12-15-2015

On the Use of the Linear Interaction Energy Method to Predict Affinities of Charged Aromatic Ammines to Naturally Occurring Clay Minerals

Milinda A. Samaraweera
University of Connecticut - Storrs, milinda.samaraweera@uconn.edu

On the Use of the Linear Interaction Energy Method to Predict Affinities of Charged Aromatic Ammines to Naturally Occurring Clay Minerals

Milinda Amila Kumara Samaraweera, Ph.D.

University of Connecticut, 2015

This study presents the use of the linear interaction energy (LIE) method as a predictive tool to approximate the free energies of sorption of organic cations to naturally occurring clay mineral montmorillonite. One objective of this thesis is to explore the applicability and the accuracy of LIE, originated in the biochemistry field, as a predictive tool to estimate the free energies of sorption of organic cations to naturally occurring aluminosilicates. For this purpose, a set of charged aromatic amines sorbing to a prototypical homoionic clay montmorillonite (MMT) with calcium ions were modeled using molecular dynamics (MD) simulations. As the LIE method enables the inclusion of both electrostatic and van der Waals interactions of the sorbate (organic cation) with the negatively charged aluminosilicate (sorbent), it provided a major improvement over existing predictive models which underestimates the sorption free energies due to exclusion of electrostatic interactions. Moreover, Use of MD simulations and electronic structure calculations provided atomistic level insight into the orientation of different organic cations inside the clay and their charge distribution. This thesis also explores the transferability of the derived LIE parameters as a function of different interlayer ions: Ca^{+2} and Na^{+} in MMT.

On the Use of the Linear Interaction Energy Method to Predict Affinities of Charged Aromatic Ammines to Naturally Occurring Clay Minerals

Milinda Amila Kumara Samaraweera

B.Sc. (Chem) Sp., University of Peradeniya 2010

A Dissertation
Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy
at the
UNIVERISTY OF CONNECTICUT
2015

APPROVAL PAGE

Doctor of Philosophy Dissertation

On the Use of the Linear Interaction Energy Method to Predict Affinities of Charged Aromatic Ammines to Naturally Occurring Clay Minerals

Presented by

Milinda Amila Kumara Samaraweera, B.Sc., (Chem) Sp.

Major Advisor_____

José A. Gascón

Associate Advisor_____

Robert R. Birge

Associate Advisor_____

Allison MacKay

Associate Advisor_____

Alfredo Angeles-Boza

Associate Advisor_____

Fatma Selampinar

University of Connecticut

2015

Acknowledgements

I would never have been able to finish my dissertation without the guidance of my major advisor, Prof. José A. Gascón. I would like to express my deepest gratitude for his patience and care.

I would like to thank Prof. Robert R. Birge for his excellent and informative teaching on topics in quantum chemistry, which sparked my interest in computational chemistry. I would also like to thank Prof. Allison MacKay for guiding my research for the past several years and helping me to develop my understanding of clay minerals and their physical properties. Special thanks to Prof. Fatma Selampinar, who was my mentor for most of the years that I worked as a teaching assistant in the chemistry department; her organization, presentation, and time management skills undeniably helped me to become a better chemistry instructor. I would like to thank Prof. Alfredo Angeles-Boza for agreeing to be on my final defense committee. I thank my parents for their support and encouragement. Finally, I would like to thank my wife Thishakya Senadeera, who cheered me up when I was down and provided good support.

Table of Contents

Chapter 1 Introduction

1.1. Background Information on Emerging Environmental Contaminants

Due to the increase in the production and use of various novel organic substances over the last decade, there has also been an increase in the accumulation of these substances in the environment; they are referred to as "emerging contaminants".[1-3] The ecotoxicological and possible health risks for humans and wildlife that are associated with the occurrence of these contaminants is still under investigation and is largely unexplored.[4-6] Emerging contaminants can be categorized into the following key subgroups of substances depending on their application: pharmaceutical and personal care products (PPCPs), hormones and steroids, quaternary ammonium compounds (QACs), illicit drugs and drugs of abuse, polar pesticides, and veterinary products.[7]

1.1.1. Pharmaceutical and Personal Care Products

Pharmaceutical and personal care products, such as x-ray contrasts, anti-inflammatory drugs, pain killers, lipid regulators, fragrances, and sun-screen agents are used to treat and prevent diseases and during daily urban activities.[8, 9] These compounds are released into the environment through domestic and hospital effluents.[10, 11] Unlike pesticides, which are used sporadically, PPCPs are regularly released into the environment.[10] Hence, trace amounts of these compounds are detected in all compartments of the environment (air, soil, and water).[10] PPCPs (Figure 1) usually contain a non-polar core (i.e. phenyl) with a polar functional group (i.e. amine).[12] Their metabolites (transformation products) may be persistent and chemically stable in the environment. For example, the active metabolite of the blood lipid regulator clofibric acid can exist in solution for up to 21 years.[13] Long-term exposure to these compounds and their metabolites are reported to cause unpredicted and adverse side effects on humans and animals alike (e.g. antibiotic

ciprofloxacin and its photodegraded products have shown genotoxic effects in human cell cultures).[14]

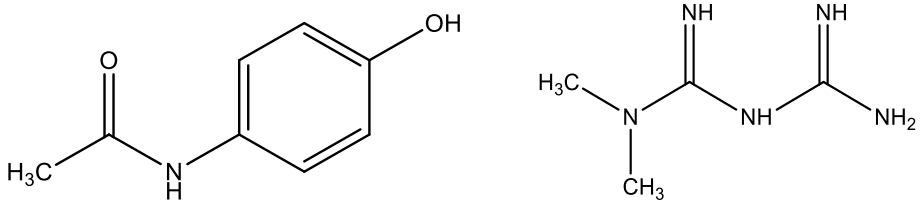


Figure 1. Chemical structures of Acetaminophen (a pain killer) and Metformin (an anti-diabetic drug) on left and right respectively.

1.1.2. Hormones and Steroids

Steroids are biologically active compounds that share a common cyclopentan-o-perhydrophenanthrene ring (Figure 2).[15] Natural steroids (e.g. progesterones, glucocorticoids, androgens) and synthetic steroids (e.g. Mestranol, Sustanon) are extracted by animals and humans, which end up

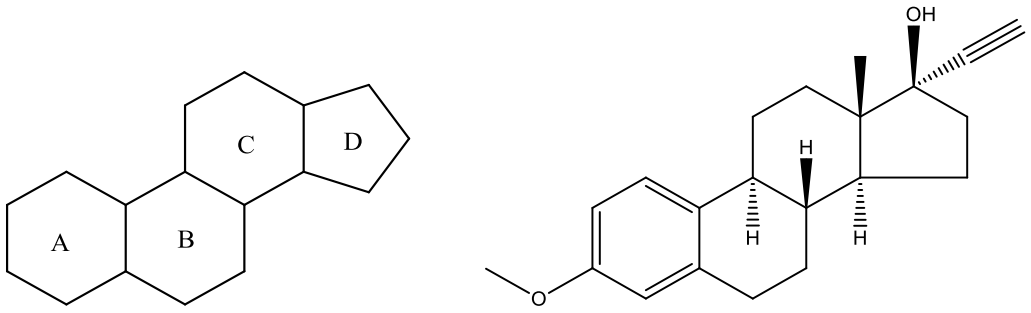


Figure 2. Chemical structure of Cyclopentan-o-perhydrophenanthrene (left) and Mestranol (right)

in the environment through animal waste disposal and sewage discharge.[16] Although these natural and synthetic steroids are excreted from human and animal bodies as inactive polar conjugates, they are activated in sewage effluent as reactive and free steroids.[17] As a consequence, these compounds can interfere with the normal function of endocrine systems, thus influencing the

reproduction and development of marine and wild life.[18-20] Plants are also affected by steroids and hormones in the environment. For example, alfalfa irrigated with sewage effluent showed elevated levels of phytoestrogens.[21]

1.1.3. Quaternary Ammonium Compounds

Quaternary ammonium compounds (QACs) are cationic surfactants that are used in products such as detergents, fabric softeners, and hair care products.[22-24] The typical structure of a QAC contains at least one hydrocarbon chain linked to a positively charged nitrogen atom and other alkyl groups containing short-chain substituents such as benzyl and methyl groups (Figure 3).

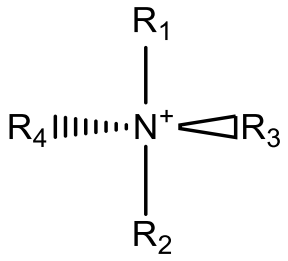


Figure 3. Typical structure of a quaternary ammonium cation. R groups may be the same or different Alkyl or Aryl groups.

The three most abundant QACs are dialkyldimethyl ammonium compounds, alkylmethyl ammonium compounds, and benzylalkyldimethyl ammonium compounds.[24] Release of effluent and sludge from sewage treatment plants is the main source of QACs that are released into the environment.[25, 26] An increase in the saturation of sewage sludge with QACs would adversely impact the anaerobic digestion process by hindering methanogenesis; this would result in an increase in volatile fatty acid accumulation, causing a decrease in the efficiency of activated sludge.[27] Residual amounts of QACs that are released into the environment are also of great concern because they are toxic to a wide range of organisms.[28] The toxicity associated with QACs is more acute than that of anionic surfactants.[27] In microbial colonies that experience long-term

exposure to QACs, the amount of antibiotic-resistant bacteria increases. Furthermore, multidrug resistant genes have been quantified in microbial colonies.[29, 30] This increase in the population and spread of antibiotic-resistant bacteria that contain multi-drug resistant genes will have adverse effects on humans and the environment.[30]

1.1.4. Illicit Drugs and Drugs of Abuse

Illicit drugs are a group of chemicals that have adverse effects on human health and negative effects on the wellbeing of society.[31] According to a world drug report (UNODC, 2013), 150-250 million people have used illicit substances at least once.[32, 33] These drugs include those that are forbidden by law for citizens to make, sell, or use. They mainly belong to the following classes of drugs: cocaine, heroin, amphetamines, and cannabis and its metabolites (Figure 4).[34, 35] Precursors and by-products associated with the manufacturing of these drugs are often illegally buried in soil or disposed of in sinks, rivers, or toilets, from which they enter the public sewage system and accumulate in marine ecosystems.[36] In contrast to legal drugs (pharmaceuticals), these types of drugs contain significant amounts of impurities in addition to the active ingredient. These contaminants pose unknown risks to the environment.[36]

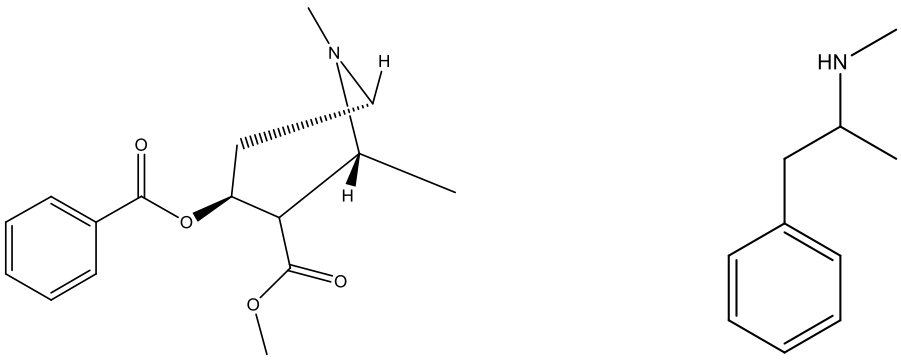


Figure 4. Chemical structures of cocaine (left) and Methamphetamine (right).

1.1.5. Polar Pesticides

There is an extensive amount of polar pesticides (e.g. Atrazine and Monolinuron) that are currently in use.[37, 38] For example, in the South Florida Water Management District it has been estimated that close to 1,415 tons of Atrazine, 622 tons of Chlorpyrifos, and 36 tons of Endosulfan are used per year.[39, 40] This widespread use of pesticides for agricultural and non-agricultural applications have resulted in an increased concentration of their residues in the aquatic environment and the soil (e.g. Atrazine is found in 94% of America's drinking water).[41] As a consequence, humans and marine organisms are constantly exposed to these chemicals.[42, 43] Of numerous studies that have investigated the effects of increased exposure to atrazine and its transformational products, three have shown the number of serious health effects, such as endocrine disruption, reproductive effects, and breast cancer.[44-46]

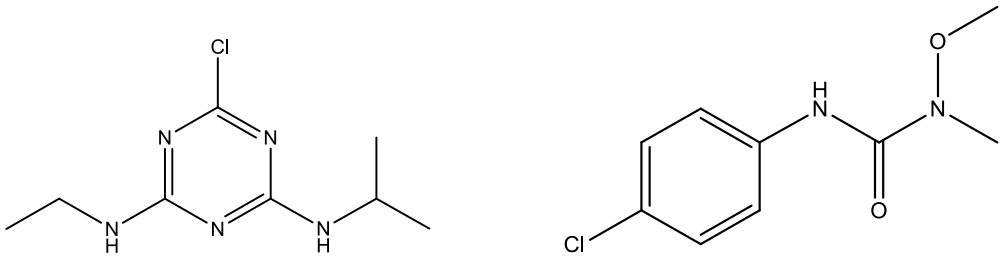


Figure 5. Chemical structure of Atrazine (left) and Monolinuron (right)

1.1.6. Veterinary Products

Veterinary medicines (antibiotics, feed additives, and supplements) are widely used today to prevent and treat diseases and protect the health of animals.[47] In 2010, in the United States only, sales of veterinary medicines totaled $6.5 billion (www.ahi.org). These compounds can enter the environment through many pathways, such as through improper disposal of unused and used medicines and their containers, from animal excrement, and during the formulation and treatment process.[48, 49] As a result, trace amounts of veterinary medicines have been detected in ground water,

surface water, and in the soil, worldwide.[49, 50] Researchers have found that these compounds are toxic to a range of organisms (e.g. fish, earthworms, and algae) and cause antibacterial resistance in the bacterial colonies that affects humans.[51, 52]

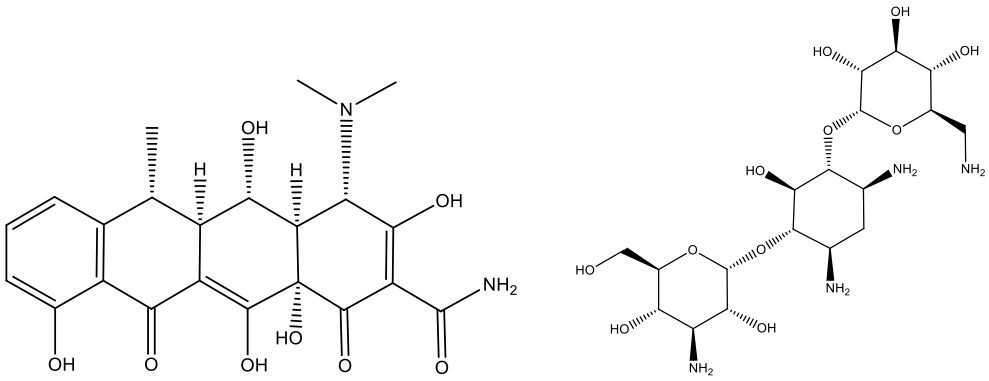


Figure 6. Chemical structure of antibiotics Doxycycline (left) and Kanamycin (Right)

Due to the plethora of adverse effects of emerging contaminants on the wellbeing of humans and animals alike, the investigation of the degradation, accumulation, and transformation (which are collectively known as environmental fate) of the these compounds in different environmental compartments is essential.[53] As discussed, these contaminants are primarily derived from industrial, municipal, and agricultural wastewater effluents.[54, 55] As a consequence, the primary route of exposure to these compounds for humans and animals is through water. Plant uptake of polluted water results in an increased concentration of these pollutants in biomass, which may cause indirect exposure to humans and animals through the consumption of vegetation.[56] The concentration, mobility, and chemo-dynamics of emerging contaminants is largely determined by their adsorption into and desorption from soil and organic matter.[57] Thus, sorption is a vital process that controls the fate of emerging contaminants and their exposure to plant and animal life.

1.2. Background Information about Soil and Clay Minerals

Soil is a complex mixture that is made of three major phases: solid, liquid, and gas. Table 1 illustrates the composition by volume of average soil (these data can vary depending on the soil type and environmental conditions).[58] Minerals and organic matter are the constituents of the solid phase, while pores or voids between the solid materials are occupied by the liquid and gaseous phases.[59]

Table 1. Composition by Volume of Average Soil[60]

Component	Composition by volume (%)
Minerals (solid phase)	45
Organic matter (solid phase)	5
Soil water (liquid phase)	25
Soil air (gaseous phase)	25

The liquid phase (soil solution), which is derived from precipitation and ground-water sources, is a complex mixture of dissolved forms (including the pollutants discussed) and suspended forms.[61] Soil air consists of gases that are derived from gases above the soil surface and from the respiration of organisms in the soil. Interactions between pollutants and soil include various physical and chemical processes that take place in solid, gas, and liquid phases, such as retention, infiltration diffusion through soil solution, and alternation (chemical changes to the structure of the pollutants).[62] Soil is nature's water filter, reducing the concentration and bioavailability of various classes of pollutants, including the emerging contaminants; thus, it reduces the transportation of these pollutants and their exposure to animals.[61]

Clay minerals are a group of hydrous aluminosilicates (size <2μm), a major constituent of the earth's crust (15% by volume).[63] Clays are the major adsorbent for organic and inorganic

adsorbates because of their high surface area and charge.[64] Clays are mainly byproducts that are formed by the weathering of primary minerals (i.e. minerals formed during the crystallization of magma) inside of the earth's crust under high temperature and pressure conditions.[65] Although primary minerals are stable inside of the earth's crust under high temperature and pressure conditions, they may undergo transformation (weathering) under ambient conditions near the Earth's surface and form secondary minerals. The term phyllosilicates is often used to refer to these types of minerals in mineralogy.[65]

The building block of phyllosilicates is the SiO_4 tetrahedron (the Si atom is covalently bonded to four oxygen atoms); SiO_4 tetrahedra form an interlocking array (Figure 7) where each tetrahedron is linked to the adjacent tetrahedrons by sharing three basal oxygen atoms to form a continuous 2D hexagonal mesh pattern (Figure 7).[66]

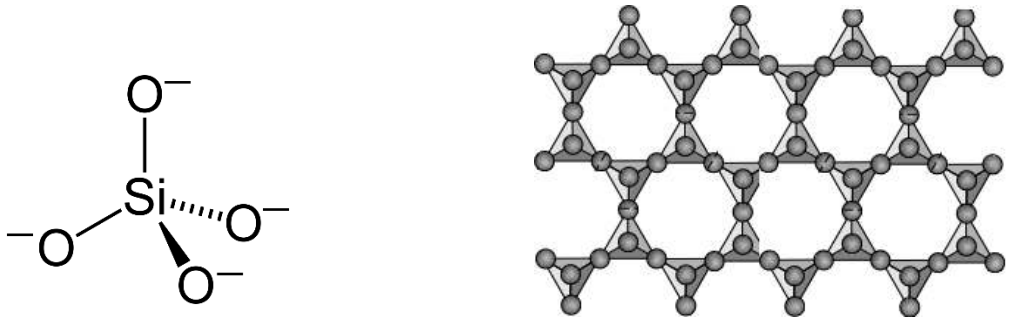


Figure 7. Structure of a silicate tetrahedron on the left and a 2D hexagonal mesh on the right.

External ions (e.g. aluminum, magnesium, and iron) usually bind to this tetrahedral sheet by coordinating with hydroxyl and oxygen anions to form an octahedron (coordinated by six oxygen atoms). The horizontal linkage of multiple octahedrons then forms an octahedral sheet. The combination of tetrahedral and octahedral sheets that are bound by sharing oxygen atoms form aluminosilicate layers that define the basic structural unit of the phyllosilicates.[66] Clay minerals

are classified into three main categories according to the number and arrangement of tetrahedral and octahedral sheets in their primitive structural unit. These categories include:

(1) 1:1 Clays - This type of clay contains one tetrahedral and one octahedral sheet in its unit cells. Clays of this type belong to the kaolin group and have a general formula of $Al_2Si_2O_5(OH)_4$.

(2) 2:1 Clays - This type of clay contains a single octahedral sheet that is sandwiched between two tetrahedral sheets, forming a three-sheet mineral. 2:1 clays include smectite, mica, talc, and vermiculite groups.

(3) 2:1:1 Clays – This clays exhibit a 2:1 layer structure that is similar to talc, except that it has an interlayer of $Mg(OH)_2$ (brucite) that forms a 2:1:1 type of structural arrangement. Chlorites are a group of minerals that belong to this category.

The composition of these clay minerals can vary, due to isomorphous substitution (replacement of one structural cation with another of similar ionic radii) of ions (Si^{+4}, Mg^{+2}) in the clay mineral structure.[67] Si^{+4} (0.039 nm) may be exchanged with Al^{+3}(0.051 nm) in the center of the tetrahedron without any significant structural rearrangements. Further, cations such as Al^{+3} and Mg^{+2} (0.066 nm) may be replaced by Zn^{+2} (0.074 nm) and $Fe^{+2/+3}$ (0.074 nm). As a result of this isomorphous substitution, clay minerals will have a negative net charge. For example, the substitution of one Mg^{+2} ion for an Al^{+3} ion in the octahedron reduces the charge by a value of one.[67]

As previously stated, Kaolinites are an example of a 1:1 type of clay; they are hydrated aluminosilicates with a chemical composition of Al_2O_3:SiO_2:H_2O = 1:2:2. In Kaolinites there is little isomorphous substitution hence they are mostly electrically neutral. The two clay sheets are held together by strong hydrogen bonding, rather than ionic forces. Kaolinites possess surface

hydroxyl groups that can be deprotonated under basic pH conditions; thus, their charge is pH dependent.[68]

Smectites, an example of a 2:1 type of clay mineral (Figure 8), have a permanent negative charge; this is primarily caused by the isomorphous substitution of Mg^{+2} with Al^{+3}. This negative charge is balanced out by cations in the interlayer space (Na^{+}, Fe^{+2}), which are hydrated (depending on the relative humidity) and exchangeable. The general structure of smectite clay is illustrated in Figure 8:[69]

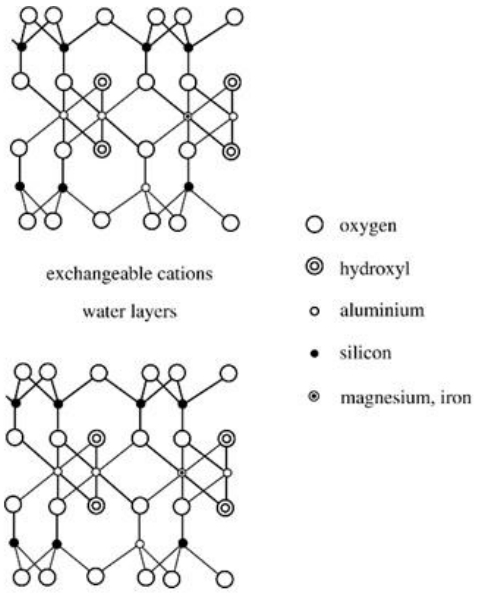


Figure 8. Typical structure of a smectite clay

Smectites have a large total surface area, close to 800 m^2/g.[70] Due to weak interactions between clay sheets, smectites undergo swelling when in contact with water. Upon dehydration, the expanded clay sheets will shrink. As a result of this swell-shrink capability, smectites have the ability to adsorb and retain inorganic cations and organic compounds. Inorganic cations can undergo cation exchange with organic cations. Thus, smectites are ideal host materials for positively charged organic cations in the environment.[71-73]

Finally, the chlorite mineral group, a 2:1:1 type of clay exists as a mixture of several clay species. Due to their wide range of chemical compositions, these minerals can vary significantly in physical and chemical properties.[74]

Smectite montmorillonite was used as the representative aluminosilicate for the studies carried out in this thesis because of its abundancy and availability of high quality crystal structures.

1.3. Environmental Risk Assessment and the Significance of the Soil Sorption Coefficient

The life cycle of a pollutant after its release into the environment (fate), and the availability, mobility, and degradability of any pollutants or mixtures of pollutants in the environment have a strong effect on their toxicity and environmental risk.[75] Further, knowledge of these pollutant's behavior and fate, such as partitioning between physical phases (e.g. solid and liquid), biodegradation, and photo degradation can provide vital insight into methods for remediation of contaminated areas.[76, 77] The process of risk assessment can be defined as: estimating the possibility of a certain event occurring under a given set of circumstances.[78] Environmental risk assessment involves an analysis of data on the fate and behavior of chemicals in the environment (e.g. water and soil), combined with an analysis of their adverse effects on the biosphere. Risk assessment is a vital tool for decision making that can be used to identify existing problems and predict potential risks, such that necessary steps can be taken to weaken the adverse effects.[79]

When a compound (any chemical species) enters the environment, it becomes distributed among three major compartments: water, air, and soil.[80] The fraction of the compound that will move into each compartment (fate) is controlled by the physio-chemical characteristics (e.g. charge, polarity, and size) of that species.[81] Since the majority of emerging contaminants are

introduced into the environment through effluents (domestic, agricultural, and industrial), studying the partition (sorption to soil from water) of these contaminants in water and soil compartments is vital. Furthermore, sorption is an important factor that determines the mobility, reactivity, and bioavailability of a compound in the environment.[12] Several mechanisms can be used to explain the sorption process of a compound to soil and soil organic matter depending on the structural characteristics of the compound (pollutant) such as functional group content, ionic nature, polarity, and solubility.[82] It is also possible that more than one of these mechanisms are involved in the sorption process, depending on the structural features of the compound.[81, 83] These mechanisms include:

(1) Hydrophobic partitioning: sorption due to hydrophobic partitioning of a molecule is driven by the coupled effects resulting from unfavorable high energy cost of cavitation in water and the favorable van der Waals interactions between the sorbate and the environmental solid. This mechanism is important for sorption of nonpolar compounds on to organic matter and poorly solvated mineral surfaces.[81]

(2) Electrostatic interactions: sorption due to attraction between oppositely charged sorbents and sorbates, electrostatic interactions can be further subdivided in to:

   I. Electron donor-acceptor (EDA) interactions: these type of interactions occurs due to transfer of electrons between electron-rich or -poor functional groups of the sorbate with the complementary electron-poor or -rich domains of the environmental solid which results in a weak electrostatic attraction.[81] Hydrogen bonding is a special case of EDA complexes.

   II. Ion exchange: occurs due to the electrostatic attraction between a charged sorbate and an oppositely charged site on an environmental solid. More precisely, this process

involves the exchange of electrostatically bound counter ions, which balances the overall surface charge of a mineral (e.g. interlayer ions in smectites) with another ion of the same charge (positive or negative).[84]

III. Cation bridging: a process where sorption of sorbates having negatively charged groups (COO^-) coordinating through sorbed cations of environmental solids (Ca^{+2}, Mg^{+2}).

(3) Surface complexation: occurs as a result of ligand exchange process where a hydroxyl group or a water molecule bound to surface Al or Fe atoms is complexed or exchanged with an organic sorbate having functional groups such as NH, COOH.

The environmental fate assessment of organic compounds requires knowledge about their sorption coefficients (i.e. the ability of a solid to sorb a certain compound).[85] The sorption coefficient, K_d (L/kg), describes the distribution of a solute between the solid and aqueous phases at equilibrium:

$$K_d = \frac{C_s}{C_w} \tag{1.1}$$

where C_s (mass/kg) is the sorbed concentration and C_w (mass/L) is the concentration of the solute in solution. In a recent work carried out by Mackay and Vasudevan *et al.,* a more complete description of the sorption coefficient considering the sorption mechanisms of various chemical compounds (polar, non- polar, and charged) and considering different sorbed forms and different modes of interaction (e.g. cation exchange and hydrophobic interactions) has been proposed:[81]

$$K_d = \frac{C_{s,Type1} + C_{s,CE} + C_{s,SC+CB}}{C_w} \tag{1.2}$$

Where $C_{s,Type1}$ is the concentration of sorbates (polar/apolar molecules) adsorbing through van der Waals interactions, solvation effects and electron donor-acceptor interactions (denoted as *Type 1*), $C_{s,CE}$ is the concentration of sorbate adsorbed by cation exchange, $C_{s,SC+CB}$ is the concentration of

sorbate adsorbed by surface complexation and cation bridging, C_w is the common dissolved concentration of the sorbate.

