eq. (26) defining the RMSTE, except that now a true prediction is performed for every molecule, since the molecule *i* does not belong to the training set. Schematic representations of the computations leading to RMSTE and LOO score are shown in Figure 2.8. The LOO computation is repeated five times for each complexity of the NN and GM based models, so that the average results are presented.
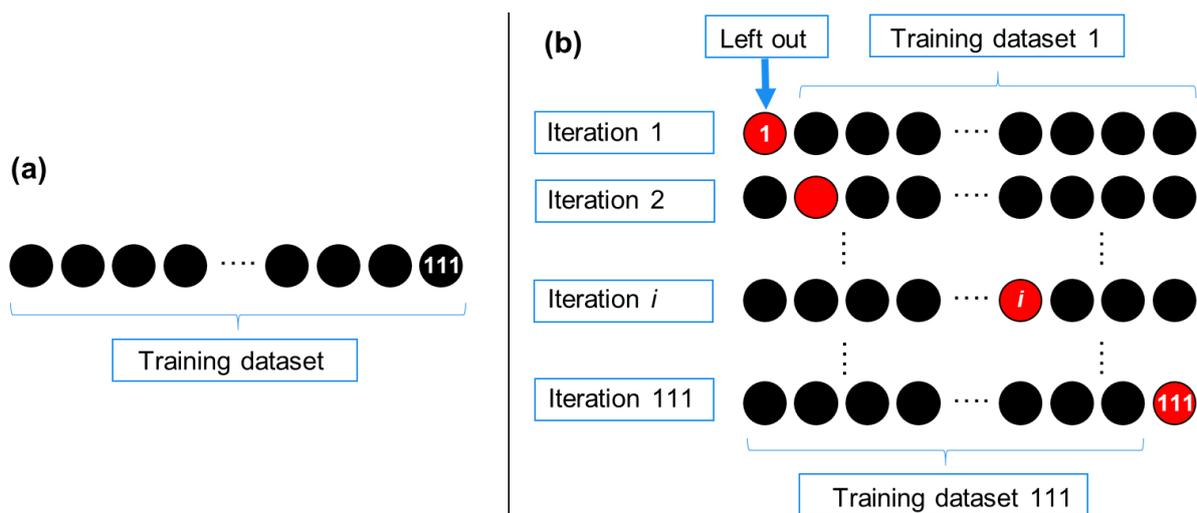
Figure 2.8. Schematic representation of (a) root mean square training error (RMSTE) calculation and (b) leave one out (LOO) score calculation. At each iteration, the model is trained using the molecules represented by black circles and the EACN value is predicted for the left-out molecule, represented by the red circles.

A few difficulties have been met with the first modeling experiments that have been addressed as follows. With the σ-moment-based NN models, a large EACN deviation was observed specifically for the 15 molecules of the *n*-alkane family, regardless of the complexity of the MLPs used. It was found that adding the number of carbon atoms (N_C) for every molecule as a fourth descriptor corrected this problem. Similarly, an important deviation from the experimental value was exclusively observed for the hexyl octanoate EACN estimation with the graph machine-based models. A thorough analysis of the graph machine construction for this molecule indicated that it was not consistent with the construction of the other linear ester graph machines (e.g. the ethyl esters). The input code for ethyl hexanoate was particularized so that the constructions were uniform for all esters. This modification is explained in Figure 2.9 where the directed graphs for hexyl octanoate and ethyl dodecanoate are shown (a). These graphs, that are encoded from their SMILES codes (b) are isomorph to the 2D formula also represented (c). Without modification of the hexyl octanoate SMILES code, the central node of the resulting graph, also called root node, would have been located on the green node. Thanks to the special bracketed tag in the hexyl octanoate SMILES code, the graph red root nodes have now a consistent position in both graphs: the root nodes are equally connected to nodes with a carbon type label and are at the same distance of the functional node, i.e. the one connected to the nodes with an oxygen type label. The position of the root node is important since it corresponds to the graph machine (not represented) output neuron that computes the estimated EACN value. As a result, the estimation of EACN was much efficient for the hexyl octanoate compound.
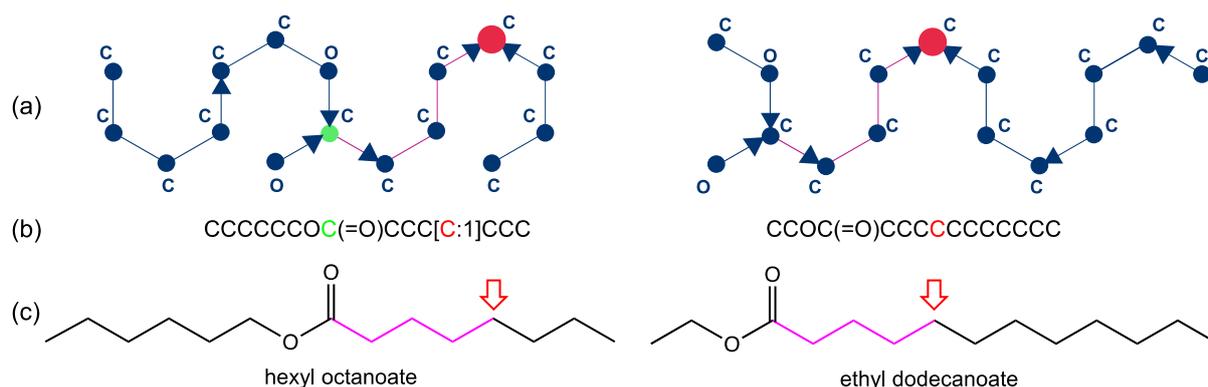
Figure 2.9. Encoding hexyl octanoate and ethyl dodecanoate into directed graphs: (a) directed graphs with root nodes in red; the root node position computed automatically for hexyl octanoate is colored in green and the path between the functional atom node and the root node in pink, (b) SMILES codes with the expected position of the root nodes indicated in red and (c) 2D formulas with expected positions for the root nodes (red arrows). The atom types C and O added on the directed graphs correspond to node labels that are inputs of the parameterized function implemented at each node of the graph to build the graph machine.

No such particular cases were encountered with the published models designed for surface tension and viscosity estimations.[135,136] This is probably due to the use of a larger training set counting many esters of various sizes and positions for the functional group. In this work, only a dataset of moderate size was built and among the dozen of linear esters that have an alkyl chain of ten carbon atoms or more, hexyl octanoate is the unique compound of similar size in the training set to have a functional group in the middle of its carbon skeleton. To illustrate this exception, a representation of the graph machine can be computed with the demo software as detailed in the Appendix A8.

4. EACN estimation using artificial intelligence methods
4.1. Graph Machine and Neural Network complexity selection

The Figure 2.10 that displays the LOO scores and RMSTE versus the number of parameters of the MLPs, i.e. the complexity, for the GM and NN models indicates that in both cases: (i) the data are correctly learned since the RMSTE are decreasing monotonously as the complexity increases and (ii) the LOO scores decrease, go through a minimum and start increasing.

The NN LOO score is clearly minimum (0.8 EACN unit) for a number of parameters equal to 37, i.e. six hidden neurons. On the contrary, very close GM LOO score values equal to 0.8 and 0.9 EACN unit are computed for 45 (5N) and 58 (6N) parameters respectively. In such a situation, the usual practice is to select the model with the lower complexity.[141] Therefore, graph machines with five hidden neurons (45 parameters) and neural networks with six hidden neurons (37 parameters) were kept for later testing.
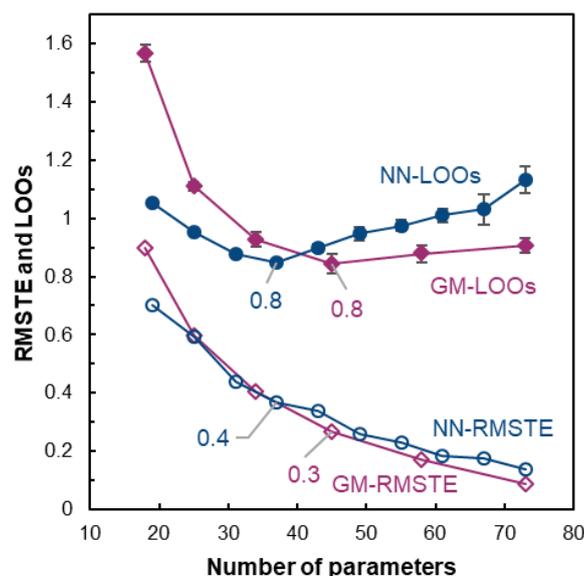
Figure 2.10. RMSTE value of the model (out of 1000) having the smallest RMSTE for the GM-based-model (pink diamonds) and NN-based-model (blue circles) for the 111 molecules of the training set, and means of the leave-one-out score values (GM orange diamonds, NN blue filled circles) computed for five different parameter initializations for the 111 molecules of the training set vs number of parameters. The error bars for the LOO scores are the standard deviations computed over the five LOO score values.

For comparison purposes, the results of the EACN leave-one-out estimations, corresponding to the LOO computation (out of five) that gives the best LOO score (0.8 EACN unit for GM and NN) for the molecule training set versus experimental EACN, are displayed in Figure 2.11 for the two preferred models.
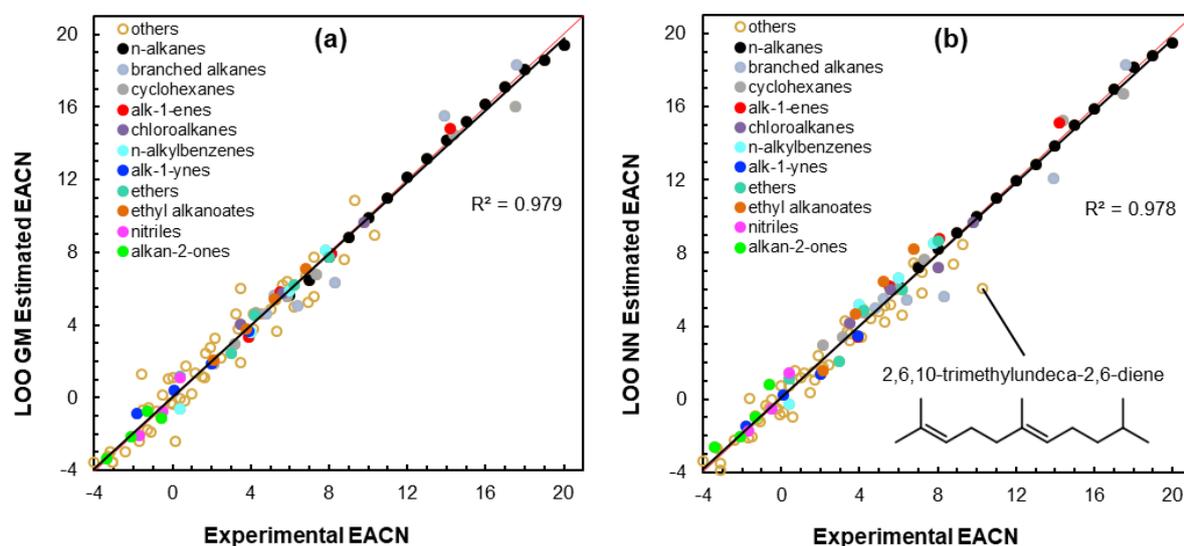
Figure 2.11. Scatter plots of LOO EACN estimations computed by graph machine from SMILES with five hidden neurons (a) and by neural networks with six hidden neurons using M_0^X, M_2^X, M_3^X and N_C as descriptors (b) for the 111 compounds of the training set vs experimental values of the EACN. The bisector and the regression lines are represented in red and black respectively.

Both models give similar results at a first glance (R^2 very close), in particular for the homologous series belonging to the chemical families indicated in the legends of Figure 2.11a and b, though slightly better estimations could be credited to the graph machine-based model

for these ten series. On the contrary the other compounds, most of them possessing several structural features, have dots that lay closer to the bisector line for the neural network-based model, indicating better results for the NN estimator with these compounds.

It is worth noting that with the NN-6N model, the EACN value of 2,6,10-trimethylundeca-2,6-diene is significantly under-estimated by the LOO calculations (Figure 2.11b). A possible explanation for this discrepancy could result from the fact that the two double bonds of 2,6,10-trimethylundeca-2,6-diene are in position 2,6 and not at the end of the chain. Indeed, the neural network "learns" the effect of double bonds on the basis of fairly rigid terpenes and a series of 1-alkenes whose double bonds are at the end of the chain. On average, each double bond in 1-alkenes decreases EACN by 4.1 units compared to the number of carbon atoms (N_C) of the molecule (entries 23 to 26 in Table 2.1). In terpenes, the effect of the unsaturation is associated to cyclization and branching that also contribute to reducing the EACN values. On average, each branching in branched alkanes decreases the EACN value by 0.3 units (entries 56 to 59, 62 and 63 in Table 2.1) compared to N_C. A rough estimation of the EACN value for 2,6,10-trimethylundeca-2,6-diene, a molecule with 14 carbons, 2 double bonds and 3 branches (see Figure 2.11) would be of about 4.9 units ($EACN = 14 - 2 \times 4.1 - 3 \times 0.3$). However, the 2 double bonds of 2,6,10-trimethylundeca-2,6-diene are less accessible than those of 1-alkenes which tend to be located close to the polar zone of the interfacial film made up of $C_{10}E_4$. As a result, instead of decreasing the EACN by about 9 units as expected, the experimentally observed decrease with respect to the corresponding *n*-alkane (tetradecane) is only 3.7 units.

According to the same reasoning, we can anticipate that double bonds in the terminal or exocyclic position have a stronger effect than endocyclic bonds. Indeed, a comparison of the experimental EACNs of citronellyl acetate (–0.2) and geranyl acetate (–0.6), two molecules that differ only in a central double bond, indicates a decrease of only 0.4 EACN units for the central bond of linalyl acetate instead of the expected 2.5 units (or more). Thus, the additivity method used to evaluate the decreasing effects of several chemical features in a molecule relative to the EACN of the alkane with the same number of carbon atoms is probably inoperative, in particular for 2,6,10-trimethylundeca-2,6-diene.

As regards to the GM-5N model in Figure 2.11a, no dot seems to be excessively far from the bisector, meaning that every EACN value of the training set is correctly estimated. In comparison with the NN model, the GM estimated EACN value of 2,6,10-trimethylundeca-2,6-diene is satisfying, which tends to confirm that the COSMO-RS descriptors used for the NN estimation correctly describe the oil in a bulk phase but is less suitable to describe this compound's behaviour in SOW systems.

4.2. Estimations with two models for a 10-molecule test set

To assess the estimation accuracies of the NN-based and the GM-based models of previously selected complexities, computation of the EACN for the ten molecules of the test set are made using the VLOO methodology previously described.[136] Briefly, for the GM and NN models selected in section 4.1, ten runs of 250 trainings each were performed with different parameter initializations. The VLOO score of each model (out of 250) was computed, and the mean of the 25 smallest VLOO scores of each run was computed. The run (out of ten) with the smallest mean VLOO score was selected. The 25 models of that sequence having the smallest VLOO scores estimated the EACN of the ten test molecules, and the mean of those 25 estimations was computed. These final estimations for both models are plotted vs the experimental values in Figure 2.12. The proximity of the dots with the bisector line shows that these estimations are close to the experimental values. Only the isododecane (2,2,4,6,6-pentamethylheptane) blue data point is somewhat further from the bisector line. These good results are confirmed by the displayed determination coefficients that are equal to 0.992 and 0.986 for the GM-5N-based and NN-6N-based models respectively.
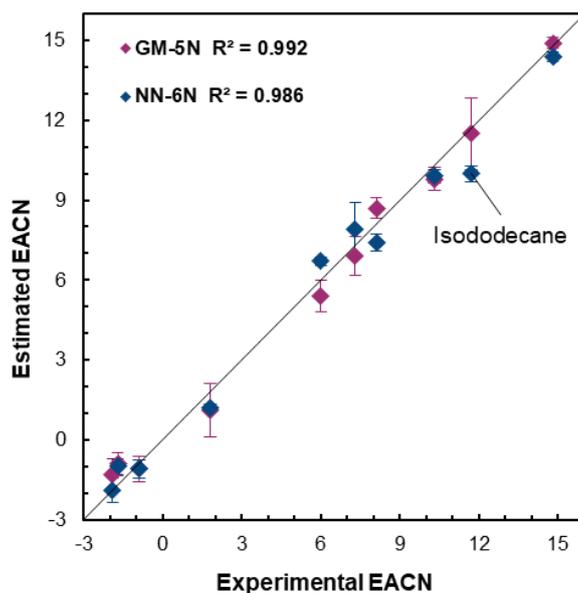
Figure 2.12. Scatter plots of EACN estimations computed by the graph-machine-based model with five hidden neurons (GM-5N, pink diamonds) and the neural-network-based model with six hidden neurons (NN-6N, blue diamonds) vs. experimental EACN values for the 10 molecules of the test set. The bisector line is represented in black, the error bars are the confidence intervals computed over the 25 selected models for the ten molecules of the test set.

The estimations errors listed in Table 2.2 for the ten molecules are indeed smaller or equal to 0.8 EACN unit but for the isododecane that exhibits a fairly large error (1.9 EACN unit) with the NN model. The computed test root mean square error values (test RMSE, bottom row) with an equation (26)-like formula, are equal to 0.5 and 0.7 EACN unit, confirming the efficiency of both models. Moreover, the estimations of the two predictors are in a good agreement since the maximum of the error deviation between the two computations is equal to 0.8 EACN unit

for eight molecules out of ten. Even for six of them, the estimation error difference is less or equal to 0.3 EACN unit. Those results were not given for granted, especially for complex molecules that have multiple features (e.g. limonene or rose oxide). Finally, it should be noted that to get such convincing results with the GM-based model, the SMILES code used to generate the octyl octanoate graph machine was also modified as explained for hexyl octanoate in section 3.4. Without taking this precaution, the prediction for octyl octanoate was clearly out of range with the GM-based model. The estimation values for the 10-molecule test set are also reported in Appendix A8.

Table 2.2. Difference between Experimental and Estimated EACN for the Test Set of Ten Molecules.

Molecule	Chemical Structure	EACN$_{exp.}$	Estimation error	
			NN-6N[a]	GM-5N[b]
Hemisqualane		14.8	+0.2	−0.1
Isododecane		11.7	+1.9	+0.2
Dioctylether		10.3	+0.2	+0.5
Octyloctanoate		8.1	+0.7	+0.6
Isopropyl myristate		7.3	−0.8	+0.3
Caryophyllene		6.0	−0.2	+0.6
Limonene		1.8	+0.7	+0.7
Linalyl acetate		-0.9	+0.1	+0.2
Rose oxide		-1.7	−0.1	−0.8
β-Ionone		-1.9	−0.4	−0.6
		test RMSE[c]	**0.7**	**0.5**

Differences between experimental and estimated EACN using [a]neural-network-based and [b]graph-machine-based models. [c]Root-mean-square test error (in EACN unit) for the ten molecules of the test.

Compared to the works of Lukowicz et al.,[133,134] the models precision are improved with a test RMSE of 0.7 and 0.5 units respectively whereas the MLP models described in the literature gave a test RMSE of 0.8 and 1.1 EACN units respectively.

4.3. Prediction of EACN for unknown compounds

One of the potential applications of the previously developed models is to predict the EACN of compounds for which there is no known experimental EACN value. Ethyl hexanoate is used here as an example. Data used as entries for computation and computation results are gathered in Table 2.3. Prediction of the EACN of any liquid of molecule containing carbon, hydrogen, oxygen, nitrogen and chloride atoms can be computed starting either from its SMILES code or its COSMO-RS σ-moments. A demonstration is available in Appendix A8. Estimations can also be calculated in homologous series of oils with alkyl chains of increasing size to evaluate and compare the effect of a given function on the EACN of oils. Lukowicz et al. observed that the EACN of several series of homologous oil increases approximately linearly with the number of carbons.[134] In comparison, the ability of the GM and NN models was tested to predict the evolution of EACNs of homologous oils. Furthermore, for practical applications, C8 to C15 phenols,[190] terpenes[130,191] and terpenoids[134] are particularly frequent in perfumery while C12 to C18 alkanes, esters and ethers are widely used as emollients to prepare cosmetic emulsions.[192] It is therefore crucial to reliably predict the EACN of oils with a relatively small number of carbons (≤ 20).

Table 2.3. SMILES code, σ-moments and number of carbon atoms necessary to compute EACN values for ethyl hexanoate.

Compound	SMILES code	M_0^X	M_2^X	M_3^X	N_C	EACN$_{NN-6N}$	EACN$_{GM-5N}$
Ethyl hexanoate	CCCC[C:1]C(OCC)=O	213.46	49.02	31.67	8	- 3.0	- 0.8

The predicted EACN values are -3.0 and -0.8 units for ethyl hexanoate with the NN-6N and the GM-5N models respectively. The same way, predictions computed for homologous series including the ethyl alkanoates are presented in Figure 2.13b. The homologous set designed to explore the effectiveness of the two models is constructed as follows. The picked homologous series are the nine chemical families already mentioned in the Figure 2.11 scatter plot legends, from cyclohexanes to alkan-2-ones. Indeed, all the molecules belonging to those families have a *n*-alkyl chain backbone of increasing size that contain either: (i) one terminal functional group (esters, ketones and nitriles), (ii) one central carbon substituted with an oxygen atom (ethers), (iii) one terminal unsaturation (alkenes and alkynes), (iv) a cycle in terminal position (cyclohexanes and benzenes), or (v) one terminal chain substitution with a chloride atom (1-chloroalkanes). For all the series the number of carbon atoms per molecule N_C is varied from six to eighteen, so that all the series contains 13 compounds each, and the whole set 117 compounds. Since 46 out of these belong *de facto* to the 111-molecule training set, they cannot be kept for prediction testing. Instead, they will be used as benchmarks to assess the accuracy of the model predictions for each series. The σ-moments for the supplementary compounds of

these series are calculated as described in section 2.2.1. The data for the 117 compounds of the homologous series is available in Appendix A7 (Table A2).

The scatter plots of the EACN predictions for the two retained models and the experimental EACN versus N_C are shown in Figure 2.13 for eight series. The alkylbenzene series plot could not be represented due to an overlap with datapoints from the alk-1-yne series or the ethyl alkanoate series. This plot is available in Figure A.36 of Appendix A7.
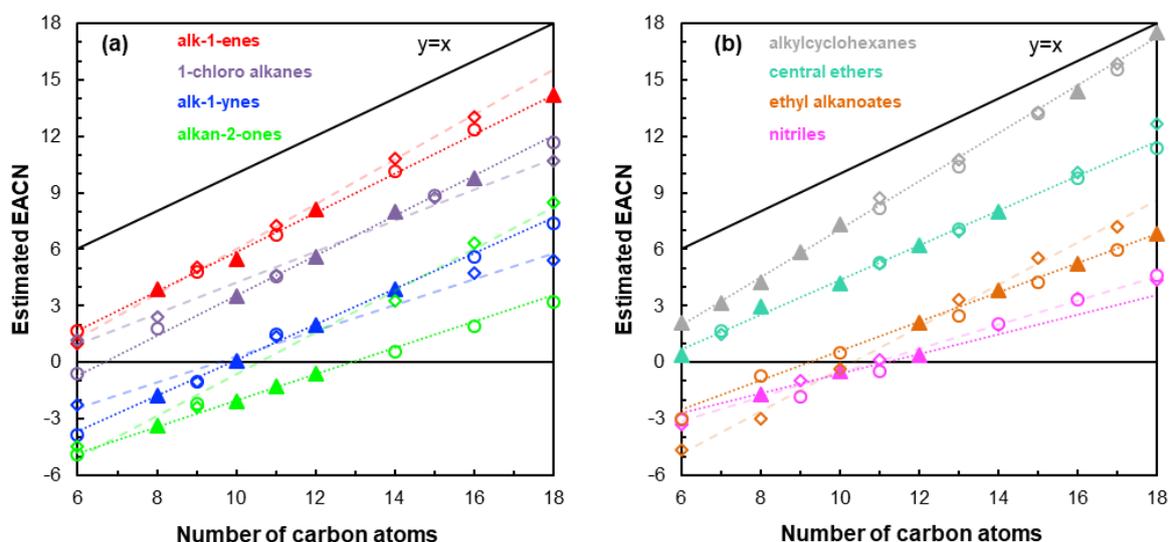
Figure 2.13. Evolution of experimental and estimated EACN with increasing number of carbon atoms (N_C) for homologous series of molecules with various chemical functions: (a) alk-1-enes, 1-chloroalkanes, alk-1-ynes and n-alkan-2-ones, (b) n-alkylcyclohexanes, central ethers, ethyl alkanoates and n-alkane nitriles. For clarity, half of the predicted values are displayed. The dotted and dashed lines indicate the experimental and neural network fits. Triangles (▲), diamonds (◇) and circles (○) are respectively markers for experimental, neural network predicted and graph machine predicted values.

As expected, the experimental linear fits (represented as dotted lines) are good for all series. The goodness of fit is further confirmed by the values of the experimental determination coefficients reported in the third column of Table 2.4, all superior to 0.99. The computed linear equations and determination coefficients for the predicted fits are also given for the nine series. With this data, the accuracy of the predictions can be analysed by comparing for all the series the proximity of the predicted points to the dotted lines (Figure 2.13), and the slopes of the GM and NN fits with the slope of the experimental fits (Table 2.4, columns 2, 4 and 6).

For the graph machine model, it can be seen in Figure 2.13 that the predictions match the experimental results quite well for seven of the nine series since most of the circles are located on or near the experimental dotted lines. Furthermore, with the exception of n-alkylbenzenes and nitriles, the slopes of the fits reported in columns 2 and 4 of Table 2.4 are very close. Regarding the 1-nitrile series, it turns out that the GM and NN models converge toward the same predictions, with almost identical slopes for their fitting equations (penultimate line of Table 2.4). Hence, we can postulate that the two model deviations from the experimental trend could be due to some experimental error. Since the experimental fit is computed with only three

Table 2.4. Linear fits for experimental and predicted EACN vs Number of Carbon Atoms N_C for homologous series. All regressions yielded R² ≥ 0.98

Family	Exp. fit[a]	GM fit[b]	NN fit[c]	$\overline{N_C - EACN_{exp}}$[d]
n-alkylcyclohexanes	1.28 N_C–5.7 (7)	1.25 N_C –5.6 (6)	1.19 N_C –4.5	2.3
alk-1-enes	1.05 N_C –4.6 (4)	1.06 N_C –4.7 (9)	1.16 N_C –5.6	4.0
central ethers	0.92 N_C –4.9 (5)	0.89 N_C –4.5 (8)	0.98 N_C –5.5	5.9
1-chloroalkanes	1.07 N_C –7.1 (4)	1.03 N_C –6.6 (9)	0.84 N_C –4.2	6.3
ethyl *n*-alkanoates	0.78 N_C –7.2 (4)	0.77 N_C –7.2 (9)	1.13 N_C –11.8	9.8
alk-1-ynes	0.95 N_C –9.4 (4)	0.93 N_C –9.1 (9)	0.70 N_C –6.6	10.0
n-alkylbenzenes	0.93 N_C –8.9 (4)	1.05 N_C –10.6 (9)	0.98 N_C –9.7	9.7
nitriles	0.53 N_C –5.9 (3)	0.62 N_C –6.9 (10)	0.64 N_C –6.9	11.5
n-alkan-2-ones	0.70 N_C –9.1 (4)	0.66 N_C –8.5 (9)	1.11 N_C –11.8	12.7

[a]In brackets, number of points used for the experimental fits. [b]In brackets, number of points used for the graph machine fits. [c]Number of points used is the same as for the graph machine fits. [d]Average value corresponds to the difference between N_C value and experimental EACN value obtained from the fit equation (N_C ranging from 6 to 18).

successive values of N_C, a small error in a fish temperature determination could induce, as already mentioned, a deviation of up to 0.3 EACN unit. Thus, such an increase in the EACN for the dodecanenitrile value (0.3) would be sufficient to make the three linear fits match. Indeed, this modification would give a modified equation equal to $0.64N_C$–6.9 for the dotted experimental line, almost identical to the two model equations. This increased experimental value for dodecanenitrile (0.7 instead of 0.4) would also be consistent with a proportional spacing for successive EACN values for nitriles. Finally, the slightly larger slope of the alkylbenzene GM fit compared to the experimental fit is mainly due to an underestimation of the EACN by the GM model for compounds with N_C lower than 11 (see Figure A.36). This behaviour can be explained by the graph machine constructions which are different in the benzenic series depending on the length of the alkyl chain. For N_C less than or equal to 12, the root node of the graph machines is located on the benzene ring while for larger N_C it is positioned on the alkyl chain. This can be shown by fitting with different lines the GM predicted points for N_C less than or equal to 12 and N_C greater than 13, as this leads to a much better fit for the two resulting lines (R² equal to 1). We need also to point out that, as for the nitrile series, a small correction in the experimental EACN (0.3) for butylbenzene would result for the experimental fit in a slope correction large enough to equal the GM slope fit. This last remark indicates that for the nitrile and alkylbenzene series the predictions computed with the graph machine model are within the experimental margin of error.

For the neural network predictive model, the results are more mixed. With four of the nine series, namely *n*-alkylcyclohexanes, ethers, nitriles and *n*-alkylbenzenes, and admitting a small measurement error for the nitrile series, the predictions are satisfactory. On the contrary for 1-alk-1-enes, 1-chloroalkanes, alk-1-ynes and ethyl *n*-alkylalkanoates a significant deviation from the dotted trend lines, up to 2 EACN unit, is observed for the predicted points located in the extrapolated regions. A larger difference in slope between the experimental and NN fits is indeed reported in Table 2.4 (columns 2 and 6), so that dashed lines for those series have been added in Figure 2.13 to materialize this divergence between the two fits. Finally, the largest deviation is computed for compounds of the alkan-2-one series for which N_C is greater than 12; the prediction becoming erroneous beyond tridecan-2-one. No explanation has yet been found for this discrepancy.

Overall, the predictions obtained with the two models are rather concordant for all homologous series, the particular case of ketones being put aside. As a result, both models can be used to predict the EACN value for a new molecule belonging to one of these series. While the graph machine model allows to obtain a result more quickly, since it is enough to use a SMILES code, the fact of computing a prediction with both models allows to anticipate an incorrect graph machine construction if the predicted results are very different.

As stated elsewhere,[128] the oils that produce a higher difference in EACN with respect to the linear alkanes are those that have a higher affinity with the interfacial film, and from the last column of Table 2.4 this decreasing order is: *n*-alkan-2-ones > 1-nitriles > alk-1-ynes > ethyl *n*-alkanoates ≈ alkylbenzenes > 1-chloroalkanes > central ethers > alk-1-enes > *n*-alkylcyclohexanes. Finally, we need to point out that specific SMILES codes for some molecules, e.g. hex-1-yne or hexan-2-one, have been used, as explained for hexyl octanoate ester, to get a consistency among the graph machine constructions in the corresponding series. With these adjustments the graph machine predictions for most of the molecules in the series are rather efficient. The construction adjustment of the hexan-2-one graph machine is explained in Appendix A8.

5. Conclusions

The experimental determination of an EACN value by the traditional fish-tail method is a tedious, costly and time-consuming task.[127] In this work, two machine-learning models were built to estimate the EACN value of oils from their molecular structure. On the basis of 111 experimental values of EACN, estimations were performed either by nonlinear regression (neural network) from COSMO-RS σ-moments, or by regression on graphs (graph machine) derived from the SMILES codes of the molecules. In each case, the selection of the appropriate model was assessed by LOO score computation. The effectiveness of the chosen NN-6N and

GM-5N models were tested on a set of ten cosmetic and perfumery molecules. It was found that both models yielded predictions with similar and satisfactory accuracies (root-mean-square estimation errors equal to 0.7 and 0.5 respectively). Molecular structures in the test set were chosen on purpose as polyfunctional molecules for which the influence of each structural feature could not be considered independently. Multilinear regressions were shown to be efficient to predict the EACN value for monofunctional molecules[134,145], but this work is the first one regarding EACN prediction of complex polyfunctional ones. It was pointed out that for homologous molecule series, the linear evolution of EACN with the increase of chain length is an appropriate model and is well tackled by the graph machine predictor. However, the neural network model based on COSMO-RS σ-moments as descriptors, met some difficulties in estimating the evolution of EACN values for the alkan-2-one series.

Overall, the graph machine model is the most convenient model since it only needs SMILES codes as input values, whereas the neural network model requires using COSMO-RS previous to EACN estimations. A demonstration of the graph machine and neural network computations, based on the Docker free software technology (available on most operating systems) is available for download (see Appendix A8). It is for example very easy and very fast (< 0.5 s) to predict the EACN value of any liquid of moderate molecular size (M < 350 Da) that contains C, H, O, N or Cl atoms using its SMILES code only. One should keep in mind, however, that automated learning methods described in this work are appropriate for chemical structures close to those represented in the learning dataset. That is why the diversity and reliability of the dataset should be carefully accounted for. Otherwise, the prediction accuracy could be importantly decreased.

Chapter 3

❖

Aggregation behaviour and dilational viscoelasticity in relation with foaming properties of 1-*O*-dodecyl diglyceryl ether
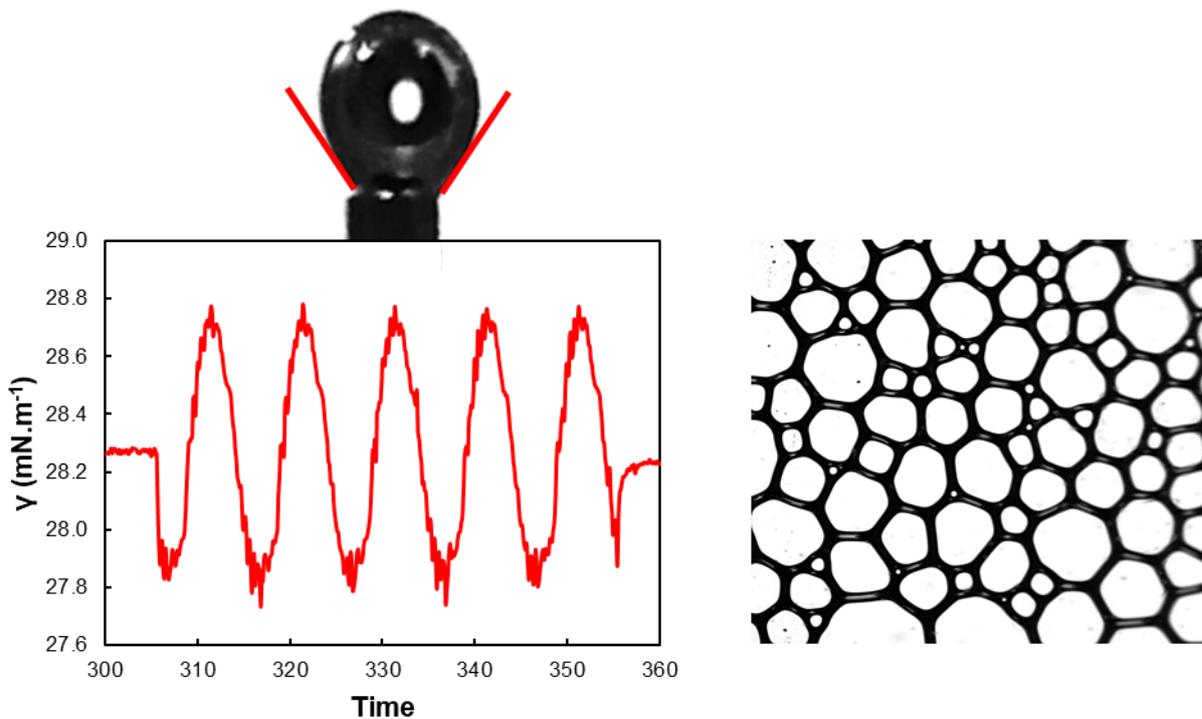
1. Introduction

Foams, consisting of gas bubbles dispersed and stabilized in a liquid or solid matrix, are found in a variety of end-use products, among which are construction materials, personal care products but also food and beverages. In an aqueous matrix, foam generation requires stabilizing the gas/water interface, usually by means of surface-active molecules, so as to achieve good foamability and foam stability.

It was shown that foamability of a surfactant solution is greatly influenced by the nature and surface properties of the surfactant. Indeed, correlation between dynamic surface tension and foamability was reported,[193–195] and the faster the surfactant adsorbs at the interface, the higher the foamability is. Also, it directly depends on the critical micelle concentration (CMC) as lower CMC corresponds to higher foamability.[196]

On the other hand, foam stability is a measure of a foam's lifetime. Three main phenomena contribute to destabilize foams, namely liquid drainage, Ostwald ripening, and bubble coalescence.[193,197,198] Slowing down or limiting those processes allows improving foam stability. First of all, it was shown that drainage (Figure 3.1a) can be limited by increasing the solution viscosity or the surface elasticity and viscosity.[193] Ostwald ripening (Figure 3.1b) is favored by large bubble size dispersity and solubility of the gas in the solution, and is thus reduced in monodisperse foams and by reducing the gas permeability through the surface surfactant layer. This is thought to be linked with surface viscoelastic properties.[199] Finally, coalescence (Figure 3.1c) can be stemmed by increasing the film resistance to rupture, which is believed to depend on the surface viscoelasticity.[194,200,201] In other words, surface viscoelastic properties intervene in all three foam destabilization phenomena, and are thus of first importance in foam studies.
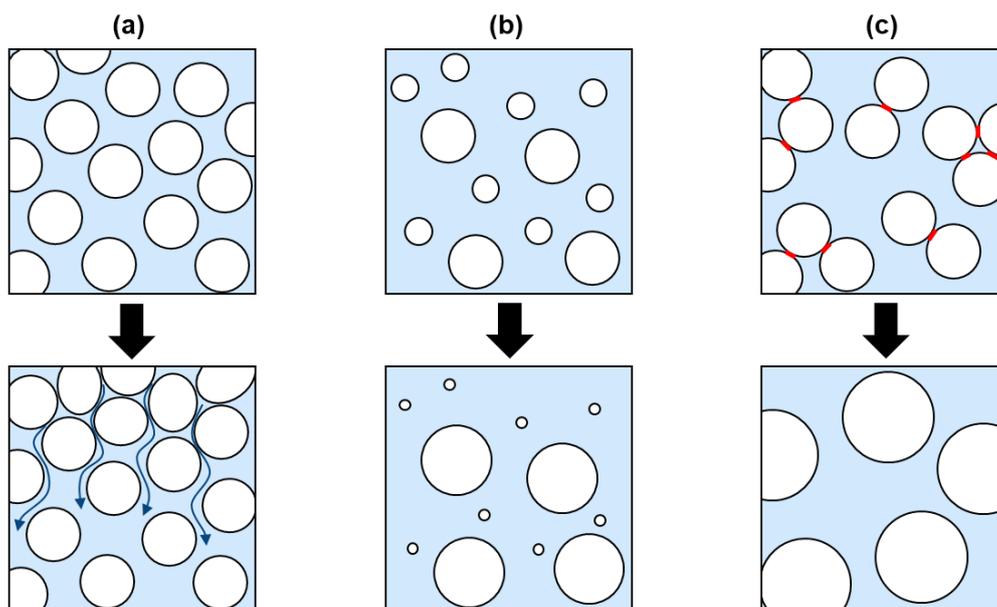
Figure 3.1. Main foam destabilization phenomena: (a) liquid drainage, (b) Ostwald ripening and (c) coalescence.

Dispersed liquid crystal (LC) phases have been shown to be efficient in stabilizing both aqueous and non-aqueous foams.[202–206] The presence of dispersed Lα lamellar liquid crystals could, indeed, stabilize foams by increasing the film viscosity and thus decreasing liquid drainage. Moreover, their contribution to foam stabilization could be attributed to the covering of bubbles by the particles, reducing gas diffusion. Finally, the presence of dispersed LC could prevent bubble coalescence by avoiding bubble collision.[205]

In this work, we report the foaming properties of the 1-O-dodecyl diglyceryl ether ($C_{12}Gly_2$) surfactant. Sagitani et al. reported the ability to form LC phase in equilibrium with aqueous solution in $C_{12}Gly_2$/water binary systems.[207] Similar structures are able to form this type of LC at low concentration, namely diglycerol fatty acid esters,[208] and were reported by Shrestha et al. as good foam stabilizers.[205] Among other nonionic surfactants, namely polyethoxylated fatty alcohols (C_iE_j) and alkylpolyglucosides (C_nGlu_m), the strength of H-bonds was found to be of importance regarding foaming properties,[209] attributed to the influence on surface viscoelasticity. Indeed, Stubenrauch et al.[210] showed the importance of intermolecular H-bonds in foam stabilization for a various C12-chain surfactants bearing different types of polar heads. Surfactants bearing oligo(ethylene oxide),[209,211–213] phosphine oxide,[214,215] trimethyl ammonium,[214,215] sarcosinate,[216] amine oxide at pH ≠ 5,[217] or carboxylic acid at pH ≠ pKa[218] as polar head produce foams that are not very stable. On the contrary, surfactants bearing a sugar-type polar head are good foam stabilizers.[209,211,212,217]

Those results point out the importance of intermolecular interactions in foam stabilization. We discuss the properties of aqueous solutions and foams stabilized by C12-chain nonionic surfactants, namely $C_{12}Gly_2$ in comparison with n-dodecyl-β-D-maltoside ($C_{12}Glu_2$) and pentaethylene glycol monododecyl ether ($C_{12}E_5$), differing by the nature of their polar head as shown in Figure 3.2, and their lyotropic behaviour.
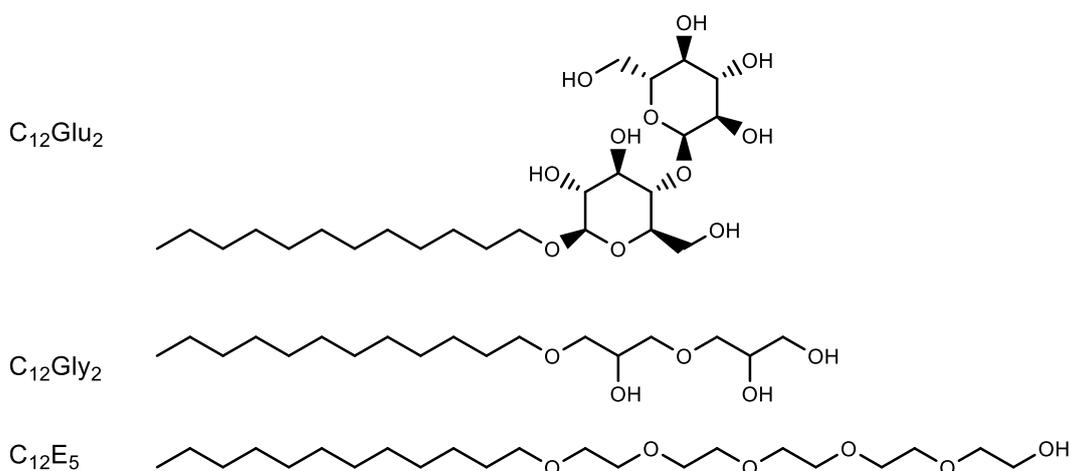
Figure 3.2. Molecular structures of n-dodecyl-β-D-maltoside ($C_{12}Glu_2$), 1-O-dodecyl diglyceryl ether ($C_{12}Gly_2$) and pentaethylene glycol monododecyl ether ($C_{12}E_5$) studied in this work.

$C_{12}E_5$ was chosen because its hydrophilicity quantified by the PIT-slope is the closest (6.8°C) to the $C_{12}Gly_2$ (4.2°C). More details about hydrophilicity quantification are developed in Chapter 4. $C_{12}Glu_2$ was chosen because its foaming and surface rheological properties are well described in literature.[200,201,209,211,213,219,220] Firstly, the aggregation behaviour in dilute solution and the behaviour of binary water/surfactant mixtures at higher concentration are discussed. Secondly, foamability and foam stability results are discussed in terms of static surface tension data as well as on dilational surface elasticities, interpreted based nature differences of the three types of polar heads and lyotropic behaviour of surfactants.

2. Materials and methods

2.1. Chemicals

Pentaethylene glycol monododecyl ether ($C_{12}E_5$, > 98.0%) was purchased from TCI chemicals and *n*-Dodecyl *β*-D-maltoside ($C_{12}Glu_2$, > 98.0%) was purchased from Sigma Aldrich. 1-*O*-dodecyl diglyceryl ether ($C_{12}Gly_2$) was synthesized in the lab according to the following procedure.

Selective 1-O-dodecyl diglyceryl ether ($C_{12}Gly_2$) synthesis (Figure 3.3). $ZnCl_2$ (8.9 g, 0.065 mol, 0.06 eq.) is dissolved in 1-dodecanol (200.5 g, 1.078 mol, 1 eq.) and heated to 100 °C. Epichlorohydrin (109.2 g, 1.180 mol, 1.1 eq.) is added dropwise for 4 h. After cooling to 50 °C, $NaOH_{aq}$ 50% (86.2 g, 1.078 mol, 1 eq.) is added dropwise for 1 h. The mixture is kept under stirring at 50 °C for 4 h. After cooling to R.T., the mixture is washed with 3x120 mL of water to remove salts. The crude is dried over $MgSO_4$, filtered and distilled under reduced pressure (7.10^{-2} mbar) between 134 and 138 °C to yield 65.2 g of a mixture of dodecyl glycidyl ether (82% GC-FID) and dodecylchlorhydrin ether (18% GC-FID). This mixture is added dropwise for 1 h to a solution of sodium solketalate prepared previously by dissolving Na(s) plates (7.4 g, 0.322 mol, 1.2 eq.) in solketal (177.0 g,1.339 mol, 5 eq.) at R.T. for 4 h then 60 °C for 20 h under N_2 flow. The mixture is stirred at 50 °C for 20 h, cooled to R.T., dissolved in 100 mL diethyl ether and washed with 3x100 mL water. The organic phase is dried over $MgSO_4$, filtered and solvent are evaporated. The crude is distilled under reduced pressure (4.10^{-2} mbar) between 160 and 165 °C to yield a colourless liquid (80.2 g, 96% GC-FID, two-steps yield = 82%). The product (35.9 g, 0.096 mol) is diluted in 120 mL methanol, trifluoroacetic acid (TFA, 1.2 mL, 0.016 mol, 1.8 g) is added and the mixture is stirred at R.T. for 96 h. The reaction is monitored by ^{1}H NMR. Once the reaction is complete, solvent and TFA are removed by rotative evaporation, yielding 1-*O*-dodecyl diglyceryl ether as a white powder (31.5 g, yield = 98%).

^{1}H NMR (300 MHz, DMSO-d_6) δ 4.55 (s broad, 3H), 3.68 (quint *a*, *J* = 5.5 Hz, 1H), 3.56 (quint *a*, *J* = 5.6 Hz, 1H), 3.44 – 3.22 (m, 10H), 1.46 (q, *J* = 6.6 Hz, 2H), 1.24 (s, 18H), 0.84 (t, *J* = 6.5 Hz, 3H).

^{1}H NMR spectra is available in Figure A.37 of the Appendix.
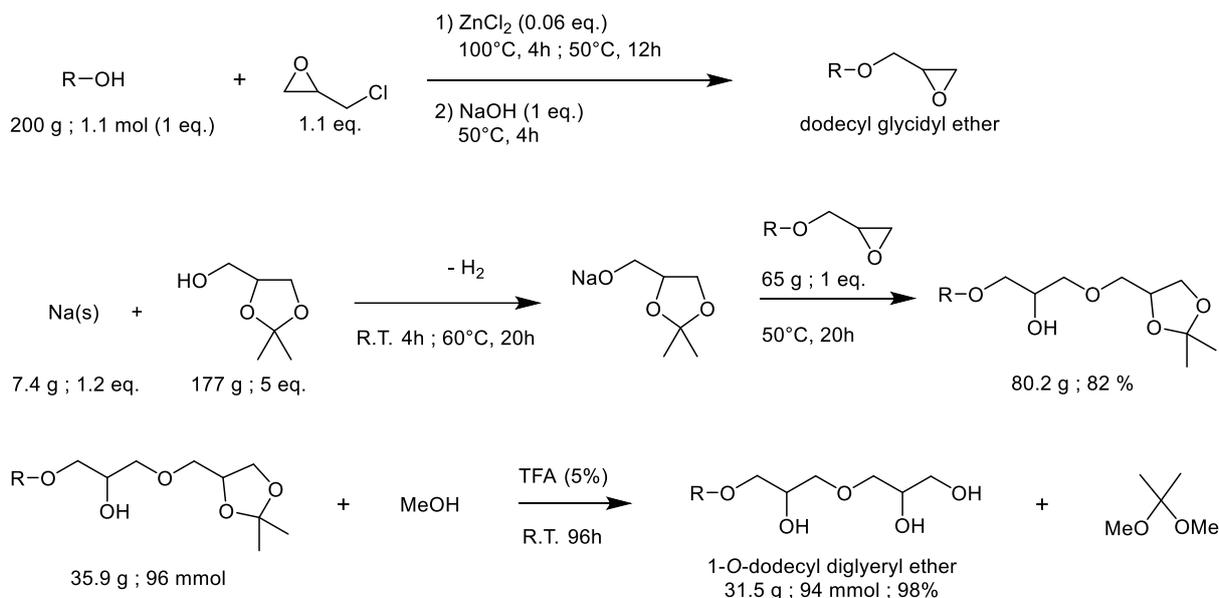
Figure 3.3. Synthetic pathway for the preparation of 1-*O*-dodecyl diglyceryl ether, with R = dodecyl.

2.2. Surface tension measurement and CMC determination

Critical micellar concentration (CMC) determination by surface tensions measurements was carried out with a Krüss K100 tensiometer (Krüss GmbH, Germany) using a Du Nouy ring. Surface tension before CMC was fitted by Langmuir-Szyszkowski equation of state given in equation (28). This model been widely applied to adsorbed surfactants, and was shown to describe well the dependence of the surface tension γ with the surfactant concentration for low molecular surfactants.[200,221] This model considers no interactions between adsorbed molecules.

$$\gamma = \gamma_0 - RT\Gamma_\infty \ln\left(1 + \frac{c}{a}\right) \qquad (28)$$

In this equation, γ is the measured surface tension, γ_0 is pure water surface tension, i.e. 72.2 mN.m^{-1}, R is the gas constant, T is the temperature, i.e. 298 K, Γ_∞ is the maximum surfactant surface concentration, c is the bulk surfactant concentration and a is the bulk concentration for which $\Gamma = \Gamma_\infty/2$. The area per molecule can then be calculated as follows.

$$A_m = \frac{1}{N_A \, \Gamma_\infty} \qquad (29)$$

where N_A is the Avogadro number.

2.3. Dilational interfacial rheology

The surface dilational rheological properties of surfactant solutions were studied at 25°C using a TRACKER™ automatic drop tensiometer (Teclis Instruments, France). Solutions at concentrations from 0.1 CMC, 0.5 CMC and 1 CMC were studied for $C_{12}E_5$ and $C_{12}Gly_2$. A

bubble of 5 µL is formed at the tip of a needle connected to a syringe in the rising drop configuration and left to rest until the interface is stabilized: the surface tension stabilization isotherm is monitored by image analysis of the contact angle between the needle and the bubble. In this work, 1 hour equilibration was sufficient. At the end of this period, 10 sinusoidal oscillations of amplitude 0.8 µL are imposed to the bubble by oscillation of the motor-driven syringe plunger. The experiment is repeated within the accessible frequency range of oscillation (10^{-2} Hz to 1 Hz), causing sinusoidal changes in the surface area and the drop shape. The changes in drop shape are monitored by a video camera, and the corresponding changes in surface tension are calculated using the TRACKER™ 2020 software. Surface tension variation with bubble area over time is processed to calculate the surface dilational visco-elasticity E according to equation (30).

$$E = \frac{d\gamma}{d\ln A} \tag{30}$$

where γ is the bubble surface tension and A is the bubble area. The surface dilational visco-elasticity E is a complex function of the perturbation frequency ν, whose real part ε_r is the dilational elasticity and the imaginary part ε_i is related to the dilational viscosity η as given in equation (31).

$$E = \varepsilon_r + i\varepsilon_i = \varepsilon_r + 2\pi\nu i\eta \tag{31}$$

In this expression, $2\pi\nu = \omega$ is the oscillation pulse.

In a diffusion-controlled adsorption system, the Lucassen-Van den Tempel model[222,223] gives the expression of the high frequency limit of dilational surface elasticity ε_0, see equation (32), and the dephasing angle φ between area deformation and surface tension variations, see equation (33).

$$\varepsilon_0 = \frac{|E|}{\sqrt{1 + 2\xi + 2\xi^2}} \tag{32}$$

$$\tan\varphi = \frac{\xi}{1 + \xi} \tag{33}$$

$$\text{with } \xi = \sqrt{\frac{\omega_0}{4\pi\nu}} \tag{34}$$

where ω_0 is the molecular exchange parameter and ν the frequency. Using equations (32) and (33), one obtains the expression of ε_0 independent of the frequency as given in equation (35).[224]

$$\varepsilon_0 = \frac{|E|}{\cos\varphi - \sin\varphi} \tag{35}$$

This ε_0 was calculated for C$_{12}$E$_5$ and C$_{12}$Gly$_2$ surfactants and for each solution, the evolution with the oscillation frequency of ε_r and η was modelled with equations (36) and (37) respectively.

$$\varepsilon_r = \varepsilon_0 \frac{1 + \xi}{1 + 2\xi + 2\xi^2} \tag{36}$$

$$\eta = \frac{\varepsilon_0}{2\pi\nu} \frac{\xi}{1 + 2\xi + 2\xi^2} \tag{37}$$

2.4. Foaming capacity and stability

Dynamic foam stability experiments are conducted using a Krüss Dynamic Foam Analyzer DFA 100 (Krüss GmbH, Germany). Foam is generated in a glass column of height 250 mm and diameter 40 mm by air sparging through a porous paper filter (pore size 12-15 µm) in 50 mL of surfactant solution at a flow rate of 0.2 L/min until a foam height of 180 mm is reached. An optical camera fixed at a height of 10 cm monitors the foam evolution (number and volume of bubbles on a certain area) for 60 minutes. The volume of both foam and solution is monitored over time, and the liquid fraction in the foam part f_{liq} is calculated as follows.

$$f_{liq}(t) = \frac{V_{i,sol} - V_{sol}(t)}{V_{foam}(t)} \tag{38}$$

with $V_{sol}(t)$ and $V_{foam}(t)$ the solution and foam volumes (mL) respectively, and $V_{i,sol}$ the initial solution volume before foam generation by air sparging. All experiments were done at least by triplicate.

3. Aggregation behaviour in aqueous solution and surface activity

The binary $C_{12}Gly_2$/water mixtures were shown to form liquid crystals (LC) in equilibrium with aqueous solution up to a concentration of 55 wt.%. Beyond this concentration, lamellar L_α phase is formed.[207] In this work, closer attention was brought to the behaviour of dilute solutions, in particular so as to approach the solubility limit of $C_{12}Gly_2$. The formation of LC phase was determined to occur between 1×10^{-4} M and 3×10^{-4} M by an increase in aggregate size measured by DLS using a Mastersizer Nano ZS (Malvern Panalytical) and shown in Figure 3.4a. Observations at the optical microscope under polarized light confirmed the presence of LC phase (Figure 3.4b and c) in equilibrium with aqueous solution and vesicle structures (Figure 3.4c). Those vesicles are formed from bilayers similar to the lamellar phase and are not thermodynamically stable, although there are exceptions, and tend evolve into a lamellar phase and a dilute aqueous solution. No change in phase behaviour was observed on a 25°C to 60°C temperature range.
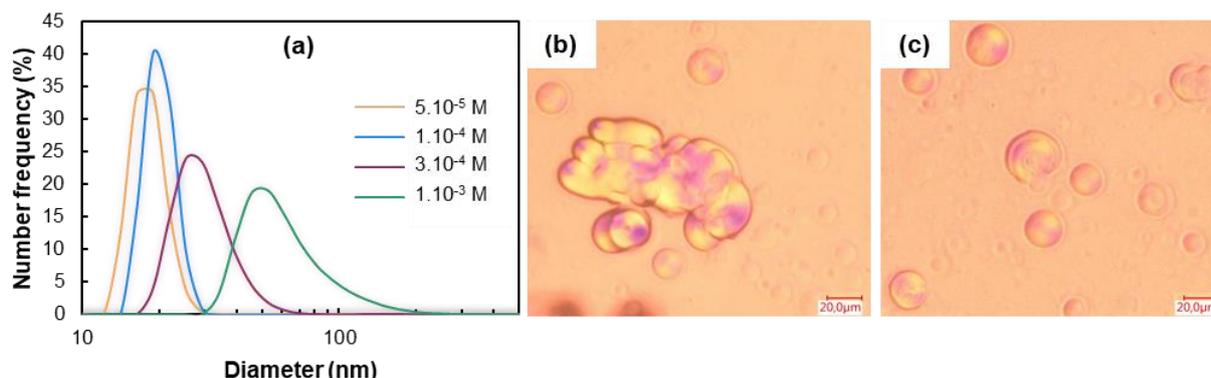
Figure 3.4. (a) Aggregate size distribution in $C_{12}Gly_2$ solutions measured by DLS at 25.0°C. Up to 1.10^{-4} M, aggregate size corresponds to micelle structures. From 3.10^{-4} M, bigger objects are formed and solutions are visibly turbid. It corresponds to the formation of LC in equilibrium with aqueous solution (b) and vesicle structures (c), observed at the optical microscope with polarized light for a 1.5×10^{-2} M solution.

The formation of LC phases is common in nonionic surfactants, in particular among C_iE_j.[225] Below the cloud point, $C_{12}E_5$ in mixture with water forms an isotropic solution L_1 at 25°C until a concentration of about 45 wt.% at which a hexagonal phase H_1 is obtained. L_α phase is formed at higher concentrations.[226] Regarding sugar-based surfactants, $C_{12}Glu_2$/water systems form isotropic solutions up to 45 wt.% and L_α phase at higher concentrations.[227] In both cases, dilute samples are isotropic solutions at room temperature. The formation of LC phases is related to interactions among polar heads, water molecules and hydrophobic chains.[228] The fact that LC form at low concentration in $C_{12}Gly_2$/water mixtures is indicative of preferential intersurfactant interactions over surfactant-water interactions, which is not the case for $C_{12}Glu_2$ and $C_{12}E_5$. Interestingly, a similar phase behaviour to that of $C_{12}Gly_2$, i.e. LC formation in equilibrium with aqueous solution at concentrations as low as 2 wt.% and vesicle formation, was reported in the case of diglycerol monolaurate, which only structural difference

from $C_{12}Gly_2$ is an ester link instead of an ether one.[208] Other oligoglycerol esters were shown to form L_α phase dispersions at low concentrations.[208,229]

The behaviour of $C_{12}Gly_2$ solutions at much more dilute concentrations allowed the determination of a critical micelle concentration (CMC), compared to that measured by Sagitani et al.[207] Both isotherms are compared in Figure 3.5 and were fitted with a Langmuir-Szyszkowski model.
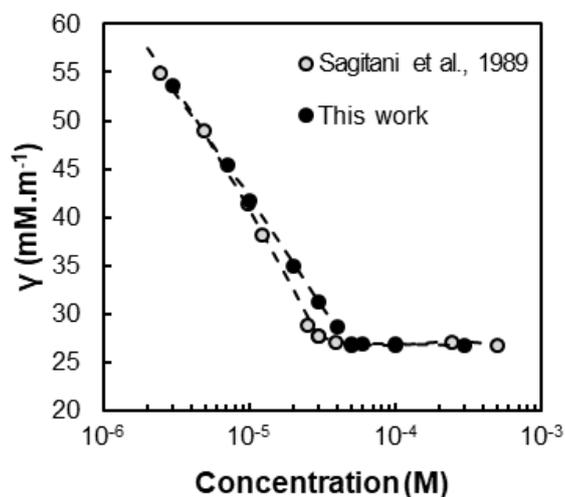
Figure 3.5. Surface tension isotherm of $C_{12}Gly_2$ at 25.0°C compared to that of Sagitani et al.[207] Data before CMC is fitted with a Langmuir-Szyszkowski model as given in equation (28).

Minimal surface tension γ_{min} = 26.8 mN.m^{-1} agrees well with the value of 27.0 mN.m^{-1} determined by Sagitani et al.[207] In comparison with $C_{12}E_5$ and $C_{12}Glu_2$, $C_{12}Gly_2$ achieves the most important surface tension diminution as γ_{min} = 30.5 mN.m^{-1} for $C_{12}E_5$,[230] and γ_{min} = 36.4 mN.m^{-1} for $C_{12}Glu_2$.[231] The CMC for $C_{12}Gly_2$ obtained here (4.7×10^{-5} M) and in the literature (3.5×10^{-5} M) are very close as can be seen in Figure 4 and they differ only by 25%. A reason for this discrepancy could be the presence of a polar impurity in the $C_{12}Gly_2$ used by Sagitani et al., responsible for faster surface tension decrease. A similar behaviour was reported by Schlarmann et al. for a $C_{10}E_4$ sample containing 1:50 molar *n*-decanol.[180] Also, the synthesis of $C_{12}Gly_2$ carried out by Sagitani et al. is not selective towards 1-*O*-dodecyl diglyceryl ether and 2-*O*-dodecyl diglyceryl ether is also formed. One could suppose that the presence of the isomer could shift the surface properties of $C_{12}Gly_2$.

Parameters of the Langmuir-Szyszkowski model and the resulting area per molecule A_m are presented in Table 3.1. The area per $C_{12}Gly_2$ molecule is higher in this work, meaning that the interface is less densely packed compared to the work of Sagitani et al., but differences in fitting parameters are small and a and Γ_∞ values are of the same orders of magnitude. Further interpretations will be based on values determined in this work.

Table 3.1. Fitting parameter of Langmuir-Szyszkowski equation for $C_{12}Gly_2$ and associated area per molecule A_m calculated from equation (29). Γ_∞ and A values are compared to those of $C_{12}Glu_2$ and $C_{12}E_5$ from literature data.

	a (M)	Γ_∞ (mol·m⁻²)	A_m (Å²)
$C_{12}Gly_2$ (This work)	5.35×10^{-7}	4.1×10^{-6}	41
$C_{12}Gly_2$ (Sagitani et al.[207])	8.95×10^{-7}	5.1×10^{-6}	33
$C_{12}Glu_2$[200]	4.59×10^{-6}	4.5×10^{-6}	37
$C_{12}E_5$[230]		3.3×10^{-6}	50

Surprisingly, the most densely packed interface is the one covered with $C_{12}Glu_2$ with the smallest area per molecule, yet the maltoside polar head is the largest of all three surfactants under study. This is indicative of important intermolecular interactions. The area per $C_{12}Gly_2$ molecule is close to that of $C_{12}Glu_2$. Finally, $C_{12}E_5$ forms the less densely packed surface.

4. Dilational surface rheology

The surface rheological properties of $C_{12}Gly_2$ and $C_{12}E_5$ solutions were determined by varying the concentration and the oscillation frequency as described in section 2.3. Data were extracted from literature for $C_{12}Glu_2$ for comparison purposes.[224] Typical examples of surface tension response to bubble volume sinusoidal oscillations at 0.5 Hz are presented in Figure 3.6 in the case of 0.1 CMC and 1 CMC $C_{12}Gly_2$ solutions. Based on those measurements, the dephasing angle φ, surface elasticity ε_r and viscosity η are calculated by Fourier transformation as described in section 2.3.
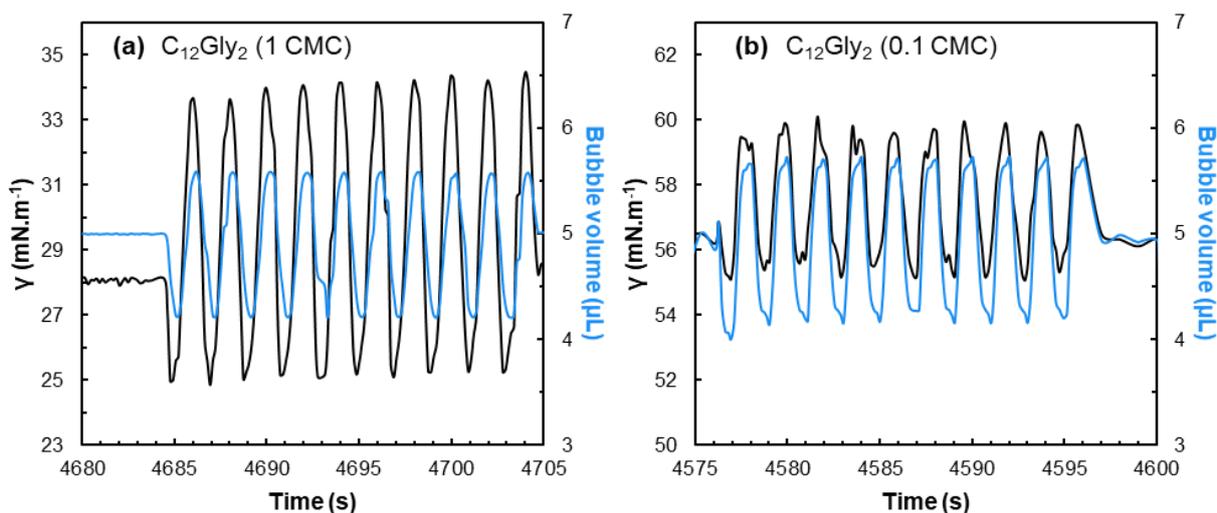
Figure 3.6. Surface tension (black) and volume (blue) of oscillating bubbles at a frequency of 0.5 Hz for (a) 1 CMC and (b) 0.1 CMC solutions of $C_{12}Gly_2$.

Both ε_r and η data presented in Figure 3.7 were fitted with the Lucassen-Van den Tempel model according to equations (36) and (37), ε_0 and ω_0 being fitting parameters. Dilational elasticities show the same tendency in ε_r evolution with the increase in frequency for all

surfactants, until reaching a plateau. This plateau, corresponding to ε_0 which values are given in Table 3.2, increases with the concentration. As the bulk concentration increases, so does the bubble surface covering and thus its elasticity at high frequencies. At the frequency range investigated in this work, the elasticity limit cannot be reached.

Similarly, the surface viscosity decreases for all three surfactants when the concentration increases. This is due to the facilitated compression of the surface at low covering, the available area per molecule being more important and the surface thus acting as a viscous surface. However, the densification of adsorbed surfactant molecules causes the need for molecules to desorb from the surface as the bubble area decreases. The evolution of both ε_r and η with increasing oscillation frequency translates the resistance of surfactant molecules towards desorption, attributed to interaction strength between molecules inside the interfacial film.
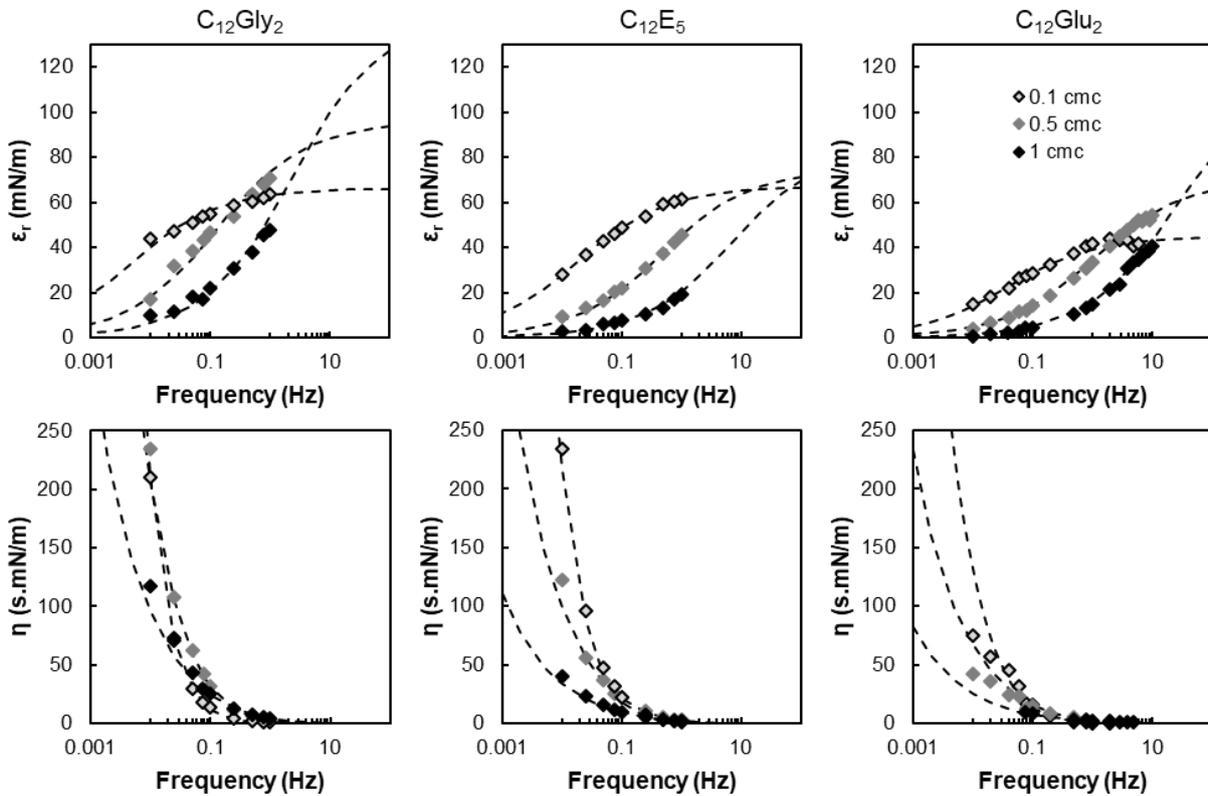
Figure 3.7. Dilational surface elasticity and viscosity of $C_{12}Gly_2$ (left, CMC = 4.7×10^{-5}M), $C_{12}E_5$ (center, CMC = 6.4×10^{-5} M)[230] and $C_{12}Glu_2$ (right, CMC = 1.5×10^{-4} M,[232] extracted and reproduced from Boos et al., 2013)[224] as a function of the oscillation frequency ν for concentrations of 0.1 CMC ($\diamond$), 0.5 CMC ($\blacklozenge$) and 1 CMC ($\blacklozenge$). For each concentration, data are fitted with equations (36) and (37).

As the bulk and surface concentration increases, so does the molecular exchange parameter. Accurate determination of ε_0 would require investigating the surface viscoelasticity at high frequencies. The elasticity limit is reached for $\nu > \omega_0/2\pi$. That is why the ε_0 plateau is reached only for the lowest concentrated solutions in Figure 3.7, as the range of frequency corresponds to the order of magnitude of ω_0. For higher concentrations than 0.5 CMC, $\omega_0/2\pi > \nu$ meaning

that the surfactant surface layer cannot be considered insoluble and purely elastic as the exchanges processes between the bulk and the surface are non-negligible.[200,233] The reader should be aware that ε_0 and ω_0 values given in Table 3.2 must be considered cautiously and are only indicative of an order of magnitude as important differences were reported for similar surfactants in similar conditions.[224]

Elasticity limit values at 1 CMC are such that $C_{12}Gly_2$ > $C_{12}Glu_2$ > $C_{12}E_5$. This trend is in accordance with the density of molecules adsorbed at the interface determined in section 3 of this chapter. However, the order of evolution is inverted for the molecular exchange parameter ω_0 which is the highest for $C_{12}Glu_2$ and lowest for $C_{12}Gly_2$. High ω_0 indicates higher molecule mobility.

Table 3.2. High frequency limit of dilational surface elasticity ε_0 (mN.m^{-1}) for $C_{12}Gly_2$, $C_{12}E_5$ and $C_{12}Glu_2$ (from Boos et al., 2013)[224] and molecular exchange parameter ω_0 (rad.s^{-1}) values for $C_{12}Gly_2$ and $C_{12}E_5$ at 0.1 CMC, 0.5 CMC and 1 CMC.

	0.1 CMC		0.5 CMC		1 CMC	
	ε_0 (mN.m^{-1})	ω_0 (rad.s^{-1})	ε_0 (mN.m^{-1})	ω_0 (rad.s^{-1})	ε_0 (mN.m^{-1})	ω_0 (rad.s^{-1})
$C_{12}Gly_2$	66.4 ± 1.5	0.030 ± 0.005	96.0 ± 3.3	0.8 ± 0.1	142.5 ± 9.5	14.7 ± 2.9
$C_{12}E_5$	67.2 ± 3.0	0.11 ± 0.06	75.1 ± 7.3	3.1 ± 0.8	84.6 ± 5.8	44.8 ± 10.8
$C_{12}Glu_2$ [224]	45.3 ± 2.7	0.24 ± 0.20[a]	70.3 ± 4.2	6.7 ± 5.4[a]	114.2 ± 6.9	149 ± 65[a]

[a] Calculated from the experimental data reported by Boos et al., 2013. By combining equations (34), (36) and (37), one obtains the following expression of the molecular exchange parameter: $\omega_0 = \frac{16\pi^3 \nu^3 \eta^2}{(\varepsilon_r - 2\pi\nu\eta)^2}$.

