View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Open Archive Toulouse Archive Ouverte

Book Title:

Puigjaner, Heyen (Eds.): Computer Aided Process & Product Engineering - CAPE

Title of the chapter

Molecular modelling for physical property prediction

Corresponding author:

Vincent GERBAUD
Laboratoire de Génie Chimique
BP 1301, 5 rue Paulin Talabot
31106 Toulouse Cedex 1
France
Vincent.Gerbaud@ensiacet.fr

Coauthor:

Xavier JOULIA
ENSIACET – Laboratoire de Génie Chimique
117 Route de Narbonne
31077 Toulouse Cedex 4
France
Xavier.Joulia@ensiacet.fr

3.2. Molecular modeling for physical property prediction

3.2.1. INTRODUCTION

Multiscale modeling is becoming the standard approach for process study in a broader framework that promotes computer aided integrated product and process design. In addition to usual purity requirements, end products must meet new constraints in terms of environmental impact, safety of goods and people, specific properties. Engineering achievements can be startling from the user side like aqueous solvent paint still washable after drying!

This can only be done by improving process knowledge and performance at all scales, down to the atomic one. Current experimental and modeling approaches assess with difficultly such sub-micronic scales. In experiments, how to conceive experimental devices small enough and introduce them in molecular systems without affecting irreversibly the phenomena that they look at? In modeling and simulation, which hypotheses are still relevant? how to handle boundary effects? Numerical difficulties may arise along with the necessity of defining new parameters… that will be adjustable ones as no experiments can obtain them. This latter statement is particularly true for energetic interaction parameters like binary interaction parameters in current liquide – vapor equilibrium macroscopic thermodynamic models based on activity coefficient approach or on equation of state approach. In all processes, study of phenomena attributed to energetic interactions has always been left over for a time … that has come:

Indeed, molecular modeling is a field of study that in interested in the behavior of atomic and molecular systems subject to energetic interactions. It is then a natural complement of experimental and modeling approaches to expand multiscale approaches towards smaller scales. Besides, process flows concern primarily molecules from raw materials to end products. Therefore, at any process development step, the challenge of knowing the physical properties and thermodynamical state of molecules is critical. But, the future of this challenge is dim when one thinks about the millions of chemical compounds referenced in the chemical abstract series. Neither experimental approaches nor current themodynamic models can handle the combination of properties needed. In some cases experiments are not even practical because of materials decomposition or safety issues. Universal group contribution methods are a pipe dream and existing ones are efficient but are restricted in use to specific areas like petrochemical and small molecular systems.

As providers of accurate physico-chemical data, molecular modeling methods offer an alternative to an intensive and expensive experimental campaign, once molecular models are available, which is becoming increasingly the case [CAS 04]. But this first goal is nothing compared to the main interest of molecular methods that is, probing the matter at the molecular scale [CHE 02 ; DEP 02, SAN 03]. Indeed, molecular modeling can be seen as a “third way to explore real matter” [ALL 87]. Like a theoretical approach, it is based on a model system of the real one. But unlike theory, no hypothesis, no transcription of key phenomena into equations or correlations is performed. Rather, molecular modeling performs a numerical experiments to simulate directly the behavior of the model system.

The concept of numerical experiment is strong. Fisrt, the model system is made of a boundered molecular system and of an interaction model analogous to an experimental sensor that enables to compute the internal energy of the model system. Second, think of the pseudo constant thermometer temperature and of the brownian motion of atoms in a liquid that generates a fluctuating temperature. More generally, any macroscopic property value measured by an experimental probe is a time average over many instantaneous fluctuating values. Statistical thermodynamic postulates that this time average equals an ensemble average over a statistically significant numbers of configurations of the model system. Molecular modeling generates them numerically using methods like molecular dynamics or Monte Carlo Methods. Any property of interest is then derived using thermodynamical laws from instantaneous property value averages and correlation factors. Thirdly, numerical standard deviation associated to the ensemble average is the equivalent of experimental accuracy.

This chapter presents molecular modeling concepts so as to demystify them and stress their interests for chemical engineers. Multiscale approach including molecular modeling are not illustrated due to restricted space. Rather, routine examples on the use of several molecular techniques suitable to get accurate vapor – liquid equilibrium data when no data is available are provided.

3.2.2. WHAT IS MOLECULAR MODELING?

Molecular modeling includes computer theoretical chemistry and molecular simulation.

Computer theoretical chemistry calculations are carried out at 0K and solve Schrödinger equation to obtain nuclear and electronic properties such as conformation, orbital, density of load, electrostatic potential surface in a fundamental or excited states. Computation time is huge, being proportional at best to $N_{electrons}^{2.5}$ which restricts its use to small systems. The precision of the results is significant because the only assumptions are linked to approximations carried out to solve the equation of Schrödinger. In particular, there are no adjustable parameters. Besides, it provides crucial information on the electronic distribution that enables to evaluate electrostatic interactions in molecular simulation.

Molecular simulation is a numerical technique to get the physicochemical properties of macroscopic systems from the description on an atomic scale of the elementary interactions and from the application of statistical thermodynamics principles. It concerns the calculation of a model system internal energy at a positive temperature. Computation time is proportional to $N_{molecules}$, which makes it a technique adapted to the study of real systems: phase properties, interfaces, reaction, transport phenomena.... Molecular simulation carries out a dynamic modeling of the system subjected to realistic temperature and pressure conditions thanks to an adequate sampling of the system configurations. A configuration is a set of particles coordinates and connections. Inaccuracy may arise from the energetic models that contain fitted but physically meaningful parameters or from system configurations sampling techniques that must comply with statistical thermodynamic principles.

Molecular simulation offers the most potentialities for process engineering. Wherever energetic interaction related phenomena have a prevalent place, molecular simulation deserves to be considered to study and look further into the knowledge of the phenomena in the heart of the processes. In particular, it is suitable for the study of phase equilibrium, interfacial properties (specific adsorption on catalyst…), transport coefficients, chemical reactivity, activity coefficients ….

3.2.2.1 Scientific challenges of molecular modeling in Process Engineering.

The use of molecular simulation in Process Engineering lies mainly in the difficulty of establishing the link between the macroscopic properties and their energetic description or that of significant parameters at mesoscopic or molecular scale. The micro – macro relation can be simple: in distillation the knowledge of phase equilibrium data enables to run an extensive study and design of the process. In tablet processing, the relation is more complex: the tablet properties (compactness, friability, dissolution) are related to the pellets cohesion and to the substrates solubility. Obviously energetic interaction is a key phenomena and is taken into account through solubility parameters which can be broken down into primarily energy contributions (Van der Waals repulsion - attraction, Coulombic interaction…), precisely the applicability of molecular simulation. But particle size and solvent effects on the aggregates size and homogeneity are equally important notwithstanding operational process parameters and are still difficult to address at a molecular scale. So, identifying the limiting phenomena is a priority before any molecular simulation.

Size of the model systems is not an unsolvable problem as periodic boundary conditions can be applied to replicate the original system box and mimic a homogeneous macroscopic phase. Rather, the scientific challenges concern issues often encountered in experiments, the sensor challenge, the sampling challenge and the multiscale challenge.

3.2.2.1.1 The sensor challenge

For data-oriented simulations, accurate force fields / sensors are needed to evaluate precisely energetic interactions. Study of highly polar systems, reliable and relevant extrapolation of carefully set force field parameters, absence of temperature dependency of these parameters are key improvements of molecular simulation models over existing macroscopic models.

The model system is usually a parallelepiped box filled with particles which energetic interactions are described by a force field enabling to compute the system internal energy. In order to mimic a homogeneous phase, the box is usually replicated in 3-D by applying periodic boundary conditions. Typical size ranges from 20 to 1000 Å and may vary during simulations. Edge effects are to be envisaged and can be attenuated by increase the box size.