The experimental determination of K_d is labor intensive and time consuming; it is estimated that close to 50,000 -100,000 different chemicals are produced for commercial purposes each year and thousands are added.[53] Therefore, alternate approaches to determining the values of K_d have been explored. One plausible approach is the use of predictive models. Predictive models use statistical methods to foretell future outcomes, based on patterns and trends observed in existing data.[86-88]

The extent of the contribution from each term in equation 1.2 depends on the structural features of the sorbate (e.g. polarity, functional group content, ionic nature) and the composition of the soil.[81] For compounds that are apolar or polar, where van der Waals and dipole-dipole types of interactions are dominant, first and second terms have the highest contribution to K_d (other terms are negligible).[81] Numerous predictive models have been developed to successfully approximate K_d values of these compounds based on octanol-water partition coefficient and Abrahams solvation models.[89, 90].

Emerging contaminants such as PPCPs and illicit drugs involve bases (amine groups) with a pK_a above 10 and the pH of natural water falls is between 6.5 and 8.5.[73] As a consequence, most of these compounds are protonated and exist in their cationic form.[73] Moreover, compounds like oxytetracycline that possess both amine and carboxylic groups (with pK_as around 3.27) exist as zwitter ions. These organic solutes are adsorbed to negatively charged substrates, such as smectites and natural organic matter through cation exchange, surface complexation, and cation bridging.[81] For these types of compounds, sorption through ion exchange is significant in contrast to the non-polar and apolar compounds discussed above. Thus, empirical models (or predictive models) that

are currently utilized (octanol-water partition coefficient and Abrahams solvation models) to approximate K_d for polar and apolar compounds, will severely underestimate the value of K_d if used for charged compounds, since these predictive models does not correctly account electrostatic interactions, which is a major driving force of sorption. Thus, the determination of K_d for organic cations is solely based on experimental methods.[81]

To overcome this problem, Mackay and Vasudevan have recently proposed the initial steps for developing predictive models to estimate the partition coefficient (K_d) by incorporating mathematical functions that reflect structural features (e.g. charge, functional group content) of the sorbate and key characteristics of the environmental solid (e.g. cation exchange capacity) through a chemical probe.[81] They also demonstrated that by using a homologous series of cationic ammines, structural changes impart a regular increase (linear) in the free energy of sorption through cation exchange, due to coupled hydrophobic and electronic effects.[81] However, experimental observations lack the capability to decouple the contributions from each mode of interaction (hydrophobic and electronic); thus, knowledge of these interactions is limited at a microscopic level. Subsequently, they were unable to derive the exact form of the equation used to predict K_d, as a better understanding of individual contributions from hydrophobic and electronic effects was required. Thus, the necessity of an alternate approach to deciphering this problem was indicated.[81]

The free energy of a process is related to the equilibrium constant K (K_d)[91] through the following equation:

$$\Delta G = -RT \ln K \tag{1.3}$$

where R is the gas constant and T is absolute temperature. According to equation 1.3, the value of any equilibrium constant (K_d) can be easily calculated if the free energy associated with the process

can be determined (i.e. the sorption of a chemical from a solution to the interlayer of a clay). Computational biochemists, in the last decade, have developed numerous highly accurate methods for calculating the free energy in ligand binding processes. The microscopic (atomistic) nature of these methods allows scientists to magnify into the molecular level and decouple contributions from different structural moieties (e.g. electronic, polar, and apolar) to the overall binding energy, which was not feasible in macroscopic experimental methods. In the next chapter, a brief introduction to the core theoretical concepts of free energy, the statistical mechanics basis of free energy, and the computational tools utilized in biochemistry to obtain the free energy of binding are discussed. Furthermore, the applicability of such methods to computing the free energy of the sorption of compounds from the solution phase to the solid phase (clays) is explored.

Chapter 2 Theoretical Background

2.1. Statistical Mechanical Basis of Free Energy

Free energy is an important factor that determines the feasibility of a chemical process such as protein-ligand binding, and the extent of partitioning of a compound among two or more phases. In thermodynamics, the free energy of a system is defined as usable energy that can be converted to work.[91] Depending on the thermodynamic state of the system, free energy can be expressed by either Gibbs (G) or Helmholtz (A) functions. For a system at constant temperature and pressure, Gibbs free energy is used, while for a system at constant temperature and volume, Helmholtz free energy is used.[92]

Helmholtz free energy is defined as:

$$A = U - TS \tag{2.1}$$

where A is Helmholtz free energy, U is internal energy, T is the temperature, and S is the entropy. The Gibbs free energy is the Helmholtz free energy plus the additional pV term, where p is pressure and V is volume:

$$G = A + pV \tag{2.2}$$

In many cases Gibbs and Helmholtz free energies are exchangeable with respect to the end result. Thus, the Helmholtz free energy is used for the remainder of this discussion.[92]

Thermodynamic relations like equation 2.1 provide an average description of the system, but does not provide detailed understanding into the atomistic level insight which generates these observables.[93] Statistical thermodynamics provides the missing link that connects observable thermodynamic properties and their molecular properties. In statistical mechanics, thermodynamic observables are described as the average property of the large number of particles that a system is composed of.[93] In order to define a microscopic system with N number of particles, the coordinates

and velocities of each particle must be determined. At any given instant, a system will have a new set of coordinates and velocities due to collisions between particles. Thus, the properties of each particle will change at any given point in time. In practice, such a description of a system is not feasible, due to the extremely high number of possible configurations.[93] However, each configuration can be ranked according to the probability. Thus, some configurations of the system will be more probable than others. For a large *N*, as in an actual thermodynamic system, the most probable number of configurations will dominate other configurations that are less probable, and will represent the average properties of the system.[94]

The Boltzmann distribution can be used to calculate the average number of particles in the most probable configuration. Assuming the constraints for the total energy and the number of particles in the system to be constant,[94] the fraction of particles n_i/N in configuration i with energy E_i is given by:

$$\frac{n_i}{N} = \frac{e^{-\beta E_i}}{\sum_i e^{-\beta E_i}} \tag{2.3}$$

Where $\beta = \frac{1}{k_b T}$, the Boltzmann constant is (k_b), and absolute temperature is (T); n_i will decrease exponentially with an increase in energy. $\sum_i e^{-\beta E_i}$, in the denominator, is termed the molecular partition function, q:

$$q = \sum_i e^{-\beta E_i} \tag{2.4}$$

The magnitude of q provides an indication of the number of thermally accessible states of a molecule at a given temperature. Using equation 2.4, total energy can be expressed as a function of its molecular properties:

$$E = \sum_i n_i E_i \tag{2.5}$$

$$E = \frac{N}{q} \sum_i E_i e^{-\beta E_i} \tag{2.6}$$

At a first approximation, the molecular partition function (q) is derived for a system of non-interacting molecules; hence, E is constant. To treat systems with interacting particles, the concept of an ensemble is used. An ensemble is a large collection of replications of a system.[91] Even though the energy of an individual microstate can fluctuate due to the interaction of particles, the total energy of the ensemble will remain constant. Therefore, the average energy (E) of each member will be constant.[90] Equations 2.5 and 2.6 can be re-written as:

$$E = \frac{1}{N} \sum_i E_i \tag{2.7}$$

$$E = \frac{1}{Q} \sum_i E_i e^{-\beta E_i} \tag{2.8}$$

where E_i is the energy of a single replicate of collection of N replications, and Q is the canonical partition function. Q is related to the molecular partition function (q) by: $Q = q^N$ for independent and distinguishable particles, and $Q = \frac{q^N}{N!}$, for independent and indistinguishable particles. The Boltzmann distribution provides the most probable configurations of the ensemble, where n_i number of members of the ensemble have energy E_i.

The total energy (E) corresponds to the value of the internal energy (U) of the system relative to its value at T=0: $U=U(0)+E$. For simplicity, assume that $U(0)=0$, and therefore $U=E$. From thermodynamics, at constant volume V, a reversible change in the heat of the system (dq_{rev}) is equal to the change in internal energy (dU) and dq_{rev} is related to the infinitesimal change in the entropy of the system (dS) at temperature T:

$$dU = dq_{rev} = TdS \tag{2.9}$$

By combining equations 2.9 and 2.8, it is possible to describe entropy in terms of the microscopic properties of a system:

$$S = \frac{U}{T} + k \ln Q \tag{2.10}$$

Once the partition function of a system is known, all of the other properties of the system can be expressed as a function of Q. Thus, statistical thermodynamics provides a molecular theory of the equilibrium properties of macroscopic systems.[95] Using equation 2.10, the equation for Helmholtz free energy (A), can be derived as:

$$A = -kT \ln Q \tag{2.11}$$

Q is the partition function for a system with a constant number of particles, volume, and temperature. In reality, absolute free energy is not used; rather, free energy changes associated with changes in state are determined:

$$\Delta A = -kT \ln \frac{Q_1}{Q2} \tag{2.12}$$

where Q_1 and Q_2 are the partition functions that correspond to states 1 and 2. ΔA is the change in free energy associated with a transition between the two states. The measure of change in free energy is an important quantity for predicting the spontaneity of a process. A spontaneous process often leads to a reduction in the free energy of a system. Thus, ΔA will have a negative value for a spontaneous reaction. The higher the negative value is, the more feasible the process is.[91] Both experimental (macroscopic) and molecular modeling (based on statistical mechanics) approaches have been utilized extensively to obtain free energy changes, as knowledge about the free energy of a process is important.[96, 97]

Molecular modeling methods are used to mimic the behavior of molecular systems. This is achieved with the aid of computer modeling. Over the last decade, because of the increase in computational power, a wide range of molecular systems have been modeled.[98, 99] Depending on the size and type of the system simulated, different methods can be applied. The two most widely used methods are force field and electronic structure methods. Electronic structure methods, also known as quantum mechanical methods, are used to model systems with high accuracy that account for the quantum behavior of matter. Because of the limitations of computational resources, only a small system can be modeled (a system with less than 100 atoms) using quantum mechanical methods. In contrast, force field (molecular mechanics) methods use a simple set of equations to define the interactions of molecules. Thus, they can be used to model systems that have thousands of atoms. A time evaluation of a system is achieved using integration algorithms.[100]

2.2. Background Information about Molecular Mechanics

In molecular mechanics (MM) models, the motion of the electrons is ignored and the atoms are treated as a set of interacting spheres (balls connected using springs).[101] Thus, in MM models, the quantum nature of nuclear motion is ignored and the dynamics of atoms are described by classical Newtonian mechanics.[102] Thus, MM methods are not suitable for modeling properties that are dependent on electron dynamics and that are quantum in nature, such as chemical reactions, excited states, and charge-transfer interactions.[103] In an MM system, the term force field (FF) is used to name the set of equations that describes the interactions among the atoms in the system.[101] The total energy of an MM system (E_{MM}) is calculated as the sum of different energy terms:

$$E_{MM} = E_{BOND} + E_{ANGLE} + E_{TOR} + E_{VDW} + E_{EL} \tag{2.13}$$

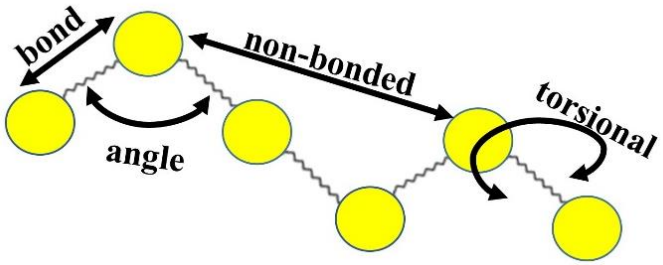


Figure 9. Fundamental force field energy terms

where E_{BOND} is the energy associated with the stretching of a bond, E_{ANGLE} is the energy associated with the bending of an angle, E_{TOR} is the energy associated with the rotation of a bond, and E_{EL} and E_{VDW} represent electrostatic and van der Waals interaction energies between non-bonded atom pairs.

Each energy term in an FF (equation 2.13) can be expressed using a simple equation that is a function of the atomic coordinates (equation 2.14). In most of the FFs that exist today (e.g. OPLS, AMBER, DREIDING), stretching and bending energies are modeled using simple harmonic functions; torsional energies are modeled using periodic functions that capture the periodic nature of torsional energies; and Columbic and Van der Waals interactions (non-bonded interactions) are modeled using Coulomb and Lennard-Jones potentials, respectively.[104-106] An example functional form of a force field is given below:

$$E_{FF} = \sum_{bonds} k_b(b-b_0)^2 + \sum_{angles} k_\theta(\theta-\theta_0)^2 + \sum_{dihedrals} \frac{V_1}{2}[1+\cos(\phi-\phi_0)] + \frac{V_2}{2}[1-\cos(2\phi-\phi_0)] + \frac{V_3}{2}[1+\cos(3\phi-\phi_0)] + \sum_{i>j} 4\varepsilon[(\frac{\sigma_{ij}}{r_{ij}})^{12} - (\frac{\sigma_{ij}}{r_{ij}})^{6}] + \sum_{i>j} \frac{q_i q_j}{D r_{ij}} \quad (2.14)$$

where k_b is bond force constant, $(b\text{-}b_0)$ is the distance from equilibrium that an atom has moved, k_θ is the angle force constant, $(\theta\text{-}\theta_0)$ is the angle from the equilibrium between three bonded atoms,

k_α is the dihedral force constant, n is the multiplicity of the function, and δ is the phase shift. In non-bonded interactions ij represents two atoms, σ_{ij} and ε are van der Waals constants, r is the distance between two atoms, q represents the charge, and D is the effective dielectric function for the medium.

One of the most important steps in FF development after figuring out an appropriate functional form is the determination of FF parameters (e.g. equilibrium bond lengths and force constants).[107] Typically, these parameters are fitted to reproduce reference data from experimental or electronic structure calculations. MM simulations assume that molecules are composed of structural units that are similar in different molecules. Using this assumption, FF parameters (force constants and Leannard-Jones parameters) can be developed for small molecules that can also be used in large molecules (polymers, proteins). Thus, for example, C-H bond parameters that are derived for methane (CH_4) molecules can be used to model a C-H bond of a decane ($C_{10}H_{22}$) molecule. The study described in this thesis utilized OPLS-AA, CLAYFF, and SPC FFs. A brief introduction to these FFs is provided in sections 2.2.1 and 2.2.2.

2.2.1. Background Information about OPLS-AA and CLAYFF Force Fields

The optimized potentials for a liquid simulation (OPLS) force field was developed by Jorgensen *et al*.[108] The OPLS force field can be divided into two classes: the OPLS-AA (all-atom) and OPLS-UA (united-atom). OPLS-UA is primarily used for engineering applications, since it is computationally cheaper than the all-atom version. The OPLS-AA force field is parameterized for small organic molecules and is intended to simulate biomolecules.[108] The OPLS-AA force field has been parameterized to simulate a broad range of functional groups and molecules such as alcohols, thiols, sulfides, ketones, amides, ammines, hydrocarbons, pyrrole, and diazoles.[108] In OPLS-AA, bond stretching and angle bending parameters are modeled using harmonic potentials, the torsional motion is modeled using a Fourier series that is truncated after the third term, van der

Waals interactions are modeled using a Lennard-Jones 12-6 potential, and electrostatic interactions are modeled using point charges. The LJ and coulomb potentials are used to describe inter and intra molecular interactions (atoms which are three or more bonds apart are scaled by a factor of 0.5).[108] In OPLS-AA, bond stretching and angle bending parameters were taken from the AMBER94 and CHARMM force fields, while torsion potentials were obtained by fitting to ab initio calculations (MP2).[108] The parameterization of LJ parameters and partial charges was conducted to reproduce the saturated liquid density and enthalpy of vaporization.[108] In later versions of the force field, partial charges were fitted to the ab initio electrostatic potential with the CHELPG method.[109]

The terms used in the OPLS-AA force field are:

$$E_{OPLS} = \sum_{bonds} k_b (b - b_0)^2 + \sum_{angles} k_\theta (\theta - \theta_0)^2 + \sum_{dihedrals} \frac{V_1}{2}[1 + \cos(\phi - \phi_0)] + \frac{V_2}{2}[1 - \cos(2\phi + \phi_0)] + \frac{V_3}{2}[1 + \cos(3\phi + \phi_0)] + \sum_{i>j} f_{ij}[4\varepsilon[(\frac{\sigma_{ij}}{r_{ij}})^{12} - (\frac{\sigma_{ij}}{r_{ij}})^6] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}] \quad (2.15)$$

where k_b and k_θ are the force constants, r is the bond distance, θ is the angle, ∅ is the torsion angle, A and B are Lennard-Jones parameters, q is partial charge, ε_0 is permittivity in a vacuum, and i and j indicate any two atoms. The subscript 0 symbolizes the equilibrium value. f_{ij} is a 'fudge factor', where f_{ij} = 0.5 for atoms that are four bonds away and f_{ij} =1.0 for all other non-bonded interactions.

CLAYFF is a force field developed by Cygan *et al.* to model mineral systems and their interfaces with aqueous solutions.[110] In the CLAYFF force field, metal-oxygen interactions are treated as non-bonded (using Lennard-Jones potential with electrostatics). Non-bonded parameters are optimized on the basis of known mineral structures, and partial charges are derived from

quantum mechanical simulations using model oxide, hydroxide, and oxyhydroxide structures. The flexible simple point charge (SPC) water model is used to model water and hydroxyl groups.[110]

CLAYFF has been successfully applied to model silicate, aluminate, and other metal oxide and hydroxide systems.[111-114] The functional form of the CLAYFF force field is:

$$E_{CLAYFF} = \sum_{bonds} k_b (r - r_0)^2 + \sum_{angles} k_\theta (\theta - \theta_0)^2 + \frac{e^2}{4\pi\varepsilon_0} \sum_{i>j} \frac{q_i q_j}{r_{ij}} + \sum_{i>j} D_{0,ij} [(\frac{R_{0,ij}}{r_{ij}})^{12} - (\frac{R_{0,ij}}{r_{ij}})^6] \quad (2.16)$$

where k_b and k_θ are the force constants, r is the bond distance, θ is the angle, R and D are Lennard-Jones parameters, q is partial charge, ε_0 is permittivity in a vacuum, and i and j indicate any two atoms. The subscript 0 symbolizes the equilibrium value.

2.2.2. Common Water Models

In computational chemistry, classical water models are used for accurate simulation of the liquid state of water.[115] Due to its abundance and extraordinary chemical properties, the effects of water are important for the processes involved in biochemistry, organic chemistry, and environmental science.[116, 117] Due to the presence of two lone pairs in the oxygen atom, water deviates from its tetrahedral structure (angle of H-O-H 109°) to a bent structure that has an H-O-H angle of a 105°.[118] Because of the high electronegativity of the O atom, there is a higher probability of finding the H-electrons near the O atom. Hence, water is a polar molecule. This asymmetric electron distribution causes H atoms to be partially positively charged. Thus, H-atoms can form strong electrostatic interactions with O-atoms in another water molecule (or with other molecules that have an electronegative atom such as nitrogen or fluorine; this is referred to as hydrogen bonding.[119]

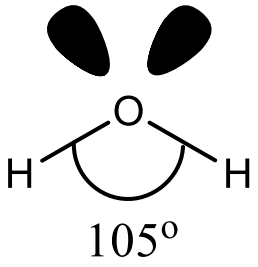


Figure 10. Structure of water

For an accurate simulation of water, a model that embodies the polarity of the water molecules and permits the formation hydrogen bonds is vital. Several water models have been developed in the last decade, ranging from simple models such as TIP3P and flexible SPC/SPC-E (three-site models), to more complex models such as TIP4P (four-site models) and TIP5P (five-site models).[120, 121] Selection of the proper water model is an important step for successful simulation. Since the CLAYFF force field uses a well-tested flexible SPC water model, the flexible SPC force field was used to model water in this study.

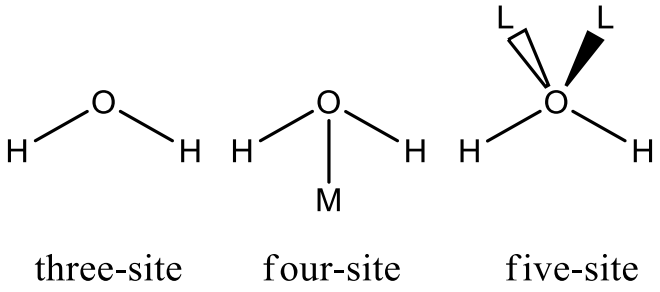


Figure 11. Illustration of simple water models

Simple point charge (SPC) is one of the most accurate three-site water models (Figure 11) and it was developed by Berendsen *et al.*[122] One of the initial drawbacks of this model was that water was treated as rigid molecules (intramolecular degrees of freedom are frozen). To overcome this limitation, Voth *et al.* developed a flexible variation of the SPC water model that introduced

flexibility into the original model.[123] Flexibility was achieved by introducing bond stretching and bending parameters to the FF. Flexible SPC accurately reproduces bulk properties of water and is used extensively in studies that are reported in the literature.[123-125] This model uses tetrahedral geometry to model the H-O-H angle (H-O-H angle is set to 109.5°); the O-H distance is set to 0.1 nm with partial charges on the oxygen, and hydrogen atoms are set to -0.82*e* and 0.41*e*.[123]

2.3. Background Information about Molecular Dynamics Simulations

Molecular dynamics (MD) includes deterministic methods that are used to study the dynamical properties of an *N*-body system in equilibrium; it is based on FF methods.[126] During the last decade MD has become a vital tool, not only in theoretical chemistry, but in material science, environmental engineering, and biochemistry, as MD provides insight into molecular processes that are difficult to obtain during an experiment.[127-129] MD uses classical or Newtonian mechanics, where the motions of atoms are described by:

$$\frac{d\mathbf{r}_i}{dt} = \mathbf{v}_i \qquad (i = 1, 2, 3, \ldots) \tag{2.16}$$

$$\mathbf{F}_i = m_i \frac{d^2\mathbf{r}_i}{dt^2} \quad (i = 1, 2, 3, \ldots) \tag{2.17}$$

where $\mathbf{r_i}$, $\mathbf{v_i}$, and m_i are the position, velocity, and mass of an atom; $\mathbf{F_i}$ is the force acting on an atom *i*. The differential form of equation 2.17 (Newton's second equation) is:

$$m_i \frac{d^2\mathbf{r}_i}{dt^2} = -\frac{dU_i}{d\mathbf{r}_i} \qquad (i = 1, 2, 3, \ldots) \tag{2.18}$$

Here *U(i)* is the potential energy at position **r**, which is obtained using a force field, as discussed above. Thus, the right side of the equation contains the negative of the energy gradient, which provides the force (**F**) on the particles.[130]

Using a Taylor expansion, the new positions of a set of particles after a small time step Δt can be obtained, given the initial positions $\mathbf{r_i}$ and time t:

$$\mathbf{r}(t+\Delta t) = \mathbf{r}_t + \frac{\delta \mathbf{r}}{\delta t}(\Delta t) + \frac{1}{2}\frac{\delta^2 \mathbf{r}}{\delta t^2}(\Delta t)^2 + \frac{1}{6}\frac{\delta^3 \mathbf{r}}{\delta t^3}(\Delta t)^3 + \ldots \tag{2.19}$$

the above equation as can be rewritten as:

$$\mathbf{r}(t+\Delta t) = \mathbf{r}_t + \mathbf{v_t}(\Delta t) + \frac{1}{2}\mathbf{a_t}(\Delta t)^2 + \frac{1}{6}\mathbf{b_t}(\Delta t)^3 + \ldots \tag{2.20}$$

where velocity (v_i) is the first derivative of the position with respect to time at time t_i, acceleration (a_i) is the second derivative of the position with respect to time at time t_i, and hyper-acceleration (b_i) is the third derivative of the position with respect to time at time t_i. Using a similar approach, the previous position of a set of particles a small time step Δt can be obtained, given the initial positions r_i and time t:

$$\mathbf{r}(t-\Delta t) = \mathbf{r}_t - \frac{\delta \mathbf{r}}{\delta t}(\Delta t) + \frac{1}{2}\frac{\delta^2 \mathbf{r}}{\delta t^2}(\Delta t)^2 - \frac{1}{6}\frac{\delta^3 \mathbf{r}}{\delta t^3}(\Delta t)^3 + \ldots \tag{2.21}$$

Adding equations 2.20 and 2.21 provides a formula that predicts the position one time step Δt after the current and previous positions and the current acceleration:

$$\mathbf{r}(t+\Delta t) = (2\mathbf{r_t} - \mathbf{r_t} - \Delta t) + \mathbf{a}(\Delta t)^2 + \ldots \tag{2.22}$$

This method is known as the Verlet algorithm for solving Newton's equations numerically. Although previous positions are unavailable at the initial point, they can be estimated from a first-order approximation of equation 2.20:

$$\mathbf{r}_{-\Delta t} = \mathbf{r}_0 - \mathbf{v}_0 \Delta t \tag{2.23}$$

Acceleration is evaluated at each time step using the forces (equation 2.18), which then permits the propagation of the atomic positions in time, generating a trajectory.

The Verlet algorithm suffers from two major drawbacks:

1) A numerical disadvantage since Δt^2 (a small value) is added to $2\mathbf{r}(t) - \mathbf{r}(t - \Delta t)$ (a large value) to obtain new positions. This might cause truncation errors due to finite precision.
2) Velocities do not appear explicitly, which is a problem when generating ensembles with a constant temperature.

To overcome these problems, a velocity verlet (VV) algorithm was introduced.[131] In VV positions, velocities and accelerations at time $t + \Delta t$ are obtained at time t using the following method:

$$\mathbf{r}(t + \Delta t) = \mathbf{r}_t + \mathbf{v}_t(\Delta t)^2 \tag{2.24}$$

$$\mathbf{v}(t + \Delta t) = \mathbf{v}_t + \mathbf{a}(t) + \frac{\mathbf{a}(t + \Delta t)}{2}(\Delta t) \tag{2.25}$$

The velocity Verlet algorithm includes the following steps:

1) Calculate: $\mathbf{r}(t + \Delta t) = \mathbf{r}_t + \mathbf{v}_t(\Delta t)^2 + \frac{1}{2}\mathbf{a}_t(\Delta t)^2$
2) Derive $\mathbf{a}_t(t + \Delta t)$ from the gradient of the potential using $\mathbf{r}(t + \Delta t)$
3) Calculate: $\mathbf{v}(t + \Delta t) = \mathbf{v}(t) + \mathbf{a}(t) + \frac{\mathbf{a}(t + \Delta t)}{2}(\Delta t)$

The time step (Δt) is an important parameter that determines the accuracy of a simulation. It should be selected smaller than the fastest process in the system. The fastest molecular motions, molecular rotations, and vibrations typically occur in the frequency range of 10^{11}-10^{14} s^{-1}. Thus, a time step of 1 *fs* is usually used to carry out MD simulations. A standard MD simulates a microcanonical ensemble (number of particles (N), volume (V), and energy (E) are constant).[131]

As discussed in section 2.1, the thermodynamic properties of a system can be derived if the partition function *Q* is known. Since a molecular dynamics simulation generates the most probable configurations of the ensemble, those configurations can be used to estimate *Q*. In this setting, the

expressions for the internal energy (U) and the Helmholtz free energy (A) in equations 2.7 and 2.8, respectively, can be rewritten in terms of an ensemble average (denoted by < >) over N configurations of the system:

$$\langle U \rangle_N = \frac{1}{N}\sum_i^N E_i = \langle E \rangle_N \tag{2.26}$$

$$\langle A \rangle_N = kt \ln(\frac{1}{N}\sum_i^N e^{\frac{E_i}{kt}}) = kt \ln\left\langle e^{\frac{E}{kt}} \right\rangle_N \tag{2.27}$$

2.3.1 Thermostats and Barostats

As stated above, a generic MD simulation produces an isolated micro-canonical ensemble (a microcanonical ensemble defines a thermodynamic system that has a constant number of particles (N), volume (V), and energy (E): NVE ensemble). Unfortunately, in most experiments, temperature and pressure are important control variables, as the temperature, pressure, and volume of a system is maintained at specific values. Therefore, the integration algorithm (e.g. velocity Verlet) used in an MD simulation must be modified to include the influence of external temperature and pressure. This can be achieved by introducing thermostats and barostats. Only after such a modification to the integration algorithm, can the canonical (with a constant number of particles (N), volume (V), and temperature (T)) or the isothermal-isobaric (with a constant number of particles (N), pressure (P), and temperature (T)) ensembles be generated.