5. Foamability and foam stability

The first step of forming a foam is the trapping of air bubbles into the solutions, also called foamability. In this work, foaming properties are studied at a concentration of 10 CMC so that the surface concentration is sufficient to attain fast air/water interface covering. All three surfactant solutions studied showed the same foamability at the air flow rate under study (Q = 0.2 L.min^{-1}), *i.e.* the target total volume of 180 mL was attained after the same duration of air injection. The characteristics of the resulting foams and their stability are analysed based on image analysis. Some examples of foam pictures are shown in Figure 3.8 for all three surfactants under study. Foam characteristics evolution over time is depicted in Figure 3.9a-c. $C_{12}E_5$ forms the less stable foam with a total collapse after about 15 min (Figure 3.9a), a quick foam drainage in about 250 s causing bubbles to break until no more bubbles are detected after 15 min. Figure 3.8 also shows very large bubbles at 500 s, and the absence of remaining foam after 300 s. On the contrary, $C_{12}Glu_2$ forms the most stable and dense foam with almost no V_{foam} decrease, highest fraction of liquid f_{liq} and the smallest bubbles over 1 hour. The behaviour of $C_{12}Gly_2$ stabilized foam is, however, not so trivial to interpret. Indeed, V_{foam} is

maintained to over 70% of its initial value after 1 hour, but Figure 3.9b and c show that the foam is the less dense of all three with little f_{liq} and the biggest bubble size, but stable over time. Those results are not intuitive as one would expect a foam with big bubbles and thin film to break quickly.
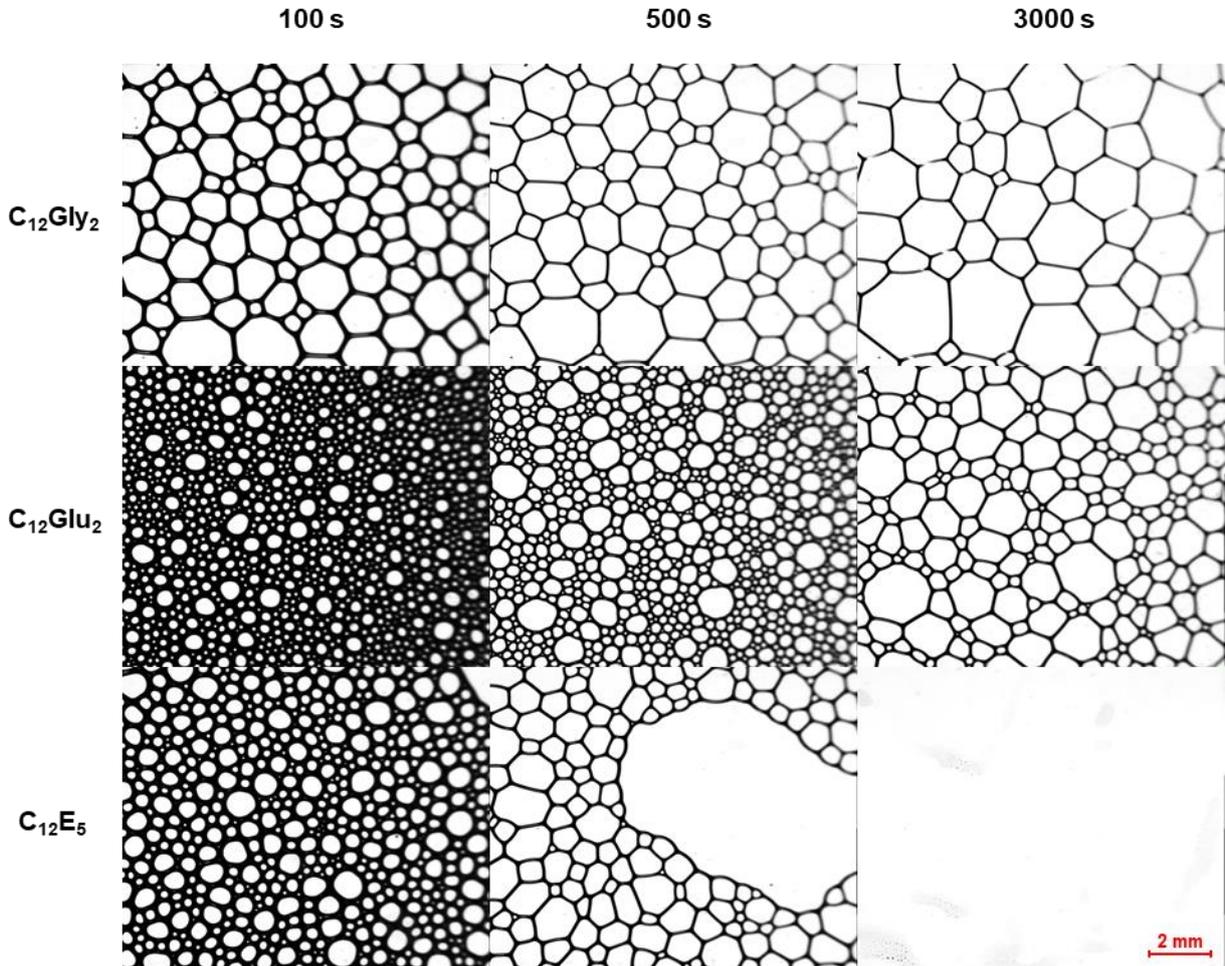
Figure 3.8. Original pictures of foam bubbles from DFA 100 data for $C_{12}Gly_2$ (top), $C_{12}Glu_2$ (middle) and $C_{12}E_5$ (bottom) solutions at 10 CMC, taken after 100 s (left), 500 s (centre) and 3000 s (right).

The formation of bubbles by air sparging is believed to be linked with the diffusion of surfactants to the newly created interface.[193] The bubbles created are the smallest for $C_{12}Glu_2$ and $C_{12}E_5$, and the largest ones are obtained with $C_{12}Gly_2$ solution as shown in Figure 3.9c. Foamability of $C_{12}Gly_2$ reported in this work is in contradiction with Stubenrauch's hypothesis that "the higher the surface elasticity the higher the resistance against shear caused by the gas flow, which, in turn, increases the foamability."[209] As shown in section 4, the elasticity limit ε_0 is the highest for $C_{12}Gly_2$, yet the initial bubble mean area is the largest compared to $C_{12}Glu_2$ and $C_{12}E_5$. One relevant parameter is, however, the molecular exchange parameter ω_0 which is the lowest for $C_{12}Gly_2$ indicating slow exchanges between bulk and interface, then increases for $C_{12}E_5$ then again for $C_{12}Gly_2$, in accordance with the trend observed in bubble initial size. Diffusion phenomena are, however, irrelevant regarding foam stability as liquid drainage was shown to be faster than molecular diffusion.[234]
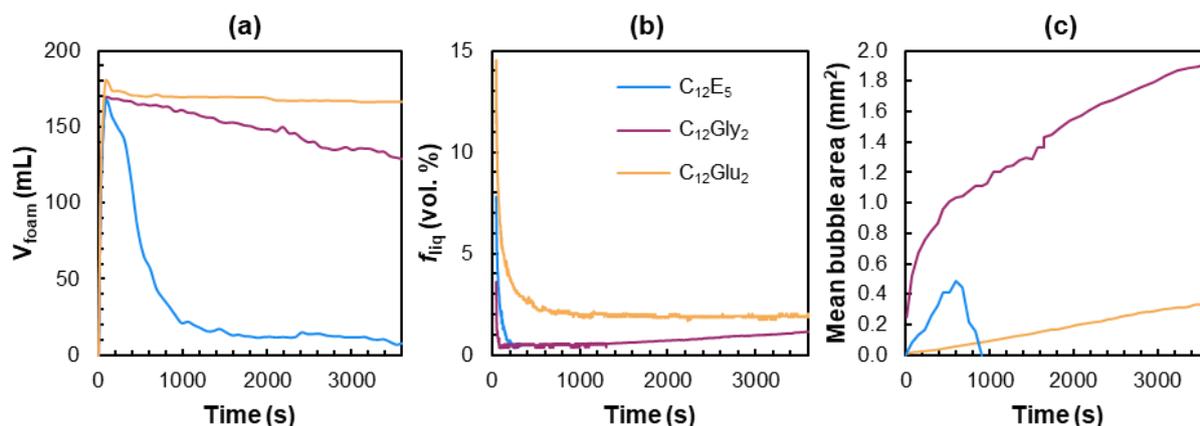
Figure 3.9. (a) Foam volume, (b) foam liquid fraction f_{liq} and (c) mean bubble area evolution over time for $C_{12}Glu_2$, $C_{12}Gly_2$ and $C_{12}E_5$ 10 CMC solutions at 25.0°C. Foam is generated and analyzed using a Dynamic Foam analyzer DFA 100 (Krüss). Procedure is descibed in section 2.4.

Elasticity limit at high frequencies ε_0 is, however, relevant as regards foam stability but results should be interpreted cautiously. Based on the hypothesis that thermally induced thickness and concentration fluctuations are responsible for foam film rupture, the high oscillation frequency range (200-800 Hz) should be of relevance in dilational rheology experiments, which would require other equipments.[235–237]

The comparison of ε_0 values suggests $C_{12}Gly_2$ should be a better foam stabilizing agent than $C_{12}Glu_2$ and $C_{12}E_5$. V_{foam} evolution (Figure 3.9a) indicates otherwise as it slightly decreases for $C_{12}Gly_2$ but remains almost constant for $C_{12}Glu_2$. Similarly, bubble size is bigger over the whole experiment duration for $C_{12}Gly_2$, partly due to initial bigger bubble size and poor foamability of $C_{12}Gly_2$. It results that one bubble rupture has more impact on V_{foam} than if bubbles were smaller. Comparatively between both surfactants, the mean bubble area increases by 1600% in $C_{12}Glu_2$ foam and by only 340% in $C_{12}Gly_2$ foam over 1 hour, implying better film stabilization against coalescence and Ostwald ripening, both phenomena contributing simultaneously to bubble size increase. It was shown that Ostwald ripening is controlled by the low frequency elasticity and coalescence by the high frequency elasticity,[238] and that for $\varepsilon_0 > \gamma/2$ Ostwald ripening is restrained.[239] Consequently, only coalescence occurs for both $C_{12}Gly_2$ and $C_{12}Glu_2$, and coalescence rate is faster for $C_{12}Glu_2$ accordingly to ε_0 values. Also, the quick stopping of liquid drainage is illustrated in Figure 3.9b by the slight increase in f_{liq} for $C_{12}Gly_2$, due to the diminution of V_{foam} while the liquid content of the foam remains constant. No discussion was based on $C_{12}E_5$ as mean bubble area evolution over 1 hour could not be calculated due to quick collapse.

Intermolecular H-bonds were shown by several studies as a crucial factor impacting foam stability, in relation with increased viscoelastic properties.[209–212,217,240,241] H-bonds are formed between hydroxyl groups as they can act as both H-bond acceptors and donors, but ether groups also act as H-bond acceptors. In this work, the maltoside polar head bears the most

hydroxyl groups (7) and forms about 5 intermolecular H-bonds and 5 H-bonds with water.[240] The diglycerol polar head in $C_{12}Gly_2$ bears 3 hydroxyls and 2 ether groups, reducing potential interactions compared to a maltoside. However, the low solubility of $C_{12}Gly_2$ points out its poor affinity for water, thus promoting intersurfactant interactions. No data were reported yet as to the number of H-bonds formed between chains. Finally, $C_{12}E_5$ only bears 1 hydroxyl and 5 ether groups, in accordance with the fact that the interface presents the largest area per molecule, the smallest surface elasticity limit and the poorest foam stability. Similar behaviour was reported for $C_{12}E_6$.[224]

One interesting point regarding the behaviour of $C_{12}Gly_2$ stabilized foam is the formation of dispersed LC as described by Sagitani et al.[207] and developed in section 3 of this chapter. At a concentration of 10 CMC (4.7×10^{-4} M), the formation of LC contributes to reducing the effective bulk concentration available for stabilizing the interface. On the other hand, as developed in the introduction, dispersed LC formed by diglycerol and oligoglycerol monoesters forming stable foams were shown to contribute to film stability by increasing the film visco-elasticity and adsorbing to the interface, thus reducing its permeability to gas.[202–206] The contribution of LC to visco-elasticity cannot be observed in dilational rheology experiments as described in section 4 given that LC are formed over the CMC and the diffusion-controlled hypothesis would not be verified.

6. Conclusions

Aggregation behaviour in water/$C_{12}Gly_2$ binary systems revealed the formation of vesicles, which are metastable structures, in dilute solutions from concentrations as low as ~10 CMC, and dispersed lamellar LC phase in equilibrium with aqueous solution. The spontaneous formation of vesicles in the dilute region indicates the applicability of $C_{12}Gly_2$ in fields where vesicles are desired. The contribution of LC to surface elasticity could not be observed due to necessity of measuring viscoelastic properties in diffusion-controlled conditions, i.e., at c ≤ CMC. However, the contribution of LC dispersion to foam stabilization was shown in diglycerol monoesters, which present very similar structures to the surfactant under study.[205]

The CMC values and minimal surface tension attained by the $C_{12}E_5$, $C_{12}Glu_2$ and $C_{12}Gly_2$ surfactants under study was not correlated to either foamability or foam stability. Dilational parameters at 1 CMC suggest low foamability but excellent foam stability, which was verified by foaming experiments by air sparging at 10 CMC. Comparison with a polyethoxylated fatty alcohol ($C_{12}E_5$) and a maltoside ($C_{12}Glu_2$) confirmed the trend that high ω_0 enhances foamability and high ε_0 enhances foam stability. The natures of polar heads support the hypothesis that the presence of intermolecular H-bonds strength accounts for the surface elastic behaviour.[209–212,217,240,241] Indeed, $C_{12}Gly_2$ and $C_{12}Glu_2$ present the most densely packed

air/water interfaces, the highest elasticity limits and the highest number of hydroxyl groups per molecule. The main foam destabilization phenomenon was identified to be bubble coalescence, which rate was inferior in the case of $C_{12}Gly_2$ compared to $C_{12}Glu_2$ and was not measurable for $C_{12}E_5$ due to quick foam break.

Further precision on high frequency elasticity limit ε_0 could be obtained using equipment able to reach higher oscillation frequency such as the capillary pressure tensiometer (CPT) method reaching frequencies up to 100 Hz.[242] Using $C_{12}Gly_2$ as foaming agent would require overcoming the poor foamability of $C_{12}Gly_2$. This could be investigated by varying the bubble generation process, e.g. by reducing the air flow or nucleating gas bubbles from the solution by gas dissolution.[193]

Chapter 4

❖

Amphiphilic behaviour of alkyl (di)glyceryl ethers within the HLD$_N$ framework: rationalization of emulsifying properties

Emulsion	O/W	unstable	W/O
HLD$_N$	< 0	0	> 0
Equilibrium	W I	W III	W II
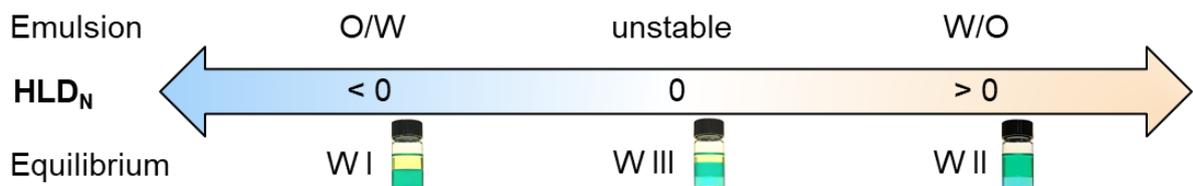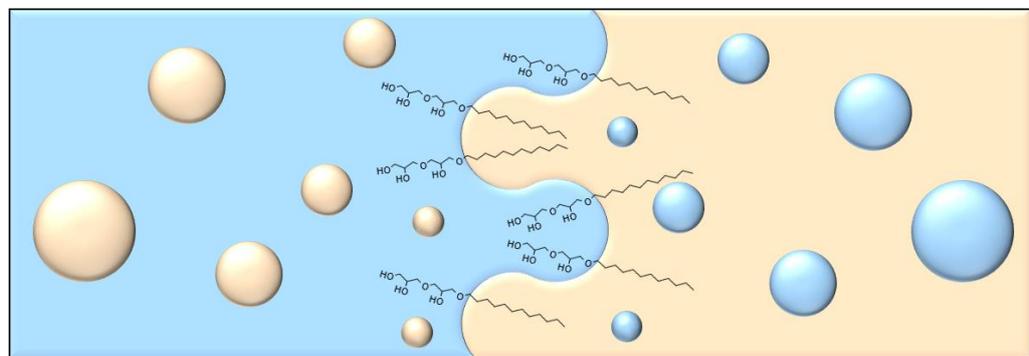
1. Introduction

The selection of a surfactant is a key factor for emulsion and microemulsion formulations. The 1-O-dodecyl diglyceryl ether (C$_{12}$Gly$_2$) surfactant, investigated in the previous chapter as foam stabilizer, was previously reported as an effective solubilizing agent forming Winsor III microemulsions at low concentration and as a promising emulsifying agent.[207] Little information is available regarding the use of alkyl (di)glyceryl ethers as emulsifiers, and determining the scope of application of a surfactant calls for quantitative and non-empirical evaluations of amphiphilic properties.

Several theories and approaches to describe the behaviour of Surfactant/Oil/Water (SOW) systems were developed. In the past century, the works of Ostwald showed that the Water-to-Oil Ratio (WOR) influences emulsion morphology, in particular when WOR is significantly different from 1, the emulsion continuous phase tends to be the phase present in larger amount.[243] Bancroft showed that at WOR close to 1, the emulsion morphology is determined by the amphiphilic agent behaviour at equilibrium and its affinity for the oil or aqueous phase. As a result, when agitating SOW systems where the surfactant is hydrophilic (or lipophilic), the resulting emulsion morphology tends to be O/W (or W/O).[244,245] Ever since, many theoretical and practical tools were developed to try and quantify this relative surfactant hydrophilicity in SOW systems, the most widely used still being the empirical Hydrophilic-Lipophilic Balance (HLB). Ontiveros et al.[246] developed an alternative surfactant classification scale based on the so-called PIT-slope, consisting of the C$_{10}$E$_4$ / n-octane / water system phase inversion temperature (PIT) disturbance. Alternatively, a similar approach based on salinity-phase-inversion (SPI) was developed by Lemahieu et al..[247] In this chapter, a comparative study of three nonionic surfactant families, namely C$_i$E$_j$, C$_n$Glu$_m$ and C$_n$Gly$_m$ is presented based on amphiphilicity quantification by both PIT-slope and SPI-slope methods.

In a second time, a further investigation of C$_{12}$Gly$_2$ emulsifying properties is undertaken within the Normalized Hydrophilic-Lipophilic-Deviation (HLD$_N$) framework for designing emulsions with desired features, i.e. morphology and stability. The HLD$_N$, which expression is given by equation (39) for nonionic surfactants has been shown to be an efficient approach for emulsion behaviour rationalization.[128,248,249]

$$HLD_N = PACN - EACN + \tau(T - 25) + \delta S \tag{39}$$

where PACN is characteristic of the surfactant and is equal to the length of the n-alkane forming a Winsor III microemulsion at 25 °C, EACN represents the oil hydrophobicity, T is the temperature (°C) and S the salinity (wt.% NaCl), τ and δ reflect the surfactant sensitivity towards temperature and salinity respectively. Both temperature and salinity sensitivity are properties of interest for the formulation of end-use products. Indeed, stability must be ensured over a range of temperature for storage. Moreover, the sensitivity of surfactants towards salts is a parameter for formulating personal care products or detergents as adjusting salinity

modifies the surfactant packing parameter, changing from micellar structures to vermicular ones in detergents, increasing viscosity to facilitate the use of the product. The HLD$_N$ equation considers the contribution of formulation variables to the relative affinity of surfactant for either the aqueous or the oil phase. When HLD$_N$ < 0, affinity is stronger towards the aqueous phase, when HLD$_N$ > 0, the affinity is stronger towards the oil phase and when HLD$_N$ = 0, the system is at the so-called "optimum formulation".[158] It concurs with a zero interfacial curvature, characteristic of an equal affinity of the surfactant for both the aqueous and the oil phase and resulting in a Winsor III microemulsion when the system equilibrates.[250] Experimentally, it also corresponds to a minimal interfacial tension between the oil and aqueous phases[251–254] and minimal viscosity of the system.[255] When temperature is the formulation variable, HLD$_N$ = 0 is reached at an equilibrium temperature T*, which is very close to the phase inversion temperature (PIT) introduced by Shinoda et al.,[256] also called the HLB temperature (T$_{HLB}$). In the same way, when salinity is the formulation variable, HDL$_N$ = 0 is reached at the salinity S*, close to the salinity of phase inversion (SPI) in a dynamic system. The parameters of temperature (τ) and salinity (δ) sensitivity are usually determined from the SOW-T and SOW-S fish diagrams studied with a series of *n*-alkanes as oils, but reaching the thermodynamic equilibrium is a long process and such experimental determination could take months. Instead, the faster dynamic inversion approaches were chosen in this work to determine PACN, τ and δ for C$_{12}$Gly$_2$.

Figure 4.1. Illustration of SOW systems behaviour with HLD$_N$ evolution when emulsified (top) and at equilibrium (bottom).

According to equation (39), the nature of the oil can be adjusted to obtain either O/W (HLD$_N$ < 0) or W/O (HLD$_N$ > 0) emulsions for which stability evolution can be anticipated, as shown in Figure 4.1. Other factors such as the emulsification process, the WOR and the surfactant concentration, not considered in the HLD$_N$ approach, were also studied for comparison purposes. This way, the scope of application of the glycerol-based 1-*O*-dodecyl diglyceryl ether (C$_{12}$Gly$_2$) surfactant was investigated: the resulting emulsions were characterized in terms of morphology, granulometry and stability monitored using a multiple light scattering device (Turbiscan®) showing the phenomena involved in emulsion destabilization.[93]

2. Description of Surfactant / Oil / Water (SOW) systems behaviour

2.1. Ostwald's and Bancroft's theories of emulsification

In 1910, Ostwald was one of the first to study SOW emulsified systems and showed that for a water-to-oil-ratio (WOR) very different from unity, the system tends to form emulsions for which the continuous phase is the one present in the largest volume, regardless of the surfactant affinity for water or oil phase.[243] In the 1910's, Bancroft also contributed as a pioneer in the comprehension of SOW emulsified systems by observing the behaviour of such systems in many conditions. He stated that "A hydrophile colloid will tend to make water the dispersing phase while a hydrophobe colloid will tend to make water the disperse phase"[244], and showed that for systems with a WOR close to 1, the emulsion morphology is determined by the amphiphilic agent behaviour at equilibrium. As a result, when agitating SOW systems where the surfactant is hydrophilic (or hydrophobic), the resulting emulsion morphology will be O/W (or W/O).[244,245] More recent works showed that the surfactant affinity for each phase and their volume proportions are not the only factors impacting the emulsion morphology: the surfactant concentration and the emulsification process are also important factors to consider.[257] The limitations of Bancroft's theory pushed scientists to develop new tools, concepts and theories to understand involved phenomena and design emulsions with desired features.

2.2. Hydrophilic-Lipophilic Balance (HLB)

The HLB concept was first introduced in 1949 by Griffin as a practical scale to determine the behaviour of a surfactant based on its relative hydrophilia and lipophilia.[123,124] In a first publication, Griffin reports some HLB values for some usual nonionic surfactants, on a scale ranging from 0 (most lipophilic) to 20 (most hydrophilic).[123] This evaluation was achieved by observation of emulsion stability formed with the surfactants but the emulsification process was not described. In 1954, the HLB calculation for a given polyethoxylated alcohol was defined, according to Griffin [124], by the relation in equation (40).

$$HLB_{Griffin} = 20 \times \frac{M_{hydrophilic}}{M_{molecule}} \tag{40}$$

with $M_{hydrophilic}$ and $M_{molecule}$ the molar weight of the hydrophilic moiety and the total molecular weight of the molecule respectively. However, this equation does not apply to other types of surfactants such as ionic ones.

In 1957, Davies proposed a competitive coalescence kinetics model to rationalize emulsion morphology and a group contribution HLB calculation so as to reflect the different natures of surfactants (nonionic, anionic and cationic).[258] According to Davies, HLB can be calculated from equation (41).

$$HLB_{Davies} = \sum H_{h,i} - \sum H_{l,i} + 7 \tag{41}$$

where $H_{h,i}$ and $H_{l,i}$ are the respective contributions of hydrophilic and lipophilic moieties in the molecule. However, contribution values are not available for every type of surfactants, *e.g.*, phospholipids or sucrose esters. Moreover, HLB values for polyethoxylated fatty alcohols (C$_i$E$_j$) calculated from equations (40) and (41) can differ from several units.

The HLB value and HLB scale is still widely used in industry due to its simplicity. Indeed, there exist a commonly accepted HLB scale corresponding to various surfactant applications, as shown in Figure 4.2. For instance, a surfactant with an HLB value ranging from 13 to 15 is, supposedly, well addressed for detergency.
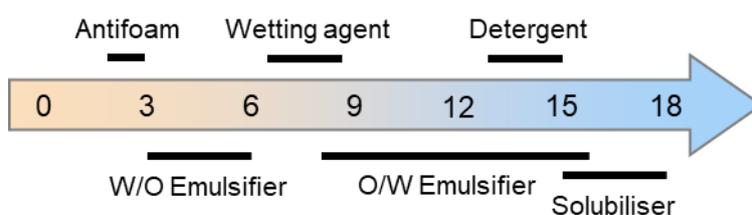
Figure 4.2. HLB scale for surfactants sorted by addressed application.

Even though the HLB value is widely used in industry, it remains an arbitrary value that hardly applies to innovative types of surfactants with new polar group as hydrophilic part.

2.3. Winsor's R-ratio

Winsor had a more thermodynamic approach of surfactant interaction with water and oil phases. Indeed, in 1948 he introduced the R-ratio, which he modified in 1954, that is defined as the ratio of surfactant interaction energy with oil, and surfactant interaction energy with water in a SOW system.[150,182] The equation was finally modified by Bourrel *et al.* in 1983 to take into account interactions between surfactant tails and polar heads as stated in equation (42).[259]

$$R = \frac{A_{so} - A_{oo} - A_{ll}}{A_{sw} - A_{ww} - A_{hh}} = \frac{A_{L,so} + A_{H,so} - A_{oo} - A_{ll}}{A_{L,sw} + A_{H,sw} - A_{ww} - A_{hh}} \tag{42}$$

with $A_{so} = A_{L,so} + A_{H,so}$ the total interaction energy between the surfactant and oil, which can be split into $A_{L,so}$ and $A_{H,so}$ that are the respective lipophilic (van der Waals interactions) and hydrophilic (hydrogen bonding) components. Similarly, A_{sw} is defined as the surfactant interaction with the water phase with $A_{L,sw}$ and $A_{H,sw}$ the lipophilic and hydrophilic components. The terms A_{oo} and A_{ww}, added in 1954, correspond to interactions between oil molecules and water molecules respectively. Finally, A_{ll} and A_{hh} correspond to the surfactant tails and polar heads interactions respectively. All considered interactions are summed up in Figure 4.3a.
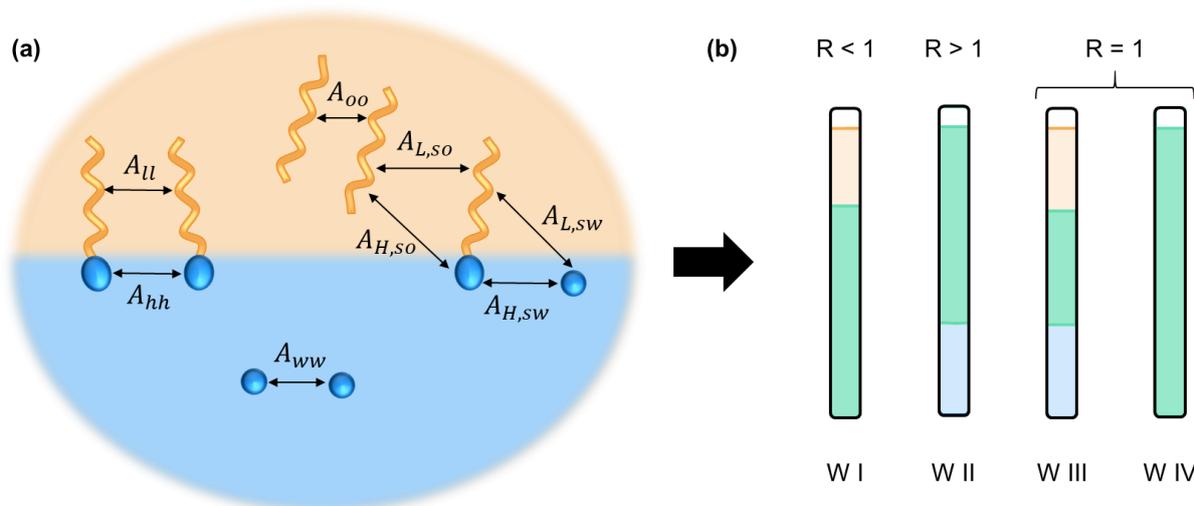
Figure 4.3. (a) Surfactant-Oil-Water interactions considered in the R-ratio calculation and (b) equilibrated SOW systems and Winsor phases associated to R-ratio values for WOR = 1. Oil phase is represented in yellow, water phase is represented in blue and microemulsion phases are represented in green.

Figure 4.3b illustrates the behaviour of SOW systems at thermodynamic equilibrium. If hydrophilic interactions are predominant over lipophilic ones, R < 1 and the surfactant will most likely be in the aqueous phase, solubilizing a part of the oil phase in it. Such microemulsion is called a Winsor I-type microemulsion (W I). In the same way, if lipophilic interactions are predominant, *i.e.* R > 1, the surfactant has greater affinity for the oil phase and the resulting microemulsion is a Winsor II (W II). When the surfactant affinity is equal for both the aqueous and the oil phase, R = 1 and the resulting microemulsion contains equal volumes of oil and water. Depending on the surfactant concentration, the microemulsion is either a Winsor III (W III) with excess water and oil or a Winsor IV (W IV) with total solubilization of both oil and water.

The concept of R-ratio can be used to rationalize the behaviour of equilibrated SOW systems. However, the individual contributions to the R-ratio calculation cannot be obtained independently from one another and only the resulting microemulsion phases can be observed.

2.4. Surfactant packing parameter

The surfactant affinity for either the aqueous phase or the oil phase was approached from a geometrical point of view by Israelachvili. In 1976, he described the packing parameter *p* that describes the geometrical arrangement of surfactant molecules aggregates into micelles, vermicular structures or bilayers based on their structure.[260] The packing parameter *p* is given by equation (43).

$$p = \frac{v_0}{a_0 \times l} \tag{43}$$

with v_0 the volume of the surfactant molecule lipophilic moiety and l its length whereas a_0 is the surface of the polar head when equilibrated at the interface. Figure 4.4 gives the schematic representation of a surfactant in Israelachvili's approach.
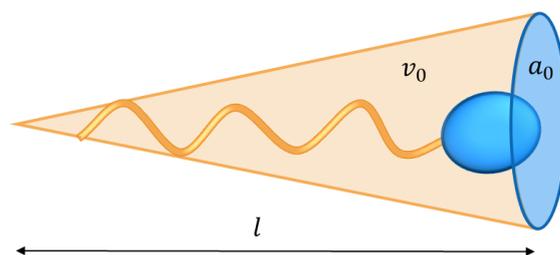
Figure 4.4. Schematic representation of a surfactant molecule in Israelachvili's packing parameter approach.

Depending on the chemical structure of the molecule and the formulation variables such as salinity or temperature, the packing parameter differs. For instance, a_0 increases with the number of ethoxylates in a C$_i$E$_j$ surfactant, but is also reduced when the temperature increases for such surfactants.[250] As a result, self-assembly of surfactants leads to structures ranging from direct micelles ($p < 1/3$), to cylindrical micelles ($1/3 < p < 1/2$), vesicles ($1/2 < p < 1$), planar bilayers ($p = 1$) and inverted micelles ($p > 1$).[22,260]

The parameters of equation (43) can be approximated by some expressions. The volume v_0 and length l can be estimated from the number of carbon atoms in the hydrophobic chain as given in equations (44) and (45), N_C being the number of atom carbons in the hydrophobic tail.[260,261] The a_0 value must, however, be experimentally determined by adsorption models for instance. The parameters for one surfactant can change depending on formulation variables, thus a precise measurement would be necessary but this would require hardly accessible scattering techniques.

$$v_0 = 27.4 + 26.9(N_C - 2) \tag{44}$$

$$l = 1.5 + 1.265(N_C - 2) \tag{45}$$

The concept of packing parameter illustrates efficiently the behaviour of surfactant molecules at the interface, and was then extended to the "effective packing parameter" concept for SOW systems by Tchakalova *et al.* so as to consider the oil penetration in the interfacial film.[262,263]

2.5. Phase Inversion Temperature (PIT), T$_{HLB}$

In 1964, Shinoda proposed a quantitative and more accessible method to evaluate the relative hydrophilicity and lipophilicity of a C$_i$E$_j$ surfactant based on its phase inversion temperature (PIT).[256] Indeed, such surfactants interact with water molecules *via* hydrogen bonds, which can be broken by increasing the temperature. As a result, the hydrophilic character of the molecule decreases when temperature increases, shifting its affinity towards the oil phase. In practice, the PIT is measured by heating a SOW system under agitation and detecting the phase

inversion. Many methods have been described to detect the PIT. The most widely used is based on conductimetry: an O/W emulsion has a significant conductivity in the presence of a small amount of electrolytes whereas a W/O emulsion, being much less dielectric, has a very weak conductivity.[264,265] Therefore, the PIT is associated to a sudden drop in conductivity which has to be monitored throughout the heating experiment. Other methods were developed, based on light scattering techniques[266] or viscosity monitoring.[267] Indeed, the phase inversion is associated to a decrease in turbidity and to a minimum in viscosity due to the minimal interfacial tension between oil and water phases, facilitating droplet deformation. At this particular temperature, the surfactant has the same affinity for both oil and water.

This method offers a quick and quantitative amphiphilicity characterization of a surfactant, and is well adapted to the C$_i$E$_j$ type of surfactants that are highly sensitive to temperature variations. The PIT measurements are however limited to the accessible temperature range and to temperature-sensitive surfactants. For instance, many ionic ones or very hydrophile nonionic ones do not enter the scope of PIT measurements.

2.6.   PIT-slope and SPI-slope methods

The principle of the PIT-slope method, developed by Ontiveros et al.,[246] is to observe the modification of the PIT value, in the same way as Shinoda defined it, of a reference SOW system with increasing amounts of various second surfactants S$_2$. The considered reference system is pure C$_{10}$E$_4$/n-octane/water, with a PIT close to room temperature.[163] Successive dynamic inversions are carried out and monitored by continuous conductivity measurements. The advantage of the PIT-slope method is that it is applicable to the assessment of the hydrophilic lipophilic balance of all kinds of surfactant: well defined pure surfactants as well as commercial mixtures. It may be considered as an alternative classification scale to the empirical HLB one. Surfactants including C$_i$E$_j$, tween, span, monoglycerides, sucrose esters, lecithins, alkyl isosorbides, alkylglucuronamides, sugar-based and glycerol-based surfactants, carboxylic acids, as well as numerous anionic and cationic surfactants were characterized accordingly.[246,268–272]

A similar method based on salinity inversions was developed by Lemahieu et al.,[247] based on the same C$_{10}$E$_4$/n-octane/water reference system. After adding a given amount of the surfactant under study S$_2$, inversions are induced by the simultaneous continuous addition of concentrated NaCl solution, and of n-octane and S$_1$ (C$_{10}$E$_4$), so as to maintain the water/oil fraction $f_w = 0.5$ and the S$_1$ concentration equal to 3 wt.%.

2.7.   Hydrophilic-Lipophilic-Deviation (HLD) theory

As briefly developed in the introduction of this chapter, the HLD$_N$ is a measure of the relative hydrophilicity and lipophilicity of a SOW system and reflects the difference in behaviour

comparatively to the optimal formulation. The first developments of the HLD concept were achieved in a context of enhanced oil recovery (EOR), for which attaining very low interfacial tension, and thus reaching the optimal formulation (HLD = 0), is crucial.[248]

The concept was first approached by Salager,[158] and named Surfactant Affinity Difference (SAD). Two expressions, corresponding respectively to nonionic polyethoxylates (equation (46)) or ionic (equation (47)) surfactants were developed.[248]

$$HLD = \frac{SAD}{RT} = \alpha - EON + bS - kACN - \varphi(A) + c_T\Delta T \tag{46}$$

$$HLD = \frac{SAD}{RT} = \sigma + \ln S - kACN - f(A) - a_T\Delta T \tag{47}$$

where S is the salinity in wt.% of NaCl, ACN or Alkane Carbon Number is characteristic of the oil phase, $f(A)$ and $\varphi(A)$ are functions of the alcohol type and the concentration respectively. σ and α are the characteristic parameters of surfactant structure. EON is the average number of ethylene oxide groups per molecule of nonionic surfactant. ΔT is the temperature deviation measured from the reference (25°C), k, a_T and c_T are empirical constants that depend on the system.

In 1982, Salager et al. showed that there exists a direct relation between the behaviour of SOW systems at equilibrium and the properties of the corresponding emulsion. By screening different formulation variables, namely the nature of the oil, the alcohol, the salinity and the proportion of two surfactants in the system, a sudden change in conductivity indicated a continuous phase inversion from water to oil or from oil to water. Close to this composition, the resulting emulsions presented a minimum in stability.[273]

More recently, a new and simpler expression was proposed. In a nonionic surfactant system containing no alcohol, expression (46) can be divided by k to yield the HLD$_N$ expression as given in equation (39), where PACN corresponds to $(\alpha - EON)/k$, $\tau = c_T/k$ and $\delta = b/k$.[128]

$$HLD_N = PACN - EACN + \tau(T - 25) + \delta S \tag{39}$$

The resulting HLD$_N$ value is thus expressed in carbon atoms units, such as the PACN and EACN values, and can be related to the characteristics of emulsions such as morphology, stability and granulometry.[128,248,249]

3. Experimental section

3.1. Chemicals

Pentaethylene glycol monododecyl ether (C$_{12}$E$_5$, > 98.0%) was purchased from TCI chemicals and *n*-Dodecyl *β*-D-maltoside (C$_{12}$Glu$_2$, > 98.0%) was purchased from Sigma Aldrich. Cyclohexane (> 99.5%), cyclooctane (> 99%), *n*-octane (98%), *n*-nonane (99%), and *n*-hexadecane (99%) were purchased from Sigma-Aldrich. *n*-hexane (> 99%) was supplied by Acros organics, *n*-heptane (99%), *n*-decane (99%) and *n*-dodecane (99%) were supplied from Alfa Aesar, *n*-undecane (> 99%), *n*-tetradecane (> 99%) and squalane (> 98%) were obtained from TCI. Octyl octanoate (>98%) was purchased from SAFC®. Pure tetraethylene glycol monodecyl ether (C$_{10}$E$_4$) used as the reference surfactant was synthesized according to a method described elsewhere.[178,179] Its purity was assessed by GC-MS analysis (> 99%) and by comparing its cloud point temperature (20.4 °C at 2.6 wt%) with the reference value (20.6 °C at 2.6 wt%).[180] 1-*O*-dodecyl diglyceryl ether (C$_{12}$Gly$_2$) was selectively synthesized in the lab according to the procedure described in section 2.1 of Chapter 3.

3.2. Phase Inversion Temperature (PIT)

S$_1$ / Oil / Water system

S$_1$ (0.085 g, 1 wt.% for C$_{12}$Gly$_2$ or 0.26 g, 3 wt.% for C$_{10}$E$_4$), *n*-alkane (4.25 g) and NaCl 10^{-2} M (4.25 g) are introduced in a double-jacketed cylindrical tube (d = 2.5 cm, h = 20 cm). The system is briefly stirred and left to pre-equilibrate at room temperature. The system is cooled down to 18 °C for 10 minutes and is kept under stirring at 500 rpm using a 2 cm-square-cross magnetic stirrer during the whole experiment. Two heating and cooling cycles are then applied at a rate of 1 °C/min by circulating water in the vessel using a HUBER 125 Ministat. Conductivity is recorded using a CDM210 conductivity meter from MeterLab® with a coupled conductivity-temperature electrode CDC641T from Radiometer Analytical®. Conductivity data is processed with the Labview software.

S$_1$+S$_2$ / Oil / Water system

S$_1$ (0.26 g, 3 wt.%), *n*-alkane (4.25 g) and NaCl 10^{-2} M (4.25 g) and are introduced in a double-jacketed cylindrical tube (d = 2.5 cm, h = 20 cm). Increasing S$_2$ fractions are added to the system so as x$_2$ (molar fraction of S$_2$ in the S$_1$/S$_2$ system) remains inferior to 0.5. The system is briefly stirred and left to pre-equilibrate at room temperature and inverted as described in the previous paragraph after each S$_2$ addition.

3.3. Salinity of Phase Inversion (SPI)

S$_1$ / Oil / Water system

The general procedure and experimental vessel used for dynamic salinity phase inversion was described by Lemahieu et al..[247,274] C$_{12}$Gly$_2$ (0.085 g), *n*-alkane (4.25 g) and water (4.25 g) are

introduced in a double-jacketed cylindrical tube (d = 2.5 cm, h = 20 cm). This C$_{12}$Gly$_2$/*n*-alkane/water mixture is briefly stirred and left to pre-equilibrate 1 h at 20.0 °C. Dynamic phase inversions are induced by increasing or decreasing continuously the aqueous phase salinity. NaCl 25 wt.% solution or pure water and C$_{12}$Gly$_2$ (2%) in *n*-alkane are added at constant rate so as to maintain the WOR equal to 1 and the surfactant concentration equal to 1%. The mixture is stirred using a 2 cm-cross magnetic stirrer and temperature is maintained at 20.0 °C by circulating water controlled by a HUBER® 125 Ministat. Phase inversions are monitored by electrical conductivity measurement as described in the previous paragraph.

S$_1$+S$_2$ / Oil / Water system

S$_1$ (0.26 g, 3 wt.%), *n*-alkane (4.25 g) and NaCl 10^{-2} M (4.25 g) and are introduced in a double-jacketed cylindrical tube (d = 2.5 cm, h = 20 cm). Increasing S$_2$ fractions are added to the system so as x$_2$ (molar fraction of S$_2$ in the S$_1$/S$_2$ system) remains inferior to 0.5. This S$_1$/S$_2$/*n*-alkane/water mixture is briefly stirred and left to pre-equilibrate 1 h at 20.0 °C. Dynamic phase inversions are induced by increasing or decreasing continuously the aqueous phase salinity. NaCl 25 wt.% solution or pure water and S$_1$ (6%) in *n*-alkane are added at constant rate so as to maintain the WOR equal to 1 and the surfactant concentration equal to 1 wt.%. The system is inverted as described in the previous paragraph after each S$_2$ addition.

3.4. Emulsions

The surfactant is dissolved in oil and the mixture surfactant/oil is sonicated if necessary and the NaCl 10^{-2} M solution (aqueous phase) is slowly added. In the first series of experiments, the mixture is agitated using an Ultra-turrax® (IKA T18/S18N-10G) at 3 krpm for 5 minutes. For studying the influence of emulsification process, a second series of experiments was carried out using a phase inversion procedure. In that case, the mixture is kept under stirring at 500 rpm using a 2 cm-square-cross magnetic stirrer. For emulsions containing *n*-hexane, *n*-heptane or *n*-octane, the temperature is set to PIT+5 °C, cooled down to PIT-5°C at a rate of 1°C/min and heated up to 25°C at 2°C/min. For emulsions containing *n*-nonane, *n*-decane, *n*-dodecane or *n*-tetradecane, the temperature is set to PIT-5 °C, increased to PIT+5°C at a rate of 1°C/min and cooled down to 25°C at 2°C/min. In all cases, emulsion morphology is assessed by conductivity measurements. A 1 mL sample is taken for size measurements and the emulsion is placed in a Turbiscan® AGS from Formulaction (see Figure 4.5a) for 14 to 28 days.

Droplet size distribution is measured using a Mastersizer® 3000 laser granulometer from Malvern Panalytical when an O/W emulsion is formed. When a W/O emulsion is obtained, the emulsion is observed at the optical microscope (Keyens VHX-900F). In that case, size distribution is calculated from at least 500 droplets per emulsion, grouped in 100 size intervals. D$_{[4,3]}$ is given as $\sum N_i D_i^4 / \sum N_i D_i^3$ where N_i is the number of observations in the size interval i of mean diameter D_i. For smaller emulsions with droplets about 1 µm, size is also measured by

Dynamic Light Scattering using a Zetasizer Nano® ZS from Malvern Panalytical. In that case, $D_{[4,3]}$ corresponds to the mean D_v diameter value. After stability monitoring, emulsions are gently re-homogenized by hand and size distributions are re-measured according to the procedures described above.

3.5. Multiple light scattering

Samples are scanned top to bottom by a laser beam (λ = 880 nm) using a Turbiscan® AGS, shown in Figure 4.5a. Detectors placed at angles of 180° and 45° record the transmitted (*TR*) and back-scattered (*BS*) light along the sample height as represented in Figure 4.5b. 54 samples can be stored at controlled temperature and monitored at the same time. At regular time intervals, samples are taken by the automatic robotic arm and placed in the analysis chamber, and placed back in the 25.0 °C storage station until the next analysis. Stability data is then processed using the Turbisoft treatment software, from which many destabilization indicators can be computed. *TR* and *BS* light signals allow visualizing the evolution of opaque and clear area over time as shown in Figure 4.5c. Indeed, the intensity of *TR* and *BS* directly depend on the concentration and size of light scattering objects according to equations (9) and (10) developed in section 2.3.2 of chapter 1.[93]
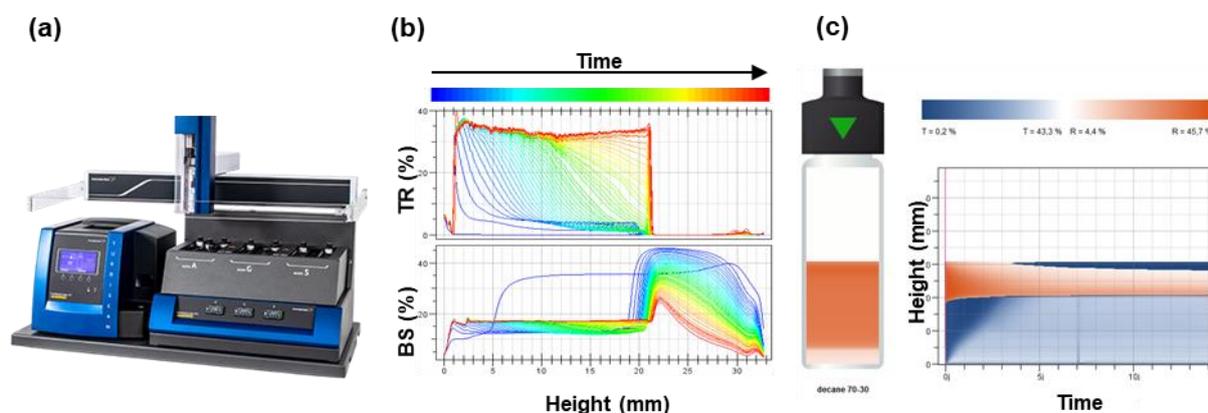
Figure 4.5. (a) Turbiscan® AGS stability analyzer, (b) transmitted (*TR*) and back-scattered (*BS*) light signals for an O/W emulsion and (c) example of virtual emulsion evolution over time generated using the Turbisoft data treatment software: blue corresponds to clear parts, i.e. transmission different from zero, and orange corresponds to opaque parts of the sample.

In this work, the evolution of internal phase released over time was calculated as an indicator of emulsion destabilization. In practice, it is obtained by measuring the peak width of *TR* signal at a threshold of $TR = TR_{max}/10$ at the top (O/W) or at the bottom (W/O) of the sample over time. The relative released volume (%) is calculated as the ratio between the volume of release internal phase and the initial volume introduced in the emulsion.

4. Amphiphilicity of alkyl (di)glyceryl ethers
4.1. Influence of nonionic polar heads (C$_i$E$_j$, C$_n$Glu$_m$ and C$_n$Gly$_m$)

Nonionic series based on sugar polar heads, namely D-glucosides C$_n$Glu and β-D-maltosides C$_n$Glu$_2$ and (di)glyceryl ethers, namely 1-O-alkyl glyceryl ethers C$_n$Gly and 1-O-alkyl diglyceryl ethers C$_n$Gly$_2$ were compared in terms of temperature and salinity sensitivities. PIT and SPI evolutions with the fraction of S$_2$ surfactant is presented Figure 4.6a-f and were measured as described in sections 3.2 and 3.3. Values are gathered in Table 4.1. They consist of the deviation behaviour from that of C$_{10}$E$_4$, the reference surfactant, due to the introduction of the second surfactant S$_2$. The homologous C$_{12}$E$_j$ series was represented for comparison purposes.[268]
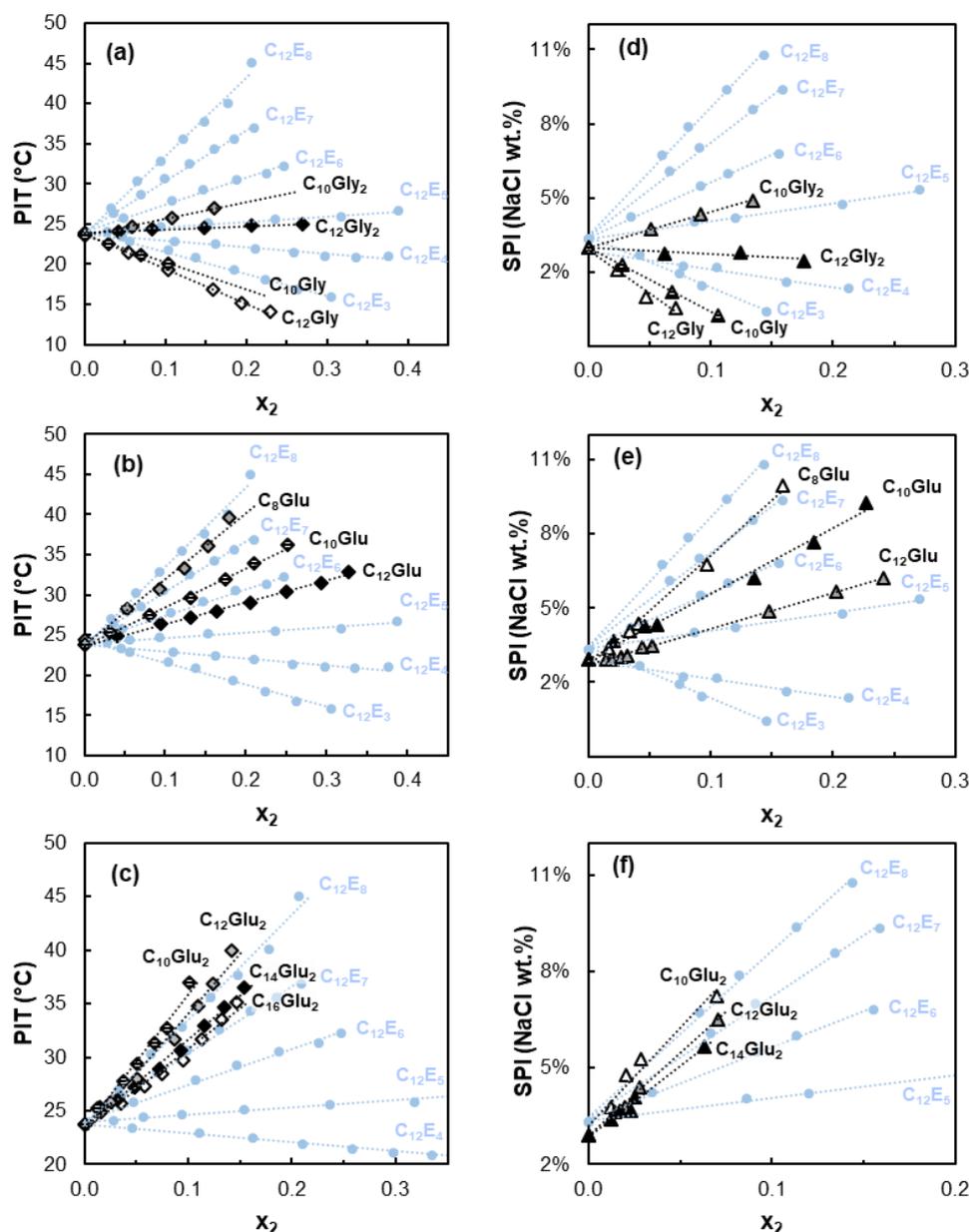
Figure 4.6. Left: PIT evolution with x$_2$ in 3 wt.% C$_{10}$E$_4$/S$_2$/n-octane/NaCl 10^{-2} M for (a) 1-O-alkyl glyceryl ethers C$_n$Gly and 1-O-alkyl diglyceryl ethers C$_n$Gly$_2$, (b) D-glucosides C$_n$Glu and (c) β-D-maltosides C$_n$Glu$_2$. Right: SPI evolution with x$_2$ in 3 wt.% C$_{10}$E$_4$/S$_2$/n-octane/water for (d) 1-O-alkyl glyceryl ethers C$_n$Gly and 1-O-alkyl diglyceryl ethers C$_n$Gly$_2$, (e) D-glucosides C$_n$Glu and (f) β-D-maltosides C$_n$Glu$_2$. C$_{12}$E$_j$ are represented as references.[247,268] PIT-slope data for C$_{12}$Gly, C$_{12}$Gly$_2$, C$_n$Glu and C$_n$Glu$_2$ were extracted from literature data.[268,271]

Table 4.1. Numerical PIT-slope and SPI-slope values for glycerol and sugar-based surfactants investigated.

Surfactant S$_2$	$dPIT/dx_2$ (°C)	$dSPI/dx_2$ (NaCl wt.%)
C$_{12}$Gly	-43	-35
C$_{10}$Gly	-34	-26
C$_{12}$Gly$_2$	4	-3
C$_{10}$Gly$_2$	21	14
C$_{12}$Glu	27	7
C$_{10}$Glu	49	24
C$_{16}$Glu$_2$	76	Not measured
C$_{14}$Glu$_2$	81	44
C$_8$Glu	82	41
C$_{12}$Glu$_2$	112	51
C$_{10}$Glu$_2$	124	60

A negative PIT-slope is indicative of a temperature sensitivity superior to that of C$_{10}$E$_4$, while a positive PIT-slope indicates inferior sensitivity to that of C$_{10}$E$_4$. The more hydrophilic the surfactant, the higher the PIT-slope is expected to be as reaching the phase inversion requires breaking more water-surfactant interactions by temperature increase.[246]

The SPI evolution with S$_2$ fraction is indicative of H-bonds interactions of water with the S$_2$ polar heads compared to that for C$_{10}$E$_4$. By NaCl salt addition, the activity of water is reduced: hydration of Na$^+$ and Cl$^-$ ions requires several water molecules per ion, disadvantaging H-bonds interactions of water with surfactants' polar heads.[275] Similarly to the PIT-slope, a negative SPI-slope is indicative of a higher salinity sensitivity to that of C$_{10}$E$_4$, whereas a positive SPI-slope indicates inferior sensitivity to that of C$_{10}$E$_4$ as reaching the phase inversion requires more NaCl addition.[247,274]

An increase of both PIT-slope and SPI-slope is observed for the C$_n$Glu$_2$ series compared to the C$_n$Glu series. The same trend in observed between the C$_n$Gly and C$_n$Gly$_2$ series. In all four series, the increase in alkyl chain length decreases PIT-slope and SPI-slope values, meaning that hydrophilicity decreases with increasing alkyl chain length. What comes out of results presented in Figure 4.6 is that C$_n$Glu and C$_n$Glu$_2$, regardless of the alkyl chain length, are more hydrophilic than C$_n$Gly and C$_n$Gly$_2$ with higher PIT-slope and SPI-slope values. C$_{12}$Gly and C$_{10}$Gly are the most lipophilic investigated surfactants with negative PIT-slope and SPI-slope. C$_{12}$Gly$_2$ has a close to zero slope in both PIT-slope and SPI-slope, meaning that its behaviour is equivalent to that of C$_{10}$E$_4$ in a water/n-octane mixture. Whereas its behaviour is similar to that of C$_{12}$E$_5$ as regards temperature, this is not verified in salinity: the diglyceryl moiety dehydrates like EO$_5$ with temperature increase but almost like an EO$_4$ with NaCl concentration. The classification order slightly varies between the PIT-slope and SPI-slope scales. Indeed, C$_{10}$Gly$_2$ responds quite similarly to temperature variations as C$_{12}$Glu, but is less sensitive to

salt addition, with an increased SPI-slope: more NaCl is required to reach the inversion. This type of surfactant could find its applications where salt tolerance is required.

As discussed by Ontiveros et al.,[246] the PIT-slope method is a practical tool to compare surfactants with one-another and it can be used as a reliable classification scale instead of the empirical HLB scale. However, it is not an absolute value as it indicates how the surfactant S$_2$ disturbs the 3% C$_{10}$E$_4$/n-octane/water system: it either increases or decreases the hydrophobicity of the system.

4.2. Non-linearity of PIT and SPI with the addition of S$_2$

The non-ideality of the PIT-slope method can be enlightened by studying surfactant binary mixtures. Among two C$_i$E$_j$ surfactants, their respective preferential affinity for the interfacial film influences both the PIT and SPI values. The example of the binary C$_{10}$E$_4$/C$_{12}$E$_6$ mixture is given in Figure 4.7, the total concentration of surfactant being maintained at 3 wt.%. Both the PIT and SPI evolutions with the increase in C$_{12}$E$_6$ derive from the ideal behaviour as a hysteresis is observed. The deviation is limited in SPI experiments compared to PIT ones. Indeed, the increase in temperature during PIT experiments modifies the solubilities of surfactants in the aqueous and oil phases; C$_{10}$E$_4$ being more sensitive to temperature increase due to a smaller number of ethylene oxide (EO), it solubilizes more easily in the oil phase rather than in the interfacial film and the behaviour tends to that of C$_{12}$E$_6$. The thermal partitioning of S$_2$ surfactants in the reference C$_{10}$E$_4$/n-octane/water system in PIT-slope and SPI-slope experiments was discussed by Lemahieu et al..[247]
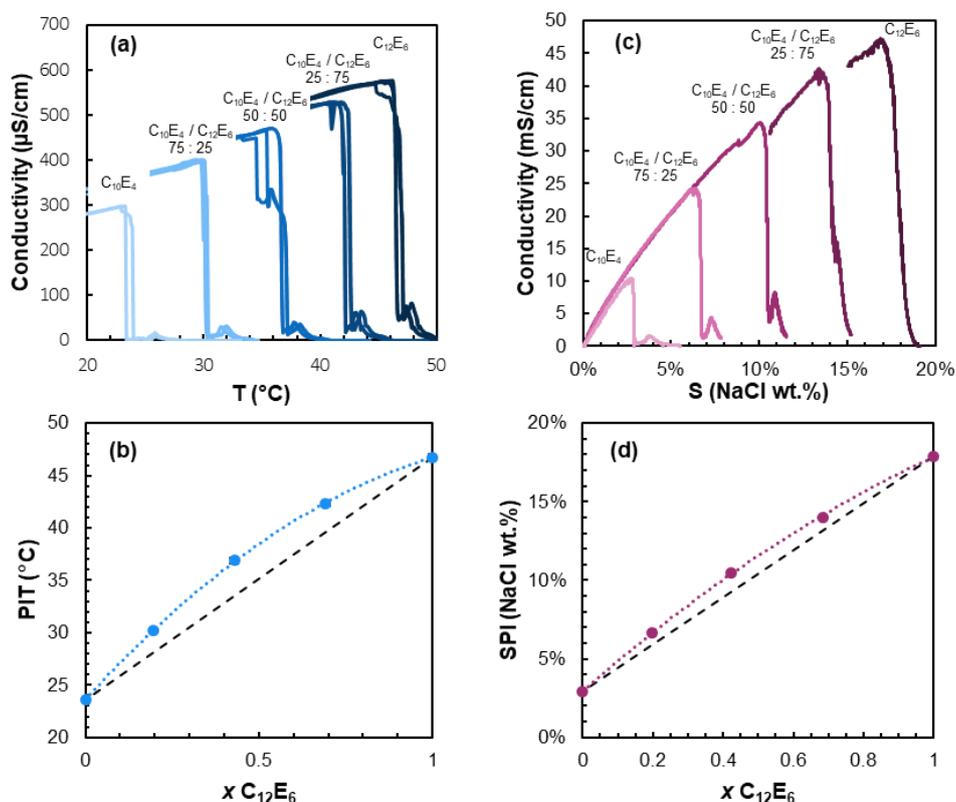
Figure 4.7. Conductivity profiles (a) and PIT evolution (b) with the increase in C$_{12}$E$_6$ content in C$_{10}$E$_4$ / C$_{12}$E$_6$ 3% wt. / n-octane / NaCl 10^{-2} M (f_w = 0.5), and conductivity profiles (c) and SPI evolution at 20.0°C (d) with the increase in C$_{12}$E$_6$ content in C$_{10}$E$_4$ / C$_{12}$E$_6$ 3% wt. / n-octane / water (f_w = 0.5). Dotted black lines in (b) and (d) indicate an ideal mixture behaviour.

The deviation from ideality is here illustrated in the case of two C_iE_j, co-solubilizing each other, and changing the properties of the interfacial film. Aubry et al.[128] reported that typical τ sensitivity values can be considered similar for C_iE_j with an alkyl chain comprising 8 to 12 carbon atoms then increases with the alkyl chain length, but can differ greatly with the nature of the polar head. Deviations are expected to increase as the nature of the polar heads of S_1 and S_2 differ.

The thermosensitivity and salino-resistance can be approached in a more accurate manner by the characteristic τ and δ parameters from equation (30). So as to measure such parameters for the alkyl (di)glyceryl ethers, $C_{12}Gly_2$ seems to be the ideal candidate. It behaves quite similarly to $C_{10}E_4$ in the *n*-octane/water (f_w=0.5) system by reaching the optimal formulation close to room temperature and in an accessible range of salinity. Indeed, extrapolation of both PIT-slope and SPI-slope values at $x_2 = 1$ predict $PIT(C_{12}Gly_2) = 27.7°C$ and $SPI(C_{12}Gly_2, 20°C) = 0.4$ wt.% in the *n*-octane/water (f_w=0.5) reference system.

4.3. Dynamic determination of HLD$_N$ parameters for C$_{12}$Gly$_2$

1-*O*-dodecyl diglyceryl ether was synthesized according to the procedure described in section 2.1 of Chapter 3 so as to obtain the pure regioisomer. However, Shi et al.[276] described a catalytic reductive etherification of diglycerol with linear aldehydes to produce a mixture of 1-*O*-alkyl diglyceryl ethers and 2-*O*-alkyl diglyceryl ethers (selectivity > 9/1). The presence of 2-*O*-dodecyl diglyceryl ether causes slight changes in physicochemical properties, which are presented in Figure A.38 and Figure A.39 of the Appendix. Chemical structures of both isomers are illustrated in Figure 4.8.

1-*O*-dodecyl diglyceryl ether   **2-*O*-dodecyl digylceryl ether**
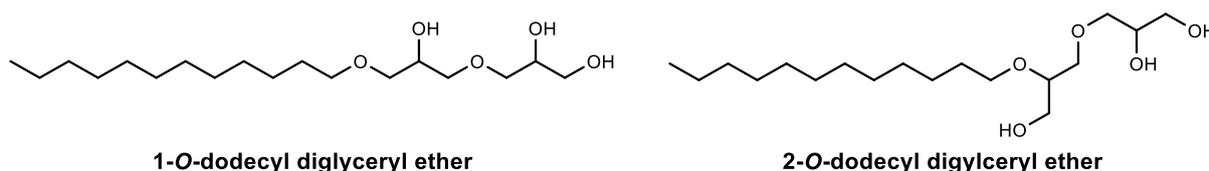
Figure 4.8. Chemical structures of 1-*O*-dodecyl diglyceryl ether and 2-*O*-dodecyl diglyceryl ether.

In this section, the HLD$_N$ parameters of $C_{12}Gly_2$ are determined by PIT and SPI in water/n-alkane systems. Indeed, τ and δ correspond to the HLD$_N$ variation triggered by a change of 1°C or 1 wt.% NaCl and are determined by studying the T and S conditions for which the phase inversion is reached (HLD$_N$ = 0).

Dynamic PIT conductivity profiles are shown in Figure 4.9a for WOR 50-50 emulsions obtained with linear alkanes ranging from *n*-octane to *n*-dodecane. As expected, the PIT value increases with ACN, as the oil hydrophobicity increases: intermolecular interactions between alkane molecules are stronger and the penetration of the surfactant is decreased. Compared to PIT values for $C_{10}E_4$ surfactant, the evolution of $C_{12}Gly_2$ PIT values is much higher when ACN increases, resulting in smaller τ value (see Figure 4.9a). PIT evolution with the linear alkanes'

length also leads the surfactant PACN value which is the ACN number corresponding to a PIT value of 25 °C. The PACN was also referred to as "N_{min}" by Wade et al. to indicate that it corresponds to the carbon number of the *n*-alkane leading to a minimum interfacial tension, i.e. the optimum formulation.[277] It should be noticed that both $C_{10}E_4$ and $C_{12}Gly_2$ have similar values of PACN (8.1-8.3 for $C_{10}E_4$[127,183,247] and 8.2 for $C_{12}Gly_2$) as they form spontaneously a Winsor III system with *n*-octane. However, their behaviour differs greatly when temperature varies. $C_{12}Gly_2$ being less "sensitive", with a τ value of 0.14 °C^{-1} against 0.40 °C^{-1} for $C_{10}E_4$, inversions require more thermic energy to occur. All C_iE_j surfactants with an alkyl chain length of at least 10 carbons have τ and δ parameters of the same orders of magnitude,[128,247] but those results show that not all nonionic polar heads have the same impact on temperature sensitivity. Ontiveros et al.[278] reported a $C_{12}Gly_2$ PACN value of 8 determined by phase equilibrium, 7.2 by the PIT-slope method, assuming that temperature coefficient was identical to that of $C_{10}E_4$, and 7.3 by the PIT-slope method assuming temperature coefficients were different but the surfactant mixture followed a linear mixing rule. Those three values agree well with ours as the $C_{12}Gly_2$ concentration is low in the PIT-slope method, limiting the error to an acceptable margin.
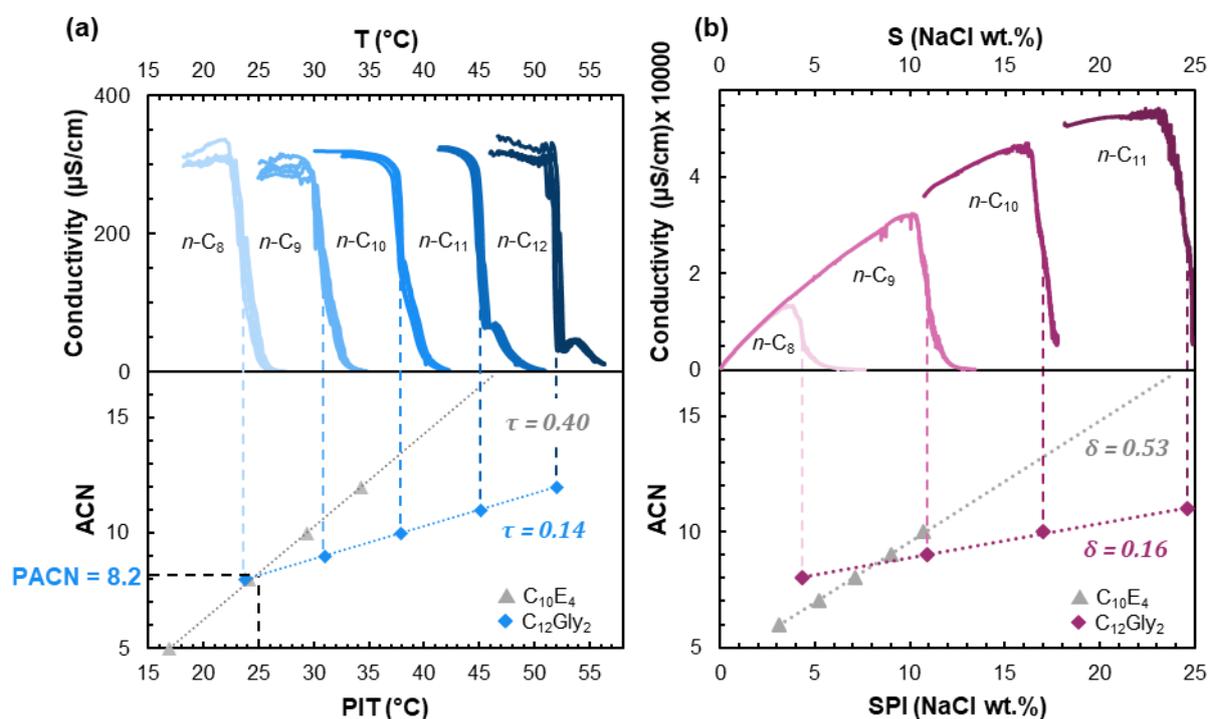
Figure 4.9. Conductivity monitored dynamic phase inversion triggered by (a) temperature (PIT) and (b) salinity (SPI) variations and their evolution depending on the n-alkane length (ACN). Emulsions are prepared at WOR=1 and the oil nature is varied from *n*-octane (ACN = 8) to *n*-dodecane (ACN = 12) or *n*-undecane in salinity screening due to NaCl solubility limitations. The τ coefficient obtained by PIT[247] and δ coefficient obtained by SPI[247] are represented for $C_{10}E_4$ for comparison.

Temperature sensitivity of nonionic surfactants is due to the formation/cleavage of hydrogen bonds between the polar head and water. As T increases, so does molecular agitation, breaking hydrogen bonds and thus reducing surfactant-water affinity and aqueous solubility of

the surfactant. Around the optimal formulation temperature, the system energy is such that water-surfactant hydrogen bonds are broken, but is insufficient for the surfactant molecules to penetrate the oil phase. Indeed, it requires breaking the oil-oil hydrophobic interactions, increasing with the *n*-alkane length. That is why the system needs more thermal energy to reach phase inversion as ACN increases.

In C$_i$E$_j$ surfactants, ether -O- groups are mostly responsible for hydrophilicity as H-bond acceptors (HBA). For instance, C$_{10}$E$_4$ contains four ether groups and one terminal hydroxyl -OH acting as both HBA and H-bond donor (HBD). In contrast, C$_{12}$Gly$_2$ contains two ether bonds and three hydroxyl groups. The relative HBD/HBA strength of those molecules can be estimated using COSMO-RS σ-profiles shown in Figure 4.10. Details about COSMO-RS calculations are given in section 2.2.1 of Chapter 2. It appears that the HBD region (negative surface charge σ) is higher in the case of C$_{12}$Gly$_2$ than for C$_{10}$E$_4$. The σ-surfaces of molecules also show more blue regions, corresponding to electron-poor regions, in the case of C$_{12}$Gly$_2$. Indeed, C$_{10}$E$_4$ and C$_{12}$Gly$_2$ have HBD σ-moments of 0.5 e and 1.2 e respectively and HBA σ-moments of 7.4 e and 6.0 e respectively. The presence of hydroxyl groups acting as both HBA and HBD increases the types of H-bonds a surfactant molecule can form with water molecules, increasing the required thermal energy to dehydrate the molecule.[279]
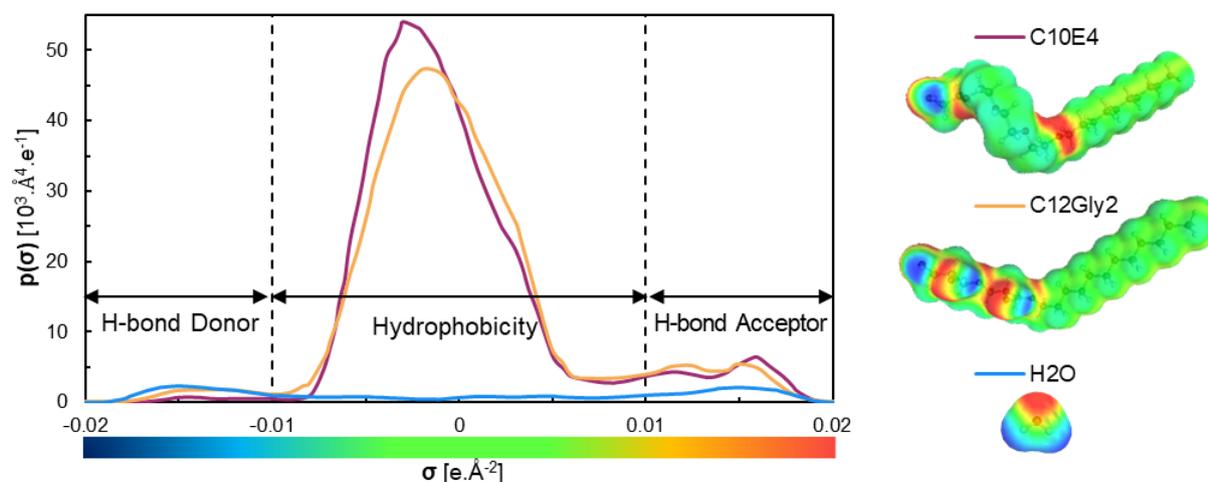
Figure 4.10. σ-profiles and σ-surfaces of C$_{10}$E$_4$ (in pink), C$_{12}$Gly$_2$ (in yellow) and water (in blue). The colour gradient corresponds to the surface charge density σ on the σ-surfaces of molecules.

The dynamic SPI evolution presented in Figure 4.9b leads to δ value. The SPI evolution with ACN for C$_{12}$Gly$_2$ is compared to that for C$_{10}$E$_4$, for which salinity sensitivity is significantly higher (δ C$_{12}$Gly$_2$ = 0.16 wt.%$^{-1}$ against 0.53 wt.%$^{-1}$). By NaCl salt addition, the activity of water is reduced: hydration of Na$^+$ and Cl$^-$ ions requires several water molecules per ion, disadvantaging H-bonds interactions of water with surfactants' polar heads.[275] As more H-bond interactions are formed between C$_{12}$Gly$_2$ and water than between C$_{10}$E$_4$ and water, the salinity sensitivity is about 3 times more important for C$_{10}$E$_4$ than C$_{12}$Gly$_2$.