The development of a force field requires a strong collaboration with theoretical chemists and physicists. Indeed, different kinds of force fields can arise: some based on quantum chemistry concepts, some based on molecular mechanics (Sandler, 2003).

Quantum-based models are used in static modeling and naturally in computer theoretical chemistry calculations. Solving the Schrödinger equation, they provide the nuclear and electronic properties system and consequently the true energy of the system (e.g. the energy of ionization) physically measurable. Molecular mechanics models are used in molecular simulation to calculate intensive properties (T, P) and extensive ones among which the internal energy of the system which is not measurable directly by an experiment but enables to calculate other thermodynamic properties by using thermodynamic laws. Properties like vaporization enthalpy connected to differences in internal energy are computed and can be compared to experiments.

Quantum (QM) models are practical on a few tens of atoms at best and more and more used in combination with molecular mechanics models for some part of the system where accurate electronic distribution are needed, e.g. a reactive zone or to provide a description of the electronic distribution.

Molecular mechanics (MM) models are the most used and are based on a springs and beads mechanistic description of the intermolecular interactions and of intramolecular bonds. They allow calculations on several hundreds of particles which enable to model real systems in a satisfactory way. They contain physical parameters evaluated from quantum calculation but also empirical parameters which must be regressed from experimental data. However, this empiricism is attenuated by some physical significance attributed to the parameters. Moreover, MM force fields show amazing properties: valid over a large pressure and temperature range, they can be used to compute many properties and all molecules can be described from a small set of parameters if careful parameterization is conducted, which constitutes the first challenge.

3.2.2.1.2 The sampling challenge

The second challenge requires a strong involvement of process engineers: novel and smart methods must be developed to sample specific states of the model system which are of great interest for Process Engineering: transition states that set the reaction energetic barrier, azeotropes that affect strongly the distillation process feasibility and design, dew points …. Usually, existing molecular simulation methods sample non specific states like a vapor – liquid equilibrium point. Unlike measurement time, its experimental equivalent, numerical sampling can be advantageously biased to sample the specific state of interest but it requires expertise to comply with statistical thermodynamics principles which enable to bridge the microscopic and the macroscopic scales.

Furthermore, for existing methods based on molecular dynamic or Monte Carlo methods, sampling efficiency should be improved in particular for complex molecules like macromolecules even if the alternate solution of running more simulations is still the leading choice as computer power increases.

With this second challenge, Process Engineering finds a new use for molecular modeling: it cannot be solely data-oriented but also discovery-oriented and assumes its status of numerical experiment.

3.2.2.1.3 Molecular modeling in a multiscale approach

The integration of molecular modeling in applicable models for the study of macroscopic systems and their properties is of the utmost importance for process engineering. Indeed, often considered as decisive, phenomena related to energetic interactions have often been left aside during a process study because of a lack of suitable tools or incorporated into parameters. Thermodynamic models used in phase equilibrium calculations are a good example: binary interaction parameters must be found empirically despite their solid physical meaning. The first illustrative example addresses the issue of calculating binary parameters by molecular modeling methods.

Process Engineering models are knowledge-based models. In most domains, process study requires a multiscale approach. As a technique of experimentation, molecular modeling makes it possible to visualize on a molecular scale physicochemical phenomena. It can thus be used to develop or revisit theories, models or parameters of models and therefore improve our knowledge of processes and increase the capacity of predictions and extrapolation of existing models.

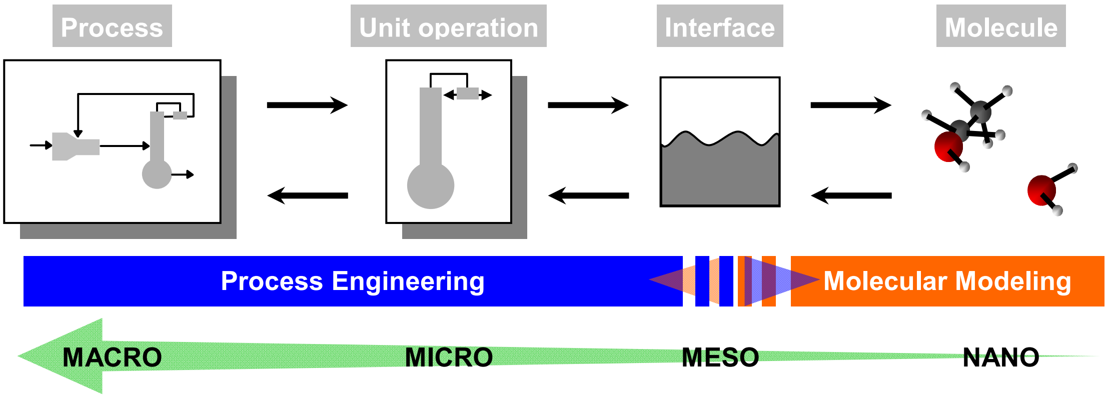


Figure 1. Process Engineering and Molecular Modeling.

3.2.3. STATISTICAL THERMODYNAMIC BACKGROUND

Suggested readings:

D.A. Mc Quarry, *Statistical thermodynamics*, Harper and Collins Publishers, New York, USA, 1976.

M.P. Allen and D.J. Tildesley, *Computer Simulation of Liquids*, Oxford University Publications, UK, 1987. [new edition in 2000].

D. Frenkel, B. Smit , *Understanding Molecular Simulation. From Algorithms to Applications*, Academic Press, San Diego, USA, 1996. [new edition in 2002].

3.2.3.1 A microscopic description of macroscopic properties

Traditional thermodynamics and statistical thermodynamics are interested in the same problems but differ in their approach: thermodynamics provides general relations without any consideration of the intrinsic constitution of the matter while statistical thermodynamics supposes the existence of atoms, molecules, of particles, to calculate and interpret thermodynamic properties at the molecular level.

The objective of statistical thermodynamics is to describe the behavior of a macroscopic system in terms of microscopic properties of a system of molecular entities.

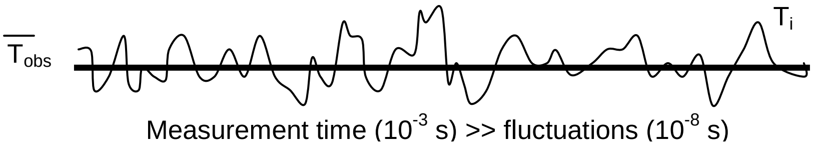

$\overline{T}_{obs}$
T_i
Measurement time (10^{-3} s) >> fluctuations (10^{-8} s)

Figure 2. Measurement of a ‘mean’ temperature and its relation with instantaneous temperature

The main idea is to evaluate an average property value and its standard deviation from a statistically significant number of configurations, much like a real experiment. Indeed, the temperature reading on a thermometer appears falsely constant. At the molecular level, a positive temperature is the resultant of atomic vibrations and collisions occurring at a time scale (e.g. 10^{-8} s.) much lower than the sampling period of the experimental sensor (e.g. 10^{-3} s.) (Figure 2). Using statistical thermodynamic concepts, molecular simulation will do the same and perform a numerical experiment. Each instantaneous configuration (atomic positions and moments) of the system exists according to a probability distribution. The most probable will have the largest contribution to the computed average value. For the experimental system the macroscopic property X value is a time average over a set of configurations Γ(t) sampled during the measurement time t_{meas}:

$$X_{macro} = \overline{X(\Gamma(t))}_{time} = \lim_{t_{meas} \to \infty} \frac{1}{t_{meas}} \int_0^{t_{meas}} X(\Gamma(t))dt \qquad [1]$$

But knowing all configurations Γ(t) is impractical because the number of particles ($6.023x10^{23}$ for a mole) and thus the number of positions and moments are incommensurable. Statistical thermodynamics was developed to solve this problem statistically.