In the canonical ensemble, temperature (T), volume (V), and number of particles (N) are fixed and energy is allowed to fluctuate (NVT). A straightforward approach to controlling the temperature is scaling the velocities of the particles (velocity scaling method). In velocity scaling, the velocities are multiplied by a factor of $\lambda = \sqrt{T_0/T(t)}$, where $T(t)$ is the temperature at time t, calculated from kinetic energy, and T_0 is the desired temperature. Unfortunately, this approach

does not permit fluctuations in kinetic energy, which cause an oscillation in the system's total energy.

A more accurate approach is the use of the Berensden thermostat.[132] In this method, the temperature is maintained by coupling the system with an external heat bath that has a fixed temperature T_0.[132] At each iteration of the integration algorithm, velocities are scaled such that the rate of change of temperature is proportional to the change in temperature:

$$\frac{dT(t)}{dt} = \frac{1}{\tau}(T_0 - T(t)) \tag{2.28}$$

τ is a coupling parameter that is used to set the tightness of the system with the external heat bath. The change in temperature of each iteration is given as:

$$\Delta T = \frac{\delta t}{\tau}[T_0 - T(t)] \tag{2.29}$$

The scaling factor for the velocities is given as:

$$\lambda^2 = 1 + \frac{\delta t}{\tau}[T_0 - T(t)] \tag{2.30}$$

The Berendsen thermostat is usually used to relax the system to a target temperature. However, it does not produce a correct canonical ensemble.[133] To produce a correct canonical ensemble, methods like Nose-Hooover, Anderson type thermostats are applied.[134, 135] The isothermal-isobaric ensemble is a thermodynamic system that has a constant number of particles (N), constant pressure (P), and constant temperature (T) (known as an NPT ensemble). The Berendsen barostat produces an NPT ensemble by weakly coupling the system with an external bath, utilizing the principle of least local perturbations.[132] Similar to the Berensden thermostat, an additional term is added to the integration algorithm that causes a pressure change:

$$\left(\frac{dp}{dt}\right)_{bath} = \frac{(p_0 - p)}{\tau_p} \tag{2.31}$$

The term τ_p is the coupling constant, p is pressure at time t, and p_0 is the target pressure. Scaling of the particle coordinates and the volume is applied simultaneously to minimize local disturbances. An additional term is added to the integration algorithm:

$$\dot{\mathbf{x}} = \mathbf{v} + \mathbf{ax} \tag{2.32}$$

where **x** and $\dot{\mathbf{x}}$ are coordinates before and after the pressure coupling, and ***v*** is the velocity of the particle.

The volume changes concomitantly:

$$\dot{V} = 3\alpha V \tag{2.33}$$

where:

$$\alpha = -\beta(p_0 - p)3\tau_p \tag{2.34}$$

and where β is the compressibility.

The modified equation of motion is:

$$\dot{\mathbf{x}} = \mathbf{v} - \frac{\beta(p_0 - p)}{3\tau_p}\mathbf{x} \tag{2.35}$$

which represents a concomitant scaling of coordinates.

Other barostats that are commonly used include Anderson, Parinello-Rahaman, and Nose-Hoover barostats.[134, 136]

2.3.2. Periodic Boundary Conditions

Although MD simulations are conducted to provide information about physical systems (several moles of different molecules), only a small portion of the sample (thousands of molecules) is simulated in practice, due to computational limitations. As a consequence, the ratio of surface molecules to that of the total number of molecules in the model will be higher than the actual system that is modelled.[127] Periodic boundary conditions (PBC) are used in MD simulations to simulate the bulk phase (e.g. water and crystal) and to minimize surface effects.[127] In the PBC

method, the MD box where solvent and solute molecules are located (user defined) is replicated in all directions in space (Figure 12), (i.e. the central box is enclosed by 26 duplicate boxes, which are again enclosed by 98 boxes, and so forth). If a solute or solvent molecule leaves the central box through the left wall, its image will enter the box through the right wall from the neighboring box.[127]

The utilization of PBC has a major disadvantage due to the exponential increase in the computational cost. Without PBC, to calculate the forces and energy on an atom (*i*) in the simulation box, the sum of its interactions with all of the atoms in the simulation box is only required (several hundreds). In contrast, when PBC is introduced, the sum of all of the interactions of the atom (*i*) with the rest of the atoms in the simulation box, and all of the atoms in the images is required (infinite);[107] this is impossible in practice. To overcome this problem, an approximation known as minimum image convention (MIC) is utilized.[134] MIC is applied to potential functions that are short ranged (van der Waals interactions) in nature. In MIC, an atom (*i*) in the simulation box will only interact with atoms in the periodic images that lie within the simulation box. To facilitate this, a cut-off distance (r_{cut}) is introduced to truncate the short range potentials. For any atom (*i*), the interactions with all other atoms (*j*) are within a cut-off distance and all other interactions are ignored. r_{cut} is defined such that it is less than half of the box dimension. If r_{cut} violates this rule, any atom might interact with more than one periodic image of another atom, which is incorrect.[137] Introduction of the cut-off distance aids in reducing the number of interaction terms, thus lowering the computational cost. Truncation of a potential might lead to errors in the calculation of the potential energy; this error can be reduced by choosing an r_{cut} that is as large as possible.[138]

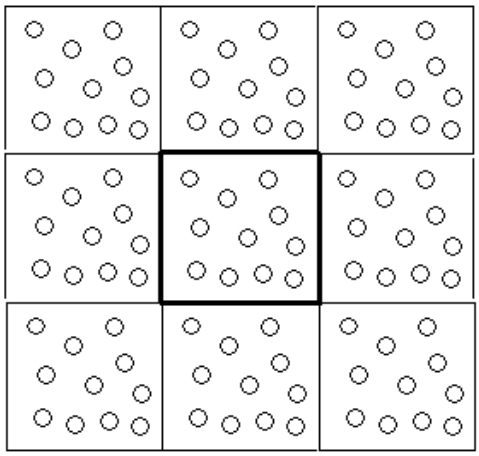


Figure 12. Illustration of PBC in 2D

This treatment is valid for potential energy functions that decay more rapidly than r^{-3} in three dimensions, such as the van der Waals function, which decays at a rate of r^{-6}.[138] For columbic interactions such truncation methods cannot be used (they decay in a rate of r^{-1}).[139] Hence, in PBC, Ewald summation is used to account for the electrostatic interactions.[140]

2.3.3 Ewald Summation

Ewald summation was introduced in 1921 as a method for computing long-range electrostatic interactions between particles and all of their replicate images.[140] The basic idea behind this method is to partition the interactions into a short-range and long-range contribution. The short-range contribution is computed by directly obtaining particle-particle interactions. In contrast, the long-range contribution assumes a Gaussian distribution of charge, which can be efficiently computed in the Fourier space.[140] In the Ewald method, point charges are screened by a Gaussian charge distribution of the opposite sign, making short-range electrostatic interactions. This added charge is then neutralized by adding a Gaussian charge distribution with the same sign as the point charge (Figure 13).

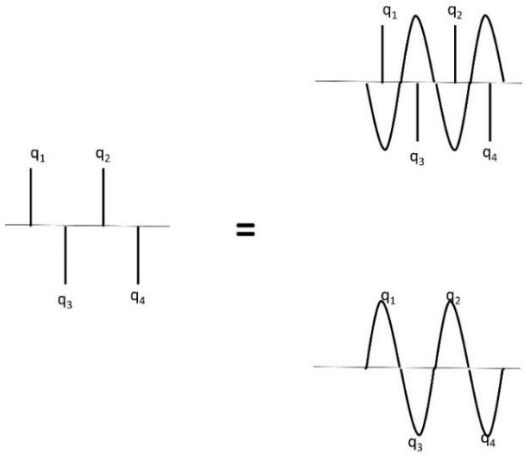


Figure 13. Schematic representation of the Ewald summation.

The short-range contribution is calculated using the following equation:

$$E_{real} = \frac{1}{4\pi\varepsilon_0} \sum_{i>j} \frac{q_i q_j erfc(\alpha_{ij})}{r_{ij}} \tag{2.36}$$

where *erfc(x)* is the error function, r_{ij} is the distance between the particles, q_i and q_j are the charges on atoms *i* and *j* respectively, α is a constant damping factor, ε_0 is the permittivity of the vacuum, and ε_r is the relative dielectric constant. The real space part of Ewald summation and van der Waals interactions can be calculated simultaneously using the same cut-off radius, although most of the popular MD codes allow users to define cut-off values separately from van der Waals and Ewald sum methods.[141]

The self-energy of the added Gaussian charges is equal to:

$$E_{real} = -\frac{1}{4\pi\varepsilon_0\varepsilon_r} \frac{\alpha}{\sqrt{\pi}} \sum_{i=1}^{N} q_i^2 \tag{2.37}$$

The Fourier part of the energy has the following form:

$$E_{Fourier} = -\frac{1}{4\pi\varepsilon_0\varepsilon_r}\frac{1}{2V}\sum_{k\neq 0}^{\infty}\exp\left(\frac{-k^2}{\frac{4\alpha^2}{k^2}}\right)\left|\sum_{j=1}^{N} q_j \exp(-ik.r_j)\right|^2 \tag{2.38}$$

where V is the volume of the system and the vectors k are linear combinations of the reciprocal basis vectors of the system.

2.3.4 Error Analysis in Molecular Dynamics Simulations

Since MD simulations are of finite length, calculated thermodynamic properties (averaged) always involve statistical uncertainty. This uncertainty can be reported in terms of a standard deviation of the average value, and will be inversely proportional to the square root of the number of sampling points.[107] It is important that these points are not correlated; thus, nearby points in simulations are not used. Therefore, to estimate statistical uncertainty, the production time of an MD simulation is divided into blocks, such that the equivalent points in two blocks are not correlated.[107] To recognize an appropriate block size, the variance of the averages that are calculated over the blocks must be calculated. When the variance of increasingly larger sizes of blocks reaches a plateau, averages become independent of the block size and the minimum block size of the independent sampling points are determined.[107]

Alternatively, to determine the block size, direct evaluation of the autocorrelation function for a desired variable can be implemented. The autocorrelation function will be zero for the interval of points, which corresponds to the relaxation interval of such a variable.[142] The advantage of this method is that if the relaxation time of a certain process is known, the required simulation time can also be estimated in advance.

Initial geometries of molecular models that are used in MD simulations are usually obtained through x-ray structures or by minimized geometries using quantum mechanical (QM) methods.

QM calculations are also used to obtain the partial charges and other parameters (equation 2.14) that will be used in a force field. In section 2.4, a brief introduction to QM calculations, the derivation of atomic partial charges, and the density functional theory is given.

2.4. Background Information about Quantum Mechanical Simulations

In quantum mechanical (QM) simulations the energy of a system is described as a function of nuclear and electronic coordinates.[143] The main difference between MM and QM simulations is that, in QM models electrons are explicitly accounted for. This allows QM simulations to be used to model properties that are dependent on electron distribution and dynamics (such as charge transfer interactions and chemical reactions).[143] Matter possesses wave and particle properties, due to its small mass and high velocity of electrons; thus, the wave nature must be taken into account to correctly describe its properties. As such, wave functions (φ) are introduced to describe its static and dynamic behavior. Wavefunctions are solutions to the Schrödinger equation:

$$H\varphi = E\varphi \qquad \text{(time independent)} \tag{2.39}$$

where $\hat{H}$ is the Hamiltonian operator, φ is the wave function, and E is the energy. The square modulus of the wave function multiplied by the infinitesimal volume factor gives the probability of finding an electron in that volume.[143] This is an observable property that is directly related to the electron density, as observed using x-ray diffraction methods. The Schrödinger equation cannot be solved precisely for systems involving more than one electron. Thus, approximations are made to simplify the problem at hand. One of the most important approximations made is the Born-Oppenheimer approximation, which states that nuclear kinetic energy is negligible compared to electron kinetic energy that allows us to solve electron and nuclei wave functions separately.[144]

The Hartree-Fock density functional theory and the Moller plesset perturbation theory are two popular wave-based approaches that are used today to solve the Schrödinger equation.[107]

2.4.1 Density Functional Theory

Over the last decade, density functional theory (DFT) has been the method of choice for studying condensed matter systems such as proteins, organic molecules, semi-conductors, clay, and minerals.[145-147] The popularity of DFT is due to its ability to accurately predict and describe material properties for a relatively low computational cost (compared to the Hartree-Fock and Moller plesset theories). DFT describes a many-body interacting system through its particle density, not through a many-body wave function (as in Hartree-Fock).[148] This causes a significant reduction in the degrees of freedom to be simulated, from 3N to three spatial coordinates through particle density.

The foundation of DFT is Hohenberg and Khon theorems, which states that the ground state electron energy (E) is a unique functional of the electron density (ρ).[148] Given the wave function (φ), electron density can be defined for an N-body system as:

$$\rho(x) = N\int \left|\varphi(x, x_2, ..., x_N\right|^2 dx_2 ... dx_N \tag{2.40}$$

where $\rho(x)$ is the probability of finding an electron at position x.

If the exact functional form is known, the ground state energy can be readily obtained by a minimization procedure of the electron density. However, HK theorem only states the existence of such a function, without providing the full details of its actual form.[148]

A study by Kohn and Sham was the first to approximate the functional form of interacting electrons by an energy functional of non-interacting electrons with a correction term that is known as the exchange-correlation functional $E_{XC}[\rho]$.[149] The Kohn-Sham DFT is formally written as:

$$E_{KS}[\varphi_i] = \int E_K[\varphi_i] + \int V_{ext}(x)\rho(x)dx + \frac{1}{2}\iint \rho(x)\rho(y)dxdy + E_{XC}[\rho] \tag{2.41}$$

where the first term computes the electron kinetic energy of non-interacting electrons, the second term computes the electron-nuclei interaction, and the third term computes the electron-electron interaction. $E_{XC}[\rho]$ is an universal functional of the electron density that is known as the exchange-correlational functional. To obtain the ground state properties of a system, equation 2.41 should be optimized over all possible electron orbitals φ_i under the following constraint:

$$N = \int \rho(X) dX \tag{2.42}$$

where N is the total number of electrons in the system (e.g. in an H_2 molecule, N is equal to 2)

One of the most important steps in conducting a DFT calculation is the proper selection of the exchange-correlation functional. There are numerous approximations of the exchange correlational functional, such as local density approximation (LDA), generalized gradient approximation (GGA), meta-GGA, and hybrids of LDA and GGA. Unfortunately, none of these functionals appear to work well for all chemical systems. A functional should be carefully selected, such that it will provide optimal results for the chemical system of interest. Among the plethora of functionals, B3LYP has been benchmarked and extensively used to obtain accurate ground state geometries of simple organic molecules.[150]

2.4.2 Quantum Mechanical Derivation of Partial Atomic Charges

The use of high-quality partial charges for atoms when conducting an MD simulation is an important factor that determines the accuracy of the results obtained from it.[151] As atomic charges are not observables and there is no specific criterion to validate the accuracy of partial charges that are derived through various methods (e.g. Mulliken and Lowdin population analysis), accuracy is often determined by their ability to reproduce experimental multipole moments.[152]

The most widely used method for calculating atomic charges is to calculate the electrostatic potential (ESP) that is generated using QM methods on a grid surrounding the molecule and

reproducing this potential by fitting charges to chosen sites on a molecule.[153] Several charge derivation methods, such as Merz-Kollman (MK), restrained electrostatic potential (RESP), and charges from electrostatic potentials using a grid (CHELPG), are based on this core concept.[151, 154]

The RESP method, developed by Kollman *et al.*, is considered the most accurate method among the aforementioned ESP fitting schemes.[154] RESP charges are less dependent on conformation and are transferable between functional groups. RESP also eliminates the poor fitting of charges on buried atoms in a molecule.[154] In the RESP method, the first step is the charge fitting process, with QM ESP V_i evaluated for each point (i) in a set of points that are fixed in solvent-accessible regions around the molecule. These points are selected such that they lie outside of the van der Waals radius of the molecule. Subsequently, the least squares fitting procedure is then used to fit the charge (q_i) to each atomic center (j) of the molecule, with a restraining function, χ^2_{rsrt} .

The calculated ESP $\widehat{V_i}$ is given by the equation:

$$V_i = \sum_i \frac{q_j}{r_{ij}} \tag{2.43}$$

And the least squares function χ^2 can be written as:

$$\chi^2 = \chi^2_{esp} + \chi^2_{rsrt} \tag{2.44}$$

where $\chi^2_{esp} = \sum_i (v_i - v_i)^2$ and $\chi^2_{rsrt} = a\sum_j (q_0 - q_j)^2$, a is a scale factor, q_0 is the target charge of the restraint, and a is the restraint.

The minimum of the equation 2.44 can be defined as:

$$\frac{\partial(\chi^2)}{\partial q_j} = \frac{\partial(\chi^2_{esp})}{\partial q_j} + \frac{\partial(\chi^2_{rsrt})}{\partial q_j} = 0 \tag{2.45}$$

By solving the equation 2.45 using a matrix, partial charges can be obtained.

The next section of this chapter focuses on the application of the molecular modeling methods in the specific case of the calculation of free energy of ligand binding in biological systems.

2.5. Introduction to Protein Systems and the Importance of Protein-Ligand Interactions

Proteins (Figure 14) are large, three-dimensional biomolecules that are composed of one or more long chains of amino acid residues. Proteins do most of the work within living organisms, such as DNA replication, catalyzing enzymatic reactions, responding to stimuli, and transporting molecules.[155] Proteins primarily differ from each other by the sequence of amino acids in their polypeptide chains. There are twenty different amino acids in nature, and each has different chemical properties.[155] As a result, different combinations of amino acids will elicit proteins with different three-dimensional structures and chemical properties. Thus, the structure-function relationship of a protein must be known to understand the processes and mechanisms associated with those processes.[156]

Ligands are species that form complexes with proteins, mainly through noncovalent molecular interactions. A ligand can be another protein, organic molecule, or membrane.[157] One of the first models used to describe the behavior of ligands interacting with proteins was the lock and key model, which states that there is a complimentary relationship between the shape of a ligand and that of the active site of a protein; hence, the analogy of the lock and key.[158] In proteins, a multitude of functions are mediated by the interaction of the protein with certain ligands. The result of such an interaction is the formation of a protein-ligand complex. The affinity of a specific ligand to a protein is measured by the overall change in the free energy of the interactions that occur upon binding. A high negative value of free energy indicates a high affinity. One of the greatest

achievements of the twentieth century was the development of methods that allow scientists to discover small organic molecules (ligands) that are used to treat people who suffer from various diseases.[159] Measuring the affinity of a ligand to a protein is an important step in the early stages of drug discovery and in elucidating gene functions.[159] Discovering a small molecule that binds tightly to a specific protein can be a time consuming and costly challenge if done experimentally.[160]

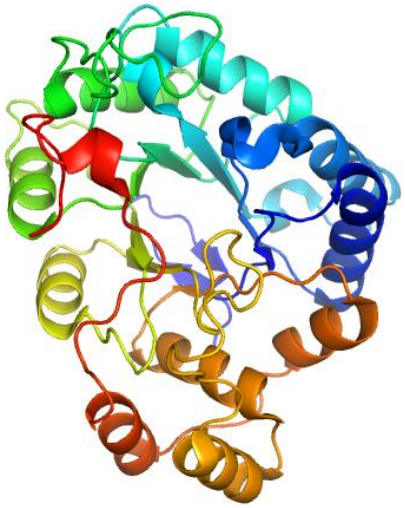


Figure 14. Illustration of a three dimensional structure of the protein Aldose Reductase

To overcome the aforementioned adversities, computational chemists have developed a multitude of approaches that are robust and accurate. These approaches are based on molecular mechanics methods, such as thermodynamic integration, free energy perturbation, potential of mean force, empirical scoring functions, and the linear interaction energy method.[160]

2.6. Rigorous Free Energy Calculation Methods

2.6.1 Free Energy Perturbation Method

The free energy perturbation method (FEP) was initially developed by Zwanzig *et al.* to calculate the thermodynamic properties of systems with argon and nitrogen using MD

simulations.[161] It was later used to calculate the binding free energy of ligands to protein active sites. The derivation of the FEP method is outlined below:

To calculate the energy difference between two systems *X* and *Y*, which are described by two different energy functions E_x and E_y, the difference in free energy is:

$$A(X) - A(Y) = -kT \ln \frac{Q_X}{Q_Y} \tag{2.46}$$

where *A(X)* and *A(Y)* are free energy values and Q_x and Q_y are the partition functions of states *X* and *Y*.

Equation 2.46 can be transcribed as an ensemble average:

$$A(X) - A(Y) = kT \ln \left\langle e^{(E_X - E_Y)/kT} \right\rangle_X \tag{2.47}$$

where the subscript *X* specifies that the average is over an ensemble of configurations that are generated using MD methods and are representative of the preliminary state *X*. We can write down a similar expression, where the configurations are averaged over the ensemble that corresponds to the final state (*Y*). As long as the energy difference $(E_X\text{-}E_Y)$ is comparable to *kT*, equation 2.47 can yield a good approximation of the free energy. If *X* and *Y* do not overlap in phase space, the phase space of *Y* will not be adequately sampled when simulating *X*. In this case, the energy difference between *X* and *Y* will be larger than *kT* and the free energy changes will be not accurate. In such cases, the intermediate states between *X* and *Y* are introduced and described in terms of a coupling parameter λ. The simplest approach involves a linear interpolation:

$$E_\lambda = \lambda E_X + (1-\lambda) E_Y \tag{2.48}$$

and equation 2.47 becomes:

$$A(X) - A(Y) = kT \sum_\lambda \ln \left\langle e^{(\Delta E_\lambda / kT)} \right\rangle_\lambda \tag{2.49}$$

The intermediate steps involved in the transformation between *X* and *Y* by variable λ may not correspond to a physical transformation. However, this does not matter since free energy is a state function and does not depend on the path of the transformation, as long as it is reversible.[107] It is vital that there is adequate overlap between successive λ states. In general, a transformation that involves high energy barriers will require smaller increases of λ to guarantee the reversibility of the process.

For larger systems such as proteins, because of the poor overlap between the thermodynamic phase space densities, the sampling problem becomes increasingly severe and results in complications with the assessment of the reliability of the calculated free energy values. Therefore, excessive sampling is often necessary, which requires extensive computer power and time.[162]

2.6.2 Thermodynamics Integration Method

In the thermodynamics integration (TI) method, free energy is written as a function of a coupling parameter λ, with values between 0 and 1:

$$A(\lambda) = -kT \ln Q(\lambda) \tag{2.50}$$

Differentiation of this equation with respect to λ yields:

$$\frac{\delta A}{\delta \lambda} = -\frac{kT}{Q}\frac{\delta Q}{\delta \lambda} = \frac{\delta E}{\delta \lambda} \tag{2.51}$$

Replacing the right side of the equation with an ensemble average and integrating over λ, the following equation is obtained:

$$A(1) - A(0) = \int_0^1 \left\langle \frac{\delta E(\lambda)}{\delta \lambda} \right\rangle_\lambda d\lambda \tag{2.52}$$

Desired free energy change is obtained from the left side of the equation, and the right side may be approximated by a discrete sum:

$$\Delta A = \sum_i \left\langle \frac{\delta E(\lambda)}{\delta \lambda} \right\rangle_N \Delta \lambda_i \tag{2.53}$$

The main difference between equations 2.49 and 2.53 is that in free energy perturbation (FEP), the ensemble average is determined over finite differences in energy functions, while in TI the average is determined over a differentiated energy function. For this reason, unless the transformation involves little change, TI is the method of choice for bimolecular systems.

2.6.3 Thermodynamic Cycles

Since free energy is a state function and it is independent of the path used to reach a specific state (as long as the intermediate states are at equilibrium), thermodynamic cycles are used to dramatically reduce time in molecular simulations.[91, 163] In a typical case where one wants to determine the relative binding affinity between two compounds A and B, a thermodynamic cycle can be devised as shown in Figure 15:

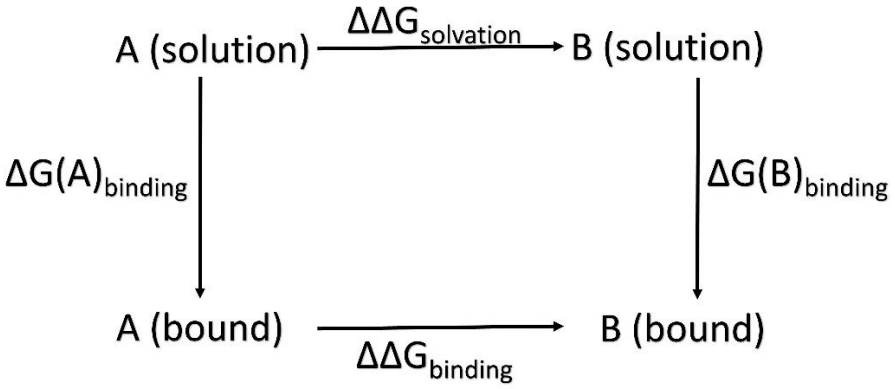


Figure 15. Thermodynamics cycle for the calculation of the relative binding energy of two compounds, A and B

In the thermodynamic cycle shown in Figure 15 above, $\Delta\Delta G_{solvation}$ and $\Delta\Delta G_{binding}$ denote the differences in free energy of solvation and binding respectively, between A and B. $\Delta G_{binding}$

represents the binding free energy. An equation for the relationship shown above in the thermodynamic cycle is derived as:

$$\Delta\Delta G(A)_{binding} - \Delta G(B)_{binding} = \Delta\Delta G_{binding} - \Delta\Delta G_{solvation} \quad (2.54)$$

If FEP/TI is used to evaluate the relative binding affinity of A to that of B, MD or Monte Carlo (MC) simulations must be conducted to obtain the free energy of a mutation from A to B in a solution and in protein (through sampling of the configuration space).[164, 165] This is difficult because the process involves changes in the molecular charge distribution and the creation or obliteration of atoms.[165] To gain converged free energy values, extensive sampling is required, which is often time consuming and computationally expensive.

2.6.4. Potential of Mean Force

The potential of mean force[166] (PMF) is used to evaluate changes in the free energy of a system (using MD methods), as a function of a specific reaction coordinate (distance between two residues of a protein, distance between two proteins, etc).[167] Since calculations involving PMF are physically achievable, the point with the highest energy on the potential energy surface corresponds to the transition state. Therefore, PMF is also used to derive kinetic parameters such as rate constants. In PMF simulations, umbrella sampling is often used to enforce adequate sampling of the configurational space as it proceeds.

2.7. Approximate Free Energy Methods

2.7.1. Empirical Scoring Functions

Computational approaches such as FEP and TI that are used to calculate ligand binding energies require excessive sampling that demands computational power and time. Modern drug developers screen libraries with thousands of molecules; hence, the use of FEP and TI methods are not possible. Thus, the need for rapid search algorithms to predict the affinities of small/large

organic molecules that may bind to target receptor molecules of interest is a vital step in drug discovery processes. If the three-dimensional structure of an active site is available, a process called docking is used to generate a variety of conformations, orientations, and positions in a user defined active site of the receptor.[168] To determine the conformation of the most energetically favorable pose, it is necessary to evaluate and assign a score to each pose based on the complementary to the target in terms of shape and favorable molecular interactions.[169] In summary, in order to rank the binding affinities of a homologous set of organic molecules to a ligand, the most favorable pose for each molecule must first be ranked. Then the energy of this pose for different molecules is compared by ranking each molecule based on its affinity to the receptor site. To score each pose, empirical scoring functions (ESF) are used. Many scoring functions fall into two main classes: knowledge-based scoring functions and schemes based on physical interaction terms.

Knowledge-based scoring functions assign a score to a particular ligand pose based on a large database of crystal structures of protein-ligand complexes.[170] The method acquires frequencies of interatomic contact and distances of functional groups in a given protein-ligand complex, and compares them with the data base. If these molecular interactions are found to be lower in frequency than the data base, it assigns a lower score and vice versa.