5. Emulsifying properties rationalized by the HLD$_N$ theory

Little documentation is available regarding the use of alkyl glyceryl ethers as emulsifiers. In 1989, Sagitani et al.[207] compared the efficiency of 3 wt.% C$_{12}$Gly$_2$ and C$_{12}$EO$_6$ as emulsifiers in *n*-dodecane / H$_2$O (2:8 wt.) systems. They showed that the droplet size was much smaller (0.47 µm against 1.28 µm) in the case of C$_{12}$Gly$_2$, even though the emulsification protocol is unclear. Further understanding of this behaviour was investigated in this work. The HLD$_N$ value of a SOW systems gives precious information regarding the surfactant affinity towards the aqueous or the oil phase, and the resulting Winsor phase behaviour at equilibrium. Salager et al.[249] argues that the morphology and stability of emulsions, outside the equilibrium conditions, can be predicted by the HLD$_N$ value.

5.1. Granulometry and stability rationalized by HLD$_N$ evolution

The influence of the oil, characterized by the EACN value, was investigated in the framework of the HLD$_N$ theory. The water-to-oil ratio (WOR) was varied and emulsions were agitated using an Ultra-turrax® according to the procedure described in section 3.4. Stability monitoring of the emulsions was achieved using a Turbiscan so as to detect the different processes involved in emulsion destabilization.

Droplet granulometry is an indicator of phase mixing efficiency: the lower the interfacial tension, the smaller the resulting droplets.[280] D$_{[4,3]}$ measured in emulsions formed with oils of varying EACN from 2 to 16, are presented in Table 4.2. As expected, droplets are smaller close to HLD$_N$ = 0, i.e. EACN = 8, corresponding to an interfacial film with zero curvature, favouring an efficient deformation of phases. HLD$_N$ = 0 also corresponds to a minimum in interfacial tension and viscosity, improving the mixing efficiency.[251–255,281] The slow evolution of C$_{12}$Gly$_2$ emulsions after preparation allowed measuring droplet size including with *n*-octane although the destabilization is supposedly much faster when HLD$_N$ is close to 0.[282]

Table 4.2. Droplet D$_{[4,3]}$ of emulsions containing 1 wt.% C$_{12}$Gly$_2$ and prepared by varying the nature of the oil and the Water-to-Oil Ratio (wt./wt.). D$_{[4,3]}$ are measured 5 minutes after preparation.

Oil	EACN	HLD$_N$	D$_{[4,3]}$ (µm) at WOR				
			20-80	30-70	50-50	70-30	80-20
cyclohexane	2.1	6.1	48.8	55.9	34.8	15.2	7.5
cyclooctane	4.1	4.1	36.9	31.4	21.5	11.4	5.2
n-hexane	6	2.2	18.2	17.9	16.8	9.8	3.5
n-octane	8	0.2	2.9	1.9	1.7	1.5	2.0
n-decane	10	-1.8	6.5	7.6	21.4	11.8	14.2
n-dodecane	12	-3.8	8.6	13.7	28.8	19.3	24.4
n-tetradecane	14	-5.8	11.5	17.6	27.1	30	33.8
n-hexadecane	16	-7.8	12.3	20.7	32.8	34.4	36.0

Based on D$_{[4,3]}$ measurements, iso-granulometry curves shown in Figure 4.11 were calculated and presented in a formulation / composition cartography. Interfacial tension being minimal for EACN = PACN, phases are efficiently mixed and small droplets are formed. This is confirmed as for every WOR investigated, the smallest droplets are obtained with *n*-octane (ACN=8) which is the closest *n*-alkane compared to C$_{12}$Gly$_2$ PACN (8.2). In the same way, droplet size increases as EACN differs from PACN at all WOR. Interestingly, no catastrophic phase inversion was observed including for high internal phase emulsions (HIPE), i.e. WOR 20-80 and 80-20. Those HIPE present smaller droplets than emulsions using the same oil at others WORs.
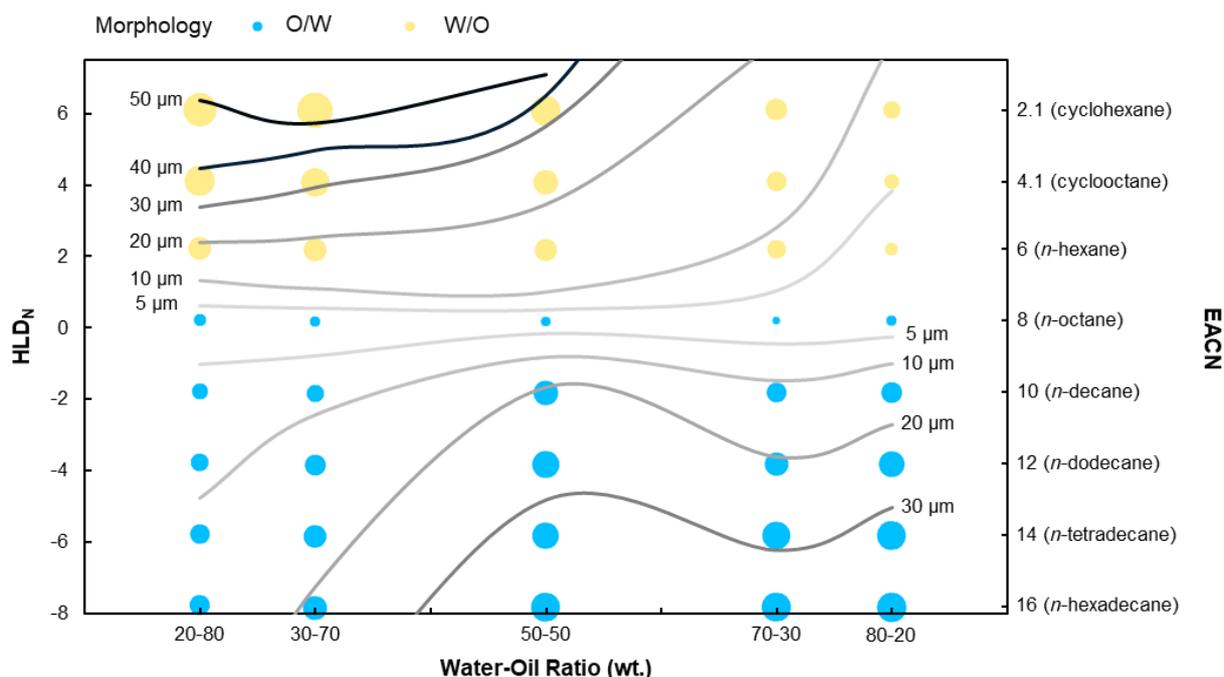
Figure 4.11. Iso-granulometry curves of emulsions containing 1 wt.% C$_{12}$Gly$_2$ and prepared by varying the nature of the oil (EACN) and the Water-to-Oil Ratio (wt./wt.). Droplet D$_{[4,3]}$ are measured 5 minutes after preparation. O/W and W/O emulsions are represented by blue dots and yellow dots respectively and symbol size is varied with D$_{[4,3]}$ values.

Stability monitoring with static multiple light scattering (Turbiscan®) allows identifying the phenomena involved in emulsion destabilization. First of all, due to density differences, emulsions tend to cream (O/W) or sediment (W/O). This sedimentation or creaming front can be seen in *BS* signals evolution,[283,284] leading eventually to an increase in transmitted light in the continuous phase region as droplets migrate, i.e. bottom for O/W and top for W/O. Secondly, aggregation and coalescence of droplets also contribute to emulsion destabilization. Droplet aggregation corresponds to droplets sticking together but no increase in diameter. On the other hand, coalescence corresponds to droplets merging together, forming bigger droplets until the internal phase is eventually released as a separate phase. Both aggregation and coalescence phenomena cause the apparent number of dispersed objects to decrease as a droplet agglomerate will scatter light the same way as one object would. Additionally, Ostwald

ripening, causing smaller droplets to diffuse into larger ones until total dissolution also contribute to decreasing the droplet number. Consequently, *BS* light signal, directly dependent on the number of dispersed objects, decreases. Examples in the case of cyclohexane W/O and dodecane O/W emulsions at WOR 50-50 are given in Figure 4.12. In both those emulsions, no internal phase release is visible. This would correspond to *TR* signal increasing at the bottom (W/O) or at the top (O/W) of the sample.
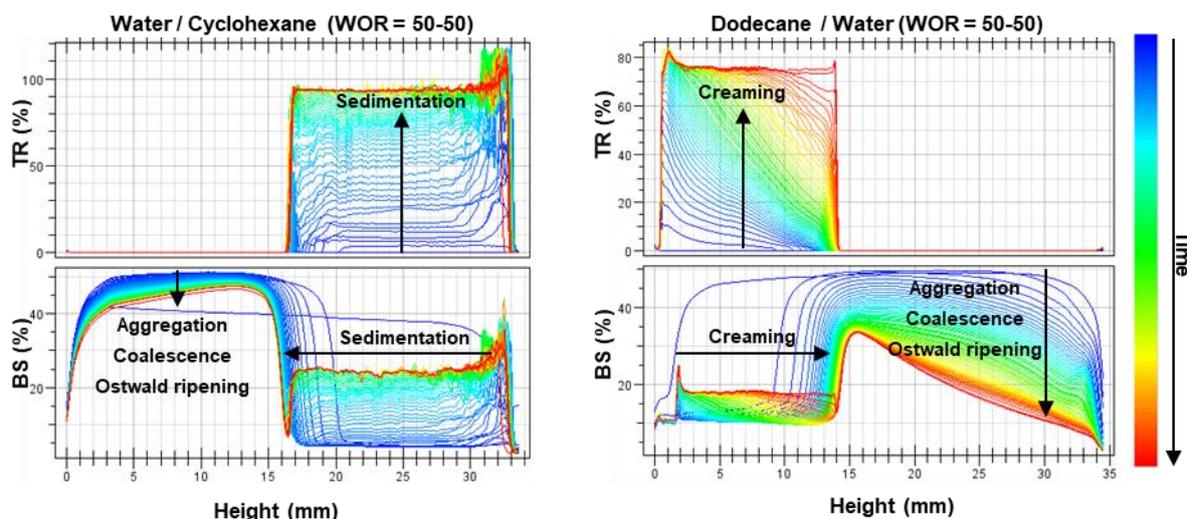
Figure 4.12. *TR* and *BS* light signals over time for water / cyclohexane (left) and dodecane / water (right) emulsions (WOR = 50-50, 1% C$_{12}$Gly$_2$). Creaming and sedimentation fronts are represented by horizontal arrows on *BS* signals and vertical arrows on *TR* evolution. Coalescence and/or aggregation and/or Ostwald ripening are visible as *BS* decreases over time.

Figure 4.13a shows the evolution of internal phase separation over 14 days. For every WOR investigated in this work, the most unstable emulsions are the ones formed with *n*-octane: fast increase in *TR* light at the top of the sample is observed in O/W emulsions for EACN = 8 (*n*-octane) evolving towards a W III microemulsion system for WOR 30-70, 50-50 and 70-30. At low water content (WOR 20-80), excess oil is quickly released due to facilitated droplet coalescence. Inversely, at high water content (WOR 80-20), oil droplets cream but the oil content being low, phase separation is only slightly visible with non-zero *TR* light in the upper part of the sample. Also, emulsions prepared with *n*-hexane (HLD$_N$ = 2.2) and *n*-decane (HLD$_N$ = -1.8), i.e. with EACN close to PACN, show fast internal phase separation.

W/O emulsions (cyclohexane and cyclooctane) showed significant internal phase release at low WOR (20-80) corresponding to the largest droplet sizes shown in Figure 4.11. On the contrary, O/W emulsions (dodecane, tetradecane and hexadecane) at high WOR (80-20) did not even though droplet sizes were comparable with those of W/O emulsions for EACN very different from PACN. O/W droplet interfaces are thus better stabilized than W/O ones, probably due to poor solubility of C$_{12}$Gly$_2$ in water compared to its solubility in cyclohexane and cyclooctane. Indeed, solubilization of the surfactant in the bulk continuous phase reduces the effective concentration of surfactant adsorbed at the droplets interface, destabilizing the

droplets. This is supported by the absence of significant internal phase release for WOR higher than 20-80 in W/O emulsions: when the amount of oil is decreased, the fraction of C$_{12}$Gly$_2$ solubilized in the continuous phase is also reduced and droplet interface stability is improved.
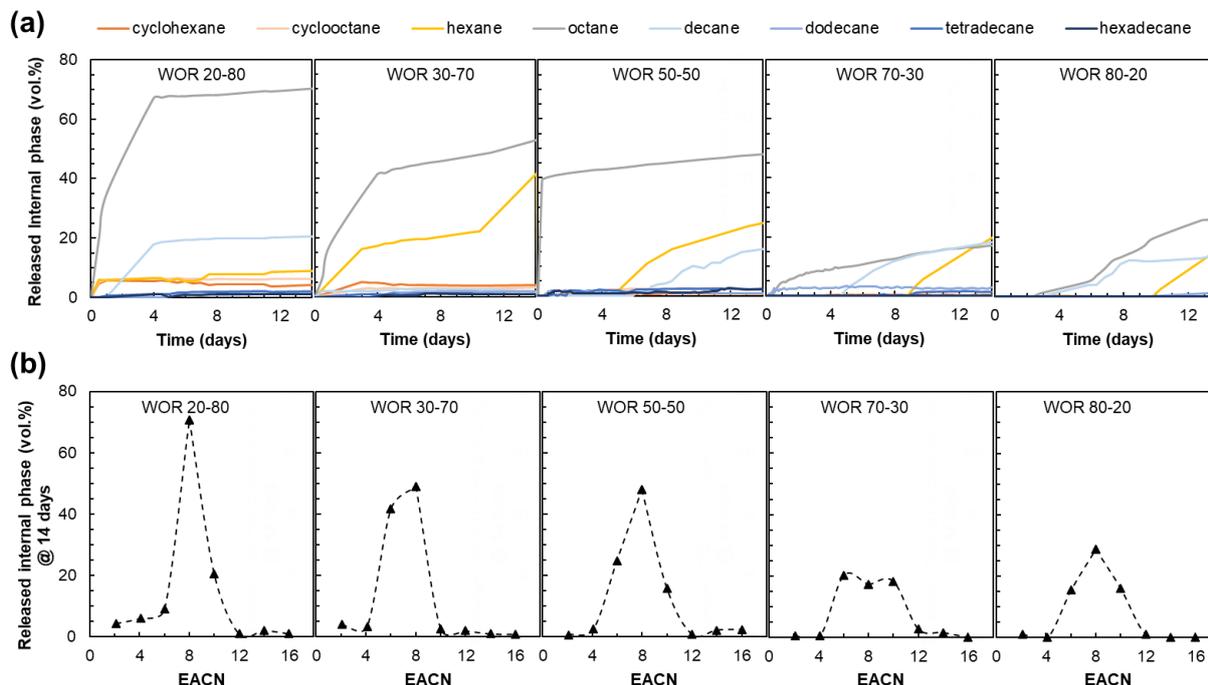
Figure 4.13. (a) Relative volume of released internal phase over time for emulsions prepared with 1% C$_{12}$Gly$_2$ and by varying the nature of the oil and the Water-to-Oil Ratio (WOR) and (b) corresponding relative volume of released internal phase after 14 days.

The volume percentage of released internal phase after 14 days shown in Figure 4.13b is in accordance with the expected evolution in the frame of HLD$_N$ theory.[249,282] The PACN of C$_{12}$Gly$_2$ being of 8.2, the most unstable emulsions are formed with n-octane (HLD$_N$ = 0.2). This is verified at WOR ranging from 20-80 to 80-20. For WOR 70-30, the volume of released internal phase is comparable for *n*-hexane, *n*-octane and *n*-decane emulsions. However, the kinetics of destabilization shown in Figure 4.13a confirms that the *n*-octane emulsion is more unstable than the *n*-hexane and *n*-decane ones.

Phenomena of coalescence and/or flocculation and/or Ostwald ripening can be quantified by looking at the *BS* light variations. A decrease in *BS* while *TR* remains null indicates a decrease in the number of light scattering objects as expressed in equation (10), caused by combinations of these objects. A way to discriminate those phenomena is to re-measure droplet size after stability monitoring: if droplet size is unchanged, droplets flocculate without coalescing or dissolving into one-another. After smoothly re-agitating the emulsions to disperse the droplets homogeneously in the sample, size was re-measured. Results are presented in Figure 4.14.

In most samples, droplet size did not change significantly over 14 days, meaning that droplet coalescence and Ostwald ripening were not preponderant destabilization phenomena.

Significant droplet size changes are, however, observed in *n*-hexane W/O emulsions (WOR 20-80, 30-70 and 80-20), *n*-decane O/W emulsions (WOR 20-80 and 30-70) and *n*-octane O/W emulsions (WOR 20-80 and 30-70). Surprisingly, one would expect the size to evolve importantly in *n*-octane emulsions as the fast destabilization was previously shown, but size re-measured after 14 days was not significantly different at WOR 50-50 to 80-20. Due to low interfacial tension, the smooth re-agitation could have been sufficient to re-form very small droplets and restore the initial state of the emulsion.
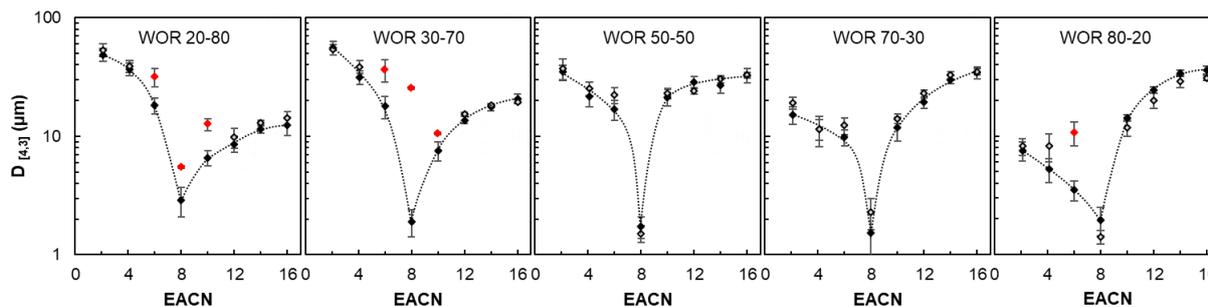
Figure 4.14. Droplet D$_{[4,3]}$ of emulsions containing 1 wt.% C$_{12}$Gly$_2$ and prepared by varying the nature of the oil (EACN) and the Water-to-Oil Ratio at initial state (◆) and after 14 days (◇). Significantly different diameters are represented in red (◆).

5.2. Process and surfactant concentration influence on emulsification

Emulsions at WOR 50-50 were prepared either by mechanical agitation using an Ultra-turrax® or by dynamic temperature phase inversion (see procedures in section 3.4). The D$_{[4,3]}$ emulsion droplet diameters are compared in Figure 4.15a for oils of EACN ranging from 6 to 14. Indeed, due to low PIT value (< 0°C, see Table 4.3), emulsions with EACN under 6 could not be inverted for practical reasons.

Table 4.3. PIT values for oils ranging from EACN 2.1 to 14 calculated from the PIT evolution with ACN presented in Figure 4.9a. Cyclohexane and cyclooctane emulsions could not be inverted as the PIT value cannot be attained.

Oil	EACN	PIT (calculated)
cyclohexane	2.1	-17.8°C
cyclooctane	4.1	-3.7°C
***n*-hexane**	6	9.7°C
***n*-heptane**	7	16.8°C
***n*-octane**	8	23.9°C
***n*-nonane**	9	30.9°C
***n*-decane**	10	38.0°C
***n*-dodecane**	12	52.1°C
***n*-tetradecane**	14	66.2°C

Phase inversion induces the change of curvature from negative to positive, passing by the 0-curvature point, i.e. the PIT. At this particular temperature, phases are sheared very efficiently due to minimal interfacial tension of the order of 10^{-2}-10^{-5} mN.m^{-1},[253,285] and viscosity[255] and the resulting droplets are much smaller than for mechanically emulsified systems.[286,287] In the case of n-octane, droplet size is comparable with both processed as the HLD$_N$ = 0 point is reached at room temperature when mechanically emulsifying. Droplet size of PIT emulsified systems tends to decrease as HLD$_N$ differs from 0. This is due to the quicker destabilization close to HLD$_N$ = 0 between preparation and size measurement (5 minutes). Destabilization kinetics is also visible in Figure 4.15b with the increase of released internal phase over time: a peak in instability is reached in Figure 4.15c for n-octane emulsion, which dephases more and faster than other emulsions for which stability is comparable with little dephasing (< 5%) observed after 14 days. Indeed, the smaller size of droplets contributes to decreasing the creaming and sedimenting processes while Brownian motion contributes to avoiding agglomeration and thus coalescence.[288]

On the other hand, increasing the surfactant concentration provides a better covering of droplets surface. Two cosmetic oils, namely octyl octanoate (EACN = 8.1)[147] which is an ester used in skin care products as an emollient and as flavouring agent (fruity, sweet taste) in food products,[289–291] and squalane (EACN = 24.4)[127] which is a branched alkane used as an emollient in skin care products, were used as oil phases in this series of emulsions. The granulometry evolution with the oil EACN is in accordance with the previously observed tendency at 1% $C_{12}Gly_2$. By increasing $C_{12}Gly_2$ concentration from 1% to 3%, not only droplet size is decreased as shown in see Figure 4.15d, but stability is also improved. The interface being better stabilized, significant phase separation (> 5 vol.%) was observed after 14 days only in the case of n-octane and octyl octanoate (HLD$_N$ ~ 0) while other emulsions were better stabilized. Also, diameter evolution after 14 days shows that no significant coalescence occurs, except for HLD$_N$ ~ 0. As depicted in Figure 4.15f, the main stability improvement compared to the 1% $C_{12}Gly_2$ emulsion series is observed for n-hexane and n-decane emulsions.

Octyl octanoate (HLD$_N$ = 0.1) initial droplet size concurs with that of n-octane (HLD$_N$ = 0.2) droplets. However, the octyl octanoate emulsion destabilizes slower: droplet size increases significantly but internal phase release appears only after 11 days, see Figure 4.15d and e. For that particular oil, the droplet size evolution after 14 days is important. This may be attributed to the higher solubility of octyl octanoate in water compared to alkanes, due to its polarity. Solubility of the dispersed phase into the continuous one favours Ostwald ripening as destabilization phenomenon. The HLD$_N$ value of the squalane emulsion being of -16.2, the formed O/W droplets are relatively big, in accordance with the general tendency observed with

other alkanes and cycloalkanes, and the emulsion is the most stable. No significant droplet size evolution or internal phase release are observed.
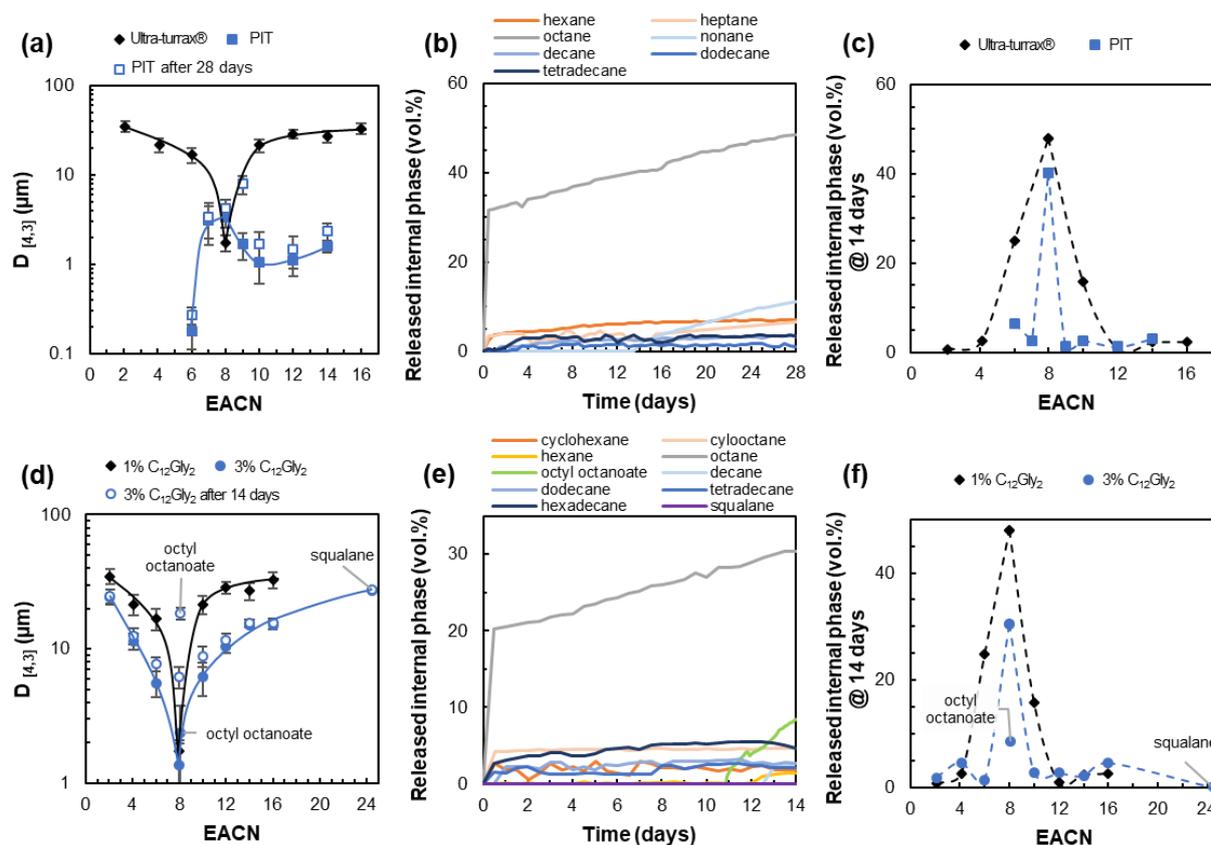
Figure 4.15. Evolution of (a) droplets D$_{[4,3]}$, (b) relative volume of released internal phase over time, (c) corresponding relative volume of released internal phase after 14 days in emulsions with oils of various EACN prepared by varying the emulsification process (WOR 50-50, 1% C$_{12}$Gly$_2$), and evolution of (d) droplets D$_{[4,3]}$, (e) relative volume of released internal phase over time and (f) corresponding relative volume of released internal phase after 14 days in emulsions with oils of various EACN prepared by varying the C$_{12}$Gly$_2$ surfactant concentration (WOR 50-50, emulsified using the Ultra-turrax® procedure).

6. Conclusions

Based on the PIT-slope classification, the glyceryl alkyl ethers studied are surfactants equilibrated between hydrophilicity and lipophilicity. Both monoglyceryl ethers C$_{10}$Gly and C$_{12}$Gly decrease the hydrophilicity of the C$_{10}$E$_4$ / n-octane / water reference system whereas diglyceryl ethers C$_{10}$Gly$_2$ and C$_{12}$Gly$_2$ tend to increase it, even if the effect of C$_{12}$Gly$_2$ is only weakly positive because its amphiphilicity at 25°C is similar to that of C$_{10}$E$_4$. By looking at the SPI-slope classification, the general trend seems to concur with the PIT-slope one. However, and interesting outcome is that for similar temperature sensitivity of C$_{10}$Gly$_2$ and C$_{12}$Glu, the diglyceryl ether has greater salt tolerance with a higher SPI-slope value. This indicates this type of surfactants for salt tolerant applications, e.g. EOR dealing with highly salted sea water. The reader should keep in mind that, even though those classification scales are quantitative, good alternative to the empirical HLB scale, and rely on experimental interfacial properties of SOW systems, they remain only comparative and refer to a reference system.

Considering the HLD$_N$ approach for designing emulsions of desired morphology, granulometry and stability was shown to be an efficient, quantitative methodology. The full HLD$_N$ equation coefficients were determined for the 1-O-dodecyl diglyceryl ether surfactant. Results outlined that temperature and salinity sensitivities are reduced by almost 3 times between SOW systems containing C$_{12}$Gly$_2$ compared to C$_i$E$_j$ surfactants (τ = 0.14°C^{-1} against 0.40°C^{-1} and δ = 0.53 wt.% against 0.16 wt.% for C$_{12}$Gly$_2$ and C$_{10}$E$_4$ respectively). A PACN value of 8.2 for C$_{12}$Gly$_2$ was determined by PIT evolution with ACN, in accordance with literature values using the PIT-slope method.[278] Differences in temperature and salinity sensitivity between C$_{12}$Gly$_2$ and C$_{10}$E$_4$ are attributed to hydroxyl groups acting as both hydrogen-bond donors and acceptors, whereas ethers bonds only act as hydrogen-bond acceptors.

A general tendency in the evolution of emulsion granulometry and stability was demonstrated with alkanes and cycloalkanes and verified with two more complex oils, namely octyl octanoate (EACN = 8.1, HLD$_N$ = 0.1) and squalane (EACN = 24.4, HLD$_N$ = -16.2). Emulsions prepared by varying the nature of the oil (EACN) intervening directly in the HLD$_N$ expression, showed significant differences in granulometry and stability. The minimum of interfacial tension between oil and water being attained for EACN = PACN, the formation of smaller droplets (about 1 μm) is favoured when using n-octane (ACN = 8) or octyl octanoate (EACN = 8.1). At every investigated emulsion WOR, the droplet size increases as oil EACN differs from the surfactant PACN. For HLD$_N$ = 0, the affinity of C$_{12}$Gly$_2$ is equivalent for both water and oil at 25°C, and the interfacial curvature is close to 0, causing droplets to destabilize quickly. Also, the kinetics of internal phase release is faster with n-octane than any other investigated oil, regardless of the emulsion WOR. The evolution of stability with EACN is in accordance with HLD$_N$ as stability increases as HLD$_N$ differs from 0 and EACN differs from 8.2.

Both the surfactant concentration and the emulsification process are factors not considered in the HLD$_N$ approach, although very influent on the granulometry and stability of emulsions. Increasing the surfactant concentration from 1% to 3% had little impact on n-octane emulsion granulometry but reduced by more than 10 μm the droplet size for other emulsions due to better droplet interface covering. Stability was also improved with less than 5 vol.% of released internal phase after 14 days, except for HLD$_N$ ~ 0 emulsions. In the same way, emulsions prepared by PIT allowed forming droplets inferior to 2 μm for HLD$_N$ ≠ 0, much more stable over time than emulsions prepared by Ultra-turrax® as droplet flocculation and coalescence is limited by Brownian motion.

General Conclusion

The multiple aims of this thesis were to take up challenges in formulation science, in relation with the design and rationalization of stability in complex, multiphasic systems. Target applications comprise solid dispersions, foams, emulsions and microemulsions. To meet those problematics, applied to nanoparticles (NPs), oils and surfactants, conceptual and predictive tools were developed. This way, relations between the chemical composition, the physicochemical properties and the applicative properties were established: from the raw material to the potential application.

In a first chapter, the complementarity of Hansen Solubility Parameters (HSP) approach and DLVO theory was examined in the case of TiO_2 NPs in aqueous and non-aqueous dispersions. In a second chapter, the hydrophobicity of oils, characterized by the Equivalent Alkane Carbon Number (EACN) value, was modelled and predicted using machine learning techniques, namely neural networks and graph machines. Finally, the amphiphilic behaviour of an innovative surfactant, $C_{12}Gly_2$, was studied in Chapter 3 in aqueous solution for application as foam stabilizer, in relation with its surface rheology. In the last chapter, this same surfactant was investigated in Surfactant / Oil / Water (SOW) systems within the Normalized Hydrophilic-Lipophilic Deviation (HLD_N) framework, indicated for the rationalization and design of emulsions and microemulsions.

The study of TiO_2 nanoparticles presented in Chapter 1 led to the development of a standardized method for the analysis of solid dispersions, based on analytical sedimentation and electrostatic interactions measurements. The analytical methodology, first developed in aqueous media, was extended to non-aqueous solvents in the scope of HSP calculation. Indeed, using HSP for particles would facilitate the solvent choice for particle synthesis and catalysis, but also to design composite materials for instance.

In aqueous media and in methanol, both the *TSI* value and Stokes diameter, obtained by Turbiscan monitoring of dispersions destabilization, were correlated to particles' zeta potential ζ, in accordance with the strength of electrostatic interactions described by the DLVO theory. Extending the study to a set of 15 additional non-aqueous solvents, the stability monitored using that same methodology could not be rationalized by either the DLVO theory, nor the HSP specific solvent / surface interactions individually. Indeed, some dispersions with ζ ~ 0 mV remained stable over time. In this context, the complementarity of both theories was explored. By considering the HSP approach only when DLVO interactions were negligible, i.e. $(V_R + V_A)_{max} < 0$, it was made possible to discriminate the involved stabilizing phenomena for each sample. Both zeta potential measurement and Turbiscan monitoring of stability were

complementary. The Hansen sphere of TiO$_2$ P25 NPs was determined in the $(V_R + V_A)_{max} < 0$ conditions by two quantitative stability descriptors avoiding an empirical visual rating of dispersions stability: $RTSI_{2h}^{top}$ and d$_{Stokes}$, yielding similar results shown in Figure 5.1.

(a)

Fit 0.90   R = 8.8

$\bar{\delta}_d$ = 15.1 ± 0.5 $\bar{\delta}_p$ = 15.5 ± 0.8 $\bar{\delta}_h$ = 14.1 ± 0.5

(b)

Fit 0.88   R = 8.8

$\bar{\delta}_d$ = 15.3 ± 0.5 $\bar{\delta}_p$ = 14.4 ± 0.6 $\bar{\delta}_h$ = 13.8 ± 0.5
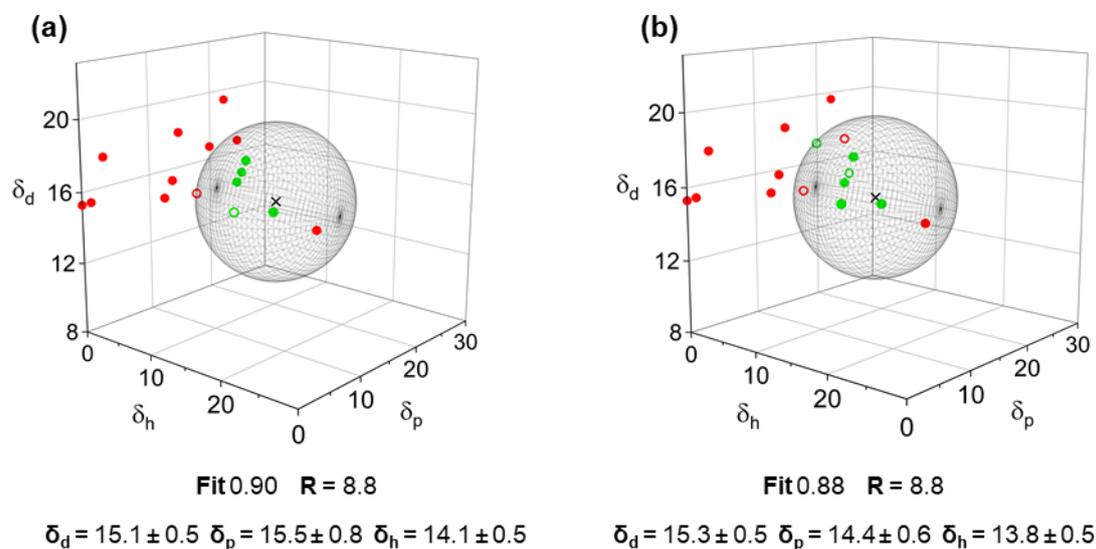
Figure 5.1. Solubility sphere of TiO$_2$ P25 1 g/L in 17 solvents based on (a) Stokes diameters d$_{Stokes}$ and (b) $RTSI_{2h}^{top}$ at 25 °C.

The standardized method developed in this first chapter could be extended to other types of particles and metal oxides. However, as the analytical method based on light scattering requires light diffusing particles, it would thus be ineffective for absorbing ones. Preliminary tests on graphene carried out in the lab tend to confirm this limitation. In that case, other detection methods should be implemented to detect destabilization kinetics. Moreover, the deviation from the spherical shape approximation could result in significant errors for Stokes diameter calculation and zeta potential measurements. This could be an additional challenge for sheet-like and worm-like particles.

A guideline to use HSP for particles efficiently is to evaluate the strength of electrostatic interactions first. By doing so, one can use the solubility sphere as an indication of potential stabilizing solvents, i.e., those comprises within the sphere. Also, coordinates located inside the sphere can be obtained with solvent mixtures as dispersing media.

The problematic of oil hydrophobicity prediction through the modelling of EACN values was addressed in Chapter 2. Such hydrophobicity scale refers to the equivalent length of the linear alkane that would behave similarly to the oil under study in a SOW system. However, the standard experimental determination of EACN is costly and time-consuming.[127] EACN is relevant in applications including emulsions, microemulsions, perfume solubilization or enhanced oil recovery. Predictive tools, i.e. nonlinear regression (Neural Network, NN) based on COSMO-RS σ-moments, and regression on graphs (Graph Machine, GM) derived from the

SMILES codes, were applied to model EACN values from a set of experimentally studied compounds.

A reliable database was constituted and enriched with experimental measurements, in particular regarding branched compounds that were under-represented in literature data. A particular attention must be brought to the database constitution as ambiguous or erroneous values result in lowering the models' predictive ability. By successive iterations of increasing complexity, the best models were found to be NN-6N and GM-5N, with the lowest LOO scores of 0.8 and RMSTE of 0.4 and 0.3 respectively, preventing overfitting to experimental data and ensuring optimal predictive ability. Both these models were used to estimate EACN values for new compounds without experimentation and yielded equivalently reliable modelling and predictive ability. On a test set of 10 molecules, voluntarily chosen as polyfunctional and non-trivial to estimate, the estimation error was of 0.7 and 0.5 EACN units for NN and GM respectively, which is improved compared to previously reported models for EACN estimation based on multiple linear regressions or genetic algorithms.[130,133,134]
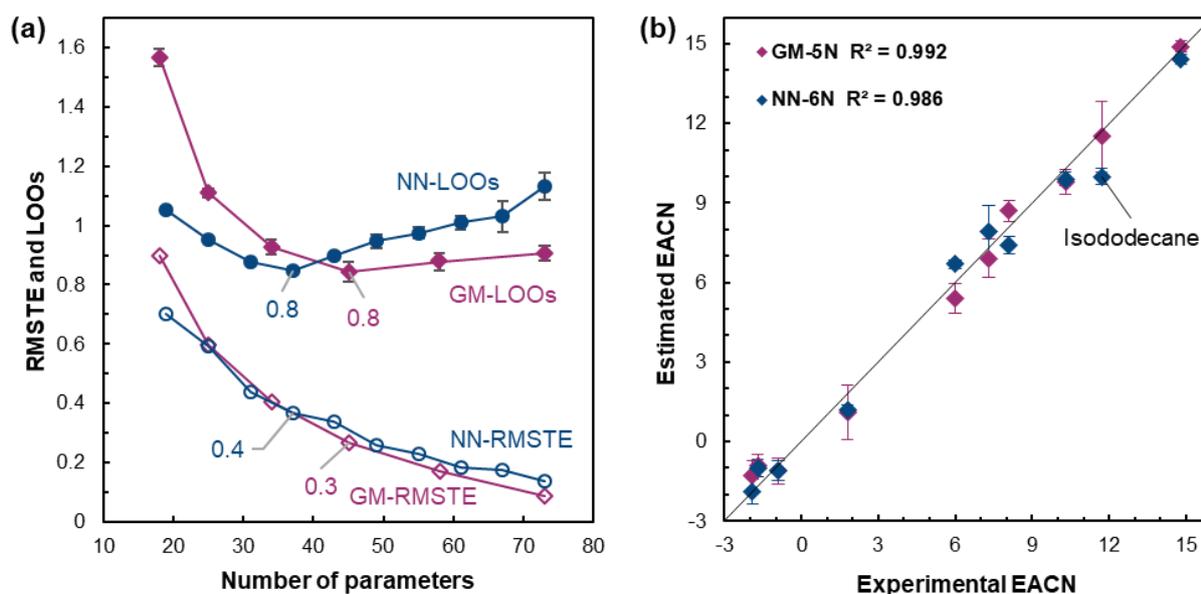
Figure 5.2. (a) RMSTE and LOO scores of NN (blue) and GM (pink) models of increasing complexity for the 111 molecules of the training set and (b) EACN estimations computed by the GM-5N (45 parameters, pink diamonds) and the NN-6N (37 parameters, blue diamonds) vs. experimental EACN values for the 10 molecules of the test set.

Evolution of EACN values in homologous series can be approached by a linear model,[134] which is well estimated by the GM-5N model, but deviations are observed with the NN-6N one. For practical considerations, the GM model is more convenient as it only requires readily available SMILES codes as entry values, whereas the NN model requires the use of COSMO-RS for σ-moments calculations.

The demonstration programs, available in Appendix A8, constitute powerful tools to predict EACN values for any molecule containing C, H, O, N or Cl atoms. However, one should keep

in mind that methods described in Chapter 2 are appropriate for chemical structures close to those represented in the learning dataset. Otherwise, the prediction accuracy could be importantly decreased. Diversification of the chemical structures in the training database could constitute a potential improvement for the extension of these predictive models to a wider range of oils, such as silicones for instance.

As regards to the 1-*O*-dodecyl diglyceryl ether ($C_{12}Gly_2$) surfactant, a detailed evaluation of its behaviour in aqueous solutions and at the air / water and oil / water interfaces was carried out. Chapter 3 focused on aqueous solutions and air / water interface in view of potential use as foam stabilizer, in comparison with two other nonionic surfactants, namely $C_{12}E_5$ and $C_{12}Glu_2$, differing only by the nature of the polar head. Indeed, foams intervene in a variety of end-use products, among which are construction materials, personal care products but also food and beverages.

In contrast with polyethoxylated fatty alcohols and alkyl polyglucosides, $C_{12}Gly_2$ has the ability to form vesicles and lamellar LC in equilibrium with aqueous solutions at concentrations as low as 10 times the CMC. The adsorption isotherm showed an interestingly low minimal surface tension (γ_{min} = 26.8 mN.m^{-1}), and allowed calculating an area per molecule close to that of $C_{12}Glu_2$, and lower than that of $C_{12}E_5$. Then, the rheological response of those surfactants adsorbed layers under surface oscillations revealed a higher value of the high frequency elasticity limit ε_0 at 1 CMC in the case of $C_{12}Gly_2$, indicative of strong intermolecular interactions and attributed to the presence of H-bond-forming -OH groups.
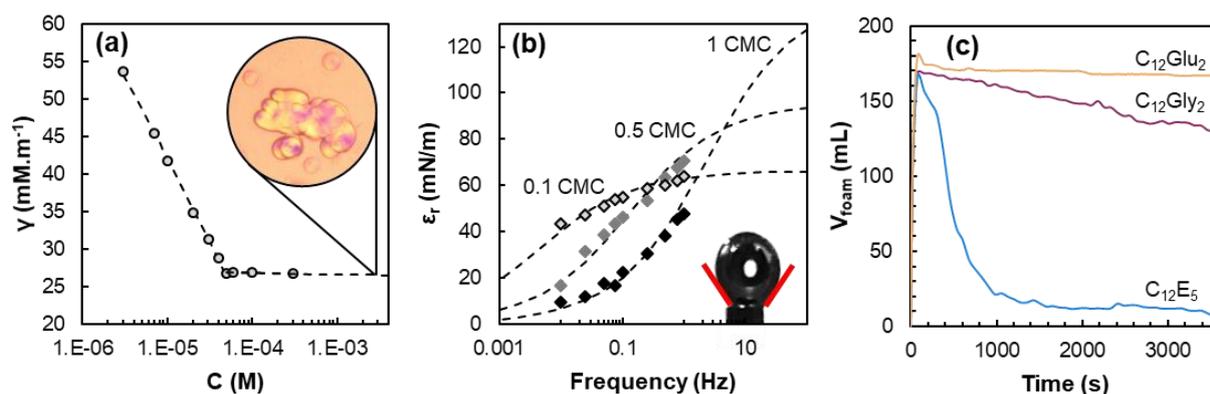
Figure 5.3. (a) Surface tension isotherm of $C_{12}Gly_2$ at 25.0°C and LC phase formed at 3×10^{-3} M observed at the optical microscope with polarized light, (b) dilational surface elasticity $C_{12}Gly_2$ at 0.1 ($\diamond$), 0.5 ($\blacklozenge$) and 1 CMC ($\blacklozenge$), and (c) foam volume evolution for the three surfactants under study.

The high surface elasticity accounts for the good foam stability of $C_{12}Gly_2$ solutions as thermally induced surface deformations are more easily absorbed, preventing film rupture. The contribution of LC to surface elasticity could not be observed due to necessity of measuring viscoelastic properties in diffusion-controlled conditions, i.e., at c ≤ CMC. However, LC

dispersion were shown to contribute to foam stabilization in diglycerol monoesters, the structures of which are very similar to $C_{12}Gly_2$.[205] On the other hand, the low foamability of $C_{12}Gly_2$ was rationalized by its low value of molecular exchange parameter ω_0, indicative of slow exchanges between bulk and surface, in particular during foam generation by air sparging. Concerning foam ageing, the main destabilization phenomenon was identified to be bubble coalescence, which rate was inferior for $C_{12}Gly_2$ compared to $C_{12}Glu_2$ and was not measurable for $C_{12}E_5$ due to quick foam break.

Using $C_{12}Gly_2$ as foaming agent would require overcoming its poor foamability. This could be investigated by varying the bubble generation process, e.g. by reducing the air flow or nucleating gas bubbles from the solution by gas dissolution.[193] Potential association with another surfactant could also be an option to improve foamability while maintaining satisfactory stability.

In the final chapter, $C_{12}Gly_2$ properties are further investigated in SOW systems, in particular regarding the formation of emulsions and microemulsions. The relative amphiphilicity of C_nGlu_m, C_nGly_m and C_iE_j in SOW systems were compared in terms of temperature and salinity sensitivities by the PIT-slope and SPI-slope methods. In a second time, the influence of formulation variables on the stabilization of the oil / water interface stabilized by $C_{12}Gly_2$ was quantified. The full HLD_N equation coefficients then being known for $C_{12}Gly_2$, emulsion characteristics were put in relation with HLD_N values.

Both PIT-slope and SPI-slope classification scales agree well for the studied surfactants. C_nGlu_m appear as the most hydrophilic ones, with highly positive values, due to the important number of -OH on the polar head. C_nGly_m, however, are more moderate amphiphiles. When comparing temperature and salinity classifications, $C_{10}Gly_2$ and $C_{12}Glu$ respond similarly to temperature variations, but the diglyceryl ether has greater salt tolerance with a higher SPI-slope value, indicating this type of surfactant for salt-tolerant applications.

By means of dynamic PIT and SPI experiments, carried out varying the *n*-alkane length, temperature and salinity sensitivities are shown to be reduced by almost 3 times between SOW systems containing $C_{12}Gly_2$ compared to C_iE_j surfactants. The observed differences in sensitivity are attributed to hydroxyl groups acting as both hydrogen-bond donors and acceptors, whereas ethers bonds only act as hydrogen-bond acceptors.

The emulsions formed using $C_{12}Gly_2$ as surfactant for different types of oils showed a minimum in granulometry, corresponding to a maximum in destabilization kinetics. This corresponds to the oil which EACN is the closest to $C_{12}Gly_2$ PACN (8.2), regardless of the oil and water proportions. Interestingly, both W/O and O/W emulsions could be formed, and their relative stability and granulometry follows the same trend on both sides away from $HLD_N = 0$. The

granulometry evolution tendency observed in alkanes (cyclic and linear) is also verified in two more complex oils, namely octyl octanoate and squalane. Minimal interfacial tension between oil and water being attained for EACN = PACN, the formation of smaller droplets (about 1 µm) by mechanical agitation is favoured when using *n*-octane (ACN = 8) or octyl octanoate (EACN = 8.1). As HLD_N deviates from 0, droplet size increases but so does the stability. Indeed, emulsions monitored by light scattering over two weeks showed creaming and sedimentation phenomena, but coalescence and Ostwald ripening get close to null in both O/W and W/O emulsions with higher (or lower) HLD_N values, indicating an efficient stabilization of the interface.
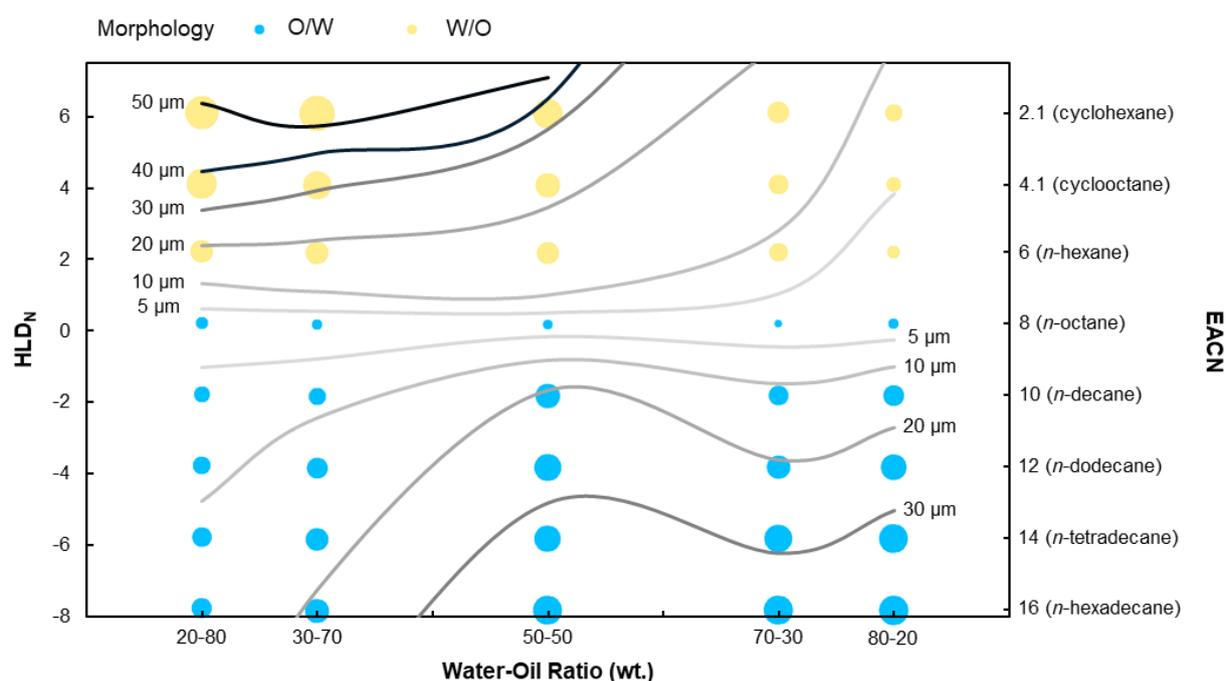
Figure 5.4. Iso-granulometry curves of emulsions containing 1 wt.% $C_{12}Gly_2$ and prepared by varying the nature of the oil (EACN) and the Water-to-Oil Ratio (wt./wt.). O/W and W/O emulsions are represented by blue dots and yellow dots respectively and symbol size is varied with $D_{[4,3]}$ values.

Both the surfactant concentration and the emulsification process are factors not considered in the HLD_N approach, although very influent. Increasing the surfactant concentration from 1% to 3% reduced by more than 10 µm the droplet size due to better droplet interface covering. Stability was also improved with less than 5 vol.% of released internal phase after 14 days, except for HLD_N ~ 0 emulsions. In the same way, emulsions prepared by PIT allowed forming droplets inferior to 2 µm for $HLD_N \neq 0$, much more stable over time than emulsions prepared by Ultra-turrax® as droplet flocculation and coalescence is limited by Brownian motion.

Knowing the HLD_N of a SOW system facilitates the design of emulsions with predictable features. Regarding $C_{12}Gly_2$, further investigations on the potential catastrophic inversions at WOR > 80-20 or WOR < 20-80 could extend the types of systems that could be formed, e.g. multiple emulsions. A rheological study of emulsions would also increase the range of specifications to be put in relation with HLD_N values for practical use.

References

(1) *Non-Ionic Surfactants Market Size, Share | 2022 to 2027*; Market Data Forecast, 2022; p 175.

(2) *Commission recommendation of 18 October 2011 on the definition of nanomaterial text with EEA relevance*; 32011H0696; 2011; Vol. 275.

(3) Monica, R. C.; Cremonini, R. Nanoparticles and higher plants. *Caryologia* **2009**, *62* (2), 161–165.

(4) Whitesides, G. M. Nanoscience, nanotechnology, and chemistry. *Small* **2005**, *1* (2), 172–179.

(5) Schneider, S. L.; Lim, H. W. A review of inorganic UV filters zinc oxide and titanium dioxide. *Photodermatol Photoimmunol Photomed* **2019**, *35* (6), 442–446.

(6) Yashwant Singh, S. *Nano titanium dioxide market by application (paints & coatings, pigments, cosmetics, plastics, energy and others) - Global opportunity analysis and industry forecast, 2014 - 2022*; 2016; p 119.

(7) Papp, J.; Soled, S.; Dwight, K.; Wold, A. Surface acidity and photocatalytic activity of TiO_2, WO_3/TiO_2, and MoO_3/TiO_2 photocatalysts. *Chem. Mater.* **1994**, *6* (4), 496–500.

(8) Pilotek, S.; Gossmann, K.; Tabellion, F.; Steingröver, K.; Homann, H. Tailored nanoscaled TiO_2 dispersions for photocatalytic applications. In *Technical proceedings of the 2009 clean technology conference and trade show*; Houston, USA, 2009; pp 304–307.

(9) Product information AEROXIDE® TiO_2 P25, 2019.

(10) Schulz, J.; Hohenberg, H.; Pflücker, F.; Gärtner, E.; Will, T.; Pfeiffer, S.; Wepf, R.; Wendel, V.; Gers-Barlag, H.; Wittern, K.-P. Distribution of sunscreens on skin. *Adv. Drug Deliv. Rev.* **2002**, *54*, S157–S163.

(11) Li, G.; Lv, L.; Fan, H.; Ma, J.; Li, Y.; Wan, Y.; Zhao, X. S. Effect of the agglomeration of TiO_2 nanoparticles on their photocatalytic performance in the aqueous phase. *J. Colloid Interface Sci.* **2010**, *348* (2), 342–347.

(12) Jiang, J.; Oberdörster, G.; Biswas, P. Characterization of size, surface charge, and agglomeration state of nanoparticle dispersions for toxicological studies. *J Nanopart Res* **2009**, *11* (1), 77–89.

(13) Qin, J.; Wang, X.; Jiang, Q.; Cao, M. Optimizing dispersion, exfoliation, synthesis, and device fabrication of inorganic nanomaterials using Hansen solubility parameters. *ChemPhysChem* **2019**, *20* (9), 1069–1097.

(14) Süß, S.; Sobisch, T.; Peukert, W.; Lerche, D.; Segets, D. Determination of Hansen parameters for particles: a standardized routine based on analytical centrifugation. *Adv Powder Technol* **2018**, *29* (7), 1550–1561.

(15) Buron, H.; Mengual, O.; Meunier, G.; Cayré, I.; Snabre, P. Optical characterization of concentrated dispersions: applications to laboratory analyses and on-line process monitoring and control. *Polym Int* **2004**, *53* (9), 1205–1209.

(16) Woo, S. H.; Min Gu, L.; Rhee, C. K. Sedimentation properties of TiO_2 nanoparticles in organic solvents. *Solid State Phenom.* **2007**, *119*, 267–270.

(17) Liu, Z. Q.; Yang, X.; Zhang, Q. Turbiscan: history, development, application to colloids and dispersions. *Adv Mat Res* **2014**, *936*, 1592–1596.

(18) Luo, M.; Qi, X.; Ren, T.; Huang, Y.; Keller, A. A.; Wang, H.; Wu, B.; Jin, H.; Li, F. Heteroaggregation of CeO_2 and TiO_2 engineered nanoparticles in the aqueous phase: application of turbiscan stability index and fluorescence excitation-emission matrix (EEM) spectra. *Colloids Surf. A* **2017**, *533*, 9–19.

(19) Wieneke, J. U.; Kommoß, B.; Gaer, O.; Prykhodko, I.; Ulbricht, M. Systematic investigation of dispersions of unmodified inorganic nanoparticles in organic solvents with focus on the Hansen solubility parameters. *Ind. Eng. Chem. Res.* **2012**, *51* (1), 327–334.

(20) Küchler, S.; Detloff, T.; Sobisch, T.; Lerche, D. Direct and accelerated characterization of ceramic dispersions. *Ceramic Forum International* **2011**, *cfi/Ber. DKG 88 (2011)*, E27–E31.

(21) Xu, L.; Liang, H.-W.; Yang, Y.; Yu, S.-H. Stability and reactivity: positive and negative aspects for nanoparticle processing. *Chem. Rev.* **2018**, *118* (7), 3209–3250.

(22) Israelachvili, J. N. *Intermolecular and surface forces*, Revised third edition.; Academic Press, 2011.

(23) Hosokawa, M.; Nogi, K.; Naito, M.; Yokoyama, T. *Nanoparticle technology handbook*, Second edition.; Elsevier: Amsterdam, 2012.

(24) Hansen, C. M. *The Three Dimensional Solubility Parameter and Solvent Diffusion Coefficient*; Danish Technical Press: Copenhagen, 1967.

(25) Hansen, C.; Abbott, S.; Yamamoto, H. HSP Examples: Nanoparticles | Hansen Solubility Parameters https://www.hansen-solubility.com/HSP-examples/nanoparticles.php (accessed Apr 9, 2020).

(26) Wheeler, L. M.; Kramer, N. J.; Kortshagen, U. R. Thermodynamic driving force in the spontaneous formation of inorganic nanoparticle solutions. *Nano Lett.* **2018**, *18* (3), 1888–1895.

(27) Stauch, C.; Süß, S.; Luxenhofer, R.; Binks, B. P.; Segets, D.; Mandel, K. Quantifying surface properties of silica particles by combining Hansen parameters and Reichardt's dye indicator data. *Part. Part. Syst. Charact.* **2018**, *35* (12), 1800328.

(28) Launay, H.; Hansen, C. M.; Almdal, K. Hansen solubility parameters for a carbon fiber/epoxy composite. *Carbon* **2007**, *45* (15), 2859–2865.

(29) Hansen, C. M.; Smith, A. L. Using Hansen solubility parameters to correlate solubility of C60 fullerene in organic solvents and in polymers. *Carbon* **2004**, *42* (8), 1591–1597.

(30) Hernandez, Y.; Lotya, M.; Rickard, D.; Bergin, S. D.; Coleman, J. N. Measurement of multicomponent solubility parameters for graphene facilitates solvent discovery. *Langmuir* **2010**, *26* (5), 3208–3213.

(31) Ham, H. T.; Choi, Y. S.; Chung, I. J. An explanation of dispersion states of single-walled carbon nanotubes in solvents and aqueous surfactant solutions using solubility parameters. *J. Colloid Interface Sci.* **2005**, *286* (1), 216–223.

(32) Maleski, K.; Mochalin, V. N.; Gogotsi, Y. Dispersions of two-dimensional titanium carbide MXene in organic solvents. *Chem. Mater.* **2017**, *29* (4), 1632–1640.

(33) Mathioudaki, S.; Barthélémy, B.; Detriche, S.; Vandenabeele, C.; Delhalle, J.; Mekhalif, Z.; Lucas, S. Plasma treatment of metal oxide nanoparticles: development of core–shell structures for a better and similar dispersibility. *ACS Appl. Nano Mater.* **2018**, *1* (7), 3464–3473.

(34) Titanium dioxide https://www.essentialchemicalindustry.org/chemicals/titanium-dioxide.html (accessed Jan 26, 2021).

(35) *Titanium Dioxide Market by Grade, Process, Application, Region - Trends and Forecasts up to 2026*; Market research report CH 5111; Markets and Markets, 2021.

(36) Titanium Dioxide (TiO2) Production and Manufacturing Process https://www.icis.com/explore/resources/news/2007/11/07/9076547/titanium-dioxide-tio2-production-and-manufacturing-process (accessed Jan 26, 2021).

(37) Hanaor, D. A. H.; Sorrell, C. C. Review of the Anatase to Rutile Phase Transformation. *J Mater Sci* **2011**, *46* (4), 855–874.

(38) *CRC Handbook of Chemistry and Physics, 84th Edition*, 84th edition.; Lide, D. R., Ed.; CRC Press: Boca Raton, 2003.

(39) Titanium Dioxide for Coatings - Product Overview C-10416-2 (2/19), 2019.

(40) Eremin, Y. A. Scattering Theory. In *Encyclopedia of Modern Optics*; Guenther, R. D., Ed.; Elsevier Academic Press: Oxford, 2005; pp 326–330.

(41) Li, H.; Hu, Z.; Zhang, S.; Gu, X.; Wang, H.; Jiang, P.; Zhao, Q. Effects of Titanium Dioxide on the Flammability and Char Formation of Water-Based Coatings Containing Intumescent Flame Retardants. *Progress in Organic Coatings* **2015**, *78*, 318–324.

(42) Duquesne, S.; Bachelet, P.; Bellayer, S.; Bourbigot, S.; Mertens, W. Influence of Inorganic Fillers on the Fire Protection of Intumescent Coatings. *Journal of Fire Sciences* **2013**, *31* (3), 258–275.

(43) Carp, O.; Huisman, C. L.; Reller, A. Photoinduced Reactivity of Titanium Dioxide. *Prog. Solid State Chem.* **2004**, *32* (1), 33–177.

(44) Cassar, L. Photocatalysis of Cementitious Materials: Clean Buildings and Clean Air. *MRS Bulletin* **2004**, *29* (5), 328–331.

(45) Chen, J.; Poon, C. Photocatalytic Construction and Building Materials: From Fundamentals to Applications. *Build. Environ.* **2009**, *44* (9), 1899–1906.

(46) Guo, S.; Wu, Z.; Zhao, W. TiO$_2$-Based Building Materials: Above and beyond Traditional Applications. *Chin. Sci. Bull.* **2009**, *54* (7), 1137–1142.

(47) Linkous, C., A. Photocatalytic Surfacing Agents for Inhibiting Algae Growth. US005518992A, May 21, 1996.

(48) Murata, Y.; Tawara, H.; Obata, H.; Murata, K. NOx-Cleaning Paving Block. EP0786283B1, January 2, 2003.

(49) Wang, R.; Hashimoto, K.; Fujishima, A.; Chikuni, M.; Kojima, E.; Kitamura, A.; Shimohigoshi, M.; Watanabe, T. Light-Induced Amphiphilic Surfaces. *Nature* **1997**, *388* (6641), 431–432.

(50) Wang, R.; Hashimoto, K.; Fujishima, A.; Chikuni, M.; Kojima, E.; Kitamura, A.; Shimohigoshi, M.; Watanabe, T. Photogeneration of Highly Amphiphilic TiO$_2$ Surfaces. *Adv. Mater.* **1998**, *10* (2), 135–138.

(51) Janczarek, M.; Klapiszewski, Ł.; Jędrzejczak, P.; Klapiszewska, I.; Ślosarczyk, A.; Jesionowski, T. Progress of Functionalized TiO2-Based Nanomaterials in the Construction Industry: A Comprehensive Review. *Chemical Engineering Journal* **2022**, *430*, 132062.

(52) Ashikaga, T.; Wada, M.; Kobayashi, H.; Mori, M.; Katsumura, Y.; Fukui, H.; Kato, S.; Yamaguchi, M.; Takamatsu, T. Effect of the Photocatalytic Activity of TiO2 on Plasmid DNA1This Article Was Communicated by the Mammalian Mutagenicity Study Group (MMS) of the Japanese Environmental Mutagen Society.1. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis* **2000**, *466* (1), 1–7.

(53) Carbon Black, Titanium Dioxide, and Talc. *IARC Monogr Eval Carcinog Risks Hum* **2010**, *93*, 1–413.

(54) Grande, F.; Tucci, P. Titanium Dioxide Nanoparticles: A Risk for Human Health? *Mini Rev Med Chem* **2016**, *16* (9), 762–769.

(55) Weir, A.; Westerhoff, P.; Fabricius, L.; Hristovski, K.; von Goetz, N. Titanium Dioxide Nanoparticles in Food and Personal Care Products. *Environ. Sci. Technol.* **2012**, *46* (4), 2242–2250.

(56) *Commission Regulation (EU) 2019/1857 of 6 November 2019 Amending Annex VI to Regulation (EC) No 1223/2009 of the European Parliament and of the Council on Cosmetic Products (Text with EEA Relevance)*; 2019; Vol. 286.

(57) Bagheri, S.; Muhd Julkapli, N.; Bee Abd Hamid, S. Titanium Dioxide as a Catalyst Support in Heterogeneous Catalysis. *The Scientific World Journal* **2014**, *2014*, e727496.

(58) Liang, G.; He, L.; Cheng, H.; Li, W.; Li, X.; Zhang, C.; Yu, Y.; Zhao, F. The Hydrogenation/Dehydrogenation Activity of Supported Ni Catalysts and Their Effect on Hexitols Selectivity in Hydrolytic Hydrogenation of Cellulose. *Journal of Catalysis* **2014**, *309*, 468–476.

(59) Luo, Q.; Beller, M.; Jiao, H. Formic Acid Dehydrogenation on Surfaces — a Review of Computational Aspect. *J. Theor. Comput. Chem.* **2013**, *12* (07), 1330001.

(60) Palcheva, R.; Dimitrov, L.; Tyuliev, G.; Spojakina, A.; Jiratova, K. TiO$_2$ Nanotubes Supported NiW Hydrodesulphurization Catalysts: Characterization and Activity. *Applied Surface Science* **2013**, *265*, 309–316.

(61) Ammal, S. C.; Heyden, A. Origin of the Unique Activity of Pt/TiO2 Catalysts for the Water–Gas Shift Reaction. *Journal of Catalysis* **2013**, *306*, 78–90.

(62) Hinojosa-Reyes, M.; Rodríguez-González, V.; Zanella, R. Gold Nanoparticles Supported on TiO2–Ni as Catalysts for Hydrogen Purification via Water–Gas Shift Reaction. *RSC Adv.* **2013**, *4* (9), 4308–4316.

(63) Kominami, H.; Kato, J.; Takada, Y.; Doushi, Y.; Ohtani, B.; Nishimoto, S.; Inoue, M.; Inui, T.; Kera, Y. Novel Synthesis of Microcrystalline Titanium(IV) Oxide Having High Thermal Stability and Ultra-High Photocatalytic Activity: Thermal Decomposition of Titanium(IV) Alkoxide in Organic Solvents. *Catalysis Letters* **1997**, *46* (3), 235–240.

(64) Smirniotis, P. G.; Sreekanth, P. M.; Peña, D. A.; Jenkins, R. G. Manganese Oxide Catalysts Supported on TiO$_2$, Al$_2$O$_3$, and SiO$_2$: A Comparison for Low-Temperature SCR of NO with NH$_3$. *Ind. Eng. Chem. Res.* **2006**, *45* (19), 6436–6443.

(65) Hua, H.; Hu, C.; Zhao, Z.; Liu, H.; Xie, X.; Xi, Y. Pt Nanoparticles Supported on Submicrometer-Sized TiO2 Spheres for Effective Methanol and Ethanol Oxidation. *Electrochimica Acta* **2013**, *105*, 130–136.

(66) Yu, J.; Dai, G.; Xiang, Q.; Jaroniec, M. Fabrication and Enhanced Visible-Light Photocatalytic Activity of Carbon Self-Doped TiO2 Sheets with Exposed {001} Facets. *J. Mater. Chem.* **2011**, *21* (4), 1049–1057.

(67) Liu, G.; Yang, H. G.; Wang, X.; Cheng, L.; Pan, J.; Lu, G. Q. (Max); Cheng, H.-M. Visible Light Responsive Nitrogen Doped Anatase TiO2 Sheets with Dominant {001} Facets Derived from TiN. *J. Am. Chem. Soc.* **2009**, *131* (36), 12868–12869.

(68) Nolan, M. Modifying Ceria (111) with a TiO$_2$ Nanocluster for Enhanced Reactivity. *J. Chem. Phys.* **2013**, *139* (18), 184710.

(69) Padikkaparambil, S.; Narayanan, B.; Yaakob, Z.; Viswanathan, S.; Tasirin, S. M. Au/TiO$_2$ Reusable Photocatalysts for Dye Degradation. *Int. J. Photoenergy* **2013**, *2013*, e752605.

(70) Grunwaldt, J.-D.; Kiener, C.; Wögerbauer, C.; Baiker, A. Preparation of Supported Gold Catalysts for Low-Temperature CO Oxidation via "Size-Controlled" Gold Colloids. *J. Catal.* **1999**, *181* (2), 223–232.

(71) Amores, J. M. G.; Escribano, V. S.; Busca, G. Anatase Crystal Growth and Phase Transformation to Rutile in High-Area TiO2, MoO3–TiO2 and Other TiO2-Supported Oxide Catalytic Systems. *J. Mater. Chem.* **1995**, *5* (8), 1245–1249.

(72) Ajaikumar, S.; Ahlkvist, J.; Larsson, W.; Shchukarev, A.; Leino, A.-R.; Kordas, K.; Mikkola, J.-P. Oxidation of α-Pinene over Gold Containing Bimetallic Nanoparticles Supported on Reducible TiO2 by Deposition-Precipitation Method. *Applied Catalysis A: General* **2011**, *392* (1), 11–18.

(73) Bowker, M.; Morton, C.; Kennedy, J.; Bahruji, H.; Greves, J.; Jones, W.; Davies, P.; Wells, P.; Dimitratos, N. Hydrogen Production by Photoreforming of Biofuels Using Au, Pd and Au-Pd/TiO$_2$ Photocatalysts. *J. Catal.* **2014**, *310*, 10–15.

(74) Chen, Y.; Li, D.; Wang, X.; Wu, L.; Wang, X.; Fu, X. Promoting Effects of H2 on Photooxidation of Volatile Organic Pollutants over Pt/TiO2. *New J. Chem.* **2005**, *29* (12), 1514–1519.

(75) Brown, R. A Brief Account of Microscopical Observations Made in the Months of June, July and August, 1827, on the Particles Contained in the Pollen of Plants; and on the General Existence of Active Molecules in Organic and Inorganic Bodies. *Philosophical Magazine* **1828**, *4*, 161–173.

(76) Feynman, R. P.; Leighton, R. B.; Sands, M. L. The Brownian Movement. In *The Feynman lectures on physics*; Addison-Wesley Pub. Co.: Reading, Mass., 1963; Vol. 1.

(77) Derjaguin, B.; Landau, L. Theory of the stability of strongly charged lyophobic sols and of the adhesion of strongly charged particles in solutions of electrolytes. *Acta Physicochim. U.R.S.S.* **1941**, *14*, 633–662.

(78) Verwey, E. J. W.; Overbeek, J. T. G.; Nes, K. van. *Theory of the stability of lyophobic colloids: the interaction of sol particles having an electric double layer*; Elsevier Publishing Company: Amsterdam, 1948.

(79)   Behrens, S. H.; Christl, D. I.; Emmerzael, R.; Schurtenberger, P.; Borkovec, M. Charging and aggregation properties of carboxyl latex particles: experiments versus DLVO theory. *Langmuir* **2000**, *16* (6), 2566–2575.

(80)   Hückel, E.; Debye, P. Zur Theorie der Elektrolyte. I. Gefrierpunktserniedrigung und verwandte Erscheinungen. *Phys. Z* **1923**, *24*, 185–206.

(81)   Hamaker, H. C. The London-van der Waals attraction between spherical particles. *Physica* **1937**, *4* (10), 1058–1072.

(82)   Hogg, R.; Healy, T. W.; Fuerstenau, D. W. Mutual Coagulation of Colloidal Dispersions. *Trans. Faraday Soc.* **1966**, *62* (0), 1638–1651.

(83)   Tadros, T. Chapter 2 - Colloid and interface aspects of pharmaceutical science. In *Colloid and interface science in pharmaceutical research and development*; Ohshima, H., Makino, K., Eds.; Elsevier: Amsterdam, 2014; pp 29–54.

(84)   Vakarelski, I. U.; Higashitani, K. Dynamic Features of Short-Range Interaction Force and Adhesion in Solutions. *J. Colloid Interface Sci.* **2001**, *242* (1), 110–120.

(85)   Vakarelski, I. U.; Ishimura, K.; Higashitani, K. Adhesion between Silica Particle and Mica Surfaces in Water and Electrolyte Solutions. *J. Colloid Interface Sci.* **2000**, *227* (1), 111–118.

(86)   Ishida, N.; Inoue, T.; Miyahara, M.; Higashitani, K. Nano Bubbles on a Hydrophobic Surface in Water Observed by Tapping-Mode Atomic Force Microscopy. *Langmuir* **2000**, *16* (16), 6377–6380.

(87)   Ishida, N.; Sakamoto, M.; Miyahara, M.; Higashitani, K. Attraction between Hydrophobic Surfaces with and without Gas Phase. *Langmuir* **2000**, *16* (13), 5681–5687.

(88)   Gutmann, V. *The Donor-Acceptor Approach to Molecular Interactions*; Plenum Press: New York, 1978.

(89)   Taft, R. W.; Pienta, N. J.; Kamlet, M. J.; Arnett, E. M. Linear Solvation Energy Relationships. 7. Correlations between the Solvent-Donicity and Acceptor-Number Scales and the Solvatochromic Parameters Π*, α, and β. *J. Org. Chem.* **1981**, *46* (4), 661–667.