The first postulate of statistical thermodynamics is that "*the value of any macroscopic property is equal to its average value over a sample of the model system configurations*".

$$\overline{X(\Gamma(t))}_{time} = \langle X \rangle_{ensemble} = \frac{1}{\tau_{total}} \sum_{\tau=1}^{\tau_{total}} X(\Gamma(\tau)) \qquad \text{First postulate} \qquad [2]$$

where τ_{total} is the number of sampled configurations. The notation $\langle\rangle_{ensemble}$ refers to a statistical ensemble. By definition it consists in a significant number of sub-ensembles having the same macroscopic properties. The thermodynamic state of a macroscopic system is perfectly specified by a few parameters, for example the number of mole N, the pressure P and the temperature T. From them, one can derive a great number of properties (density, chemical potential, heat capacity, diffusion coefficient, viscosity coefficient …) through equations of state and other thermodynamic relations. Reproducing conditions occurring in experiments, the « canonical » NVT, and the « isobar - isothermal » NPT are quite useful. Notations, NVT and NPT means respectively that the number of moles N + volume V + the temperature T; the number of moles N + the pressure P + the temperature T are kept constant for each system configuration during simulations run in those ensembles.

One considers that the postulate of statistical thermodynamics apply during simulations in a statistical ensemble on systems with a few thousands of particles replicated by periodic boundary conditions and that averages are made on a few million configurations. Sampling size and quality are often the Achilles' heels of molecular simulations.

3.2.3.2 Probability density

Equation 2 states that configurations have the same weight in the average, the same probability of existence. This is the second postulate of statistical thermodynamics: "*All the accessible and distinct quantum states from a closed system of fixed energy (« microcanonic » NVE) are equiprobable*".

Then equation 2 is rewritten:

$$\overline{X(\Gamma(t))}_{\text{time}} = \langle X \rangle_{\text{ensemble}} = \sum_{\Gamma(\tau)} \left(X(\Gamma(\tau)) \cdot \rho_{\text{ensemble}}(\Gamma(\tau)) \right) \qquad \text{Second postulate} \qquad [3]$$

where $\rho_{\text{ensemble}}(\Gamma)$ is the probability density, that is the probability of finding a configuration with positions and moments $\Gamma(\tau)$. In the NVT ensemble, any configuration probability density is connected to its energy E and to Q_{NVT} the total partition function, namely the sum over *all* configurations, by the Boltzmann formula:

$$\rho_{NVT} = \frac{1}{Q_{NVT}} \exp\left(\frac{-E}{k_B T} \right) \qquad [4]$$

Two points are noteworthy:

1. The knowledge of the partition function would allow calculating all thermodynamic properties. But this can never be done fully but rather imperfectly through the generation of a statistically representative number of configurations.
2. A model is required to evaluate any configuration energy in order to calculate the partition function. This is done through a force field.

3.2.3.3 Average, fluctuations and correlation functions

Equation 2 is the usual mean formula to calculate an average value (molar fractions, …). Other properties (heat capacity, …) are calculated from the variance expressing the fluctuations around the mean:

$$\sigma_X^2 = \frac{1}{\tau_{\text{total}}} \sum_{\tau=1}^{\tau_{\text{total}}} \left(\left(X(\tau) - \langle X \rangle_{\text{ensemble}} \right)^2 \right) = \left\langle X^2 \right\rangle - \langle X \rangle^2 \qquad \text{Variance} \qquad [5]$$

Correlation coefficients give access to properties describing the dynamic state of the system. The non normalized form of the correlation coefficient over τ configurations is:

$$\text{correl}_{XX}(\tau) = \left\langle X(\tau) \cdot X(\tau_0) \right\rangle = \frac{1}{\tau_{\text{total}}} \sum_{\tau_0=1}^{\tau_{\text{total}}} X(\tau_0) \cdot X(\tau_0 + \tau) \qquad \text{Correlation coefficient} \qquad [6]$$

The integration of the non normalized correlation coefficients enables to calculate directly macroscopic transfer coefficients (diffusion, viscosity or thermal diffusivity coefficient). Their Fourier transform can be compared with experimental spectra.

3.2.3.4 Statistical error

Molecular simulation is a numerical experiment. Consequently, the results are prone to systematic and statistical errors. The systematic errors must be evaluated then eliminated. They are caused by size effects, bad random number generation, insufficient equilibration period (see further). Statistical errors are inversely proportional to sampling and are thus null for infinite sampling.

On the assumption that the Gauss law applies, the statistical error is the variance (equation 5). However, sampling a large but finite number of configurations induces a correlation between the τ_{total} configurations which persists during a certain number of successive configurations. A statistical factor of inefficiency s is introduced to evaluates the number of correlated successive configurations. The τ_{total} configurations are cut into n_b blocs of τ_b configurations upon which the average $\langle X \rangle_b$ and its variance $\sigma^2(\langle X \rangle_b$ are computed. By selecting several increasing values of τ_b, the statistical inefficiency s and the statistical error $\sigma^2(\langle X \rangle_{\text{total}})$ is evaluated:

$$s = \lim_{\tau_b \to \infty} \frac{\tau_b \cdot \sigma^2\left(\langle X\rangle_b\right)}{\sigma^2(X)_{total}} \tag{7}$$

$$\sigma\left(\langle X\rangle_{total}\right) = \sqrt{\frac{s}{\tau_{total}}} \cdot \sigma(x)_{total} \tag{8}$$

3.2.4. NUMERICAL SAMPLING TECHNIQUES

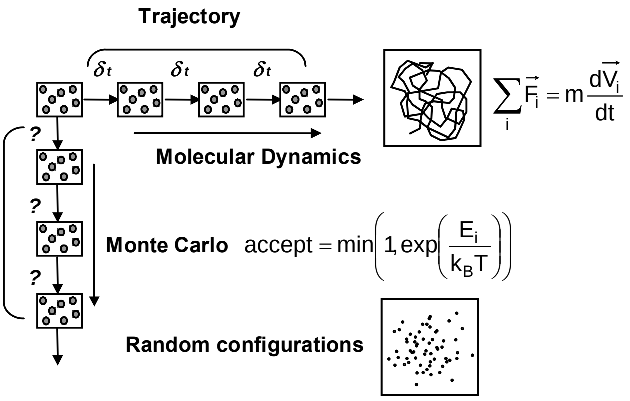

Trajectory
δt
δt
δt
$\sum_i \vec{F}_i = m\frac{d\vec{V}_i}{dt}$
?
Molecular Dynamics
?
Monte Carlo
$accept = \min\left(1, \exp\left(\frac{E_i}{k_B T}\right)\right)$
?
Random configurations

Figure 3. Basic concepts of the Monte Carlo method and Molecular Dynamics

The generation of a statistically representative sample of the model system configurations is mainly done by two techniques, molecular dynamics and the Monte Carlo method. They obey the principles summarized on figure 3. Both methods differ in their applications: the Monte Carlo method is adapted for the study of static phenomena (equilibrium, static interface) while molecular dynamics is adapted to the study of dynamic phenomena (shear induced flow): a phase equilibrium easily computed with Monte Carlo methods would be extremely difficult to reach in molecular dynamics because of the time needed and of boundary effects near the interface.

3.2.4.1 Molecular Dynamics

Molecular dynamics generates a trajectory by integrating the classical equation of motion over time steps δt starting from an initial configuration whose particle positions and velocity are known (figure 3). The nth configuration can be traced down the initial one by reverse integration. In the equation of motion (figure 3), the forces F_i acting on the particle of mass m_i are equal to the derivative of the $V_i(r)$ potential describing the interactions of the particle i with its surrounding

The integration of the differential set of equations is carried out mainly by Verlet-like algorithms rather than by Gear-like algorithms which are widespread in process engineering. The Verlet algorithm calculates the new particle positions r(t) using a 3rd order taylor expansion and replacing the second derivative by the forces thanks to the equation of motion, one gets a formula with no velocity term:

$$r(t+\delta t) = 2r(t) - r(t-\delta t) + \frac{F(t)}{m} \cdot \frac{\delta t^2}{2!} + O\left(\delta t^4\right) \tag{9}$$

Velocities are computed afterwards:

$$v(t) = \frac{dr(t)}{dt} = \frac{r(t+\delta t) - r(t-\delta t)}{2\delta t} + O\left(\delta t^2\right) \tag{10}$$

This algorithm shows several interesting characteristics. (i) It is symmetrical versus δt, which makes the trajectory reversible over the time. (ii) It preserves the total energy of the system over long periods of integration, a key point to get long trajectories and deduce with accuracy some correlation functions. In particular, it is more precise than the Gear-like algorithms for large δt (the reverse is true for small δt), what makes it suitable to simulate long trajectories, which is our goal. (iii) It requires less data storage than Gear-like algorithms.