Scoring functions based on physical interaction terms that are based on the assumption that the change in free energy upon binding can be expressed as a sum of individual interaction terms[171]:

$$\Delta G_{BIND} = \Delta G_{INT} + \Delta G_{SOLV} + \Delta G_{CONF} + \Delta G_{MOTION} \tag{2.55}$$

where ΔG_{INT} represents specific ligand-receptor interactions, ΔG_{SOLV} is the interaction of the ligand with the solvent, ΔG_{CONF} is the conformational changes in the protein and the ligand, and ΔG_{MOTION} is the motion of the protein and the ligand during the complex formation. These values

are not calculated by considering an ensemble average, but by considering a single structure and additivity. In biochemical processes this is not strictly valid. Despite these limitations, scoring functions are widely used to calculate the relative binding affinities of putative ligands. [171]

2.7.2. Linear Interaction Energy Method

The linear interaction energy (LIE) method that was developed by Åqvist *et al.* hypothesizes that the free energy of binding can be expressed as a linear combination of weighted energy estimates of the interactions between a ligand and the rest of the system.[172] In this method, the binding energy can be written as the sum of the ensemble averaged difference of Lennard-Jones (V^{LJ}) and Coulomb (V^{el}) interactions between the ligand and its surroundings in the bound (ligand-protein complex) and free (ligand in solvent) state as:

$$\Delta G = \alpha \left\langle V^{LJ} \right\rangle + \beta \left\langle V^{el} \right\rangle \tag{2.56}$$

Parameters α and β are obtained empirically using a fitting procedure. As the LIE method only involves sampling in the bound and free state of the ligand, it is computationally inexpensive compared to methods such as TI and FEP.

2.8. Objectives and Significance of this Study

As discussed above, there is a void in the approximation of K_d values of cationic compounds that are sorbed through cation exchange to environmentally relevant solids. As a solution, K_d models that account for the structural features of organic solutes combined with key features of the environmental solids have been hypothesized. Due to the hardship of decoupling the different modes of interactions that are involved of the compounds with the environmental solid, mainly through hydrophobic and electrostatic interactions, it was difficult to determine an exact form of

an equation for K_d that includes a precise contribution from each of these interactions using experimental methods.

The primary objective of this thesis is to translate molecular modeling tools used in biochemistry to predict the free energy of the sorption of charged organic cations to environmentally relevant solids. It was proposed that the LIE method would be best suited for this purpose among the methods discussed in sections 2.6 and 2.7, given its inexpensive computational cost (compared to FEP/TI) and its reduced constraints of shape complementarity for mineral sorption (compared to docking and scoring methods). For this purpose, experimental K_d values of a homologous series of ammines sorbing to pure phase clay Ca-montmorillonite (a smectite clay having Ca^{+2} interlayer ions) was used. To the best of the author's knowledge, such an investigation has not been conducted before.

Another objective of this thesis is to analyze LIE parameters as a function of different interlayer ions in montmorillonite; this was evaluated using Ca and Na ions. Experimentally, it has been shown, that there is a very small difference in K_d values between these two ions, however such difference is consistent for a large number of compounds. Thus, it is important to determine whether or not LIE will capture these deviations. Finally, the LIE method will be applied to model organic cations' sorption to a surface of montmorillonite clay; this is done to model the system without considering the d-space between two montmorillonite sheets, thus simplifying the problem at hand.

3. Linear Interaction Energy Method

3.1. Background

The LIE method is a semi-empirical method developed by Åqvist *et al.* for calculating the free energies of binding from MD simulations to assist in drug design processes.[172, 173] At the time this method was developed, scientists employing FEP and TI methods could only propose timid binding candidates that had size-shape complementary and favorable electrostatic properties.[174] Thus, absolute binding constants were obtained using costly and time consuming experimental methods. Furthermore, computational approaches have only been used to rank smaller drug candidates (due to convergence issues and the need for extensive sampling).[175] Since larger molecules were probed as drug candidates, the need for more efficient methods of determining binding free energy were required.[176]

3.2. Derivation of the LIE Equation

The LIE method is based on the linear approximation of electrostatic forces, which states that polar solutions will yield quadratic free energy functions in response to changes in electric fields. This assumption is also the basis for the Marcus theory of electron transfer.[177] For a system with two states, A and B, that is given by two potential energy functions, V_A and V_B, quadratic free energy functions of equal curvature can be illustrated as demonstrated in Figure 16.

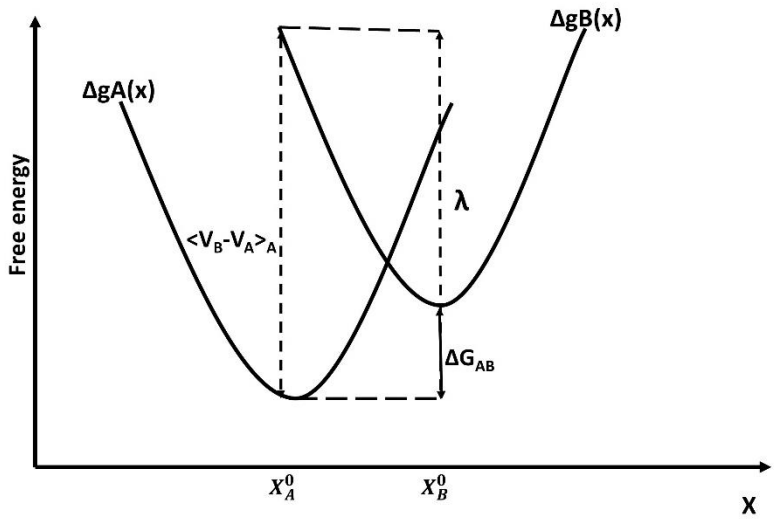


Figure 16. Illustration of free energy curves for a system that obeys the linear response approximation.

ΔG_{AB} is the associated change in free energy, ΔgA(x) and ΔgB(x) are free energy functions, and λ is the reorganization energy.

An assumption can be made that the two states, A and B, only involve purely electrostatic variables and the difference between V_A and V_B are strictly electrostatic. Using the FEP equation, free energy change between two states, A and B, can be written as:

$$\Delta G = -\beta^{-1} \ln \left\langle e^{-\beta(V_B - V_A)} \right\rangle_A \tag{3.1}$$

Expansion of the exponent and logarithm yields:

$$\Delta G = -\beta^{-1} \ln < 1 - \beta(V_B - V_A) + (\beta^2 / 2)(V_B - V_A)^2 - ... >_A \tag{3.2}$$

$$\Delta G = -\beta^{-1} \ln[1 - \beta < (V_B - V_A) >_A + (\beta^2 / 2) < (V_B - V_A)^2 >_A - ...] \tag{3.3}$$

$$\begin{aligned} \Delta G = -\beta^{-1}[&-\beta < (V_B - V_A) >_A + (\beta^2 / 2) < (V_B - V_A)^2 >_A \\ &-\frac{1}{2}(-\beta < (V_B - V_A) >_A + (\beta^2 / 2) < (V_B - V_A)^2 >_A)^2 + ...] \end{aligned} \tag{3.4}$$

The above equation can be arranged as:

$$\Delta G = < \Delta V >_A - \left(\frac{\beta}{2}\right) < (\Delta V - < \Delta V >_A)^2 >_A + \left(\frac{\beta^2}{6}\right) < (\Delta V - < \Delta V >_A)^3 >_A$$
$$- \left(\frac{\beta^3}{24}\right) < (\Delta V - < \Delta V >_A)^4 >_A - 3 < (\Delta V - < \Delta V >_A)^2 >_A^2 + \ldots \quad (3.5)$$

where $\Delta V = V_B - V_A$

In the same manner, Gibbs free energy can be obtained by averaging the potential energy surface V_B :

$$\Delta G = < \Delta V >_B + \left(\frac{\beta}{2}\right) < (\Delta V - < \Delta V >_B)^2 >_B - \left(\frac{\beta^2}{6}\right) < (\Delta V - < \Delta V >_B)^3 >_B + \ldots \quad (3.6)$$

By adding equation 3.5 and 3.6 the following equation is obtained:

$$\Delta G = \frac{1}{2}\{< \Delta V >_A + < \Delta V >_B\} - \frac{\beta}{4}\{< (\Delta V - < \Delta V >_A)^2 >_A - < (\Delta V - < \Delta V >_B)^2 >_B\} + \ldots \quad (3.7)$$

Since an assumption that the two free energy functions, ΔgA(x) and ΔgB(x), have equal force constants (equal curvature) was made, mean square fluctuations in the energy gaps on the two surfaces will be equal. Thus, the terms that are second order and higher in equations 3.6 and 3.7 will be cancelled upon addition and free energy change can be written as:

$$\Delta G_{AB} = \frac{1}{2}(< \Delta V >_A + < \Delta V >_B) \quad (3.8)$$

where $\Delta V = V_B - V_A$

If the hydration of a single ion is considered, equation 3.8 can be written as:

$$\Delta G_{sol}^{el} = \frac{1}{2} < V_{i-s}^{el} > \quad (3.9)$$

where, electrostatic contribution to the solvation energy is equal to half of the associated ion-solvent interaction energy (V_{i-s}^{el}). For the protein-inhibitor binding problem, where two states of

the ligand are in solution (A) and bound to protein (B), the binding free energy can be approximated to the following equation:

$$\Delta G_{bind}^{el} = \frac{1}{2} < \Delta V_{i-s}^{el} > \quad (3.10)$$

where Δ refers to the difference in interaction energy of ligand in solution and protein.

To test the validity of equation 3.10, Åqvist *et al.* conducted FEP/MD simulations to calculate the free energy of charging (ΔG_{sol}^{el}) Na^+ and Ca^{+2} ions in solution. He obtained values of 0.49 and 0.52 respectively for terms connecting ΔG_{sol}^{el} to ensemble averaged ion-solvent electrostatic terms ($< V_{i-s}^{el} >$) .[172] In a separate study by Jorgensen *et al.*, the charging of a methanol molecule in water with the OPLS FF yielded a ratio of $\Delta G_{sol}^{el} : < V_{i-s}^{el} >$ of 0.43. Thus, both of the results showed similar results to the LIE predicted value of ½. Since equation 3.10 was derived with the electrostatic interactions alone, incorporation of the contribution of non-polar interactions and hydrophobic effects on the free energy of binding (ΔG_{bind}^{vdW}) was necessary. An analytical treatment was not feasible, due to the heterogeneity of the protein active sites.[172] It was observed that, experimental free energy of solvation for hydrocarbon compounds such as n-alkanes approximately depend on the chain length, as[178]:

$$\Delta G_{Sol} = kn + l \quad (3.11)$$

where *n* is the number of carbon atoms in the chain and k and *l* are constants.

This equation yielded accurate hydration energies for hydrocarbons in non-polar solutions and in polar solutions such as water. Åqvist *et al.* conducted a set of MD simulations for n-alkanes that were solvated in water and in a non-polar solvent. He observed that the mean solute-solvent interaction was roughly linear with the number of carbon chains.[172] Thus, he deduced that ΔG_{bind}^{vdW} can be approximated using a similar approach to equation 3.11:

$$\Delta G_{sol}^{vdW} = \alpha < \Delta V_{i-s}^{vdW} > + \gamma \tag{3.12}$$

where α and γ can be fitted to reproduce experimental ΔG_{sol}^{vdW}

Hence, the LIE equation can be summarized by including ΔG_{sol}^{el} and and ΔG_{sol}^{vdW} as:

$$\Delta G_{bind} = \frac{1}{2} < V_{i-s}^{el} > + \alpha < V_{i-s}^{vdW} > + \gamma \tag{3.13}$$

where α and γ are determined by empirical fitting.

In the original paper, the accuracy of the LIE method was tested using a system of endothiapepsin binding to five different inhibitors, where experimental binding constants were known.[172] The LIE parameters were calibrated using the following two approaches:

Approach 1: Fixing β to ½ and optimizing α in such a way that root mean square (rms) deviation between experimental binding energies and LIE predicted values were minimized (γ was not considered).

Approach 2: Optimizing α and γ using experimental binding energies while keeping β fixed at ½, to minimize the rms values.

Since there was no significant improvement of the fitting after the addition of an extra parameter γ, it was discarded. Thus, approach 1 was used and the final value of α that was obtained was 0.161. The predicted free energy values were closely related to the experimental values, as the mean unsigned error was 0.39 Kcal mol^{-1}.[172] Thus, the final LIE equation was:

$$\Delta G_{bind} = \frac{1}{2} < \Delta V_{i-s}^{el} > + 0.161 < \Delta V_{i-s}^{vdw} > \tag{3.14}$$

The LIE parameters 0.161 and 1/2 were successfully used to predict the free energy of the binding of sugar to a periplasmic glucose receptor and of several inhibitors to HIV-1 protease.[179-181]

3.3. A Review of Modifications Made to the Original LIE Method

Although the value of β was calculated as ½ in accordance with the linear response theory, Hansson *et al.* found that when charging neutral dipolar solutes (e.g alcohols and ammines) in water, the solvent did not respond linearly to the electrostatic field exerted by the solute (β was not equal to ½) .[182] The calculated value of β varied between 0.33 and 0.43.[182] They reasoned that the nonlinear response was because of water-solute hydrogen bonding, since same set of solutes in aprotic solvents displayed β values closer to ½.[182] It was also observed that in solutes that formed extensive amounts of hydrogen bonding (compounds with amine and hydroxyl groups), the ratio of $\Delta G_{sol}^{el} : < \Delta V_{i-s}^{el} >$ significantly deviated from ½.[182] Thus, Hansson *et al.* proposed a novel treatment for the electrostatic contribution in the LIE method, where the difference in ligand-surrounding electrostatic interaction energies was scaled by a factor of β rather than a factor of ½:[182]

$$\Delta G_{sol}^{el} = \beta < \Delta V_{i-s}^{el} > \tag{3.15}$$

A study by Almlof *et al.,* showed that β was solute specific and must be optimized to increase the accuracy of the LIE predicted free energy values (Table 2).[183]

Table 2. The optimal parameters of β obtained using different classes of compounds[182]

Compound class	β
Alcohols	0.37
1^0, 2^0 ammines	0.39
Carboxylic acids	0.40
Cations	0.52

In the original LIE method (equation 2.22), α was parameterized using a set of five endothiapepsin inhibitors, yielding a value of 0.16 when β was fixed to ½. Subsequent modification of the parameter β by Hansson *et al.* yielded an optimum value of $\alpha = 0.18$. Since the parameter α depends on the characteristics of the binding site since different proteins have different inhomogeneous environments, α has different values depending on the protein-ligand pair.

Thus, a modified version of the LIE equation was introduced:

$$\Delta G_{bind} = \beta < \Delta V_{i-s}^{el} > + \alpha < \Delta V_{i-s}^{vdW} > \tag{3.16}$$

where α and β were both optimized as in quantitative structure activity relationship (QSAR) methods to reduce the rms deviation of calculated binding free energies to that of experimental values.[183]

Although the parameter γ was not included in equations 3.14 and 3.16 due to the lack of increase in the quality of the fitting in the original LIE equation, subsequent investigations that used LIE to calculate ligand binding free energy required a nonzero γ to reproduce absolute binding free energy, leading to the LIE equation[184, 185]:

$$\Delta G_{bind} = \beta < \Delta V_{i-s}^{el} > + \alpha < \Delta V_{i-s}^{vdW} > + \gamma \tag{3.17}$$

A study by Jorgensen *et al.* indicated that the LIE method (equation 3.14) was not capable of reproducing the free energy of the hydration (free energy difference between the two states: a molecule in solution phase and in vacuum) of molecules such as methane and ethane, due to the positive value of the free energy.[184] Since α and β were positive and the interaction energies calculated for these molecules were negative, the positive free energy of hydration was unobtainable. They outlined the need for a term that accounts for the price of solute cavity formation. Thus, the expansion of the LIE equation to consider a cavitation term that is proportional to the solvent-accessible surface area (SASA) was proposed:

$$\Delta G_{bind} = \beta < \Delta V_{i-s}^{el} > + \alpha < \Delta V_{i-s}^{vdw} > + \gamma < \Delta SASA > \quad (3.18)$$

where $< \Delta SASA >$ is the ensemble averaged change in a SASA in solution and in protein. Jorgensen *et al.* found improvement in the predicted free energy values compared to the experimental values (lower rms relative to experimental values) from equation 3.18 over the original LIE equation (equation 3.14).

Chapter 4. Atomistic Prediction of Sorption Free Energies of Cationic Aromatic Ammines on Montmorillonite: A Linear Interaction Energy Method

4.1. Background

Due to the increase in the use of polar organic compounds, such as pharmaceuticals, hormones, toxins, antibiotics, antidepressants, and beta-blockers, and their adverse health effects in humans and freshwater organisms, accurate models are necessary to improve the risk assessment of such trace contaminants by determining their bioavailability.[2, 15, 31, 53] These compounds exist as positively charged ions, due to protonation of the amine groups at environmentally relevant pH (>7) conditions. Thus, they can easily interact with clays and soil organic matter, due to the occurrence of negatively charged surface sites.[73] Because of the lack of experimentally derived soil sorption affinities, risk assessment can be based on predictive sorption models.[81]

The affinity of a compound for a solid phase such as clay and soil organic matter can be influenced by its structural features (e.g. charge, substituted groups, and aromaticity).[81] The transport and fate of these compounds in soil involve complex mechanisms that can be influenced by leaching, volatilization, adsorption/desorption, ion-exchange, and biological and chemical degradation.[57] Sorption is an important factor that determines the ensuing bioavailability, reactivity, and mobility in the environment of that compound.[50] Although a large number of publications divulge qualitative trends with respect to structural criteria, influencing polar and organic compound sorption through cation exchange and surface complexation[81], there have been limited attempts to develop predictive models that use these changes in specific electrostatic and covalent interactions, as The European Union Technical Guidance Document (EU-TGD) in 2006 specifies that there is no sorption model for cationic ammines.[186]

Empirical models for the sorption of non-heterocyclic cationic ammines has been proposed, considering molar volume and amine charge area[73]; these models lack the ability to correctly include structural features that are derived from small sets of organic compounds that have similar sub-structural features, such as additional aromatic rings, -COOH, and –OH. Since the sorption of cationic ammines is a coupled effect of electrostatic and hydrophobic interactions, there is a need to explicitly account for electronic and hydrophobic interactions.[81]

The study described in this thesis is based on the assumption that the cation exchange (the process of the exchange of ions of the same charge between an insoluble solid and a solution in contact) is the driving force behind the sorption of organo-cations to environmental solids as aluminosilicates. Cation exchange is mediated by a single key functional group (the cationic amine) on the sorbate structure and driven by its interaction with the negatively charged surface sites of the sorbent (e.g., aluminosilicate); this negative charge is largely insensitive to changes in the environmentally-relevant pH values.[85] Aluminosilicates such as pyrophillites, micas, and montmorillonites are abundant in soil, and characterized by a layer-like structure, and are the preferred host materials for incorporating these compounds.[60] Montmorillonite (MMT) is a layered aluminosilicate, where the central sheet is composed of octahedrally coordinated Al atoms, and the adjacent sheets are composed of tetrahedrally coordinated Si atoms. These sheets are made up of 2:1 or tetrahedral-octahedral-tetrahedral (TOT) layers. These layers are negatively charged due to the substitution of divalent metals (e.g., Mg^{+2}, Fe^{+2}) with the octahedrally coordinated Al atoms, or the substitution of Si atoms with Al atoms. The negative charges of the MMT layers are balanced by interlayer cations (e.g. Ca^{+2}), which are usually hydrated to allow the expansion or contraction of the interlayer spacing, depending on the relative humidity.[187] Since these interlayer cations are

exchangeable, clays are ideal host materials in soils for positively charged organo/inorganic ions through ion exchange.[54]

On the other hand, in computational biochemistry and biology there has been a great expansion in the development and utilization of highly sophisticated computational tools, in order to understand the role of structural effects that influence the binding of organic ligands to proteins.[174, 183] This thesis proposes that these computational tools could be translated and used in the environmental science area to understand and predict the trends, and the influence of structural criteria on organic compound sorption to clays and soil organic matter.

In computational biochemistry, a wide variety of computational tools have been used to calculate free energy (ΔG) from computationally expensive methods (needs extensive sampling), such as FEP, TI to high speed docking, and scoring functions.[107] Neither approach is suitable in its application to compute ΔG for environmental sorption. The former method is not sufficient because the computational cost prevents routine screening of compound libraries, and the later method is not sufficient because of the lack of analogous docking confirmation for sorbates with environmental solids. On the other hand, LIE, a method that has been underutilized in biochemistry, is best suited for the issue at hand.[172] The LIE method is more affordable than FEP/TI and is suited for the sorption to environmental solids that lack size-shape complementarity as protein-ligand interactions. This method involves conducting two MD simulations, one with a ligand in solution and one bound to the protein to sample electrostatic (*el*) and van der Waals (*vdW*) forces. The form of the LIE equation we utilized was:

$$\Delta G_{LIE} = \alpha < \Delta E_{vdW} > + \beta < \Delta E_{el} > + \delta \qquad (4.1)$$

where < > denotes MD averages of the electrostatic (el) and van der Waals (vdW) interactions of the solute with its surrounding environment. Δ denotes the change in these averages when

transferring the solute from the solution to the interlayer of the clay. The coefficients α and β are scaling factors, while δ is a constant.

In the study described in this thesis, the LIE method is applied to calculate the free energy of the sorption of organic cations to environmentally relevant solids. Additionally, the influences from electrostatic and van der Waals interactions to the free energy of sorption were decoupled. For this purpose, Ca-montmorillonite (Ca-MMT) was chosen as a representative pure phase sorbent to evaluate the applicability of the LIE method using a set of homologous organic cations (Figure 17).

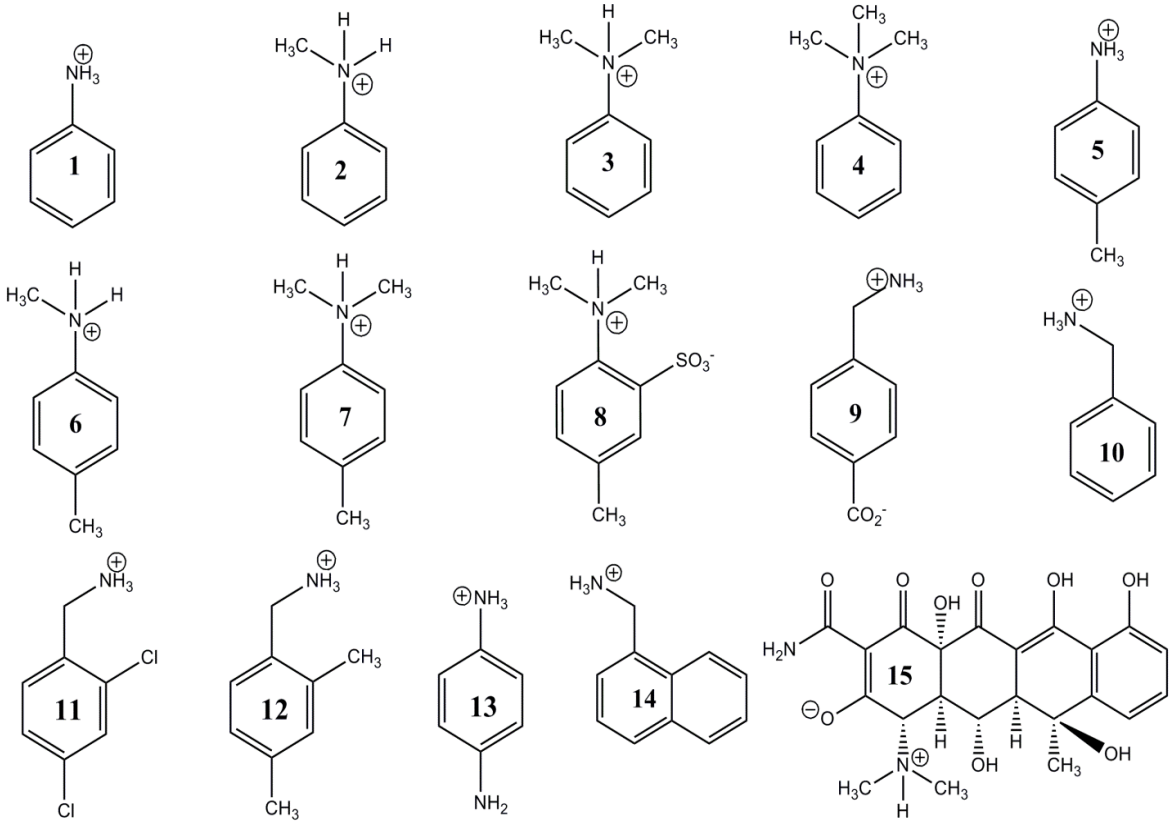


Figure 17. Chemical structures of compounds 1-15

4.2. Computational Details

The atomistic models of Ca-MMT were created using the previously reported unit cell of a Texas type MMT, with an unit cell formula of $(Ca)_{0.33}(Si_{7.75}Al_{0.25})(Al_{3.5}Mg_{0.5})O_{20}\text{-}(OH)_4.nH_2O$, with isomorphic octahedral Al^{+3}/Mg^{+2} and tetrahedral (obeying the Lowenstien's rule for isomorphic substitution) Si^{+4}/Al^{+3} substitutions.[188, 189] The unit cell was replicated four times in the x and y directions and two times in the z direction (4×4×2), creating a supercell with 32 unit cells. The ensuing super cell had two montmorillonite slabs (TOT) with an overall charge of -24.00*e* (-0.75*e* per unit cell). Density functional theory (DFT) was used to obtain the optimized geometries of organic cations that were used in this study (Figure 17). DFT optimizations were conducted using a b3lyp/6-31G* level of theory in the gas phase using the software Gaussian 09.[190, 191]

To model the solute (organic cation) in solution, a single solute ion was placed in a 60 Å × 60 Å ×60 $Å^3$ box and water molecules were added such that the density of the box reached 1.0 g/cm^3. Solute molecules in Ca-MMT were modeled using the constructed supercell. A single solute molecule was placed in the interlayer spacing and water molecules were subsequently added to form a double layer hydrate; water molecules that sterically clashed with the organic cation were removed.

To model the Ca-MMT, parameters from the CLAYFF force field was used, which has been used previously to model interactions between organic molecules with montmorillonite clays.[110] Organic cations were modeled using the OPLS-AA force, which has been widely used to model condensed phase simulations.[104] Water molecules were modeled using the SPC water model, due to its simplicity.[123] Partial charges for the organic cations were derived using the RESP method with the DFT b3lyp/6-31g(d) level of theory in the gas phase, in agreement with the philosophy of the OPLS-AA force field.[154]

All of the molecular dynamics (MD) simulations were performed using the MD code DL_POLY CLASSIC[192] version 1.8, and the Nose-Hoover isothermal-isobaric (NPT) ensemble with a coupling constant of 0.5 ps, a temperature of 298 K, and a pressure of 1 atm.[193] A time step of 1fs was used for the simulations. Electrostatic interactions were computed using the Ewald summation method, which accounts for long-range electrostatic interactions.[140] Fuchs correction was applied to the Ewald summation to account for the positive overall charge in solution and clay.[194] The model with solute in water was equilibrated for 100 ps, followed by a production time of 300 ps. The clay/solute system was equilibrated for 1000ps followed by a production time of 300ps. The length of the production was determined by the converged values in the average interaction energies or using similar production times as those used in previous studies found in literature. Statistical errors associated with different interaction energies from MD simulations ($\Delta< >$ values in equation 4.1) were computed using the block average method. More precisely, the total production time in solution and clay was divided into 50ps blocks, and averages were computed in each block. Error was reported using the standard deviation of the obtained average values, propagated as:

$$\Delta <>_{error} = \sqrt{< \partial clay >^2_{stdev} + < \partial solution >^2_{stdev}} \tag{4.2}$$

where $\Delta <>_{error}$ is the error associated with different interaction energies, $< \partial clay >^2_{stdev}$ is the standard deviation obtained by the block average method in clay, and $< \partial solution >^2_{stdev}$ is the standard deviation obtained by the block average method in solution.