(90)   Shaw, D. J. *Introduction to Colloid and Surface Chemistry*, 4th ed.; Butterworth-Heinemann: Oxford ; Boston, 1992.

(91)   Light Scattering. In *Particle Characterization: Light Scattering Methods*; Xu, R., Scarlett, B., Eds.; Particle Technology Series; Kluwer Academic Publishers: New York, 2002; pp 56–110.

(92)   Dynamic Light Scattering DLS | Malvern Panalytical https://www.malvernpanalytical.com/en/products/technology/light-scattering/dynamic-light-scattering (accessed Feb 24, 2022).

(93)   Mengual, O.; Meunier, G.; Cayré, I.; Puech, K.; Snabre, P. TURBISCAN MA 2000: Multiple Light Scattering Measurement for Concentrated Emulsion and Suspension Instability Analysis. *Talanta* **1999**, *50* (2), 445–456.

(94)   Detloff, T.; Sobisch, T.; Lerche, D. Particle Size Distribution by Space or Time Dependent Extinction Profiles Obtained by Analytical Centrifugation. *Part. Part. Syst. Charact.* **2006**, *23* (2), 184–187.

(95)   AEROXIDE®, AERODISP® and AEROPERL® titanium dioxide as photocatalyst - Technical information 1243, 2015.

(96)   Technical overview 13 - AEROXIDE® - Fumed metal oxides, 2017.

(97)   Application note TS_STAB_60 Explanation to the TSI calculation, 2019.

(98)   Mills, P.; Snabre, P. Settling of a suspension of hard spheres. *EPL* **1994**, *25* (9), 651.

(99)   Application note - Zeta potential measurements of non-aqueous particulate suspensions, 2010.

(100)   Tucker, I. M.; Corbett, J. C. W.; Fatkin, J.; Jack, R. O.; Kaszuba, M.; MacCreath, B.; McNeil-Watson, F. Laser Doppler Electrophoresis applied to colloids and surfaces. *Curr Opin Colloid Interface Sci* **2015**, *20* (4), 215–226.

(101)   Morrison, I. D.; Ross, S. *Colloidal dispersions: suspensions, emulsions, and foams*; Wiley, 2002.

(102)   Hunter, R. J. *Zeta potential in colloid science: principles and applications*; Academic Press: New York, 1981.

(103)   Kosmulski, M. Zeta potentials in nonaqueous media: how to measure and control them. *Colloids Surf. A* **1999**, *159* (2), 277–281.

(104)   Kosmulski, M.; Eriksson, P.; Rosenholm, J. B. Application of zetametry to determine concentrations of acidic and basic impurities in analytical reagents. *Anal. Chem.* **1999**, *71* (13), 2518–2522.

(105)   Kosmulski, M. *Surface charging and points of zero charge*; CRC Press, 2009.

(106)   Bourikas, K.; Hiemstra, T.; Van Riemsdijk, W. H. Ion pair formation and primary charging behaviour of titanium oxide (anatase and rutile). *Langmuir* **2001**, *17* (3), 749–756.

(107)   Foissy, A.; M'Pandou, A.; Lamarche, J. M.; Jaffrezic-Renault, N. Surface and diffuse-layer charge at the TiO_2-electrolyte interface. *Colloids Surf.* **1982**, *5* (4), 363–368.

(108)   Suttiponparnit, K.; Jiang, J.; Sahu, M.; Suvachittanont, S.; Charinpanitkul, T.; Biswas, P. Role of surface area, primary particle size, and crystal phase on titanium dioxide nanoparticle dispersion properties. *Nanoscale Res Lett* **2010**, *6* (1), 27.

(109)   Qi, J.; Ye, Y. Y.; Wu, J. J.; Wang, H. T.; Li, F. T. Dispersion and stability of titanium dioxide nanoparticles in aqueous suspension: effects of ultrasonication and concentration. *Water Science and Technology* **2013**, *67* (1), 147–151.

(110)   Beranek, R. (Photo)electrochemical methods for the determination of the band edge positions of TiO_2-based nanomaterials. *Adv. Phys. Chem.* **2011**.

(111)   Herrmann, M.; Boehm, H. P. Über die Chemie der Oberfläche des Titandioxids. ii. Saure Hydroxylgruppen auf der Oberfläche. *Z Anorg Allg Chem* **1969**, *368* (1–2), 73–86.

(112)   Butler, M. A.; Ginley, D. S. Prediction of flatband potentials at semiconductor-electrolyte interfaces from atomic electronegativities. *J. Electrochem. Soc.* **1978**, *125* (2), 228.

(113)   Kormann, C.; Bahnemann, D. W.; Hoffmann, M. R. Photolysis of chloroform and other organic molecules in aqueous titanium dioxide suspensions. *Environ. Sci. Technol.* **1991**, *25* (3), 494–500.

(114)   Lyklema, J. Principles of the stability of lyophobic colloidal dispersions in non-aqueous media. *Adv. Colloid Interface Sci.* **1968**, *2* (2), 67–114.

(115)   Rosenholm, J. B.; Dahlsten, P. Influence of charge exchange in acidic aqueous and alcoholic titania dispersions on viscosity. *Advances in Colloid and Interface Science* **2015**, *226*, 138–165.

(116)   Burrows, N. D.; Kesselman, E.; Sabyrov, K.; Stemig, A.; Talmon, Y.; Lee Penn, R. Crystalline nanoparticle aggregation in non-aqueous solvents. *CrystEngComm* **2014**, *16* (8), 1472–1481.

(117)   Izutsu, K. Acid–base reactions in nonaqueous solvents. In *Electrochemistry in nonaqueous solutions*; John Wiley & Sons, Ltd, 2009; pp 63–87.

(118)   Barbosa, J.; Bosch, C. M.; Sanz-Nebot, V. Effect of the Solvent on the Equilibria of Acid-Base Indicators in Aprotic and Amphiprotic Solvents. *Mikrochim. Acta* **1992**, *106* (3), 327–337.

(119)   Popovych, O. Transfer Activity Coefficients of Ions in Methanol-Water Solvents Based on the Tetraphenylborate Assumption. *J. Phys. Chem.* **1984**, *88* (18), 4167–4170.

(120)   Kosmulski, M.; Matijević, E. Zeta potential of anatase (TiO_2) in mixed solvents. *Colloids Surf.* **1992**, *64* (1), 57–65.

(121)   Janusz, W.; Sworska, A.; Szczypa, J. The structure of the electrical double layer at the titanium dioxide/ethanol solutions interface. *Colloids Surf. A* **1999**, *152* (3), 223–233.

(122)   Leo, A.; Hansch, C.; Elkins, D. Partition Coefficients and Their Uses. *Chem. Rev.* **1971**, *71* (6), 525–616.

(123)   Griffin, W. C. Classification of surface-active agents by "HLB." *J. Soc. Cosmet. Chem.* **1949**, *1* (5), 311–326.

(124)   Griffin, W. C. Calculation of HLB values of non-ionic surfactants. **1954**, *5* (4), 249–256.

(125) Wade, W. H.; Morgan, J. C.; Jacobson, J. K.; Schechter, R. S. Low Interfacial Tensions Involving Mixtures of Surfactants. *Soc. Pet. Eng. J.* **1977**, *17* (02), 122–128.

(126) Cash, L.; Cayias, J. L.; Fournier, G.; Macallister, D.; Schares, T.; Schechter, R. S.; Wade, W. H. The Application of Low Interfacial Tension Scaling Rules to Binary Hydrocarbon Mixtures. *Journal of Colloid and Interface Science* **1977**, *59* (1), 39–44.

(127) Queste, S.; Salager, J. L.; Strey, R.; Aubry, J. M. The EACN Scale for Oil Classification Revisited Thanks to Fish Diagrams. *J. Colloid Interface Sci.* **2007**, *312* (1), 98–107.

(128) Aubry, J.-M.; Ontiveros, J. F.; Salager, J.-L.; Nardello-Rataj, V. Use of the Normalized Hydrophilic-Lipophilic-Deviation (HLD_N) Equation for Determining the Equivalent Alkane Carbon Number (EACN) of Oils and the Preferred Alkane Carbon Number (PACN) of Nonionic Surfactants by the Fish-Tail Method (FTM). *Adv. Colloid Interface Sci.* **2020**, *276*, 102099.

(129) Creton, B.; Lévêque, I.; Oukhemanou, F. Equivalent Alkane Carbon Number of Crude Oils: A Predictive Model Based on Machine Learning. *Oil Gas Sci. Technol. – Rev. IFP Energies nouvelles* **2019**, *74*, 30.

(130) Bouton, F.; Durand, M.; Nardello-Rataj, V.; Borosy, A. P.; Quellet, C.; Aubry, J.-M. A QSPR Model for the Prediction of the "Fish-Tail" Temperature of C_iE_4/Water/Polar Hydrocarbon Oil Systems. *Langmuir* **2010**, *26* (11), 7962–7970.

(131) Klamt, A.; Eckert, F.; Hornig, M. COSMO-RS: A Novel View to Physiological Solvation and Partition Questions. *J. Comput. Aided Mol. Des.* **2001**.

(132) Klamt, A. *COSMO-RS: from quantum chemistry to fluid phase thermodynamics and drug design*, 1st edition.; Elsevier: Amsterdam, 2005.

(133) Lukowicz, T.; Benazzouz, A.; Nardello-Rataj, V.; Aubry, J.-M. Rationalization and Prediction of the Equivalent Alkane Carbon Number (EACN) of Polar Hydrocarbon Oils with COSMO-RS σ-Moments. *Langmuir* **2015**, *31* (41), 11220–11226.

(134) Lukowicz, T.; Illous, E.; Nardello-Rataj, V.; Aubry, J.-M. Prediction of the equivalent alkane carbon number (EACN) of aprotic polar oils with COSMO-RS sigma-moments. *Colloids Surf. A Physicochem. Eng. Asp.* **2018**, *536*, 53–59.

(135) Goussard, V.; Duprat, F.; Gerbaud, V.; Ploix, J.-L.; Dreyfus, G.; Nardello-Rataj, V.; Aubry, J.-M. Predicting the Surface Tension of Liquids: Comparison of Four Modeling Approaches and Application to Cosmetic Oils. *J. Chem. Inf. Model.* **2017**, *57* (12), 2986–2995.

(136) Goussard, V.; Duprat, F.; Ploix, J.-L.; Dreyfus, G.; Nardello-Rataj, V.; Aubry, J.-M. A New Machine-Learning Tool for Fast Estimation of Liquid Viscosity. Application to Cosmetic Oils. *J. Chem. Inf. Model.* **2020**, *60* (4), 2012–2023.

(137) Saldana, D. A.; Starck, L.; Mougin, P.; Rousseau, B.; Pidol, L.; Jeuland, N.; Creton, B. Flash Point and Cetane Number Predictions for Fuel Compounds Using Quantitative Structure Property Relationship (QSPR) Methods. *Energy Fuels* **2011**, *25* (9), 3900–3908.

(138) Goulon, A.; Picot, T.; Duprat, A.; Dreyfus, G. Predicting Activities without Computing Descriptors: Graph Machines for QSAR. *SAR QSAR Environ. Res.* **2007**, *18* (1–2), 141–153.

(139) Goulon, A.; Duprat, A.; Dreyfus, G. Graph Machines and Their Applications to Computer-Aided Drug Design: A New Approach to Learning from Structured Data. In *Unconventional Computation*; Calude, C. S., Dinneen, M. J., Păun, G., Rozenberg, G., Stepney, S., Eds.; Lecture Notes in Computer Science; Springer: Berlin, Heidelberg, 2006; Vol. 4135, pp 1–19.

(140) Porcheron, F.; Jacquin, M.; Hadri, N. E.; Saldana, D. A.; Goulon, A.; Faraj, A. Graph Machine Based-QSAR Approach for Modeling Thermodynamic Properties of Amines: Application to CO2 Capture in Postcombustion. *Oil Gas Sci. Technol. – Rev. IFP Energies nouvelles* **2013**, *68* (3), 469–486.

(141) Dioury, F.; Duprat, A.; Dreyfus, G.; Ferroud, C.; Cossy, J. QSPR Prediction of the Stability Constants of Gadolinium(III) Complexes for Magnetic Resonance Imaging. *J. Chem. Inf. Model.* **2014**, *54* (10), 2718–2731.

(142) Goulon, A.; Faraj, A.; Pirngruber, G.; Jacquin, M.; Porcheron, F.; Leflaive, P.; Martin, P.; Baron, G. V.; Denayer, J. F. M. Novel Graph Machine Based QSAR Approach for the Prediction of the Adsorption Enthalpies of Alkanes on Zeolites. *Catal. Today* **2011**, *159* (1), 74–83.

(143) Duprat, A. F.; Huynh, T.; Dreyfus, G. Toward a Principled Methodology for Neural Network Design and Performance Evaluation in QSAR. Application to the Prediction of LogP. *J. Chem. Inf. Comput. Sci.* **1998**, *38* (4), 586–594.

(144) Goulon-Sigwalt-Abram, A.; Duprat, A.; Dreyfus, G. From Hopfield Nets to Recursive Networks to Graph Machines: Numerical Machine Learning for Structured Data. *Theor. Comput. Sci.* **2005**, *344* (2), 298–334.

(145) Lukowicz, T. Synergistic Solubilisation of Fragrances in Binary Surfactant Systems and Prediction of Their EACN Value with COSMO-RS. PhD thesis, Lille 1, 2015.

(146) Bouton, F.; Durand, M.; Nardello-Rataj, V.; Serry, M.; Aubry, J.-M. Classification of Terpene Oils Using the Fish Diagrams and the Equivalent Alkane Carbon (EACN) Scale. *Colloids Surf. A Physicochem. Eng. Asp.* **2009**, *338* (1), 142–147.

(147) Ontiveros, J. F.; Pierlot, C.; Catté, M.; Molinier, V.; Pizzino, A.; Salager, J.-L.; Aubry, J.-M. Classification of Ester Oils According to Their Equivalent Alkane Carbon Number (EACN) and Asymmetry of Fish Diagrams of $C_{10}E_4$/Ester Oil/Water Systems. *J. Colloid Interface Sci.* **2013**, *403*, 67–76.

(148) Silverstein, T. P. The Real Reason Why Oil and Water Don't Mix. *J. Chem. Educ.* **1998**, *75* (1), 116.

(149) Prausnitz, J. M.; Lichtenthaler, R. N.; Azevedo, E. G. de. *Molecular Thermodynamics of Fluid-Phase Equilibria*; Pearson Education, 1998.

(150) Winsor, P. A. Hydrotropy, Solubilisation and Related Emulsification Processes. *Trans. Faraday Soc.* **1948**, *44* (0), 376–398.

(151) Robbers, J. E.; Bhatia, V. N. Technique for the Rapid Determination of HLB and Required-HLB Values. *J. Pharm. Sci.* **1961**, *50* (8), 708–709.

(152) Zajic, J. E.; Seffens, W. Approaches to Hydrophilic-Lipophilic Balance Evaluation of Microbes. In *Biotechnology for the Oils and Fats Industry*; Ratledge, C., Dawson, P. S. S., Rattray, J., Eds.; The American Oil Chemists Society: Urbana, IL, 1984.

(153) Orafidiya, L. O.; Oladimeji, F. A. Determination of the Required HLB Values of Some Essential Oils. *Int. J. Pharm.* **2002**, *237* (1), 241–249.

(154) Dos Santos, O. D. H.; Miotto, J. V.; de Morais, J. M.; da Rocha-Filho, P. A.; de Oliveira, W. P. Attainment of Emulsions with Liquid Crystal from Marigold Oil Using the Required HLB Method. *J. Dispers. Sci. Technol.* **2005**, *26* (2), 243–249.

(155) Moreira de Morais, J.; David Henrique dos Santos, O.; Delicato, T.; Azzini Gonçalves, R.; Alves da Rocha-Filho, P. Physicochemical Characterization of Canola Oil/Water Nano-emulsions Obtained by Determination of Required HLB Number and Emulsion Phase Inversion Methods. *J. Dispers. Sci. Technol.* **2006**, *27* (1), 109–115.

(156) Meher, J. G.; Yadav, N. P.; Sahu, J. J.; Sinha, P. Determination of Required Hydrophilic-Lipophilic Balance of Citronella Oil and Development of Stable Cream Formulation. *Drug. Dev. Ind. Pharm.* **2013**, *39* (10), 1540–1546.

(157) Lee, Y.-Y.; Yoon, K.-S. Determination of Required HLB Values for Citrus Unshiu Fruit Oil, Citrus Unshiu Peel Oil, Horse Fat and Camellia Japonica Seed Oil. *J. Cosmet. Sci.* **2020**, *71* (6), 411–424.

(158) Salager, J. L.; Morgan, J. C.; Schechter, R. S.; Wade, W. H.; Vasquez, E. Optimum Formulation of Surfactant/Water/Oil Systems for Minimum Interfacial Tension or Phase Behaviour. *SPE J.* **1979**, *19* (02), 107–115.

(159) Kanei, N.; Tamura, Y.; Kunieda, H. Effect of Types of Perfume Compounds on the Hydrophile–Lipophile Balance Temperature. *J. Colloid Interface Sci.* **1999**, *218* (1), 13–22.

(160) Tchakalova, V.; Fieber, W. Classification of Fragrances and Fragrance Mixtures Based on Interfacial Solubilization. *J. Surfact. Deterg.* **2012**, *15* (2), 167–177.

(161) Hyde, S.; Blum, Z.; Landh, T.; Lidin, S.; Ninham, B. W.; Andersson, S.; Larsson, K. *The Language of Shape: The Role of Curvature in Condensed Matter: Physics, Chemistry and Biology*; Elsevier, 1996.

(162) Graciaa, A.; Lachaise, J.; Cucuphat, C.; Bourrel, M.; Salager, J. L. Interfacial Segregation of an Ethyl Oleate/Hexadecane Oil Mixture in Microemulsion Systems. *Langmuir* **1993**, *9* (6), 1473–1478.

(163) Pizzino, A.; Molinier, V.; Catté, M.; Ontiveros, J. F.; Salager, J.-L.; Aubry, J.-M. Relationship between phase behaviour and emulsion inversion for a well-defined surfactant ($C_{10}E_4$)/n-octane/water ternary system at different temperatures and water/oil ratios. *Ind. Eng. Chem. Res.* **2013**, *52* (12), 4527–4538.

(164) Wormuth, K. R.; Kaler, E. W. Microemulsifying Polar Oils. *J. Phys. Chem.* **1989**, *93* (12), 4855–4861.

(165) Brown, A. C.; Fraser, T. R. On the connection between chemical constitution and physiological action; with special reference to the physiological action of the salts of the ammonium bases derived from strychnia, brucia, thebaia, codeia, morphia, and nicotia. *J Anat Physiol* **1868**, *2* (2), 224–242.

(166) Hansch, Corwin.; Fujita, Toshio. ρ-σ-π Analysis. A method for the correlation of biological activity and chemical structure. *J. Am. Chem. Soc.* **1964**, *86* (8), 1616–1626.

(167) Free, S. M.; Wilson, J. W. A mathematical contribution to structure-activity studies. *J. Med. Chem.* **1964**, *7* (4), 395–399.

(168) *Guidance document on the validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] models*; ENV/JM/MONO(2007)2; Organisation for Economic Co-operation and Development (OECD): Brussels, 2007.

(169) Nieto-Draghi, C.; Fayet, G.; Creton, B.; Rozanska, X.; Rotureau, P.; de Hemptinne, J.-C.; Ungerer, P.; Rousseau, B.; Adamo, C. A General Guidebook for the Theoretical Prediction of Physicochemical Properties of Chemicals for Regulatory Purposes. *Chem. Rev.* **2015**, *115* (24), 13093–13164.

(170) Goulon-Sigwalt-Abram, A. Une nouvelle méthode d'apprentissage de données structurées : applications à l'aide à la découverte de médicaments. PhD thesis, Université Pierre et Marie Curie - Paris VI, 2008.

(171) Dupeux, T.; Gaudin, T.; Marteau-Roussy, C.; Aubry, J.-M.; Nardello-Rataj, V. COSMO-RS as an Effective Tool for Predicting the Physicochemical Properties of Fragrance Raw Materials. *Flavour Frag. J.* **2022**, *37* (2), 106–120.

(172) Gaudin, T.; Aubry, J.-M. Prediction of Pourbaix Diagrams of Quinones for Redox Flow Battery by COSMO-RS. *J. Energy Storage* **2022**, *49*, 104152.

(173) Klamt, A. Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena. *J. Phys. Chem.* **1995**, *99* (7), 2224–2235.

(174) McCulloch, W. S.; Pitts, W. A Logical Calculus of the Ideas Immanent in Nervous Activity. *Bulletin of Mathematical Biophysics* **1943**, *5* (4), 115–133.

(175) Dreyfus, G. *Neural Networks: Methodology and Applications*; Springer: Berlin, 2005.

(176) Peterson, K. L. Artificial Neural Networks and Their Use in Chemistry. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; John Wiley & Sons, Ltd, 2000; pp 53–140.

(177) Hammer, B. Recurrent Networks for Structured Data - A Unifying Approach and Its Properties. *Cogn. Syst. Res.* **2002**, *3* (1–4), 145–165.

(178) Gibson, T. Phase-Transfer Synthesis of Monoalkyl Ethers of Oligoethylene Glycols. *J. Org. Chem.* **1980**, *45* (6), 1095–1098.

(179) Lang, J. C.; Morgan, R. D. Nonionic Surfactant Mixtures. I. Phase Equilibria in $C_{10}E_4$–H_2O and Closed-loop Coexistence. *J. Chem. Phys.* **1980**, *73* (11), 5849–5861.

(180) Schlarmann, J.; Stubenrauch, C.; Strey, R. Correlation between Film Properties and the Purity of Surfactants. *Phys. Chem. Chem. Phys.* **2003**, *5* (1), 184–191.

(181) Schubert, K.-V.; Strey, R.; Kahlweit, M. A New Purification Technique for Alkyl Polyglycol Ethers and Miscibility Gaps for Water-CiEj. *J. Colloid Interface Sci.* **1991**, *141* (1), 21–29.

(182) Winsor, P. A. *Solvent Properties of Amphiphilic Compounds*; Butterworth: London, 1954.

(183) Burauer, S.; Sachert, T.; Sottmann, T.; Strey, R. On Microemulsion Phase Behaviour and the Monomeric Solubility of Surfactant. *Phys. Chem. Chem. Phys.* **1999**, *1* (18), 4299–4306.

(184) Kahlweit, M.; Strey, R.; Busse, G. Effect of Alcohols on the Phase Behaviour of Microemulsions. *J. Phys. Chem.* **1991**, *95* (13), 5344–5352.

(185) Kunieda, H.; Shinoda, K. Evaluation of the Hydrophile-Lipophile Balance (HLB) of Nonionic Surfactants. I. Multisurfactant Systems. *J. Colloid Interface Sci.* **1985**, *107* (1), 107–121.

(186) Deen, G. R.; Pedersen, J. S. Phase Behaviour and Microstructure of $C_{12}E_5$ Nonionic Microemulsions with Chlorinated Oils. *Langmuir* **2008**, *24* (7), 3111–3117.

(187) Bouton, F. Influence of Terpenes and Terpenoids on the Phase Behaviour of Micro- and Macro-Emulsions. PhD thesis, Lille 1, 2010.

(188) Lade, O.; Beizai, K.; Sottmann, T.; Strey, R. Polymerizable Nonionic Microemulsions: Phase Behaviour of H_2O-n-Alkyl Methacrylate−*n*-Alkyl Poly(Ethylene Glycol) Ether (C$_i$E$_j$). *Langmuir* **2000**, *16* (9), 4122–4130.

(189) Monari, G.; Dreyfus, G. Local Overfitting Control via Leverages. *Neural Comput.* **2002**, *14* (6), 1481–1506.

(190) Ontiveros, J. F.; Bouton, F.; Durand, M.; Pierlot, C.; Quellet, C.; Nardello-Rataj, V.; Aubry, J.-M. Dramatic Influence of Fragrance Alcohols and Phenols on the Phase Inversion Temperature of the Brij30/n-Octane/Water System. *Colloids Surf. A Physicochem. Eng. Asp.* **2015**, *478*, 54–61.

(191) Lukowicz, T.; Maldonado, R. C.; Molinier, V.; Aubry, J.-M.; Nardello-Rataj, V. Fragrance Solubilization in Temperature Insensitive Aqueous Microemulsions Based on Synergistic Mixtures of Nonionic and Anionic Surfactants. *Colloids Surf. A Physicochem. Eng. Asp.* **2014**, *Complete* (458), 85–95.

(192) Goussard, V.; Aubry, J.-M.; Nardello-Rataj, V. Bio-Based Alternatives to Volatile Silicones: Relationships between Chemical Structure, Physicochemical Properties and Functional Performances. *Adv. Colloid Interface Sci.* **2022**, 102679.

(193) Pugh, R. J. Foaming, Foam Films, Antifoaming and Defoaming. *Adv. Colloid Interface Sci.* **1996**, *64*, 67–142.

(194) Prins, A. Surface Rheology and Practical Behaviour of Foams and Thin Liquid Films. *Chem. Ing. Tech.* **1992**, *64* (1), 73–75.

(195) Engels, Th.; von Rybinski, W.; Schmiedel, P. Structure and Dynamics of Surfactant-Based Foams. In *Structure, Dynamics and Properties of Disperse Colloidal Systems*; Rehage, H., Peschel, G., Eds.; Progress in Colloid & Polymer Science; Steinkopff: Darmstadt, 1998; Vol. 111, pp 117–126.

(196) Rosen, M. J.; Kunjappu, J. T. Foaming and Antifoaming by Aqueous Solutions of Surfactants. In *Surfactants and Interfacial Phenomena*; John Wiley & Sons, Ltd, 2012; pp 308–335.

(197) *Foams and Emulsions*; Sadoc, J. F., Rivier, N., Eds.; NATO ASI Series E; Kluwer Academic: Dordrecht, 1999; Vol. 354.

(198) Exerowa, D.; Kruglyakov, P. M. *Foam and Foam Films: Theory, Experiment, Application*, 1st edition.; Möbius, D., Miller, R., Eds.; Elsevier Science: Amsterdam; New York, 1997.

(199) Wilson, A. J. Experimental Techniques for the Characterization of Foams. In *Foams: Theory Measurements, and Applications*; Prud'homme, R. K., Khan, S. A., Eds.; Surfactant Science Series; M. Dekker: New York, 1996; Vol. 57.

(200) Stubenrauch, C.; Miller, R. Stability of Foam Films and Surface Rheology: An Oscillating Bubble Study at Low Frequencies. *J. Phys. Chem. B* **2004**, *108* (20), 6412–6421.

(201) Santini, E.; Ravera, F.; Ferrari, M.; Stubenrauch, C.; Makievski, A.; Krägel, J. A Surface Rheological Study of Non-Ionic Surfactants at the Water–Air Interface and the Stability

of the Corresponding Thin Foam Films. *Colloids Surf. A Physicochem. Eng. Asp.* **2007**, *298* (1), 12–21.

(202) Friberg, S. E.; Solans, C. Surfactant Association Structures and the Stability of Emulsions and Foams. *Langmuir* **1986**, *2* (2), 121–126.

(203) Friberg, S. E. Foams from Non-Aqueous Systems. *Curr. Opin. Colloid Interf. Sci.* **2010**, *15* (5), 359–364.

(204) Shrestha, L. K.; Acharya, D. P.; Sharma, S. C.; Aramaki, K.; Asaoka, H.; Ihara, K.; Tsunehiro, T.; Kunieda, H. Aqueous Foam Stabilized by Dispersed Surfactant Solid and Lamellar Liquid Crystalline Phase. *J. Colloid Interface Sci.* **2006**, *301* (1), 274–281.

(205) Shrestha, L. K.; Saito, E.; Shrestha, R. G.; Kato, H.; Takase, Y.; Aramaki, K. Foam Stabilized by Dispersed Surfactant Solid and Lamellar Liquid Crystal in Aqueous Systems of Diglycerol Fatty Acid Esters. *Colloids Surf. A Physicochem. Eng. Asp.* **2007**, *293* (1), 262–271.

(206) Kunieda, H.; Shrestha, L. K.; Acharya, D. P.; Kato, H.; Takase, Y.; Gutiérrez, J. M. Super-Stable Nonaqueous Foams in Diglycerol Fatty Acid Esters—Non Polar Oil Systems. *J. Dispers. Sci. Technol.* **2007**, *28* (1), 133–142.

(207) Sagitani, H.; Hayashi, Y.; Ochiai, M. Solution Properties of Homogeneous Polyglycerol Dodecyl Ether Nonionic Surfactants. *JAOCS* **1989**, *66* (1), 146–152.

(208) Shrestha, L. K.; Shrestha, R. G.; Iwanaga, T.; Aramaki, K. Aqueous Phase Behaviour of Diglycerol Fatty Acid Esters. *J. Dispers. Sci. Technol.* **2007**, *28* (6), 883–891.

(209) Stubenrauch, C.; Shrestha, L. K.; Varade, D.; Johansson, I.; Olanya, G.; Aramaki, K.; Claesson, P. Aqueous Foams Stabilized by n-Dodecyl-β-D-Maltoside, Hexaethyleneglycol Monododecyl Ether, and Their 1: 1 Mixture. *Soft Matter* **2009**, *5* (16), 3070–3080.

(210) Stubenrauch, C.; Hamann, M.; Preisig, N.; Chauhan, V.; Bordes, R. On How Hydrogen Bonds Affect Foam Stability. *Adv. Colloid Interface Sci.* **2017**, *247*, 435–443.

(211) Stubenrauch, C.; Claesson, P. M.; Rutland, M.; Manev, E.; Johansson, I.; Pedersen, J. S.; Langevin, D.; Blunk, D.; Bain, C. D. Mixtures of n-Dodecyl-β-D-Maltoside and Hexaoxyethylene Dodecyl Ether - Surface Properties, Bulk Properties, Foam Films, and Foams. *Adv. Colloid Interface Sci.* **2010**, *155* (1–2), 5–18.

(212) Boos, J.; Drenckhan, W.; Stubenrauch, C. Protocol for Studying Aqueous Foams Stabilized by Surfactant Mixtures. *J. Surfact. Deterg.* **2013**, *16* (1), 1–12.

(213) Saulnier, L.; Boos, J.; Stubenrauch, C.; Rio, E. Comparison between Generations of Foams and Single Vertical Films – Single and Mixed Surfactant Systems. *Soft Matter* **2014**, *10* (29), 5280–5288.

(214) Carey, E.; Stubenrauch, C. Foaming Properties of Mixtures of a Non-Ionic (C12DMPO) and an Ionic Surfactant (C12TAB). *J. Colloid Interface Sci.* **2010**, *346* (2), 414–423.

(215) Carey, E.; Stubenrauch, C. Free Drainage of Aqueous Foams Stabilized by Mixtures of a Non-Ionic (C12DMPO) and an Ionic (C12TAB) Surfactant. *Colloids Surf. A Physicochem. Eng. Asp.* **2013**, *419*, 7–14.

(216) Raykundaliya, N.; Bordes, R.; Holmberg, K.; Wu, J.; Somasundaran, P.; Shah, D. O. The Effect on Solution Properties of Replacing a Hydrogen Atom with a Methyl Group in a Surfactant. *Tenside surfactant deter.* **2015**, *52* (5), 369–374.

(217) Schellmann, K.; Preisig, N.; Claesson, P.; Stubenrauch, C. Effects of Protonation on Foaming Properties of Dodecyldimethylamine Oxide Solutions: A pH-Study. *Soft Matter* **2015**, *11* (3), 561–571.

(218) Kanicky, J. R.; Poniatowski, A. F.; Mehta, N. R.; Shah, D. O. Cooperativity among Molecules at Interfaces in Relation to Various Technological Processes: Effect of Chain Length on the pKa of Fatty Acid Salt Solutions. *Langmuir* **2000**, *16* (1), 172–177.

(219) Stubenrauch, C.; Schlarmann, J.; Strey, R. A Disjoining Pressure Study of N-Dodecyl-β-D-Maltoside Foam Films. *Phys. Chem. Chem. Phys.* **2002**, *4* (18), 4504–4513.

(220) Schlarmann, J.; Stubenrauch, C. Stabilization of Foam Films with Non-Ionic Surfactants: Alkyl Polyglycol Ethers Compared with Alkyl Polyglucosides. *Tenside surfactant deter* **2003**, *40* (4), 190–195.

(221) Chang, C.-H.; Franses, E. I. Adsorption Dynamics of Surfactants at the Air/Water Interface: A Critical Review of Mathematical Models, Data, and Mechanisms. *Colloids Surf. A Physicochem. Eng. Asp.* **1995**, *100*, 1–45.

(222) Lucassen, J.; Van den Tempel, M. Longitudinal Waves on Visco-Elastic Surfaces. *J. Colloid Interface Sci.* **1972**, *41* (3), 491–498.

(223) Lucassen, J.; Van den Tempel, M. Dynamic Measurements of Dilational Properties of a Liquid Interface. *Chem. Eng. Sci.* **1972**, *27* (6), 1283–1291.

(224) Boos, J.; Preisig, N.; Stubenrauch, C. Dilational Surface Rheology Studies of n-Dodecyl-β-D-Maltoside, Hexaoxyethylene Dodecyl Ether, and Their 1:1 Mixture. *Adv. Colloid Interface Sci.* **2013**, *197–198*, 108–117.

(225) Mitchell, D. J.; Tiddy, G. J. T.; Waring, L.; Bostock, T.; McDonald, M. P. Phase Behaviour of Polyoxyethylene Surfactants with Water. Mesophase Structures and Partial Miscibility (Cloud Points). *J. Chem. Soc., Faraday Trans. 1* **1983**, *79* (4), 975–1000.

(226) Strey, R.; Schomäcker, R.; Roux, D.; Nallet, F.; Olsson, U. Dilute Lamellar and L3 Phases in the Binary Water–C12E5 System. *J. Chem. Soc., Faraday Trans.* **1990**, *86* (12), 2253–2261.

(227) Warr, G. G.; Drummond, C. J.; Grieser, F.; Ninham, B. W.; Evans, D. F. Aqueous Solution Properties of Nonionic N-Dodecyl-β-D-Maltoside Micelles. *J. Phys. Chem.* **1986**, *90* (19), 4581–4586.

(228) Hishida, M.; Tanaka, K. Transition of the Hydration State of a Surfactant Accompanying Structural Transitions of Self-Assembled Aggregates. *J. Phys.: Condens. Matter* **2012**, *24* (28), 284113.

(229) Izquierdo, P.; Acharya, D. P.; Hirayama, K.; Asaoka, H.; Ihara, K.; Tsunehiro, T.; Shimada, Y.; Asano, Y.; Kokubo, S.; Kunieda, H. Phase Behaviour of Pentaglycerol Monostearic and Monooleic Acid Esters in Water. *J. Dispers. Sci. Technol.* **2006**, *27* (1), 99–103.

(230) Rosen, M. J.; Cohen, A. W.; Dahanayake, M.; Hua, X. Y. Relationship of Structure to Properties in Surfactants. 10. Surface and Thermodynamic Properties of 2-Dodecyloxypoly(Ethenoxyethanol)s, C12H25(OC2H4)XOH, in Aqueous Solution. *J. Phys. Chem.* **1982**, *86* (4), 541–545.

(231) Drummond, C. J.; Warr, G. G.; Grieser, F.; Ninham, B. W.; Evans, D. F. Surface Properties and Micellar Interfacial Microenvironment of N-Dodecyl-β-D-Maltoside. *J. Phys. Chem.* **1985**, *89* (10), 2103–2109.

(232) Shinoda, K.; Yamaguchi, T.; Hori, R. The Surface Tension and the Critical Micelle Concentration in Aqueous Solution of β-D-Alkyl Glucosides and Their Mixtures. *BCSJ* **1961**, *34* (2), 237–241.

(233) Jayalakshmi, Y.; Ozanne, L.; Langevin, D. Viscoelasticity of Surfactant Monolayers. *J. Colloid Interface Sci.* **1995**, *170* (2), 358–366.

(234) Sonin, A. A.; Bonfillon, A.; Langevin, D. Thinning of Soap Films: The Role of Surface Viscoelasticity. *J. Colloid Interface Sci.* **1994**, *162* (2), 323–330.

(235) Bergeron, V. Disjoining Pressures and Film Stability of Alkyltrimethylammonium Bromide Foam Films. *Langmuir* **1997**, *13* (13), 3474–3482.

(236) Langevin, D. Influence of Interfacial Rheology on Foam and Emulsion Properties. *Adv. Colloid Interface Sci.* **2000**, *88* (1–2), 209–222.

(237) Bergeron, V. Forces and Structure in Thin Liquid Soap Films. *J. Phys.: Condens. Matter* **1999**, *11* (19), R215–R238.

(238) Georgieva, D.; Cagna, A.; Langevin, D. Link between Surface Elasticity and Foam Stability. *Soft Matter* **2009**, *5* (10), 2063–2071.

(239) Lucassen-Reynders, E. H. *Anionic Surfactants: Physical Chemistry of Surfactant Action*; Surfactant Science Series; M. Dekker: New York, 1981.

(240) Kanduč, M.; Schneck, E.; Stubenrauch, C. Intersurfactant H-Bonds Between Head Groups of n-Dodecyl-β-D-Maltoside at the Air-Water Interface. *J. Colloid Interface Sci.* **2021**, *586*, 588–595.

(241) Ranieri, D.; Preisig, N.; Stubenrauch, C. On the Influence of Intersurfactant H-Bonds on Foam Stability: A Study with Technical Grade Surfactants. *TSD* **2018**, *55* (1), 6–16.

(242) Ravera, F.; Ferrari, M.; Santini, E.; Liggieri, L. Influence of Surface Processes on the Dilational Visco-Elasticity of Surfactant Solutions. *Adv. Colloid Interface Sci.* **2005**, *117* (1), 75–100.

(243) Ostwald, W. Beiträge Zur Kenntnis Der Emulsionen. *Kolloid Z* **1910**, *6*, 103–109.

(244) Bancroft, W. D. The Theory of Emulsification, V. *J. Phys. Chem.* **1913**, *17* (6), 501–519.

(245) Bancroft, W. D. The Theory of Emulsification, VI. *J. Phys. Chem.* **1915**, *19* (4), 275–309.

(246) Ontiveros, J. F.; Pierlot, C.; Catté, M.; Molinier, V.; Salager, J.-L.; Aubry, J.-M. A Simple Method to Assess the Hydrophilic Lipophilic Balance of Food and Cosmetic Surfactants Using the Phase Inversion Temperature of C10E4/n-Octane/Water Emulsions. *Colloids Surf. A Physicochem. Eng. Asp.* **2014**, *458*, 32–39.

(247) Lemahieu, G.; Ontiveros, J. F.; Gaudin, T.; Molinier, V.; Aubry, J.-M. The Salinity-Phase-Inversion Method (SPI-Slope): A Straightforward Experimental Approach to Assess the Hydrophilic-Lipophilic-Ratio and the Salt-Sensitivity of Surfactants. *J. Colloid Interface Sci.* **2022**, *608*, 549–563.

(248) Salager, J.-L. Guidelines for the Formulation, Composition and Stirring to Attain Desired Emulsion Properties (Type, Droplet Size, Viscosity and Stability). In *Surfactants in Solution*; Surfactant Science Series; CRC Press: New York, 1996; pp 261–295.

(249) Salager, J.-L.; Antón, R.; Bullón, J.; Forgiarini, A.; Marquez, R. How to use the normalized hydrophilic-lipophilic deviation (HLD$_N$) concept for the formulation of equilibrated and emulsified surfactant-oil-water systems for cosmetics and pharmaceutical products. *Cosmetics* **2020**, *7* (3), 57.

(250) Kunz, W.; Testard, F.; Zemb, T. Correspondence between Curvature, Packing Parameter, and Hydrophilic−Lipophilic Deviation Scales around the Phase-Inversion Temperature. *Langmuir* **2009**, *25* (1), 112–115.

(251) Hayes, M. E.; Bourrel, M.; El-Emary, M. M.; Schechter, R. S.; Wade, W. H. Interfacial Tension and Behaviour of Nonionic Surfactants. *SPE J.* **1979**, *19* (06), 349–356.

(252) Salager, J. L.; Bourrel, M.; Schechter, R. S.; Wade, W. H. Mixing Rules for Optimum Phase-Behaviour Formulations of Surfactant/Oil/Water Systems. *SPE J.* **1979**, *19* (05), 271–278.

(253) Sottmann, T.; Strey, R. Ultralow Interfacial Tensions in Water–n-Alkane–Surfactant Systems. *J. Chem. Phys.* **1997**, *106* (20), 8606–8615.

(254) Pierlot, C.; Ontiveros, J. F.; Catté, M.; Salager, J.-L.; Aubry, J.-M. Cone–Plate Rheometer as Reactor and Viscosity Probe for the Detection of Transitional Phase Inversion of Brij30–Isopropyl Myristate–Water Model Emulsion. *Ind. Eng. Chem. Res.* **2016**, *55* (14), 3990–3999.

(255) Salager, J. L.; Miñana-Pérez, M.; Andérez, J. M.; Grosso, J. L.; Rojas, C. I.; Layrisse, I. Surfactant-Oil-Water Systems Near the Affinity Inversion Part II: Viscosity of Emulsified Systems. *J. Dispers. Sci. Technol.* **1983**, *4* (2), 161–173.

(256) Shinoda, K.; Arai, H. The Correlation between Phase Inversion Temperature In Emulsion and Cloud Point in Solution of Nonionic Emulsifier. *J. Phys. Chem.* **1964**, *68* (12), 3485–3490.

(257) Salager, J.-L.; Márquez, L.; Peña, A. A.; Rondón, M.; Silva, F.; Tyrode, E. Current Phenomenological Know-How and Modeling of Emulsion Inversion. *Ind. Eng. Chem. Res.* **2000**, *39* (8), 2665–2676.

(258) Davies, J. T. A Quantitative Kinetic Theory of Emulsion Type, I. Physical Chemistry of the Emulsifying Agent. In *Proc. 2nd Int. Congr. Surf. Act. 1*; 426; 1957.

(259) Bourrel, M.; Chambu, C. The Rules for Achieving High Solubilization of Brine and Oil by Amphiphilic Molecules. *Soc. Pet. Eng. J.* **1983**, *23* (2), 327–338.

(260) Israelachvili, J. N.; Mitchell, D. J.; Ninham, B. W. Theory of Self-Assembly of Hydrocarbon Amphiphiles into Micelles and Bilayers. *J. Chem. Soc., Faraday Trans. 2* **1976**, *72* (0), 1525–1568.

(261) Tanford, C. *The Hydrophobic Effect: Formation of Micelles and Biological Membranes*; Wiley, 1980.

(262) Tchakalova, V.; Testard, F.; Wong, K.; Parker, A.; Benczédi, D.; Zemb, T. Solubilization and Interfacial Curvature in Microemulsions: I. Interfacial Expansion and Co-Extraction of Oil. *Colloids Surf. A Physicochem. Eng. Asp.* **2008**, *331* (1), 31–39.

(263) Tchakalova, V.; Testard, F.; Wong, K.; Parker, A.; Benczédi, D.; Zemb, T. Solubilization and Interfacial Curvature in Microemulsions: II. Surfactant Efficiency and PIT. *Colloids Surf. A Physicochem. Eng. Asp.* **2008**, *331* (1), 40–47.

(264) Matsumoto, S.; Sherman, P. A DTA Technique for Identifying the Phase Inversion Temperature of O/W Emulsions. *J. Colloid Interface Sci.* **1970**, *33* (2), 294–298.

(265) Enever, R. P. Correlation of Phase Inversion Temperature with Kinetics of Globule Coalescence for Emulsions Stabilized by a Polyoxyethylene Alkyl Ether. *J Pharm Sci* **1976**, *65* (4), 517–520.

(266) Pizzino, A. Inversion de Phase Des Émulsions : Relation Avec Le Comportement à l'équilibre et Détection Par Rétrodiffusion de Lumière. These de doctorat, Lille 1, 2008.

(267) Allouche, J.; Tyrode, E.; Sadtler, V.; Choplin, L.; Salager, J.-L. Simultaneous Conductivity and Viscosity Measurements as a Technique To Track Emulsion Inversion by the Phase-Inversion-Temperature Method. *Langmuir* **2004**, *20* (6), 2134–2140.

(268) Ontiveros, J. F.; Pierlot, C.; Catté, M.; Molinier, V.; Salager, J.-L.; Aubry, J.-M. Structure–Interfacial Properties Relationship and Quantification of the Amphiphilicity of Well-Defined Ionic and Non-Ionic Surfactants Using the PIT-Slope Method. *J. Colloid Interface Sci.* **2015**, *448*, 222–230.

(269) Lebeuf, R.; Liu, C.-Y.; Pierlot, C.; Nardello-Rataj, V. Synthesis and Surfactant Properties of Nonionic Biosourced Alkylglucuronamides. *ACS Sustainable Chem. Eng.* **2018**, *6* (2), 2758–2766.

(270) Ontiveros, J. F.; Nollet, M.; Pierlot, C.; Nardello-Rataj, V. Revisiting the Influence of Carboxylic Acids on Emulsions and Equilibrated SOW Systems Using the PIT-Slope Method. *Colloids Surf. A Physicochem. Eng. Asp.* **2018**, *536*, 191–197.

(271) Lemahieu, G.; Aguilhon, J.; Strub, H.; Molinier, V.; Ontiveros, J. F.; Aubry, J.-M. Hexahydrofarnesyl as an Original Bio-Sourced Alkyl Chain for the Preparation of Glycosides Surfactants with Enhanced Physicochemical Properties. *RSC Adv.* **2020**, *10* (28), 16377–16389.

(272) Ontiveros, J. F.; Hong, B.; Aramaki, K.; Pierlot, C.; Aubry, J.-M.; Nardello-Rataj, V. Cation Effect on the Binary and Ternary Phase Behaviours of Double-Tailed Methanesulfonate Amphiphiles. *J Surfactants Deterg* **2021**, *24* (3), 401–410.

(273) Salager, J. L.; Loaiza-Maldonado, I.; Minana-Perez, M.; Silva, F. Surfactant-Oil-Water Systems Near the Affinity Inversion Part I: Relationship Between Equilibrium Phase Behaviour and Emulsion Type and Stability. *J. Dispers. Sci. Technol.* **1982**, *3* (3), 279–292.

(274) Lemahieu, G.; Ontiveros, J. F.; Terra Telles Souza, N.; Molinier, V.; Aubry, J.-M. Fast and Accurate Selection of Surfactants for Enhanced Oil Recovery by Dynamic Salinity-Phase-Inversion (SPI). *Fuel* **2021**, *289*, 119928.

(275) Chirife, J.; Resnik, S. L. Unsaturated Solutions of Sodium Chloride as Reference Sources of Water Activity at Various Temperatures. *J. Food Sci.* **1984**, *49* (6), 1486–1488.

(276) Shi, Y.; Dayoub, W.; Favre-Réguillon, A.; Chen, G.-R.; Lemaire, M. Straightforward Selective Synthesis of Linear 1-O-Alkyl Glycerol and Di-Glycerol Monoethers. *Tetrahedron Lett.* **2009**, *50* (49), 6891–6893.

(277) Wade, W. H.; Morgan, J. C.; Schechter, R. S.; Jacobson, J. K.; Salager, J. L. Interfacial Tension and Phase Behaviour of Surfactant Systems. *Soc. Pet. Eng. J.* **1978**, *18* (04), 242–252.

(278) Ontiveros, J. F.; Pierlot, C.; Catté, M.; Salager, J.-L.; Aubry, J.-M. Determining the Preferred Alkane Carbon Number (PACN) of Nonionic Surfactants Using the PIT-Slope Method. *Colloids Surf. A Physicochem. Eng. Asp.* **2018**, *536*, 30–37.

(279) Roberts, J. D.; Caserio, M. C. Alcohols and Ethers. In *Basic Principles of Organic Chemistry, second edition*; W. A. Benjamin, Inc.: Menlo Park, CA, 1977.

(280) Salager, J.-L.; Anton, R.; Aubry, J.-M. Formulation des émulsions par la méthode du HLD. *Tech. L'Ingénieur* **2006**.

(281) Bourrel, M.; Salager, J. L.; Schechter, R. S.; Wade, W. H. A Correlation for Phase Behaviour of Nonionic Surfactants. *J. Colloid Interface Sci.* **1980**, *75* (2), 451–461.

(282) Salager, J.-L. The Fundamental Basis for the Action of a Chemical Dehydrant. Influence of the Physical and Chemical Formulation on the Stability of an Emulsion. *Int. Chem. Eng.* **1990**, *30*, 103–116.

(283) Zheng, L.; Cao, C.; Li, R.-Y.; Cao, L.-D.; Zhou, Z.-L.; Li, M.; Huang, Q.-L. Preparation and Characterization of Water-in-Oil Emulsions of Isoprothiolane. *Colloids Surf. A Physicochem. Eng. Asp.* **2018**, *537*, 399–410.

(284) Peng, K.; Wang, X.; Lu, L.; Liu, J.; Guan, X.; Huang, X. Insights into the Evolution of an Emulsion with Demulsifying Bacteria Based on Turbiscan. *Ind. Eng. Chem. Res.* **2016**, *55* (25), 7021–7029.

(285) Kunieda, H.; Friberg, S. E. Critical Phenomena in a Surfactant/Water/Oil System. Basic Study on the Correlation between Solubilization, Microemulsion, and Ultralow Interfacial Tensions. *BCSJ* **1981**, *54* (4), 1010–1014.

(286) Engels, T.; Förster, T.; von Rybinski, W. The Influence of Coemulsifier Type on the Stability of Oil-in-Water Emulsions. *Colloids Surf. A Physicochem. Eng. Asp.* **1995**, *99* (2), 141–149.

(287) Friberg, S.; Solans, C. Emulsification and the HLB-Temperature. *J. Colloid Interface Sci.* **1978**, *66* (2), 367–368.

(288) Tadros, T.; Izquierdo, P.; Esquena, J.; Solans, C. Formation and Stability of Nano-Emulsions. *Adv. Colloid Interface Sci.* **2004**, *108–109*, 303–318.

(289) Joint FAO/WHO Expert Committee on Food Additives (JECFA) - Octyl octanoate https://www.fao.org/food/food-safety-quality/scientific-advice/jecfa/jecfa-flav/details/en/c/25/ (accessed Jul 27, 2022).

(290) *Commission Implementing Regulation (EU) No 872/2012 of 1 October 2012 Adopting the List of Flavouring Substances Provided for by Regulation (EC) No 2232/96 of the European Parliament and of the Council, Introducing It in Annex I to Regulation (EC) No 1334/2008 of the European Parliament and of the Council and Repealing Commission Regulation (EC) No 1565/2000 and Commission Decision 1999/217/EC Text with EEA Relevance*; 2012; Vol. 267.

(291) Zocchi, G. Skin-Feel Agents. In *Handbook of Cosmetic Science and Technology*; CRC Press: Boca Raton, 2001.

Appendix

A1. Abbreviations

a_0	Polar head surface
A	Affinity
ACN	Alkane Carbon Number
A_i	Hamaker's constant
A_m	Area per molecule
AN	Acceptor Number
BS	Backscattered light signal
C*	Optimal concentration
$C_{12}E_5$	Pentaethylene glycol monododecyl ether
$C_{12}Glu_2$	n-dodecyl-β-D-maltoside
$C_{12}Gly_2$	1-O-dodecyl diglyceryl ether
c	Bulk surfactant concentration
C_iE_j	Polyethoxylated fatty alcohol
CMC	Critical Micellar Concentration
C_nGlu_m	Alkylpolyglucosides
COSMO-RS	COnductor like Screening MOdel for Real Solvents
CPT	Capillary Pressure Tensiometer
$D_{[4,3]}$	Statistical diameter in volume
d	Diameter
DFA	Dynamic Foam Analyzer
DLS	Dynamic Light Scattering
DLVO	Derjaguin Landau Verwey Overbeek
DMF	Dimethylformamide
DMSO	Dimethylsulfoxide
DN	Donor Number
d_{Stokes}	Stokes diameter
E	Surface dilational visco-elasticity
e	Electron charge
EACN	Equivalent Alkane Carbon Number
EO	Ethylene Oxide
EON	Ethylene Oxide Number
EOR	Enhanced Oil Recovery
FID	Flame Ionization Detector
f_{liq}	Liquid fraction
f_w	Water/oil fraction
$g(d)$	Asymmetry factor

g	Gravitational constant
GC	Gas Chromatography
GM	Graph Machine
h	Planck's constant
HBA	Hydrogen Bond Acceptor
HBD	Hydrogen Bond Donor
HIPE	High Internal Phase Emulsion
HLB	Hydrophilic-Lipophilic-Balance
HLD_N	Normalized Hydrophilic-Lipophilic-Deviation
HSP	Hansen Solubility Parameters
I_0	Incident light intensity
I	Ionic strength
$I_S(\theta)$	Scattered light intensity at the angle θ
IP	Ionization Potential
K	Partition coefficient
k_B	Boltzmann constant
Lα	Lamellar liquid crystal phase
l	Length of the surfactant molecule lipophilic moiety
LC	Liquid Crystal
LOOs	Leave One Out score
M	Molar weight
M_i^X	σ-moment i
MLP	Multi-Layer Perceptron
MLR	Multiple Linear Regression
n	Refractive Index
N_A	Avogadro constant
N_C	Number of carbon atoms
N_h	Number of height positions
NMP	N-methylpyrrolidone
NMR	Nuclear Magnetic Resonance
NN	Neural Network
NP	Nanoparticle
O	Oil
O/W	Oil in Water
P	n-octanol-water partition coefficient
p	Packing parameter
PACN	Preferred Alkane Carbon Number
PEEK	Polyether-ether-ketone
PIT	Phase Inversion Temperature
$p^X(\sigma)$	σ-profile
Q	Flow rate

Q_e	Extinction efficiency
QSPR	Quantitative Structure Property Relationship
R	Molar gas constant
R	R-Ratio of Winsor
R.T.	Room Temperature
RED	Relative Energy Difference
RMSE	Root Mean Square Error
RMSTE	Root Mean Square Training Error
$RTSI$	Relative Turbiscan Stability Index
S*	Optimal salinity
S	Salinity
SAD	Surfactant Affinity Difference
SMILES	Simplified Molecular Input Line Entry Specification
SMLS	Static Multiple Light Scattering
SOW	Surfactant/Oil/Water
SPI	Salinity of Phase Inversion
T*	Optimal Temperature
T	Temperature
t	Time
TBAH	Tetrabutylammonium hydroxide
TFA	Trifluoroacetic acid
THF	Tetrahydrofuran
T_{HLB}	Hydrophilic Lipophilic Balance Temperature
TR	Transmitted light signal
TSI	Turbiscan Stability Index
u	Mobility
v_0	Volume of the surfactant molecule lipophilic moiety
V	Volume
v	Sedimentation rate
V_A	Attractive potential
VLOOs	Virtual Leave-One-Out score
V_m	Molar volume
V_R	Repulsive potential
V_T	Total interaction potential
W	Water
W I, W II, W III, W IV	Winsor I, II, III, IV phases
W/O	Water in Oil
WOR	Water-to-Oil Ratio
x_i	Molar fraction of surfactant i
z_{min}, z_{max}	Lower and upper height limits
α	Gain of the experimental setup

β	Offset of the experimental setup
Γ	Surfactant surface concentration
γ	Surface tension
ΔG_{mix}	Gibbs energy of mixture
$\Delta G_{solv.}$	Energy of solvation
ΔH_{mix}	Enthalpy of mixture
ΔH_v	Heat of vaporization
ΔS_{mix}	Entropy of mixture
δ	Salinity sensitivity coefficient
δ_d	Dispersive interaction parameter
δ_H	Hildebrand solubility parameter
δ_h	Hydrogen bonding interaction parameter
δ_p	Polar interaction parameter
ε'	Permittivity
ε'_0	Electrical permittivity of vacuum
ε'_r	Relative permittivity
ε_0	High frequency limit of dilational surface elasticity
ε_i	Imaginary part of surface dilational visco-elasticity
ε_r	Dilational elasticity
ζ	Zeta potential
η	Viscosity
κ^{-1}	Debye-Hückel distance
λ	Wavelength
ν	Frequency
ν_e	Main electronic absorption frequency
ρ	Density
τ	Temperature sensitivity coefficient
Φ	Volume fraction of dispersed phase
φ	Dephasing angle
ω_0	Molecular exchange parameter
ω	Oscillation pulse

A2. Hamaker's constants calculation

Hamaker's constant A_{12} of component 1 in medium 2 is calculated with equation (4) from each component constants A_1 and A_2 expressed by equation (5).

$$A_{12} = \left(\sqrt{A_1} - \sqrt{A_2}\right)^2 \tag{4}$$

$$A_i = \frac{3}{4} k_B T \frac{(\varepsilon'_{r,i} - 1)^2}{(\varepsilon'_{r,i} + 1)^2} + \frac{3 h \nu_e (n_i^2 - 1)^2}{16 \sqrt{2} (n_i^2 + 1)^{3/2}} \tag{5}$$

a, the particle radius, is taken to be 25 nm. Ionic strength I is taken to be 10^{-3} mol.m^{-3} in pure solvents and 1 mol.m^{-3} for samples with 10^{-3} M trifluoroacetic acid or tetrabutylammonium hydroxide. ν is the electronic absorption frequency and is calculated based on ionization potential IP as $\nu = IP/h$ where h is Planck's constant. $\varepsilon'_{r,i}$, n_i, IP and ν are listed in Table A1 along with A_i and A_{1i} values.

Table A1. Relative dielectric constant ε'_r, refractive index n, main electronic absorption frequency ν calculated based on ionization potential IP and Hamaker's constants A_i and A_{1i} for TiO$_2$ and solvents studied at 25.0 °C.

Compound	ε'_r	n	IP (eV)	ν (10^{15} s^{-1})	A_i (10^{-20} J)	A_{1i} (10^{-19} J)
TiO$_2$	48.0	2.550	9.5 [a]	2.3	30.1	
Acetone	20.7	1.359	9.69 [b]	2.3	3.3	1.3
Acetonitrile	37.5	1.344	12.22 [c]	3.0	3.9	1.2
γ-Butyrolactone	41.0	1.434	10.26 [d]	2.5	4.9	1.1
DMF	36.7	1.430	9.12 [c]	2.2	4.2	1.2
Ethanol	24.5	1.360	10.48 [c]	2.5	3.6	1.3
Ethyl Acetate	6.0	1.372	10.11 [c]	2.4	3.5	1.3
Heptane	1.9	1.387	10.08 [c]	2.4	3.6	1.3
Isopropanol	17.9	1.377	10.15 [e]	2.5	3.8	1.2
Methanol	32.7	1.315	10.85 [b]	2.6	3.0	1.4
NMP	33.0	1.470	9.17 [f]	2.2	4.9	1.1
Nitromethane	39.0	1.382	11.08 [c]	2.7	4.2	1.2
Propylene Carbonate	64.9	1.419	10.5 [g]	2.5	4.6	1.1
Pyridine	12.4	1.510	9.9 [h]	2.4	6.0	0.9
THF	7.6	1.407	9.54 [c]	2.3	4.0	1.2
Toluene	2.4	1.496	8.82 [b]	2.1	4.9	1.1
Triethylamine	2.4	1.400	7.5 [i]	1.8	2.9	1.4
Water	80.1	1.330	12.61 [j]	3.0	3.7	1.3

[a] G. Balducci, G. Gigli, M. Guido, *J. Chem. Phys.* **1985**, *83* (4), 1913–1916.
[b] K. Watanabe, *J. Chem. Phys.* **1954**, *22* (9), 1564–1570.
[c] K. Watanabe, T. Nakayama, J. Mottl, *J. Quant. Spectry. Radiative Transfer* **1962**, *2* (4), 369–382.
[d] A. D. Bain, D. C. Frost, *Can. J. Chem.* **1973**, *51* (8), 1245–1247.
[e] K. Watanabe, *J. Chem. Phys.* **1957**, *26* (3), 542–547.
[f] L. Treschanke, P. Rademacher, *J. Mol. Struct.* **1985**, *122* (1–2), 47–57.
[g] Propylene carbonate https://ionscience.com/gases/propylene-carbonate/ (accessed Jul 15, 2021).
[h] D. Stefanović, H. F. Grützmacher, *Org. Mass Spectrom.* **1974**, *9* (10), 1052–1054.
[i] K. Watanabe, J. R. Mottl, *J. Chem. Phys.* **1957**, *26* (6), 1773–1774.
[j] P. Gürtler, V. Saile, E. E. Koch, *Chem. Phys. Lett.* **1977**, *51* (2), 386–391.

A3. Turbiscan profiles of TiO$_2$ dispersions in aqueous media

In this section, the variations of transmitted (ΔT) and backscattered (ΔR) signals obtained using a Turbiscan LAB are shown for 1 g/L TiO$_2$ P25 aqueous dispersions at various pH values. pH is adjusted by adding HCl or NaOH and ionic strength is kept constant at 10^{-3} M by NaCl addition. Signals were recorded from 0 (blue curve) to 24 hours (red curve).
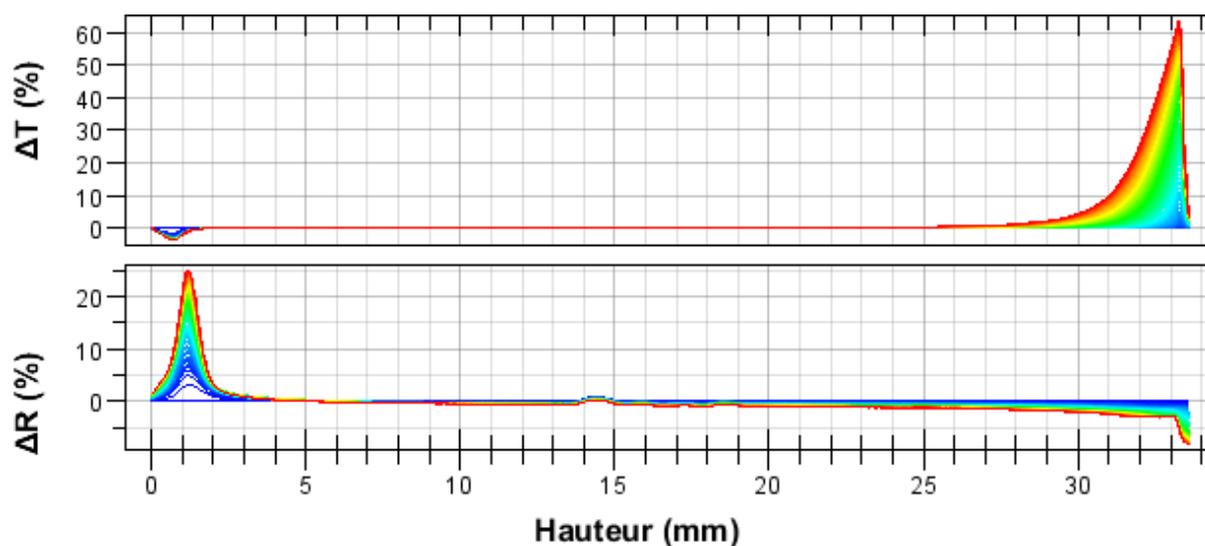
Figure A.1. Aqueous TiO$_2$ dispersion such as pH = 3.1.
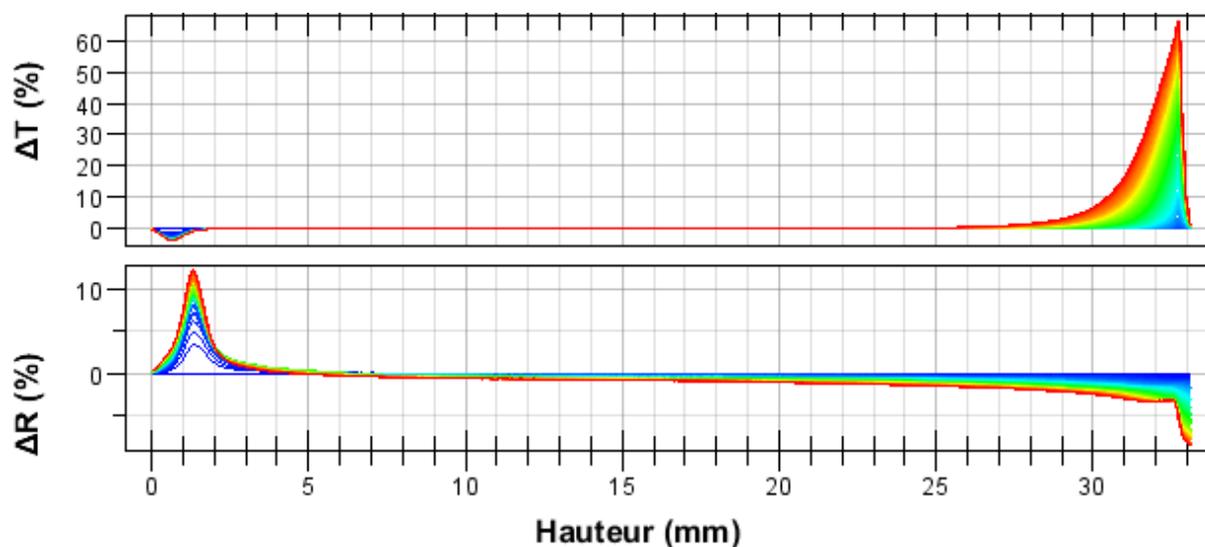
Figure A.2. Aqueous TiO$_2$ dispersion such as pH = 3.7.
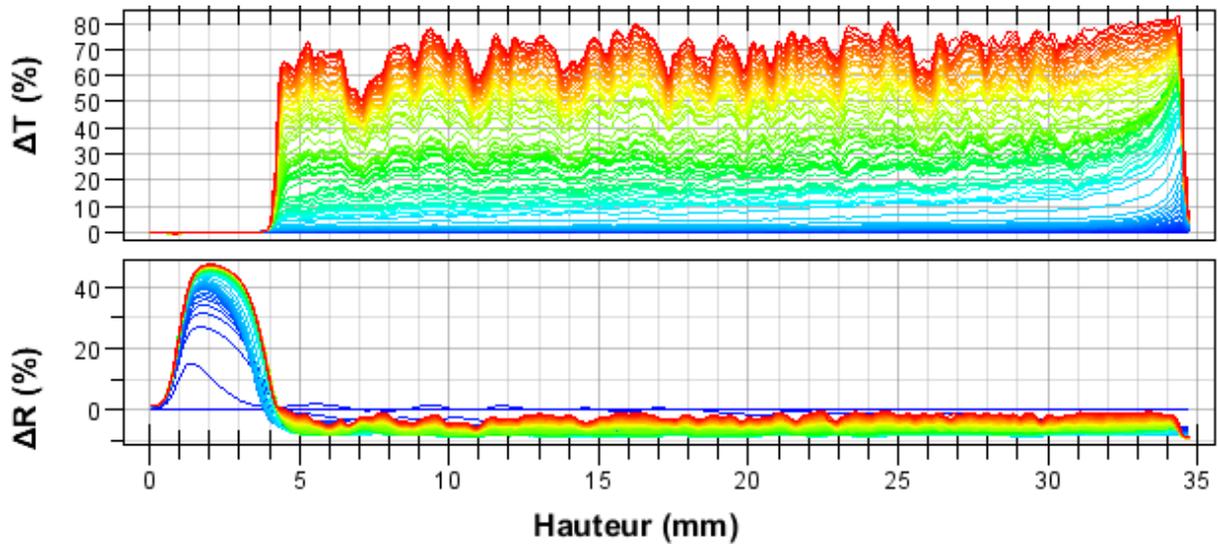
Figure A.3. Aqueous TiO$_2$ dispersion such as pH = 4.7.
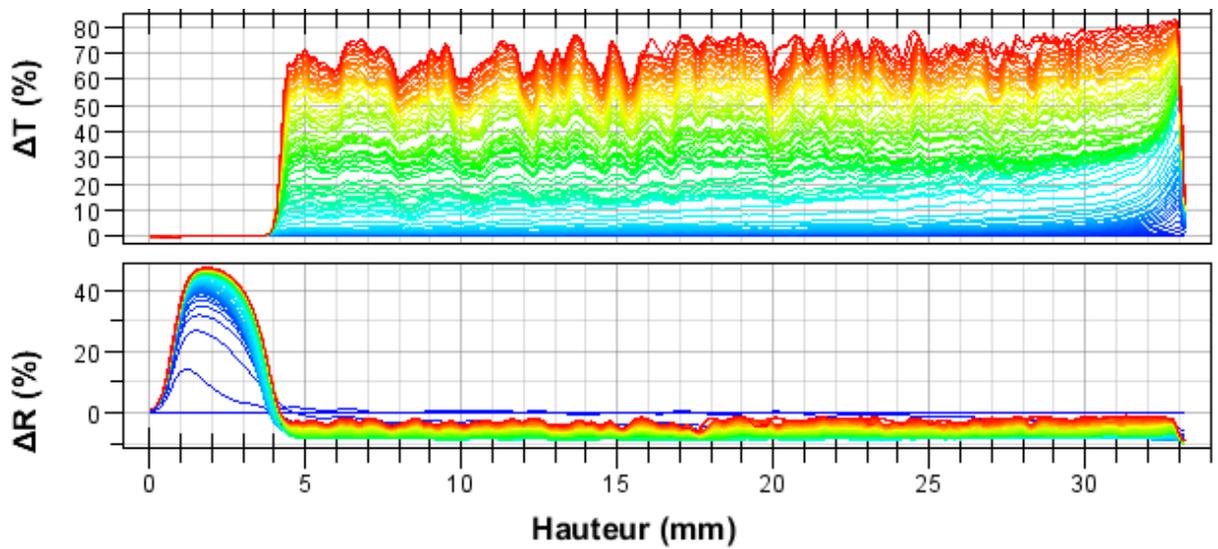
Figure A.4. Aqueous TiO$_2$ dispersion such as pH = 5.5.
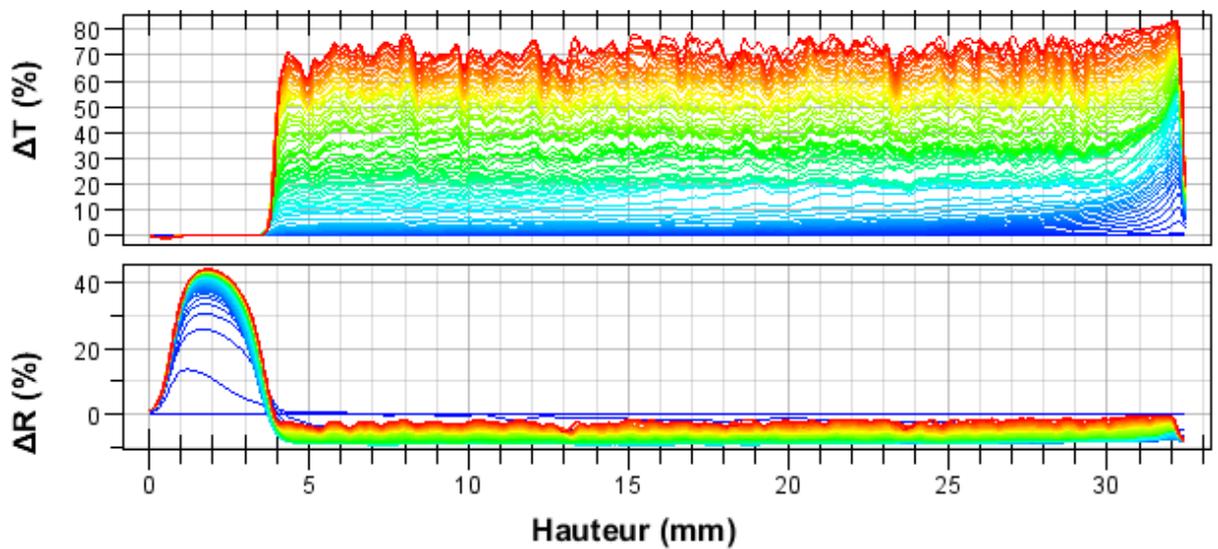
Figure A.5. Aqueous TiO$_2$ dispersion such as pH = 6.7.
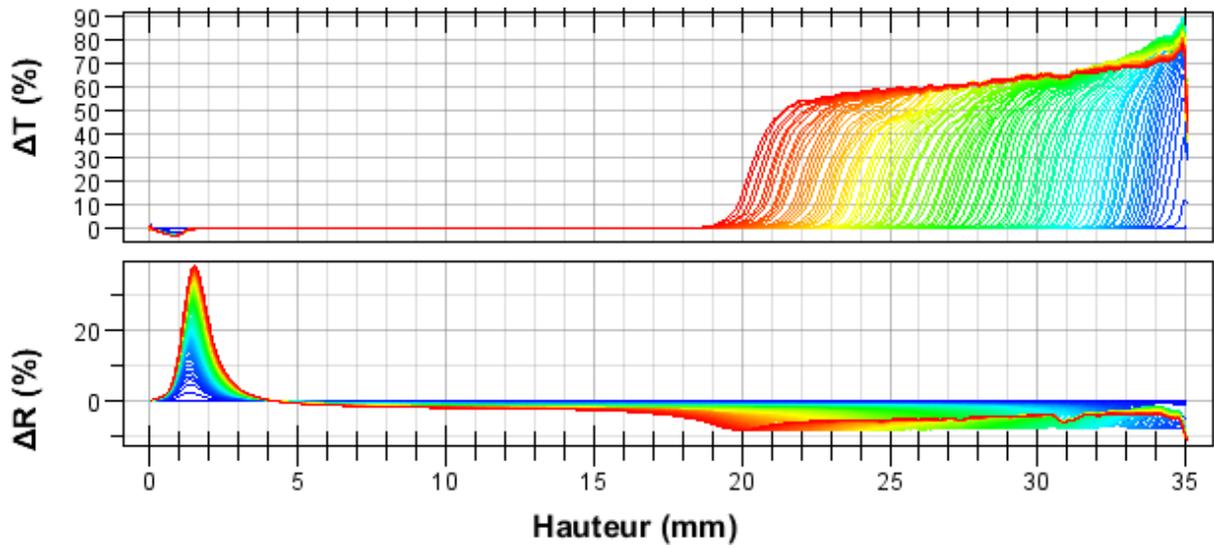
Figure A.6. Aqueous TiO_2 dispersion such as pH = 7.8.
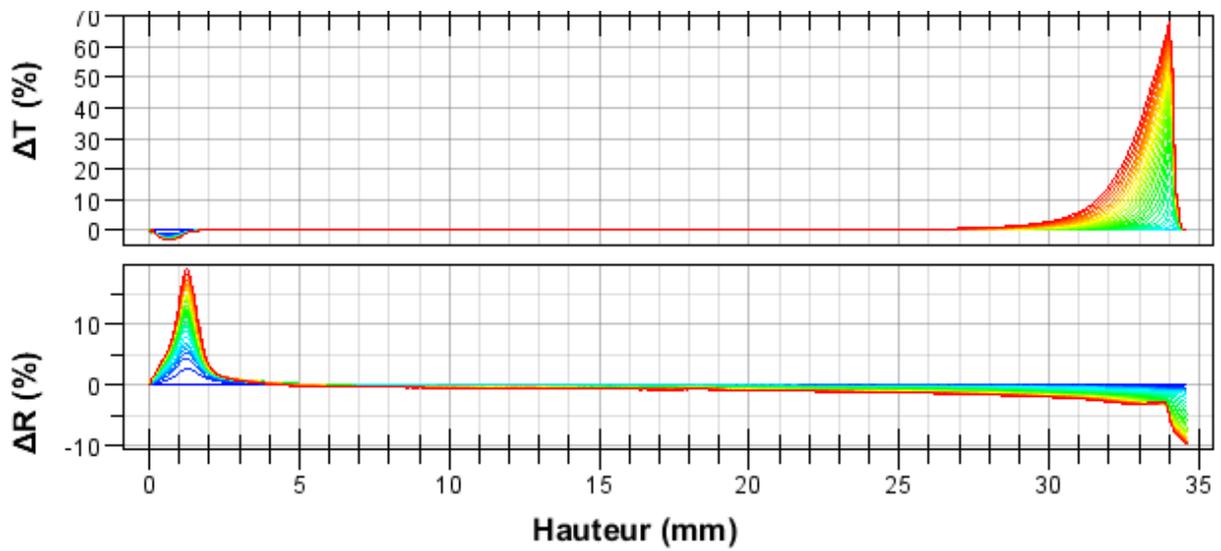
Figure A.7. Aqueous TiO_2 dispersion such as pH = 9.5.
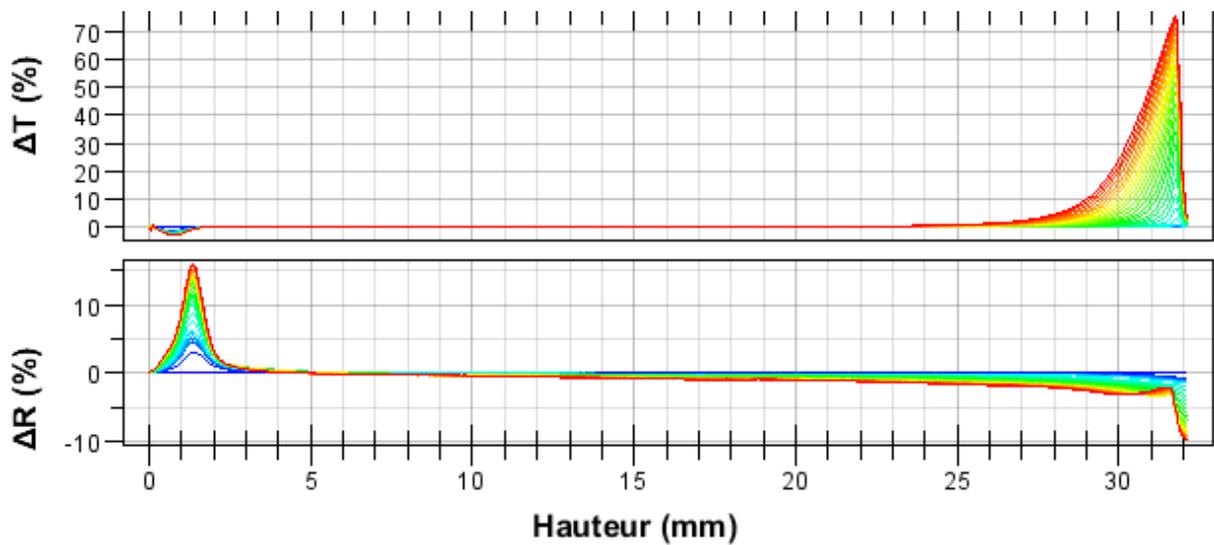
Figure A.8. Aqueous TiO_2 dispersion such as pH = 10.6.