Transport coefficients (self diffusion, thermal diffusivity and viscosity) are computed from auto-correlation coefficients, the ‘green-kubo' formulas, for instance the coefficient of self diffusion D_i is related to the relative particle velocities:

$$D_i = \frac{1}{3}\int_0^\infty \langle v_i(\tau) - v_i(0)\rangle d\tau \qquad [11]$$

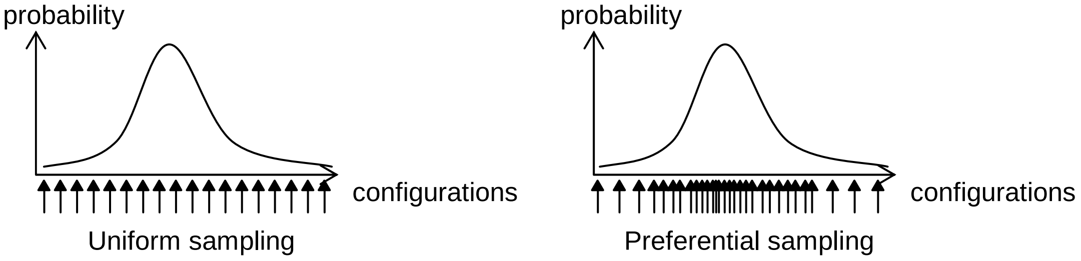
Similarly, viscosity is obtained from the shear stress tensor auto-correlation coefficient related to the pressure exerted on the particle and the thermal diffusivity is obtained from the energy flow auto-correlation coefficient.

A challenge in molecular dynamics run in a statistical ensemble where the temperature is set constant is to keep it constant when moving and interacting particles inevitably heat the system. A solution is to place the system in a large thermostated bath periodically set in contact with the model system through techniques like Andersen or Nose-Hoover methods.

3.2.4.2 Monte Carlo Method

The Monte Carlo method generates system configurations randomly. The nth configuration is related to its preceding one but it is impossible to go back to the initial configuration.

First of all, randomness is particularly critical and has given its name to the method in reference to the Monte Carlo casino. The advice is to always use a published robust random number generator and never try to build one or use the falsely random precompiled computer Ran function. Systematic deviation, repetitive sequence can be checked by running simple tests.



Uniform sampling

Preferential sampling

Figure 4. Uniform and preferential sampling

The second key issue is sampling (figure 4). Uniform sampling allows a good estimate of the partition function needed to compute all macroscopic properties, but at the expense of sampling high energy and thus improbable configurations. Preferential, or Metropolis, sampling samples the configurations with the largest contribution in the calculation of the partition function and of averages.

The disadvantage of the metropolis sampling is that the partition function (equivalent to the surface under the curve) is no longer correctly evaluated. Thus, the question arise of finding how to generate the configurations with a correct probability distribution without having to calculate the function of partition which occurs in the definition of the probability density (equation 4)? The solution is to obey the microscopic law of reversibility:

Given an old (o) and a new (n) configuration, their probability densities ρ are proportional to exp($-E_{(0)}/k_BT$) and exp($-E_{(n)}/k_BT$) in the NVT ensemble (equation 4). Defining the transition probability M(o→n) of going from (o) to (n), the microscopic reversibility states that at equilibrium the number of transitions from (o) to (n) and from (n) to (o) corrected by the probability densities must be equal.

$$\rho^{(0)}_{NVT} \cdot M(o \to n) = \rho^{(n)}_{NVT} \cdot M(n \to o) \qquad \text{microscopic reversibility} \qquad [12]$$

In addition an acceptance criterion is introduced acc(o→n) along with α(o→n) an "a priori" probability of trying to go from (o) to (n) who is supposed to be symmetrical ($\alpha(o\rightarrow n) = \alpha(n\rightarrow o)$):

$$M(o \rightarrow n) = \alpha(o \rightarrow n) \cdot acc(o \rightarrow n) \qquad [13]$$

Then by exploiting the symmetry of α, equation 12 becomes:

$$\frac{acc(o \rightarrow n)}{acc(n \rightarrow o)} = \frac{\rho_{NVT}^{(n)}}{\rho_{NVT}^{(o)}} = \exp\left(\frac{-\left(E_{(n)} - E_{(0)}\right)}{k_B T}\right) \qquad [14]$$

In this equation, the partition function so difficult to calculate no longer appears.

Metropolis idea is to choose acc(o→n) asymmetrically. As indicated on figure 5:

- if the new configuration energy is lower than the old one, the transition is always accepted.
- if the new configuration energy is higher than the old one, one picks a random number ζ between zero and one:
    - if $\zeta = \zeta_1 \leq \exp(-(E_{(n)}-E_{(0)})/k_B T)$, the transition is accepted,
    - if $\zeta = \zeta_1 > \exp(-(E_{(n)}-E_{(0)})/k_B T)$, the transition is rejected.

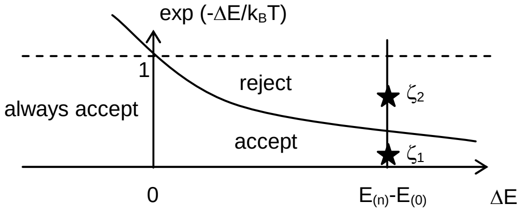


Figure 5. Metropolis preferential sampling. Criterion of acceptance

Applied to equation 14, this acceptance criterion enables to define the acceptance probability of a random displacement in general and in the NVT ensemble:

$$P_{acc,displacement} = \min\left(1, \frac{\rho^{(n)}}{\rho^{(o)}}\right) \xrightarrow[\text{in NVT}]{} \min\left(1, \exp\left(\frac{-\Delta E}{k_B T}\right)\right) \qquad [15]$$

Last, a symmetrical α(o→n) in chosen to allow a sampling effective in terms of acceptance and efficient of the configuration space. Usually one defines a maximum value associated with the transition, like a maximum displacement d_{max} that is fixed in order to satisfy a rate of 50% of accepted transitions. If several movements are possible (e.g. translation, rotation, volume), the type of movement will be chosen randomly from a predetermined statistical distribution. Again, one insists on the randomness of the choices of particle, of the type of movement… in order to respect the microscopic reversibility.

In ensembles other than NVT, probability densities are corrected with respect of the microscopic reversibility law.

3.2.4.3 Phase equilibrium calculations using Gibbs Ensemble Monte Carlo

The Gibbs Ensemble was imagined by Panagiotopoulos in 1987 to simulate vapor – liquid equilibrium. Simulations are carried out in a NVT ensemble on two microscopic boxes located within two homogeneous phases far from any interface. Each box is simulated with periodic boundary conditions. Constant total volume V and total N particles are divided between the two phases V_1, N_1 and V_2, N_2.

The temperature is set constant in the simulations and random movements are performed to satisfy the phase equilibrium conditions as described in figure 6:

- Displacements (translation, rotation) within each phase to ensure minimal internal energy.

- Change of volume proportional between the phases: $\Delta V_1 = -\Delta V_2$ so that the total volume is constant. This should satisfy the pressure equality.
- Transfer of particle from one box to the other to equalize the chemical potentials.