One problem that could not be resolved using experimental methods was the decoupling of the underlying forces that drive the sorption process. Using the LIE method, this was achieved by decoupling the ΔG_{LIE} with the following equation:

$$\Delta\Delta G_{LIE} = \Delta\Delta G_{vdW} + \Delta\Delta G_{el} \tag{4.3}$$

Where $\Delta\Delta G_{vdW} = \alpha(\Delta < E_{vdw} > -\Delta < E_{vdW} >_{Aniline})$ and $\Delta\Delta G_{el} = \beta(\Delta < E_{el} > -\Delta < E_{el} >_{Aniline})$

Aniline was used as the reference compound.

4.3. LIE Fitting

After obtaining different interaction energies from MD simulations (Δ<>), a multivariable linear regression was applied to determine the coefficients α, β, and δ (equation 4.1) that were fitted to experimental free energy values ΔG_{EXP}.[81] All of the coefficients were allowed to optimize freely. Experimental binding energies were calculated as a ratio of equilibrium concentrations, using equation 4.4, with the gas constant (R) and temperature (T):

$$\Delta G_{EXP} = -RT \ln\left(\frac{K_d}{K_0}\right) \tag{4.4}$$

where a reference state is assumed and represented by a molecule with a sorption constant K_0 of 1 L/Kg, and T was set to 298 K. Experimental K_d values were obtained from the linear portion of the sorption isotherms under definite experimental conditions (i.e. ionic strength, pH).

A leave-one-out (LOO) cross validation scheme was utilized to determine the averaged α, β, and δ values (equation 4.1) and to calculate the associated statistical error. To illustrate, LOO was implemented ten times, leaving one molecule out of the training set each time [1-10 (Table 3)]. Using the generated values of α, β, and δ of ten different data sets, averages were computed. The statistical uncertainty of the derived LIE coefficients was computed using bootstrapping statistics (due to the small sample size), with the statistical software IBM SPSS. Additional testing was conducted using a test set: calculating ΔG_{LIE} for compounds that were not included in the training set [11-15 (Table 3)].

4.4. Results and Discussion

The applicability of the LIE method was examined by evaluating the sorption affinities of 15 organic solutes that bind to clay Ca-MMT (structures of the relevant organic solutes are provided in Figure 17). The training set that was used to compute LIE parameters (compounds 1-10) was selected arbitrarily from a set of 15 compounds. The remaining compounds (11-15) were used as the test set. It was observed that by exchanging compounds from the test set to the training set and vice versa, the quality of the fitting deteriorated or improved insignificantly. Calculated differences in the solute-surrounding interaction energies ($\Delta<>$) of the solutes are shown in Table 3 and are based on averages of finite temperature MD simulations. The results for the LIE fitting are as follows: $\alpha = 0.594 \pm 0.323$, $\beta = 0.130 \pm 0.035$, and $\delta = 2.9 \pm 1.8$ kcal/mol with a coefficient of determination (r^2) of 0.83. Calculated LIE free energy values (ΔG_{LIE}) compared well with the experimental values, with a mean unsigned error (MUE) of 0.3 kcal/mol. Plots of correlation between the experimental values for ΔG_{EXP} and ΔG_{LIE} are shown in Figure 18. The robustness of the LIE method was demonstrated by the accurate prediction of free energy of compounds that were not included in the training set (test set) with high accuracy (MUE = 0.4 kcal/mol, or within a factor 2 in K_d value).

Analysis of the MD trajectory provided detailed atomistic level insight into the orientation of all aromatic ammines. It was observed that all of the ammines oriented such that the benzene ring was parallel or near parallel to the clay sheets (Figures 20.1 and 20.2). Such an orientation maximizes electrostatic (amine group) and hydrophobic interactions (benzene ring) with the MMT surface. A similar coplanar arrangement of aromatic molecules with the silicate sheet under low external concentration has been observed in numerous experimental studies.[195-197] This further validates the accuracy of the atomistic models that were used for the MD simulations.

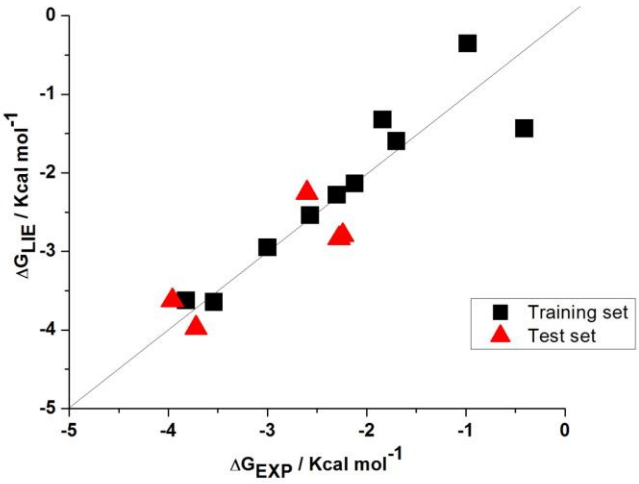


Figure 18. Comparison of fits of the LIE model (Equation 3.1) for the training set of ten compounds and the test set of five compounds on Ca-MMT. Best fit LIE parameters are as follows: $\alpha = 0.594 \pm 0.323$, $\beta = 0.130 \pm 0.035$, and $\delta = 2.9 \pm 1.8$ kcal/mol.

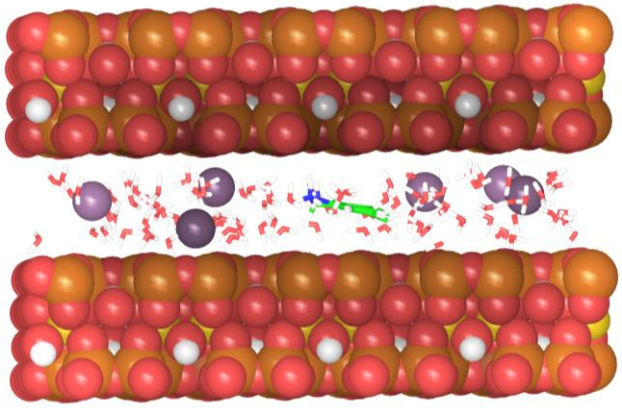


Figure 19. Typical structure of the hydrated interlayer with the organic cation is shown. (Red = oxygen, green = carbon, plum = calcium, orange = silicon, blue = nitrogen, yellow = aluminum, white/gray = hydrogen. Magnesium atoms are not seen in this view but are present).

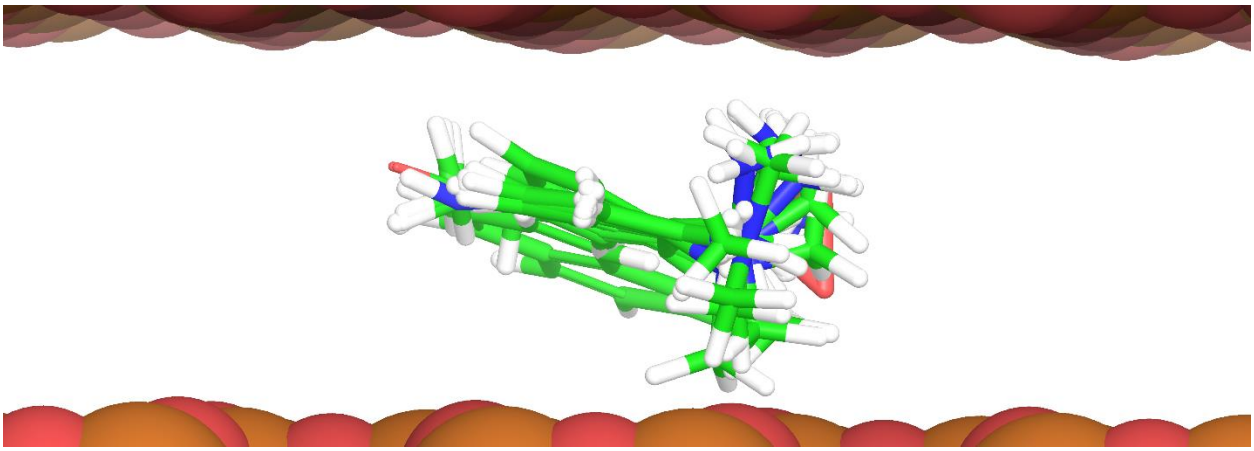


Figure 20.1. Superposition of all aromatic ammines in their most common conformation.

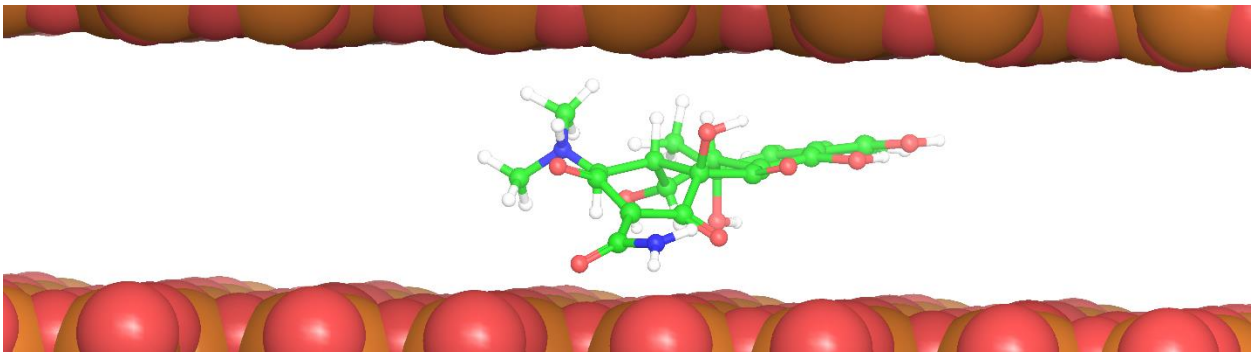


Figure 20.2. Oxytetracycline in its most common configuration.

Equation 4.3 was used to decouple the contribution of hydrophobic and electrostatic interactions to the ΔG_{LIE}. The purpose of this was to investigate and understand the structure-affinity relationships and trends within aromatic amine sub-groups (Table 3).

In the compound series 1-4 and 5-7 (derivatives of aniline and p-toluidine, from here on referred to as group A), subsequent decoupling of ΔG_{LIE} as an expansion of $\Delta\Delta G_{el}$ and $\Delta\Delta G_{vdW}$ led to the following observations: there was an equivalent contribution to the ΔG_{LIE} from $\Delta\Delta G_{el}$ and $\Delta\Delta G_{vdW}$, and there was a systematic increase in free energy as a function of the increment of methyl groups (primary, secondary, and tertiary ammines). To better understand the origin of the systematic increase of the $\Delta\Delta G_{el}$, QM calculations were used to analyze the charge distribution of

compounds in group A. After analyzing the molecular electrostatic potential (MEP), a defocusing of charge over the nitrogen's coordination sphere was evident (Figure 21); this increased the extent of interaction (higher surface area) with the negatively charged clay surface. In a previous study by Vasudevan *et al.* and Droge *et al.,* the observed trend of primary, secondary, and tertiary ammines was interpreted as a result of focusing the charge[73, 197] on the amine N and H molecules. They did not account for the charge distributions on the methyl groups.

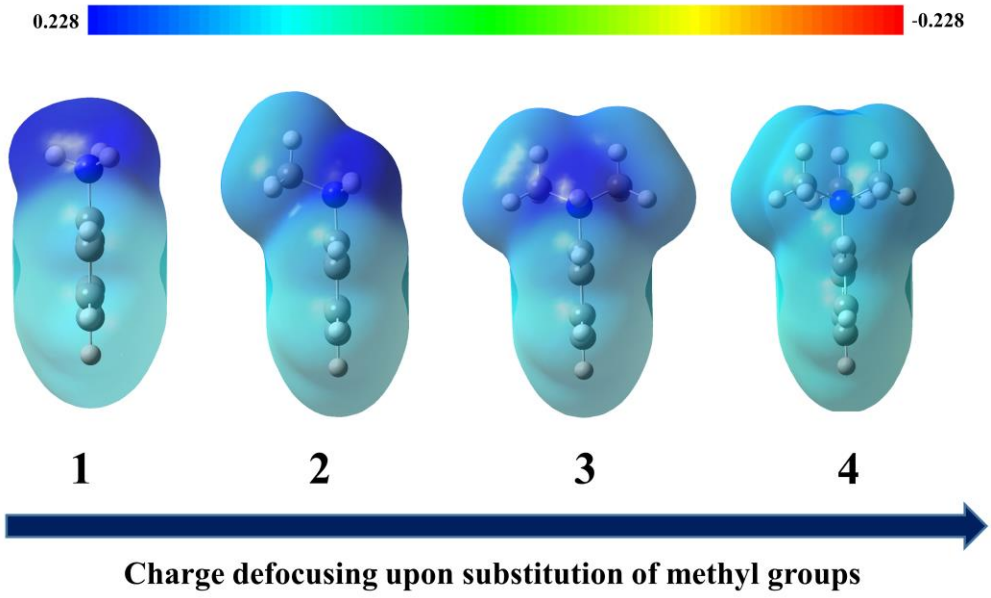


Figure 21. Illustration of molecular electrostatic potential (MEP) mapped onto the total electron density.

Conversely, the work presented in this chapter demonstrates that the spread of charge over the N, H, and methyl groups permits enhanced interaction with the negatively charged clay surface, resulting in higher sorption.[198] This analysis is consistent with previous observations of aromatic heterocyclic ammines which concluded that enhanced affinity was due to the positive charge delocalization to the ring, which enabled a molecular orientation that maximized the electrostatic attraction.[199] In class A, $\Delta\Delta G_{vdW}$ also increased with the addition of methyl substituents; this was expected, as the addition of methyl groups increases the number of hydrophobic sites.

Table 3. Calculated Free Energies via the LIE Method and Comparison with Experimental Values for the Training (1-10) and Test (11-15)[a]

	Compound	$\Delta\langle E_{vdW}\rangle$	$\Delta\langle E_{El}\rangle$	[b]ΔG_{LIE}	ΔG_{EXP}	$\Delta\Delta G_{vdW}$	$\Delta\Delta G_{el}$
1	aniline	-4.07 ± 0.40	-13.89 ± 0.36	-1.32	-1.84	0	0
2	N-methylaniline	-5.18 ± 1.05	-16.17 ± 0.82	-2.28	-2.30	-0.66	-0.29
3	N,N-methylaniline	-6.06 ± 0.46	-17.28 ± 1.78	-2.95	-3.00	-1.18	-0.44
4	Phenyltrimethylammonium	-6.09 ± 0.34	-22.36 ± 0.97	-3.63	-3.82	-1.20	-1.10
5	p-toluidine	-5.14 ± 0.36	-11.07 ± 2.08	-1.59	-1.72	-0.63	0.37
6	N-methyl-p-toluidine	-5.47 ± 0.32	-16.84 ± 1.98	-2.54	-2.57	-0.83	-0.38
7	N,N-dimethyl-ammonium	-5.55 ± 0.32	-25.00 ± 0.28	-3.65	-3.54	-0.88	-1.44
8	[c]2-(dimethylamino)-5-methylbenzenesulfonate	-5.80 ± 0.22	-6.82 ± 1.10	-1.43	-0.41	-1.03	0.92
9	(charge of 0)	-9.26 ± 0.79	17.29 ± 2.25	-0.35	-0.96	-3.08	4.06
10	[c]4-(aminomethyl) benzoic acid (charge of 0)	-5.38 ± 0.86	-14.15 ± 2.80	-2.13	-2.12	-0.78	-0.03
11	Benzylamine	-7.07 ± 0.93	-11.48 ± 1.44	-2.79	-2.24	-1.78	0.32
12	2,4-dicholorobenzylamine	-7.12 ± 0.42	-11.57 ± 3.36	-2.83	-2.28	-1.81	0.31
13	2,4-dimethylbenzylamine	-5.47 ± 0.72	-14.64 ± 1.28	-2.25	-2.60	-0.83	-0.09
14	p-phenylenediamine	-7.44 ± 0.51	-18.85 ± 1.07	-3.97	-3.72	-2.00	-0.64
15	1-napthylmethylbenzylamine	-16.10 ± 1.34	23.36 ± 3.23	-3.62	-3.96	-7.14	4.85
	[c]Oxytetracycline (charge of 0)						

[a]Experimental and calculated LIE binding free energies (Kcal/mol) are shown. Compound structures are shown in figure 17. [b]Training set mean unsigned error (MUE)=0.3 kcal/mol, and test set MUE = 0.4 kcal/mol. [c]Indicates a zwitterion.

The LIE model was also accurate at predicting the free energy of the sorption of zwitterionic compounds (8, 9, and 15 are referred to as group B) with both, positively and negatively charged groups; consequently, they have an overall neutral charge. Due to unfavorable electrostatic interactions (electrostatic repulsion of the negatively charged groups from the clay surface), van der Waals interactions were the main driving force of sorption. Electrostatic repulsion of

zwitterionic compounds can be qualitatively explained by examining the MEP of group B molecules. In compound 8, positively (amine) and negatively charged (sulfonate) groups are in close proximity (~4 Å), leading to screening of the negative charge on the sulfonate group. Conversely, in compounds 9 and 15, positively and negatively charged groups are further separated (~ 8 Å). This leads to the favorable electrostatic interactions of compound 8 with the clay surface, compared to the interactions of compounds 9 and 15.

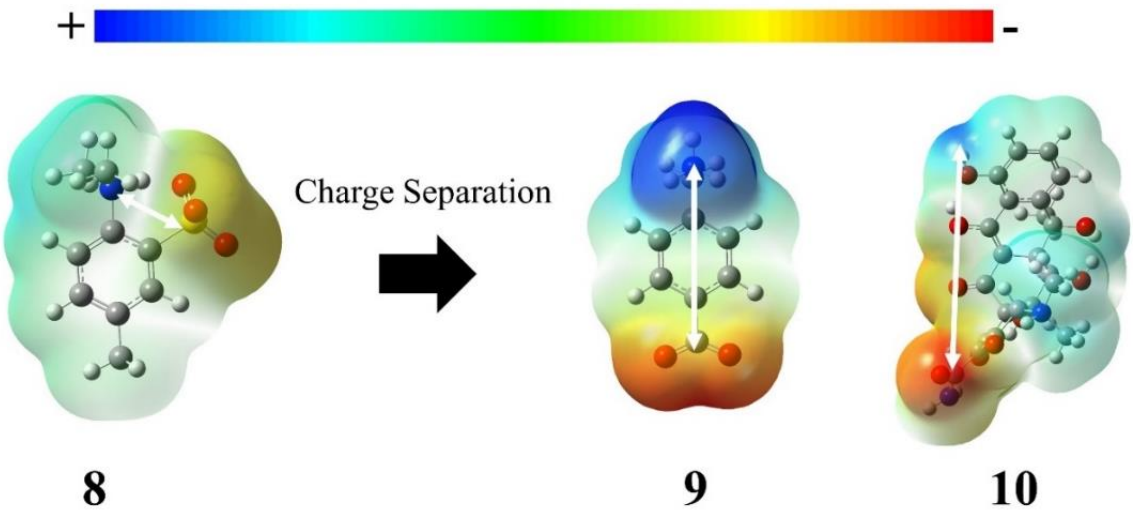


Figure 22. Illustration of the MEP of zwitterionic compounds 9, 8, and 15

4.5. Summary and Conclusions

The study presented here investigated the applicability of the well-known LIE method in combination with MD simulations to predict the sorption free energy of organic cations and zwitterions to naturally occurring clay Ca-MMT. This study was motivated by the inability of experimental methods to decouple the solute specific contribution to the sorption free energy when developing predictive models, and in contrast, the success of the LIE method in mapping biochemistry problems. The standard form and parameterization of the LIE method accurately reproduced the binding free energies of the majority of the compounds used in the compound-

library. The LIE model reproduced trends within and between homologous compound classes. LIE facilitates the application of corrective increments and decouples the contribution of polar and non-polar to sorption free energy. Calculations demonstrate that, although MD simulations can be computationally expensive (solution phase), the LIE method offers a convenient method for gaining accurate predictions of solute-clay binding free energies. Since the experimental sorption coefficients were obtained under a specific set of experimental conditions (e.g. ionic composition and sorption in the linear range of the isotherm), re-parameterization of the LIE coefficients may be required for different experimental conditions. Given these points, the LIE based methods to estimate sorption free energy have the potential to become powerful tools that can be used in the environmental risk assessment of polar organic compounds.

Chapter 5. Calculation of the Binding Affinities of Charged Ammines to Montmorillonite with Different Exchangeable Cations (Ca^{+2} and Na^{+}): A Linear Interaction Energy Method

5.1. Background

As a result of increase in use of various novel organic compounds (emerging contaminants) that are used in day to day human activities, there is also an increase in the concentration of these compounds in the environmental.[25, 75] Majority of these contaminants exist in the aqua sphere as positively charge cations due to the presence of basic amine groups (protonation of the amine group at environmentally relevant pH values).[73] Environmental risk assessment of these compounds is essential as they have been found to be toxic to human and animals alike.[1, 54] Determination of the availability of these contaminants in different environmental compartments (fate) is an important step in risk assessment models.[85]

In the previous chapter, a well-known method used in computational biochemistry to approximate free energies of ligands binding to receptor sites (proteins), known as the LIE method was applied to approximate free energies of sorption of a set of organic compounds (charged and neutral) to a representative homoionic smectite Ca-montmorillonite.[198] A mean average error of 0.3 kcal mol^{-1} with a r^2 value of 0.84 was obtained, demonstrating the accuracy and applicability of the method in approximating free energies of sorption. In the LIE method, free energies were computed as an expansion (weighted) of ensemble averaged changes of the interaction energies of the solutes with its environment (in solution and clay) which captured variety of structural effects that control the sorption process such as: defocusing of amine charge, effects of negatively charge groups, functional group proximity, etc. Ultimately, the LIE method was concluded to be a robust method for the estimation of organic cation adsorption free energies.

As clays such as montmorillonites, kaolinites carry negative charges, various cations that exist in the soil (e.g. sodium, calcium) are attracted to the clay particles. These cations are held to the clay surface by electrostatic interactions and termed as adsorbed cations, which are free to diffuse in to the liquid phase and also can get replaced (exchanged) by other cations that exist in the liquid phase.[200, 201] The ease of cation exchange often follows the lyotropic series as:

Held more strongly Held more weakly

$$H^+ \geq Al^{3+} > Ca^{2+} > Mg^{2+} > NH_4^+ = K^+ > Na^+$$

Figure 23. The Lyotropic Series

Ions which are strongly held to the negatively charged surfaces (i.e. H^+, Al^{+3}) will have high valence to hydrated radius ratio while ions which are weakly held such as K^+ and Na^+ will have lower ratio.

Table 4. Hydrated Radius of Selected Metals that Occupy Interlayer Space.[202]

Ion	Radius (Å)
K^+	5.3
Na^+	7.9
Ca^{+2}	9.6
Mg^{+2}	10.8

Because of this, different cations may have different abilities to exchange the adsorbed cations. As calcium and sodium are the most common of these exchange ions in natural environments, implementation of the LIE method to compute sorption free energies of clays having aforementioned cations is important for the development of the the LIE method as a predictive tool and to understand the trends that are observed which could be helpful to apply the method to much complex systems.

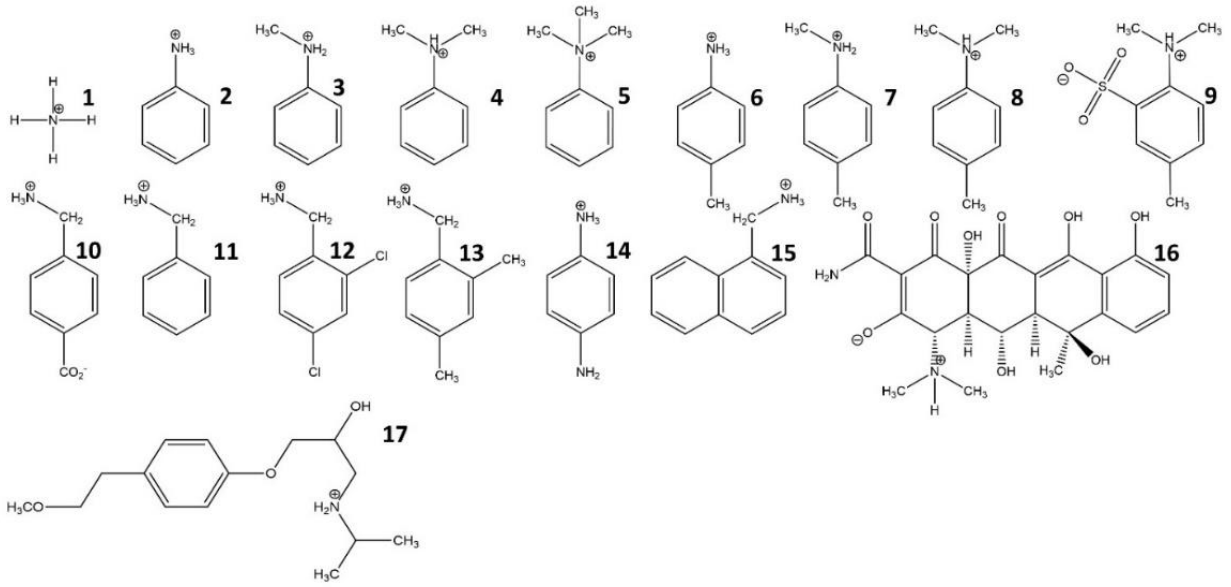


Figure 24. Chemical Structures of Compounds Studied 1-17

In this chapter, the LIE method was extended to investigate the accuracy of the predicted sorption free energies and transferability of the LIE coefficients (equation 4.1) within MMTs having similar composition only varied by different interlayer ions (calcium and sodium).To decouple the underlying forces that drives the sorption process, ΔG_{LIE} was decomposed as:

$$\Delta\Delta G_{LIE} = \Delta\Delta G_{vdW} + \Delta\Delta G_{el} \tag{5.1}$$

Where $\Delta\Delta G_{vdW} = \alpha(\Delta < E_{vdw} > -\Delta < E_{vdw} >_{Ammonium})$ and $\Delta\Delta G_{vdW} = \beta(\Delta < E_{el} > -\Delta < E_{el} >_{Ammonium})$. Ammonium was used as the reference compound.

5.2. Computational Details

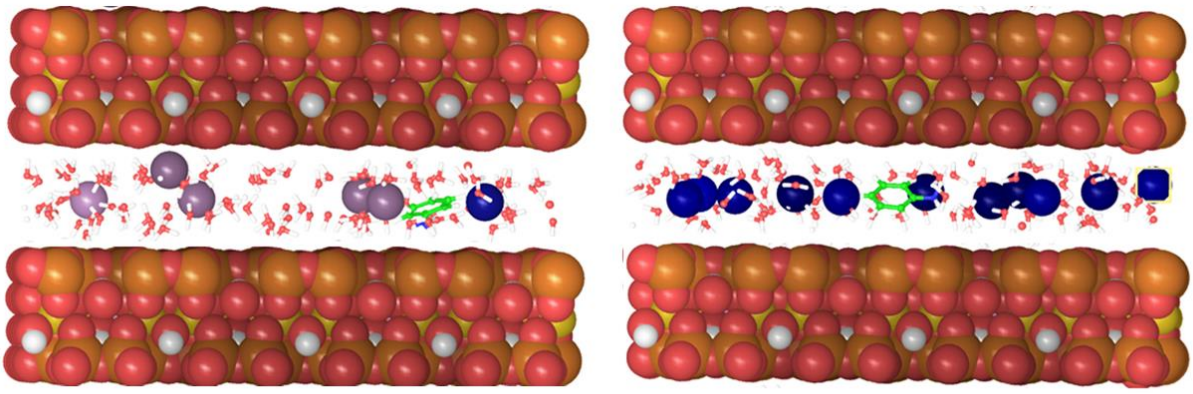
Ground state geometries of organic compounds were obtained by optimizing there structures at b3lyp/6-31G(d) level of theory in the gas phase using the Gaussian package.[190] A cubic box with an initial box length of 60 Å with periodic boundary conditions was used to simulate organic cations/zwitterions in solution. To model the compounds having a +1 overall charge, a single

organic cation was placed in the cubic box along with one Na^+ and two Cl^- ions to neutralize the overall charge. To model the zwitterions having a zero overall charge, a single zwitterion was placed in the cubic box along with a Na^+ and a Cl^- ion. Subsequently, water molecules were added so the density of the simulation box was 1.00 g cm^{-3}.