A4. Turbiscan profiles of TiO$_2$ dispersions in "pure" non-aqueous solvents

In this section, the variations of transmitted (ΔT) and backscattered (ΔR) signals obtained using a Turbiscan LAB are shown for 1 g/L TiO$_2$ P25 non-aqueous dispersions with solvents used as received. Signals were recorded from 0 (blue curve) to 2 hours (red curve). Samples are presented in the order of increasing measured d$_{Stokes}$ (same order as in Table 1.1).
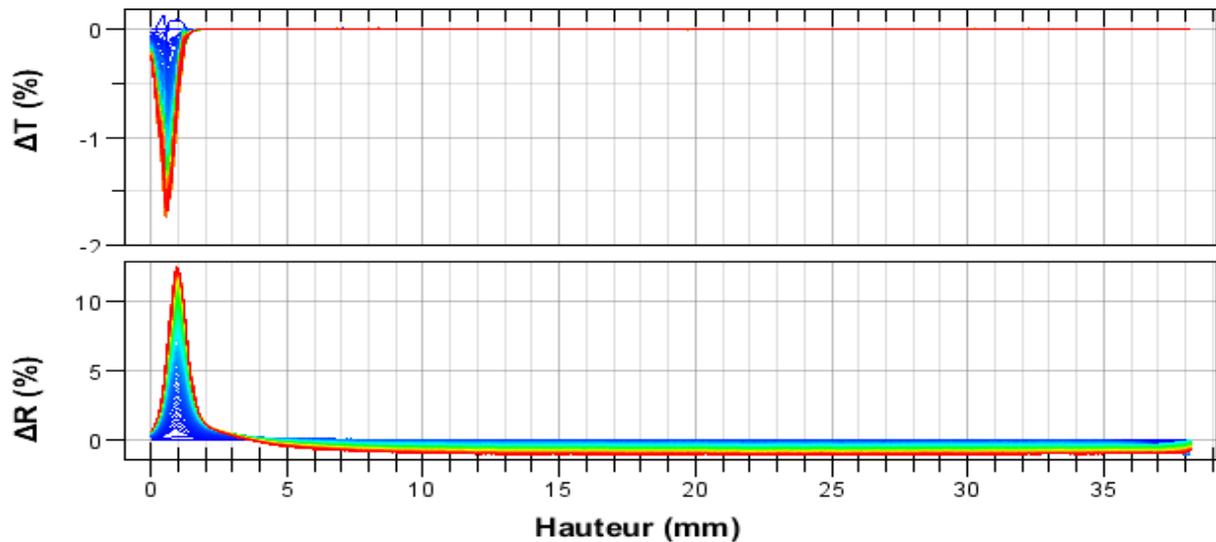
Figure A.9. Non-aqueous TiO$_2$ dispersion in nitromethane.
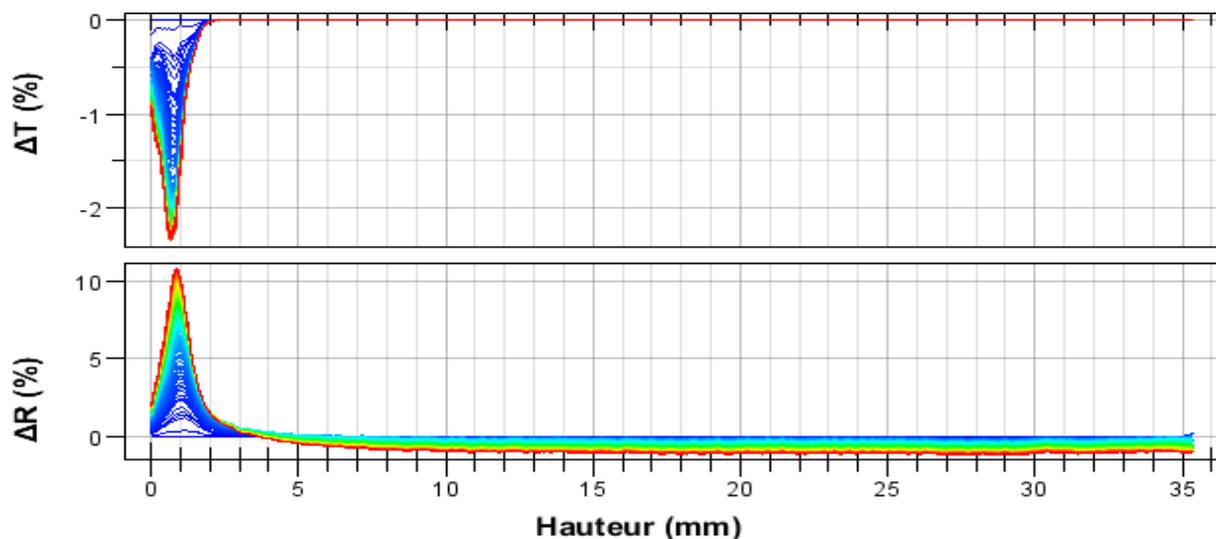
Figure A.10. Non-aqueous TiO$_2$ dispersion in ethanol.

143
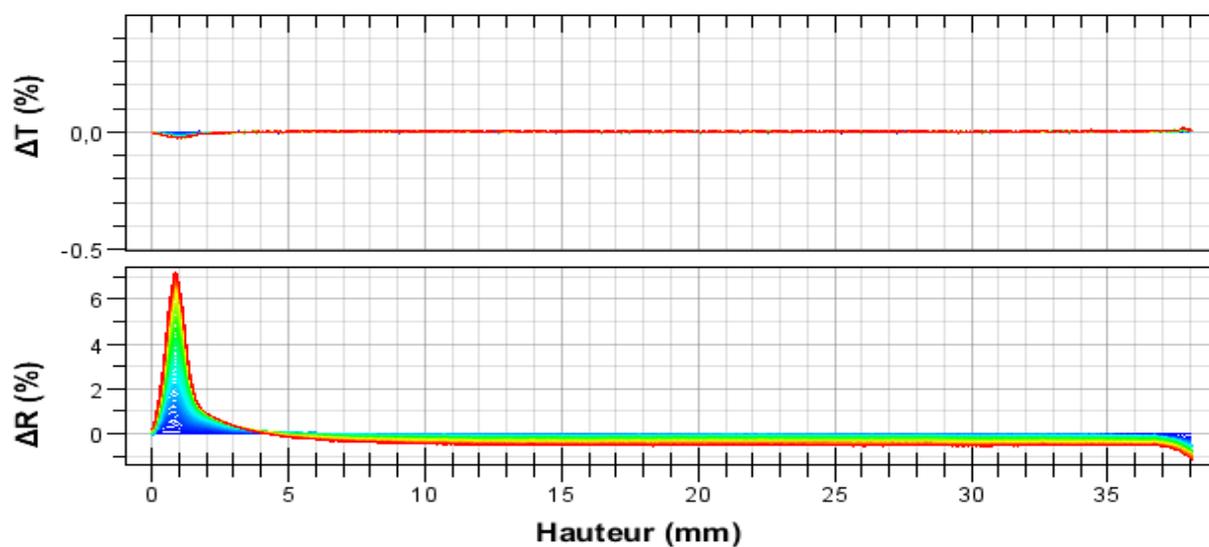
Figure A.11. Non-aqueous TiO$_2$ dispersion in pyridine.
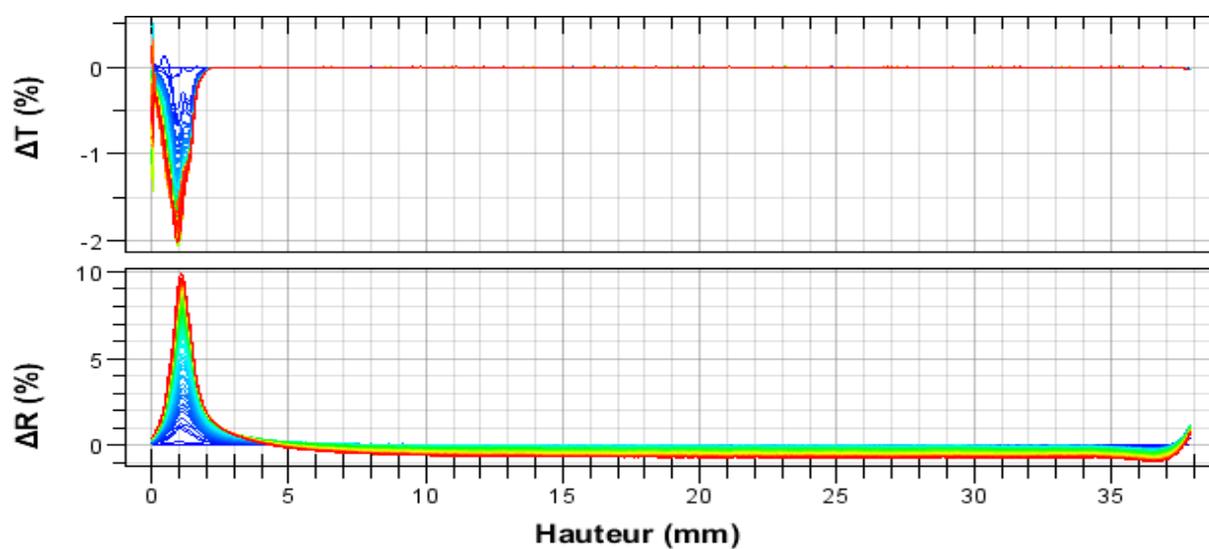
Figure A.12. Non-aqueous TiO$_2$ dispersion in isopropanol.
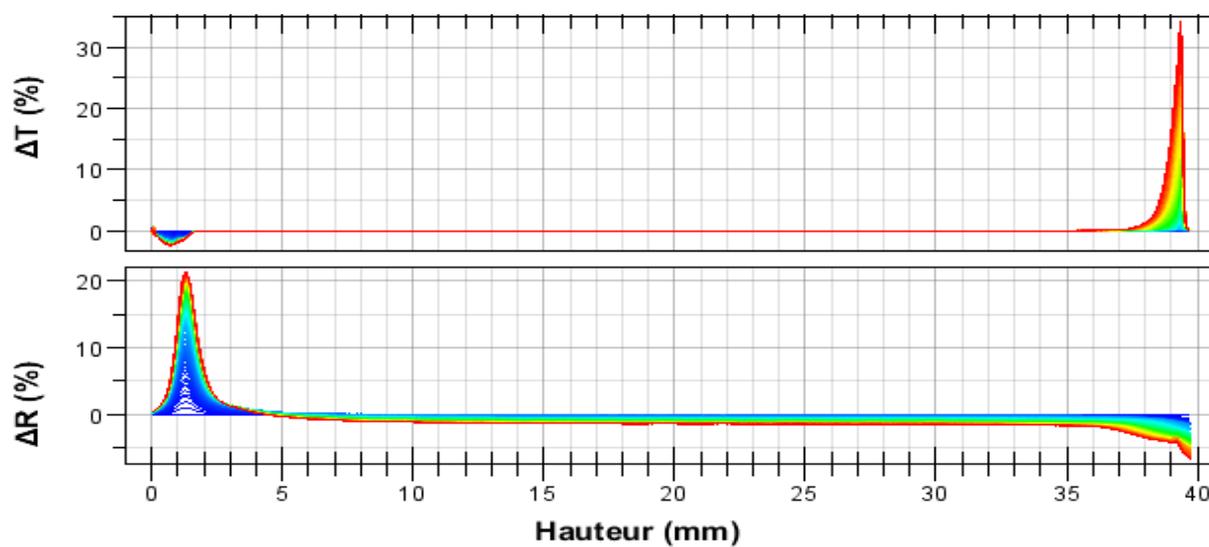
Figure A.13. Non-aqueous TiO$_2$ dispersion in acetonitrile.
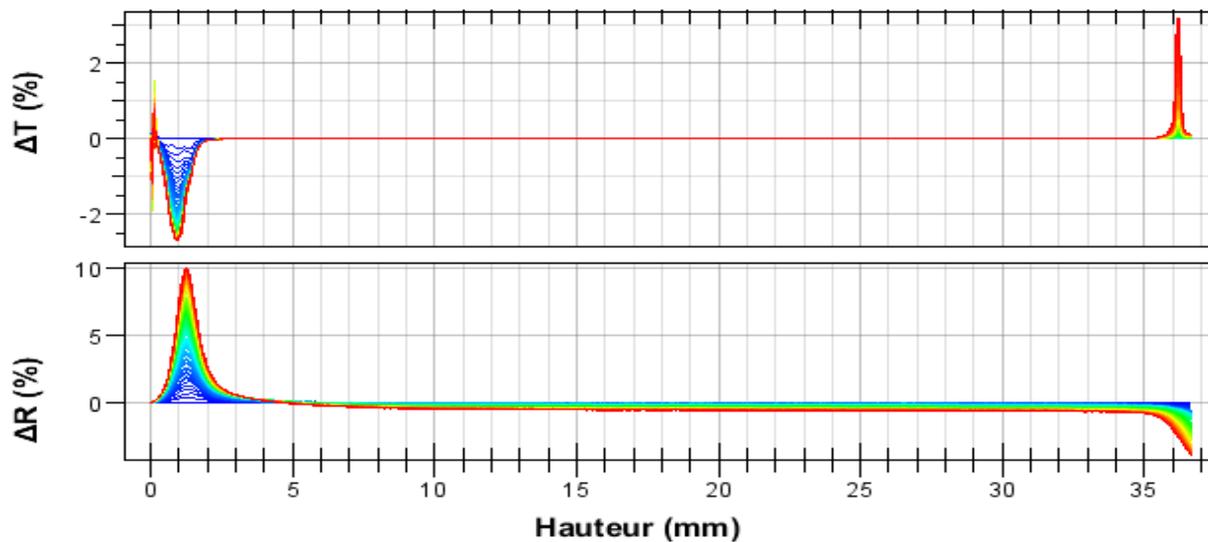
Figure A.14. Non-aqueous TiO$_2$ dispersion in dimethylformamide (DMF).
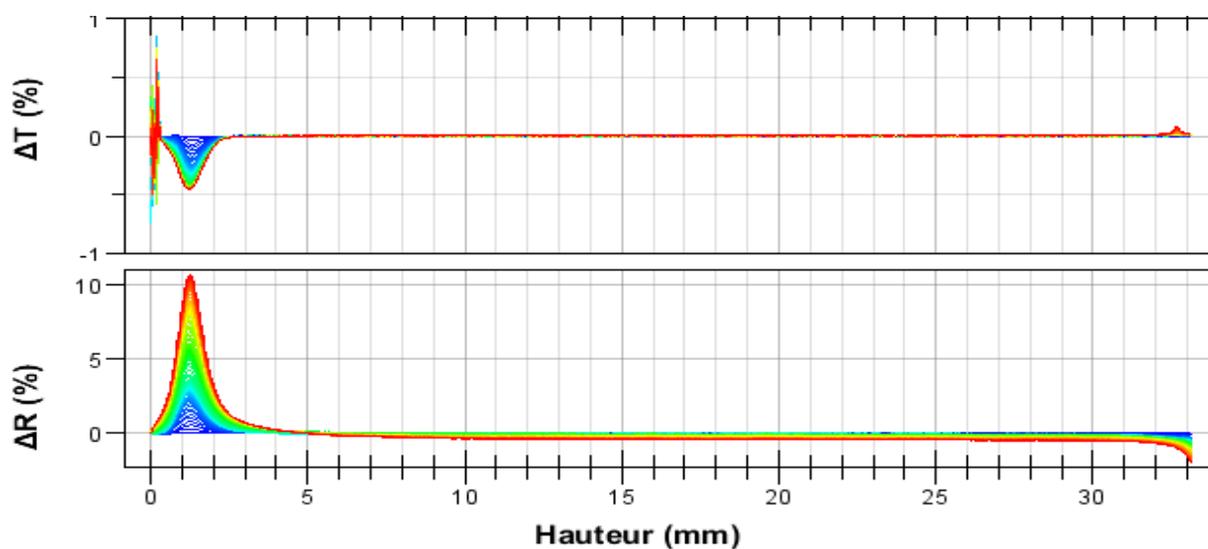
Figure A.15. Non-aqueous TiO$_2$ dispersion in N-methyl-2-pyrrolidone (NMP).
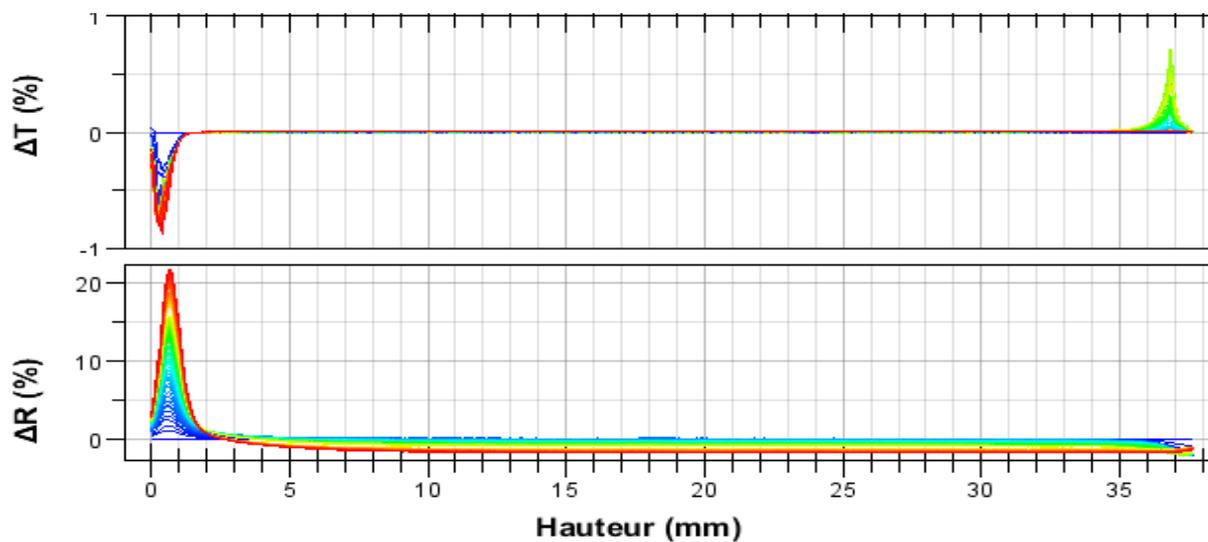
Figure A.16. Non-aqueous TiO$_2$ dispersion in tetrahydrofuran (THF).
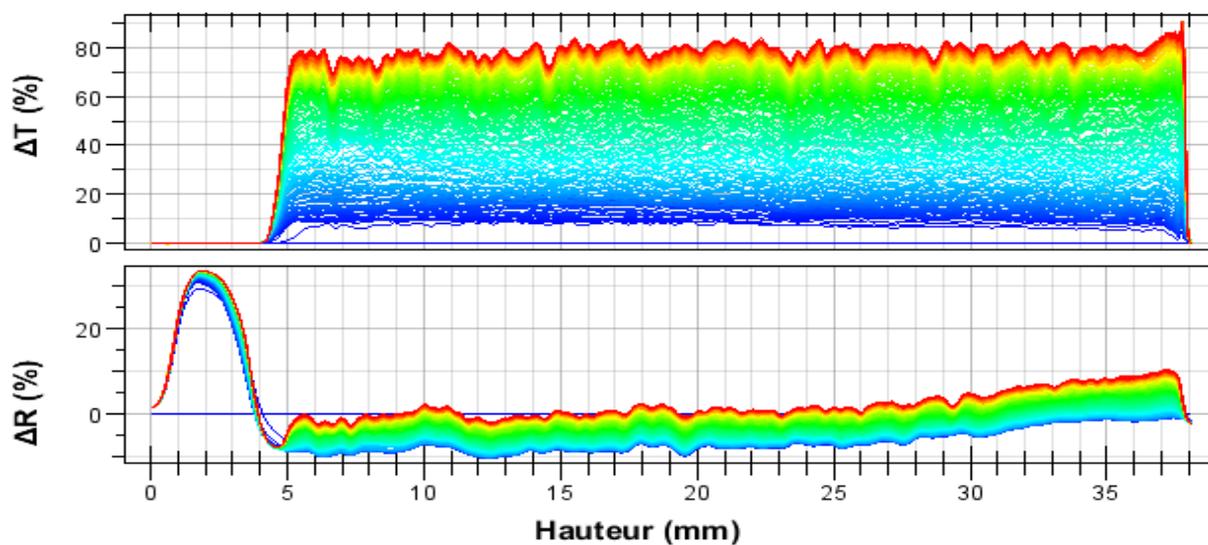
Figure A.17. Non-aqueous TiO$_2$ dispersion in acetone.
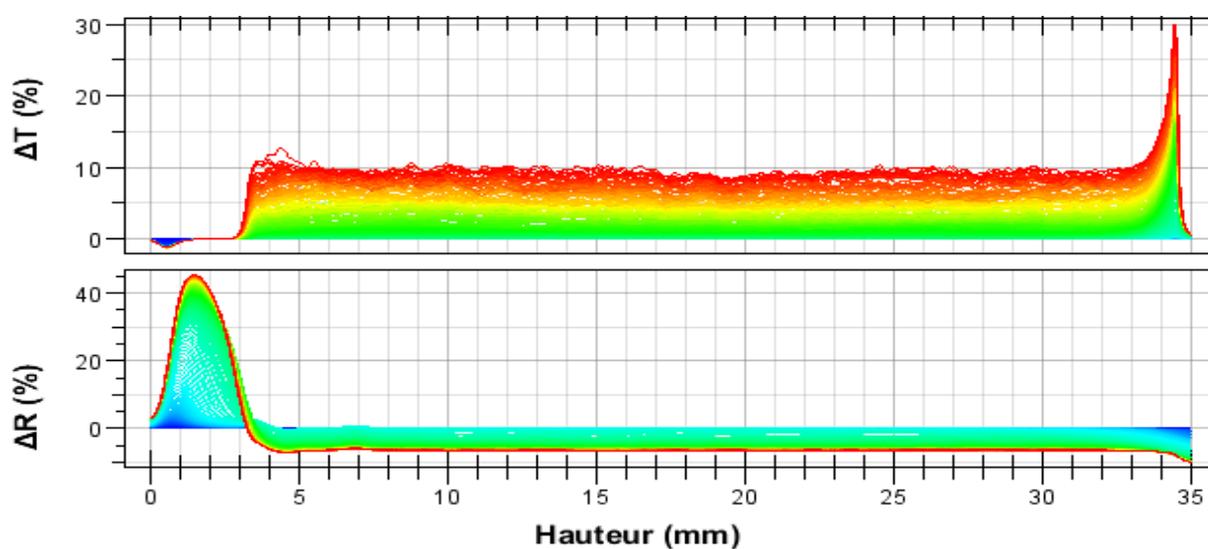
Figure A.18. Non-aqueous TiO$_2$ dispersion in γ-butyrolactone (GBL).
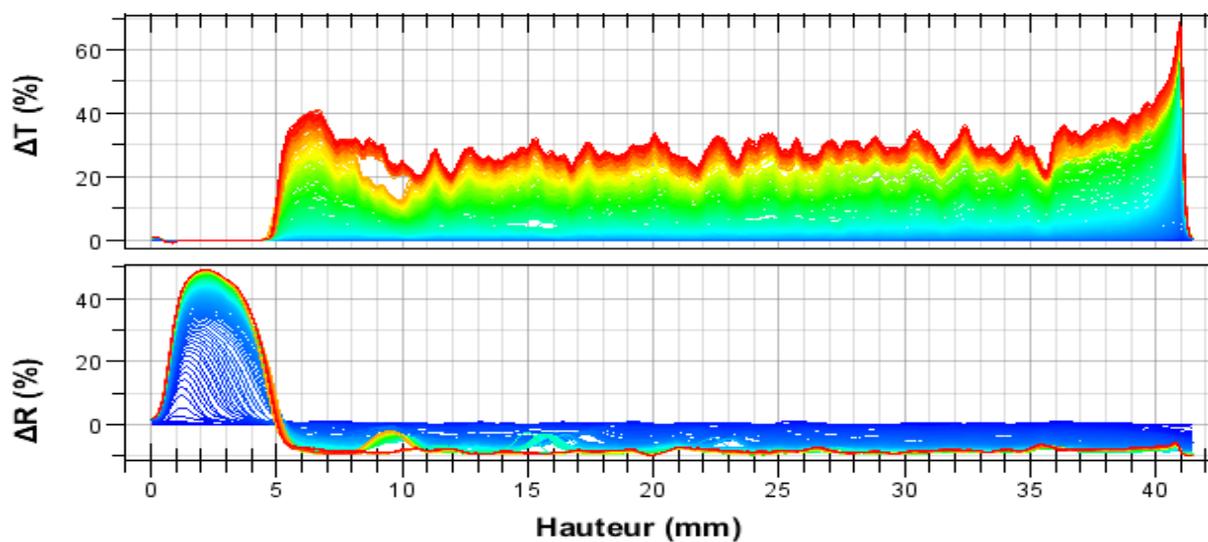
Figure A.19. TiO$_2$ dispersion in ultrapure water without electrolytes.

146
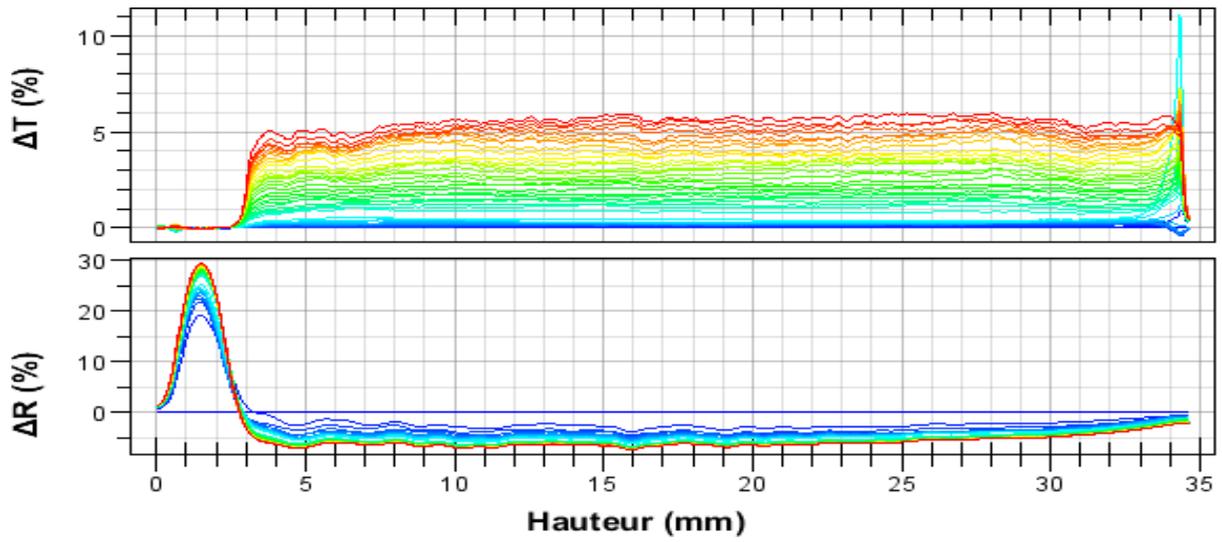
Figure A.20. Non-aqueous TiO$_2$ dispersion in ethyl acetate.
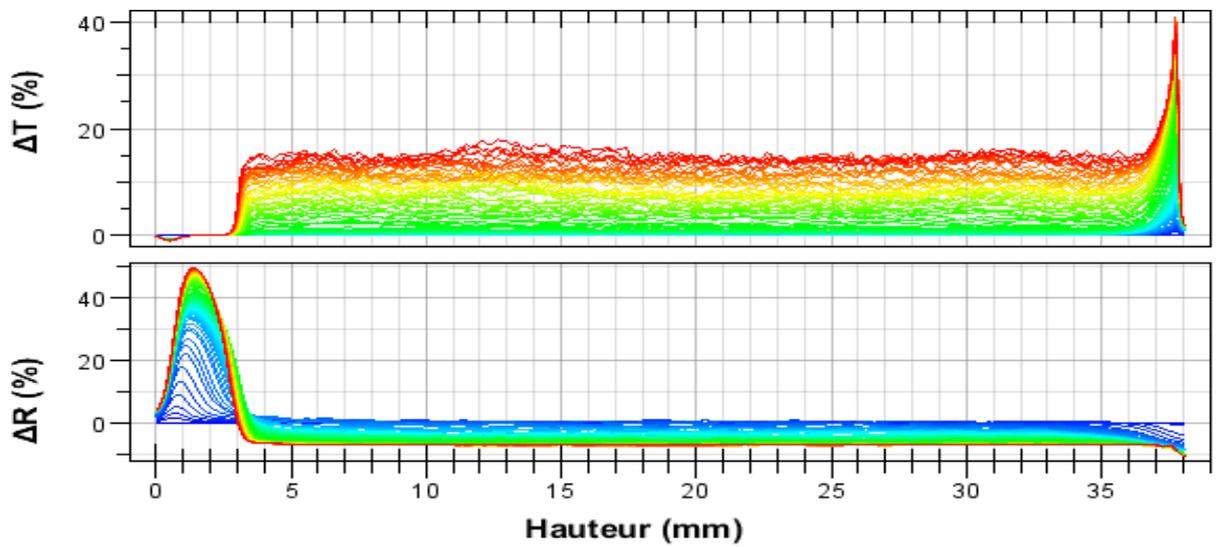
Figure A.21. Non-aqueous TiO$_2$ dispersion in propylene carbonate.
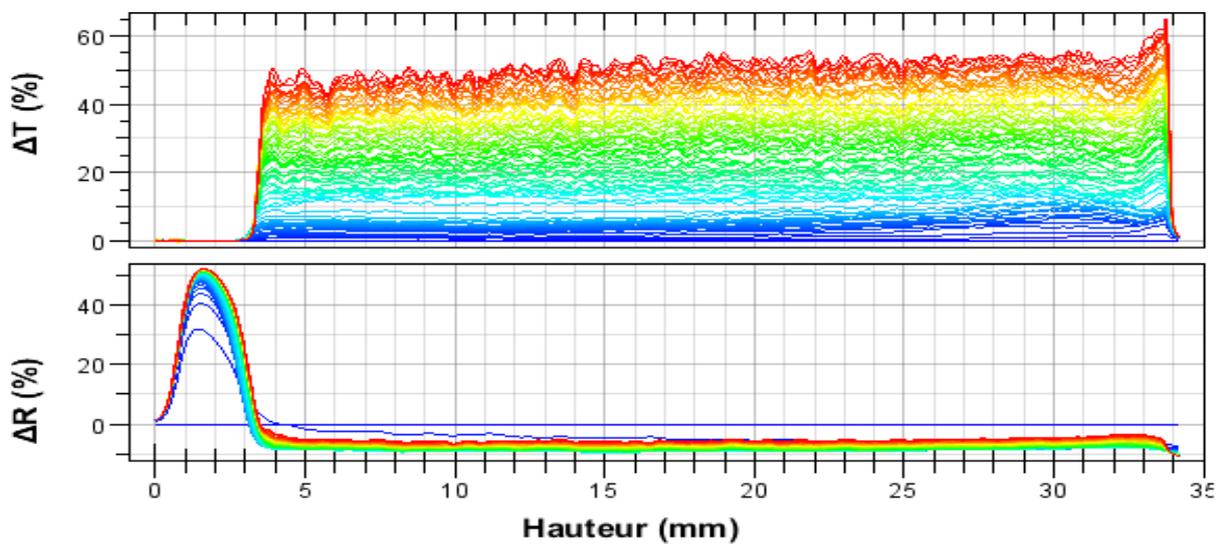
Figure A.22. Non-aqueous TiO$_2$ dispersion in methanol.
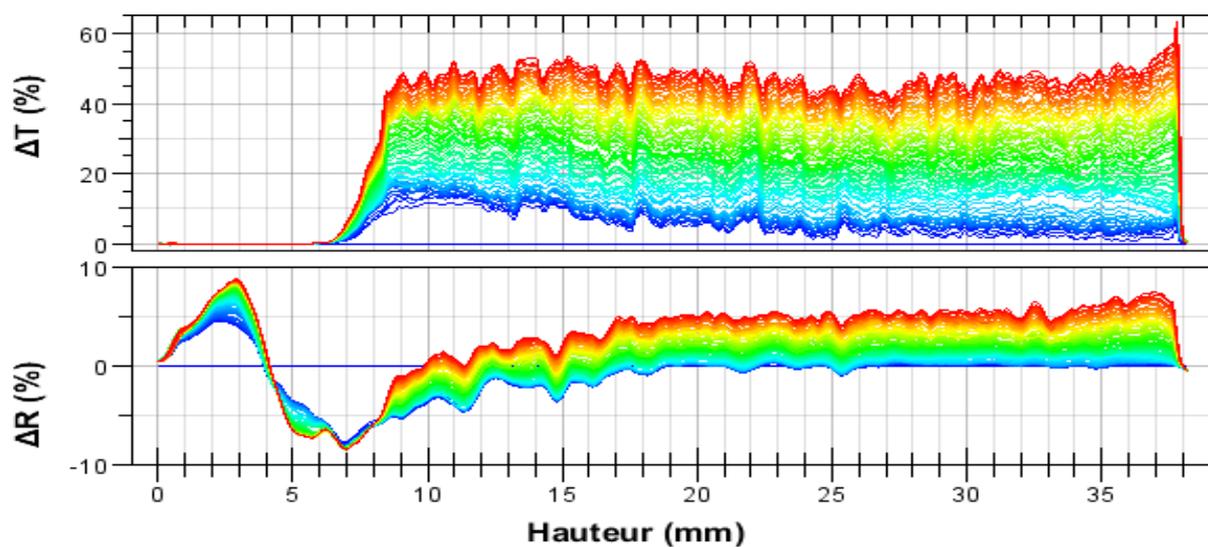
Figure A.23. Non-aqueous TiO$_2$ dispersion in heptane.
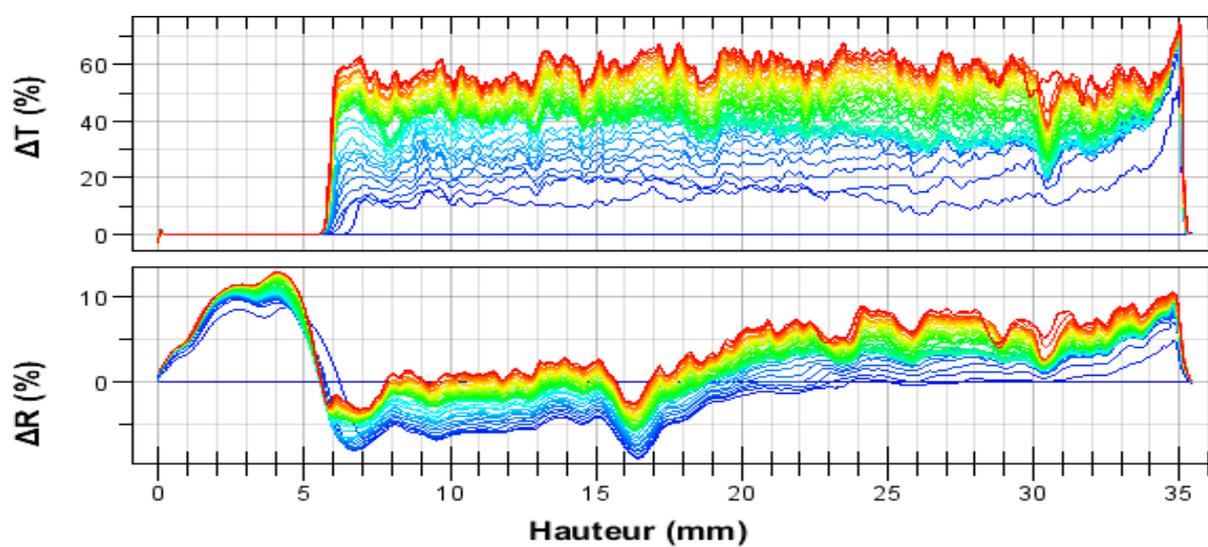
Figure A.24. Non-aqueous TiO$_2$ dispersion in toluene.
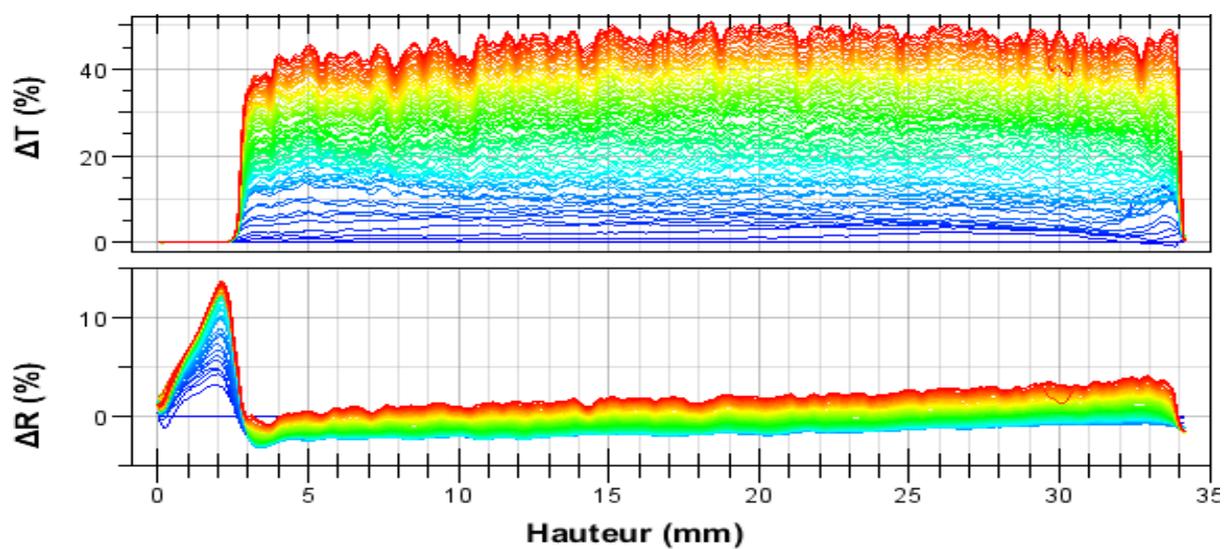
Figure A.25. Non-aqueous TiO$_2$ dispersion in triethylamine.

A5. Turbiscan profiles of TiO$_2$ dispersions in solvents with base or acid

In this section, the variations of transmitted (ΔT) and backscattered (ΔR) signals obtained using a Turbiscan LAB are shown for 1 g/L TiO$_2$ P25 non-aqueous dispersions with solvents to which tetrabutylammonium hydroxide (TBAH) or trifluoroacetic acid (TFA) were added (10^{-3} M). Signals were recorded from 0 (blue curve) to 2 hours (red curve). Samples are presented in the order of increasing measured d$_{Stokes}$ (same order as in Table 1.2).
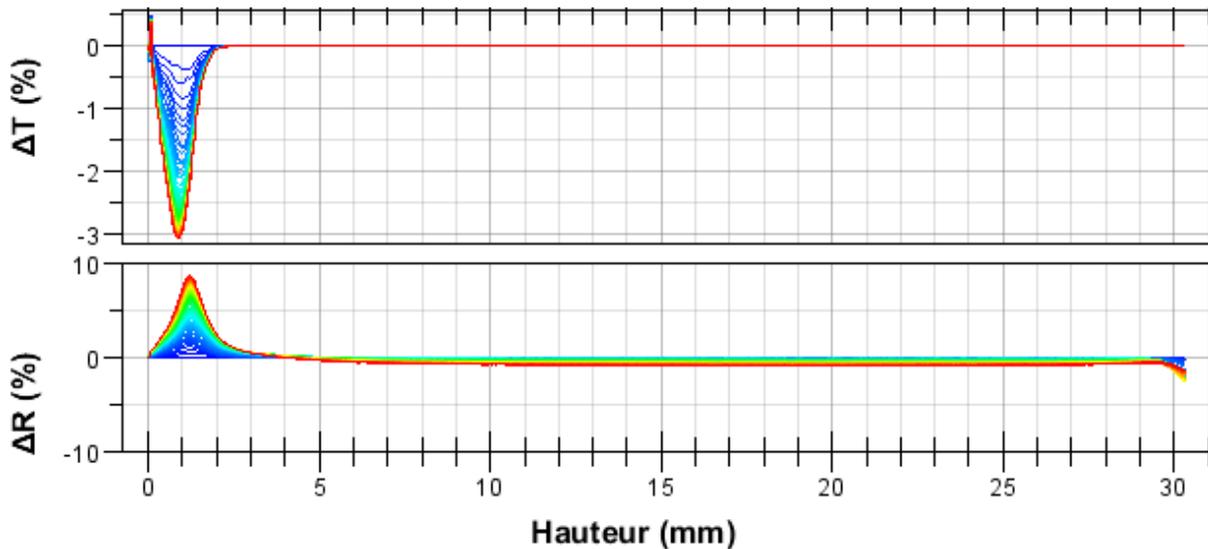
Figure A.26. Non-aqueous TiO$_2$ dispersion in nitromethane (10^{-3} M TFA).
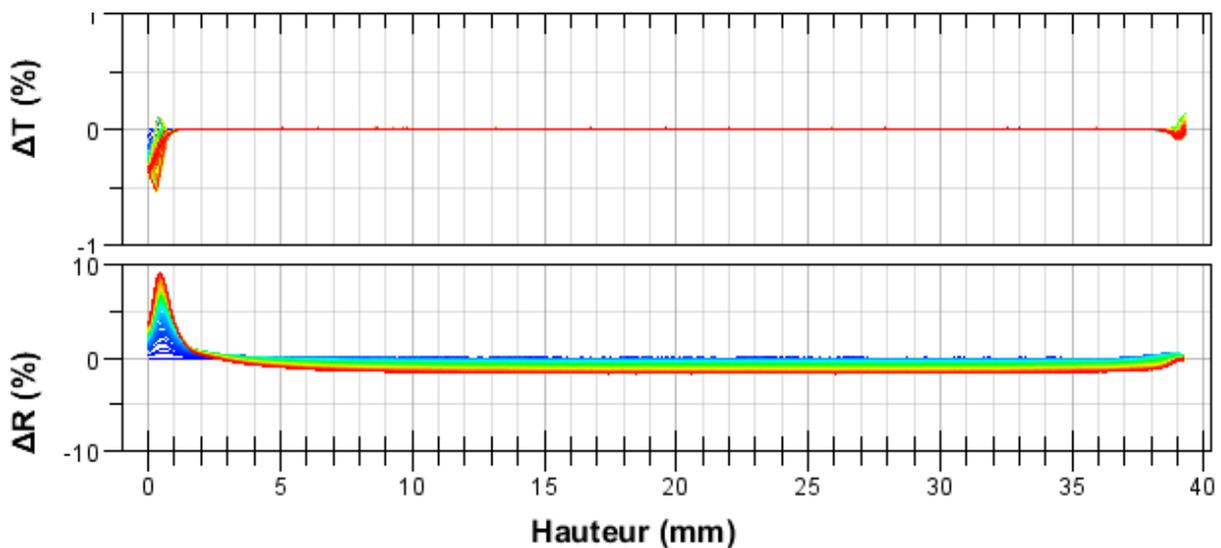
Figure A.27. Non-aqueous TiO$_2$ dispersion in acetonitrile (10^{-3} M TFA).
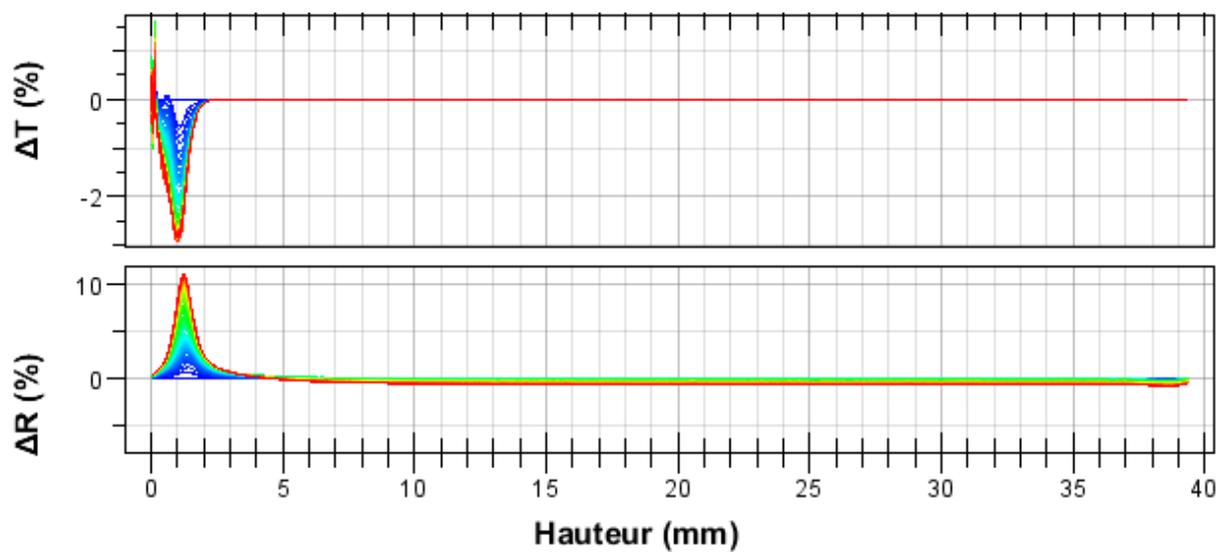
Figure A.28. Non-aqueous TiO$_2$ dispersion in isopropanol (10^{-3} M TFA).
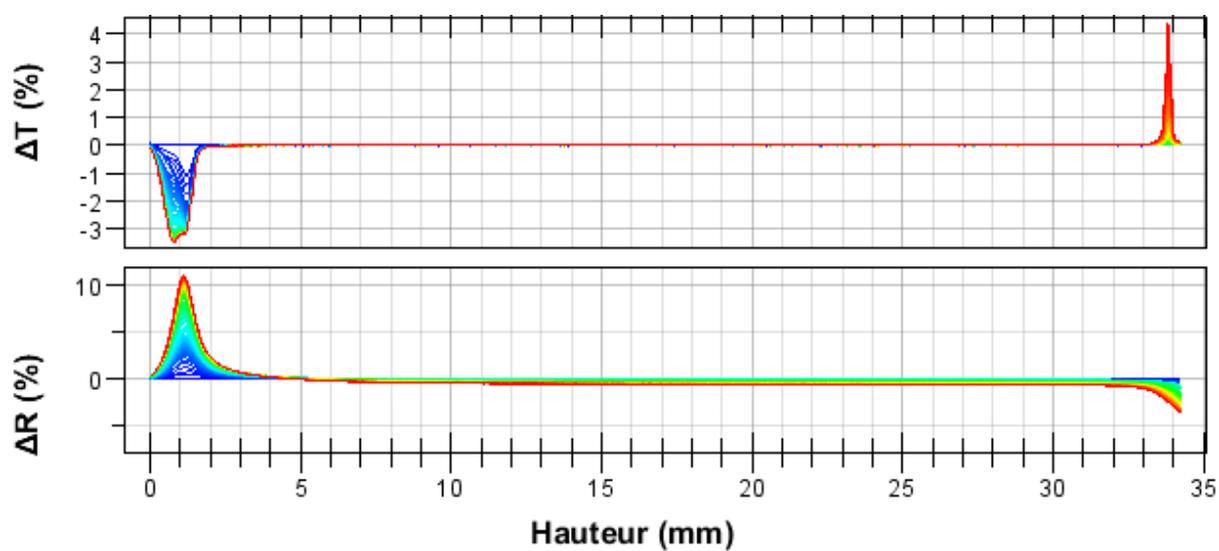
Figure A.29. Non-aqueous TiO$_2$ dispersion in dimethylformamide (DMF + 10^{-3} M TFA).
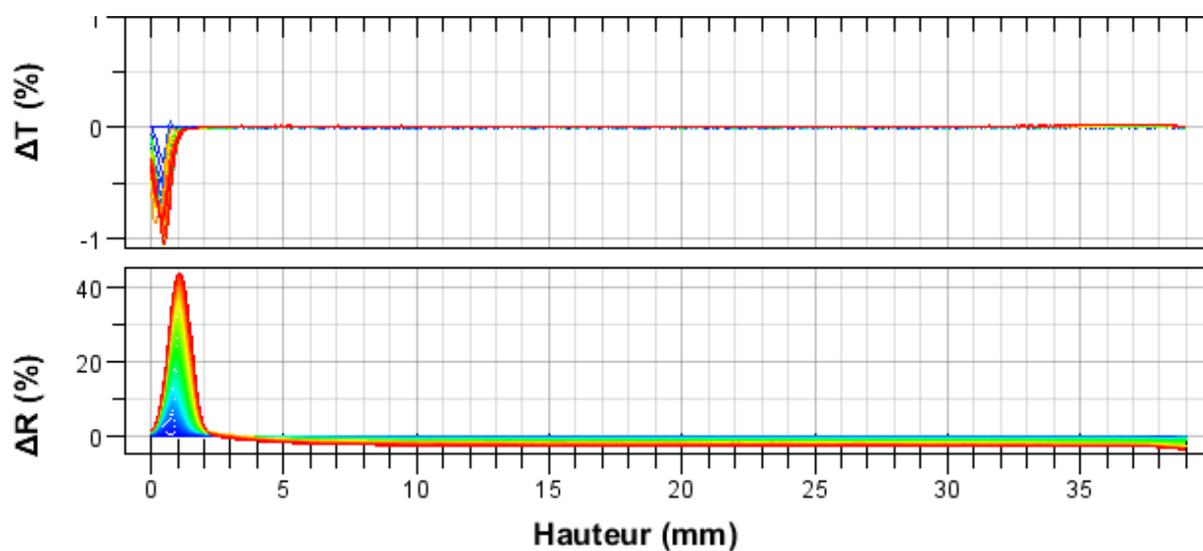
Figure A.30. Non-aqueous TiO$_2$ dispersion in ethanol (10^{-3} M TBAH).
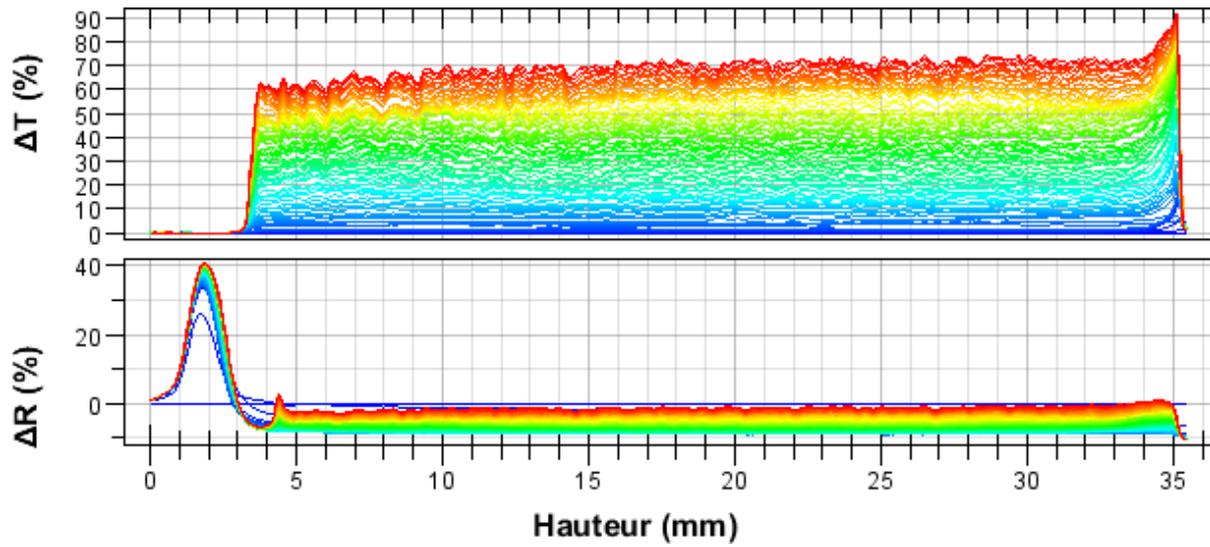
Figure A.31. Non-aqueous TiO$_2$ dispersion in acetone (10^{-3} M TFA).
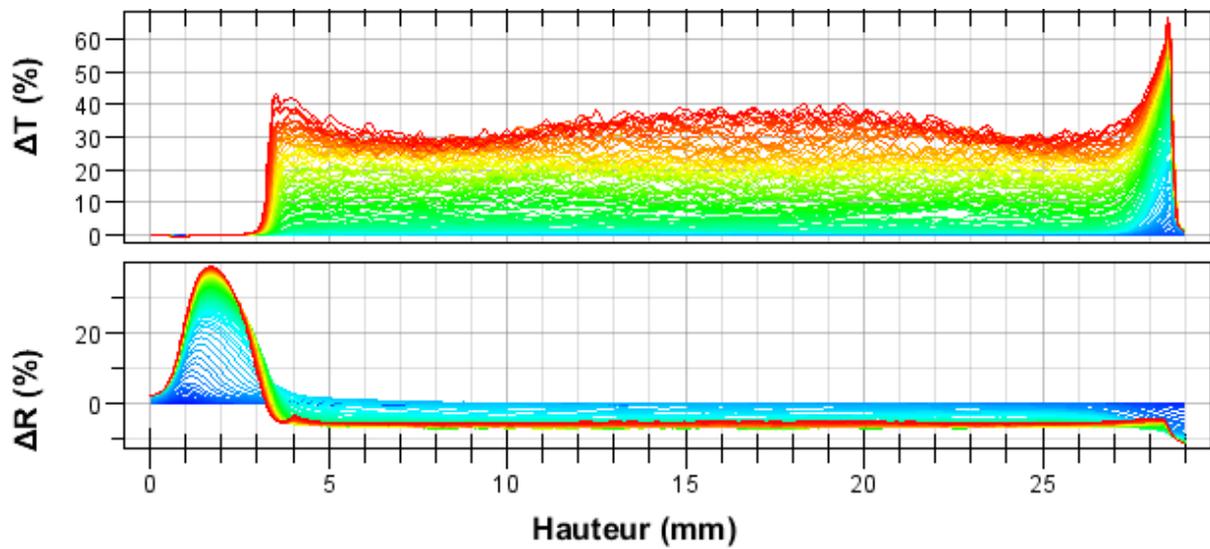
Figure A.32. Non-aqueous TiO$_2$ dispersion in propylene carbonate (10^{-3} M TFA).
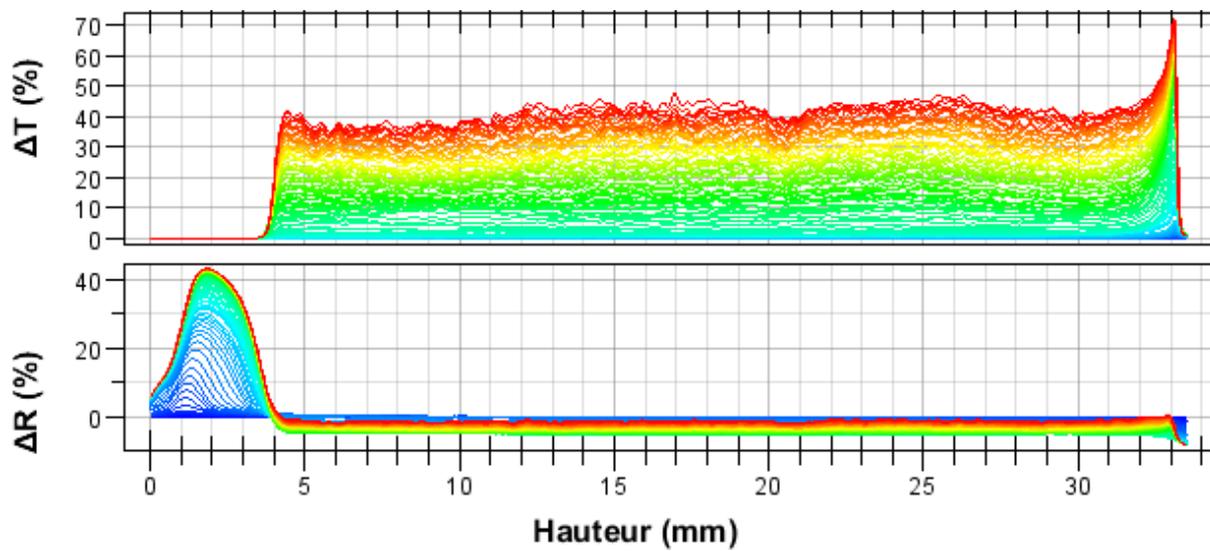
Figure A.33. Non-aqueous TiO$_2$ dispersion in pyridine (10^{-3} M TBAH).

A6. EACN determination with C_6E_4 surfactant
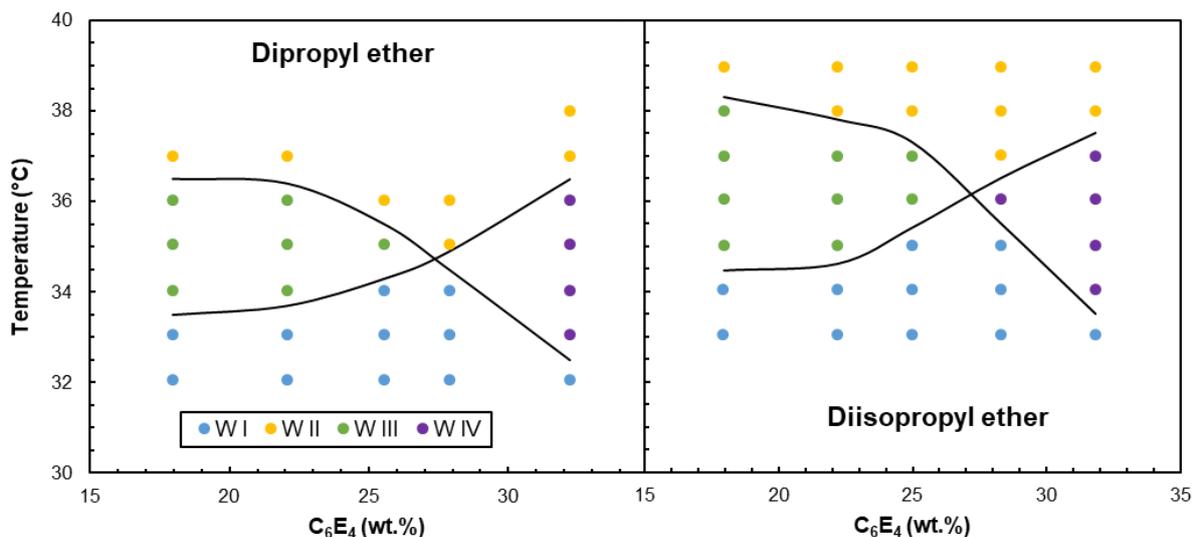
Figure A.34.Partial experimental fish plot of C_6E_4/dipropyl ether/Water-T (left) and C_6E_4/diisopropyl ether/Water-T (right) systems at Water/Oil ratio equal to 1 (w/w) used to determine EACN values.

The temperatures of the fish-tail points are 34.7 °C for dipropyl ether and 36.1 °C for diisopropyl ether. Those temperatures are reported to the calibration straight line obtained with C_6E_4/n-Alkanes/Water-T systems shown in Figure A.35. Resulting EACN values are 0.4 for dipropyl ether and 0.6 for diisopropyl ether.
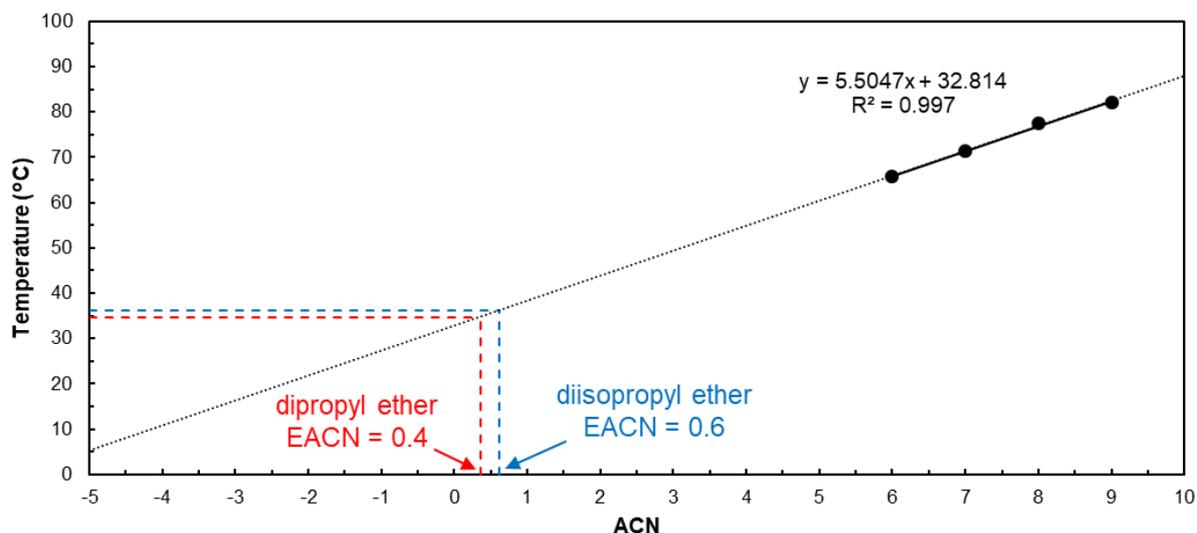
Figure A.35. Calibration straight line obtained with C_6E_4/n-Alkanes/Water-T systems at Water/Oil ratio equal to 1 (w/w) used to determine EACN values.[146]

A7. Oil homologous series data and EACN estimations

Table A2. Names, SMILES notations, three first σ-moments (different from zero) calculated with COSMO-RS, number of carbon atoms, experimental EACN and average EACN values determined from the Fish-tail-temperature T* reported in the literature for ternary systems C_iE_j/Oil/Water and estimated EACN values with the GM-5N and NN-6N models for the 56 molecules of the homologous series.