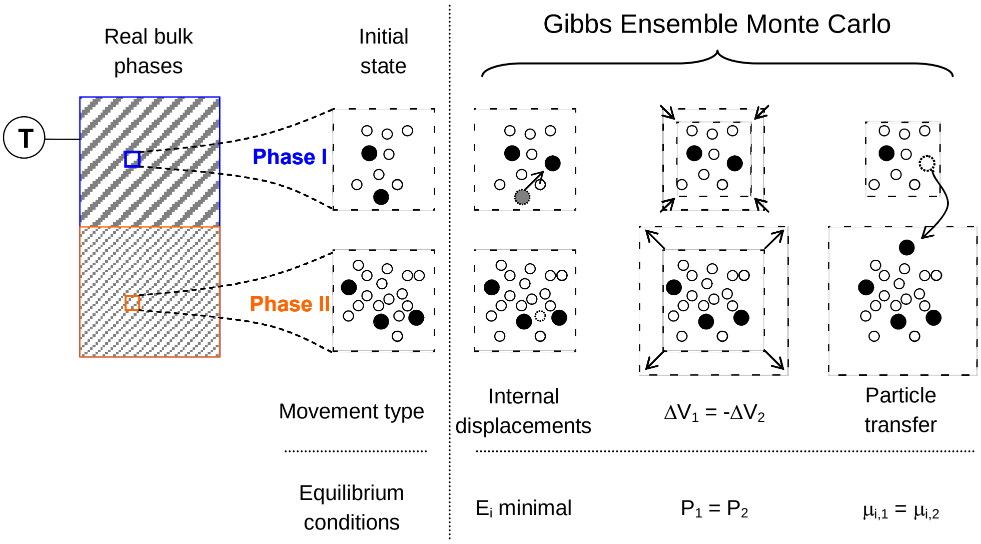


Figure 6. Principles of phase equilibrium simulations in the Gibbs Ensemble.

The acceptance probabilities of the various movements in the case of a single component system are:

For the translation in each area:

$$P_{acc,translation} = \min\left(1, \exp\left(\frac{-\Delta E}{k_B T}\right)\right) \qquad [16]$$

For the change of volume; V_1 being increased by ΔV and V_2 being decreased by as much:

$$P_{acc,volume} = \min\left(1, \exp\left(\frac{-\Delta E_1}{k_B T} - \frac{\Delta E_2}{k_B T} + N_1 \ln\frac{V_1 + \Delta V}{V_1} + N_2 \ln\frac{V_2 - \Delta V}{V_2}\right)\right) \qquad [17]$$

with ΔV chosen by generating a uniform random number ζ bertween 0 and 1 ; δV_{max} being the change of maximum volume adjusted to obtain a fixed percentage (e.g. 50%) of acceptance of the move:

$$\Delta V = \zeta \cdot \delta V_{max} \cdot \min(V_1, V_2) \qquad [18]$$

For the transfer of a particle of area 2 to area 1:

$$P_{acc,transfert} = \min\left(1, \frac{N_2 \cdot V_1}{(N_1 + 1) \cdot V_2} \exp\left(\frac{-\Delta E_1}{k_B T} - \frac{\Delta E_2}{k_B T}\right)\right) \qquad [19]$$

One of the main difficulties of the Gibbs Ensemble Monte Carlo method resides in the transfer of particles to satisfy the chemical potentials equality because of the difficulty to insert polyatomic molecules in the dense phase. An alternative is to seek open spaces where insertion is eased the particle. This affects randomness and introduces a statistical bias like the configurational bias method which consists in inserting segment by segment a molecule in a phase. The probability of acceptance of the transfer of the particle of equation 19 is then modified by introducing the energy differences ΔE_i into weighting factors W_i that represent the total energy of interaction with its surrounding of the inserted molecule. For a L segment molecule inserted in m possible directions:

$$P_{acc,transfert} = \min\left(1, \frac{N_2 \cdot V_1 \cdot W_1}{(N_1 + 1) \cdot V_2 \cdot W_2}\right) \qquad [20]$$

$$W_1 = \prod_{j=1}^{L}\left(\sum_{i=1}^{m}\exp\left(\frac{-E_{1,j}^{i}}{k_B T}\right)\right) \qquad [21]$$

More generally, the introduction of a bias consists in defining a “a priori” probability $\alpha(o \rightarrow n)$ which is no longer symmetrical. Equation 14 and 15 become:

$$\frac{acc(o \rightarrow n)}{acc(n \rightarrow o)} = \frac{\alpha(n \rightarrow o)\cdot \rho^{(n)}}{\alpha(o \rightarrow n)\cdot \rho^{(o)}} \qquad [22]$$

$$P_{acc,mouvement} = \min\left(1, \frac{\alpha(n \rightarrow o)\cdot \rho^{(n)}}{\alpha(o \rightarrow n)\cdot \rho^{(o)}}\right) \qquad [23]$$

To conclude this section, both molecular dynamics and Monte Carlo methods require the calculation of the interaction energy; molecular dynamics to derive forces exerted on the system particles; Monte Carlo method to calculate the acceptance criterion. The following section reviews the main features of the force fields enabling to calculate intra and intermolecular interaction energies.

3.2.5. INTERACTION ENERGY

Suggested readings:

A.R.Leach, *Molecular Modelling, Principles and Applications*, Longmann, Harlow, UK, 1996.

M. Karplus and R.N. Porter, *Atoms and Molecules*, Benjamin, New York, USA, 1970.

3.2.5.1 Quantum chemistry models

Quantum chemistry models are never used alone in molecular simulation because of the still prohibitive computation time. However, they must be considered as they can provide for less sophisticated molecular mechanics models with partial electronic charges and various dipoles usefull to compute Coulombic and dipolar interactions as well as with accurate values of the strain constants describing the bonding intramolecular interactions associated with the various oscillatory modes within the molecules (stretching, bending, torsion).

In quantum chemistry, only atomic nuclei surrounded by revolving electrons are considered. Calculations provide the nuclear and electronic properties system and the true total energy of the system. Total energy is related to the general time dependent wavefunction $\Psi(r,t)$ by means of the generalized equation of Schrödinger:

$$H \cdot \psi = E \cdot \psi \qquad [24]$$

where H is the Hamiltonian, a mathematical operator with kinetic and potential energetic contributions. Apart from an analytical solution for the sole hydrogen atom, the Schrödinger equation solutions are always approximative to various degree that are compromise between computation time and accuracy. Three levels of approximation are considered, namely ab-initio methods, mean field like DFT (density functionnal theory) methods and semi-empirical methods.

Among ab-initio methods, CI (configuration interaction) methods are the most accurate but the slowest. Calculated energy values have a precision comparable with experimental ones (0.001 eV). The CI solutions are obtained by minimizing a linear combination of the wavefunctions associated with the system fundamental state and all excited states.

The Self Consist Field Molecular Orbital concept considers atomic orbitals which represent wavefunctions of electrons moving within a potential generated by the nucleus and by an average effective potential generated by the other electrons. The best such wavefunctions are Hartree-Fock ones and solve the Schrödinger equation for a given electronic configuration (e.g. the fundamental state) without any empirical parameter. They can be used for CI calculations. Atomic orbitals wavefunctions are approximated using Gaussian functions, which leads to peculiar denominations like 'STO-3G hartree-fock calculations' (use of a basis set of 3 gaussian functions). The larger the basis set, the longer and the more accurate the calculation.

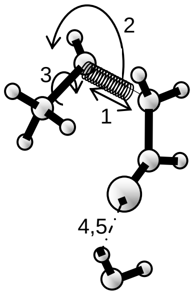
The semi-empirical methods are the most approximated quantum methods: Hückel calculations can be done on a sheet of paper; finer semi-empirical models enable to obtain with a good precision ionization energy, optimal conformations, electronic surface potential. However, they present the disadvantage to calculate approximately the wavefunctions by replacing various integrals by fitted empirical parameters.