Previously reported structure of a MMT with a unit cell formula, $(Si_{7.75}Al_{0.25})(Al_{3.5}Mg_{0.5})O_{20}(OH)_4.xH_2O$ was used to construct the atomistic models.[188] A super cell having 32 unit cells was constructed by translating the unit cell four times in the x and y directions and two times in the z direction ($4 \times 4 \times 2$). The resulting super cell had two montmorillonite slabs, each having a charge of -12.00e (-0.75e per unit cell).

Ca Montmorillonite: Atomistic models for MMT with Ca^{+2} ions (Ca-MMT) was constructed by using the aforementioned super-cell, by placing a single solute molecule (organic compound) in the interlayer space and by adding Ca^{+2} and Na^+ ions to neutralize the overall charge (-24*e*). In detail, eleven Ca^{+2} ions and a single Na^+ ion were added to clay models having compounds 1-8, 11-15, eleven Ca^{+2} ions and two Na^+ to clay models having compounds 10 and 16, and twelve Ca^{+2} ions to the clay model having compound 9. Afterwards, water molecules were added to form a double-layer hydrate (sterically clashing molecules were removed).

Na Montmorillonite: Atomistic models for MMT with Na^+ ions (Na-MMT) was constructed by using the clay models from the previous step and by replacing Ca^{+2} ions by Na^+ ions to neutralize the overall charge. In detail, 23 Na^+ ions were added to clay models having compounds 1-5 11-13, 15,17 and 24 Na^+ ions were added to clay models having compounds 9, 10 and 16. Subsequently, water molecules were added to form a double-layer hydrate (sterically clashing molecules were removed).



Typical structure of the hydrated interlayer with the organic cation with Ca^{+2} ions

Typical structure of the hydrated interlayer with the organic cation with Na^{+} ions

Figure 25. Typical Structures of the Hydrated Interlayer (viewed along xz plane) With the Organic Cation is Shown. (Red = oxygen, green = carbon, purple = calcium, orange = silicon, light blue = nitrogen, yellow = aluminum, white/gray = hydrogen, sodium = dark blue, plum = calcium, Magnesium atoms are not seen in this view but are present).

To model the Ca and Na MMT, parameters from the CLAYFF force field[110] were utilized, which has been previously utilized to model interactions between organic molecules with montmorillonite clays. Organic cations/zwitterions were modeled using the OPLS-AA force field[104] which has been widely utilized to model in the condensed phase simulations. Water molecules were modeled using the SPC water model[123] considering its simplicity. Partial charges for the organic cations/zwitterions were derived using RESP method[154] using DFT b3lyp/6-31g(d) level of theory in the gas phase, in agreement with the philosophy of the OPLS-AA force field.

All the molecular dynamics (MD) simulations were performed using the MD code DL_POLY CLASSIC version 1.8[192], using the Nose-Hoover[135] isothermal-isobaric (NPT) ensemble with a coupling constant of 0.5 ps, with the temperature of 298 K and the pressure of 1 atm. A time step of 1 fs was used for the simulations. Electrostatic interactions were computed using the Ewald

summation method which takes into account long range electrostatic interactions.[140] The model having solute in water was equilibrated for 200 ps followed by a production time of 200 ps. The clay/solute system was equilibrated for 1500ps followed by a production time of 500ps with periodic boundary conditions. The length of the production was determined by converged values in the average interaction energies and or by using similar production times used by earlier studies found in literature.

The data collection phase of the simulation was split into blocks of 50ps, for the solute-clay (Ca/Na MMT) and solute-water calculations. The average electrostatic and van der Waals interaction energies of the solute with its environment (water and clay) were calculated for each block. The overall means and standard errors were calculated from these blocks. Error assessment of the computed free energies was carried out in a similar way by splitting the production time of simulation into equal blocks (in solution splitting 200 ps into two blocks of 100ps and in clay splitting 500 ps into blocks two blocks of 250ps); the means and standard errors were computed from these blocks.

5.3. LIE fitting

After obtaining different interaction energies from MD simulations ($\Delta<>$), multivariable linear regression was applied to determine coefficients α, β and δ (equation .1) fitted to experimental free energy values ΔG_{EXP} for MMT having different interlayer ions (Ca^{+2} and Na^{+}). Experimental K_d values in Ca and Na MMT were obtained from the linear portion of the sorption isotherms and under definite experimental conditions (i.e. ionic strength, pH).

Linear regression was conducted 11 times using Ca-MMT data, leaving one molecule out of the training set [**1-11** (Table 6)] each time (LOO method). Further testing was achieved by

calculating the ΔG_{LIE} of several compounds which were not in the training set [**12-16** (Table 6)]. All LIE coefficients were allowed to optimize freely.

LIE parameters derived using Ca-MMT was then used in Na-MMT with an additional optimization of the δ parameter (fixing α and β) 10 times, leaving one molecule out of the training set [**1-5,11-13,15-16** (Table 7)] each time. Optimization was carried in such a way that mean squared error (MSE) between experimental free energy and the LIE predicted free energy was minimized (equation 4.1). Further testing was achieved by calculating the ΔG_{LIE} of several compounds which were not in the training set [**10 and 17** (Table 7)].

$$MSE = \frac{1}{n}\sum_{i=1}^{N}(\Delta G_{EXP} - \Delta G_{LIE})^2 \tag{5.2}$$

Where n is the number of predictions, ΔG_{EXP} is the experimental free energy calculated using equation 3.2 and ΔG_{LIE} is the LIE predicted free energy using equation 5.1.

5.4. Results and Discussion

Our first task is to vindicate the level of accuracy and range of applicability of the LIE model for predicting sorption coefficients of positively charged and zwitterionc organic compounds. The statistical results for all the LIE models are summarized in Table 5. It was found that the if LIE coefficients derived for Ca-MMT model were directly used in LIE model for Na-MMT, a MSE of 0.4 kcalmol^{-1} was observed; thus, separate optimization of the fitting parameter δ for the Na-MMT model keeping parameters α and β fixed was carried out. In the case of Na-MMT the optimal value for δ was 0.844 kcal mol^{-1}. The lowering of the parameter δ leads to a decrease in the MSE of Na-MMT to 0.08 kcalmol^{-1}.Thus, it was conclude that LIE coefficients (α and β) are transferable within MMT models with identical chemical composition having different counter ions with an

additional optimization of the coefficient δ. A lower value of δ for Na-MMT model may in fact be a result of changes in electrostatic environments around the organic cation in the interlayer space due to presence of different charge compensating ions Ca^{+2} and Na^{+}. Since small changes in the charge distribution around the organic cation will lead to large changes in the interaction energies due to strong electrostatic forces involved. Table 6 and 7 list the calculated difference in energy components, sorption free energies by the LIE (equation 5.1) method together with the respective experimental data for Ca and Na MMT. Plots of correlation between the experimental ΔG_{EXP} and ΔG_{LIE} are shown in Figure 26. Comparison of the observed and LIE predicted free energy values in table 6 and 7 for Ca and Na MMT clearly demonstrates that the LIE is very accurate in predicting sorption free energies. Further, R^2 value for Ca-MMT model (Table 6) was 0.818 with a mean squared error (MSE) of 0.17 $kcalmol^{-1}$ and for Na-MMT model (Table 7) R^2 value was 0.893 with a MSE of 0.08 $kcalmol^{-1}$ for the training and test sets. In view of the highly successful performance of the LIE model in predicting the sorption free energies of

Table 5. Optimized LIE Coefficients for Ca-MMT and Na-MMT

System	α	β	δ / kcal mol^{-1}
Ca-MMT	0.464 ± 0.056	0.180 ± 0.031	1.627 ± 0.548
Na-MMT	0.464 ± 0.056	0.180 ± 0.031	0.844 ± 0.030

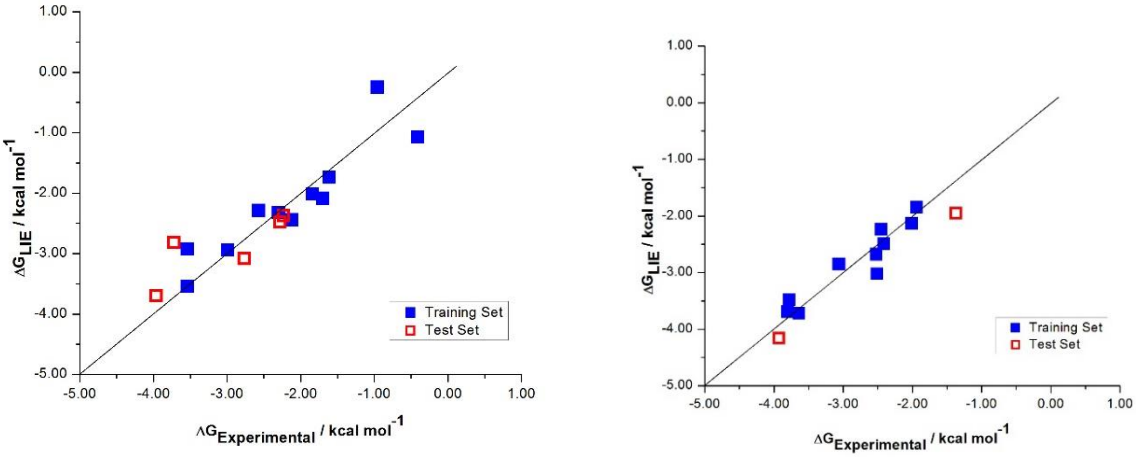


Figure 26. Free energy of sorption estimated by the LIE method (ΔG_{LIE}) versus the experimentally determined free energy of sorption ($\Delta G_{Experimental}$) for Ca-MMT (left) and Na-MMT (right)

K_{ow} coefficients are commonly accepted as accurate approximates of the soil sorption coefficients for apolar/polar organic compounds through correct approximation of K_{oc} (Equation 1.2).[95] In the case of charged organic compounds, the driving force for the sorption process is subjugated by electrostatic interactions with the negatively charged clay surfaces with lesser contributions from hydrophobic and van der Waals interactions. Models based on K_{ow} correctly takes into account hydrophobic and polar contributions through van der Waals interactions (thus K_{oc}) and does not correctly approximates the electrostatic interactions (thus K_{ce} in equation 1.2), thus underestimates the value of K_d. Further, models based on octantol/water partition coefficients does not take in to account specific features of environmental solids such as cation exchange capacity, surface areas, etc., which are important factors that determines the extent of sorption through cation exchange.[81] In poly parameter linear free energy relationships (PPLFER) even though we take into account key structural features of the sorbate to estimate the free energy changes, the method does not take into account the key features of the sorbent such as availability of exchange sites, surface area, etc. [215] As stated these which are important aspects that determine the extent of sorption through cation

exchange. Thus, we concluded that the accuracy of the LIE method is because of its ability to incorporate both: key structural features of the sorbate

Table 6. Calculated Free Energies for the Ca-MMT via the LIE method and Their Comparison with Experimental Values for the Training (**1-11**) and Test (**12-17**) Sets.

	Compound	$\Delta\langle E_{vdW}\rangle$	$\Delta\langle E_{El}\rangle$	[b]ΔG_{LIE}	[a]ΔG_{EXP}	$\Delta\Delta G_{vdW}$	$\Delta\Delta G_{el}$
1	Ammonium	-0.10 ± 0.68	-18.45 ± 2.20	-1.74 ± 0.19	-1.61	0.00	0.00
2	Aniline	-3.21 ± 0.73	-11.93 ± 1.60	-2.01 ± 0.17	-1.84	-1.44	1.17
3	N-methylaniline	-3.38 ± 0.67	-13.22 ± 1.98	-2.32 ± 0.16	-2.30	-1.52	0.94
4	N,N-dimethylaniline	-3.79 ± 0.73	-15.61 ± 1.27	-2.94 ± 0.12	-2.99	-1.71	0.51
5	phenyltrimethylammonium	-4.69 ± 0.22	-16.65 ± 1.34	-3.55 ± 0.18	-3.54	-2.13	0.32
6	p-toluidine	-3.19 ± 1.04	-12.41 ± 2.14	-2.09 ± 0.37	-1.72	-1.43	1.09
7	N-methyl-p-toluidine	-2.90 ± 0.43	-14.28 ± 1.80	-2.29 ± 0.23	-2.57	-1.30	0.75
8	N,N-dimethyl-p-toluidin	-3.59 ± 0.60	-16.01 ± 1.86	-2.92 ± 0.39	-3.54	-1.62	0.44
9	DMTS (charge 0)	-1.94 ± 0.55	-9.97 ± 2.21	-1.07 ± 0.05	-0.41	-0.85	1.53
10	4-(aminomethyl) benzoic acid (charge 0)	-1.02 ± 1.11	-7.79 ± 4.04	-0.25 ± 0.41	-0.96	-0.43	1.92
11	Benzylamine	-2.61 ± 0.58	-15.85 ± 2.06	-2.44 ± 0.16	-2.12	-1.16	0.47
12	2,4-dicholorobenzylamine	-5.54 ± 0.40	-8.52 ± 1.28	-2.48 ± 0.24	-2.28	-2.52	1.79
13	2,4-dimethylbenzylamine	-3.94 ± 0.66	-12.06 ± 1.90	-2.37 ± 0.30	-2.23	-1.78	1.15
14	p-phenylenediamine	-6.31 ± 0.47	-9.91 ± 1.78	-3.08 ± 0.09	-2.77	-2.88	1.54
15	1-napthylmethylbenzylamine	-4.68 ± 0.72	12.64 ± 3.33	-2.82 ± 0.40	-3.72	-2.12	1.05
16	Oxytetracycline (charge 0)	-16.35 ± 1.19	12.57 ± 3.00	-3.70 ± 0.56	-3.96	-7.54	5.58

[a]Experimental and calculated LIE binding free energies (Kcal/mol) are shown. Compound structures are shown in figure 24. [b]Training set mean squared error (MSE)=0.15 kcal/mol, and test set MSE = 0.21 kcal/mol. [c]Indicates a zwitterion.

Table 7. Calculated Free Energies for the Na-MMT via the LIE method after optimization of δ from Ca-MMT keeping α and β fixed, for the Training (**1-5, 11-13, 15-16**) and Test (**10 and 17**) Sets.

Compound		$\Delta\langle E_{vdW}\rangle$	$\Delta\langle E_{El}\rangle$	[b]ΔG_{LIE}	[a]ΔG_{EXP}	$\Delta\Delta G_{vdW}$	$\Delta\Delta G_{el}$
1	Ammonium	-1.68 ± 0.68	-10.62 ± 2.20	-1.85 ± 0.15	-1.94	0.00	0.00
2	Aniline	-2.18 ± 0.88	-10.94 ± 2.92	-2.14 ± 0.11	-2.01	-0.23	-0.06
3	N-methylaniline	-2.78 ± 1.01	-11.39 ± 3.61	-2.49 ± 0.76	-2.41	-0.51	-0.14
4	N,N-dimethylaniline	-3.44 ± 0.66	-11.63 ± 1.05	-2.85 ± 0.07	-3.06	-0.82	-0.18
5	phenyltrimethylammonium	-4.27 ± 0.35	-13.03 ± 1.33	-3.48 ± 0.18	-3.78	-1.20	-0.43
11	Benzylamine	-2.55 ± 0.66	-10.56 ± 2.70	-2.24 ± 0.11	-2.45	-0.40	0.01
12	2,4-dicholorobenzylamine	-5.77 ± 0.71	-6.63 ± 2.27	-3.02 ± 0.23	-2.51	-1.90	0.72
13	2,4-dimethylbenzylamine	-4.37 ± 0.85	-8.31 ± 1.52	-2.68 ± 0.29	-2.52	-1.25	-0.42
15	1-Napthylbenzylamine	-4.98 ± 1.11	-12.35 ± 1.87	-3.69 ± 0.41	-3.80	-1.53	-0.31
16	[c]Oxytetracycline (charge 0)	-13.15 ± 0.91	8.56 ± 4.07	-3.72 ± 0.86	-3.64	-5.32	3.45
10	[c]4-Aminobenzoicacid (charge 0)	-6.95 ± 0.56	2.38 ± 4.80	-1.95 ± 0.54	-1.37	-2.44	2.34
17	Metoprolol	-11.63 ± 0.63	2.16 ± 3.23	-4.16 ± 0.50	-3.93	-4.62	2.30

[a]Experimental and calculated LIE binding free energies (Kcal/mol) are shown. Compound structures are shown in figure 24. [b]Training set mean squared error (MSE)=0.05 kcal/mol, and test set MSE = 0.19 kcal/mol. [c]Indicates a zwitterion.

organic cations, second task was to examine why its predictive ability is far superior to other empirical models that exists such as methods based on octanol/water partition coefficients (K_{OW}) and polyparameter Linear Free Energy relationships (PPLFER).[89, 203] (organic cation) and sorbent (soil or clay) and other thermodynamic factors such as entropy to the free energy implicitly through changes in electrostatic and van der Waals interaction energies, weighed according to the LIE coefficients which have been derived by fitting to reproduce known sorption free energy values.

Based on the production MD trajectories sampling a NPT ensemble, the orientations of different organic molecules in solution and in Ca/Na MMT were analyzed. In solution it was observed that zwitterions 4-Aminobenzoicacid (10) and Oxytetracycline (16) formed a salt-bridge with the Na^+ ion. For this reason, during the intercalation process, there will be a high possibility that the bound Na^+ ion will be dragged into the interlayer of the clay. Bearing this in mind, in Ca-MMT model, a single Ca^{+2} ion was replaced with two Na^+ ions in clay model for 4-Aminobenzoicacid (10) and Oxytetracycline (16). It was observed that the aromatic ammines tend to orient in such a way that the phenyl ring was parallel or near parallel to the MMT sheets. As stated in the previous chapter such an orientation allows maximum electrostatic and van der Waals interactions with the clay surface and agrees with experimental studies which stated that under low surface concentrations aromatic compounds orient in such a way that the phenyl ring was parallel or near parallel to the plane of the clay surface. [195, 196]

Analysis of the energy components resulting from the decomposition of ΔG_{LIE} to $\Delta\Delta G_{el}$ and $\Delta\Delta G_{vdW}$ using equation 4.1 was carried out next. In both Ca and Na LIE models, homologous cationic compounds from 1-5 (derivatives of aniline), from here in referred as group A, there was a monotonic increase in the $\Delta\Delta G_{el}$ and $\Delta\Delta G_{vdW}$. From a structural point of view there was an increase in the number of methyl substituents at charged N atom (monotonically increases from 1 to 5). This increase can be expounded as a consequence of spread of positive charge over the substituted methyl groups and increase in hydrophobicity thus an enhanced interaction with the clay surface.[198]

In compounds 11-13 and 15 (derivative of benzyl amine), from here in referred as group B, values of $\Delta\Delta G_{vdW}$ exhibited an increase (irregular) relative to benzylamine (11). This can be expected as in compounds 12 and 13 addition of choloro and methyl groups and in compound 15

addition of a phenyl group will increase the hydrophobicity of the molecules. Thus, enhanced hydrophobic interactions with the clay surface. The effects of the aforementioned structural changes did not manifested a regular increase or decrease in $\Delta\Delta G_{el}$, as in class A, since, substitutions were further away from the positively charged amine group compared to substituted ammines.

Compounds 10, 11 and 16, from here in referred as group C, are zwitter ions which have a positive and a negative charge groups thus, are electrically neutral. In class C, zwitterions showed an unfavorable electrostatic interaction with the clay surface compared to classes A and B. The main driving force for zwitter ionic compounds was the van der Waals interactions with the clay surface in both models (Ca-MMT and Na-MMT).

It was observed experimentally that the free energies of sorption of organic cations sorbing to MMT having Na^+ as charge compensating ions were lower than that of MMT having Ca^{+2}. This was expected: as clarified by the lyatropic series: Na^+ has a lower charge to hydrated radius ratio than that of Ca^{+2}, thus, Ca^{+2} will be strongly bound to the negatively charged clay surface compared to Na^+. Hence, it will be easier to remove a Na^+ ion from MMT surface than a Ca^{+2} resulting a higher affinity of organic cations towards MMT with Na^+ ions. To further verify this assertion, ensemble averaged distance of the averaged z-coordinate (clay sheets orient in the xy plane) of the interlayer ions from the mid-point of the interlayer was calculated for both MMT models (Ca-MMT and Na-MMT). From the results it was evident that Ca^{+2} ions oriented further away from the mid-point of the interlayer (closer to the clay surface) relative to Na^+ ions. Thus agreeing with the philosophy behind the lyotropic series (Figure 23).

Table 8. Distances from the center of the MMT slab to the averaged center of the counter ions (Ca^{+2} and Na^{+})

Compound	Distance from the center of the interlayer (Ca- MMT)*	Distance from the center of the interlayer (Na-MMT)*
Ammonium (1)	1.65 ± 0.17	1.19 ±0.11
Aniline (2)	1.41 ± 0.02	1.09 ±0.04
N-Methylaniline (3)	1.37 ± 0.02	1.29 ±0.14
N, N- Methylaniline (4)	1.39 ± 0.03	1.08 ±0.07
Phenyltrimethylammonium (5)	1.43 ± 0.02	1.11 ±0.04
Benzylamine (11)	1.40 ± 0.01	1.22 ±0.13
2,4-dimethylbenzylamine (12)	1.42 ± 0.02	1.38 ±0.06
2,4-Dichlorobenzylamine (13)	1.34 ± 0.02	1.39 ±0.04
1-Napthylmethylbenzylamine (15)	1.44 ± 0.02	1.54 ±0.06
4-Aminomethyl Benzoic Acid (10)	1.38 ±0.01	1.42 ±0.07
Oxytetracycline (16)	1.16 ±0.04	0.75 ± 0.12

*All distances are in angstroms.

5.5. Summary and Conclusion

In the work summarized in this chapter, investigations were carried to determine the accuracy of the LIE method in predicting free energies of sorption considering MMT having identical chemical composition varied only by the interlayer cations (calcium and sodium). From the results of the calculations, it was evident that the LIE method was promising in predicting experimental sorption free energies with an accuracy of ±.50 kcal mol^{-1} that of experimental free energies. It was apparent that LIE parameters derived from Ca-MMT was transferable with an additional optimization of the parameter δ, stemming from the constant change in experimental free energy from ΔG_{CA} to ΔG_{NA}. Thus, we conclude that δ coefficient is counterion-dependent while

parameters α and β are transferable independent of the counter ion to predict sorption affinities at a given set of experimental conditions.

Chapter 6. Atomistic Prediction of Sorption Free Energies of Charged Ammines on to an External Montmorillonite Surface: A Linear Interaction Energy Method

6.1. Background

Due to upsurge in release of polar organic contaminants into the environment as a result of anthropogenic activities and there adverse effects on human, animal and plant lives, environmental risk assessment of aforementioned compounds is of utmost importance.[53] Determination of the sorption coefficients (K_d) of these compounds is an important step in formulating environmental risk assessment models.[202] Work carried out in this thesis so far, focused on the application of the LIE method as a predictive tool to evaluate sorption affinities of charged ammines to the prototypical aluminosilicate montmorillonite having different exchangeable ions: calcium and sodium. Results obtained from chapters four and five highlighted that the LIE method was promising with respect to the ability of predicting free energies of sorption irrespective of the interlayer ion present. It was also concluded that LIE coefficients are transferable with an additional optimization of the fitting parameter δ.

In the work summarized in chapters four and five, interaction energies of the organic cations with the MMT surface was calculated utilizing a model consisting of a single organic cation placed in the interlayer space between two clay slabs. The distance between two clay slabs (or the d-spacing) of the simulated clay models was set in such a way, that when appropriate number (with respect to mass of clay) of water molecules were added, a double layer hydrated structure was formed. Thus, to this point, we have only explored the adsorption of organic cations by the intercalation process (in a confined environment). Adsorption could also take place on the external surfaces of smectite clays at the mineral-solution interface (in an non-confined environment).[204]

To develop a through picture of molecular scale behaviors of various organic cations in such a non-confined environment and how they interact with the clay surface will require extensive spectroscopic studies if carried out experimentally.[205] Computational techniques permits an alternate approach of obtaining a detailed picture of interfacial structures or thermodynamic modelling to predict adsorption equilibria.[206]

This chapter presents a computational MD study of adsorption of a set of organic cations (Figure 24) to an external surface of homoionic MMT having sodium and calcium as counter ions. The ultimate goal of the study was to investigate whether such a model will reproduce experimental sorption free energies which strictly occurs due to sorption of organic cations in the interlayer (>90%) of MMT.

6.2. Computational Details

MMT with a unit cell formula of, $(Si_{7.75}Al_{0.25})(Al_{3.5}Mg_{0.5})O_{20}(OH)_4.xH_2O$ was used to build the atomistic models. A super cell having 32 unit cells was constructed by translating the unit cell eight times in the x direction, four times in y direction and one time in the z direction ($8 \times 4 \times 1$). The resulting super cell had a single MMT slab, having a charge of -24.00e (-0.75e per unit cell). Molecular structures of organic solutes (Figure 24) were taken from our previous work (chapter 5) optimized using DFT at 6-31G(d) level of theory using Gaussian 09 package.[190] Dimension of the z-axis was set to 37 Å, to make sure that the adjacent slabs of montmorillonite were completely separated, and thus the simulation would occur on the external surface of the montmorillonite slab. The center of mass of the MMT slab was placed in the center of the z-axis of the simulation box in such a way that the surface of the slab was parallel with the xy plane of the simulation box (Figure 27).

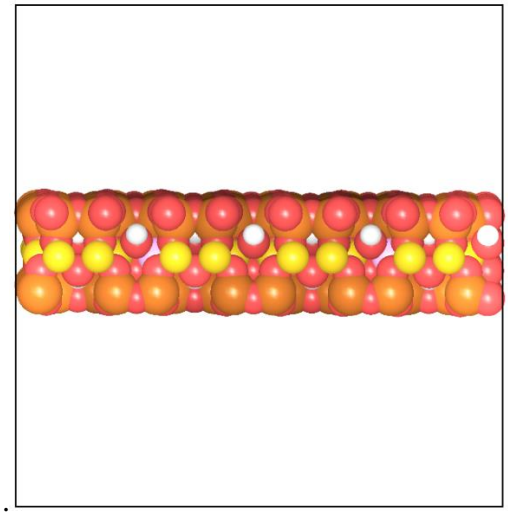
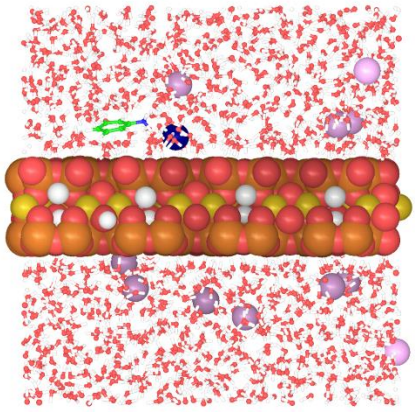


Figure 27. 2d view of the supercell (viewed along the xz plane) used for the simulations with the periodic boundaries highlighted by black lines (Red = oxygen, orange = silicon, yellow = aluminum, white/gray = hydrogen. Magnesium atoms are not seen in this view but are present).

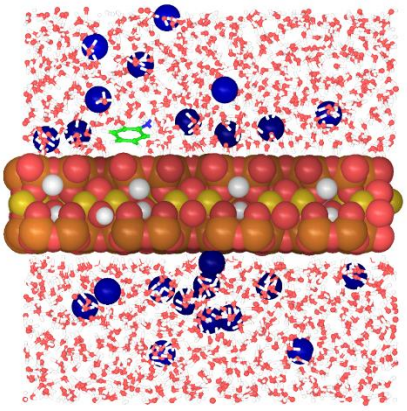
Ca Montmorillonite: Atomistic models for MMT slab with Ca^{+2} ions (Ca-MMT-surf) was constructed by using the aforementioned super-cell, by placing a single solute molecule (organic compound) in the interlayer space and by adding Ca^{+2} and Na^{+} ions to neutralize the overall charge (-24*e*). In detail, eleven Ca^{+2} ions and a single Na^{+} ion were added to clay models having compounds 1-8, 11-15, eleven Ca^{+2} ions and two Na^{+} to clay models having compounds 10 and 16, and twelve Ca^{+2} ions to the clay model having compound 9. Afterwards, water molecules were added to form a double-layer hydrate (sterically clashing molecules were removed).