Entry	Compound	SMILES code	$M_0^{X\,[a]}$	$M_2^{X\,[b]}$	$M_3^{X\,[c]}$	N_C	EACN$_{exp}$	Ref.	EACN$_{GM}^{[d]}$	EACN$_{NN}^{[e]}$
1	hex-1-ene	CCCCC=C	154.47	17.08	3.03	6	-	-	1.7	1.0
2	hept-1-ene	CCCCCC=C	174.54	17.93	3.06	7	-	-	2.8	2.6
3	oct-1-ene	CCCCCCC=C	192.38	18.77	3.59	8	3.9	133	3.6	3.7
4	non-1-ene	CCCCCCCC=C	213.71	19.66	3.2	9	-	-	4.8	5.0
5	dec-1-ene	CCCCCCCCC=C	234.26	20.46	3.17	10	5.5	133	5.6	6.2
6	undec-1-ene	CCCCCCCCCC=C	253.11	21.06	2.99	11	-	-	6.8	7.3
7	dodec-1-ene	CCCCCCCCCCC=C	274.46	22.29	3.31	12	8.1	133	7.9	8.3
8	tridec-1-ene	CCCCCCCCCCCC=C	293.79	23.18	3.23	13	-	-	9.0	9.5
9	tetradec-1-ene	CCCCCCCCCCCCC=C	314.61	24.11	3.38	14	-	-	10.1	10.8
10	pentadec-1-ene	CCCCCCCCCCCCCC=C	334.71	25.24	3.47	15	-	-	11.2	12.0
11	hexadec-1-ene	CCCCCCCCCCCCCCC=C	355.28	25.93	3.5	16	-	-	12.3	13.0
12	heptadec-1-ene	CCCCCCCCCCCCCCCC=C	374.83	27.05	3.56	17	-	-	13.3	13.7
13	octadec-1-ene	CCCCCCCCCCCCCCCCC=C	394.66	27.89	3.6	18	14	133	14.4	14.4
14	1-chlorohexane	CC[C:1]CCCCl	178.83	26.47	4.0	6	-	-	–0.6	1.2
15	1-chloroheptane	CCCCCCCCl	198.14	27.2	3.78	7	-	-	0.6	1.8
16	1-chlorooctane	CCCCCCCCCl	219.07	28.31	4.13	8	-	-	1.8	2.4
17	1-chlorononane	CCCCCCCCCCl	238.09	29.01	3.89	9	-	-	2.5	3.0
18	1-chlorodecane	CCCCCCCCCCCl	257.74	29.94	3.76	10	3.5	133,186	3.6	3.7
19	1-chloroundecane	CCCCCCCCCCCCl	278.35	30.89	3.8	11	-	-	4.5	4.6
20	1-chlorododecane	CCCCCCCCCCCCCl	297.96	31.52	3.82	12	5.6	133,186	5.7	5.7
21	1-chlorotridecane	CCCCCCCCCCCCCCl	318.51	32.66	3.9	13	-	-	6.7	6.8
22	1-chlorotetradecane	CCCCCCCCCCCCCCCl	338.17	33.58	3.92	14	8	133,186	7.8	7.9
23	1-chloropentadecane	CCCCCCCCCCCCCCCCl	358.14	34.49	3.93	15	-	-	8.9	8.7
24	1-chlorohexadecane	CCCCCCCCCCCCCCCCCl	378.22	35.15	3.98	16	9.8	133,186	9.8	9.6
25	1-chloroheptadecane	CCCCCCCCCCCCCCCCCCl	398.86	36.28	4.05	17	-	-	10.9	10.1
26	1-chlorooctadecane	CCCCCCCCCCCCCCCCCCCl	418.41	37.03	4.09	18	-	-	11.7	11.1
27	hex-1-yne	C[C:1]CCC#C	150.44	31.68	1.42	6	-	-	–3.9	–2.3
28	hept-1-yne	CC[C:1]CCC#C	170.69	32.6	1.53	7	-	-	–2.5	–1.9
29	oct-1-yne	CCCCCC#C	188.77	33.22	1.53	8	–1.8	133	–1.3	–1.8
30	non-1-yne	CCC[C:1]CCCC#C	209.6	33.93	1.37	9	-	-	–1.0	–1.0
31	dec-1-yne	CCCCCCCCC#C	230.44	35.32	1.68	10	0.1	133	0.2	0.1
32	undec-1-yne	CCCCCCCCCC#C	249.55	35.61	1.43	11	-	-	1.5	1.4
33	dodec-1-yne	CCCCCCCCCCC#C	265.87	36.64	1.52	12	2.0	133	2.0	1.8
34	tridec-1-yne	CCCCCCCCCCCC#C	290.57	38.01	1.86	13	-	-	3.2	3.2
35	tetradec-1-yne	CCCCCCCCCCCCC#C	310.75	38.72	1.96	14	3.9	133	3.8	3.8
36	pentadec-1-yne	CCCCCCCCCCCCCC#C	330.13	39.49	1.74	15	-	-	4.9	4.4
37	hexadec-1-yne	CCCCCCCCCCCCCCC#C	350.9	40.53	2.02	16	-	-	5.6	4.8
38	heptadec-1-yne	CCCCCCCCCCCCCCCC#C	371.5	41.48	2.09	17	-	-	6.6	5.1
39	octadec-1-yne	CCCCCCCCCCCCCCCCC#C	391.04	42.25	2.09	18	-	-	7.4	5.4
40	hexan-2-one	CC[C:1]CC(C)=O	162.69	47.19	35.22	6	-	-	–4.9	–4.5
41	heptan-2-one	CCCCCC(C)=O	182.02	47.29	34.83	7	-	-	–3.8	–3.7
42	octan-2-one	CCCCCC(C)=O	201.89	48.96	35.68	8	–3.4	134	–3.4	–3.2
43	nonan-2-one	CCCCCCC(C)=O	221.83	48.89	34.74	9	-	-	–2.2	–2.4
44	decan-2-one	CCCCCCCC(C)=O	242.35	50.66	35.46	10	–2.1	134	–2.2	–1.8
45	undecan-2-one	CCCCCCCCC(C)=O	-	-	-	11	–1.3	134	–1.0	–
46	dodecan-2-one	CCCCCCCCCC(C)=O	282.29	52.17	35.35	12	–0.6	134	–0.8	–0.1

153

#	name	SMILES								
47	tridecan-2-one	CCCCCCCCCCCC(C)=O	302.53	52.44	34.62	13	-	-	0.4	1.9
48	tetradecan-2-one	CCCCCCCCCCCCC(C)=O	322.06	54.01	35.42	14	-	-	0.5	3.3
49	pentadecan-2-one	CCCCCCCCCCCCCC(C)=O	342.6	54.37	34.66	15	-	-	1.6	5.1
50	hexadecan-2-one	CCCCCCCCCCCCCCC(C)=O	362.4	55.79	35.63	16	-	-	1.9	6.3
51	heptadecan-2-one	CCCCCCCCCCCCCCCC(C)=O	383.45	57.07	35.93	17	-	-	3.0	7.5
52	octadecan-2-one	CCCCCCCCCCCCCCCCC(C)=O	402.65	57.78	35.77	18	-	-	3.2	8.5
53	cyclohexane	C1CCCCC1	135.76	5.88	-0.12	6	2.1	130,159,184,185	2.0	2.4
54	methylcyclohexane	CC1CCCCC1	-	-	-	7	3.2	130	2.9	–
55	ethylcyclohexane	CCC1CCCCC1	171.53	7.91	-0.12	8	4.2	127,130	4.4	4.6
56	Propylcyclohexane	CCCC1CCCCC1	-	-	-	9	5.9	127,130	5.7	–
57	butylcyclohexane	CCCCC1CCCCC1	210.84	9.64	-0.04	10	7.3	127,130	7.1	7.4
58	pentylcyclohexane	CCCCCC1CCCCC1	230.87	10.52	0.01	11	-	-	8.2	8.7
59	hexylcyclohexane	CCCCCCC1CCCCC1	251.83	11.39	0.03	12	-	-	9.3	9.9
60	heptylcyclohexane	CCCCCCCC1CCCCC1	267.31	12.16	-0.03	13	-	-	10.4	10.8
61	octylcyclohexane	CCCCCCCCC1CCCCC1	291.22	13.25	0.13	14	-	-	12.1	12.0
62	nonylcyclohexane	CCCCCCCCCC1CCCCC1	311.7	14.12	0.18	15	-	-	13.2	13.3
63	decylcyclohexane	CCCCCCCCCCC1CCCCC1	332.16	14.95	0.24	16	14.4	127	14.5	14.6
64	undecylcyclohexane	CCCCCCCCCCCC1CCCCC1	352.12	15.76	0.3	17	-	-	15.6	15.9
65	dodecylcyclohexane	CCCCCCCCCCCCC1CCCCC1	371.81	16.81	0.29	18	17.5	127	17.1	17.1
66	1-propoxypropane	CCCOCCC	171.36	26.39	20.59	6	0.4	[f]	0.7	0.7
67	1-propoxybutane	CCCOCCCC	191.74	28.16	21.83	7	-	-	1.7	1.5
68	1-butoxybutane	CCCCOCCCC	211.38	28.47	21.23	8	3.0	127,134,164	2.6	2.7
69	1-butoxypentane	CCCCOCCCCC	231.24	29.01	20.98	9	-	-	3.5	3.7
70	1-pentoxypentane	CCCCCOCCCCC	251.19	29.57	20.44	10	4.2	134	4.4	4.5
71	1-pentoxyhexane	CCCCCOCCCCCC	271.23	30.7	21.06	11	-	-	5.3	5.2
72	1-hexyloxyhexane	CCCCCCOCCCCCC	291.31	31.96	21.34	12	6.2	134	6.2	6.0
73	1-hexyloxyheptane	CCCCCCOCCCCCCC	311.43	32.38	20.97	13	-	-	7.1	7.0
74	1-heptyloxyheptane	CCCCCCCOCCCCCCC	331.49	32.86	20.36	14	8.0	134	8.0	8.1
75	1-heptyloxyoctane	CCCCCCCOCCCCCCCC	351.57	33.84	20.58	15	-	-	8.9	9.1
76	1-octyloxyoctane	CCCCCCCCOCCCCCCCC	371.28	35.31	21.42	16	10.3	134	9.8	10.1
77	1-octyloxynonane	CCCCCCCCOCCCCCCCCC	391.55	35.95	21.17	17	-	-	10.6	11.3
78	1-nonyloxynonane	CCCCCCCCCOCCCCCCCCC	411.72	36.47	20.58	18	-	-	11.4	12.7
79	ethyl butyrate	CC[C:1]C(OCC)=O	173.82	47.52	31.71	6	-	-	–3.0	–4.7
80	ethyl valerate	CCC[C:1]C(OCC)=O	194.22	48.26	32.23	7	-	-	–1.9	–3.9
81	ethyl caproate	CCCC[C:1]C(OCC)=O	213.46	49.02	31.67	8	-	-	–0.7	–3.0
82	ethyl oenanthate	CCCCC[C:1]C(OCC)=O	234.15	50.07	32.26	9	-	-	0.4	–1.9
83	ethyl caprylate	CCCCCCCC(OCC)=O	253.0	49.88	30.92	10	-	-	0.5	–0.4
84	ethyl pelargonate	CCCCCCCCC(OCC)=O	274.18	51.89	32.1	11	-	-	1.0	0.8
85	ethyl caprate	CCCCCCCCCC(OCC)=O	294.96	53.81	33.56	12	2.1	145,147	2.1	1.9
86	ethyl undecylate	CCCCCCCCCCC(OCC)=O	314.51	54.14	33.04	13	-	-	2.5	3.3
87	ethyl laurate	CCCCCCCCCCCC(OCC)=O	333.62	54.56	32.64	14	3.8	147	3.6	4.5
88	ethyl tridecanoate	CCCCCCCCCCCCC(OCC)=O	354.72	55.83	33.02	15	-	-	4.3	5.5
89	ethyl myristate	CCCCCCCCCCCCCC(OCC)=O	375.02	56.48	32.88	16	5.2	147	5.3	6.5
90	ethyl pentadecanoate	CCCCCCCCCCCCCCC(OCC)=O	395.34	57.79	33.24	17	-	-	6.0	7.2
91	ethyl palmitate	CCCCCCCCCCCCCCCC(OCC)=O	415.24	59.07	33.25	18	6.8	147	6.9	7.7
92	capronitrile	CCCCCC#N	163.84	47.37	22.7	6	-	-	–3.1	–3.3
93	enanthonitrile	CCCC[C:1]CC#N	183.34	48.24	22.69	7	-	-	–1.9	–2.4
94	caprylonitrile	CCCCCCCC#N	204.22	49.16	22.76	8	–1.7	134	–1.9	–1.7
95	pelargononitrile	CCCCCCCCC#N	222.16	49.71	22.71	9	-	-	–1.8	–1.0
96	caprinitrile	CCCCCCCCCC#N	243.52	50.82	22.88	10	–0.5	134	–0.6	–0.4

#	Name	SMILES	M_0^X	M_2^X	M_3^X	N_C		Ref		
97	undecanenitrile	CCCCCCCCCCC#N	263.88	51.76	23.03	11	-	-	−0.5	0.1
98	lauronitrile	CCCCCCCCCCCC#N	283.99	52.55	23.07	12	0.4	134	0.7	0.5
99	tridecanenitrile	CCCCCCCCCCCCC#N	304.6	53.71	23.08	13	-	-	0.9	1.2
100	myristonitrile	CCCCCCCCCCCCCC#N	324.22	54.56	23.13	14	-	-	2.0	1.9
101	pentadecanenitrile	CCCCCCCCCCCCCCC#N	342.98	55.14	22.97	15	-	-	2.3	2.7
102	palmitonitrile	CCCCCCCCCCCCCCCC#N	364.4	56.25	23.18	16	-	-	3.3	3.4
103	heptadecanenitrile	CCCCCCCCCCCCCCCCC#N	384.41	57.18	23.27	17	-	-	3.6	3.9
104	stearonitrile	CCCCCCCCCCCCCCCCCC#N	404.51	57.93	23.31	18	-	-	4.6	4.4
105	benzene	C1=CC=CC=C1	122.42	22.98	-1.27	6	-	-	−4.5	−3.0
106	toluene	CC1=CC=CC=C1	141.81	23.64	0.06	7	-	-	−3.5	−2.7
107	ethylbenzene	CCC1=CC=CC=C1	160.64	24.31	0.44	8	-	-	−2.2	−1.9
108	propylbenzene	CCCC1=CC=CC=C1	180.59	25.28	0.67	9	-	-	−1.0	−1.3
109	butylbenzene	CCCCC1=CC=CC=C1	200.65	26.12	0.72	10	0.4	133,185	0.2	−0.3
110	pentylbenzene	CCCCCC1=CC=CC=C1	220.24	27.00	0.82	11	-	-	1.4	0.7
111	hexylbenzene	CCCCCCC1=CC=CC=C1	237.83	27.56	0.78	12	-	-	2.5	1.8
112	heptylbenzene	CCCCCCCC1=CC=CC=C1	258.5	28.41	0.80	13	-	-	2.8	3.1
113	octylbenzene	CCCCCCCCC1=CC=CC=C1	278.92	29.37	0.88	14	4.0	127,130	3.8	4.4
114	nonylbenzene	CCCCCCCCCC1=CC=CC=C1	299.80	30.47	0.96	15	-	-	5.0	5.6
115	decylbenzene	CCCCCCCCCCC1=CC=CC=C1	320.4	31.55	1.1	16	6.0	127,130	5.9	6.4
116	undecylbenzene	CCCCCCCCCCCC1=CC=CC=C1	339.84	32.10	1.04	17	-	-	7.0	7.3
117	dodecylbenzene	CCCCCCCCCCCCC1=CC=CC=C1	359.87	33.08	1.12	18	7.8	127,159,185	8.0	7.8

[a] M_0^X, expressed in Å^2, is equal to the whole surface area of molecule X; [b] M_2^X, expressed in $\text{e}^2.\text{Å}^{-2}$, reflects the polarity of molecule X; [c] M_3^X, expressed in $\text{e}^3.\text{Å}^{-4}$, reflects the electrostatic asymmetry of molecule X; [d]Estimated EACN values with the GM-5N model having the best VLOO score average (out of ten) for the 117 molecules of the homologous series; [e]Estimated EACN values with the NN-6N model having the best VLOO score average (out of ten) for the 117 molecules of the homologous series; [f]This work. For the 117 molecules the given data can be obtained in running the Docker image "espcigm/eacn:demo", see section "graph machine and neural networks results with Docker".
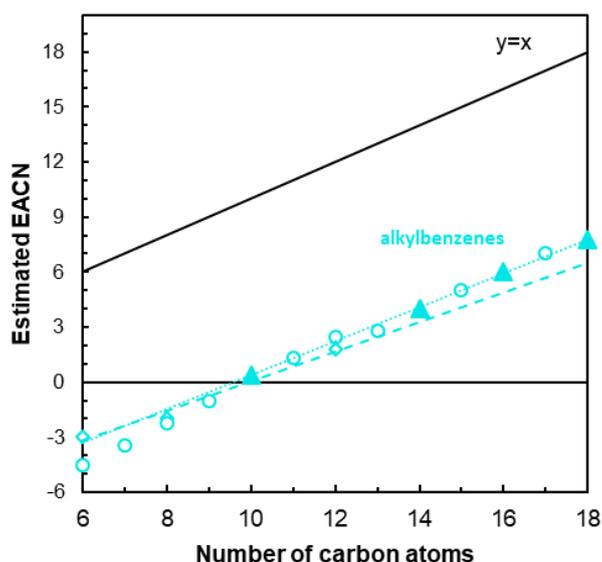
Figure A.36. Evolution of experimental and estimated EACN with increasing number of carbon atoms (N_C) for the alkylbenzene homologous series of molecules. For clarity, half of the predicted values are displayed. The dotted and dashed lines indicate the experimental and neural network fits. Triangles (▲), diamonds (◇) and circles (○) are respectively markers for experimental, neural network predicted and graph machine predicted values.

A8. Graph Machine and Neural Network demonstrations with docker containers

<u>**Installing Docker for Mac and downloading the demo image**</u>

In the following example the installation of Docker is performed with an intel version of Docker Desktop for Mac. The same operations can be done on a Mac with a M1 ARM-based system. The appropriate link to download the corresponding version is given at the end of this Section.

1) Download the Intel chip Docker application by copying and pasting the following line into a browser URL bar:

```
https://desktop.docker.com/mac/main/amd64/Docker.dmg?utm_source=docker
&utm_medium=webreferral&utm_campaign=docs-driven-download-mac-amd64
```

2) After Docker installation (with administrator privileges), launch it. You can open the Docker preferences to increase the Docker allocated memory (e.g. 40 Go), and select a number of cores for the virtual machine (more or less according to the machine resources).

3) Open a terminal window, paste the following line, and hit return:

```
docker pull espcigm/eacn:demo
```

The image used to create containers is then downloaded.

4) You can ensure that the image is genuine by checking the hash code generated at the end of the download process; it should be:

```
sha256:7a64a246509086047c59c7f4767cf717b50a60d813e5f640de7152bfcbaf2283
```

The same image is used to launch either graph machine or neural network computations. The set-up is now complete. More information on Docker client installation can be obtained from the link below and from the Docker website (docker.com).

http://pubs.acs.org/doi/suppl/10.1021/acs.jcim.0c00083/suppl_file/ci0c00083_si_003.pdf

The link to install the M1 chip version of Docker is the following:

```
https://desktop.docker.com/mac/main/arm64/Docker.dmg?utm_source=docker&
utm_medium=webreferral&utm_campaign=dd-smartbutton&utm_location=module
```

<u>**Installing Docker for Windows and downloading the demo image**</u>

The steps for the installation of the Docker Windows version and the demo image are given below.

1) Download the Windows version of Docker application by copying and pasting the following line into a browser URL bar:

```
https://desktop.docker.com/win/main/amd64/Docker%20Desktop%20Installer.exe
```

2) Install the Docker Desktop version 4.6.1 (or above) with administrator privileges

3) In the Dashboard opened at the beginning of the Docker installation verify that:

      a. the box 'Enable Hyper-V Windows feature' is checked;

      b. the box 'Add shortcut to Desktop' is checked;

      c. uncheck if necessary the box 'Install required Windows components for WSL 2'.

4) When the installation has succeeded open Docker Desktop with the desktop shortcut. You have to accept the Docker service agreement by checking the box 'I accept the terms'.

5) Docker Desktop is starting, you might skip the tutorial. Go to Docker Desktop settings and choose General, then uncheck the box 'Use the WSL 2 based engine' (very important). Click on 'Apply and restart'.

6) After a while, open a Powershell window (or a command prompt window) and to verify that Docker is running type the following command:

```
docker version
```

You can confirm in the lines returned that the Docker Desktop version is 4.6.1 (or above).

7) Open a terminal window, paste the following line, and hit return:

```
docker pull espcigm/eacn:demo
```

The image used to create containers is then downloaded.

8) You can ensure that the image is genuine by checking the hash code generated at the end of the download process; it should be:

```
sha256: 8082d2eadbf4629a1616f931ae8cf647edc5fbb357fcd566955622945ec471b9
```

The same image is used to launch either graph machine or neural network computations. The set-up is now complete.

Notes

- If at step 4 Docker does not start, it is because Hyper-V is probably not active on your system. You need to activate Hyper-V by typing in a powershell windows the following command (as admin):

```
Enable-WindowsOptionalFeature -Online -FeatureName Microsoft-Hyper-V -All
```

then restart your machine.

- Docker Desktop can be installed for a standard user. The user has to be a member of the docker-users group. This can be done with the Windows administration tools.

Loading and launching the Docker image

To open a container that will launch the default graph machine computations for the molecules of the *test set*, open a terminal window (or start a PowerShell session), and type the following line of text below (or copy and paste it), the argument demo1 being optional:

```
docker run -it --rm -v ~/docker:/host espcigm/eacn:demo demo1
```

Note: with the Docker Desktop for Windows you will have to accept the filesharing message in the Docker popup for the docker folder located in your home folder. Simply click the 'Share it' button. This can also be customized in the File sharing window of the Docker Settings (Resources menu). The first time this command is entered, you are asked to define a path for the shared folder as described below:

```
Please give the absolute path of the shared folder in the host
(for labelling output files only) [<host>]
```

Some suitable propositions for the absolute path that must be typed are given below for the Macintosh (Unix) and Windows OS, where `home` is the home folder (or user folder), and `docker` is the folder manually created in the home folder:

```
/Users/`home`/docker
```

```
C:\Users\`home`\docker
```

If, for any reason, the folder path has been incorrectly entered, you can correct it by deleting the file "host.pth" located inside the docker folder, and give the proper path at the next launch.

At the end of the computations issued by this command, the results are written in an excel file located in a shared folder mounted inside the container. In the later command line, docker is the folder used; it is automatically created in the home folder at the Docker installation step (~/docker = /Users/home/docker on macOS, see below). If this folder does not exist, no results are saved.

The shared excel file is recorded in the result subdirectory of the docker directory on the host machine ($\sim$/docker/result, where $\sim$ is a shortcut indicating the path to the connected user's home directory). The output file name is automatically incremented when the same command is issued.

Three other computations can be called with the same command line but with the arguments demo2, demo3 and demo4 instead of demo1. The explanations and the outputs of all four command lines are given in the Section 'Graph machine results with Docker'.

Notes

- The above run command is the minimal command; if for example, the "$-v$ $\sim$/docker:/host" is omitted in the command syntax, no excel file is created on the host machine, the computed results being lost when the container is deleted. In that case the path to the shared folder must also be entered for each computation.

- The computed times reported during the demo depend on the machine used. For more explanations on hyper-V you can use the following link: https://docs.microsoft.com/fr-fr/virtualization/hyper-v-on-windows/quick-start/enable-hyper-v

Predicting EACN for the test set

We explain hereafter the demo that describes the EACN computation for the ten test molecules with the selected complexity of the graph machine-based model and computation results are then given for both the GM and the NN models; the command line for launching the graph machine-based model demo is the following:

```
docker run -it --rm -v ~/docker:/host espcigm/eacn:demo
```

The data file is used to compute the EACN values of the molecules of the test set. It contains two sheets as described below:

- The first sheet named DATA, contains the training data for the 111 molecules of the training set in the DATA cell range. Every line of this sheet contains a compound name, its SMILES code, the three computed sigma-moments, the number of carbon atoms and the experimental value of the property of interest (here the dimensionless EACN). The other data present in the file, namely the entry, the molecular formula (MF), the molecular weight (MW) and the chemical abstract registry number (CAS RN) are ignored since they don't belong to the DATA range.

- The second sheet named TEST contains the data for the 10 molecules of the test set, in the TEST cell range. The same information as above is displayed in the eleven columns, but only the data present in the TEST range are taken into account.

During the training step *achieved in an earlier stage*, a graph machine model based on the SMILES code derived from the molecular structure was automatically generated for every molecule of the DATA sheet. All 111 models were then merged into a module that was trained with the desired property values. The parameters at the end of the training were stored to be used with a new model.

In the course of the present demo a graph machine model is generated for the 10 molecules (also inputed as their SMILES codes) of the TEST sheet. After the model constructions, the parameters saved during the training step are passed to the functions of the graph machines to predict the 10 EACN values of the test set molecules. At the end of the computation the file GM_test_5N.xlsx is written in the result folder (located also in the home folder). It can be read with Excel 2010-2021 (or more recent) or with LibreOffice 7.3.1 (or more recent).

The GM_test_5N.xlsx file has one sheet. For each molecule, the following quantities are displayed:

- the experimental value of the EACN;
- the mean estimated value of the EACN, computed in averaging the estimations produced by the 25 models that have the smallest VLOO training scores;
- the minimum estimation obtained for the 25 models having the smallest VLOO training scores;
- the maximum estimation obtained for the 25 models having the smallest VLOO training scores.

Results

159

For the graph machine model with five hidden neurons, the predictions for the ten molecules are obtained with the following command (demo1 can be omitted):

```
docker run -it --rm -v ~/docker:/host espcigm/eacn:demo demo1
```

The terminal output is then:

```
Run demo computing |
computing time: 0.59 s
results:
```

Table A3. EACN estimations for the test set molecules using the GM-5N model.

```
+----------------------------+------------------------------+--------+---------------+--------------+--------------+
| Compound                   | SMILES                       | EACN   | Estimated_EACN | minmin_EACN | maxmax_EACN |
|----------------------------+------------------------------+--------+---------------+--------------+--------------|
| Dioctylether               | CCCCCCCCOCCCCCCCC            | 10.30  |          9.76 |        8.12 |       12.04 |
| 2,2,4,6,6-pentamethylheptane | CC(C)(C)CC(C)CC(C)(C)C     | 11.70  |         11.47 |        6.49 |       27.22 |
| Hemisqualane               | CCC(C)CCCC(C)CCCC(C)C        | 14.80  |         14.87 |       11.21 |       17.65 |
| Limonene                   | CC(=C)C1CCC(=CC1)C           | 1.80   |          1.10 |       -1.22 |        5.47 |
| Caryophyllene              | C\C1=C/CCC(=C)C2CC(C)(C)C2CC1 | 5.95   |          5.39 |        1.30 |        9.83 |
| Octyloctanoate             | CCCCCCCCOC(=O)CCC[C:1]CCC     | 8.10   |          8.74 |        0.26 |       15.81 |
| Isopropyl myristate        | CCCCCCCCCCCCCC(=O)OC(C)C      | 7.25   |          6.94 |       -3.00 |       29.67 |
| Rose oxide                 | CC1CCOC(C1)C=C(C)C           | -1.70  |         -0.92 |       -4.62 |        2.66 |
| Linalyl acetate            | CC(C)=CCCC(C)(OC(C)=O)C=C    | -0.90  |         -1.07 |       -3.81 |        1.80 |
| beta-Ionone                | CC(=O)/C=C/C1=C(C)CCCC1(C)C  | -1.90  |         -1.32 |       -4.53 |        3.57 |
+----------------------------+------------------------------+--------+---------------+--------------+--------------+
```

```
Writing results in file "/Users/home/docker/result/GM_test_5N.xlsx".
```

For the neural network model with six hidden neurons, the predictions for the ten molecules are obtained with the following command:

```
docker run -it --rm -v ~/docker:/host espcigm/eacn:demo demo2
```

The terminal output is then:

```
Run demo 2
computing time: 0.18 s
results:
```

Table A4. EACN estimations for the test set molecules using the NN-6N model.

```
+----------------------------+--------+-------+-------+-----+--------+---------------+-------------+-------------+
| Compound                   | M0     | M2    | M3    | NC  | EACN   | Estimated_EACN | minmin_EACN | maxmax_EACN |
|----------------------------+--------+-------+-------+-----+--------+---------------+-------------+-------------|
| Dioctylether               | 371.28 | 35.31 | 21.42 | 16  | 10.30  |         10.07 |        7.91 |       12.36 |
| 2,2,4,6,6-pentamethylheptane | 244.43 | 14.90 | -0.11 | 12  | 11.70  |          9.80 |        8.07 |       13.01 |
| Hemisqualane               | 320.82 | 16.86 | -0.11 | 15  | 14.80  |         14.59 |       10.18 |       18.11 |
| Limonene                   | 197.80 | 24.85 | 5.78  | 10  | 1.80   |          1.12 |       -1.14 |        5.76 |
| Caryophyllene              | 254.08 | 25.94 | 5.64  | 15  | 5.95   |          6.16 |       -0.35 |       13.09 |
| Octyloctanoate             | 371.81 | 53.73 | 31.52 | 16  | 8.10   |          7.41 |        6.63 |        8.89 |
| Isopropyl myristate        | 392.03 | 55.25 | 31.87 | 17  | 7.25   |          8.06 |        7.31 |        9.57 |
| Rose oxide                 | 214.57 | 42.15 | 32.13 | 10  | -1.70  |         -1.64 |       -3.03 |        1.49 |
| Linalyl acetate            | 263.17 | 58.29 | 27.24 | 12  | -0.90  |         -1.03 |       -4.13 |        0.81 |
| beta-Ionone                | 247.59 | 61.65 | 48.43 | 13  | -1.90  |         -1.46 |       -5.04 |        3.45 |
+----------------------------+--------+-------+-------+-----+--------+---------------+-------------+-------------+
```

```
Writing results in file "/Users/home/docker/result/NN_test_6N.xlsx"
```

In each case, the EACN predictions are the same than those reported as estimation errors in columns 4 and 5 of Table 2.2. The predictions are also stored in the GM_test_5N.xlsx and NN_test_6N.xlsx files as explained in the 'loading and launching the Docker image' section.

Explanation of the demo command line

The execution of the graph machine demonstration can be launched from the command line. The proposed default command line is (gm mode is invoked if demo1 is omitted):

```
docker run -it --rm -v ~/docker:/host espcigm/eacn:demo demo1
```

It contains the following terms:

- "docker": calls the Docker daemon of the host machine;
- "run": launches a Docker container from the Docker image;
- "-it": opens and launches the interactive mode;
- "--rm": destroys the container at the end of the session;
- "-v ~/docker:/host": creates a volume in the container, and shares it with the ~/docker tree on the host machine;
- "espcigm/eacn:demo": name of the Docker image launched with the run command;
- "demo1": this argument invokes the gm mode, as does demo3; if demo2 or demo4 are used, the nn mode is activated. In the former case a SMILES input is expected while a list of four descriptors is required for the later. The demo computations are made with the graph machine model and the neural network model that have the numbers of hidden neurons chosen when looking for the appropriate complexity (i.e. five and six hidden neurons respectively). After completion of the demo, the container is automatically deleted. A new demo session can be started with the same command, but within a new container.

Other command line options

Two subcommands "get" or "draw" can be appended to the command line instead of using demo1 to demo4.

The subcommand "**get**" can be passed to the demonstrator to compute the property value for a single compound using either a SMILES code input (with a GM argument) or some sigma-moment inputs (with a NN argument) as follows:

```
docker run -it --rm -v ~/docker:/host espcigm/eacn:demo get (GM or NN)"id;inputs"
```

where "id" is the name of the compound and "inputs" corresponds either to a SMILES code or to a list of the four comma-separated descriptors M_0, M_2, M_3 and N_C. When a SMILES input is passed, the argument GM is used, while NN is used if the descriptors are the inputs. An excel file with the name "id".xlsx is also written in the result folder as described in the previous Section 'Predicting EACN for the test set'.

The subcommand "**draw**" can be passed to the demonstrator to generate a representation of a graph machine for a given compound. This representation is written in a svg graphic file that can be opened in any browser (or other svg compatible software). The command line used is as follows:

```
docker run -it --rm -v ~/docker:/host espcigm/eacn:demo draw GM "id;SMILES"
```

where "id" and "SMILES" are the name and the SMILES code of the compound for which the representation of the graph machine is requested. The "id_5N".svg file produced is also written in the result folder. The computed representation of the graph machine can be simplified by adding the option "-H n" at the end of the above command line, where n (<5) is the number of hidden neurons of the MLP. It is particularly useful to add the option -H 0 (or 1) to understand how a graph machine is built for a given compound. The file name produced is then "id_0N".svg (or "id_1N".svg).

The computation of the representation of the neural network with six hidden neurons is invoked with the following command:

```
docker run -it --rm -v ~/docker:/host espcigm/eacn:demo draw NN
```

It does not have any option. The name of the file written is model_6N.svg.

Estimations of the EACN values for the 111-molecule training set

For the graph machine model with five hidden neurons, the estimations for the 111 molecules of the training set are obtained with the following command:

```
docker run -it --rm -v ~/docker:/host espcigm/eacn:demo demo3
```

The terminal output is then:

```
Run demo 3 computing \
computing time: 3.25 s
results:
```

Table A5. EACN estimated values for the 111-molecule training set using the GM-5N model.

Compound	SMILES	EACN	Estimated_EACN	minmin_EACN	maxmax_EACN
hexane	CCCCCC	6.00	5.97	5.08	6.73
heptane	CCCCCCC	7.00	6.90	6.27	7.63
octane	CCCCCCCC	8.00	7.92	7.36	8.64
nonane	CCCCCCCCC	9.00	8.94	8.20	9.74
n-decane	CCCCCCCCCC	10.00	9.99	9.56	10.60
undecane	CCCCCCCCCCC	11.00	11.01	10.64	11.42
dodecane	CCCCCCCCCCCC	12.00	12.04	11.61	12.43
tridecane	CCCCCCCCCCCCC	13.00	13.05	12.51	13.49
tetradecane	CCCCCCCCCCCCCC	14.00	14.06	13.58	14.58
pentadecane	CCCCCCCCCCCCCCC	15.00	15.06	14.53	15.71
n-hexadecane	CCCCCCCCCCCCCCCC	16.00	16.06	15.41	16.66
heptadecane	CCCCCCCCCCCCCCCCC	17.00	17.06	16.47	17.60
octadecane	CCCCCCCCCCCCCCCCCC	18.00	18.04	17.52	18.57
nonadecane	CCCCCCCCCCCCCCCCCCC	19.00	18.99	18.47	19.56
eicosane	CCCCCCCCCCCCCCCCCCCC	20.00	19.86	19.18	20.71
Cyclohexane	C1CCCCC1	2.10	1.99	1.04	2.75
Methylcyclohexane	CC1CCCCC1	3.15	2.89	1.85	4.17
Ethylcyclohexane	CCC1CCCCC1	4.23	4.40	3.69	5.15
Propylcyclohexane	CCCC1CCCCC1	5.87	5.73	5.02	6.59
Butylcyclohexane	CCCCC1CCCCC1	7.33	7.05	6.15	8.22
Decylcyclohexane	CCCCCCCCCCC1CCCCC1	14.40	14.52	13.55	15.40
Dodecylcyclohexane	CCCCCCCCCCCCC1CCCCC1	17.50	17.15	15.60	18.22
1-Octene	CCCCCCC=C	3.90	3.60	2.88	4.53
1-Decene	CCCCCCCCC=C	5.50	5.57	4.92	6.40
1-Dodecene	CCCCCCCCCCC=C	8.10	7.85	6.78	8.78
1-Octadecene	CCCCCCCCCCCCCCCCC=C	14.20	14.39	13.52	15.42
Dipropylether	CCCOCCC	0.40	0.71	-0.15	1.40
Dibutylether	CCCCOCCCC	2.97	2.57	1.93	3.30
Dipentylether	CCCCCOCCCCC	4.20	4.38	3.69	4.94
Dihexylether	CCCCCCOCCCCCC	6.20	6.17	5.55	6.76
Diheptylether	CCCCCCCOCCCCCCC	8.00	7.98	7.09	9.02
1-Chlorodecane	CCCCCCCCCCCl	3.50	3.64	2.84	4.46
1-Chlorododecane	CCCCCCCCCCCCCl	5.60	5.68	5.02	6.29
1-Chlorotetradecane	CCCCCCCCCCCCCCCl	8.00	7.79	7.17	8.56
1-Chlorohexadecane	CCCCCCCCCCCCCCCCCl	9.80	9.79	8.82	10.78
Butylbenzene	CCCCc1ccccc1	0.40	0.19	-1.07	1.46
Octylbenzene	CCCCCCCCc1ccccc1	4.00	3.82	3.11	4.70
Decylbenzene	CCCCCCCCCCc1ccccc1	6.00	5.94	5.11	6.68
Dodecylbenzene	CCCCCCCCCCCCc1ccccc1	7.80	8.01	7.15	8.77
1-Octyne	CCCCCCC#C	-1.80	-1.31	-2.39	-0.27
1-Decyne	CCCCCCCCC#C	0.10	0.24	-0.56	0.95
1-Dodecyne	CCCCCCCCCCC#C	2.00	2.00	1.20	2.79
1-Tetradecyne	CCCCCCCCCCCCC#C	3.90	3.79	3.07	4.62
Ethyl decanoate	CCCCCCCCCC(=O)OCC	2.10	2.15	1.26	3.32
Ethyl dodecanoate	CCCCCCCCCCCC(=O)OCC	3.80	3.61	2.87	4.57
Ethyl myristate	CCCCCCCCCCCCCC(=O)OCC	5.20	5.29	4.79	5.84
Ethyl palmitate	CCCCCCCCCCCCCCCC(=O)OCC	6.80	6.90	6.17	7.56
Octanenitrile	CCCCCCCC#N	-1.70	-1.91	-2.63	-1.01
Decanenitrile	CCCCCCCCCC#N	-0.50	-0.63	-1.33	0.23
Dodecanenitrile	CCCCCCCCCCCC#N	0.40	0.68	-0.27	1.56
2-Decanone	CCCCCCCC(C)=O	-2.10	-2.16	-2.86	-1.37
2-Undecanone	CCCCCCCCC(C)=O	-1.30	-0.99	-1.66	-0.40
2-Dodecanone	CCCCCCCCCC(C)=O	-0.60	-0.77	-1.70	0.02
2-Octanone	CCCCCC(C)=O	-3.40	-3.43	-4.28	-2.61
2-methylpentane	CCCC(C)C	6.40	6.13	5.00	7.07
2,2,4-trimethylpentane	CC(C)CC(C)(C)C	8.30	6.98	6.15	9.01
2,3-dimethylbutane	CC(C)C(C)C	4.80	4.81	3.58	5.93
3-methylpentane	CCC(C)CC	5.20	5.18	4.05	6.08
Isohexadecane	CC(CC(C)(C)C)CC(C)(C)CC(C)(C)C	13.90	13.97	13.12	14.83
Pristane	CC(C)CCCC(C)CCCC(C)CCCC(C)C	17.60	17.64	16.68	18.44
1,2-Dimethylcyclohexane	CC1CCCCC1C	3.25	4.05	2.56	4.68
1,4-Dimethylcyclohexane	CC1CCC(C)CC1	4.55	4.31	3.20	5.60
Isopropylcyclohexane	CC(C)C1CCCCC1	5.27	5.36	3.93	6.49
Cyclooctane	C1CCCCCCC1	4.10	3.86	3.22	4.76
Cyclodecane	C1CCCCCCCCC1	5.60	5.98	5.07	6.87
cis-Decalin	C1CC[C@@H]2CCCC[C@@H]2C1	5.30	4.72	3.66	6.01
Pinane	CC1CCC2CC1C2(C)C	4.10	4.34	3.33	5.43
p-Menthane	CC1CCC(CC1)C(C)C	5.83	6.28	5.27	7.22
Cyclohexene	C1CCC=CC1	-1.20	-0.96	-1.86	-0.16
1-Methyl-1-cyclohexene	CC1=CCCCC1	0.00	0.46	-0.92	1.37
4-Methyl-1-cyclohexene	CC1CCC=CC1	0.05	-0.05	-0.78	0.68
3-Methyl-1-cyclohexene	CC1CCCC=C1	-0.50	0.00	-0.93	1.14
cis-Cyclooctene	C1CC\C=C/CC1	1.50	1.20	0.56	2.35
1,3-Cyclohexadiene	C1CC=CC=C1	-3.10	-3.47	-4.43	-2.68
1,4-Cyclohexadiene	C1C=CCC=C1	-4.00	-3.69	-4.61	-2.64
2,5-Norbornadiene	C1C2C=CC1C=C2	-3.20	-3.15	-4.32	-1.88
2,6,10-trimethylundeca-2,6-diene	CC(C)CCC=C(C)CCC=C(C)C	10.30	10.08	8.91	11.17
p-Xylene	Cc1ccc(C)cc1	-2.40	-2.81	-3.94	-1.87
Phenyl-1-butyne	CCC#Cc1ccccc1	-3.30	-3.23	-4.15	-2.30
p-Cymene	CC(C)c1ccc(C)cc1	-0.80	-0.60	-1.96	0.36
alpha-Pinene	CC1=CCC2CC1C2(C)C	3.50	2.87	1.84	4.34
p-Menth-2-ene	CC(C)C1CCC(C)C=C1	3.35	3.58	2.44	4.59
delta-3-carene	CC1=CCC2C(C1)C2(C)C	2.45	2.12	1.20	3.41
beta-Pinene	CC1(C)C2CCC(=C)C1C2	2.15	2.57	1.58	3.49
gamma-Terpinene	CC(C)C1=CCC(=CC1)C	1.65	1.25	0.32	2.50
alpha-Terpinene	CC(C)C1=CC=C(C)CC1	1.15	1.30	0.33	2.28
Terpinolene	CC(C)=C1CCC(=CC1)C	0.70	1.07	-0.08	2.33
Longifolene	CC1(C)CCCC2(C)C3CCC(C13)C2=C	6.95	6.92	5.95	7.89
Diisopropylether	CC(C)OC(C)C	0.60	0.50	-0.52	1.60
1,4-dipropoxybutane	CCCOCCCCOCCC	1.90	1.90	0.98	2.95
1,2-dipropoxyethane	CCCOCCOCCC	0.40	0.29	-0.79	1.22
1,2-dibutoxyethane	CCCCOCCOCCCC	1.70	1.92	0.86	3.06
Decyl butyrate	CCCCCCCCCCOC(=O)CCC	5.00	5.12	4.18	6.15
Hexyl octanoate	CCCCCCOC(=O)CCC[C:1]CCC	6.20	6.20	5.19	7.27
Butyl dodecanoate	CCCCCCCCCCCC(=O)OCCCC	7.20	7.45	6.55	8.54
Myristyl propionate	CCCCCCCCCCCCCCOC(=O)CC	6.80	6.68	5.51	7.58
Isoamyl laurate	CCCCCCCCCCCC(=O)OCC(C)C	8.80	8.56	7.32	9.59
Hexyl dodecanoate	CCCCCCCCCCCC(=O)OCCCCCC	9.30	9.27	8.18	10.25
Ethyl oleate	CCCCCCCC\C=C/CCCCCCCC(=O)OCC	7.20	7.14	6.19	8.09
Menthone	CC(C)C1CCC(C)CC1=O	-1.50	-1.36	-2.51	-0.21
Eucalyptol	CC12CCC(CC1)C(C)(C)O2	-1.60	-1.50	-2.47	-0.50
D-Carvone	CC(=C)[C@H]1CC=C(C)C(=O)C1	-3.10	-3.22	-4.16	-2.05
Hexyl methacrylate	CCCCCCOC(=O)C(C)=C	0.12	-0.02	-1.30	0.92
Menthyl acetate	CC(C)C1CCC(C)CC1OC(C)=O	-0.10	-0.15	-1.21	0.80
Citronellyl acetate	CC(CCOC(C)=O)CCC=C(C)C	-0.20	-0.05	-1.02	0.89
Geranyl acetate	CC(=O)OC\C=C(/C)CCC=C(C)C	-0.60	-0.69	-1.56	0.32
alpha-Damascone	C/C=C/C(=O)C1C(=CCCC1(C)C)C	-1.30	-1.20	-2.30	-0.28
Methyl dihydrojasmonate	CCCCCC1C(CCC1=O)CC(=O)OC	-1.70	-1.64	-2.47	-0.72
Ethylene brassylate	O=C1CCCCCCCCCCCC(=O)OCCO1	-1.10	-1.04	-2.05	0.14
cedryl methyl ether	CO[C@]1(C)CCC23C[C@@H]1C(C)(C)[C@@H]2CC[C@H]3C	3.50	3.54	2.56	4.45
Ambrettolid	O=C1CCCCC\C=C/CCCCCCO1	1.00	0.92	-0.18	1.97

```
Writing results in file "/Users/home/docker/result/GM_data_5N.xlsx"
```

For the neural network model with six hidden neurons, the estimations for the 111 molecules of the training set are obtained with the following command:

```
docker run -it --rm -v ~/docker:/host espcigm/eacn:demo demo4
```

The terminal output is the following:

```
Run demo 4 computing \
computing time: 1.50 s
```

163

```
results:
```

Table A6. EACN estimated values for the 111-molecule training set using the NN-6N model.

Compound	M0	M2	M3	NC	EACN	Estimated_EACN	minmin_EACN	maxmax_EACN
hexane	159.25	8.17	-0.08	6	6.00	6.04	5.18	6.82
heptane	179.32	9.14	-0.03	7	7.00	7.07	6.15	7.71
octane	198.88	10.04	-0.00	8	8.00	8.08	7.45	8.71
nonane	218.16	10.71	-0.00	9	9.00	9.07	8.44	9.65
n-decane	238.31	11.78	0.06	10	10.00	10.09	9.44	10.66
undecane	258.33	12.51	0.10	11	11.00	11.08	10.42	11.65
dodecane	278.75	13.46	0.12	12	12.00	12.09	11.49	12.72
tridecane	298.90	14.31	0.22	13	13.00	12.99	12.30	13.91
tetradecane	319.04	15.18	0.26	14	14.00	13.90	13.03	14.73
pentadecane	339.46	16.11	0.28	15	15.00	14.90	14.15	15.68
n-hexadecane	359.04	17.29	0.38	16	16.00	15.86	15.08	16.57
heptadecane	379.07	18.20	0.43	17	17.00	17.03	16.36	17.88
octadecane	399.34	18.90	0.42	18	18.00	18.23	17.50	19.08
nonadecane	419.21	20.01	0.52	19	19.00	18.96	18.21	19.80
eicosane	440.35	20.71	0.52	20	20.00	19.79	18.36	20.93
Cyclohexane	135.77	5.86	-0.12	6	2.10	2.42	1.43	3.33
Methylcyclohexane	153.55	7.16	-0.04	7	3.15	3.32	2.58	4.06
Ethylcyclohexane	171.52	7.89	-0.13	8	4.23	4.56	3.70	5.32
Propylcyclohexane	191.55	8.77	-0.09	9	5.87	6.08	5.08	7.41
Butylcyclohexane	210.84	9.62	-0.05	10	7.33	7.43	6.12	8.98
Decylcyclohexane	332.16	14.95	0.24	16	14.40	14.59	13.33	15.56
Dodecylcyclohexane	371.81	16.81	0.29	18	17.50	17.07	16.04	18.44
1-Octene	192.38	18.77	3.59	8	3.90	3.68	2.71	4.54
1-Decene	234.26	20.46	3.17	10	5.50	6.16	5.15	7.30
1-Dodecene	274.46	22.29	3.31	12	8.10	8.26	6.83	9.61
1-Octadecene	394.66	27.89	3.60	18	14.20	14.43	12.96	16.03
Dipropylether	171.36	26.39	20.59	6	0.40	0.68	-0.55	1.91
Dibutylether	211.38	28.47	21.23	8	2.97	2.68	1.61	4.02
Dipentylether	251.19	29.57	20.44	10	4.20	4.50	3.17	5.63
Dihexylether	291.31	31.96	21.34	12	6.20	5.99	4.99	6.82
Diheptylether	331.49	32.86	20.36	14	8.00	8.08	6.85	9.26
1-Chlorodecane	257.74	29.94	3.76	10	3.50	3.74	2.63	4.82
1-Chlorododecane	297.96	31.52	3.82	12	5.60	5.73	4.67	6.99
1-Chlorotetradecane	338.17	33.58	3.92	14	8.00	7.87	6.71	8.82
1-Chlorohexadecane	378.22	35.15	3.98	16	9.80	9.59	8.31	11.08
Butylbenzene	197.06	25.15	0.54	10	0.40	-0.03	-1.24	1.37
Octylbenzene	280.39	29.60	0.94	14	4.00	4.48	2.96	6.07
Decylbenzene	320.42	31.55	1.10	16	6.00	6.39	4.88	7.30
Dodecylbenzene	359.87	33.08	1.12	18	7.80	7.78	6.23	9.05
1-Octyne	188.77	33.22	1.53	8	-1.80	-1.78	-2.86	-0.61
1-Decyne	230.44	35.32	1.68	10	0.10	0.09	-1.02	1.21
1-Dodecyne	265.87	36.64	1.52	12	2.00	1.83	0.83	2.87
1-Tetradecyne	310.75	38.72	1.96	14	3.90	3.85	2.37	5.15
Ethyl decanoate	294.96	53.81	33.56	12	2.10	1.88	0.57	3.37
Ethyl dodecanoate	333.62	54.56	32.64	14	3.80	4.55	3.73	5.22
Ethyl myristate	375.02	56.48	32.88	16	5.20	6.48	5.32	7.20
Ethyl palmitate	415.24	59.07	33.25	18	6.80	7.69	6.55	8.47
Octanenitrile	204.22	49.16	22.76	8	-1.70	-1.66	-3.00	-0.69
Decanenitrile	243.52	50.82	22.88	10	-0.50	-0.35	-1.63	0.59
Dodecanenitrile	283.99	52.55	23.07	12	0.40	0.51	-1.03	2.05
2-Decanone	242.35	50.66	35.46	10	-2.10	-1.80	-2.78	-0.28
2-Undecanone	261.87	51.41	35.83	11	-1.30	-1.25	-2.59	-0.16
2-Octanone	201.89	48.96	35.68	8	-3.40	-3.18	-4.39	-1.85
2-methylpentane	154.78	8.47	-0.17	6	6.40	5.69	4.97	6.60
2,2,4-trimethylpentane	181.75	10.89	-0.13	8	8.30	6.81	5.70	8.32
2,3-dimethylbutane	149.31	8.56	-0.25	6	4.80	5.01	3.91	6.09
3-methylpentane	152.61	8.30	-0.28	6	5.20	5.52	4.69	6.48
Isohexadecane	303.13	18.48	-0.11	16	13.90	13.89	12.49	15.24
Pristane	396.57	20.64	0.06	19	17.60	17.93	16.66	19.34
1,2-Dimethylcyclohexane	168.83	8.20	-0.05	8	3.25	4.10	3.33	4.74
1,4-Dimethylcyclohexane	171.39	8.45	-0.03	8	4.55	4.38	3.55	5.15
Isopropylcyclohexane	185.48	9.03	-0.13	9	5.27	5.20	4.51	5.88
Cyclooctane	163.91	7.28	-0.04	8	4.10	3.65	2.77	4.26
Cyclodecane	192.46	9.22	-0.38	10	5.60	5.28	4.49	6.05
cis-Decalin	188.05	7.80	0.05	10	5.30	4.77	3.40	6.16
Pinane	186.28	10.46	-0.38	10	4.10	4.39	3.06	5.86
p-Menthane	200.29	10.29	-0.15	10	5.83	6.07	5.11	6.80
Cyclohexene	132.04	14.28	3.82	6	-1.20	-1.15	-2.06	-0.18
1-Methyl-1-cyclohexene	150.83	14.02	3.11	7	0.00	0.47	-0.76	1.20
4-Methyl-1-cyclohexene	149.89	15.41	3.72	7	0.05	-0.17	-1.00	0.48
3-Methyl-1-cyclohexene	150.06	15.25	3.73	7	-0.50	-0.11	-0.95	0.57
cis-Cyclooctene	160.62	14.36	3.25	8	1.50	0.97	0.27	2.44
1,3-Cyclohexadiene	127.77	20.68	3.54	6	-3.10	-3.41	-4.48	-2.14
1,4-Cyclohexadiene	128.11	21.47	4.65	6	-4.00	-3.61	-4.97	-2.73
2,5-Norbornadiene	133.62	22.95	3.78	7	-3.20	-3.25	-4.67	-2.00
2,6,10-trimethylundeca-2,6-diene	291.67	28.17	4.24	14	10.30	8.96	7.70	10.88
p-Xylene	161.17	23.96	1.22	8	-2.40	-2.44	-3.57	-1.20
Phenyl-1-butyne	195.47	35.72	-0.41	10	-3.30	-3.24	-4.39	-1.72
p-Cymene	195.88	25.88	1.52	10	-0.80	-0.50	-1.62	0.33
alpha-Pinene	186.35	14.89	1.06	10	3.50	3.21	2.05	4.69
p-Menth-2-ene	197.75	17.00	3.01	10	3.35	3.38	1.98	4.46
delta-3-carene	194.74	20.82	3.22	10	2.45	2.01	1.27	3.33
beta-Pinene	184.70	18.79	3.18	10	2.15	2.06	0.99	3.44
gamma-Terpinene	199.76	22.82	4.25	10	1.65	1.58	0.74	2.29
alpha-Terpinene	199.72	23.33	4.58	10	1.15	1.44	0.51	2.23
Terpinolene	198.78	22.97	3.89	10	0.70	1.39	0.13	2.27
Longifolene	237.83	19.09	2.20	15	6.95	7.13	5.78	8.27
Diisopropylether	165.76	31.98	26.80	6	0.60	-0.15	-1.35	1.11
1,4-dipropoxybutane	263.26	46.21	37.90	10	1.90	2.15	0.93	3.27
1,2-dipropoxyethane	216.46	43.75	38.01	8	0.40	0.42	-0.87	1.51
1,2-dibutoxyethane	264.40	50.76	43.18	10	1.70	1.58	0.30	2.79
Decyl butyrate	334.40	53.88	32.77	14	5.00	4.87	4.13	5.68
Hexyl octanoate	333.77	53.52	32.68	14	6.20	4.95	4.24	5.76
Butyl dodecanoate	374.28	55.45	33.31	16	7.20	7.11	6.36	8.54
Myristyl propionate	394.93	57.20	33.01	17	6.80	7.42	6.53	7.99
Isoamyl laurate	387.57	55.79	32.49	17	8.80	7.81	6.98	8.81
Hexyl dodecanoate	413.93	57.13	33.18	18	9.30	8.79	7.87	10.57
Ethyl oleate	448.42	66.44	35.58	20	7.20	6.53	5.28	8.29
Menthone	204.53	47.96	39.96	10	-1.50	-1.72	-3.00	-0.28
Eucalyptol	193.69	33.68	30.32	10	-1.60	-1.18	-2.84	0.71
D-Carvone	201.11	60.13	41.08	10	-3.10	-3.05	-4.39	-1.89
Hexyl methacrylate	246.73	48.51	27.65	10	0.12	-0.37	-1.36	1.03
Menthyl acetate	251.45	47.83	31.61	12	-0.10	-0.57	-1.76	0.89
Citronellyl acetate	277.20	60.77	33.79	12	-0.20	-0.31	-1.25	0.78
Geranyl acetate	274.66	65.71	33.06	12	-0.60	-0.65	-1.55	0.50
alpha-Damascone	244.47	55.35	36.19	13	-1.30	-1.22	-2.41	-0.09
Methyl dihydrojasmonate	289.12	88.67	56.94	13	-1.70	-1.85	-3.14	-0.59
Ethylene brassylate	311.19	84.83	49.39	15	-1.10	-0.99	-2.24	0.20
cedryl methyl ether	263.38	31.76	2.03	16	3.50	3.65	2.37	4.92
Ambrettolid	309.44	58.60	35.90	16	1.00	1.01	-0.07	2.27

```
Writing results in file "/Users/home/docker/result/NN_data_6N.xlsx"
```

Prediction of EACN for a single molecule of the test set

EACN prediction of a given test molecule can be replicated for the default graph machine-based model or for the default neural network-based models. Limonene is given as an example. According to Table 2.1, the following information can be extracted (entry 88) for this molecule:

Compound	SMILES code	M_0^X	M_2^X	M_3^X	N_C	EACN$_{exp}$
limonene	CC(=C)C1CCC(=CC1)C	197.80	24.85	5.78	10	1.8

The command used to predict its EACN from SMILES is the following:

```
docker run -it --rm -v ~/docker:/host espcigm/eacn:demo get GM
"limonene;CC(=C)C1CCC(=CC1)C"
```

The output produced is then:

```
computing time: 0.31 s
results:

+----------+--------------------+----------------+--------------+--------------+
| ID       | smiles             | Estimated_EACN | minmin_EACN  | maxmax_EACN  |
|----------+--------------------+----------------+--------------+--------------|
| limonene | CC(=C)C1CCC(=CC1)C |           1.10 |        -1.22 |         5.47 |
+----------+--------------------+----------------+--------------+--------------+
Writing results in file "/Users/home/docker/result/GM_limonene_5N.xlsx"
```

The results are identical to those written for limonene in the file GM_test_5N.xlsx.

The commands for predicting its EACN from the descriptors used as inputs is the following:

```
docker run -it --rm -v ~/docker:/host espcigm/eacn:demo get NN
"limonene;197.8,24.85,5.78,10"
```

The messages returned are:

```
computing time: 0.01 s
results:

+----------+--------+-------+------+------+----------------+--------------+--------------+
| ID       | M0     | M2    | M3   | NC   | Estimated_EACN | minmin_EACN  | maxmax_EACN  |
|----------+--------+-------+------+------+----------------+--------------+--------------|
| limonene | 197.80 | 24.85 | 5.78 | 10   |           1.12 |        -1.14 |         5.76 |
+----------+--------+-------+------+------+----------------+--------------+--------------+
Writing results in file "/Users/home/docker/result/NN_limonene_6N.xlsx"
```

A9. Creating a graph machine image for a compound

This command is useful to obtain the representation of a graph machine for a new compound, in particular when the estimated value has a large deviation or does not seem right, e.g. the

hexan-2-one, a compound that belongs to the 2-ketone series. In the following example, the command line to generate the graph machine image for hexan-2-one is:

```
docker run -it --rm -v ~/docker:/host espcigm/eacn:demo draw GM
"hexan-2-one;CCCCC(C)=O"
```

The messages returned are the following:

```
Writing image in file /Users/home/docker/result/hexan-2-one_5N.svg
```

The resulting svg file can be visualized with any browser. It has been converted in a png file before insertion into this document in the next page to reduce its size. The svg image is scalable and allows then to zoom on the labels of the connexions between the neurons easily.

A simpler representation of the graph machine built from the same compound but with zero hidden neuron can be obtained with the following command:

```
docker run -it --rm -v ~/docker:/host espcigm/eacn:demo draw GM
"hexan-2-one;CCCCC(C)=O" -H 0
```

The output produced is:

```
Writing image in file /Users/home/docker/result/hexan-2-one_0N.svg
```

The svg output for the first command inserted as a png image is:

hexan-2-one CCCCC(C)=O
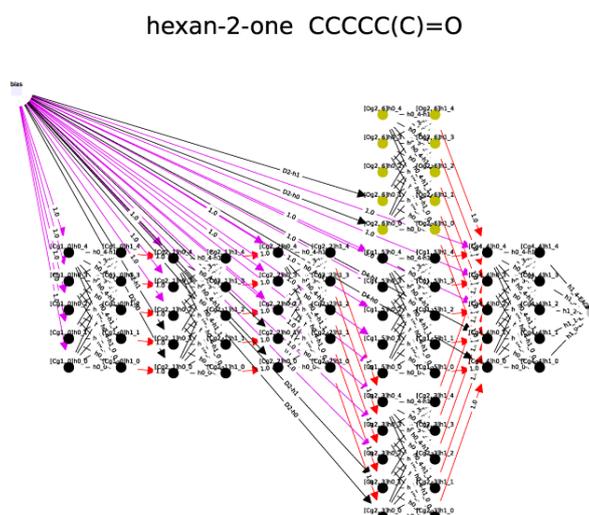
The svg output for the second command inserted as a png image is:

hexan-2-one CCCCC(C)=O
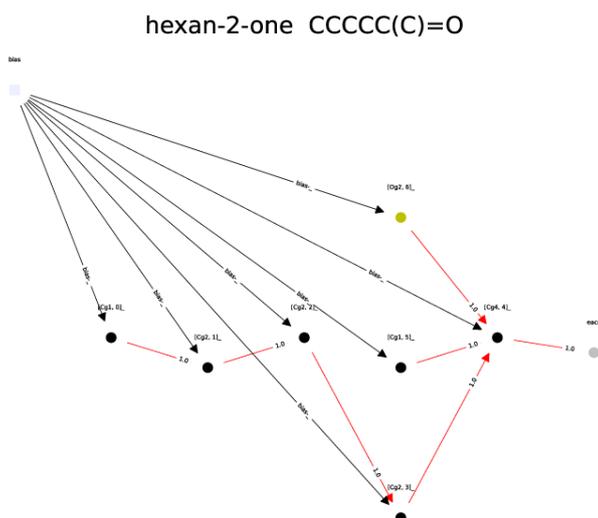
With this representation is much easier to locate the carbon root node indicated in the SMILES code with the bracketed atom. The directed graph for hexan-2-one can also be quickly drawn from this image.

With this input SMILES, the EACN estimation value for hexan-2-one computed by the following command:

```
docker run -it --rm -v ~/docker:/host espcigm/eacn:demo get GM "hexan-2-one;CCCCC(C)=O"
```

is equal to –3.5, a value almost equal to the estimation computed for octan-2-one (–3.4), so that it is probably not correct. Indeed, the data point with coordinates (N_C=8; EACN=–3.5) is above the experimental dashed line of the ketone series in Figure 11a of the paper. Seeking for an explanation of this deviation, we can compare the simplified GM of octan-2-one and decan-2-one with the hexan-2-one GM. The representation for these two ketones are the following:

octan-2-one CCCCCCC(C)=O

decan-2-one CCCCCCCCC(C)=O
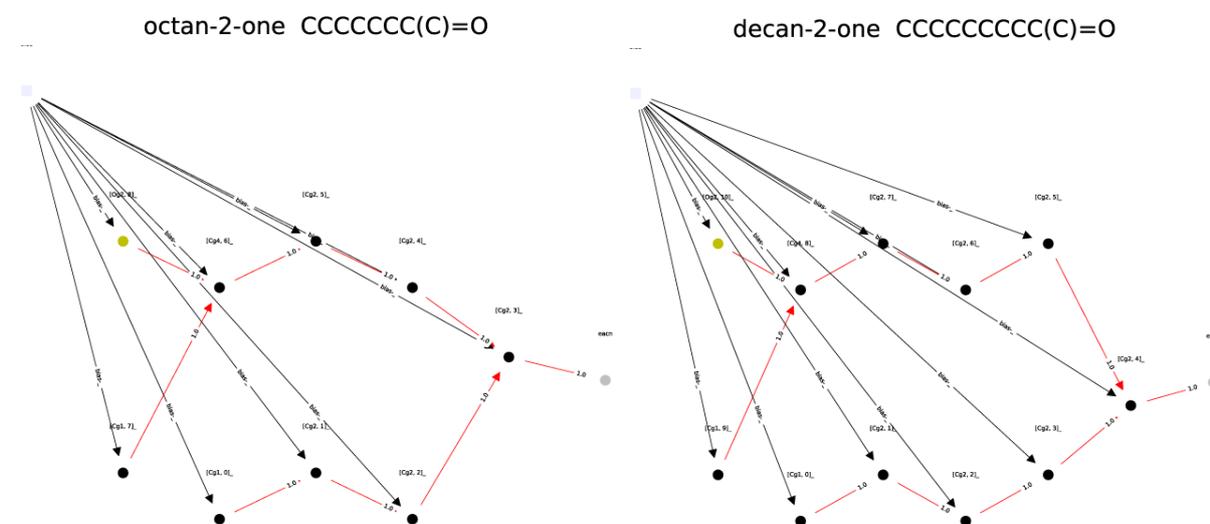
We can observe that the for these two graph machines the root node is not the node corresponding to the ketone carbon like in the previous hexan-2-one representation. For consistency it would be advisable to build a graph machine for the hexan-2-one that would be similar to the former above representation.

The command to compute such a graph machine (with 0 neuron) is then:

```
docker run -it --rm -v ~/docker:/host espcigm/eacn:demo draw GM
"hexan-2-one;CC[C:1]C(C)=O" -H 0
```

The representation of the new graph machine for the hexan-2-one is as follows:

hexan-2-one CC[C:1]CC(C)=O
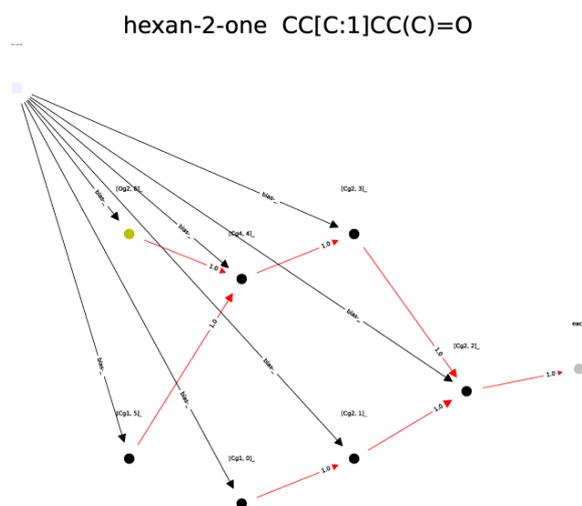
With this new SMILES, the EACN estimation value for hexan-2-one computed by the following command:

```
docker run -it --rm -v ~/docker:/host espcigm/eacn:demo get GM "hexan-
2-one;CCCCC(C)=O"
```

is equal to –4.9, a value that is in line with the computed values of the six other ketones in the series.

A10.   1-O-dodecyl diglyceryl ether supplementary characterization
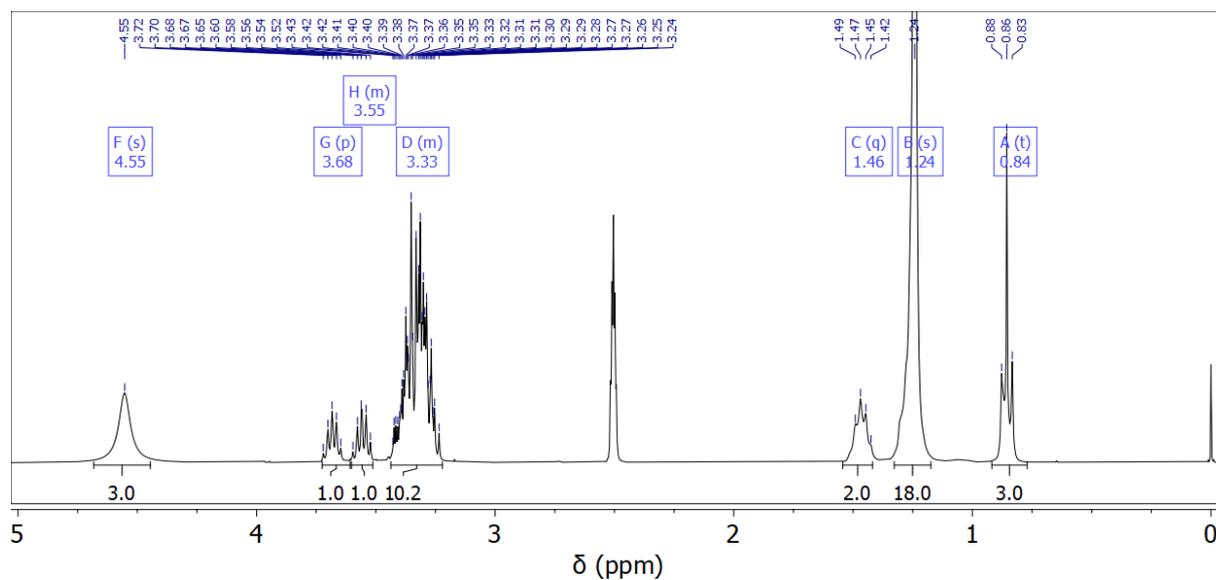
Figure A.37. ^{1}H NMR (300 MHz) spectrum of 1-O-dodecyl diglyceryl ether in d_6-DMSO. Peak at 2.5 ppm corresponds to the solvent peak.
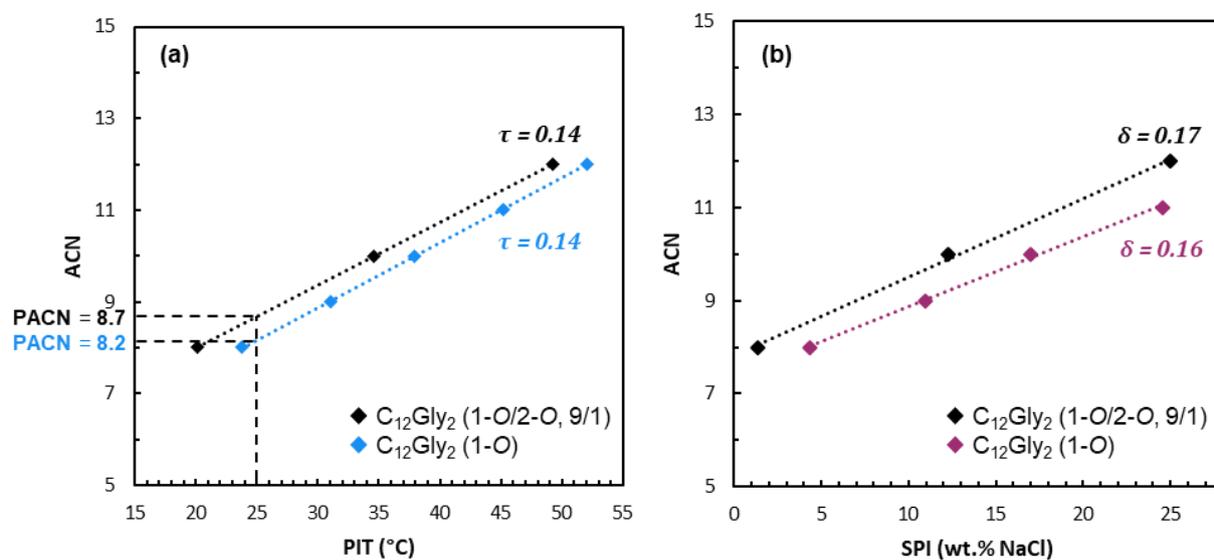
Figure A.38. Comparison of (a) PIT in $C_{12}Gly_2$ (1%)/n-alkane/NaCl 10^{-2} M (f_w = 0.5) systems and (b) SPI in $C_{12}Gly_2$ (1%)/n-alkane/water (WOR = 1) systems for pure 1-O-dodecyl diglyceryl ether and a 9/1 mixture of 1-O-dodecyl diglyceryl ether and 2-O-dodecyl diglyceryl ether.

$C_{12}Gly_2$	1-*O*/2-*O* (9/1)	1-O
PIT (°C)	20.2 ± 0.2	23.8 ± 0.3
Δ PIT	3.6 °C	
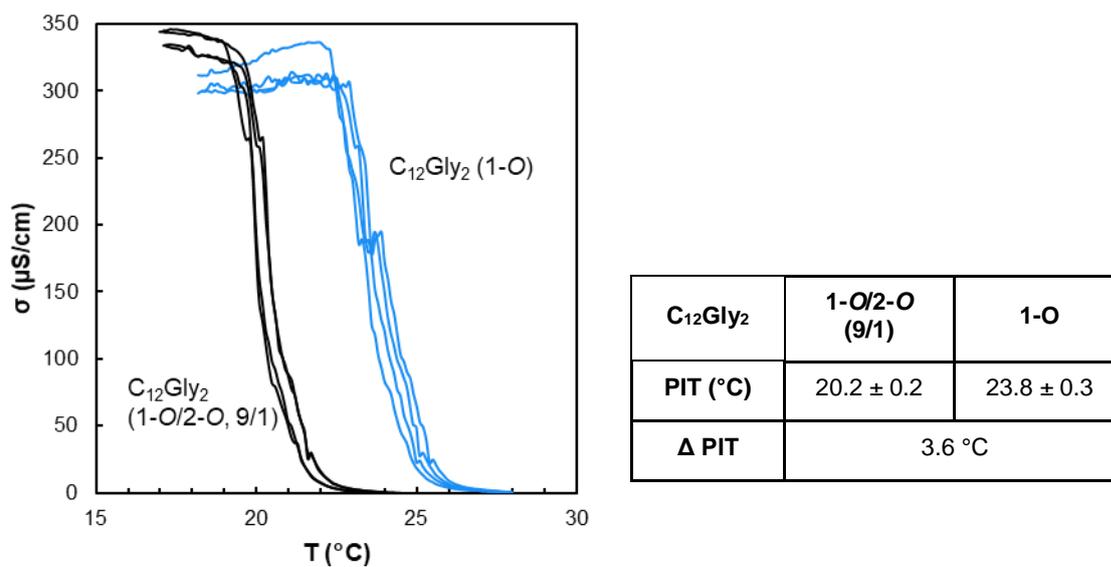
Figure A.39. Conductivity monitoring of PIT in *n*-octane/NaCl 10^{-2} M ($f_w = 0.5$) systems containing 1 wt.% of pure 1-O-dodecyl diglyceryl ether or a 9/1 mixture of 1-*O*-dodecyl diglyceryl ether and 2-*O*-dodecyl diglyceryl ether.

Publications

TiO$_2$ nanoparticle dispersions in water and nonaqueous solvents studied by gravitational sedimentation analysis: Complementarity of Hansen Parameters and DLVO interpretations

Lucie Delforce [a], Evamaria Hofmann [b], Véronique Nardello-Rataj [a,*], Jean-Marie Aubry [a,*]

[a] *Univ. Lille, CNRS, Centrale Lille, Univ. Artois, UMR 8181 – UCCS – Unité de Catalyse et Chimie du Solide, F-59000 Lille, France*
[b] *Univ. Regensburg, Institute of Theoretical and Physical Chemistry, 93053 Regensburg, Germany*

GRAPHICAL ABSTRACT
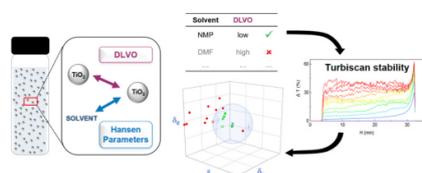
ARTICLE INFO

ABSTRACT

Hansen Solubility Parameters (HSP) have been shown to be an effective approach for rationalizing and predicting the stability of titanium dioxide nanoparticles (TiO$_2$ NPs) dispersions. However, interparticle electrostatic interactions, not considered in Hansen's approach but taken into account in the DLVO theory, are expected to play a significant role in organic solvents having a notable dielectric constant. Zeta potential ζ of TiO$_2$ P25 NPs was measured in both aqueous and organic media to identify DLVO-stabilized dispersions from those stabilized by more specific NP-solvent interaction. Stability was quantified using a Turbiscan optical device which provides Stokes diameters and Relative Turbiscan Stability Index (RTSI). When the zeta potential of NPs and the dielectric constant of the solvent are both high, the dispersion benefits from additional stabilization while when the electrostatic repulsion is negligible, only the solvents within a Hansen dispersion sphere give stable dispersions. The two interpretations are therefore complementary to describe the behavior of TiO$_2$ dispersions in organic solvents.

1. Introduction

Nanoscience occupies an important place in research and industry. Due to their size less than 100 nm [1], nanoparticles (NPs) have peculiar properties relevant for applications in coatings, cosmetics, pharmaceutics, energy and agriculture, to name a few [2,3]. Titanium dioxide

nanoparticles (TiO$_2$ NPs) are among the most widely used NPs, in particular as UV filters in sunscreen cosmetics [4] and plastics to avoid solar degradation [5], but also as photocatalysts in wastewater treatment [6], in self-cleaning transparent coatings [7], in solar cells or as silicon heat-stabilizers [8] and many other applications that require their dispersion in fluid or solid matrixes. It is of major importance that

NPs be and remain homogeneously dispersed in the matrix to achieve optimal properties and stability. For instance, the efficiency of UV filters for skin protection is highly dependent on the distribution of this filter in the sunscreen film [9]. When TiO$_2$ NPs are used as catalysts for wastewater treatment, photocatalytic degradation of pollutants increases with NPs dispersion as it is necessary that a large fraction of the catalytic area is accessible [10]. Moreover, well dispersed nano-scaled particles allow the production of transparent materials [7].

Nanoparticle dispersions can be studied by various experimental methods such as DLS (Dynamic Light Scattering) for size measurement [11–13] as well as gravitational [14–17] or centrifugal [13,18,19] sedimentations coupled with multiple light scattering methods which can provide more detailed information regarding the destabilization mechanisms of dispersions. These techniques are based on the measurement of transmitted and backscattered light over time along the whole height of a cell containing the dispersion. Both methods allow to determine hydrodynamic radius using sedimentation rate. However, agglomeration and flocculation of NPs, due to interparticle attraction, are more efficiently observed by gravitational sedimentation. Actually, when a sample is centrifuged, sedimentation is greatly accelerated whereas attractive interactions remain identical to that under gravitational field and, therefore, do not have time to induce agglomeration of the particles.

In this work, we study the evolution over time of aqueous and organic liquid dispersions of uncoated TiO$_2$ nanoparticles using an optical device. The experiments are carried out under gravitational field in order to observe both the agglomeration and flocculation phenomena that occur when formulations are stored under ordinary conditions. In aqueous media, interparticle interactions are well described by the DLVO theory, named after the researchers Derjaguin, Landau, Verwey and Overbeek. Electrostatic repulsions between charged particle surfaces and van der Waals attractions coming from the solid core of particles contribute to the overall particle interactions [20] and become dominant over gravity as particles get smaller [21,22]. However, it is recognized in the literature that organic solvents behave in a more complex manner than water as NP dispersing media because of their diversity in structure and polarity.

Hansen Solubility Parameters (HSP) approach derives from the Hildebrand solubility parameter which is the square root of cohesive energy density. Hansen's assumption is that this cohesive energy density is due to three types of interactions between molecules: polar (δ_p), dispersive (δ_d) and hydrogen bonding (δ_h) interactions [23]. Originally, Hansen's solubility parameters were developed to study and anticipate the solubilization of molecular and macromolecular compounds in organic solvents. Hansen himself proposed to apply his method to characterize the surface of dispersed particles by arguing that organic liquids which adsorb most strongly to the surfaces of particles are those in which dispersions are most stable [23,24]. It is also argued that for small nanoparticles, energy of solvation can be <0 kT [25] and thus, solvation would be thermodynamically favoured, placing NPs in the frame of Hansen Parameters [26]. However, as the physicochemical phenomena involved in the dispersion of particles are definitely different from those involved in the dissolution of organic compounds, Süß et al. proposed to use the term "Hansen's Dispersion Parameters" (HDP) instead of HSP when Hansen's approach is used to study the dispersibility and stability of particles [13]. HSP has been shown to be a versatile tool for rationalizing and predicting the stability of various types of NP dispersions such as carbon black [13], carbon nanofibers [27], fullerene [28], graphene [29] and carbon nanotubes [30] but also titanium carbides [31] and inorganic nanoparticles of ZnO, Al$_2$O$_3$, ZrO$_2$ [32], hydroxyapatite and TiO$_2$ [18]. In practice, the particles are dispersed in a series of carefully chosen solvents according to a standard protocol. Each solvent is assigned a score based on the stability of the dispersion, then a Hansen sphere including the most effective solvents is built in the 3D Hansen space. However, it can be expected that interparticle electrostatic interactions, not considered in Hansen's approach, also play a significant role, especially in organic solvents with notable dielectric constant. Many other solvent scales such a donor numbers (DN), solvatochromic Kamlet-Taft descriptors and Reichardt's E$_T$(30) could be used to quantify solvent-particle interactions. This work is based on Hansen parameters as the three complementary parameters allow a practical 3D representation and a clear visualization of effective and non-effective solvents.