Between the two levels of approximation, one finds the methods of a mean field of the popular Density Functionnal Theory. The idea is "rather than to seek to solve the exact problem Hartree-Fock in an approximate way, one could seek to solve an approximate problem in an exact way". That consists in modifying the Hamiltonian operator and replacing the term of exchange of correlation accounting for multi atomic orbital interactions by the electronic density ρ_i. The results are obtained with a satisfying accuracy and much faster than hartree-fock calculations, enabling to even study periodic systems of interesting size.

Nevertheless, all quantum mechanical calculations are performed for a static configuration of the system under 0 K conditions. But as provider of key properties like electronic distribution, they should be systematically used in any molecular simulation aiming to be quantitative.

3.2.5.2 Molecular mechanics models

Molecular simulation uses molecular mechanics models to calculate the internal energy of the system. It consists in considering that the molecules can be represented by centers of forces like beads and bonds by springs (figure 7). As figure 7 shows it, the total internal energy is the sum of intramolecular or bonding interactions and of intermolecular or non-bonding interactions. The set of molecular mechanics parameters is called a force field.



$$V(r^N) = \overbrace{\sum_{\text{bonds}} \frac{k_l}{2}(l-l_0)^2}^{\textit{1}\text{-stretching}} + \overbrace{\sum_{\text{angles}} \frac{k_\theta}{2}(\theta-\theta_0)^2}^{\textit{2}\text{-bending}} + \overbrace{\sum_{\text{torsions}} \frac{W_N}{2}\left(1+\cos(n\omega-\gamma)\right)}^{\textit{3}\text{-torsion}}$$

$$+\sum_{i=1}^{N}\sum_{j=i+1}^{N}\left(\underbrace{4\varepsilon_{ij}\cdot\left(\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12}-\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right)}_{\textit{4}\text{-Van der Waals}}+\underbrace{\frac{q_iq_j}{4\pi\varepsilon_0 r_{ij}}}_{\textit{5}\text{-Coulomb}}\right)$$

Figure 7. Typical molecular mechanic force field

Intramolecular energy takes into account vibrational phenomena between bonded centers of forces. As the beads and spring model suggests, they are described by harmonic functions and handle stretching, bending, torsion as well as improper rotation if needed. Average parameters l_0, θ_0 and harmonic constants k_i are usually fitted to accurate vibrationnal energy calculations made with quantum mechanic methods.

Intermolecular energy takes into account the 2-body interactions between the centers of forces. 3-body interactions are rarely included. Short range interactions can be described by a Van der Waals potential modeled by a 12-6 Lennard Jones function. The $1/r^{12}$ term represents the repulsive contribution which becomes significant below 3 Å. The $1/r^6$ term represents the attractive contribution related to the dispersive effect of induced dipoles. More rigorous forms may include $1/r^8$ or $1/r^{10}$ terms or other functional forms (Buckingham potential, "exponential-6" potential....)

Electrostatic interactions are a major contribution to intermolecular energy as they are long range interactions felt up to 25 Å for multicharged ions. Permanent dipole and multipole are rarely include but coulombic interaction related to partial atomic charges q_i and q_j is a must be. All electronic parameters (dipoles, partial charges) should be fitted to electronic surface potential computed by quantum mechanics to improve quantitative predictions of molecular simulations.

Hydrogen bonding interactions are either modeled explicitely by a 12-10 Lennard Jones function or assumed to be implicitely taken into account in the Van der Waals interaction.

The functional form of the molecular mechanics energy shows that it is not a true energy which could be measured experimentally. Rather, for a single molecule, it is zero at its most stable conformation whereas true zero energy corresponds to the protons, neutrons and electrons infinitely split apart. Molecular mechanics predictions of conformations are in excellent agreement with experimental ones. Nevertheless, the practical use of molecular mechanics is great because for a system of several molecules, it computes the thermodynamic internal energy from which many interesting properties can be derived.

Force fields can be of AA ' all atoms' type as in figure 7 in which there is a center of force on each atom. Their names are Dreiding, Universal Force Field, Compass, OPLS,…. But other types exists where atoms are grouped (e.g. $-CH_3$) under a single center of force in order to reduce the computing time of short range interactions. This lead to UA ' United atoms' force fields.

In all cases long range electrostatic interaction is split in as many center as possible, usually on all atoms and sometimes on virtual centers. A similar idea is at the origin of polarizable force field like the AUA ' Anisotropic United Atoms' which intends to take into account the electronic clouds shift when two particles approach: the charged center is displaced along the resultant of the nearby bonds.

Since all intramolecular parameters and electrostatic parameters are systematically derived from quantum mechanical calculations, molecular simulations using molecular mechanics force fields have greatly improved their accuracies. However, even if for a particle i, Lennard Jones parameters σ_i et ε_i are respectively associated with the colllision diameter (the distance for which energy is null) and with the potential well, they must still be fitted to some extend as will be shown later using experimental data (enthalpies, formation energies, densities…).

For multi component systems, the diameter σ_{ij} and the energy parameter ε_{ij} are obtained from pure substances ones by using traditional mixing rules like those of Lorentz-Berthelot:

$$\sigma_{ij} = \frac{\sigma_i + \sigma_j}{2} \text{ and } \varepsilon_{ij} = \sqrt{\varepsilon_i \cdot \varepsilon_j} \qquad [25]$$

These very simple rules have rarely been questionned, another proof of the strong physical basis of molecular simulation. Furthermore, they highlight that the study of a system with M different centers of forces requires only the knowledge of 2M parameters whereas a traditional approach with a thermodynamic model with binary interaction parameters would require M(M+1)/2 such parameters.

Even if the main functional forms of the potentials (stretching, bending, torsion, Van der Waals, Coulomb) are present in all quoted force fields, the choice must be made by knowing the type of experimental data which were used to regress the Lennard Jones parameters and the way electrostatic interaction are described. Mixing Lennard Jones parameters from several force field without any confrontation to experimental data is acceptable only if qualitative results are sought.

3.2.6. RUNNNING THE SIMULATIONS

How to represent the behavior of macroscopic systems when the model system contains typically only a few thousands of particles? The problem is solved by adopting periodic boundary conditions which duplicate in all directions identical images of the model system. In molecular dynamics, care should be taken that any particle moving through one wall of the main image will reenter at the opposite wall with the same velocity.

Notice that the interaction energy of a particle must include interaction with all including replicated particles. However, for long range interactions this would require too much computer effort and limiting techniques are implemented: rough ones such as a 'cutoff' distance beyond which the interaction is supposed to be null; or more accurate ones like the Ewald summation. In the case of a cutoff, it is necessary to include long range corrections.

The initial particles in the box are usually set along a periodic network to avoid overlaps that would result in an infinite energy. Then, a statistical ensemble and a sampling technique are chosen. Force field parameters are associated to all center of forces and for molecular dynamics simulation a statistical distribution of initial velocities is set. At last, simulation is launched. It consists of a phase of equilibration and a phase of production. The purpose of the phase of equilibration is to bring the system from an initial configuration to a configuration representative of the system: random distribution of the molecules and the velocities within a system with imposed thermodynamic conditions (that of the chosen statistical ensemble). In molecular dynamics under fixed temperature T, the system in gradually heated to the T set value.

The phase of production starts when key properties like potential energy, pressure and density fluctuate over mean values. Each configuration then generated is kept to calculate the macroscopic properties from averages, of fluctuations or coefficients of correlation. As statistical error decreases when the number of configuration increases, at least 10^6 configurations should be generated.

3.2.7. APPLICATIONS

Vapor – liquid equilibrium calculations are a major field of investigation because of the importance of processes like distillation. Too often, data are missing. We present two approaches that use molecular modeling to obtain such data. The first example aims at computing binary interaction parameters occurring in the UNIQUAC activity coefficient model. The second example computes directly the equilibrium compositions using a Gibbs ensemble Monte Carlo method.