Na Montmorillonite: Atomistic models for MMT slab with Na^{+} ions (Na-MMT-surf) was constructed by using the clay models from the previous step and by replacing Ca^{+2} ions by Na^{+} ions to neutralize the overall charge. In detail, 23 Na^{+} ions were added to clay models having compounds 1-5 11-13, 15, 17 and 24 Na^{+} ions were added to clay models having compounds 9, 10

and 16. Subsequently, water molecules were added to form a double-layer hydrate (sterically clashing molecules were removed).



Typical structure of the hydrated slab model with the organic cation with Ca^{+2} ions



Typical structure of the hydrated slab model with the organic cation with Na^{+} ions

Figure 28. Snapshots representing structures of the slab model viewed along the xz plane. (Red = oxygen, green = carbon, purple = calcium, orange = silicon, light blue = nitrogen, yellow = aluminum, white/gray = hydrogen, sodium = dark blue, plum = calcium, Magnesium atoms are not seen in this view but are present).

Afterwards, water molecules were added to each model in such a way that the density of the box (excluding the volume occupied by the MMT slab) was 1.00 g cm^{-3}. Water molecules that sterically clashed with the organic molecules were removed. Force field parameters for MMT were taken from the CLAYFF parameter set.[110] Parameters for the water molecules were taken from flexible SPC water model. Organic cations were modeled using the OPLS-AA force field.[104] Partial charges for the organic solutes derived using the RESP method in previous work (chapter 5) were used.[154]

MD code DL_POLY CLASSIC version 1.8 was used to carry out the simulations.[192] Periodic boundary conditions were employed to the simulation box in all (x, y and z) directions. Long range

electrostatics were treated with the Ewald summation method.[140] MD simulations were carried out in the NPT (isothermal isobaric) ensemble. The Nose-Hoover barostat was used with a relaxation time of 0.5 *ps* at a temperature T=300 K and pressure p=1.0 atm. Velocity-Verlet algorithm with a time step of 1 *fs* was used to propagate the system and to produce the atomistic trajectories. Prior to the addition of the inorganic cations (i.e. Na^{+} and Ca^{+2}) and water molecules, 0 K optimization was carried out to obtain the most stable configuration of the organic solutes on the clay surface. The clay/solute system was equilibrated for 1500ps followed by a production time of 500ps with periodic boundary conditions. The length of the production was determined by converged values in the average interaction energies and or by using similar production times used by earlier studies. During the production time solute-clay interaction energies (electrostatic and van der Waals) was recorded.

The data collection phase of the simulation was split into blocks of 50ps, for the solute-clay (Ca/Na MMT-surf) and solute-water calculations. The average electrostatic and van der Waals interaction energies of the solute with its environment (water and clay) were calculated for each block. The overall means and standard errors were calculated from these blocks. Error assessment of the computed free energies was carried out in a similar way by splitting the production time of simulation into equal blocks (in solution splitting 200 ps into two blocks of 100ps and in clay splitting 500 ps into blocks two blocks of 250ps); the means and standard errors were computed from these blocks.

6.3. LIE fitting

Interaction energies of the organic solutes in solution phase computed in the previous work (chapter 5) were used for this study. After obtaining ensemble averaged interaction energies of solute-clay from MD simulations, differences ($\Delta\langle\rangle$) were computed. Multivariable linear

regression was applied to determine coefficients α, β and δ (equation 3.1) fitted to experimental free energy values ΔG_{EXP} for MMT having different interlayer ions (Ca^{+2} and Na^{+}).

Linear regression was conducted 11 times using Ca-MMT-surf data, leaving one molecule out of the training set [**1-8, 11, 14, 15** (Table 9)] each time (LOO method). Further testing was achieved by calculating the ΔG_{LIE} of several compounds which were not in the training set [**9, 10, 12, 13, 16** (Table 9)]. All LIE coefficients were allowed to optimize freely. LIE parameters derived using Ca-MMT was then used in Na-MMT-surf with an additional optimization of the δ parameter (fixing α and β) 10 times, leaving one molecule out of the training set [**1-5,11-13,15-16** (Table 10)] each time. Optimization was carried in such a way that mean squared error (MSE) between experimental free energy and the LIE predicted free energy was minimized. Further testing was achieved by calculating the ΔG_{LIE} of several compounds which were not in the training set [**10 and 17** (Table 10)].

6.4. Results and Discussion

Results obtained from the LIE equations derived in chapters five and six exhibited an exceptional agreement with experimental free energy values across training and test sets. Furthermore, LIE parameters were transferable with an additional optimization of δ. However, when LIE method is applied based on interaction energies obtained from the slab models, such an agreement was not observed (Tables 8 and 9). The calculations yielded a modest overall mean unsigned error of 0.64 kcal mol^{-1} with a R^2 value of 0.5677 for Ca-MMT-surf and a MUE of 1.84 $kcalmol^{-1}$ and a R^2 value of 0.6376 for Na-MMT-surf. Compound 4-aminobenzoic was recognized as an outlier in Ca-MMT-surf and Na-MMT-surf; thus removed from the data set. After doing so

a MUE of 0.29 kcal mol^{-1} was observed with a R^2 value of 0.6736. The statistical results for all the LIE models are summarized in Table 11.

Possible underlying causes for the poor correlation of free energies of sorption obtained using the LIE method (using the slab models) to that of experimental free energy values obtained from adsorption isotherms was investigated. By comparing the trajectories of the organic cation in the interlayer model (organic cation between two slabs of clay) to that of surface model (organic cation interacting with the surface of a single slab of clay), it was observed that in the surface model for compounds having a phenyl group (i.e. anilines and benzylamines), the plane of the phenyl ring was no longer oriented parallelly (or near parallelly) to that of the plane of the clay surface (Figure 29) but it switched between parallel and perpendicular orientations throughout the duration of the trajectory of the production phase of the MD simulations. In our previous work, which was carried out using the interlayer model, we observed that (by analyzing the trajectories and the ensemble averaged orientations of the organic cations) the phenyl ring oriented parallelly or near parallelly to the clay surface thus assenting the experimental observations which specified that at lower surface concentrations of aromatic compounds (e.g. anilinium, nitrobenzene, methylene blue) the phenyl ring was (spectroscopically determined) oriented in such a way that it was parallel (or near parallel) to the plane of the silicate surface.[195, 196] Furthermore, a perpendicular orientation of the phenyl ring was observed (experimentally) only at high surface concentrations of the solute.[195, 196] In the interlayer model, since the organic cations were placed in a confined environment, spurious orientations which were observed in the surface mode (unconfined environment) was restricted. Thus, generating an accurate ensemble of configurations of the organic cations on the clay surface which were in agreement with the experimental results. Zwitterion 4-ABA (10) showed repulsive interactions as it drifted away to the solution phase from the clay surface.

Table 9. Calculated Free Energies for the Ca-MMT via the LIE method and Their Comparison with Experimental Values for the Training (**1-8,11,14,15**) and Test (**9,10,12,13,16**) Sets.

	Compound	$\Delta\langle E_{vdW}\rangle$	$\Delta\langle E_{El}\rangle$	[b]ΔG_{LIE}	[a]ΔG_{EXP}
1	Ammonium	-0.28 ± 1.03	-7.07 ± 2.90	-1.44 ± 0.30	-1.61
2	Aniline	-1.72 ± 0.91	-2.72 ± 2.64	-1.97 ± 0.18	-1.84
3	N-methylaniline	-2.56 ± 0.54	-1.52 ± 1.08	-2.48 ± 0.19	-2.30
4	N,N-dimethylaniline	-2.96 ± 0.37	-3.77 ± 1.23	-3.13 ± 0.17	-2.99
5	phenyltrimethylammonium	-2.92 ± 0.57	-4.45 ± 1.31	-3.20 ± 0.20	-3.54
6	p-toluidine	-2.43 ± 0.82	-3.18 ± 2.11	-2.62 ± 0.49	-1.72
7	N-methyl-p-toluidine	-2.34 ± 0.85	-2.99 ± 1.25	-2.51 ± 0.43	-2.57
8	N,N-dimethyl-ammonium	-3.24 ± 0.57	-2.89 ± 2.31	-3.23 ± 0.24	-3.54
9	[c]DMTS (charge 0)	-1.33 ± 1.12	-1.27 ± 3.57	-1.45 ± 0.11	-0.41
10	[c, d]4-(aminomethyl) benzoic acid (charge 0)	0.92 ± 1.22	6.14 ± 2.11	1.45 ± 0.68	-0.96
11	Benzylamine	-1.94 ± 0.73	-3.13 ± 1.95	-2.21 ± 0.96	-2.12
12	2,4-dicholorobenzylamine	-3.37 ± 0.70	-2.05 ± 1.76	-3.21 ± 0.07	-2.28
13	2,4-dimethylbenzylamine	-2.88 ± 1.01	-1.80 ± 0.84	-2.78 ± 0.20	-2.23
14	p-phenylenediamine	-2.05 ± 0.45	-1.70 ± 1.80	-2.10 ± 0.12	-2.77
15	1-napthylmethylbenzylamine	-2.90 ± 0.78	-3.47 ± 2.20	-3.04 ± 0.30	-3.72
16	[c]Oxytetracycline (charge 0)	-6.89 ± 0.90	10.39 ± 2.96	-4.26 ± 1.01	-3.96

[a]Experimental and calculated LIE binding free energies (kcal/mol) are shown. Compound structures are shown in figure 24. [b]Training set mean squared error (MSE)=0.19 kcal/mol, and test set MSE = 0.58 kcal/mol excluding the outliers. [c]Indicates a zwitterion. [d]indicates an outlier.

Table 10. Calculated Free Energies for the Na-MMT via the LIE method after optimization of δ from Ca-MMT keeping α and β fixed, for the Training (**1-5, 11-13, 15-16**) and Test (**10 and 17**) Sets.

Compound		$\Delta\langle E_{vdW}\rangle$	$\Delta\langle E_{El}\rangle$	[b]ΔG_{LIE}	[a]ΔG_{EXP}
1	Ammonium	-0.15 ± 0.75	-6.22 ± 1.95	-1.00 ± 0.30	-1.94
2	Aniline	-2.15 ± 0.67	-1.17 ± 2.36	-1.89 ± 0.16	-2.01
3	N-methylaniline	-2.50 ± 0.56	-2.27 ± 1.58	-2.33 ± 0.19	-2.41
4	N,N-dimethylaniline	-3.03 ± 0.54	-2.47 ± 1.28	-2.79 ± 0.17	-3.06
5	phenyltrimethylammonium	-3.04 ± 0.70	-3.69 ± 1.40	-2.97 ± 0.24	-3.78
11	Benzylamine	-2.03 ± 0.77	-1.96 ± 1.45	-1.90 ± 0.07	-2.45
12	2,4-dicholorobenzylamine	-4.97 ± 0.59	-2.17 ± 1.31	-4.31 ± 0.08	-2.51
13	2,4-dimethylbenzylamine	-2.47 ± 0.45	-6.79 ± 1.14	-2.96 ± 0.17	-2.52
15	1-Napthylbenzylamine	-3.65 ± 0.50	-2.39 ± 1.83	-3.28 ± 0.09	-3.80
16	[c]Oxytetracycaline (charge 0)	-8.85 ± 0.63	16.28 ± 2.96	-4.78 ± 0.36	-3.64
10	[c,d]4-Aminobenzoic acid (charge 0)	1.71 ± 1.08	5.39 ± 2.55	2.19 ± 0.40	-1.37
17	Metoprolol	-6.19 ± 1.06	-3.54 ± 3.72	-5.50 ± 1.50	-3.93

[a]Experimental and calculated LIE binding free energies (kcal/mol) are shown. Compound structures are shown in figure 24. [b]Training set mean squared error (MSE)=0.69 kcal/mol, and test set MSE = 2.46 kcal/mol. [c]Indicates a zwitterion. [d]indicates an outlier.

In the surface model, since the organic cations were placed in an unconfined environment, facilitating the free rotation of the organic cations around the center of mass of the molecule, thus generating an ensemble of configurations which were different to that of actual orientations (experimental) of the organic solutes having low surface concentration. Hence interaction energies computed from the surface model will be incorrect, propagating the errors to the LIE method.

Table 11. Optimized LIE Coefficients for Ca-MMT-surf and Na-MMT-surf

System	α	β	δ / kcal mol^{-1}
Ca-MMT	0.809 ± 0.095	0.145 ± 0.058	-0.189 ± 0.341
Na-MMT	0.809 ± 0.095	0.145 ± 0.058	0.021 ± 0.030

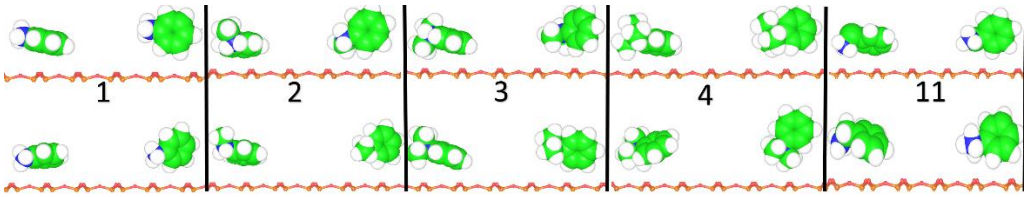


Figure 29. Equilibrium snapshots representing parallel and perpendicular orientations of organic compounds relative to the MMT surface in Ca-MMT-surf (top row) and Na-MMT-surf (bottom row).

6.5. Summary and Conclusions

The adsorption of organic cations on to a montmorillonite surface was simulated using MD simulations. Using the interaction energies generated from the production phase, LIE method was applied to calculate sorption free energies using a similar protocol used in chapters four and five. A modest agreement between experimental and LIE predicted free energy values was obtained. By analyzing the trajectories from the simulations, it was concluded that, such a poor agreement was obtained due to configurations sampled during the MD simulations due to non-confined environment that existed on the mineral-solution interfaces. It was apparent that the phenyl ring of the aromatic ammines no longer oriented parallelly but switched between perpendicular and parallel orientations which was conflicting with the experimental observations which stated that

under low surface concentration of the aromatic compounds the aromatic ring held parallel to the clay surface.

7. Concluding Remarks

Work presented in this thesis demonstrates the ability of the LIE method to predict the sorption free energies of different organic cations to homoionic smectite montmorillonite. In chapter four, we applied the LIE method for the prediction of organic cations sorption affinities to a representative clay mineral: Ca-montmorillonite. Molecular dynamics (MD) simulations were utilized to sample interactions energies of the organic cations with their environment (in solution and clay) and linear regression models were used to derive LIE coefficients (α, β and δ) fitting to reproduce experimental free energies of sorption using a training set having 10 compounds. It was found that the method was capable of predicting accurate free energies matching well with the experimental results; with a mean average error of 0.3 kcal mol^{-1} and a correlation coefficient of 0.84. LIE predicted free energy values were within 0.50 kcal mol^{-1} of experimental sorption free energies. From the average trajectories of each aromatic amine studied, it was observed that these compounds oriented in the interlayer in such a way that the phenyl ring always oriented parallelly or near parallelly to the clay surface. Such an orientation has been well documented in previous experimental work carried out validating the simulation models used. The direct inclusion of interaction energies to expand the free energies in the LIE equation enabled to capture variety of structural effects that mediated the sorption including defocusing of the positive charge on N-substituted ammines as a function of increased methyl groups which enhanced sorption through equal contributions from electrostatic and van der Waals interaction energies. Electrostatic interaction energies of zwitterionic compounds unveiled repulsive interactions with the clay surface thus lowering the free energies of sorption due to presence of negatively charged functional groups. The major driving force of sorption for zwitterionic compounds was the van der Waals interactions with the clay surface.

In chapter five, we investigated the accuracy and the transferability of LIE method and its parameters in predicting sorption affinities of an extended set of organic cations sorbing to homoionic smectite montmorillonite having calcium and sodium as charge compensating ions separately. Work carried out in this chapter differ from the previous chapter (chapter four) as the bound (in clay) and unbound (in solution) states were modeled to have a neutral charge. LIE was capable of predicting the sorption affinities with high accuracy with a mean square error of 0.17 kcal mol^{-1} for Ca-MMT and 0.08 kcal mol^{-1} for Na-MMT. It was concluded that LIE parameters derived with Ca-MMT data were transferable with an additional optimization of the parameter δ.

In chapter six, we modeled the sorption of organic cations to the mineral-solution interface by means of a model consisting a slab of montmorillonite in bulk solution. LIE method was applied to replicate the experimental free energies of sorption having different charge compensating ions: calcium and sodium. In contrast to results obtained from clay models used in chapters five and six, where organic cation was placed in between two slabs of clay (intercalated), we obtained moderate accuracy with 0.29 kcal mol^{-1} for the model with calcium ions and 0.94 kcal mol^{-1} for the model with sodium ions. It was concluded that such poor agreement was perceived due to sampling of configurations which were unrealistic at low surface concentrations which promoted by unconfined nature of the slab model.

In summary, we reason that LIE method can be used as a robust and accurate tool to predict sorption affinities of charged organic compounds sorbing through cation exchange with high accuracy. This is a major improvement over existing predictive models which underestimates binding affinities due to exclusion of electrostatic interactions. However to increase the computational efficiency, use of implicit solvation models must be considered in solution phase.

A bigger compound library having diverse functional groups should be employed to further evaluate the robustness of the LIE method.

Bibliography

1. Petrović, M.; Gonzalez, S.; Barceló, D., Analysis and removal of emerging contaminants in wastewater and drinking water. *TrAC Trends in Analytical Chemistry* **2003,** *22*, (10), 685-696.

2. Bolong, N.; Ismail, A.; Salim, M. R.; Matsuura, T., A review of the effects of emerging contaminants in wastewater and options for their removal. *Desalination* **2009,** *239*, (1), 229-246.

3. Pruden, A.; Pei, R.; Storteboom, H.; Carlson, K. H., Antibiotic resistance genes as emerging contaminants: studies in northern Colorado. *Environmental Science & Technology* **2006,** *40*, (23), 7445-7450.

4. Richardson, S. D.; Ternes, T. A., Water analysis: emerging contaminants and current issues. *Analytical chemistry* **2011,** *83*, (12), 4614-4648.

5. Brausch, J. M.; Rand, G. M., A review of personal care products in the aquatic environment: environmental concentrations and toxicity. *Chemosphere* **2011,** *82*, (11), 1518-1532.

6. Chen, Z.-F.; Ying, G.-G., Occurrence, fate and ecological risk of five typical azole fungicides as therapeutic and personal care products in the environment: A review. *Environment International* **2015,** *84*, 142-153.

7. Richardson, S. D.; Ternes, T. A., Water analysis: emerging contaminants and current issues. *Analytical chemistry* **2014,** *86*, (6), 2813-2848.

8. Yang, X.; Flowers, R. C.; Weinberg, H. S.; Singer, P. C., Occurrence and removal of pharmaceuticals and personal care products (PPCPs) in an advanced wastewater reclamation plant. *Water Research* **2011,** *45*, (16), 5218-5228.

9. Evgenidou, E. N.; Konstantinou, I. K.; Lambropoulou, D. A., Occurrence and removal of transformation products of PPCPs and illicit drugs in wastewaters: a review. *Science of the Total Environment* **2015,** *505*, 905-926.

10. Boxall, A.; Rudd, M. A.; Brooks, B. W.; Caldwell, D. J.; Choi, K.; Hickmann, S.; Innes, E.; Ostapyk, K.; Staveley, J. P.; Verslycke, T., Pharmaceuticals and personal care products in the environment: what are the big questions? *Environmental Health Perspectives* **2012,** *120*, (9), 1221-1229.

11. Onesios, K. M.; Jim, T. Y.; Bouwer, E. J., Biodegradation and removal of pharmaceuticals and personal care products in treatment systems: a review. *Biodegradation* **2009,** *20*, (4), 441-466.

12. Bowers, N.; Grant, K.; Xiong, A.; Gayfield, B., Emerging Pharmaceutical and Personal Care Products In Our Water System. **2015**.

13. Packer, J. L.; Werner, J. J.; Latch, D. E.; McNeill, K.; Arnold, W. A., Photochemical fate of pharmaceuticals in the environment: Naproxen, diclofenac, clofibric acid, and ibuprofen. *Aquatic Sciences* **2003,** *65*, (4), 342-351.

14. Herbold, B. A.; Brendler-Schwaab, S. Y.; Ahr, H. J., Ciprofloxacin: in vivo genotoxicity studies. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis* **2001,** *498*, (1), 193-205.

15. Ying, G.-G.; Kookana, R. S.; Ru, Y.-J., Occurrence and fate of hormone steroids in the environment. *Environment international* **2002,** *28*, (6), 545-551.

16. Kolpin, D. W.; Furlong, E. T.; Meyer, M. T.; Thurman, E. M.; Zaugg, S. D.; Barber, L. B.; Buxton, H. T., Pharmaceuticals, hormones, and other organic wastewater contaminants in US streams, 1999-2000: A national reconnaissance. *Environmental science & technology* **2002,** *36*, (6), 1202-1211.

17. Zhang, Q. Q.; Zhao, J.-L.; Ying, G.-G.; Liu, Y.-S.; Pan, C.-G., Emission estimation and multimedia fate modeling of seven steroids at the river basin scale in China. *Environmental science & technology* **2014,** *48*, (14), 7982-7992.

18. Oury, F.; Sumara, G.; Sumara, O.; Ferron, M.; Chang, H.; Smith, C. E.; Hermo, L.; Suarez, S.; Roth, B. L.; Ducy, P., Endocrine regulation of male fertility by the skeleton. *Cell* **2011,** *144*, (5), 796-809.

19. Wit, J.; Camacho-Hübner, C., Endocrine regulation of longitudinal bone growth. **2011**.

20. Won, E. T.; Borski, R. J., Endocrine regulation of compensatory growth in fish. *Frontiers in endocrinology* **2013,** *4*.

21. Zeitoun, M.; Alsoqeer, A.-R., Detection of Sex Steroid Hormones in Alfalfa and Some Rangeland Native Species in Saudi Arabia and Their Subsequent Effects on Camel Reproduction. **2014**.

22. Xu, Q.; Cao, M.-l.; Ma, W.-h.; SUN, D.-f.; WANG, W., Synthesis and properties of dissymmetric gemini quaternary ammonium salts cationic surfactants. *CHINA SURFACTANT DETERGENT AND COSMETICS* **2004,** *34*, (5; ISSU 201), 280-282.

23. Mishra, S.; Tyagi, V., Ester quats: the novel class of cationic fabric softeners. *Journal of oleo science* **2007,** *56*, (6), 269-276.

24. Dery, M., Quaternary Ammonium Compounds. *Kirk-Othmer Encyclopedia of Chemical Technology* **2000**.

25. Garcıa, M.; Ribosa, I.; Guindulain, T.; Sanchez-Leal, J.; Vives-Rego, J., Fate and effect of monoalkyl quaternary ammonium surfactants in the aquatic environment. *Environmental Pollution* **2001,** *111*, (1), 169-175.

26. Kümmerer, K., Drugs in the environment: emission of drugs, diagnostic aids and disinfectants into wastewater by hospitals in relation to other sources–a review. *Chemosphere* **2001,** *45*, (6), 957-969.

27. Zhang, C.; Cui, F.; Zeng, G.-m.; Jiang, M.; Yang, Z.-z.; Yu, Z.-g.; Zhu, M.-y.; Shen, L.-q., Quaternary ammonium compounds (QACs): A review on occurrence, fate and toxicity in the environment. *Science of the Total Environment* **2015,** *518*, 352-362.

28. Buffet-Bataillon, S.; Tattevin, P.; Bonnaure-Mallet, M.; Jolivet-Gougeon, A., Emergence of resistance to antibacterial agents: the role of quaternary ammonium compounds—a critical review. *International journal of antimicrobial agents* **2012,** *39*, (5), 381-389.

29. Hegstad, K.; Langsrud, S.; Lunestad, B. T.; Scheie, A. A.; Sunde, M.; Yazdankhah, S. P., Does the wide use of quaternary ammonium compounds enhance the selection and spread of antimicrobial resistance and thus threaten our health? *Microbial Drug Resistance* **2010,** *16*, (2), 91-104.

30. Sidhu, M. S.; Sørum, H.; Holck, A., Resistance to quaternary ammonium compounds in food-related bacteria. *Microbial Drug Resistance* **2002,** *8*, (4), 393-399.

31. Brecher, E. M., *Licit and illicit drugs*. Little, Brown Boston: 1972.

32. Burns, L., World Drug Report 2013 By United Nations Office on Drugs and Crime New York: United Nations, 2013ISBN: 978-92-1-056168-6, 151 pp. Grey literature. *Drug and Alcohol Review* **2014,** *33*, (2), 216-216.

33. Saffer, H.; Chaloupka, F., The demand for illicit drugs. *Economic inquiry* **1999,** *37*, (3), 401-411.

34. Zuccato, E.; Castiglioni, S.; Bagnati, R.; Chiabrando, C.; Grassi, P.; Fanelli, R., Illicit drugs, a novel group of environmental contaminants. *Water Research* **2008,** *42*, (4), 961-968.

35. Gledhill-Hoyt, J.; Lee, H.; Strote, J.; Wechsler, H., Increased use of marijuana and other illicit drugs at US colleges in the 1990s: results of three national surveys. *Addiction* **2000,** *95*, (11), 1655-1667.

36. Kasprzyk-Hordern, B.; Dinsdale, R. M.; Guwy, A. J., The removal of pharmaceuticals, personal care products, endocrine disruptors and illicit drugs during wastewater treatment and its impact on the quality of receiving waters. *Water research* **2009,** *43*, (2), 363-380.

37. Graymore, M.; Stagnitti, F.; Allinson, G., Impacts of atrazine in aquatic ecosystems. *Environment international* **2001,** *26*, (7), 483-495.

38. Scheepmaker, J.; Vonk, J., Environmental risk limits for monolinuron. *Bilthoven, The Netherlands: National Institute for Public Health and the Environment (RIVM). RIVM Letter report* **2008**, (601716009), 21.

39. Miles, C.; Pfeuffer, R., Pesticides in canals of South Florida. *Archives of Environmental Contamination and Toxicology* **1997,** *32*, (4), 337-345.

40. Harman-Fetcho, J. A.; Hapeman, C. J.; McConnell, L. L.; Potter, T. L.; Rice, C. P.; Sadeghi, A. M.; Smith, R. D.; Bialek, K.; Sefton, K. A.; Schaffer, B. A., Pesticide occurrence in selected South Florida canals and Biscayne Bay during high agricultural activity. *Journal of agricultural and food chemistry* **2005,** *53*, (15), 6040-6048.

41. Organization, W. H., *Guidelines for drinking-water quality: recommendations*. World Health Organization: 2004; Vol. 1.

42. DeLorenzo, M. E.; Scott, G. I.; Ross, P. E., Toxicity of pesticides to aquatic microorganisms: a review. *Environmental Toxicology and Chemistry* **2001,** *20*, (1), 84-98.

43. Solomon, K. R.; Carr, J. A.; Du Preez, L. H.; Giesy, J. P.; Kendall, R. J.; Smith, E. E.; Van Der Kraak, G. J., Effects of atrazine on fish, amphibians, and aquatic reptiles: a critical review. *Critical reviews in toxicology* **2008,** *38*, (9), 721-772.

44. Hopenhayn-Rich, C.; Stump, M.; Browning, S. R., Regional assessment of atrazine exposure and incidence of breast and ovarian cancers in Kentucky. *Archives of Environmental Contamination and Toxicology* **2002,** *42*, (1), 127-136.

45. Kniewald, J.; Jakominić, M.; Tomljenović, A.; Šimić, B.; Romac, P.; Vranešić, Đ.; Kniewald, Z., Disorders of male rat reproductive tract under the influence of atrazine. *Journal of Applied Toxicology* **2000,** *20*, (1), 61-68.

46. Fan, W.; Yanase, T.; Morinaga, H.; Gondo, S.; Okabe, T.; Nomura, M.; Komatsu, T.; Morohashi, K.-l.; Hayes, T. B.; Takayanagi, R., Atrazine-lnduced aromatase expression is SF-1 dependent: Implications for endocrine disruption in wildlife and reproductive cancers in humans. *Environmental Health Perspectives* **2007**, 720-727.