Herein we discuss the respective contributions of DLVO and non-DLVO interactions in the stability of TiO$_2$ P25 nanoparticle dispersions, with a special emphasis on the relevance of the HSP concept to rationalize non-DLVO interactions in organic solvents. Zeta potential measurements in organic and aqueous media are carried out to identify the solvents in which stability can be explained by the DLVO theory from those for which the stability results from more specific NP-solvent interaction. These latter solvents are used to determine the Hansen sphere of TiO$_2$ P25 with a Turbiscan as a stability analyser.

2. Experimental

2.1. Chemicals

The nanoparticles AEROXIDE ® TiO$_2$ P25 (titanium dioxide, purity $\geq$ 99.5%) were obtained from Acros Organics (Thermo Fisher Scientific Inc., Geel, Belgium). These uncoated TiO$_2$ nanoparticles (NPs) had a specific surface area of 35 – 65 m^2/g [8] and an average primary particle diameter of 21 nm [33]. The crystal structure was mainly anatase (85%) and rutile (15%) [34].

Organic solvents used as dispersion media were supplied by the companies Sigma-Aldrich Chemie GmbH (St. Louis, USA), VWR International GmbH (Radnor, Pennsylvania), Acros Organics (Thermo Fisher Scientific Inc., Geel, Belgium), Alfa Aesar (Thermo Fisher Scientific Inc., Heysham, UK), Honeywell (Honeywell International Inc., Morristown, USA), Verbièse (Laboratoire Verbièse, Merville, France) and TCI (Tokyo Chemical Industry Co. Ltd., Tokyo, Japan). They were all of the highest purity available and used as such. Trifluoroacetic acid (99%) was supplied by Alfa Aesar and tetrabutylammonium hydroxide (1 M in methanol) was supplied by Sigma-Aldrich. Ultrapure water was obtained using a Thermo Scientific Barnstead MicroPure Ultrapure water system with a resistivity of 18.2 MΩ.cm.

2.2. Protocol for dispersing TiO$_2$ P25 nanoparticles

2.2.1. Aqueous dispersions

pH-controlled solutions were prepared by mixing NaCl 10^{-3} M and either NaOH or HCl 10^{-3} M solutions in order to maintain ionic strength at a constant value of 10^{-3} M. 20 mg of TiO$_2$ was placed in a borosilicated glass cell (from Formulaction Company, 27.5 mm diameter), then 20 mL NaCl 10^{-3} M and HCl or NaOH 10^{-3} M were added. The cell was placed in a 23.0 °C thermoregulated bath and sonicated for 12 min using an ultrasonic probe Sonotrode S26d2 (2 mm diameter) immersed by 5 mm in the liquid and operated by the ultrasonic processor UP200St (both from Hielscher). The sonotrode pulse was fixed at 50% and the amplitude at 20%. Thermoregulation was carried out by means of a Huber Ministat 125 circulating water bath. pH was measured by immersing the electrode directly in the cell and a sample was taken to measure particle size and zeta-potential ζ using a Zetasizer Nano ZS from Malvern Panalytical. The cell was wiped to remove water drops from the outside wall and was scanned by a thermoregulated Turbiscan LAB from Formulaction.

2.2.2. Dispersions in organic solvents

In organic solvents, 20 mL of solvent ("pure" in the first series of experiments and containing 10^{-3} M trifluoroacetic acid or tetrabutylammonium hydroxide afterwards) were sampled with a graduated pipette and added to a borosilicated glass cell containing 20 mg of TiO$_2$. The particles were then dispersed and analysed according to the protocol described above.

2

2.3. Turbiscan measurement

Dispersions were scanned every 30 s for 2–72 h at 25.0 °C. Measured signals are transmitted light (*T*) and backscattered light (*BS*) represented as a function of the sample height. Variations in T and BS can be seen by deducting the first scan to all the following scans yielding ΔT and ΔBS. Typical resulting spectra are shown in Fig. 1.

Experimental data were processed using the TurbiSoft Lab software (2.3.1.125 FAnalyser) and *TSI* (Turbiscan Stability Index) was computed according to the following equation [35].

$$TSI(t) = \frac{1}{N_h} \sum_{t_i=1}^{t_{max}} \sum_{z_i=z_{min}}^{z_{max}} |BS, T(t_i, z_i) - BS, T(t_{i-1}, z_i)| \tag{1}$$

with t_{max} the measurement time at which the *TSI* is calculated, z_{min} and z_{max} the lower and upper selected height limits respectively, $N_h = (z_{max} - z_{min})/\Delta h$ the number of height positions in the selected zone of the scan and *BS,T* the considered signal (backscattering *BS* if $T < 0.2\%$ or transmission *T* otherwise). Consequently, $TSI = 0$ for t = 0 and increases as the sample gets destabilized. High *TSI* values are characteristic of unstable dispersions. In organic media, a Relative Turbscan Stability Index (*RTSI*) was defined according to Eq. (2) to compare solvents with each other, so as to consider their viscosity η [cP] and their density ρ_i [g.cm^{-3}] which modify the sedimentation rate according to Stokes' law.

$$RTSI = TSI \times \frac{\eta}{(\rho_{particle} - \rho_{solvent})} \tag{2}$$

Moreover, the variations of *T* (or *BS* when the sample is opaque) allow the calculation of mean particle size via Stokes sedimentation.
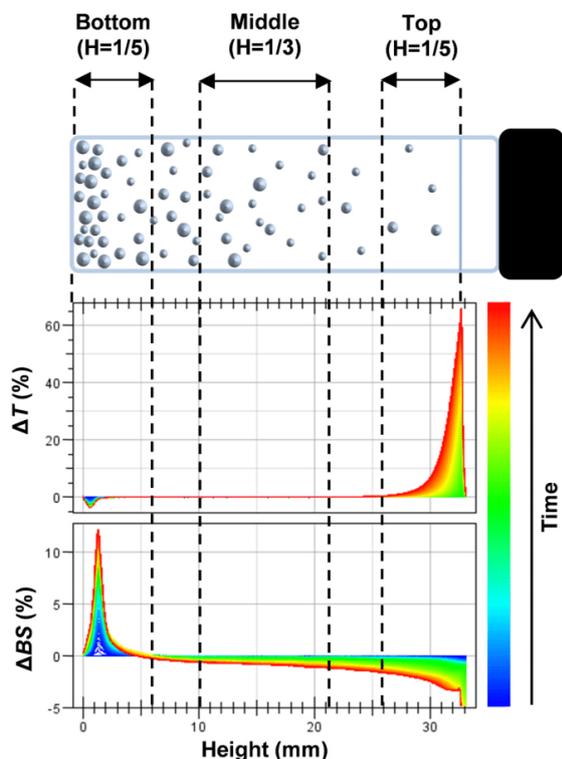
Fig. 1. Change of transmitted light (ΔT) and backscattered light (ΔBS) from Turbiscan data for a partially sedimented dispersion. *T* increases at the top and *BS* increases at the bottom as particles sediment over time.

Using the width evolution at a threshold of the ΔT clarification peak at the top of the cell (see Fig. 2), the migration rate of particles corrected by the viscosity and density of the solvent is linked to particle size by Eq. (3) [36]. A threshold of $\Delta T = \Delta T_{max}/10$ was chosen to measure d_{Stokes} in all samples. When variation of the peak width was not linear, due to sedimentation of different size populations in the sample, the initial slope was used. A comparison with DLS size measurement was made for TiO$_2$ P25 dispersions in non-aqueous solvents and is available in SI. In the case of non-monomodal population, significant differences between d_{DLS} and d_{Stokes} occur. In this work, we chose to consider d_{Stokes} to evaluate particle size.

$$v = \frac{d^2 \quad \times \quad g \quad \times \quad (\rho_p - \rho_f)}{18 \quad \eta} \times \frac{1 - \varphi}{1 + \frac{4.6 \quad \varphi}{(1-\varphi)^3}} \tag{3}$$

with v the sedimentation rate, ϕ the solid volume fraction, d the diameter, g the gravitational constant, ρ_p and ρ_f the density of particles and fluid respectively and η the viscosity.

2.4. Zeta potential measurement

Immediately after dispersing the NPs, 1 mL of dispersion was introduced in a folded capillary zeta cell (aqueous dispersions) or a dip cell (dispersions in organic solvents) and the zeta potential of TiO$_2$ suspensions was measured using a Zetasizer Nano ZS from Malvern Panalytical. In organic solvents, measuring ζ is more challenging than in water. Indeed, the usual folded capillary cell made out of polycarbonate cannot be used for two reasons. Firstly, capillary zeta cells are made of polycarbonate that can be dissolved by some organic solvents. Being made of polyether-ether-ketone (PEEK) polymer and a glass cuvette with excellent chemical compatibility, the dip cell is more addressed for this type of measurements. Secondly, low conductivity media require a higher applied field to observe electrophoretic mobility. The dip cell electrodes are positioned closely together on the PEEK probe, producing greater electric fields with lower voltage and thus avoiding local heating and electrolysis effects [37,38].

Zeta potential is the potential at the surface between a rigidly adsorbed layer and the freely flowing solvent molecules [39]. The functional groups at the particle surface interact with the dispersing medium. Dipole-dipole interactions induce the creation of electric charges that affect the electrostatic forces of particles with one-another. The measurement is based on laser Doppler electrophoresis. The zetasizer measures the frequency difference between the incident and the backscattered beams on the chamber detectors. This Doppler effect is induced by the particle movement under the difference of electric potential. Zeta potential ζ is then calculated using Hückel's equation [40]:

$$u_{H\"uckel} = \frac{2 \quad \varepsilon_0 \quad \varepsilon_r \quad \zeta}{3 \quad \eta} \tag{4}$$

with u the particle mobility, ε_r the relative permittivity of the solvent, ε_0 the electrical permittivity of vacuum and η the viscosity.

Three concordant zeta measurements were achieved for each sample, and values reported in this work are the average zeta values. Uncertainty of measurement is estimated to be the standard deviation. In organic solvents, zeta deviation is generally larger than in aqueous media. Zeta potentials measured in solvents without the addition of electrolytes should be considered specific to this study as it may differ according to experimental conditions, namely the solvent purity, supplier or batch. Zeta potential of TiO$_2$ in nonaqueous solvents is very sensitive to the presence and nature of impurities [41,42].

2.5. Hansen sphere calculation

Two HSP spheres of TiO$_2$ P25 were determined based on either one of two distinct stability criteria, namely d_{Stokes} and RTSI. Solvents were
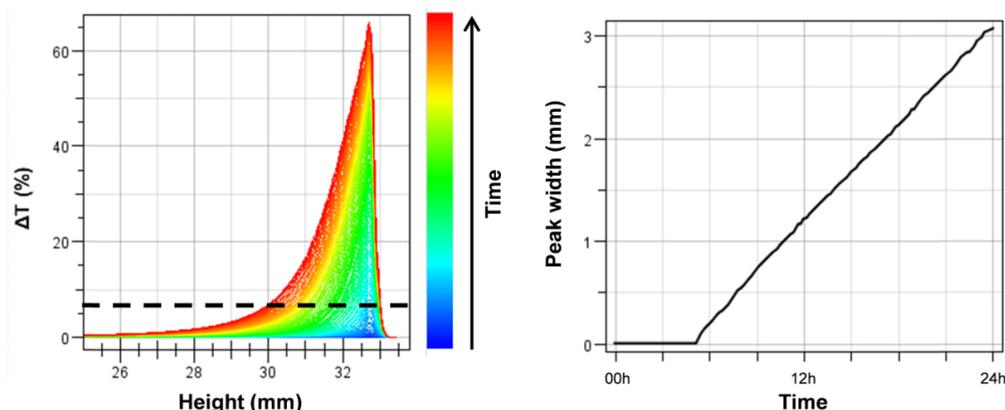
3

Fig. 2. ΔT clarification peak and threshold at $\Delta T = \Delta T_{max}/10$ (left). Peak width evolution at the 1/10 threshold (right): the slope corresponds to the sedimentation rate v.

rated according to one of these parameters: score 1 for "good" dispersing media (stable dispersion) and score 2 for "bad" dispersing media (unstable dispersion). A threshold was defined for d_{Stokes} and RTSI as follows.

Score 1: $d_{Stokes} < 0.31$ μm or $RTSI_{2h}^{top} < 1.0$.

Score 2: $d_{Stokes} > 0.31$ μm or $RTSI_{2h}^{top} > 1.0$.

Using HSPiP software and the scores attributed to each solvent, the spheres are computed so as to include scores 1 and exclude scores 2. The fit indictor reflects the quality of the sphere computation: it decreases if scores 1 are excluded and scores 2 are included in the sphere. The center of the sphere, represented by the three coordinates δ_d, δ_p and δ_h, corresponds to the Hansen Solubility Parameters of TiO$_2$.

3. Results and discussion

The stability of TiO$_2$ nanoparticle dispersions was studied in various aqueous and organic liquid media and at different pH by gravitational sedimentation analysis using the optical analyser Turbiscan. Two complementary approaches were used to analyse and interpret the results. The first, based on DLVO theory, is well suited to dispersions in water and in highly polar organic solvents in which electrostatic repulsions play a major role, while the second, based on Hansen's solubility parameters, makes account for non-DLVO interactions in organic solvents when repulsive electrostatic interactions are weak.

3.1. Aqueous TiO$_2$ dispersions

Zeta potential ζ is the key parameter to rationalize the stability of aqueous TiO$_2$ dispersions [11,43–45]. Fig. 3 (filled dots ●) shows the evolution of ζ as a function of pH in aqueous solution at constant ionic strength.

When dispersed in solution, NPs collide with each other due to the Brownian movement. They can either agglomerate or rebound depending on the relative strength of the van der Waals attraction and the electrostatic repulsion. The DLVO theory which quantitatively accounts for these two types of interactions, was initially developed to rationalize the stability of dispersions in aqueous media [46,47].

Surface charge effects of TiO$_2$ dispersions in water have been extensively studied [11,43–45,48,49]. The pH of the aqueous solution plays an important role as it influences ion exchanges between NPs and water, modifying the surface charge and accordingly, the zeta potential. Indeed, the presence of acidic Ti^{IV} sites on the surface causes water dissociation by adsorption, creating -OH functional groups [48]. NP surface is then modified via reaction with H$^+$ or HO$^-$ ions according to
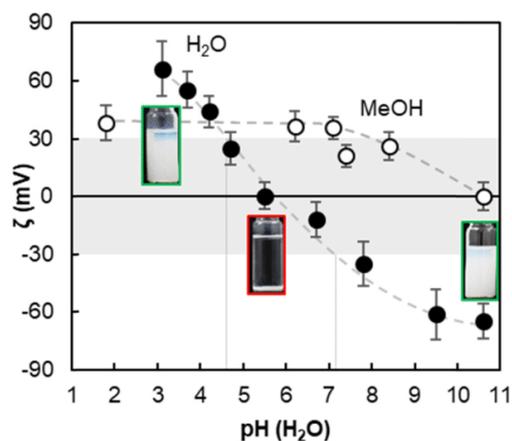
Fig. 3. Evolution of 1 g/L TiO$_2$ P25 nanoparticle zeta potential with pH in water (●) and methanol (x̄) at 25.0 °C referred to the aqueous pH scale. pH is adjusted with HCl and NaOH. Ionic strength is kept constant at 10^{-3} M with NaCl. Instability area is marked in grey, corresponding to pH 4.6–7.2 in water. Pictures of dispersions in water at pH 3.1, 5.5 and 10.6 (left to right) are taken after 24 h.

equilibria as depicted in Fig. 4 [48,50].

Bidentate bridge OH between two Ti atoms (Ti-OH$^+$-Ti) and monodentate terminal H$_2$O adsorbed on 5-fold Ti sites have pK_a values of 2.9 and 12.7 respectively [51]. The reported values for the isoelectric point of TiO$_2$ P25 are comprised between 5.8 and 6.6 [48,51–53]. As shown in Fig. 3 (●), an isoelectric point of 5.9 was found in accordance with the literature data. Electrolyte concentration is also known to strongly impact the zeta potential as more counter-ions can screen surface charges when its concentration increases. Those differences between measured zeta potential and surface potential are minimized at low electrolyte concentration and with monovalent ions [46,47,54]. In this work, pH was adjusted by NaOH or HCl addition and the ionic strength was then adjusted to a constant value of 10^{-3} M by NaCl addition. pH ranging from 4.6–7.2 causes ζ to be comprised within the − 30–30 mV interval (Fig. 3 ●) leading to quickly destabilized dispersions. Such destabilizations can be finely analysed using a Turbiscan through the detection of the transmission T and backscattered BS light signals. Indeed, using T and BS signals, the average TSI value (Eq. (1)) can be
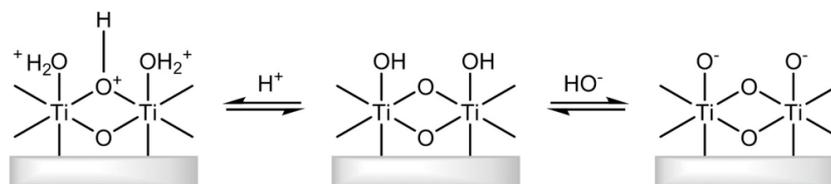
Fig. 4. Simplified illustration of electric charges formation on TiO$_2$ surface by acido-basic reactions in water.

calculated on any portion of the cell and at any time of the analysis (see experimental section). It is worth noting that phenomena differ depending on the height: at the top, *T* increases faster than below as clarification occurs. Fig. 5 shows the evolution of the transmitted *T* and backscattered *BS* lights of two aqueous TiO$_2$ nanoparticle dispersions at pH 6.7 and pH 7.8.

When DLVO repulsion is strong enough to avoid particle agglomeration, the typical profile is the one in Fig. 5b. *T* signal shows no variation except at the top, where slow sedimentation occurs. With time, the sedimentation front (also visible in *BS* light) would eventually reach the bottom of the cell. *BS* light increases as particles accumulate at the bottom. Those samples are easily re-dispersed with a simple re-agitation. On the contrary, when DLVO repulsion is weak (i.e., zeta potential is less than 30 mV), as in Fig. 5a, destabilization is fast. *BS* decreases and *T* increases at the top and in the middle of the cell due to particle agglomeration. Indeed, the particle concentration decreases as they agglomerate. *T* even turns into an irregular signal when agglomerates tend to be individually distinguishable. Sedimented particles accumulate faster if DLVO repulsion is low: the increasing size of agglomerates, in turn, accelerates their sedimentation rate. All those variations over time can be accounted for by the *TSI* value. *TSI* represented in Fig. 6a has been calculated at the top of the cell in order to detect the very early changes even for the most stable samples. After 2 h, TSI_{2h}^{top} values match the expected dispersions behaviour in accordance with the DLVO theory. When $|\zeta|\langle20mV$, TSI_{2h}^{top} is significantly higher than outside these boundaries. Destabilization is faster within this interval whereas

solutions with high $|\zeta|$ remain stable. In Fig. 6b, the Stokes diameters have been reported as a function of ζ. They reach a maximum around the isoelectric point and are noticeably smaller as $|\zeta|$ increases. All these findings clearly show that Turbiscan TSI_{2h}^{top} and Stokes diameters appear as reliable values to quantify dispersion stability. In particular, the maximum d$_{Stokes}$ being attained for zeta potentials close to 0 mV is in accordance with the increase of coagulation rate when electrostatic repulsion is low. In that case, there exists little or no energetic barrier to prevent particles from colliding and coagulation occurs rapidly [55].

3.2. TiO$_2$ dispersions in methanol and other nonaqueous solvents

DLVO theory was initially developed to quantitatively account for interactions between particles in aqueous media but it can also be applied to organic dispersing media [15,56,57]. The main differences between aqueous and organic dispersing media concern dielectric constants and electrolyte concentrations. Both those factors impact the electrostatic repulsion potential.

The stability of TiO$_2$ dispersion was investigated in a common polar and protic solvent, namely methanol. Fig. 3 (empty dots x̂) shows the variation of the zeta potential at different pH values measured using a glass electrode standardized in aqueous buffer. The pH range accessible, based on the water scale, is -1.8–17.2 and depends on the dissociation constants of methanol [58]. The lower limit is given by the transfer activity coefficient of H$^+$ ions in methanol $p\gamma_{H_2O\rightarrow MeOH}^{t}(H^+) = -1.8$ and the upper limit is given by the autoprotolysis constant of methanol

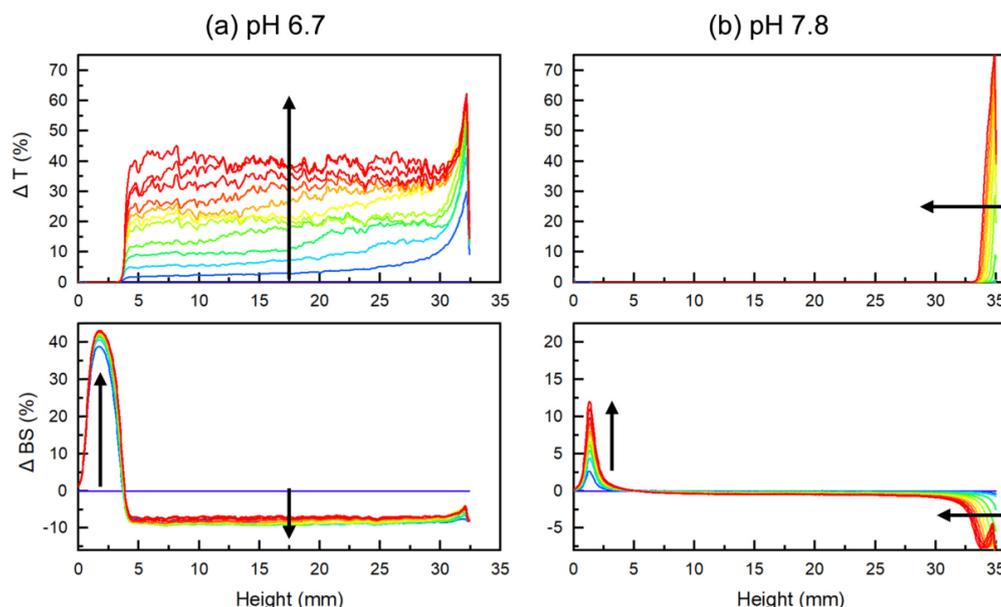**Fig. 5.** Variations over 2 h of transmitted *T* and backscattered *BS* light of aqueous TiO$_2$ P25 1 g/L dispersions (T = 25.0 °C) at pH 6.7 and ζ = - 8 mV (a) and at pH 7.8 and ζ = - 23 mV (b).
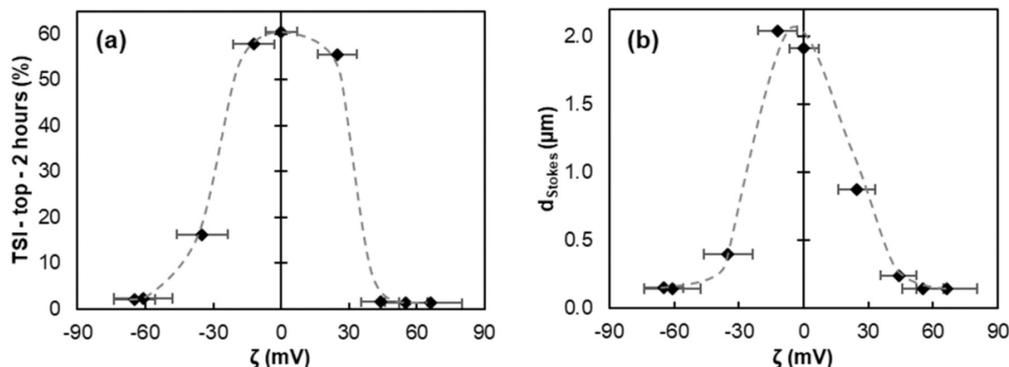
Fig. 6. Evolution at 25.0 °C of (a) *TSI* at the top of the cell after 2 h and (b) Stokes diameters determined using the Turbiscan with zeta potential ζ of 1 g/L TiO$_2$ P25 aqueous dispersions. Ionic strength is kept constant at 10^{-3} M by NaCl addition, pH is adjusted using NaOH and HCl solutions.

$pK_{MeOH} = 17.2$ [58–60].

First of all, in contrast to aqueous media, it appears that whatever the pH value, ζ remains positive. A quite similar observation was made by Kosmulski et al. for TiO$_2$ in water/methanol mixtures: zeta potential became closer to 0 mV with increasing amounts of methanol [61]. On the other hand, it slightly increases with the addition of HCl, whereas NaOH addition brings ζ closer to 0 mV but not in the negative values. Dissociative adsorption of methanol on a hydrated TiO$_2$ surface is described by the following equations [62].

$$MeOH \rightarrow MeO^-_{adsorbed} + H^+ \tag{5}$$

$$H^+ + O^{2-}_{surface} \rightarrow OH^-_{adsorbed} \tag{6}$$

$$MeOH + O^{2-}_{surface} \rightarrow MeO^-_{adsorbed} + OH^-_{adsorbed} \tag{7}$$

As a result, surface hydroxyl groups (shown in Fig. 4) are partially replaced by methoxy groups, which are not able to donate protons. The creation of negative surface charges in the presence of NaOH is thus reduced. Moreover, in the presence of electrolytes, there exists a charge screening from Na$^+$ and Cl$^-$ ions. It is assumed that Cl$^-$ ions interact with the surface through TiOH$_2^+$Cl$^-$ interactions, whereas Na$^+$ ions interact in a non-specific way with negatively charged groups on TiO$_2$ surface [62]. When NaOH is added to methanol dispersions, TiOH$_2^+$ disappears from the surface and Cl$^-$ ion adsorption decreases while Na$^+$ ion adsorption ability remains identical. This may explain why the zeta potential remains positive whereas surface potential may be negative. Fig. 7 displays the evolution of TSI_{2h}^{top} and Stokes diameters determined with the

Turbiscan as a function of the zeta potential ζ in methanol.

As in water, TSI_{2h}^{top} and particle diameters vary depending on the zeta potential suggesting heavily that the stability of TiO$_2$ dispersions in methanol is mainly influenced by electrostatic repulsions as in water. Actually, it is well established that DLVO theory is also applicable in polar organic solvents and may explain the stability of dispersions provided that the dielectric constant is high enough. However, for solvents of lower permittivity ε, electrostatic repulsions significantly decrease even when zeta potential is relatively high. Indeed, the repulsion potential V_R between two spheres is given by the following expression [63].

$$V_R = 2 \quad \pi \quad \varepsilon \quad a \quad \zeta^2 \log(1 + exp(-R\kappa)) \tag{8}$$

where ε is the solvent permittivity, a is the particle radius, ζ is the zeta potential, R is the distance between two spheres and κ^{-1} the Debye-Hückel distance defined by [64].

$$\kappa^{-1} = \sqrt{\frac{\varepsilon k_B T}{2 N_A e^2 I}} \tag{9}$$

where k_B is the Boltzmann constant, N_A is the Avogadro constant, e is the electron charge and I is the ionic strength. The attractive component V_A, described by Hamaker [65], is given by Eq. (10).

$$V_A = \frac{-A_{12}}{12} \frac{a}{R} \tag{10}$$

Hamaker's constant A$_{12}$ of component 2 in medium 1 is calculated
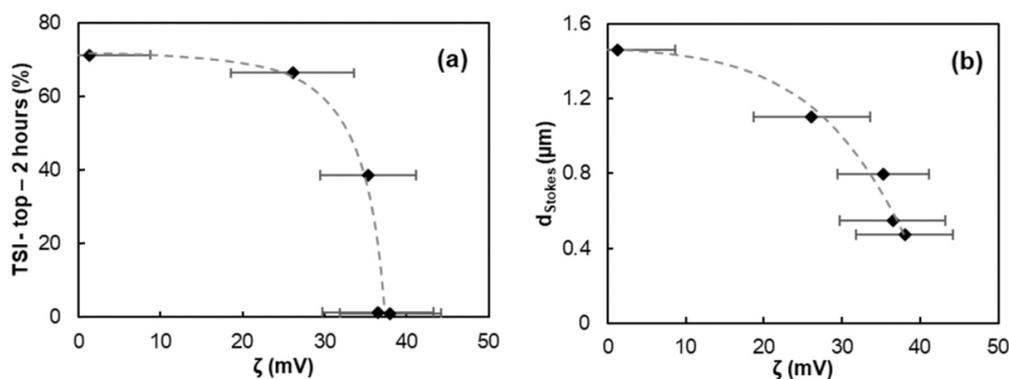
Fig. 7. Evolution of (a) TSI at the top of the cell after 2 h and (b) Stokes diameters d$_{Stokes}$ as a function of the zeta potential of 1 g/L TiO$_2$ P25 particles in methanol. Ionic strength is kept constant at 10^{-3} M with NaCl and pH is adjusted with NaOH or HCl.

6

with Eq. (11) from each component constants A_1 and A_2 expressed by Eq. (12) [21].

$$A_{12} = \left(\sqrt{A_1} - \sqrt{A_2} \right)^2 \tag{11}$$

$$A_i = \frac{3}{4} k_B T \frac{(\varepsilon_{r,i} - 1)^2}{(\varepsilon_{r,i} + 1)^2} + \frac{3 h \nu_e (n_i^2 - 1)^2}{16\sqrt{2}(n_i^2 + 1)^{3/2}} \tag{12}$$

ε_r is the relative permittivity, h is Planck's constant, ν_e is the main electronic absorption frequency for the dielectric permittivity (calculated based on ionization energies, details are given in SI) and n is the refractive index.

When the total interaction energy $(V_R - V_A)$ is below zero no energy barrier is opposed to the coagulation and the dispersion is extremely unstable. On the contrary, when $(V_R - V_A)$ is above 25 $k_B T$, the elec-

trostatic repulsion is so predominant over the Van der Waals attraction that the dispersion is kinetically stable although the thermodynamically stable state corresponds to particles in contact for lyophobic colloids [66]. Between these extreme cases, the energy barrier more or less stabilizes the dispersions without completely preventing aggregation.

According to formulas (8) and (9), the electrostatic repulsion between two particles depends mainly on the dielectric constant of the medium and the zeta potential of the particles. These two parameters were therefore considered for assessing the DLVO contribution to the stability of the dispersions. Zeta potential ζ was measured in samples containing 1 g/L TiO$_2$ P25 nanoparticles dispersed in a series of 17 "pure" solvents, i.e., of the highest purity commercially available and free of any additional compound. The analysis of the transmitted and backscattered signals recorded by the Turbiscan for two hours provides the average diameter of the aggregates (d$_{Stokes}$) as well as the so-called

Table 1
Stability results and physicochemical characteristics of TiO$_2$ dispersions (1 g/L) in various "pure" solvents at 25 °C: viscosity (η), density (ρ), dielectric constant (ε_r), zeta potential (ζ), Relative Turbiscan Stability Index (RTSI), Stokes diameters (d$_{Stokes}$) and total interaction energy $(V_R\text{-}V_A)_{max}$ calculated according to Eqs. (8)–(12). Calculation details are given in SI.

Solvent	Structure	η (cP)	ρ (g/cm^3)	ε_r	ζ (mV)	RTSI$_{2h}^{top}$	d$_{Stokes}$ (μm)	$\dfrac{(V_R - V_A)_{max}}{k_B T}$
Nitromethane		0.67	1.13	39.0	-31±9	0.3	0.13	2.1
Ethanol		1.22	0.82	24.5	31±1	0.4	0.14	1.1
Pyridine		0.88	0.98	12.4	44±10	0.2	0.18	0.9
Isopropanol		2.1	0.79	17.9	-70±9	0.8	0.19	5.8
Acetonitrile		0.37	0.79	37.5	-34±10	0.9	0.21	2.5
DMF		0.92	0.94	36.7	43±14	0.8	0.26	4.6
NMP		1.67	1.03	33.0	-7±5	0.3	0.40	< 0
THF		0.95	0.98	7.6	19±2	1.4	0.53	< 0
Acetone		0.32	0.79	20.7	-26±2	7.2	0.81	0.2
γ-Butyrolactone		1.75	1.13	41.0	-4±2	13.1	0.89	< 0
Water		0.89	1.00	80.1	-5±4	14.5	1.11	< 0
Ethyl Acetate		0.46	0.9	6.0	44±12	2.2	1.14	< 0
Propylene Carbonate		2.8	1.2	64.9	-46±6	28.4	1.34	11.0
Methanol		0.54	0.79	32.7	11±13	10.8	2.88	< 0
Heptane		0.42	0.68	1.9	-10±3	5.6	8.05	< 0
Toluene		0.55	0.89	2.4	-30±8	9.7	9.47	< 0
Triethylamine		0.36	0.73	2.4	6±6	4.8	9.59	< 0

Relative Turbiscan Stability Index ($RTSI_{2h}^{top}$). This corrected index is calculated from the *TSI* by taking into account the viscosity and density of the solvents (Eq. (2)). The solvents are listed in Table 1 according to the increasing size of the aggregates.

Most of the observed stabilities can be rationalized by the DLVO theory. It thus appears that the six solvents leading to the finest particles ($d_{Stokes} < 0.3$ µm), namely nitromethane, ethanol, pyridine, isopropanol, acetonitrile and dimethylformamide, correspond to particles with a high zeta potential ($|\zeta| > 30$ mV) dispersed in quite polar solvents ($\varepsilon_r > 10$). On the contrary, the seven solvents in which the aggregates are the largest ($0.9 < d_{Stokes} < 10$ µm) have either a very low dielectric constant ($\varepsilon_r < 10$), namely ethyl acetate, heptane, toluene and triethylamine, or a low zeta potential ($|\zeta| < 6$ mV) insufficient for the electrostatic repulsion to dominate the Van der Waals attraction, namely, γ-butyrolactone, water and methanol. Three solvents (N-methyl pyrrolidone, tetrahydrofuran and acetone) have an intermediate behavior because either their dielectric constant or the zeta potential of the particles is slightly lower than the threshold values defined above.

Likewise, the evolution $RTSI_{2h}^{top}$ (Table 1) follows the same trend as the best solvents exhibit low values ($RTSI_{2h}^{top} < 1$) whereas the less efficient solvents have very high values ($RTSI_{2h}^{top} > 10$). However, two solvents fail to fall into this general framework of explanation: N-methyl pyrrolidine (NMP) in which the dispersion is stable although the electrostatic repulsion is negligible ($|\zeta| = 2.3$ mV), and propylene carbonate which meets both stability criteria ($\varepsilon_r = 64.9$ and $|\zeta| = 51.8$ mV) but in which the dispersion is extremely unstable. The presence of these outliers indicates that other phenomena not considered in the DLVO theory are involved in the stability of TiO$_2$ dispersions. To analyze the influence of non-DLVO forces, the alternative approach of Hansen's solubility parameters (HSP) is investigated below.

3.3. Stability of TiO$_2$ dispersions in organic solvents interpreted by Hansen's approach

The Hansen Solubility Parameters (HSP) method is a pragmatic and versatile tool originally developed to facilitate the finding of solvents able to dissolve paint resins. The principle of the method is based on the idea that "like dissolves like", which means that a solvent should effectively dissolve a solute provided it resembles it. This concept was quickly extended to solid/liquid dispersions to help formulators in designing the most suitable media for dispersing pigments. This variant is based on the "like disperses like" principle assuming that particles disperse better in solvents having high affinity for the surface.

In the HSP approach, only three types of interaction between the particle and the surrounding medium are considered, namely hydrogen bonding, dipolar interactions and dispersive interactions due to London forces. The attractive interactions of Van der Waals are considered in both theories HSP and DLVO. On the other hand, electrostatic repulsions are only considered in DLVO theory while hydrogen and dipole bonds are only considered in Hansen's approach. The decisive impact of electrostatic stabilization is clearly established in the case of charged TiO$_2$ particles dispersed in polar solvents, but it is unable to explain the stability observed for some TiO$_2$ dispersions when $|\zeta|$ is low.

The estimation of the respective contributions of these complementary interactions would provide valuable information to rationalize the experimental results usually interpreted on the basis of only one of these theories. The difficulty to quantify the stability of a dispersion has been brought up in the literature and Süß et al. [13] proposed to use analytical centrifugation and quantify the sedimentation rate. This technique allows analyzing a large number of samples in a timely manner compared to gravitational sedimentation, but interparticle attraction and coagulation cannot be detected by this technique. In this work, we have chosen to use the Turbiscan device based on gravitational sedimentation which detects the two phenomena involved in the destabilization of dispersions under ordinary storage conditions. Namely, on the one hand,

the coagulation of particles resulting from Brownian motion and Van der Waals attraction and, on the other hand, the sedimentation of particles and aggregates under the effect of gravity.

To focus on the influence of non-DLVO forces, the contribution of electrostatic repulsion to stabilization was minimized by decreasing the zeta potential of particles and removing the energy barrier. Thus, whenever $(V_R - V_A)_{max}/k_BT$ in "pure" solvents was positive (Table 1), 10^{-3} M acid or base was added and the zeta potential was measured again to verify that it is weak enough. Trifluoroacetic acid (TFA) and tetrabutylammonium hydroxide (TBAH) were chosen to allow counter-ion solubility in organic solvents and avoid ionic adsorption at the particle surface. TFA was expected to increase ζ whereas TBAH was expected to decrease ζ. Of course, when solvents included in Table 1 already have a negative $(V_R - V_A)_{max}/k_BT$, the stability of dispersions was interpreted according to HSP without adding TFA or TBAH. Table 2 summarizes the experimental stabilities of the non-electrostatically stabilized dispersions which are interpreted on the basis of HSP. The solvents are listed in

Table 2 according to the increasing size of the aggregates. Each solvent is then assigned a score of 1 (stable) or 2 (unstable) based on the RTSI and the diameter of the aggregates:

Score 1: $d_{Stokes} < 0.31$ µm or $RTSI_{2h}^{top} < 1.0$.

Score 2: $d_{Stokes} > 0.31$ µm or $RTSI_{2h}^{top} > 1.0$.

Two HSP spheres were determined using HSPiP software considering each of the stability criteria, i.e., d_{Stokes} or $RTSI_{2h}^{top}$ calculated based on different methods described hereafter. The first one based on the size of the aggregates, d_{Stokes}, is shown in Fig. 8. The "fit" indicator, which can vary between 0 and 1, expresses the effectiveness of the sphere for modelling experimental results. The more misplaced solvents, the more the fit decreases. Here the fit is very good (0.90), since only one effective solvent (isopropanol) is excluded from the sphere and one poor solvent (acetone) is included.

The second sphere calculated from $RTSI_{2h}^{top}$ values is somewhat less satisfactory (fit = 0.88) since two effective solvents are outside the sphere and two poor solvents are inside (see Fig. S2 in SI). In both cases the radius of the sphere is 8.8 but the first sphere including all dispersions with $d_{Stokes} < 0.31$ µm appears to be the most reliable. In a study on the influence of different dispersion conditions on the size of aggregates, Jiang et al. showed that the ultrasonic probe Sonotrode provided aggregates of 155 nm [11] but was unable to further break the TiO$_2$ P25 aggregates down to the elementary particle (25 nm). In this work, we consider that particles size up to twice this minimal size provide the most stable dispersions. Beyond this size, a competition takes place between NP-solvent and NP-NP interactions that prevent the formation of smaller particles.

It is worth noting that with the first sphere (Fig. 8), the two misplaced solvents are both located near the boundary separating effective and poor solvents. Thus, the poor solvent acetone is erroneously slightly inside the sphere with a RED of 0.99 while isopropanol (effective solvent) is slightly outside with a RED of 1.11. RED is the "Relative Energy Difference" defined by the relationship RED = R_a / R where R_a is the distance between the solvent and the sphere centre and R is the sphere radius. So, when RED < 1, the solvent is inside the sphere and when RED > 1, the solvent is outside the sphere.

Other authors have characterized TiO$_2$ particles [23] and nanoparticles [18] using HSP. However, none of them considered electrostatic repulsion as a possible stabilization phenomenon. Comparative results are displayed in Table 3. Characteristic parameters of the sphere calculated by Hansen regarding TiO$_2$ pigments are very different from those determined in this study. However, Hansen investigated the paint pigment TiO$_2$ Kronos RN 57 at a concentration 20 times greater than that of the present work [23] and no information is available regarding coating and particle diameter. Actually, it is likely that this white pigment has a diameter close to 0.3 µm to maximize its opacifying power and that it has undergone a surface modification to avoid that the

Table 2

TiO$_2$ dispersions (1 g/L) in various solvents at 25 °C for which zeta potential (ζ) and total interaction energy $(V_R$ - $V_A)_{max}$ are minimized by addition, when necessary, of 10^{-3} M TFA (a) or TBAH (b): Relative Turbiscan Stability Index ($RTSI_{2h}^{top}$), Stokes diameters (d$_{Stokes}$), zeta potential (ζ), Hansen solvent parameters and Relative Energy Difference (RED) for the Hansen's sphere based on d$_{Stokes}$. Calculation details are given in SI.

Solvent	Structure	$RTSI_{2h}^{top}$	d$_{Stokes}$ (µm)	$\frac{(V_R - V_A)_{max}}{k_B T}$	ζ (mV)	δ$_d$	δ$_p$	δ$_h$	RED
Nitromethane [a]		0.30	0.12	< 0	-19±6	15.8	18.8	6.1	0.99
Acetonitrile [a]		0.20	0.14	< 0	-11±7	15.3	18	6.1	0.95
Isopropanol [a]		0.50	0.18	< 0	15±6	15.8	6.1	16.4	1.11
DMF [a]		0.90	0.25	< 0	12±1	17.4	13.7	11.3	0.64
Ethanol [b]		1.00	0.31	< 0	-14±11	15.8	8.8	19.4	0.99
NMP		0.30	0.4	< 0	-7±5	18.0	12.3	7.2	1.00
THF		1.40	0.53	< 0	19±2	16.8	5.7	8.0	1.36
Acetone [a]		8.10	0.58	< 0	-7±4	15.5	10.4	7.0	0.99
γ-Butyrolactone		13.10	0.89	< 0	-4±2	18.0	16.6	7.4	1.01
Propylene Carbonate [a]		4.50	1.09	< 0	-18±2	20.0	18	4.1	1.61
Water		14.50	1.11	< 0	-5±4	15.5	16	42.3	3.21
Ethyl Acetate		2.20	1.14	< 0	44±12	15.8	5.3	7.2	1.41
Methanol		10.80	2.88	< 0	11±13	14.7	12.3	22.3	1.01
Pyridine [b]		14.60	3.34	< 0	-24±11	19.0	8.8	5.9	1.49
Heptane		5.60	8.05	< 0	-10±3	15.3	0.0	0.0	2.38
Toluene		9.70	9.47	< 0	-30±8	18.0	1.4	2.0	2.21
Triethylamine		4.80	9.59	< 0	6±6	15.5	0.4	1.0	2.27

[a] the solvent contains 10^{-3} M trifluoroacetic acid

[b] the solvent contains 10^{-3} M tetrabutylammonium hydroxide

photocatalytic properties of TiO$_2$ degrade the organic matrix of the coating. Hence, the spheres are hardly comparable.

Differences with Wieneke's results could come from the nature of the solvents used in both cases. Moreover, Wieneke et al. [18] studied smaller TiO$_2$ NPs with an average particle size of 5.3 nm composed at 95% of anatase and 5% of rutile with a specific surface area of 265 m^2/g which is much greater than that of TiO$_2$ P25. However, the sphere obtained by Wieneke et al. is mostly included inside the one calculated in this work and the sphere radius is more than twice smaller. Also, as TiO$_2$ NPs get smaller, interparticle forces get stronger, making dispersion harder to achieve at the expense of solvent-particle interactions [22]. Another impacting factor is the sample concentration which was of 0.015% for Wieneke's study as the visual changes in stability are easier

to detect with the naked eye in dilute samples. In their case, coagulation was evaluated through the colour change from bluish to white [18]. Using the Turbiscan, quantitative descriptors allow detecting destabilization before the naked eye could detect anything.

The Turbiscan-based method developed in this work is more reliable than those reported in the literature to determine HSP of TiO$_2$ nanoparticles. Indeed, the phenomena occurring during the destabilization, namely the coagulation of particles and the sedimentation of aggregates, can be observed separately. Moreover, this work clearly shows that, when particles are charged and dispersed in polar solvents, the electrostatic repulsion must absolutely be taken into account to rationalize the observed stabilities using DLVO theory. As electrostatic repulsions are not considered in Hansen's theory, only dispersions of weakly
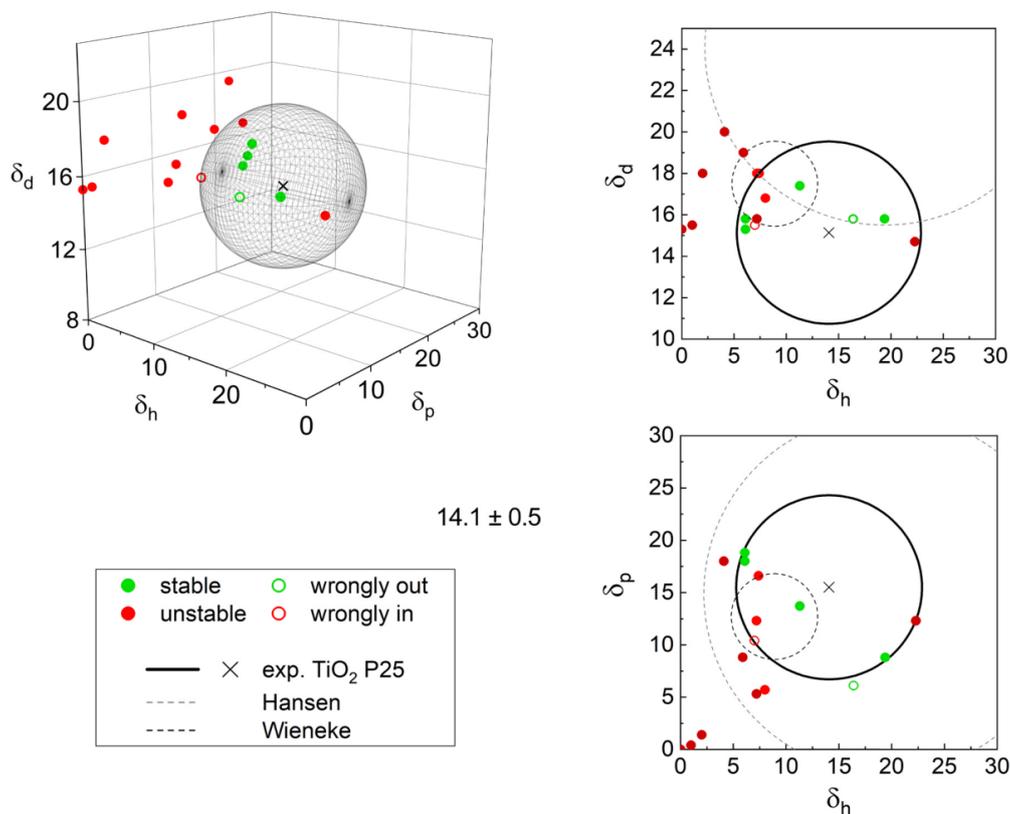
Fig. 8. HSP sphere and 2D projections of TiO$_2$ P25 1 g/L in 17 organic solvents based on Stokes diameters d$_{Stokes}$ at 25 °C.

Table 3

Experimentally determined Hansen parameters (δ_d, δ_p, δ_h), of the center and the radius (R) of the Hansen solubility sphere and calculated Hildebrand parameter (δ_H) compared to literature values.

Method	δ_d [MPa$^{1/2}$]	δ_p [MPa$^{1/2}$]	δ_h [MPa$^{1/2}$]	δ_H [MPa$^{1/2}$]	R [MPa$^{1/2}$]
d$_{Stokes}$	15.1 ± 0.5	15.5 ± 0.8	14.1 ± 0.5	25.8 ± 0.9	8.8
RTSI$_{2h}^{top}$	15.3 ± 0.5	14.4 ± 0.6	13.8 ± 0.5	25.1 ± 0.8	8.8
Hansen [23]	24.1	14.9	19.4	34.3	17.2
Wieneke [18]	17.5	12.7	8.9	23.4	4.1

charged particles in media of low polarity can correctly be interpreted on the basis of HSP. Therefore, it makes no sense to interpret the stability of DLVO stabilized samples using HSP theory. On the other hand, the analysis of the stability of the dispersions jointly by Hansen's and DLVO theories provides complementary information allowing a more accurate interpretation of the results and highlight the solvents exhibiting a significant affinity for the particles' surface.

4. Conclusion

Hansen Solubility Parameters (HSP) have been used for rationalizing and predicting the stability of titanium dioxide particles dispersions [13, 18,23]. However, interparticle electrostatic interactions, not considered in Hansen's approach but considered in the DLVO theory, play a

significant role in organic solvents having a notable dielectric constant [15,56,57]. HSP interpretation for NPs dispersions can be achieved provided that electrostatic repulsion is accounted for, in particular regarding inorganic NPs that can easily acquire surface charges.

In water, it was shown that the Turbiscan detected high *TSI* (Turbiscan Stability Index) and Stokes diameters when zeta potential was elevated, in accordance with the DLVO theory. In organic solvents, it was intended to interpret the stability using HSP. This model was shown insufficient to describe with fidelity the dispersions stability since no sphere could fit the experimental observations. The relevance of DLVO interpretation came in to play as total repulsive energy *(V$_R$ - V$_A$)$_{max}$* was non-negligible in numerous solvents. When the zeta potential of NPs and the dielectric constant of the solvent are both high, the dispersion benefits from additional stabilization while when the electrostatic repulsion is negligible, only the solvents within a Hansen dispersion sphere give stable dispersions. The two interpretations are therefore complementary to describe the behavior of TiO$_2$ dispersions in organic solvents.

Using a zeta dip cell, with higher chemical compatibility than the regular folded capillary cells, it was possible to measure zeta potential in organic solvents and assess the strength of the electrostatic contribution to stabilization. Solvents were scored according to Turbiscan indicators, namely *RTSI$_{2h}^{top}$* (Relative Turbiscan Stability Index after 2 h at the top of the cell) and d$_{Stokes}$ calculated from the sedimentation rate to yield the TiO$_2$ P25 Hansen Solubility Sphere. The sphere (R = 8.8, δ_d = 15.1, δ_p = 15.5, δ_h = 14.1 MPa$^{1/2}$) was compared to that of Hansen [23] and that of Wieneke et al. [18] for TiO$_2$, both determined based on visual scoring. In this study, both zeta measurement and Turbiscan monitoring of stability were complementary. *RTSI$_{2h}^{top}$* and Stokes diameters are quantitative descriptors avoiding an empirical visual rating of dispersions stability.

10

Future HSP determination for other types of NPs such as metal oxides or carbon-based NPs should validate this method.

CRediT authorship contribution statement

Lucie Delforce: Investigation, Writing – original draft. **Evamaria Hofmann:** Investigation. **Véronique Nardello-Rataj:** Funding acquisition, Project administration. **Jean-Marie Aubry:** Conceptualization, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.colsurfa.2021.127333.

References

[1] Commission recommendation of 18 October 2011 on the definition of nanomaterial text with EEA relevance, 2011. http://data.europa.eu/eli/reco/2011/696/oj/eng (Accessed 3 April 2020).

[2] R.C. Monica, R. Cremonini, Nanoparticles and higher plants, Caryologia 62 (2009) 161–165, https://doi.org/10.1080/00087114.2004.10589681.

[3] G.M. Whitesides, Nanoscience, nanotechnology, and chemistry, Small 1 (2005) 172–179, https://doi.org/10.1002/smll.200400130.

[4] S.L. Schneider, H.W. Lim, A review of inorganic UV filters zinc oxide and titanium dioxide, Photodermatol. Photomed. 35 (2019) 442–446, https://doi.org/10.1111/phpp.12439.

[5] S. Yashwant Singh , Nano titanium dioxide market by application (paints & coatings, pigments, cosmetics, plastics, energy and others) - Global opportunity analysis and industry forecast, 2014 - 2022, 2016. https://www.alliedmarketresearch.com/nano-titanium-dioxide-market (Accessed 6 April 2020).

[6] J. Papp, S. Soled, K. Dwight, A. Wold, Surface acidity and photocatalytic activity of TiO₂, WO₃/TiO₂, and MoO₃/TiO₂ photocatalysts, Chem. Mater. 6 (1994) 496–500, https://doi.org/10.1021/cm00040a026.

[7] S. Pilotek , K. Gossmann , F. Tabellion , K. Steingröver , H. Homann , Tailored nanoscaled TiO2 dispersions for photocatalytic applications, in: Technical proceedings of the 2009 clean technology conference and trade show, Houston, USA, 2009: pp. 304–307.

[8] Product information AEROXIDE® TiO₂ P25, 2019. https://products-re.evonik.com/www2/uploads/productfinder/AEROXIDE-TiO2-P-25-EN.pdf (Accessed 24 March 2020).

[9] J. Schulz, H. Hohenberg, F. Pflücker, E. Gärtner, T. Will, S. Pfeiffer, R. Wepf, V. Wendel, H. Gers-Barlag, K.-P. Wittern, Distribution of sunscreens on skin, Adv. Drug Deliv. Rev. 54 (2002) S157–S163, https://doi.org/10.1016/S0169-409X(02)00120-5.

[10] G. Li, L. Lv, H. Fan, J. Ma, Y. Li, Y. Wan, X.S. Zhao, Effect of the agglomeration of TiO₂ nanoparticles on their photocatalytic performance in the aqueous phase, J. Colloid Interface Sci. 348 (2010) 342–347, https://doi.org/10.1016/j.jcis.2010.04.045.

[11] J. Jiang, G. Oberdörster, P. Biswas, Characterization of size, surface charge, and agglomeration state of nanoparticle dispersions for toxicological studies, J. Nanopart. Res. 11 (2009) 77–89, https://doi.org/10.1007/s11051-008-9446-4.

[12] J. Qin, X. Wang, Q. Jiang, M. Cao, Optimizing dispersion, exfoliation, synthesis, and device fabrication of inorganic nanomaterials using Hansen solubility parameters, ChemPhysChem 20 (2019) 1069–1097, https://doi.org/10.1002/cphc.201900110.

[13] S. Süß, T. Sobisch, W. Peukert, D. Lerche, D. Segets, Determination of Hansen parameters for particles: a standardized routine based on analytical centrifugation, Adv. Powder Technol. 29 (2018) 1550–1561, https://doi.org/10.1016/j.apt.2018.03.018.

[14] H. Buron, O. Mengual, G. Meunier, I. Cayré, P. Snabre, Optical characterization of concentrated dispersions: applications to laboratory analyses and on-line process monitoring and control, Polym. Int 53 (2004) 1205–1209, https://doi.org/10.1002/pi.1231.

[15] S.H. Woo, L. Min Gu, C.K. Rhee, Sedimentation properties of TiO₂ nanoparticles in organic solvents, Solid State Phenom. 119 (2007) 267–270, https://doi.org/10.4028/www.scientific.net/SSP.119.267.

[16] Z.Q. Liu, X. Yang, Q. Zhang, Turbiscan: history, development, application to colloids and dispersions, Adv. Mat. Res. 936 (2014) 1592–1596, https://doi.org/10.4028/www.scientific.net/AMR.936.1592.

[17] M. Luo, X. Qi, T. Ren, Y. Huang, A.A. Keller, H. Wang, B. Wu, H. Jin, F. Li, Heteroaggregation of CeO₂ and TiO₂ engineered nanoparticles in the aqueous phase: application of turbiscan stability index and fluorescence excitation-emission matrix (EEM) spectra, Colloids Surf. A 533 (2017) 9–19, https://doi.org/10.1016/j.colsurfa.2017.08.014.

[18] J.U. Wieneke, B. Kommoß, O. Gaer, I. Prykhodko, M. Ulbricht, Systematic investigation of dispersions of unmodified inorganic nanoparticles in organic solvents with focus on the Hansen solubility parameters, Ind. Eng. Chem. Res. 51 (2012) 327–334, https://doi.org/10.1021/ie201973u.

[19] S. Küchler, T. Detloff, T. Sobisch, D. Lerche, Direct and accelerated characterization of ceramic dispersions, Ceram. Forum Int. CFI Ber. DKG 88 (2011) (2011) E27–E31.

[20] L. Xu, H.-W. Liang, Y. Yang, S.-H. Yu, Stability and reactivity: positive and negative aspects for nanoparticle processing, Chem. Rev. 118 (2018) 3209–3250, https://doi.org/10.1021/acs.chemrev.7b00208.

[21] J.N. Israelachvili, Intermolecular and Surface Forces, third ed., Academic Press, 2011. Revised.

[22] M. Hosokawa, K. Nogi, M. Naito, T. Yokoyama, Nanoparticle Technology Handbook, Second ed., Elsevier, Amsterdam, 2012.

[23] C.M. Hansen, The Three Dimensional Solubility Parameter and Solvent Diffusion Coefficient, Danish Technical Press, Copenhagen, 1967. https://hansen-solubility.com/contents/HSP1967-OCR.pdf.

[24] C. Hansen , S. Abbott , H. Yamamoto , HSP Examples: Nanoparticles | Hansen Solubility Parameters, 2020. https://www.hansen-solubility.com/HSP-examples/nanoparticles.php (Accessed 9 April 2020).

[25] L.M. Wheeler, N.J. Kramer, U.R. Kortshagen, Thermodynamic driving force in the spontaneous formation of inorganic nanoparticle solutions, Nano Lett. 18 (2018) 1888–1895, https://doi.org/10.1021/acs.nanolett.7b05187.

[26] C. Stauch, S. Süß, R. Luxenhofer, B.P. Binks, D. Segets, K. Mandel, Quantifying surface properties of silica particles by combining hansen parameters and Reichardt's dye indicator data, Part. Part. Syst. Charact. 35 (2018), 1800328, https://doi.org/10.1002/ppsc.201800328.

[27] H. Launay, C.M. Hansen, K. Almdal, Hansen solubility parameters for a carbon fiber/epoxy composite, Carbon 45 (2007) 2859–2865, https://doi.org/10.1016/j.carbon.2007.10.011.

[28] C.M. Hansen, A.L. Smith, Using Hansen solubility parameters to correlate solubility of C60 fullerene in organic solvents and in polymers, Carbon 42 (2004) 1591–1597, https://doi.org/10.1016/j.carbon.2004.02.011.

[29] Y. Hernandez, M. Lotya, D. Rickard, S.D. Bergin, J.N. Coleman, Measurement of multicomponent solubility parameters for graphene facilitates solvent discovery, Langmuir 26 (2010) 3208–3213, https://doi.org/10.1021/la903188a.

[30] H.T. Ham, Y.S. Choi, I.J. Chung, An explanation of dispersion states of single-walled carbon nanotubes in solvents and aqueous surfactant solutions using solubility parameters, J. Colloid Interface Sci. 286 (2005) 216–223, https://doi.org/10.1016/j.jcis.2005.01.002.

[31] K. Maleski, V.N. Mochalin, Y. Gogotsi, Dispersions of two-dimensional titanium carbide MXene in organic solvents, Chem. Mater. 29 (2017) 1632–1640, https://doi.org/10.1021/acs.chemmater.6b04830.

[32] S. Mathioudaki, B. Barthélémy, S. Detriche, C. Vandenabeele, J. Delhalle, Z. Mekhalif, S. Lucas, Plasma treatment of metal oxide nanoparticles: development of core–shell structures for a better and similar dispersibility, ACS Appl. Nano Mater. 1 (2018) 3464–3473, https://doi.org/10.1021/acsanm.8b00645.

[33] AEROXIDE®, AERODISP® and AEROPERL® titanium dioxide as photocatalyst - Technical information 1243, 2015. https://www.aerosil.com/sites/lists/RE/DocumentsSI/TI-1243-Titanium-Dioxide-as-Photocatalyst-EN.pdf (Accessed 3 April 2020).

[34] Technical overview 13 - AEROXIDE® - Fumed metal oxides, 2017. https://www.aerosil.com/product/aerosil/downloads/to-aeroxide-en.pdf (Accessed 24 March 2020).

[35] Application note TS_STAB_60 Explanation to the TSI calculation, 2019.

[36] P. Mills, P. Snabre, Settling of a suspension of hard spheres, Europhys. Lett. EPL 25 (1994) 651–656, https://doi.org/10.1209/0295-5075/25/9/003.

[37] Application note - Zeta potential measurements of non-aqueous particulate suspensions, 2010.

[38] I.M. Tucker, J.C.W. Corbett, J. Fatkin, R.O. Jack, M. Kaszuba, B. MacCreath, F. McNeil-Watson, Laser Doppler Electrophoresis applied to colloids and surfaces, Curr. Opin. Colloid Interface Sci. 20 (2015) 215–226, https://doi.org/10.1016/j.cocis.2015.07.001.

[39] I.D. Morrison, S. Ross, Colloidal Dispersions: Suspensions, Emulsions, and Foams, Wiley, 2002.

[40] R.J. Hunter, Zeta Potential in Colloid Science: Principles and Applications, Academic Press, New York, 1981.

[41] M. Kosmulski, Zeta potentials in nonaqueous media: how to measure and control them, Colloids Surf. A 159 (1999) 277–281, https://doi.org/10.1016/S0927-7757(99)00273-3.

11

[42] M. Kosmulski, P. Eriksson, J.B. Rosenholm, Application of zetametry to determine concentrations of acidic and basic impurities in analytical reagents, Anal. Chem. 71 (1999) 2518–2522, https://doi.org/10.1021/ac9806052.

[43] M. Kosmulski, Surface Charging and Points of Zero Charge, CRC Press, 2009, https://doi.org/10.1201/9781420051896.

[44] K. Bourikas, T. Hiemstra, W.H. Van Riemsdijk, Ion pair formation and primary charging behavior of titanium oxide (anatase and rutile), Langmuir (2001) 749–756, https://doi.org/10.1021/la000806c.

[45] A. Foissy, A. M'Pandou, J.M. Lamarche, N. Jaffrezic-Renault, Surface and diffuse-layer charge at the TiO$_2$-electrolyte interface, Colloids Surf. 5 (1982) 363–368, https://doi.org/10.1016/0166-6622(82)80046-2.

[46] B. Derjaguin, L. Landau, Theory of the stability of strongly charged lyophobic sols and of the adhesion of strongly charged particles in solutions of electrolytes, Acta Physicochim. U. R. S. S 14 (1941) 633–662.

[47] E.J.W. Verwey, J.T.G. Overbeek, K. van Nes, Theory of the Stability of Lyophobic Colloids: The Interaction of Sol Particles Having an Electric Double Layer, Elsevier Publishing Company, Amsterdam, 1948.

[48] K. Suttiponparnit, J. Jiang, M. Sahu, S. Suvachittanont, T. Charinpanitkul, P. Biswas, Role of surface area, primary particle size, and crystal phase on titanium dioxide nanoparticle dispersion properties, Nanoscale Res. Lett. 6 (2011) 27, https://doi.org/10.1007/s11671-010-9772-1.

[49] J. Qi, Y.Y. Ye, J.J. Wu, H.T. Wang, F.T. Li, Dispersion and stability of titanium dioxide nanoparticles in aqueous suspension: effects of ultrasonication and concentration, Water Sci. Technol. 67 (2013) 147–151, https://doi.org/10.2166/wst.2012.545.

[50] R. Beranek, Photo)electrochemical methods for the determination of the band edge positions of TiO2-based nanomaterials, Adv. Phys. Chem. 2011 (2011) 1–20, https://doi.org/10.1155/2011/786759.

[51] M. Herrmann, H.P. Boehm, Über die Chemie der Oberfläche des Titandioxids. ii. Saure Hydroxylgruppen auf der Oberfläche, Z. Anorg. Allg. Chem. 368 (1969) 73–86, https://doi.org/10.1002/zaac.19693680111.

[52] M.A. Butler, D.S. Ginley, Prediction of flatband potentials at semiconductor-electrolyte interfaces from atomic electronegativities, J. Electrochem. Soc. 125 (1978) 228–232, https://doi.org/10.1149/1.2131419.

[53] C. Kormann, D.W. Bahnemann, M.R. Hoffmann, Photolysis of chloroform and other organic molecules in aqueous titanium dioxide suspensions, Environ. Sci. Technol. 25 (1991) 494–500, https://doi.org/10.1021/es00015a018.

[54] J. Lyklema, Principles of the stability of lyophobic colloidal dispersions in non-aqueous media, Adv. Colloid Interface Sci. 2 (1968) 67–114, https://doi.org/10.1016/0001-8686(68)85001-8.

[55] R. Hogg, T.W. Healy, D.W. Fuerstenau, Mutual coagulation of colloidal dispersions, Trans. Faraday Soc. 62 (1966) 1638–1651, https://doi.org/10.1039/TF9666201638.

[56] J.B. Rosenholm, P. Dahlsten, Influence of charge exchange in acidic aqueous and alcoholic titania dispersions on viscosity, Adv. Colloid Interface Sci. 226 (2015) 138–165, https://doi.org/10.1016/j.cis.2015.10.007.

[57] N.D. Burrows, E. Kesselman, K. Sabyrov, A. Stemig, Y. Talmon, R. Lee Penn, Crystalline nanoparticle aggregation in non-aqueous solvents, CrystEngComm 16 (2014) 1472–1481, https://doi.org/10.1039/C3CE41584H.

[58] K. Izutsu, Acid–base reactions in nonaqueous solvents, in: Electrochemistry in Nonaqueous Solutions, John Wiley & Sons Ltd, 2009, pp. 63–87, https://doi.org/10.1002/9783527629152.ch3.

[59] J. Barbosa, C.M. Bosch, V. Sanz-Nebot, Effect of the solvent on the equilibria of acid-base indicators in aprotic and amphiprotic solvents, Mikrochim. Acta 106 (1992) 327–337, https://doi.org/10.1007/BF01242105.

[60] O. Popovych, Transfer activity coefficients of ions in methanol-water solvents based on the tetraphenylborate assumption, J. Phys. Chem. 88 (1984) 4167–4170, https://doi.org/10.1021/j150662a063.

[61] M. Kosmulski, Zeta potential of anatase (TiO$_2$) in mixed solvents, Colloids Surf. 64 (1992) 57–65, https://doi.org/10.1016/0166-6622(92)80162-U.

[62] W. Janusz, A. Sworska, J. Szczypa, The structure of the electrical double layer at the titanium dioxide/ethanol solutions interface, Colloids Surf. A 152 (1999) 223–233, https://doi.org/10.1016/S0927-7757(98)00341-0.

[63] S.H. Behrens, D.I. Christl, R. Emmerzael, P. Schurtenberger, M. Borkovec, Charging and aggregation properties of carboxyl latex particles: experiments versus DLVO theory, Langmuir 16 (2000) 2566–2575, https://doi.org/10.1021/la991154z.

[64] E. Hückel, P. Debye, Zur Theorie der Elektrolyte. I. Gefrierpunktserniedrigung und verwandte Erscheinungen, Phys. Z. 24 (1923) 185–206.

[65] H.C. Hamaker, The London-van der Waals attraction between spherical particles, Physica 4 (1937) 1058–1072, https://doi.org/10.1016/S0031-8914(37)80203-7.

[66] T. Tadros, Chapter 2 - Colloid and interface aspects of pharmaceutical science, in: H. Ohshima, K. Makino (Eds.), Colloid and Interface Science in Pharmaceutical Research and Development, Elsevier, Amsterdam, 2014, pp. 29–54, https://doi.org/10.1016/B978-0-444-62614-1.00002-8.

182

ACS OMEGA

http://pubs.acs.org/journal/acsodf

Article

Fast Prediction of the Equivalent Alkane Carbon Number Using Graph Machines and Neural Networks

Lucie Delforce, François Duprat,* Jean-Luc Ploix, Jesus Fermín Ontiveros, Valentin Goussard, Véronique Nardello-Rataj,* and Jean-Marie Aubry*
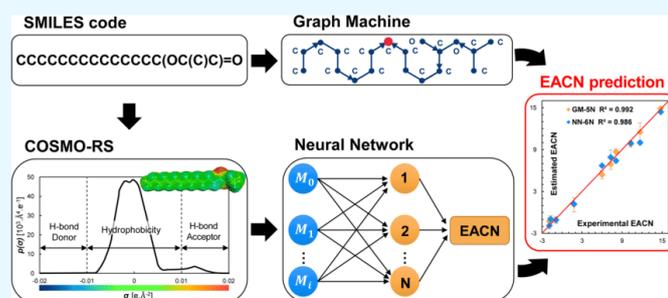
Cite This: https://doi.org/10.1021/acsomega.2c04592

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The hydrophobicity of oils is a key parameter to design surfactant/oil/water (SOW) macro-, micro-, or nano-dispersed systems with the desired features. This essential physicochemical characteristic is quantitatively expressed by the equivalent alkane carbon number (EACN) whose experimental determination is tedious since it requires knowledge of the phase behavior of the SOW systems at different temperatures and for different surfactant concentrations. In this work, two mathematical models are proposed for the rapid prediction of the EACN of oils. They have been designed using artificial intelligence (machine-learning) methods, namely, neural networks (NN) and graph machines (GM). While the GM model is implemented from the SMILES codes of a 111-molecule training set of known EACN values, the NN model is fed with some σ-moment descriptors computed with the COSMOtherm software for the 111-molecule set. In a preliminary step, the leave-one-out algorithm is used to select, given the available data, the appropriate complexity of the two models. A comparison of the EACNs of liquids of a fresh set of 10 complex cosmetic and perfumery molecules shows that the two approaches provide comparable results in terms of accuracy and reliability. Finally, the NN and GM models are applied to nine series of homologous compounds, for which the GM model results are in better agreement with the experimental EACN trends than the NN model predictions. The results obtained by the GMs and by the NN based on σ-moments can be duplicated with the demonstration tool available for download as detailed in the Supporting Information.

1. INTRODUCTION

A large diversity of natural and synthetic liquid compounds are grouped under the generic term "oils" due to their non-miscibility with water. They are key components of surfactant/oil/water (SOW) systems such as swollen micelles, micro-emulsions, or emulsions which are found in numerous end-use products and various fields such as cosmetics, pharmaceutics, food, crude oil, and so forth. The quantitative evaluation of the hydrophobicity of the oil is extremely important because it allows for choosing the most effective SOW system composition and, *in fine*, optimizing its performances in applications.

Several concepts have thus been developed to characterize the hydrophobicity/polarity of oils. The best known are log P, that is, the logarithm of the *n*-octanol−water partition coefficient[1] and the "required HLB" (hydrophilic−lipophilic

balance).[2,3] Log P is widely used in environmental and medical sciences since it expresses the ability of a non-ionizable substance to partition between aqueous and lipophilic compartments of organisms, but it gives no indication regarding its behavior at the O/W interfaces in SOW systems.[1] The required HLB value of an oil, introduced by Griffin,[2,3] was designed to that aim. It corresponds to the HLB value of the mixture of surfactants providing the most stable emulsion with
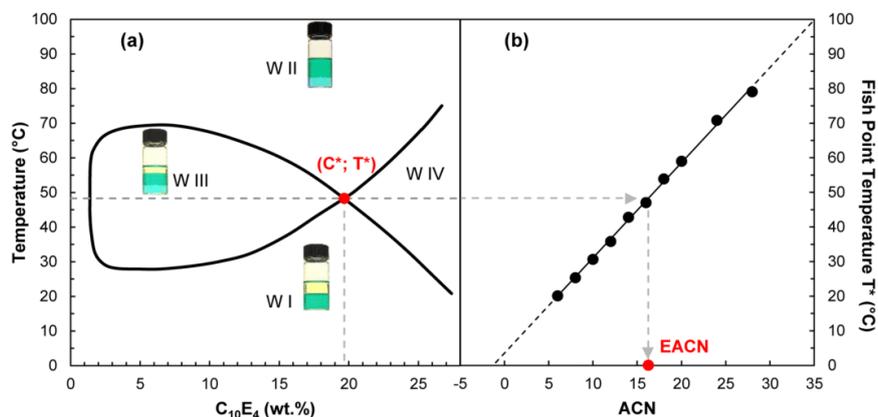
A

Figure 1. Determination of the EACN of an oil from the fish plot of the $C_{10}E_4$/oil/water-T system (a). The temperature of the fish-tail point indicated in red is reported to the calibration straight line obtained with $C_{10}E_4$/n-alkanes/water-T systems (b).

the oil under study.[2,4] This method is sometimes used to classify complex mixtures such as essential, vegetable, or animal oils.[5−9] However, the required HLBs thus obtained are imprecise and poorly reproducible because they depend on the mixture of surfactants and on the emulsification process. Furthermore, the required HLB concept is based on the HLB of surfactants which is itself an approximate empirical parameter only reliable for polyethoxylated nonionic surfactants.

A more relevant concept to quantify the hydrophobicity of an oil in SOW systems was introduced in 1977 by Wade et al. as a dimensionless number: the so-called equivalent alkane carbon number (EACN).[10] It corresponds to the number of carbon atoms of the n-alkane that exhibit a phase behavior similar to that of the oil under consideration.[11] In practice, measuring accurately the EACN value of an oil is tedious. The standard method is based on the elaboration of the so-called "fish diagrams," which represent the phase behavior of equilibrated SOW systems where S is a well-defined polyethyleneglycol monoalkyl ethers (C_iE_j for $CH_3(CH_2)_{i−1}O$-$(CH_2CH_2O)_jH$) and the water-to-oil ratio (WOR) is equal to 1.[12]

A typical fish diagram of a system $C_{10}E_4$/oil/water is given in Figure 1a as an example. When varying the temperature and the surfactant concentration, the C_iE_j/oil/water-T systems provide different types of microemulsion behaviors, called Winsor phases (Winsor I, II, III, and IV) depending on the affinity of the surfactant for water and oil. When it is balanced, a three-phase system (Winsor III) is formed, giving a diagram shaped like a fish. The characteristic temperature T^* at the intersection of the Winsor I to IV (i.e., one single phase microemulsion) regions is then compared to the T^* values of a series of n-alkanes (Figure 1b) to determine the EACN of the oil which expresses its hydrophobicity.[11,13]

While reliable and accurate, the experimental determination of EACNs from fish tail diagrams is, however, a lengthy process which is limited by experimentally accessible conditions in terms of temperature ($T \approx 5−80$ °C). Thus, in silico estimation of the EACN values of oils without any experiments would be considerably time-saving.