3.2.7.1 example n°1 : validation of the UNIQUAC theory

3.2.7.1.1 Overview of UNIQUAC model

The practical calculation of vapor – liquid equilibrium (equation 26) involves an activity coefficient γ_i to describe the non ideality of the liquid phase due to energetic interactions.

$$y_i = \frac{\gamma_i(T,P,x_i)\cdot f_i^{0L}(T,P)}{\phi_i^V(T,P,y_i)\cdot P}\cdot x_i \qquad \text{activity coefficient approach of phase equilibrium} \qquad [26]$$

By applying the thermodynamic relation of Gibbs-Duhem, one connects the individual coefficients of activity γ_i with the excess gibbs energy G^E:

$$RT\ln\gamma_i = \left(\frac{\partial G^E}{\partial n_i}\right)_{T,P,n_j,i\neq j} \qquad [27]$$

The UNIQUAC model proposes an expression for the G^E with two contributions: a combinatorial part which describes the dominant entropic contribution, a residual part which is due mainly to the intermolecular forces which are responsible for the mixing enthalpy. The combinatorial part is related to the composition x_i and to the molecule shape and size. It requires only pure component data. The residual part depends in addition to the interaction forces embedded into two binary interaction parameters A_{ij} and A_{ji}.

$$\frac{G^E}{RT}=\left(\frac{G^E}{RT}\right)_{combinatorial}+\left(\frac{G^E}{RT}\right)_{residual} \qquad [28]$$
$$\frac{G^E}{RT}=f_{combinatorial}(x_i,r_i,q_i)+f_{résidual}(x_i,q_i',A_{ij},A_{ji})$$

Parameters r_i, q_i et q_i' are molecular constants for each pure component i, related respectively to its size, its external geometrical surface and its interaction surface. q' can be different from q, in particular for polar molecules. The model system upon which the UNIQUAC theory was developped consider interacting molecules. Then, the two binary interaction parameters A_{ij} et A_{ji} can be expressed in terms of interaction energies U_{ij} between dissimilar molecules i and j, and U_{ii} between similar molecules i.

$$A_{ij}=\exp\left(\frac{-\frac{1}{2}\left(U_{ij}-U_{jj}\right)\cdot N_A}{RT}\right) \qquad [29]$$
$$A_{ji}=\exp\left(\frac{-\frac{1}{2}\left(U_{ji}-U_{ii}\right)\cdot N_A}{RT}\right)$$

where N_A is the Avogadro number.

The WILSON activity coefficient model also proposes to use two binary interaction parameters. It is a simplification of the UNIQUAC model in which parameters r, q et q' are all set to unity, interaction surfaces are not taken into account and molar volumes are eliminated in the equation.

The relation between the WILSON and UNIQUAC parameters is as follows:

$$A_{ij}^{Wilson}=A_{ji}^{UNIQUAC}q_i+T\cdot\ln\frac{V_j}{V_i} \qquad [30]$$

where V_i et V_j are the components i and j molar volumes.

The traditional approach consists in regressing A_{ij} et A_{ji} from experimental data, with all drawbacks associated with this approach: data specific parameters, poor extrapolation capacity, temperature and pressure dependency, need of experimental data. A few years ago, an attempt to calculate directly the binary interaction parameters has been made and is reported below.

3.2.7.1.2 Calculation using molecular mechanics UNIQUAC binary interaction parameters

In 1994, Jonsdottir et al. [JON 94] computed interaction energies between isolated couples of molecules. They use molecular mechanics models not in a molecular simulation perspective but rather like a quantum mechanical approach: For a given orientation of the two molecules, a energy minimization is run to reach a stable conformation. Many orientations are selected and the mean interaction energy U_{ii} and U_{ij} is evaluated by weighting each value using its Boltzmann factor exp(-U_{ij}/k_BT).

This corresponds to a rought sampling obviously not statistically representative as a few hundred couples are investigated. This questions the validity of the first statistical thermodynamic postulate that equals ensemble average and macroscopic time average value. Rightfully, the authors claim to perform a molecular static approach in between quantum mechanics and molecular simulation approaches. The 'Consistent Force Field' parameters are optimized for the alkanes and ketones that are the molecules of interest but no value in particular no partial atomic charges values are provided. But alkane conformers are taken into account, an advantage of molecular modeling approaches over classical parameter fitting.

The U_{ii} and U_{ij} interaction energies are computed as the difference between the molecules couple energy and the energy of each molecule isolated:

$$U_{ij}=E_{i\ldots j}-\left(E_i+E_j\right) \qquad [31]$$

Simulation results are used with the UNIQUAC equation to predict vapor – liquid equilibrium data (equation 26) which are compared with experimental ones:

- For the alkane/alkane systems (n-butane/n-pentane ; n-hexane/cyclohexane (figure 8) and n-pentane/n-hexane), the relative error ranges from 1.1 to 4% for the pressure and the absolute error tandis ranges from 0,011 to 0,042 for the molar fractions.
- For the alkane/ketones (n-pentane/acétone ; acétone/cyclohexane (figure 8); cyclohexane / cyclohexanone), the relative error ranges from 4.3 to 17.6% for the pressure and the absolute error ranges from 0,016 to 0,042 for the molar fractions.

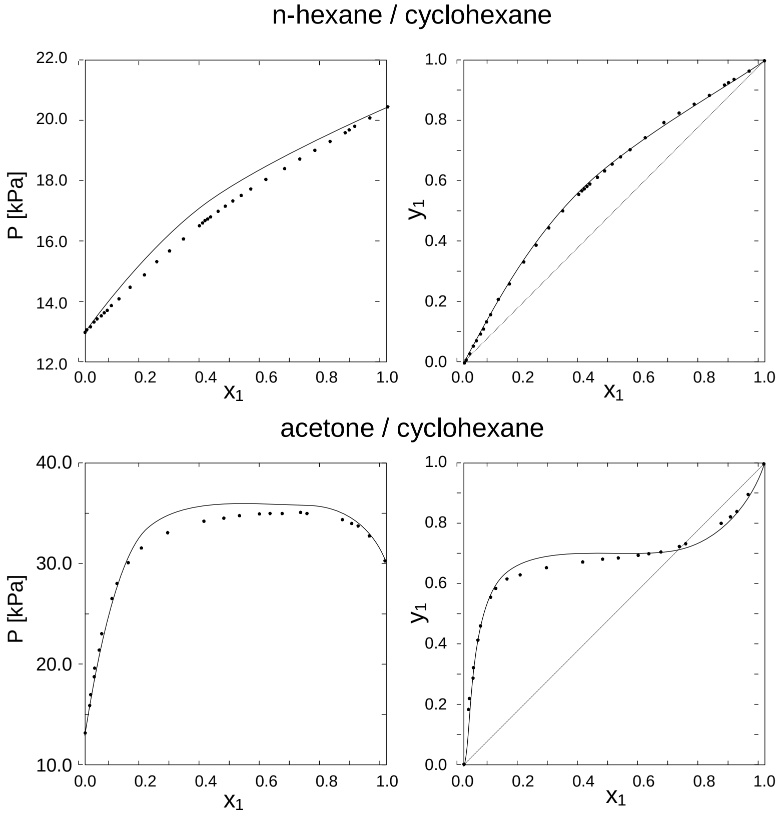


Figure 8. Bubble curve at 298.15 K for the n-hexane / cyclohexane (top figures) and acetone/cyclohexane (bottom figures) systems. Full line: simulation. Stars: experimental data (Reprinted from *Fluid Phase Equil.*, Vol. 100, 1994, Jonsdottir S.O., Rasmussen K. Fredenslund A. 'UNIQUAC parameters determined by Molecular Mechanics'. 121-138. Copyright 1994, with permission from Elsevier).