47. Ettinger, S. J.; Feldman, E. C., *Textbook of veterinary internal medicine*. Elsevier Health Sciences: 2009.

48. Boxall, A.; Fogg, L.; Blackwell, P.; Blackwell, P.; Kay, P.; Pemberton, E.; Croxford, A., Veterinary medicines in the environment. In *Reviews of environmental contamination and toxicology*, Springer: 2004; pp 1-91.

49. Henderson, K. L.; Coats, J. R., *Veterinary pharmaceuticals in the environment*. American Chemical Society; Distributed by Oxford University Press: 2009.

50. Tolls, J., Sorption of veterinary pharmaceuticals in soils: a review. *Environmental science & technology* **2001,** *35*, (17), 3397-3406.

51. Teuber, M., Veterinary use and antibiotic resistance. *Current opinion in microbiology* **2001,** *4*, (5), 493-499.

52. Levy, S. B.; Marshall, B., Antibacterial resistance worldwide: causes, challenges and responses. *Nature medicine* **2004,** *10*, S122-S129.

53. Eisenreich, S.; Baker, J.; Franz, T.; Swanson, M.; Rapport, R.; Strachan, W.; Hites, R., Fate of Pesticides and Chemicals in the Environment. In John Wiley and Sons: New York: 1992.

54. Schwarzenbach, R. P.; Gschwend, P. M.; Imboden, D. M., *Environmental organic chemistry*. John Wiley & Sons: 2005.

55. Manivanan, R., *Recycling of Industrial effluents*. New India Publishing: 2006.

56. Calderón-Preciado, D.; Matamoros, V.; Bayona, J. M., Occurrence and potential crop uptake of emerging contaminants and related compounds in an agricultural irrigation network. *Science of the total environment* **2011,** *412*, 14-19.

57. Chesworth, W., *Encyclopedia of soil science*. Springer Science & Business Media: 2008.

58. Mitchell, J. K.; Soga, K., *Fundamentals of soil behavior*. Wiley New York: 1976.

59. Dixon, J. B.; Weed, S. B., *Minerals in soil environments*. Soil Science Society of America Inc.(SSSA). 1989.

60. Tan, K. H., *Principles of soil chemistry*. CRC Press: 2010.

61. Yerima, B. P.; Van Ranst, E., *Introduction to soil science: soils of the tropics*. Trafford Publishing: 2005.

62. Yong, R. N.; Mohamed, A.-M. O.; Warkentin, B. P., *Principles of contaminant transport in soils*. Elsevier Science Publishers: 1992.

63. McDonough, W. F.; Sun, S.-S., The composition of the Earth. *Chemical geology* **1995,** *120*, (3), 223-253.

64. Dominguez, E. A., A Clay Odyssey.

65. Bergaya, F.; Lagaly, G., *Handbook of clay science*. Newnes: 2013; Vol. 5.

66. Cairns-Smith, A. G.; Hartman, H., *Clay minerals and the origin of life*. CUP Archive: 1986.

67. Moore, D. M.; Reynolds, R. C., *X-ray Diffraction and the Identification and Analysis of Clay Minerals*. Oxford university press Oxford: 1989; Vol. 378.

68. Saikia, N.; Bharali, D.; Sengupta, P.; Bordoloi, D.; Goswamee, R.; Saikia, P.; Borthakur, P., Characterization, beneficiation and utilization of a kaolinite clay from Assam, India. *Applied clay science* **2003,** *24*, (1), 93-103.

69. Odom, I., Smectite clay minerals: properties and uses. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* **1984,** *311*, (1517), 391-409.

70. Thomas, J.; Bohor, B. F., Surface area of montmorillonite from the dynamic sorption of nitrogen and carbon dioxide. *Clays and Clay Minerals* **1968,** *16*, (1), 83-91.

71. Low, P. F., Structural component of the swelling pressure of clays. *Langmuir* **1987,** *3*, (1), 18-25.

72. Boyd, S. A.; Mortland, M. M.; Chiou, C. T., Sorption characteristics of organic compounds on hexadecyltrimethylammonium-smectite. *Soil Science Society of America Journal* **1988,** *52*, (3), 652-657.

73. Droge, S. T.; Goss, K.-U., Sorption of organic cations to phyllosilicate clay minerals: CEC-normalization, salt dependency, and the role of electrostatic and hydrophobic effects. *Environmental science & technology* **2013,** *47*, (24), 14224-14232.

74. Brindley, G., Chlorite minerals. *The X-ray identification and crystal structures of clay minerals* **1961,** *6*, 242-296.

75. Fawell, J. K.; Hunt, S., *Environmental toxicology: organic pollutants*. E. Horwood: 1988.

76. Perriello, F. A., Bioreactor for remediation of pollutants with butane utilizing bacteria. In Google Patents: 2000.

77. Suresh, S., Reductive remediation of pollutants using metals. *Open Waste Management Journal* **2009,** *2*, 6-16.

78. von der Ohe, P. C.; Dulio, V.; Slobodnik, J.; De Deckere, E.; Kühne, R.; Ebert, R.-U.; Ginebreda, A.; De Cooman, W.; Schüürmann, G.; Brack, W., A new risk assessment approach for the prioritization of 500 classical and emerging organic microcontaminants as potential river basin specific pollutants under the European Water Framework Directive. *Science of the Total Environment* **2011,** *409*, (11), 2064-2077.

79. Smith, K., *Environmental hazards: assessing risk and reducing disaster*. Routledge: 2013.

80. Kaveeshwar, R., Characterisation and evaluation of some inorganic pollutants in the environment. **2014**.

81. MacKay, A. A.; Vasudevan, D., Polyfunctional ionogenic compound sorption: challenges and new approaches to advance predictive models. *Environmental science & technology* **2012,** *46*, (17), 9209-9223.

82. Naidu, R.; Kookana, R. S.; Oliver, D.; Rogers, S.; McLaughlin, M., *Contaminants and the Soil Environment in the Australasia-Pacific Region: Proceedings of the First Australasia-Pacific Conference on Contaminants and Soil Environment in the Australasia-Pacific Region, Held in Adelaide, Australia, 18–23 February 1996*. Springer Science & Business Media: 2012.

83. Jardine, P.; McCarthy, J.; Weber, N., Mechanisms of dissolved organic carbon adsorption on soil. *Soil Science Society of America Journal* **1989,** *53*, (5), 1378-1385.

84. Kelley, W. P., Cation exchange in soils. *Cation Exchange in Soils.* **1948**.

85. Gerstl, Z., Estimation of organic chemical sorption by soils. *Journal of contaminant hydrology* **1990,** *6*, (4), 357-375.

86. Burkhard, L. P., Estimating dissolved organic carbon partition coefficients for nonionic organic chemicals. *Environmental Science & Technology* **2000,** *34*, (22), 4663-4668.

87. Razzaque, M. M.; Grathwohl, P., Predicting organic carbon–water partitioning of hydrophobic organic chemicals in soils and sediments based on water solubility. *Water research* **2008,** *42*, (14), 3775-3780.

88. Muñoz, J.; Felicísimo, Á. M.; Økland, R., Comparison of statistical methods commonly used in predictive modelling. *Journal of Vegetation Science* **2004,** *15*, (2), 285-292.

89. Nguyen, T. H.; Goss, K.-U.; Ball, W. P., Polyparameter linear free energy relationships for estimating the equilibrium partition of organic compounds between water and the natural organic matter in soils and sediments. *Environmental science & technology* **2005,** *39*, (4), 913-924.

90. Goss, K.-U.; Schwarzenbach, R. P., Linear free energy relationships used to evaluate equilibrium partitioning of organic compounds. *Environmental science & technology* **2001,** *35*, (1), 1-9.

91. Atkins, P.; De Paula, J., *Elements of physical chemistry*. Oxford University Press: 2013.

92. Bevan, J.; Boerio-Goates, J., Chemical Thermodynamics: Principles and Applications. *Chemical Thermodynamics: Principles and Applications* **2000**.

93. Tahir-Kheli, R., *General and Statistical Thermodynamics*. Springer Science & Business Media: 2011.

94. Chandler, D., Introduction to modern statistical mechanics. *Introduction to Modern Statistical Mechanics, by David Chandler, pp. 288. Foreword by David Chandler. Oxford University Press, Sep 1987. ISBN-10: 0195042778. ISBN-13: 9780195042771* **1987,** *1*.

95. Khokhlov, A. R., *Statistical physics of macromolecules*. Amer Inst of Physics: 1994.

96. Pierce, M. M.; Raman, C.; Nall, B. T., Isothermal titration calorimetry of protein–protein interactions. *Methods* **1999,** *19*, (2), 213-221.

97. Jørgensen, S. E.; Nielsen, S. N.; Mejer, H., Emergy, environ, exergy and ecological modelling. *Ecological modelling* **1995,** *77*, (2), 99-109.

98. Stone, J. E.; Hardy, D. J.; Ufimtsev, I. S.; Schulten, K., GPU-accelerated molecular modeling coming of age. *Journal of Molecular Graphics and Modelling* **2010,** *29*, (2), 116-125.

99. Leach, A. R., *Molecular modelling: principles and applications*. Pearson education: 2001.

100. Germann, T. C.; Lomdahl, P. S., Recent advances in large-scale atomistic materials simulations. *Computing in Science & Engineering* **1999,** *1*, (2), 10-11, 16.

101. Burkert, U.; Allinger, N. L., *Molecular mechanics*. American Chemical Society Washington, DC: 1982; Vol. 177.

102. Hoover, W. G. In *Molecular dynamics*, Molecular Dynamics, 1986; 1986.

103. Karplus, M.; Petsko, G. A., Molecular dynamics simulations in biology. *Nature* **1990,** *347*, (6294), 631-639.

104. Jorgensen, W. L., OPLS force fields. *Encyclopedia of computational chemistry* **1998**.

105. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and testing of a general amber force field. *Journal of computational chemistry* **2004,** *25*, (9), 1157-1174.

106. Mayo, S. L.; Olafson, B. D.; Goddard, W. A., DREIDING: a generic force field for molecular simulations. *Journal of Physical Chemistry* **1990,** *94*, (26), 8897-8909.

107. Jensen, F., *Introduction to computational chemistry*. John Wiley & Sons: 2013.

108. Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J., Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society* **1996,** *118*, (45), 11225-11236.

109. Kony, D.; Damm, W.; Stoll, S.; Van Gunsteren, W. F., An improved OPLS–AA force field for carbohydrates. *Journal of computational chemistry* **2002,** *23*, (15), 1416-1429.

110. Cygan, R. T.; Liang, J.-J.; Kalinichev, A. G., Molecular models of hydroxide, oxyhydroxide, and clay phases and the development of a general force field. *The Journal of Physical Chemistry B* **2004,** *108*, (4), 1255-1266.

111. Kirkpatrick, R.; Kalinichev, A.; Wang, J., Molecular dynamics modelling of hydrated mineral interlayers and surfaces: structure and dynamics. *Mineralogical Magazine* **2005,** *69*, (3), 289-308.

112. Cygan, R. T.; Romanov, V. N.; Myshakin, E. M., Natural materials for carbon capture. *Sandia report* **2010**.

113. Kalinichev, A. G.; Kirkpatrick, R. J., Molecular dynamics modeling of chloride binding to the surfaces of calcium hydroxide, hydrated calcium aluminate, and calcium silicate phases. *Chemistry of Materials* **2002,** *14*, (8), 3539-3549.

114. Padma Kumar, P.; Kalinichev, A. G.; Kirkpatrick, R. J., Hydration, swelling, interlayer structure, and hydrogen bonding in organolayered double hydroxides: Insights from molecular dynamics simulation of citrate-intercalated hydrotalcite. *The Journal of Physical Chemistry B* **2006,** *110*, (9), 3841-3844.

115. Mark, P.; Nilsson, L., Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. *The Journal of Physical Chemistry A* **2001,** *105*, (43), 9954-9960.

116. Lindström, U. M., Stereoselective organic reactions in water. *Chemical Reviews* **2002,** *102*, (8), 2751-2772.

117. Zaks, A.; Klibanov, A. M., The effect of water on enzyme action in organic media. *Journal of Biological Chemistry* **1988,** *263*, (17), 8017-8021.

118. Soper, A.; Phillips, M., A new determination of the structure of water at 25 C. *Chemical Physics* **1986,** *107*, (1), 47-60.

119. Jeffrey, G. A.; Jeffrey, G. A., *An introduction to hydrogen bonding*. Oxford university press New York: 1997; Vol. 12.

120. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics* **1983,** *79*, (2), 926-935.

121. Toukan, K.; Rahman, A., Molecular-dynamics study of atomic motions in water. *Physical Review B* **1985,** *31*, (5), 2643.

122. Berendsen, H. J.; Postma, J. P.; van Gunsteren, W. F.; Hermans, J., Interaction models for water in relation to protein hydration. In *Intermolecular forces*, Springer: 1981; pp 331-342.

123. Wu, Y.; Tepper, H. L.; Voth, G. A., Flexible simple point-charge water model with improved liquid-state properties. *The Journal of chemical physics* **2006,** *124*, (2), 024503.

124. Habershon, S.; Markland, T. E.; Manolopoulos, D. E., Competing quantum effects in the dynamics of a flexible water model. *The journal of chemical physics* **2009,** *131*, (2), 024501.

125. Wang, J.; Kalinichev, A. G.; Kirkpatrick, R. J., Effects of substrate structure and composition on the structure, dynamics, and energetics of water at mineral surfaces: A molecular dynamics modeling study. *Geochimica et cosmochimica acta* **2006,** *70*, (3), 562-582.

126. Haile, J., *Molecular dynamics simulation*. Wiley, New York: 1992; Vol. 18.

127. Rapaport, D. C., *The art of molecular dynamics simulation*. Cambridge university press: 2004.

128. Karplus, M.; Kuriyan, J., Molecular dynamics and protein function. *Proceedings of the National Academy of Sciences of the United States of America* **2005,** *102*, (19), 6679-6685.

129. Yamakov, V.; Wolf, D.; Phillpot, S. R.; Mukherjee, A. K.; Gleiter, H., Dislocation processes in the deformation of nanocrystalline aluminium by molecular-dynamics simulation. *Nature materials* **2002,** *1*, (1), 45-49.

130. Alder, B. J.; Wainwright, T., Studies in molecular dynamics. I. General method. *The Journal of Chemical Physics* **1959,** *31*, (2), 459-466.

131. Verlet, L., Computer" experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Physical review* **1967,** *159*, (1), 98.

132. Berendsen, H. J.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J., Molecular dynamics with coupling to an external bath. *The Journal of chemical physics* **1984,** *81*, (8), 3684-3690.

133. Lemak, A.; Balabaev, N., On the Berendsen thermostat. *Molecular Simulation* **1994,** *13*, (3), 177-187.

134. Andersen, H. C., Molecular dynamics simulations at constant pressure and/or temperature. *The Journal of chemical physics* **1980,** *72*, (4), 2384-2393.

135. Hoover, W. G., Canonical dynamics: equilibrium phase-space distributions. *Physical Review A* **1985,** *31*, (3), 1695.

136. Parrinello, M.; Rahman, A., Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied physics* **1981,** *52*, (12), 7182-7190.

137. Allen, M. P.; Tildesley, D. J., *Computer simulation of liquids*. Oxford university press: 1989.

138. Schreiber, H.; Steinhauser, O., Cutoff size does strongly influence molecular dynamics results on solvated polypeptides. *Biochemistry* **1992,** *31*, (25), 5856-5860.

139. Davis, M. E.; McCammon, J. A., Electrostatics in biomolecular structure and dynamics. *Chemical Reviews* **1990,** *90*, (3), 509-521.

140. Ewald, P. P., Ewald summation. *Ann. Phys.* **1921,** *64*, 253-371.

141. Smith, W.; Forester, T.; Todorov, I.; Cheshire, U., The DL POLY Classic user manual. *STFC, STFC Daresbury Laboratory, Daresbury, Warrington, Cheshire, WA4 4AD, United Kingdom, version* **2010,** *1*.

142. Harp, G.; Berne, B., Time-correlation functions, memory functions, and molecular dynamics. *Physical Review A* **1970,** *2*, (3), 975.

143. Szabo, A.; Ostlund, N. S., *Modern quantum chemistry: introduction to advanced electronic structure theory*. Courier Corporation: 2012.

144. Born, M., Born-Oppenheimer Approximation. *Quantum* **1927,** *2*, (2014/12), 4.

145. Himo, F.; Lovell, T.; Hilgraf, R.; Rostovtsev, V. V.; Noodleman, L.; Sharpless, K. B.; Fokin, V. V., Copper (I)-catalyzed synthesis of azoles. DFT study predicts unprecedented reactivity and intermediates. *Journal of the American Chemical Society* **2005,** *127*, (1), 210-216.

146. Carey, F. A.; Sundberg, R. J., *Advanced Organic Chemistry: Part A: Structure and Mechanisms*. Springer Science & Business Media: 2007.

147. Kresse, G.; Hafner, J., Ab initio molecular-dynamics simulation of the liquid-metal–amorphous-semiconductor transition in germanium. *Physical Review B* **1994,** *49*, (20), 14251.

148. Hohenberg, P.; Khon, W., Density Functional Theory. *Physical Review* **1964,** *136*, B864-B871.

149. Kohn, W.; Sham, L. J., Self-consistent equations including exchange and correlation effects. *Physical Review* **1965,** *140*, (4A), A1133.

150. Bachrach, S. M., *Computational organic chemistry*. John Wiley & Sons: 2007.

151. Singh, U. C.; Kollman, P. A., An approach to computing electrostatic charges for molecules. *Journal of Computational Chemistry* **1984,** *5*, (2), 129-145.

152. Cho, S. J., Calculation and application of partial charges. *Journal of the Chosun Natural Science* **2010,** *3*, (4), 226-230.

153. Breneman, C. M.; Wiberg, K. B., Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis. *Journal of Computational Chemistry* **1990,** *11*, (3), 361-373.

154. Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A., A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *The Journal of Physical Chemistry* **1993,** *97*, (40), 10269-10280.

155. Kessel, A.; Ben-Tal, N., *Introduction to proteins: structure, function, and motion*. CRC Press: 2010.

156. Creighton, T. E., *Proteins: structures and molecular properties*. Macmillan: 1993.

157. Mannhold, R.; Kubinyi, H.; Folkers, G.; Böhm, H.-J.; Schneider, G., *Protein-ligand interactions: from molecular recognition to drug design*. John Wiley & Sons: 2006; Vol. 19.

158. Berg, J. M.; Tymoczko, J. L.; Stryer, L., Biochemistry. *475–477](WH Freeman and Co., New York, 2011)* **2002**.

159. Drews, J., Drug discovery: a historical perspective. *Science* **2000,** *287*, (5460), 1960-1964.

160. Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J., Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced drug delivery reviews* **2012,** *64*, 4-17.

161. Zwanzig, R. W., High-temperature equation of state by a perturbation method. I. nonpolar gases. *The Journal of Chemical Physics* **1954,** *22*, (8), 1420-1426.

162. Goette, M.; Grubmüller, H., Accuracy and convergence of free energy differences calculated from nonequilibrium switching processes. *Journal of computational chemistry* **2009,** *30*, (3), 447-456.

163. Wayner, D. D.; Parker, V. D., Bond energies in solution from electrode potentials and thermochemical cycles. A simplified and general approach. *Accounts of chemical research* **1993,** *26*, (5), 287-294.

164. Lyubartsev, A.; Martsinovski, A.; Shevkunov, S.; Vorontsov-Velyaminov, P., New approach to Monte Carlo calculation of the free energy: Method of expanded ensembles. *The Journal of chemical physics* **1992,** *96*, (3), 1776-1783.

165. Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P., AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Computer Physics Communications* **1995,** *91*, (1), 1-41.

166. Kirkwood, J. G., Statistical mechanics of fluid mixtures. *The Journal of Chemical Physics* **1935,** *3*, (5), 300-313.

167. Mitchell, J. B.; Laskowski, R. A.; Alex, A.; Forster, M. J.; Thornton, J. M., BLEEP—potential of mean force describing protein–ligand interactions: II. Calculation of binding energies and comparison with experimental data. *Journal of computational chemistry* **1999,** *20*, (11), 1177-1185.

168. Krumrine, J.; Raubacher, F.; Brooijmans, N.; Kuntz, I., Principles and methods of docking and ligand design. *Structural Bioinformatics, Volume 44* **2005**, 441-476.

169. Bissantz, C.; Folkers, G.; Rognan, D., Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. *Journal of medicinal chemistry* **2000,** *43*, (25), 4759-4767.

170. Gohlke, H.; Hendlich, M.; Klebe, G., Knowledge-based scoring function to predict protein-ligand interactions. *Journal of molecular biology* **2000,** *295*, (2), 337-356.

171. Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P., Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *Journal of computer-aided molecular design* **1997,** *11*, (5), 425-445.

172. Åqvist, J.; Medina, C.; Samuelsson, J.-E., A new method for predicting binding affinity in computer-aided drug design. *Protein engineering* **1994,** *7*, (3), 385-391.

173. Hansson, T.; Marelius, J.; Åqvist, J., Ligand binding affinity prediction by linear interaction energy methods. *Journal of computer-aided molecular design* **1998,** *12*, (1), 27-35.

174. Ring, C. S.; Sun, E.; McKerrow, J. H.; Lee, G. K.; Rosenthal, P. J.; Kuntz, I. D.; Cohen, F. E., Structure-based inhibitor design by using protein models for the development of antiparasitic agents. *Proceedings of the national academy of Sciences* **1993,** *90*, (8), 3583-3587.

175. Miyamoto, S.; Kollman, P. A., Absolute and relative binding free energy calculations of the interaction of biotin and its analogs with streptavidin using molecular dynamics/free energy perturbation approaches. *PROTEINS: Structure, Function, and Genetics* **1993,** *16*, (3), 226-245.

176. Horton, N.; Lewis, M., Calculation of the free energy of association for protein complexes. *Protein Science* **1992,** *1*, (1), 169-181.

177. Marcus, R. A., Electron transfer reactions in chemistry. Theory and experiment. *Reviews of Modern Physics* **1993,** *65*, (3), 599.

178. Ben-Naim, A.; Marcus, Y., Solvation thermodynamics of nonionic solutes. *The Journal of chemical physics* **1984,** *81*, (4), 2016-2027.

179. Carlsson, J.; Boukharta, L.; Åqvist, J., Combining docking, molecular dynamics and the linear interaction energy method to predict binding modes and affinities for non-nucleoside inhibitors to HIV-1 reverse transcriptase. *Journal of medicinal chemistry* **2008,** *51*, (9), 2648-2656.

180. Hansson, T.; Åqvist, J., Estimation of binding free energies for HIV proteinase inhibitors by molecular dynamics simulations. *Protein Engineering* **1995,** *8*, (11), 1137-1144.

181. Mowbray, S. L., Sugar Recognition by a Glucose/Galactose Receptor EVALUATION OF BINDING ENERGETICS FROM MOLECULAR DYNAMICS SIMULATIONS. *Journal of Biological Chemistry* **1995,** *270*, (17), 9978-9981.

182. Åqvist, J.; Hansson, T., On the validity of electrostatic linear response in polar solvents. *The Journal of Physical Chemistry* **1996,** *100*, (22), 9512-9521.

183. Almlöf, M.; Brandsdal, B. O.; Åqvist, J., Binding affinity prediction with different force fields: examination of the linear interaction energy method. *Journal of computational chemistry* **2004,** *25*, (10), 1242-1254.

184. Carlson, H. A.; Jorgensen, W. L., An extended linear response method for determining free energies of hydration. *The Journal of Physical Chemistry* **1995,** *99*, (26), 10667-10673.

185. Mishra, S. K.; Sund, J.; Åqvist, J.; Koča, J., Computational prediction of monosaccharide binding free energies to lectins with linear interaction energy models. *Journal of computational chemistry* **2012,** *33*, (29), 2340-2350.

186. Worth, A.; Bassan, A.; De Bruijn, J.; Gallegos Saliner, A.; Netzeva, T.; Pavan, M.; Patlewicz, G.; Tsakovska, I.; Eisenreich, S., The role of the European Chemicals Bureau in promoting the regulatory use of (Q) SAR methods†. *SAR and QSAR in Environmental Research* **2007,** *18*, (1-2), 111-125.

187. Norrish, K., The swelling of montmorillonite. *Discussions of the Faraday society* **1954,** *18*, 120-134.

188. Zhou, Q.; Lu, X.; Liu, X.; Zhang, L.; He, H.; Zhu, J.; Yuan, P., Hydration of methane intercalated in Na-smectites with distinct layer charge: Insights from molecular simulations. *Journal of colloid and interface science* **2011,** *355*, (1), 237-242.

189. Loewenstein, W., The distribution of aluminum in the tetrahedra of silicates and aluminates. *American Mineralogist* **1954,** *39*, (1-2), 92-96.

190. Frisch, M.; Trucks, G.; Schlegel, H.; Scuseria, G.; Robb, M.; Cheeseman, J.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G., Gaussian 09, revision B. 01. *Gaussian Inc., Wallingford, CT* **2010**.

191. Becke, A. D., Density-functional thermochemistry. III. The role of exact exchange. *The Journal of chemical physics* **1993,** *98*, (7), 5648-5652.

192. Smith, E.; Todorov, I., DL_POLY Classic Molecular Simulation Package. In Daresbury Laboratory: Cheshire, UK: 2012.

193. Martyna, G. J.; Tobias, D. J.; Klein, M. L., Constant pressure molecular dynamics algorithms. *The Journal of Chemical Physics* **1994,** *101*, (5), 4177-4189.

194. Hall, G. L., Correction to Fuchs' calculation of the electrostatic energy of a Wigner solid. *Physical Review B* **1979,** *19*, (8), 3921.

195. Greene-Kelly, R., Sorption of aromatic organic compounds by montmorillonite. Part 1.—Orientation studies. *Transactions of the Faraday Society* **1955,** *51*, 412-424.

196. Yariv, S.; Heller, L.; Sofer, Z.; Bodenheimer, W., Sorption of aniline by montmorillonite. *Israel Journal of Chemistry* **1968,** *6*, (5), 741-756.

197. Vasudevan, D.; Arey, T. A.; Dickstein, D. R.; Newman, M. H.; Zhang, T. Y.; Kinnear, H. M.; Bader, M. M., Nonlinearity of cationic aromatic amine sorption to aluminosilicates and soils: role of intermolecular cation− π interactions. *Environmental science & technology* **2013,** *47*, (24), 14119-14127.

198. Samaraweera, M.; Jolin, W.; Vasudevan, D.; MacKay, A. A.; Gascón, J. A., Atomistic Prediction of Sorption Free Energies of Cationic Aromatic Amines on Montmorillonite: A Linear Interaction Energy Method. *Environmental Science & Technology Letters* **2014,** *1*, (6), 284-289.

199. Figueroa-Diva, R. A.; Vasudevan, D.; MacKay, A. A., Trends in soil sorption coefficients within common antimicrobial families. *Chemosphere* **2010,** *79*, (8), 786-793.

200. Wu, L.; Liao, L.; Lv, G., Influence of interlayer cations on organic intercalation of montmorillonite. *Journal of colloid and interface science* **2015,** *454*, 1-7.

201. Staunton, S.; Roubaud, M., Adsorption of 137Cs on montmorillonite and illite: Effect of charge compensating cation, ionic strength, concentration of Cs, K and fulvic acid. *Clays and clay minerals* **1997,** *45*, (2), 251-260.

202. Evangelou, V., Environmental soil and water chemistry. **1998**.

203. Karickhoff, S. W.; Brown, D. S.; Scott, T. A., Sorption of hydrophobic pollutants on natural sediments. *Water research* **1979,** *13*, (3), 241-248.

204. Violante, A., Elucidating mechanisms of competitive sorption at the mineral/water interface. *Advances in agronomy* **2013,** *118*, 111-176.

205. Siebecker, M.; Li, W.; Khalid, S.; Sparks, D., Real-time QEXAFS spectroscopy measures rapid precipitate formation at the mineral–water interface. *Nature communications* **2014,** *5*.

206. Teich-McGoldrick, S. L.; Greathouse, J. A.; Cygan, R. T., Molecular dynamics simulations of uranyl adsorption and structure on the basal surface of muscovite. *Molecular Simulation* **2014,** *40*, (7-9), 610-617.