To date, a few predictive models of EACN values have been reported. The EACN value of complex oil mixtures, that is, crude oil, was predicted by Creton et al. using an evolutionary algorithm coupled to data mining.[14] Bouton et al. built a QSPR model by applying genetic algorithms to structural molecular descriptors of polar hydrocarbon oils.[15] A multilinear regression based on the σ-moments calculated by the conductor-like screening model for real solvents (COSMO-RS) approach[16,17] was applied to polar hydrocarbons and aprotic polar oils by Lukowicz et al.[18,19] These works showed that depending on the chemical functions of molecules, the relevant descriptors differ and EACN estimations were less satisfactory in the case of polar oils. Building a QSPR model relies on finding the best relation between a group of descriptors and a target property. Those models require the construction of a reliable database, consisting of entry/output pairs where entries are molecular descriptors and outputs are the target properties. Numerous predictive methods, based on linear and nonlinear approaches, have been applied to a diversity of physical and chemical properties.[20] The linear methods, such as MLR, principal component regression, and partial least-squares regression, are often used. However, nonlinear models can be adapted to translate more complex relations between descriptors and the predicted property, making them more efficient predictors in some cases.[20−22] Rather than using methods that select descriptors from a large pull of automatically generated variables,[21] we have chosen to start with descriptors that have a physical meaning, such as σ-moments (computed with COSMOtherm) and 2D structures of molecules used as descriptors for graph machines (GM).[23] Already, a variety of chemical and physicochemical properties such as boiling point of halogenated hydrocarbons,[24] surface tension,[25] viscosity,[26] flash point, cetane number of fuels,[27] bioactivity of drugs,[24,28] and other thermodynamic properties[29−31] can be predicted accurately with GM and neural networks (NN) inputted with 2D-structures and σ-moments, respectively. Both these theoretical tools are non-linear models that learn a pathway from input values to a resulting output. For NNs, that are basically standard multi-layer perceptrons (MLPs), the inputs are either measured or computed from molecular simulations, while for GMs, the inputs are the 2D molecular structures entered as their SMILES (simplified molecular input line entry specification) codes.

In this work, we report on two approaches for predicting the EACN of functionalized oils using NN and GM. To that goal, a set of 111 molecules with a reliable experimental EACN was

184

gathered either from literature or from our database.[11,15,18,19,32−34] A GM regression based solely on the readily accessible molecular SMILES codes and a NN regression using as inputs COSMO-RS-computed σ-moments are designed for the 111 molecules. After a selection step of the optimal model in each case, predictions are performed on a test set of 10 cosmetic or perfumery molecules for which experimental EACNs have been determined. The respective reliability of the two models is finally evaluated by predicting the EACN of compounds belonging to nine homologous series.

2. MATERIALS AND METHODS

2.1. Database Construction. A set of 121 compounds with reliable EACN values either extracted from literature or determined experimentally in our laboratory was assembled. These averaged EACN values are reported in Table S1 of the Supporting Information. The whole set includes *n*-alkanes, esters, ethers, ketones, alkenes, alkynes, cyclic hydrocarbons, aromatics, branched hydrocarbons, nitriles, chloroalkanes, and consequently compounds containing carbon, hydrogen, oxygen, nitrogen, and chloride atoms. In addition to the commercially available series which are mainly linear compounds, the data set contains molecules of varying complexity, which are characterized by their structural features, that is, the presence of a double bond (alkene), ring, branching, and functional group. For training and testing purposes, this set was divided into a training set of 111 compounds and a test set of 10 compounds, including five cosmetic-type oils and five perfume-type oils. Among the five cosmetic oils, three compounds (hemisqualane, dioctylether, and isopropyl myristate) were already members of the test sets of our previous papers.[25,26] Isododecane, a petroleum-based cosmetic, has recently been synthetized under environmentally friendly conditions, becoming thus an important molecule for cosmetic formulators and manufacturers. Finally, octyloctanoate was selected because it is the only ester with a functional group in the middle of its carbon skeleton. As for the five fragrance molecules, they were chosen because they have several structural characteristics: (i) one or two cycles, double bonds, and branching (limonene and caryophyllene), (ii) double bonds, branching, and an ester group (linalyl acetate), and (iii) a cycle, one or two double bonds, branching, and an ether or ketone function (rose oxide and β-ionone). In addition, the EACN values of these 10 compounds, which have been determined in our group to have consistent values, are fairly well distributed over the range of property values (−4 to 20). The distributions of the structural features present in both data sets are displayed in Figure 2, indicating clearly the challenging complexity of the test set molecules.

2.2. EACN Experimental Determination. *2.2.1. Chemicals.* Oils for which EACN values were measured in this work are presented in Table 1 and were used as such. Pure tetraethyleneglycol monodecyl ether ($C_{10}E_4$) was synthesized according to a method described elsewhere.[35,36] Its purity was assessed by GC−MS analysis (>99%) and by comparing its cloud point temperature at 2.6 wt % (20.4 vs 20.6 °C) with the reference value.[37] Tetraethyleneglycol monohexyl ether (C_6E_4) was synthetized using an analogous method to C10E4, and its cloud point temperature (66.2 °C at 16.4 wt %) was compared to the reference value (66.1 °C at 16.4 wt %).[38]

2.2.2. Phase Diagrams. In order to enrich the EACN values database, experimental phase diagrams were built, in particular
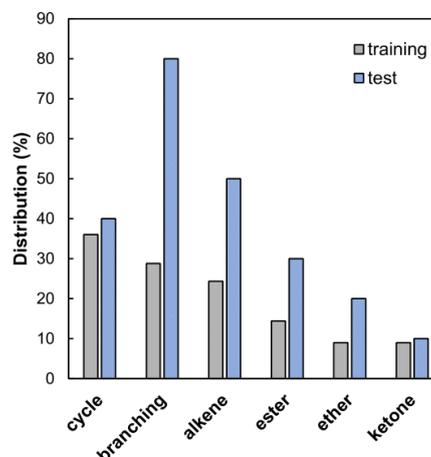

Figure 2. Distribution of common structural features (in percent) of molecules in the training and test sets of the EACN database.

Table 1. Oils for Which EACN Values Were Measured in This Work, Commercial Name, Supplier, and Purity

compound	supplier	purity
isoamyl laurate, JOLEE 7750	Oleon	100%
hemisqualane, Neossance	Amyris	>95%
isohexadecane, 2,2,4,6,8-heptamethylnonane	Sigma-Aldrich	98%
pristane, 2,6,10,14-tetramethylpentadecane	TCI	>95%
isododecane, 2,2,4,6,6-pentamethylheptane	TCI	>98%
2-methylpentane	Sigma-Aldrich	>99%
3-methylpentane	Sigma-Aldrich	>99%
2,3-dimethylbutane	Sigma-Aldrich	98%
isooctane, 2,2,4-trimethylpentane	Sigma-Aldrich	>99%
dipropyl ether	Sigma-Aldrich	>99%
diisopropyl ether	Sigma-Aldrich	>98.5%

in the case of branched alkanes that were under-represented in the literature data. The experimental EACN value was determined by establishing the phase behavior of 50 ± 0.2 wt % water/oil mixtures at different $C_{10}E_4$ or C_6E_4 concentrations as a function of temperature. The Winsor systems were determined by visual observation.[39] The most volatile oil samples (2-methylpentane, 3-methylpentane, and 2,3-dimethylbutane) were weighed in glass tubes, placed in liquid nitrogen, and then sealed with a flame. Other samples were prepared in glass tubes closed by screw caps. Samples were first shaken gently several times and left in a thermoregulated bath at T ± 0.1 °C until equilibration. The point (C^*; T^*) corresponding to the intersection of the Winsor III and the Winsor IV phases was used to determine the oil's EACN: its T^* value was reported on the T^* versus ACN reference straight line for linear alkanes using either $C_{10}E_4$ or C_6E_4 as the surfactant.[11] The fish diagram of hemisqualane is given as an example in Figure 3a. The fish diagram lower concentration limit was determined by extrapolation of Winsor III phase relative volume as described by Burauer et al.[40] Other experimentally determined (C^*; T^*) points using $C_{10}E_4$ are represented in Figure 3b. Fish diagrams of dipropylether and diisopropyl ether determined with C_6E_4 as a surfactant are available in Figures S1 and S2 of the Supporting Information.
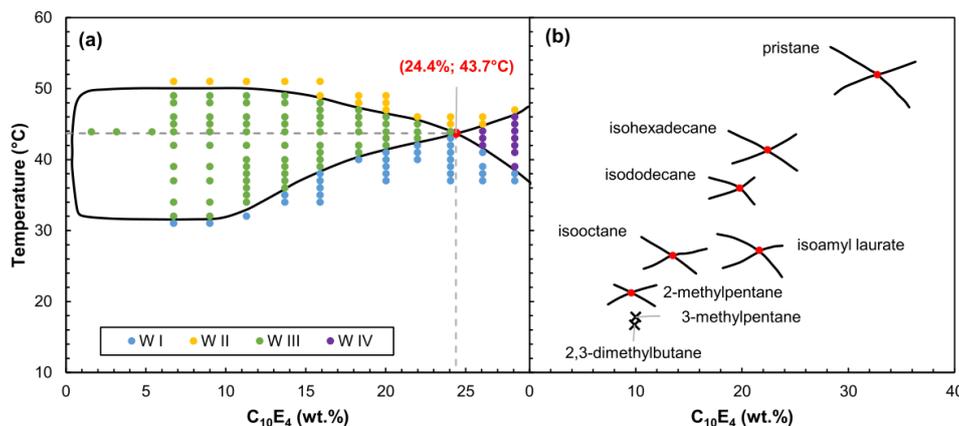
C

Figure 3. (a) Experimental fish plot of the $C_{10}E_4$/hemisqualane/water-T system at a water/oil ratio equal to 1 (w/w) and (b) partial fish plot and fish tail points $(C^*; T^*)$ determined with $C_{10}E_4$ for pristane (2,6,10,14-tetramethylpentadecane), isohexadecane (2,2,4,4,6,6,8-heptamethylnonane), isododecane (2,2,4,6,6-pentamethylheptane), isooctane (2,2,4-trimethylpentane), isoamyl laurate (3-methylbutyl dodecanoate), and 2-methylpentane. Fish tail points $(C^*; T^*)$ for 3-methylpentane and 2,3-dimethylbutane are represented by cross-marks for clarity.
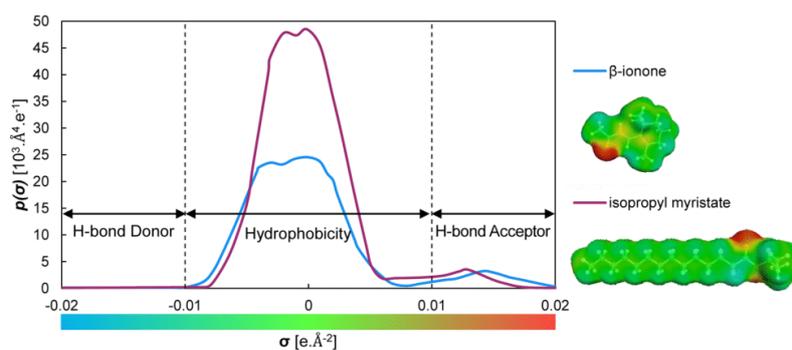
Figure 4. σ-Profiles and σ-surfaces of β-ionone (in blue) and isopropyl myristate (in purple). The color gradient corresponds to the surface charge density σ.

2.3. COSMO-RS σ-Moment Calculation. COSMO-RS is a first-principles theoretical model based on a combination of quantum chemistry and statistical thermodynamics that serves to estimate, without any prior experience, a large number of chemical properties.[17,41] Due to the presence of polar covalent bonds, molecules carry a surface charge density σ on its so-called "σ-surface", which corresponds to the slightly inflated van der Waals surface. The "σ-profile" $p^X(\sigma)$ of a molecule X is the curve obtained by smoothing the histogram of surface portions grouped by charge density in the interval $[\sigma - d\sigma/2, \sigma + d\sigma/2]$.[16] Examples in the case of β-ionone and isopropyl myristate are represented in Figure 4. Using the COSMO-conf software (version 4.3), the lower energy conformations in the bulk liquid state are calculated for all molecules. These conformations are then used as inputs in the COSMOtherm software (version 19.0.4), allowing for the calculation of the σ-surface, σ-potential, and σ-moments. Klamt[17] has shown that any partition coefficient K can be very well expressed as a Taylor-like development of σ-moments as defined by eq 1. It is estimated that a development up to m equal to six σ-moments is sufficient to satisfactorily express the partition coefficient K according to eq 1.

$$RT \ \ln K = c_{\text{acc}}M_{\text{acc}}^X + c_{\text{don}}M_{\text{don}}^X + \sum_{i=0}^{m} c_i M_i^X \tag{1}$$

The σ-moments M_i^X are calculated from the σ-profile $p^X(\sigma)$ of the studied compound X according to eqs 2–4.

$$M_{\text{acc}}^X = \int_{+\sigma_{\text{HB}}}^{+\infty} p^X(\sigma)(\sigma - \sigma_{\text{HB}}) \, d\sigma \tag{2}$$

$$M_{\text{don}}^X = \int_{-\infty}^{-\sigma_{\text{HB}}} p^X(\sigma)(-\sigma - \sigma_{\text{HB}}) \, d\sigma \tag{3}$$

$$M_i^X = \int_{-\infty}^{+\infty} p^X(\sigma)\sigma^i \, d\sigma \tag{4}$$

The first σ-moments have a simple physical meaning: the zero-order σ-moment M_0^X is the surface area of the molecule, expressed in Å^2. The first-order one M_1^X is the polarization charge of this surface, expressed in e. For uncharged molecules, this moment is equal to zero. The second-order σ-moment M_2^X, expressed in e$^2 \cdot$Å^{-2}, is the polarity of the molecule.[42] The third-order M_3^X represents the asymmetry of the σ-profile $p^X(\sigma)$. The other σ-moments up to M_6^X have no particular
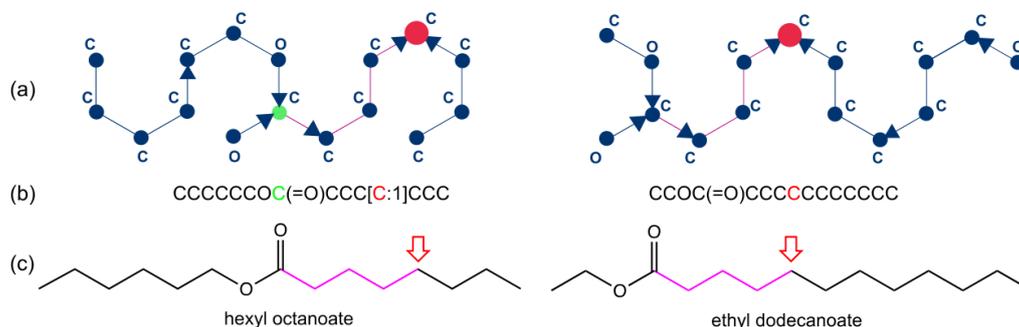
D

Figure 5. Encoding hexyl octanoate and ethyl dodecanoate into directed graphs: (a) directed graphs with root nodes in red; the root node position computed automatically for hexyl octanoate is colored in green and the path between the functional atom node and the root node in pink, (b) SMILES codes with the expected position of the root nodes indicated in red and (c) 2D formulas with expected positions for the root nodes (red arrows). The atom types C and O added on the directed graphs correspond to node labels that are inputs of the parameterized function implemented at each node of the graph to build the GM.

physical meanings. Finally, M_{acc}^X and M_{don}^X, expressed in e (unit equal to the charge of one electron), are the "hydrogen-bonding" σ-moments representing the ability of the molecule to interact with hydrogen-bond acceptors and donors, respectively. Their value is non-zero when the σ-profile outranges the $[-\sigma_{HB}, +\sigma_{HB}]$ interval, where σ_{HB}, the hydrogen-bond threshold, is equal to 0.01 e·Å^{-2}, as shown in Figure 4.

Neither β-ionone nor isopropyl myristate exhibit Lewis acidity corresponding to the hydrogen-bond donor region. However, both of them have a Lewis basicity with non-zero value σ-profile in the hydrogen-bond acceptor region. This is due to the presence of the ester and carbonyl functions inducing locally electron-rich surface areas (in red in both molecules according to the color scale in Figure 4). Finally, the central part of the σ-profile shows higher hydrophobicity in the case of isopropyl myristate than for β-ionone, which is in accordance with its longer alkyl moiety.

2.4. GM and NN Model Selection. As briefly stated in the introduction, GM are regression or classification models that estimate a property directly from the topological information provided by their SMILES codes. In these models, molecules are described as directed acyclic graphs derived from their 2D structures and the parameterized functions that compute the estimate of the property of interest reflect the molecular structures of the compounds.[28,43] As usual in regression or classification models, GM parameters are computed by learning from examples present in an experimental value database.[28] Basically, NN models are multiple non-linear regressions that estimate an output value of a property of interest from some input descriptors values, hereafter three σ-moments selected from a pull of eight σ-moments, all computed with COSMO-RS according to a procedure described in Section 2.3. [A selection of the σ-moments was performed with Metagen, a homemade software package written in Python. Feature selection by the random probe method showed that for our data, M_0^X, M_2^X, and M_3^X are most relevant for EACN estimation; n was also selected as relevant when added to the pull of σ-moment descriptors.][44] Both NN and GM models are built from MLPs that contain a single hidden layer of neurons. The complexity of the models is consequently dependent on the number of neurons of that layer, and along with this, on the number of parameters of the

models. Since, for a given number of neurons in the MLPs, NN and GM models have a different number of parameters, the latter variable will be preferred as a complexity equivalent in the model complexity selection (Section 3.2).

The selection of a model is a key step in machine learning model design: it consists of finding the model complexity, given the available data for designing it, that will result in the best generalization. To that end, with the 111-molecule set available, trainings are carried out with an increasing number of MLP hidden neurons. The ability of both models to account for the training data is monitored with the root mean square training error (RMSTE) that is computed as follows

$$\text{RMSTE} = \sqrt{\frac{1}{111}\sum_{i=1}^{111}\left(\text{EACN}_{\text{exp.}}^i - \text{EACN}_{\text{est.}}^i\right)^2} \quad (5)$$

where $\text{EACN}_{\text{exp.}}^i$ is the EACN value determined experimentally for molecule i, and $\text{EACN}_{\text{est.}}^i$ is the EACN value estimated by the model for molecule i at the end of the training.

The estimation of the generalization error for model selection is usually performed by two methods: the computation of the leave-one-out (LOO) score and the computation of the virtual LOO (VLOO). The computation of the LOO score was chosen for the determination of the optimal complexity, since the VLOO score, that is a first order approximation of the LOO score, is less accurate for small size data sets.[26,45] At the end of the LOO process, the LOO is computed as

$$\text{LOO score} = \sqrt{\frac{1}{111}\sum_{i=1}^{111}\left(\text{EACN}_{\text{exp.}}^i - \text{EACN}_{\text{pred.}}^i\right)^2} \quad (6)$$

where $EACN_{\text{exp.}}^i$ is the EACN value determined experimentally for molecule i, and $EACN_{\text{pred.}}^i$ is the average EACN prediction value computed for the left out molecule i with 50 models having different initialization parameters. The abovementioned equation is the same as eq 5 defining the RMSTE, except that now a true prediction is performed for every molecule, since the molecule i does not belong to the training set. The LOO computation is repeated five times for each complexity of the NN and GM based models, so that the average results are presented.
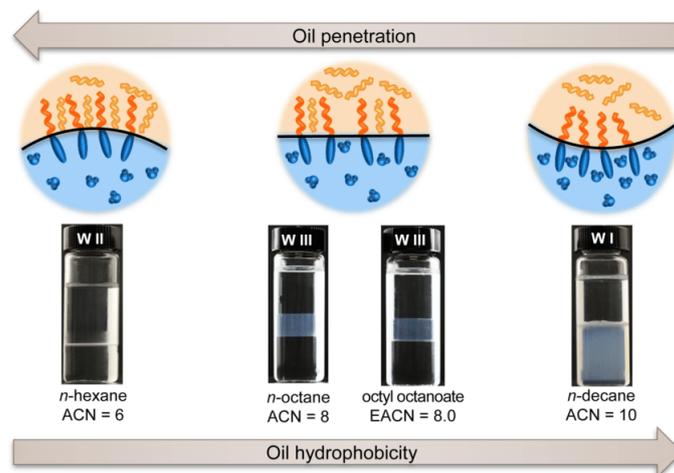
Figure 6. Effect of oil penetration on the spontaneous curvature of the interfacial film (top) and $C_{10}E_4$/oil/water [WOR = 1 (v/v)] microemulsion systems equilibrated at 25.0 °C yielding Winsor II, Winsor III, and Winsor I microemulsions (bottom). Systems with *n*-alkanes contain 3% C10E4, and the one with octyl octanoate contains 7% $C_{10}E_4$.

A few difficulties have been met with the first modeling experiments that have been addressed as follows. With the σ-moment-based NN models, a large EACN deviation was observed specifically for the 15 molecules of the *n*-alkane family, regardless of the complexity of the MLPs used. We found that adding the number of carbon atoms (n) for every molecule as a fourth descriptor corrected this problem. Similarly, an important deviation from the experimental value was exclusively observed for the hexyl octanoate EACN estimation with the GM-based models. A thorough analysis of the GM construction for this molecule indicated that it was not consistent with the construction of the other linear ester GMs (e.g., the ethyl esters). The input code for ethyl hexanoate was particularized so that the constructions were uniform for all esters. This modification is explained in Figure 5 where the directed graphs for hexyl octanoate and ethyl dodecanoate are shown (a). These graphs, which are encoded from their SMILES codes (b), are isomorph to the 2D formula also represented (c). Without modification of the hexyl octanoate SMILES code, the central node of the resulting graph, also called root node, would have been located on the green node. Thanks to the special bracketed tag in the hexyl octanoate SMILES code, the graph red root nodes have now a consistent position in both graphs: the root nodes are equally connected to nodes with a carbon type label and are at the same distance of the functional node, which is the one connected to the nodes with an oxygen-type label. The position of the root node is important since it corresponds to the GM (not represented) output neuron that computes the estimated EACN value. As a result, the estimation of the EACN was much efficient for the hexyl octanoate compound.

We did not meet such a particular case with our published models designed for surface tension and viscosity estimations.[25,26] This is probably because we used a larger training set counting many esters of various sizes and positions for the functional group. For this work, we could only build a data set of moderate size, and among the dozen of linear esters that have an alkyl chain of 10 carbon atoms or more, hexyl octanoate is the unique compound of similar size in the training set to have a functional group in the middle of its carbon skeleton. To illustrate this exception, a representation of the GM can be computed with the demo software as detailed in the Supporting Information.

3. RESULTS AND DISCUSSION

The EACN of an oil quantitatively expresses its hydrophobicity and corresponds to the related but obsolete concept of "required HLB" of oils introduced by Griffin in 1949.[46] The more hydrophobic the oil, the higher its EACN. In particular, for the *n*-alkane series, the EACN is, by definition, equal to the carbon number of the alkane and is denoted ACN. Conversely, the more polar the oil, the lower its EACN and the EACN can even be negative for short oils bearing a polar function such as a ketone or a nitrile.

3.1. Theoretical Versus Experimental EACNs. In colloidal physical chemistry, the EACN value of an oil expresses its ability to penetrate the interfacial film of SOW systems and to modify its spontaneous curvature.[47−49] In the case where the surfactant is a polyethoxylated fatty alcohol C_iE_j, some molecules of oil penetrate the interfacial film according to their affinity for C_iE_j molecules. In particular, when the oil has a polar function, its affinity for the film is stronger than apolar oils and its EACN is much lower than n, its number of carbon atoms. Indeed, Figure 6 illustrates the identical Winsor phase behavior of octyl octanoate ($n = 16$) and *n*-octane ($n = 8$), which is the linear alkane having an ACN equal to the EACN of the ester.
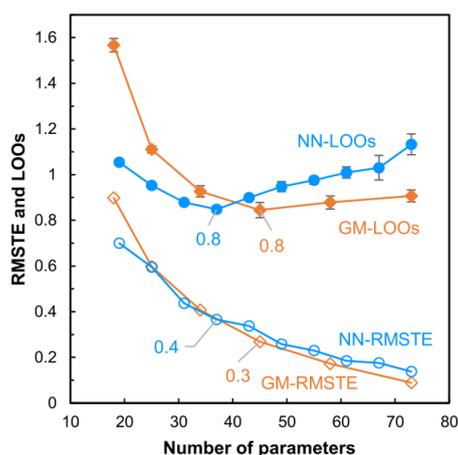
The EACN concept is of interest only if the values assigned to oils do not depend on the nature of the C_iE_j surfactant used for its measurement. This key issue has been checked by Bouton et al. who showed that the EACN values of 26 terpenes and non-linear (branched, unsaturated, cyclic) hydrocarbons were identical within 0.3 unit regardless of the surfactant used, namely, C_6E_4, C_8E_4, or $C_{10}E_4$.[15] However, for very polar oils, two major problems decrease the accuracy of EACN measurements. The first one stems from the fact that for oils having an EACN lower than 6, the calibration curve established with *n*-alkanes must be extrapolated to the dotted

parts of the regression straight line (see Figure 1b). Accordingly, the lower the EACN, the greater the uncertainty over its estimated value. The second problem arises from the monomeric solubility of C_iE_j surfactants in the oil phase which increases the apparent polarity of the oil. As a result, the EACNs measured with short C_iE_j such as C_6E_4 tend to be lower than the EACNs measured with a long C_iE_j whose monomeric concentration in the oil phase is significantly lower. This issue is particularly acute when the oils are very polar and the C_iE_j is very short. We encountered this difficulty while seeking to model the EACN of diisopropyl ether, for which we had previously assigned an EACN equal to 2.2[12] on the basis of the fish diagram determined by Wormuth and Kaler with the $C_{12}E_6$/diisopropyl ether/water system.[50] According to our very first models (GM and NN), the EACN of this oil appeared as an outlier. We therefore measured the EACN of this ether again using the same amphiphile (C_6E_4) that was used to measure the EACNs of most other highly polar oils.[34] The new value of the EACN thus determined (0.6, see Supporting Information) is, as expected, significantly lower than the previous value and perfectly consistent with the EACNs of other very polar oils, as they were determined with the same surfactant (C_6E_4). This revised EACN value has therefore been used to fit our GM and NN models.

3.2. GM and NN Complexity Selection. Figure 7 which displays the LOO scores and RMSTE versus the number of

Figure 7. RMSTE value of the model (out of 1000) having the smallest RMSTE for the GM-based model (orange empty diamonds) and NN-based model (blue circles) for the 111 molecules of the training set and means of the LOO values (GM orange diamonds, NN blue-filled circles) computed for five different parameter initializations for the 111 molecules of the training set vs number of parameters. The error bars for the LOO scores are the standard deviations computed over the five LOO score values.

parameters of the MLPs, that is the complexity, for the GM and NN models, indicates that in both cases (i) the data are correctly learned since the RMSTE are decreasing monotonously as the complexity increases and (ii) the LOO scores decrease, go through a minimum, and start increasing. The NN LOO score is clearly minimum (0.8 EACN unit) for a number of parameters equal to 37, that is, six hidden neurons. On the contrary, very close GM LOO score values equal to 0.8 and 0.9

EACN unit are computed for 45 (5N) and 58 (6N) parameters, respectively. In such a situation, the usual practice is to select the model with the lower complexity.[30] Therefore, GMs with five hidden neurons (45 parameters) and NNs with six hidden neurons (37 parameters) were kept for later testing.

An alternative to the LOO score computation is the leave-many-out (LMO) score computation, for which several molecules (e.g., k) are removed from the training set instead of one. Since the data set is small, it is advisable to train as many examples as possible to avoid overfitting. Indeed, when the number of parameters of the trained model exceeds half the number of examples in the training set, overfitting occurs and prediction performance starts deteriorating (above 60 parameters in Figure 7). With LOO, more examples are trained than with LMO (110 rather than 111-k), which reduces the risk of overfitting for a given number of parameters. Furthermore, with LOO, the maximum amount of information is used for training in the case of each removed example, so that the highest possible accuracy for its predicted value can be expected. Therefore, LOO was preferred to LMO, even if it requires a little more computation time.

For comparison purposes, the results of the EACN LOO estimations, corresponding to the LOO computation (out of five) that gives the best LOO score (0.8 EACN unit for GM and NN) for the molecule training set versus experimental EACN, are displayed in Figure 8 for the two preferred models.

Both models give similar results at a first glance (averaged cross validation R^2 are very close), in particular for the homologous series belonging to the chemical families indicated in the legends of Figure 8a,b, though slightly better estimations could be credited to the GM-based model for these 10 series. On the contrary, the other compounds, most of them possessing several structural features, have dots that lay closer to the bisector line for the NN-based model, indicating better results for the NN estimator with these compounds.

It is worth noting that with the NN-6N model, the EACN value of 2,6,10-trimethylundeca-2,6-diene is under-estimated by the LOO calculations (Figure 8b). A possible explanation for this significant discrepancy could result from the fact that the two double bonds of 2,6,10-trimethylundeca-2,6-diene are in position 2,6 and not at the end of the chain. Indeed, the NN "learns" the effect of double bonds on the basis of fairly rigid terpenes and a series of 1-alkenes whose double bonds are at the end of the chain. On average, each double bond decreases the EACN by 2.5−4.5 units and each branching decreases the EACN by 0.3−0.8 units. It is therefore logical that the EACN predicted by the NNs for this molecule with 14 carbons, 2 double bonds, and 3 branches (see Figure 8) is equal to 5 ± 1.5 units. However, the two double bonds of 2,6,10-trimethylundeca-2,6-diene are less accessible than those of 1-alkenes which tend to be located close to the polar zone of the interfacial film made up of $C_{10}E_4$. As a result, instead of decreasing the EACN by 9 units as expected, the experimentally observed decrease with respect to the corresponding n-alkane (tetradecane) is only 3.7 units.

This is probably due to the fact that double bonds in terminal or exocyclic position have a much greater effect than endocyclic bonds. Indeed, a comparison of the experimental EACNs of citronellyl acetate (−0.2) and geranyl acetate (−0.6), two molecules that differ only in a central double bond, indicates a decrease of only 0.4 EACN units for the central bond of linalyl acetate instead of the expected units 2.5 (or more). Thus, the additivity method used to evaluate the
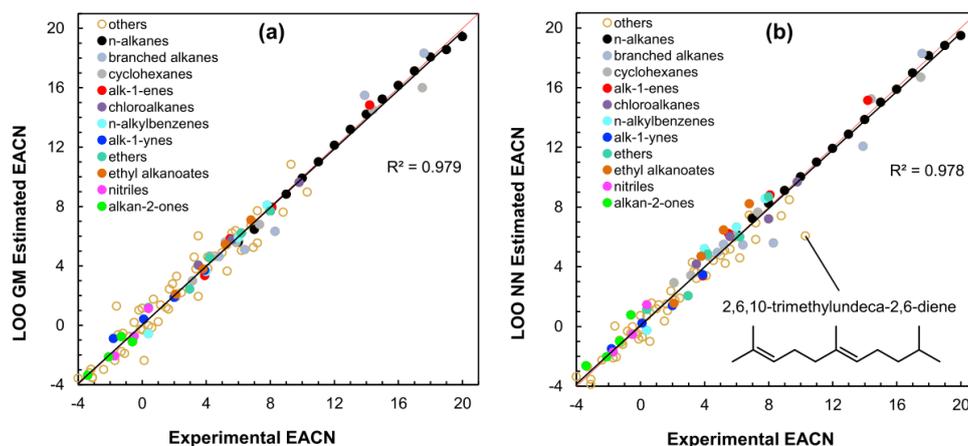
G

Figure 8. Scatter plots of LOO EACN estimations computed by GM from SMILES with five hidden neurons (a) and by NNs with six hidden neurons using M_0^X, M_2^X, M_3^X, and n as descriptors (b) for the 111 compounds of the training set vs experimental values of the EACN. The bisector and the regression lines are represented in red and black, respectively.

decreasing effects of several chemical features in a molecule relative to the EACN of the alkane with the same number of carbon atoms is probably inoperative, in particular for 2,6,10-trimethylundeca-2,6-diene.

As regards the GM-5N model in Figure 8a, no dot seems to be excessively far from the bisector, meaning that every EACN value of the training set is correctly estimated. In comparison with the NN model, the GM estimated EACN value of 2,6,10-trimethylundeca-2,6-diene is satisfying, which tends to confirm that the COSMO-RS descriptors used for the NN estimation fail to describe this compound's behavior.

3.3. Comparison of the Two Methods on the 10-Molecule Set. Hydrophobicity can be seen as the difficulty for the surfactant molecules to penetrate in the oil phase and is related to intermolecular forces between oil molecules and in the interfacial film between oil and surfactant molecules. In linear alkanes, London interactions between chains induce cohesive interchain interactions and hydrophobicity. Inter-molecule cohesion is reduced in cyclic, branched, and unsaturated molecules due to the steric constraint. In polar functionalized oil molecules, Keesom and Debye interactions occur, resulting in similarities between oil and surfactant molecules. Penetrating the oil interfacial film is then made easier for a surfactant molecule. Both phenomena, namely, a decrease in cohesive energy in the bulk oil due to steric constraint and an increase of favorable interactions between the surfactant and oil molecules contribute to reducing the hydrophobicity of an oil, illustrated here by its EACN. In complex molecules bearing several types of topological features and chemical functions, predicting the EACN is not trivial and the models presented should allow for considering the influence of every factor.

To assess the estimation accuracies of the NN-based and the GM-based models of previously selected complexities, computation of the EACN for the 10 molecules of the test set are made using the VLOO methodology previously described.[26] Briefly, for the GM and NN models selected in Section 3.2, 10 runs of 250 trainings each were performed with different parameter initializations. The VLOO score of each model (out of 250) was computed, and the mean of the 25

smallest VLOO scores of each run was computed. The run (out of ten) with the smallest mean VLOO score was selected. The 25 models of that sequence having the smallest VLOO scores estimated the EACN of the 10 test molecules, and the mean of those 25 estimations was computed. These final estimations for both models are plotted versus the experimental values in Figure 9. The proximity of the dots with the bisector line shows that these estimations are close to the experimental values. Only the isododecane (2,2,4,6,6-pentamethylheptane) blue data point is far from the bisector line. These good results are confirmed by the displayed determination coefficients that are equal to 0.992 and 0.986 for the GM-5N-based and NN-6N-based models, respectively.
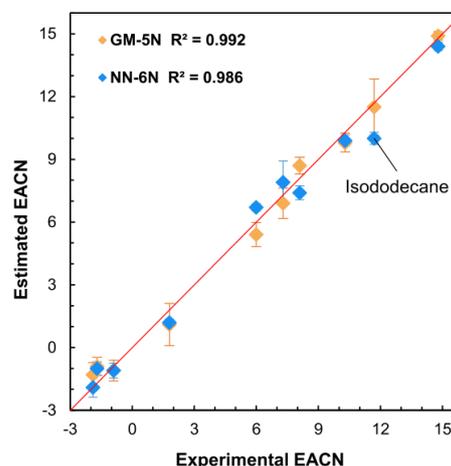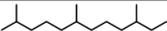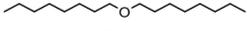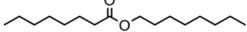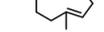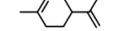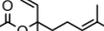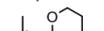

Figure 9. Scatter plots of EACN estimations computed by the graph-machine-based model with five hidden neurons (GM-5N, orange diamonds) and the neural-network-based model with six hidden neurons (NN-6N, blue diamonds) vs experimental EACN values for the 10 molecules of the test set. The bisector line is represented in red, and the error bars are the confidence intervals computed over the 25 selected models for the 10 molecules of the test set.

Table 2. Difference between Experimental and Estimated EACNs for the Test Set of 10 Molecules

Molecule	Chemical Structure	EACNexp.	Estimation error	
			NN-6N[a]	GM-5N[b]
Hemisqualane		14.8	+0.2	−0.1
Isododecane		11.7	+1.9	+0.2
Dioctylether		10.3	+0.2	+0.5
Octyloctanoate		8.1	+0.7	+0.6
Isopropyl myristate		7.3	−0.8	+0.3
Caryophyllene		6.0	−0.2	+0.6
Limonene		1.8	+0.7	+0.7
Linalyl acetate		-0.9	+0.1	+0.2
Rose oxide		-1.7	−0.1	−0.8
β-Ionone		-1.9	−0.4	−0.6
		test RMSE[c]	**0.7**	**0.5**

[a]Differences between experimental and estimated EACNs using the neural-network-based model. [b]Differences between experimental and estimated EACNs using graph-machine-based models. [c]Root-mean-square test error (in EACN unit) for the 10 molecules of the test.
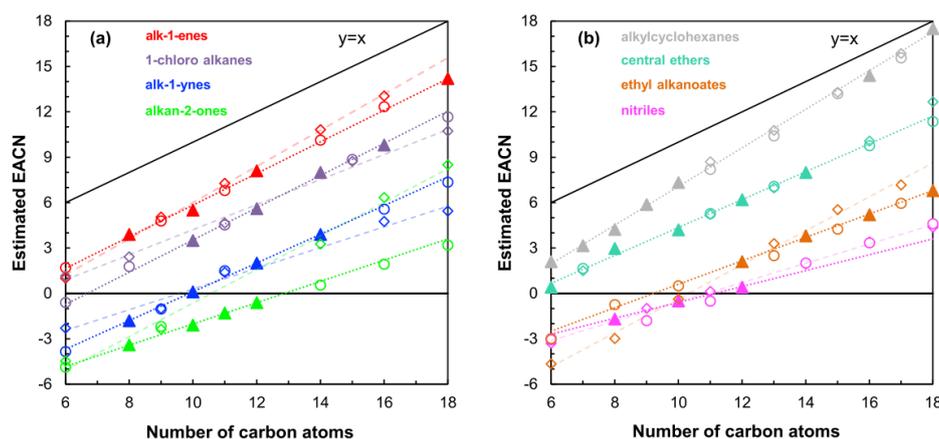

Figure 10. Evolution of experimental and estimated EACN with an increasing number of carbon atoms for homologous series of molecules with various chemical functions: (a) alk-1-enes, 1-chloroalkanes, alk-1-ynes, and n-alkan-2-ones and (b) n-alkylcyclohexanes, central ethers, ethyl alkanoates, and n-alkane nitriles. For clarity, the n-alkylbenzene series is not represented (superposition with alk-1-yne or ester series) and half of the predicted values are displayed. The dotted and dashed lines indicate the experimental and NN fits, respectively. Triangles (▲), diamonds (◇), and circles (○) are markers for experimental, NN-predicted, and GM-predicted values, respectively.

The estimations errors listed in Table 2 for the 10 molecules are indeed smaller or equal to 0.8 EACN unit but for the isododecane that exhibits a fairly large error with the NN model. The computed test root mean square error values (test RMSE, bottom row) with an eq 5-like formula, are equal to 0.5 and 0.7 EACN unit, confirming the efficiency of both models. Moreover, the estimations of the two predictors are in good agreement since the maximum of the error deviation between the two computations is equal to 0.8 EACN unit for 8 molecules out of 10. Even for six of them, the estimation error difference is less or equal to 0.3 EACN unit. Those results were not given for granted, especially for complex molecules that have multiple features (e.g., limonene or rose oxide). Finally, it should be noted that to get such convincing results with the GM-based model, the SMILES code used to generate the octyl octanoate GM was also modified as explained for hexyl octanoate in Section 2.4. Without taking this precaution, the prediction for octyl octanoate was clearly out of range. The estimation values for the 10-molecule test set are also reported in the Supporting Information (Table S1, columns 10 and 11). Therefore, the two selected models can be used in tandem to predict the EACN of compounds in homologous series, while keeping in mind that the EACN for branched molecules will be probably under-estimated.

Table 3. Family Linear Fits for Experimental and Predicted EACN Versus Number of Carbon Atoms n

family	exp. fit[a]	exp. R^2	GM fit[b]	GM R^2	NN fit[c]	NN R^2	EACN$_{exp}$(10)[d]
n-alkylcyclohexanes	$1.28n-5.7$ (7)	1	$1.25n-5.6$ (6)	1	$1.19n-4.5$	1	7.3
alk-1-enes	$1.05n-4.6$ (4)	1	$1.06n-4.7$ (9)	1	$1.16n-5.6$	1	6.4
central ethers	$0.92n-4.9$ (5)	0.99	$0.89n-4.5$ (8)	1	$0.98n-5.5$	1	4.9
1-chloroalkanes	$1.07n-7.1$ (4)	1	$1.03n-6.6$ (9)	1	$0.84n-4.2$	0.99	3.9
ethyl n-alkanoates	$0.78n-7.2$ (4)	1	$0.77n-7.2$ (9)	0.99	$1.13n-11.8$	1	0.8
alk-1-ynes	$0.95n-9.4$ (4)	1	$0.93n-9.1$ (9)	1	$0.70n-6.6$	0.98	0.6
n-alkylbenzenes	$0.93n-8.9$ (4)	1	$1.05n-10.6$ (9)	0.99	$0.98n-9.7$	0.99	0.1
nitriles	$0.53n-5.9$ (3)	0.99	$0.62n-6.9$ (10)	0.98	$0.64n-6.9$	1	−0.9
n-Alkan-2-ones	$0.70n-9.1$ (4)	1	$0.66n-8.5$ (9)	0.99	$1.11n-11.8$	0.98	−2.1

[a]In brackets, number of points used for the experimental fits. [b]In brackets, number of points used for the GM fits. [c]Number of points used is the same as for the GM fits. [d]EACN$_{exp}$ calculated with $n = 10$.

3.4. EACN Prediction of Homologous Series. One of the potential applications of the previously developed models is to predict the EACN for homologous series of oils with alkyl chains of increasing size. Indeed, the comparison of homologous oils having the same number of carbons but carrying various chemical functions makes it possible to deduce the effect of a given function on the EACN of oils. Furthermore, for practical applications, C8 to C15 phenols,[51] terpenes,[15,52] and terpenoids[19] are particularly frequent in perfumery, while C12 to C18 alkanes, esters, and ethers are widely used as emollients to prepare cosmetic emulsions.[53] It is therefore crucial to reliably predict the EACN of oils with a reduced number of carbons (≤ 20).

In a previous work, we empirically observed that the EACN of several series of homologous oil increases approximately linearly with the number of carbons.[19] We therefore tested the ability of the GM and NN models to predict the evolution of EACNs of homologous oils. The homologous set designed to explore the effectiveness of the two models is constructed as follows. The picked homologous series are the nine chemical families already mentioned in the Figure 8 scatter plot legends, from cyclohexanes to alkan-2-ones. Indeed, all the molecules belonging to those families have a n-alkyl chain backbone of increasing size that contain one of the following: (i) one terminal functional group (esters, ketones, and nitriles), (ii) one central carbon substituted with an oxygen atom (ethers), (iii) one terminal unsaturation (alkenes and alkynes), (iv) a cycle in the terminal position (cyclohexanes and benzenes), or (v) one terminal chain substitution with a chloride atom (1-chloroalkanes). For all the series, the number of carbon atoms per molecule n is varied from 6 to 18, so that all the series contains 13 compounds each, and the whole set 117 compounds. Since 46 out of these belong de facto to the 111-molecule training set, they cannot be kept for prediction testing. Instead, they will be used as benchmarks to assess the accuracy of the model predictions for each series. The σ-moments for the supplementary compounds of theses series are calculated as described in Section 2.3. The data for the 117 compounds of the homologous series are available in Table S2 of the Supporting Information.

The scatter plots of the EACN predictions for the two retained models and the experimental EACN versus n are shown in Figure 10 for eight series. The alkylbenzene series plot could not be represented due to an overlap with datapoints from the alk-1-yne series or the ethyl alkanoate series. This plot is shown, as well as those of the other series with all points shown, in Figures S3 and S4 of the Supporting Information.

As expected, the experimental linear fits (represented as dotted lines) are good for all series. The goodness of fit is further confirmed by the values of the experimental determination coefficients reported in the third column of Table 3, all superior to 0.99. In this last table, the computed linear equations and determination coefficients for the predicted fits are also given for the nine series. With these data, the accuracy of the predictions can be analyzed by comparing for all the series the proximity of the predicted points to the dotted lines (Figure 10) and the slopes of the GM and NN fits with the slope of the experimental fits (Table 3, columns 2, 4, and 6).

For the GM model, it can be seen in Figure 10 that the predictions match the experimental results quite well for seven of the nine series since most of the circles are located on or near the experimental dotted lines. Furthermore, with the exception of n-alkylbenzenes and nitriles, the slopes of the fits reported in columns 2 and 4 of Table 3 are very close. Regarding the 1-nitrile series, it turns out that the GM and NN models converge toward the same predictions, with almost identical slopes for their fitting equations (penultimate line of Table 3). Hence, we can postulate that the two model deviations from the experimental trend could be due to some experimental error. Since the experimental fit is computed with only three successive values of n, a small error in a fish temperature determination could induce, as already mentioned, a deviation of up to 0.3 EACN unit. Thus, such an increase in the EACN for the dodecanenitrile value (0.3) would be sufficient to make the three linear fits match. Indeed, this modification would give a modified equation equal to $0.64n-6.9$ for the dotted experimental line, almost identical to the two model equations. This increased experimental value for dodecanenitrile (0.7 instead of 0.4) would also be consistent with a proportional spacing for successive EACN values for nitriles. Finally, the slightly larger slope of the alkylbenzene GM fit compared to the experimental fit is mainly due to an underestimation of the EACN by the GM model for compounds with n lower than 11 (see Figure S4 of the Supporting Information). This behavior which has also been observed for the LOO EACN estimations of p-xylene and p-cymene (n equal to 8 and 10; entries 81 and 83 in Table S1 of the Supporting Information) can be explained by the GM constructions which are different in the benzenic series depending on the length of the alkyl chain. For n less than or equal to 12, the root node of the GMs is located on the benzene ring, while for larger n, it is positioned on the alkyl chain. This can be shown by fitting with different lines the GM-predicted points for n less than or equal to 12 and n

J

greater than 13, as this leads to a much better fit for the two resulting lines (R^2 equal to 1). Nevertheless, to obtain a coherence in the construction of the GMs for the benzene derivatives, it would be necessary to modify the SMILES codes of several molecules of the training set, which is beyond the scope of this work. We need also to point out that, as for the nitrile series, a small correction in the experimental EACN (0.3) for butylbenzene would result for the experimental fit in a slope correction large enough to equal the GM slope fit. This last remark indicates that for the nitrile and alkylbenzene series, the predictions computed with the GM model are within the experimental margin of error.

For the NN predictive model, the results are more mixed. With four of the nine series, namely, n-alkylcyclohexanes, ethers, nitriles, and n-alkylbenzenes, and admitting a small measurement error for the nitrile series, the predictions are satisfactory. On the contrary, for 1-alk-1-enes, 1-chloroalkanes, alk-1-ynes, and ethyl n-alkylalkanoates, a significant deviation from the dotted trend lines, up to 2 EACN unit, is observed for the predicted points located in the extrapolated regions. A larger difference in slope between the experimental and NN fits is indeed reported in Table 3 (columns 2 and 6), so that dashed lines for those series have been added in Figure 10 to materialize this divergence between the two fits. Finally, the largest deviation is computed for compounds of the alkan-2-one series, for which n is greater than 12; the prediction is becoming erroneous beyond tridecanone. No explanation has yet been found for this discrepancy.

Overall, the predictions obtained with the two models are rather concordant for all homologous series, the particular case of ketones being put aside. As a result, both models can be used to predict the EACN value for a new molecule belonging to one of these series. While the GM model allows us to obtain a result more quickly, since it is enough to use a SMILES code, computing a prediction with both models allows us to anticipate an incorrect GM construction if the predicted results are very different.

As stated elsewhere,[12] the oils that produce a higher difference in EACN with respect to the linear alkanes with the same n are those that have a higher affinity with the interfacial film, and from the last column of Table 3, the decreasing order is n-alkan-2-ones > 1-nitriles > alkylbenzenes > ethyl n-alkanoates ≈ alk-1-ynes > 1-chloroalkanes > central ethers > alk-1-enes > n-alkylcyclohexanes. Finally, we need to point out that specific SMILES codes for some molecules, for example, hex-1-yne or hexan-2-one, have been used, as explained for hexyl octanoate ester, to get a consistency among the GM constructions in the corresponding series. With these adjustments, the GM predictions for most of the molecules in the series are rather efficient. The construction adjustment of the hexan-2-one GM is explained in the Supporting Information.

4. CONCLUSIONS

The experimental determination of an EACN value by the traditional fish-tail method is a tedious and time-consuming task.[11] In this work, two machine-learning models were built to estimate the EACN value of oils from their molecular structure. On the basis of 111 experimental values of EACN, estimations were performed either by nonlinear regression (NN) from COSMO-RS σ-moments or by regression on graphs (GM) derived from the SMILES codes of the molecules. In each case, the selection of the appropriate model was assessed by LOO score computation. The

effectiveness of the chosen NN-6N and GM-5N models was tested on a set of 10 cosmetic and perfumery molecules. It was found that both models yielded predictions with similar and satisfactory accuracies (root-mean-square estimation errors equal to 0.7 and 0.5, respectively). Molecular structures in the test set were chosen on purpose as polyfunctional molecules, for which the influence of each structural feature could not be considered independently. Multilinear regressions were shown to be efficient to predict the EACN value for monofunctional molecules,[18,19] but this work is the first one regarding EACN prediction of complex polyfunctional ones. It was pointed out that for homologous molecule series, the linear evolution of the EACN with the increase of chain length is an appropriate model and is well tackled by the GM predictor. However, the NN model based on COSMO-RS σ-moments as descriptors met some difficulties in estimating the evolution of EACN values for the alkan-2-one series.

Overall, the GM model is the most convenient model since it only needs SMILES codes as input values, whereas the NN model requires COSMO-RS before EACN estimations. A demonstration of the GM and NN computations, based on the Docker free software technology (available on most operating systems), is available for download (see the Supporting Information). It is, for example, very easy and very fast (<0.5 s) to predict the EACN value of any liquid of moderate molecular size (M < 350 Da) that contains C, H, or O atoms using its SMILES code only.

■ ASSOCIATED CONTENT

ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.2c04592.

> Training and test compound database DEMOEACN (XLSX)
>
> Names, SMILES notations, three first σ-moments (different from zero) calculated with COSMO-RS, number of carbon atoms, experimental EACN, average EACN values determined from the fish-tail-temperature T^* reported in the literature for ternary systems C_iE_j/oil/water, estimated EACN values with the GM-5N and NN-6N models, molecular formulas and CAS RNs for the 111 molecules of the training set and the 10 molecules of the test set and the 117 molecules of the homologous series, fish diagrams of dipropyl ether and diisopropyl ether using C_6E_4 as a surfactant, evolution of experimental and estimated EACNs with increasing number of carbon atoms for nine homologous series, GM and NN demonstrations with Docker containers, and GM and NN results with Docker (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

François Duprat − Laboratoire de Chimie Organique, CNRS, ESPCI Paris, PSL Research University, 75005 Paris, France; ◉ orcid.org/0000-0002-2889-1701; Phone: 33-1-40794465; Email: arthur.duprat@espci.psl.eu; Fax: 33-1-40794466

Véronique Nardello-Rataj − University of Lille, CNRS, Centrale Lille, Université d'Artois, UMR 8181—UCCS—Unité de Catalyse et Chimie du Solide, F-59000 Lille, France; Phone: 33-3-20336369; Email: veronique.rataj-nardello@univ-lille.fr

Jean-Marie Aubry − *University of Lille, CNRS, Centrale Lille, Université d'Artois, UMR 8181—UCCS—Unité de Catalyse et Chimie du Solide, F-59000 Lille, France*; Phone: 33-3-20336364; Email: jean-marie.aubry@univ-lille.fr

Authors

Lucie Delforce − *University of Lille, CNRS, Centrale Lille, Université d'Artois, UMR 8181—UCCS—Unité de Catalyse et Chimie du Solide, F-59000 Lille, France*

Jean-Luc Ploix − *Laboratoire de Chimie Organique, CNRS, ESPCI Paris, PSL Research University, 75005 Paris, France*

Jesus Fermín Ontiveros − *University of Lille, CNRS, Centrale Lille, Université d'Artois, UMR 8181—UCCS—Unité de Catalyse et Chimie du Solide, F-59000 Lille, France*

Valentin Goussard − *University of Lille, CNRS, Centrale Lille, Université d'Artois, UMR 8181—UCCS—Unité de Catalyse et Chimie du Solide, F-59000 Lille, France*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acsomega.2c04592

Notes

The authors declare no competing financial interest.
The established database **DEMOEACN** is provided in the excel file named "DEMOEACN". A pdf file named Table S2 contains the EACN prediction results for the 117 compounds of 9 homologous series by using a Docker image available via the Docker software. Another pdf file named DEMO explains how to install Docker on your system and launch the Docker image used to compute the predictions. The Docker software is available at https://www.docker.com/get-started.

■ **REFERENCES**

(1) Leo, A.; Hansch, C.; Elkins, D. Partition Coefficients and Their Uses. *Chem. Rev.* **1971**, *71*, 525−616.
(2) Griffin, W. C. Classification of surface-active agents by "HLB". *J. Soc. Cosmet. Chem.* **1949**, *1*, 311−326.
(3) Griffin, W. C. *Calculation of HLB values of non-ionic surfactants* **1954**, *5*, 249−256.
(4) Robbers, J. E.; Bhatia, V. N. Technique for the Rapid Determination of HLB and Required-HLB Values. *J. Pharm. Sci.* **1961**, *50*, 708−709.
(5) Orafidiya, L. O.; Oladimeji, F. A. Determination of the Required HLB Values of Some Essential Oils. *Int. J. Pharm.* **2002**, *237*, 241.
(6) dos Santos, O. D. H.; Miotto, J. V.; de Morais, J. M.; da Rocha-Filho, P. A.; de Oliveira, W. P. Attainment of Emulsions with Liquid Crystal from Marigold Oil Using the Required HLB Method. *J. Dispersion Sci. Technol.* **2005**, *26*, 243−249.
(7) Moreira de Morais, J.; David Henrique dos Santos, O.; Delicato, T.; Azzini Gonçalves, R.; Alves da Rocha-Filho, P. Physicochemical Characterization of Canola Oil/Water Nano-emulsions Obtained by Determination of Required HLB Number and Emulsion Phase Inversion Methods. *J. Dispersion Sci. Technol.* **2006**, *27*, 109−115.
(8) Meher, J. G.; Yadav, N. P.; Sahu, J. J.; Sinha, P. Determination of Required Hydrophilic-Lipophilic Balance of Citronella Oil and Development of Stable Cream Formulation. *Drug Dev. Ind. Pharm.* **2013**, *39*, 1540−1546.
(9) Lee, Y.-Y.; Yoon, K.-S. Determination of Required HLB Values for Citrus Unshiu Fruit Oil, Citrus Unshiu Peel Oil, Horse Fat and Camellia Japonica Seed Oil. *J. Cosmet. Sci.* **2020**, *71*, 411−424.
(10) Wade, W. H.; Morgan, J. C.; Jacobson, J. K.; Schechter, R. S. Low Interfacial Tensions Involving Mixtures of Surfactants. *Soc. Pet. Eng. J.* **1977**, *17*, 122−128.
(11) Queste, S.; Salager, J. L.; Strey, R.; Aubry, J. M. The EACN Scale for Oil Classification Revisited Thanks to Fish Diagrams. *J. Colloid Interface Sci.* **2007**, *312*, 98−107.
(12) Aubry, J.-M.; Ontiveros, J. F.; Salager, J.-L.; Nardello-Rataj, V. Use of the Normalized Hydrophilic-Lipophilic-Deviation (HLD_N) Equation for Determining the Equivalent Alkane Carbon Number (EACN) of Oils and the Preferred Alkane Carbon Number (PACN) of Nonionic Surfactants by the Fish-Tail Method (FTM). *Adv. Colloid Interface Sci.* **2020**, *276*, 102099.
(13) Pizzino, A.; Molinier, V.; Catté, M.; Ontiveros, J. F.; Salager, J.-L.; Aubry, J.-M. Relationship between phase behavior and emulsion inversion for a well-defined surfactant ($C_{10}E_4$)/n-octane/water ternary system at different temperatures and water/oil ratios. *Ind. Eng. Chem. Res.* **2013**, *52*, 4527−4538.
(14) Creton, B.; Lévêque, I.; Oukhemanou, F. Equivalent Alkane Carbon Number of Crude Oils: A Predictive Model Based on Machine Learning. *Oil Gas Sci. Technol.—Rev. IFP Energies nouvelles* **2019**, *74*, 30.
(15) Bouton, F.; Durand, M.; Nardello-Rataj, V.; Borosy, A. P.; Quellet, C.; Aubry, J.-M. A QSPR Model for the Prediction of the "Fish-Tail" Temperature of C_iE_4/Water/Polar Hydrocarbon Oil Systems. *Langmuir* **2010**, *26*, 7962−7970.
(16) Klamt, A.; Eckert, F.; Hornig, M. COSMO-RS A Novel View to Physiological Solvation and Partition Questions. *J. Comput. Aided Mol. Des.* **2001**, *15*, 355.
(17) Klamt, A. *COSMO-RS: from quantum chemistry to fluid phase thermodynamics and drug design*, 1st ed.; Elsevier: Amsterdam, 2005.
(18) Lukowicz, T. Synergistic Solubilisation of Fragrances in Binary Surfactant Systems and Prediction of Their EACN Value with COSMO-RS. *Lille* **2015**, *1*.
(19) Lukowicz, T.; Illous, E.; Nardello-Rataj, V.; Aubry, J.-M. Prediction of the equivalent alkane carbon number (EACN) of aprotic polar oils with COSMO-RS sigma-moments. *Colloids Surf. A Physicochem. Eng. Asp.* **2018**, *536*, 53−59.
(20) Katritzky, A. R.; Kuanar, M.; Slavov, S.; Hall, C. D.; Karelson, M.; Kahn, I.; Dobchev, D. A. Quantitative Correlation of Physical and Chemical Properties with Chemical Structure: Utility for Prediction. *Chem. Rev.* **2010**, *110*, 5714−5789.
(21) Nieto-Draghi, C.; Fayet, G.; Creton, B.; Rozanska, X.; Rotureau, P.; de Hemptinne, J.-C.; Ungerer, P.; Rousseau, B.; Adamo, C. A General Guidebook for the Theoretical Prediction of Physicochemical Properties of Chemicals for Regulatory Purposes. *Chem. Rev.* **2015**, *115*, 13093−13164.
(22) Duprat, A. F.; Huynh, T.; Dreyfus, G. Toward a Principled Methodology for Neural Network Design and Performance Evaluation in QSAR. Application to the Prediction of LogP. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 586−594.
(23) Goulon-Sigwalt-Abram, A.; Duprat, A.; Dreyfus, G. From Hopfield Nets to Recursive Networks to Graph Machines: Numerical Machine Learning for Structured Data. *Theor. Comput. Sci.* **2005**, *344*, 298−334.
(24) Goulon, A.; Duprat, A.; Dreyfus, G. Graph Machines and Their Applications to Computer-Aided Drug Design: A New Approach to Learning from Structured Data. In *Unconventional Computation*; Calude, C. S., Dinneen, M. J., Păun, G., Rozenberg, G., Stepney, S., Eds.; *Lecture Notes in Computer Science*; Springer: Berlin, Heidelberg, 2006; Vol. *4135*, pp 1−19.
(25) Goussard, V.; Duprat, F.; Gerbaud, V.; Ploix, J.-L.; Dreyfus, G.; Nardello-Rataj, V.; Aubry, J.-M. Predicting the Surface Tension of Liquids: Comparison of Four Modeling Approaches and Application to Cosmetic Oils. *J. Chem. Inf. Model.* **2017**, *57*, 2986−2995.
(26) Goussard, V.; Duprat, F.; Ploix, J.-L.; Dreyfus, G.; Nardello-Rataj, V.; Aubry, J.-M. A New Machine-Learning Tool for Fast

924 Estimation of Liquid Viscosity. Application to Cosmetic Oils. *J. Chem.*
925 *Inf. Model.* **2020**, *60*, 2012−2023.
926 (27) Saldana, D. A.; Starck, L.; Mougin, P.; Rousseau, B.; Pidol, L.;
927 Jeuland, N.; Creton, B. Flash Point and Cetane Number Predictions
928 for Fuel Compounds Using Quantitative Structure Property Relation-
929 ship (QSPR) Methods. *Energy Fuels* **2011**, *25*, 3900−3908.
930 (28) Goulon, A.; Picot, T.; Duprat, A.; Dreyfus, G. Predicting
931 Activities without Computing Descriptors: Graph Machines for
932 QSAR. *SAR QSAR Environ. Res.* **2007**, *18*, 141−153.
933 (29) Porcheron, F.; Jacquin, M.; El Hadri, N. E.; Saldana, D. A.;
934 Goulon, A.; Faraj, A. Graph Machine Based-QSAR Approach for
935 Modeling Thermodynamic Properties of Amines: Application to CO2
936 Capture in Postcombustion. *Oil Gas Sci. Technol.—Rev. IFP Energies*
937 *nouvelles* **2013**, *68*, 469−486.
938 (30) Dioury, F.; Duprat, A.; Dreyfus, G.; Ferroud, C.; Cossy, J.
939 QSPR Prediction of the Stability Constants of Gadolinium(III)
940 Complexes for Magnetic Resonance Imaging. *J. Chem. Inf. Model.*
941 **2014**, *54*, 2718−2731.
942 (31) Goulon, A.; Faraj, A.; Pirngruber, G.; Jacquin, M.; Porcheron,
943 F.; Leflaive, P.; Martin, P.; Baron, G. V.; Denayer, J. F. M. Novel
944 Graph Machine Based QSAR Approach for the Prediction of the
945 Adsorption Enthalpies of Alkanes on Zeolites. *Catal. Today* **2011**,
946 *159*, 74−83.
947 (32) Bouton, F.; Durand, M.; Nardello-Rataj, V.; Serry, M.; Aubry,
948 J.-M. Classification of Terpene Oils Using the Fish Diagrams and the
949 Equivalent Alkane Carbon (EACN) Scale. *Colloids Surf. A*
950 *Physicochem. Eng. Asp.* **2009**, *338*, 142−147.
951 (33) Ontiveros, J. F.; Pierlot, C.; Catté, M.; Molinier, V.; Pizzino, A.;
952 Salager, J.-L.; Aubry, J.-M. Classification of Ester Oils According to
953 Their Equivalent Alkane Carbon Number (EACN) and Asymmetry of
954 Fish Diagrams of $C_{10}E_4$/Ester Oil/Water Systems. *J. Colloid Interface*
955 *Sci.* **2013**, *403*, 67−76.
956 (34) Lukowicz, T.; Benazzouz, A.; Nardello-Rataj, V.; Aubry, J.-M.
957 Rationalization and Prediction of the Equivalent Alkane Carbon
958 Number (EACN) of Polar Hydrocarbon Oils with COSMO-RS σ-
959 Moments. *Langmuir* **2015**, *31*, 11220−11226.
960 (35) Gibson, T. Phase-Transfer Synthesis of Monoalkyl Ethers of
961 Oligoethylene Glycols. *J. Org. Chem.* **1980**, *45*, 1095−1098.
962 (36) Lang, J. C.; Morgan, R. D. Nonionic Surfactant Mixtures. I.
963 Phase Equilibria in $C_{10}E_4$−H_2O and Closed-loop Coexistence. *J.*
964 *Chem. Phys.* **1980**, *73*, 5849−5861.
965 (37) Schlarmann, J.; Stubenrauch, C.; Strey, R. Correlation between
966 Film Properties and the Purity of Surfactants. *Phys. Chem. Chem. Phys.*
967 **2003**, *5*, 184−191.
968 (38) Schubert, K.-V.; Strey, R.; Kahlweit, M. A New Purification
969 Technique for Alkyl Polyglycol Ethers and Miscibility Gaps for Water-
970 CiEj. *J. Colloid Interface Sci.* **1991**, *141*, 21−29.
971 (39) Winsor, P. A. *Solvent Properties of Amphiphilic Compounds*;
972 Butterworth: London, 1954.
973 (40) Burauer, S.; Sachert, T.; Sottmann, T.; Strey, R. On
974 Microemulsion Phase Behavior and the Monomeric Solubility of
975 Surfactant. *Phys. Chem. Chem. Phys.* **1999**, *1*, 4299−4306.
976 (41) Dupeux, T.; Gaudin, T.; Marteau-Roussy, C.; Aubry, J.-M.;
977 Nardello-Rataj, V. COSMO-RS as an Effective Tool for Predicting the
978 Physicochemical Properties of Fragrance Raw Materials. *Flavour*
979 *Fragrance J.* **2022**, *37*, 106−120.
980 (42) Klamt, A. Conductor-like Screening Model for Real Solvents: A
981 New Approach to the Quantitative Calculation of Solvation
982 Phenomena. *J. Phys. Chem.* **1995**, *99*, 2224−2235.
983 (43) Hammer, B. Recurrent Networks for Structured Data - A
984 Unifying Approach and Its Properties. *Cogn. Syst. Res.* **2002**, *3*, 145−
985 165.
986 (44) Dreyfus, G. *Neural Networks: Methodology and Applications*;
987 Springer: Berlin, 2005.
988 (45) Monari, G.; Dreyfus, G. Local Overfitting Control via
989 Leverages. *Neural Comput.* **2002**, *14*, 1481−1506.
990 (46) Winsor, P. A. Hydrotropy, Solubilisation and Related
991 Emulsification Processes. *Faraday Soc. Trans.* **1948**, *44*, 376−398.

992 (47) Kanei, N.; Tamura, Y.; Kunieda, H. Effect of Types of Perfume
993 Compounds on the Hydrophile−Lipophile Balance Temperature. *J.*
994 *Colloid Interface Sci.* **1999**, *218*, 13−22.
995 (48) Tchakalova, V.; Fieber, W. Classification of Fragrances and
996 Fragrance Mixtures Based on Interfacial Solubilization. *J. Surfactants*
997 *Deterg.* **2012**, *15*, 167−177.
998 (49) Hyde, S.; Blum, Z.; Landh, T.; Lidin, S.; Ninham, B. W.;
999 Andersson, S.; Larsson, K. *The Language of Shape: The Role of*
1000 *Curvature in Condensed Matter: Physics, Chemistry and Biology*;
1001 Elsevier, 1996.
1002 (50) Wormuth, K. R.; Kaler, E. W. Microemulsifying Polar Oils. *J.*
1003 *Phys. Chem.* **1989**, *93*, 4855−4861.
1004 (51) Ontiveros, J. F.; Bouton, F.; Durand, M.; Pierlot, C.; Quellet,
1005 C.; Nardello-Rataj, V.; Aubry, J.-M. Dramatic Influence of Fragrance
1006 Alcohols and Phenols on the Phase Inversion Temperature of the
1007 Brij30/n-Octane/Water System. *Colloids Surf. A Physicochem. Eng.*
1008 *Asp.* **2015**, *478*, 54−61.
1009 (52) Lukowicz, T.; Company Maldonado, R. C.; Molinier, V.;
1010 Aubry, J.-M.; Nardello-Rataj, V. Fragrance Solubilization in Temper-
1011 ature Insensitive Aqueous Microemulsions Based on Synergistic
1012 Mixtures of Nonionic and Anionic Surfactants. *Colloids Surf. A*
1013 *Physicochem. Eng. Asp.* **2014**, *458*, 85−95.
1014 (53) Goussard, V.; Aubry, J.-M.; Nardello-Rataj, V. Bio-Based
1015 Alternatives to Volatile Silicones: Relationships between Chemical
1016 Structure, Physicochemical Properties and Functional Performances.
1017 *Adv. Colloid Interface Sci.* **2022**, *304*, 102679.

Résumé

Les objectifs de la thèse sont multiples et s'inscrivent dans le cadre de la mise en place d'une nouvelle plateforme technologique HT-SMART-FORMU dédiée à la formulation. L'accent est mis sur le développement de méthodes expérimentales fiables et d'outils théoriques et prédictifs, afin d'établir des relations entre lla composition chimique, les propriétés physicochimiques et les propriétés applicatives.

Les dispersions de particules sont le premier type de systèmes étudiés. Leur est traditionnellement rationalisée à l'aide de la théorie DLVO, mais des études récentes proposent l'utilisation des Paramètres de Solubilité de Hansen pour décrire leur stabilité en milieu non aqueux. Dans le premier chapitre, une méthode analytique systématique basée sur la mesure du potentiel zêta et le suivi de la stabilité par diffusion de la lumière permet de déduire une complémentarité des deux théories pour décrire la stabilité des dispersions de TiO_2 en solvants non aqueux.

La problématique de la prédiction de l'hydrophobie des huiles par la modélisation d'EACN, qui est un paramètre clé pour concevoir des systèmes dispersés tensioactif/huile/eau (SOW), est abordée dans le deuxième chapitre. Deux modèles mathématiques, conçus à l'aide de méthodes de machine-learning, sont proposés pour la prédiction rapide de l'EACN des huiles, à savoir les réseaux de neurones (NN) et les machines graphiques (GM). Alors que le modèle GM est implémenté à partir des codes SMILES, le modèle NN est alimenté par des descripteurs σ-moments calculés avec le logiciel COSMOtherm. La fiabilité des prédictions des deux modèles est discutée sur la base d'un ensemble de test de 10 molécules complexes.

Dans les chapitres 3 et 4, l'étendue des propriétés applicatives d'un tensioactif non ionique à base de glycérol ($C_{12}Gly_2$) est étudiée. Le chapitre 3 se porte sur son agrégation en solution aqueuse et la formation de cristaux liquides (CL) à faible concentration, en comparaison avec les alcools gras polyéthoxylés et les alkyl polyglucosides. L'influence de ses propriétés physico-chimiques, en particulier la viscoélasticité de dilatation de l'interface air/eau, est mise en relation avec la faible moussabilité et la stabilité de mousse durable observées.

Enfin, les propriétés du $C_{12}Gly_2$ en systèmes SOW sont étudiées dans le chapitre 4. La déviation hydrophile-lipophile normalisé (HLD_N), un outil théorique puissant, est considéré comme un moyen de rationaliser les caractéristiques des émulsions et des microémulsions. Ainsi, une quantification approfondie de l'amphiphilie des tensioactifs, de leur sensibilité à la température et de leur tolérance au sel est présentée. L'utilisation du $C_{12}Gly_2$ comme émulsifiant H/E et E/H est ensuite étudiée : la granulométrie et la stabilité des émulsions obtenues en faisant varier l'huile concordent avec les valeurs du HLD_N. Un minimum est observé à $HLD_N = 0$, puis la granulométrie et la stabilité augmentent pour des valeurs de HLD_N négatives et positives.

<u>Mots-clés</u> : Particules ; Huiles ; Tensioactifs ; Méthodes Prédictives ; Mousse ; HLD

Abstract

The multiple aims of this thesis fall within the implementation of a new HT-SMART-FORMU technological platform dedicated to formulation science. The focus is set on the development of reliable experimental methods, theoretical and predictive tools, so as to establish relations between chemical composition, physicochemical properties and applicative properties.

The first type of systems investigated in this work consists of solid dispersions. Their stability has long been approached by the DLVO theory, but more recent studies suggest the use of Hansen Solubility Parameters to describe their stability in non-aqueous media. In the first chapter, a systematic analytical method based on zeta potential measurement and light scattering stability monitoring allows deducing a complementarity of both theories to describe the stability of TiO_2 dispersions in non-aqueous solvents.

The problematic of oil hydrophobicity prediction through EACN values modelling, which is a key parameter to design surfactant/oil/water (SOW) dispersed systems, is addressed in the second chapter. Two mathematical models, designed using machine-learning methods, are proposed for the rapid prediction of the EACN of oils, namely Neural Networks (NN) and Graph Machines (GM). While the GM model is implemented from the SMILES codes, the NN model is fed with σ-moments descriptors computed with the COSMOtherm software. The prediction reliability of both models is discussed based on a complex 10-molecule test set.

In chapters 3 and 4, the scope of applicative properties of a nonionic glycerol-based surfactant are investigated. Firstly, chapter 3 focuses on its aggregation behaviour in aqueous solutions and the formation of liquid crystals (LC) at low concentration, in comparison with the benchmark polyethoxylated fatty alcohols and alkyl polyglucosides. The influence of its physicochemical properties, in particular the air / water interface dilational viscoelasticity, is put in relation with the observed poor foamability and long-lasting foam stability.

Finally, $C_{12}Gly_2$ properties in SOW systems are investigated in chapter 4. The Normalized Hydrophilic-Lipophilic Deviation (HLD_N), a powerful theoretical tool, is regarded as a way to rationalize the characteristics of both emulsions and microemulsions. In this way, a thorough quantification of surfactants amphiphilicity, temperature sensitivity and salt-tolerance are presented. The potential use of $C_{12}Gly_2$ as O/W and W/O emulsifier is then investigated: the granulometry and stability of emulsions obtained by varying the oil concurs with HLD_N values. A minimum is observed at $HLD_N = 0$ and increases for negative and positive HLD_N values.

<u>Keywords</u>: Particles; Oils; Surfactants; Predictive methods; Foam; HLD