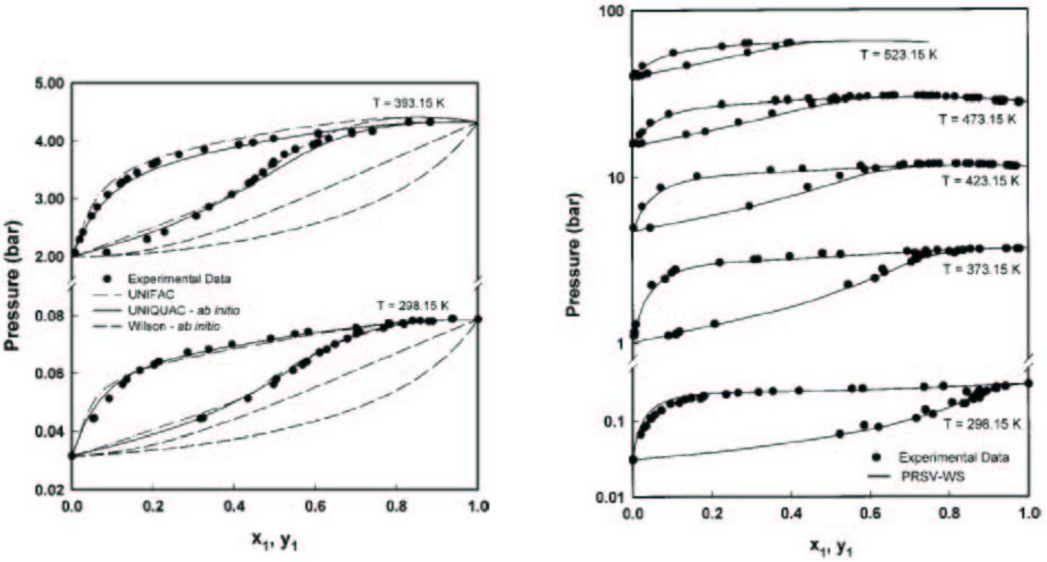
In conclusion, the error increases along with the molecule polarity. One may question the force field ability to handled electrostatic interaction in addition to likely insufficient sampling of the system configurations. Finally, the authors tested the WILSON model and found errors 4 times greater for the system N-pentane/acetone. They concluded that the UNIQUAC equation has better physical basis that that of Wilson. That was foreseeable since the WILSON model is a simplified form of the UNIQUAC model.

3.2.7.1.3 *ab-initio* calculation of UNIQUAC binary interaction parameters

Compared to Jonsdottir's et al. work [JON 94], Sum and Sandler [SUM 99] uses quantum mechanics models so as to improve the representation of electrostatic interactions. As often in such calculations, no rigorous sampling is performed, the accent being made on minimizing the system total energy.

For each binary system, 8 molecules are considered, 4 of each and a stable system conformation is found by minimization using semi-empirical methods. Then the energy is minimized using an *ab-initio* methods (Hartree-Fock method with an extended basis set 6-311**G(3d,2p)). Then couples of molecules are isolated and their interaction energy is computed according to equation 31. The average interaction energy is computed at best on 10 pairs of molecules. Then binary interaction parameters are derived for the UNIQUAC and WILSON equation which enable to compute vapor – liquid equilibrium data from equation 26. Systems studied are higly polar : water – methanol; water - ethanol; water – formic acid; eau – acetic acid and water - acetone. Simulation data are then compared with experimental data and with prediction using the activity coefficient group contribution method UNIFAC.

The results obtained for the WILSON model are never quantitative and even qualitative wrong: as it does not manage to reproduce the azeotropic behavior of the water - ethanol mixture (figure 9). On the other hand, despite the poor sampling, simulations with the UNIQUAC model give good quantitative results, comparable with experimental data and UNIFAC predicted data. Two points are significant: no experimental data where used at any stage, no temperature or pressure conditions where set, an advantage over regressed binary parameters. Indeed, the same set of water - acetone parameters is used to generate accurate data over a large temperature and pressure range (figure 9).



water / ethanol

water / acetone

Figure 9. vapor – liquid equilbrium at 298.15 K for the water – ethanol and water – acetone systems (Reprinted with permission from Sum A.K. and Sandler S.I. ‘A novel approach to phase equilibria predictions using *ab-initio* methods’. *Ind. Chem. Eng. Res.* 38, 2849-2855, 1999. Copyright 1999. American Chemical Society).

In conclusion of this example, when no experimental data are available, vapor – liquid equilibrium data can be predicted using UNIQUAC binary interaction parameters directly computed with molecular modeling methods. If the sampling issue is not yet settled, quantum mechanics methods which descirbe accurately electronic distribution have demonstrated their interest while force field approaches did not for polar systems.

The next example shows that accurate predictions can be made with carefully set force field approaches using efficient sampling of phase equibrium systems thanks to the Gibbs Ensemble Monte Carlo method.

3.2.7.2 Example n°2: direct prediction of nitrile vapor – liquid equilibria

As highlighted before, the development of molecular simulation as a systematic provider of accurate physico-chemical data is impeded by the availability of accurate force fields. Force fields of the 80's were derived to reproduce physicochemical data of monophasic systems or devoted to macromolecules of biological interest (amino acids, proteins…). But the simulation of multiphasic systems was neglected until the Gibbs Ensemble Monte Carlo method and the active development of new force fields of the type AA, AUA et UA : OPLS of Jorgensen [JOR 84], Trappe of Martin and Siepmann [MAR 98], NERD of Nath, Escobedo and de Pablo [NAT 98], Exp6 of Errington and Panagiotopoulos [ERR 99] and AUA of Toxvaerd [TOX 90, TOX 97]) participate to this worldly effort.

As stated before, force field development consists in deriving short range Van der Waals interaction parameters like the σ and ε Lennard Jones parameters. But the challenge is to obtain generic values which can be used for many molecules much like in a group contribution approach and for many properties with various sampling techniques: phase equilibrium data (Gibbs Ensemble Monte Carlo, transport coefficients (Molecular dynamics), absorption isotherm (Monte Carlo). No existing macroscopic model could compute such a wide variety of properties using so few parameters.

In the AUA4 model, generic parameters have been derived for linear, branched and cyclic alkanes, aromatics, hydroxyl, carboxyl and thiol groups [DEL 99, UNG 00, DEL 00, BOU 02a, BOU 02b]. For the nitrile group -C≡N, we proceeded as follow [HAD 03]:

1. quantum mechanics calculations using DFT for the acetonitrile molecule, for which many experimental data are available, to find a stable conformation, determine harmonic constants for the intramolecular contribution of the force field potential, determine discrete partial atomic charges from quantum electrostatic surface potential.
2. setting up the acetonitrile (CH_3CN) force field for which CH_3 Lennard Jones parameters are taken from the generic databank of the AUA4 force field. The same general expression than in figure 7 is used.
3. running Gibbs Ensemble Monte Carlo simulations to identify missing (ε_N,σ_N) and (ε_C,σ_C) Lennard Jones parameters of the nitrile group. The acetonitrile molecule is fully flexible and long range electrostatic interactions where evaluated with a cut-off and tail corrections. Reference experimental data [FRA 75; CHA 84; KRA 85; WAR 94] are the saturated vapor pressure $\ln(P^{sat})$ at 433.15 and 453,15K the vaporization enthalpy ΔH_{vap} and the liquid density ρ_{liq} at 273.15, 298.15, 433.15 and 453,15K. The optimization method is a simple gradient method and the objective function used is a square mean root function with uncertainty values set equal to 0,1for ln(Psat), 1 kJ.mol-1 for ΔH_{vap} and 10 kg.m-3 for ρ_{liq}.
4. Once (ε_N,σ_N) and (ε_C,σ_C) values reproduce acetonitrile data accurately, their genericity is evaluated by predicting with no further parameter adjustment vapor – liquid equilibrium data of other linear nitriles (propionitrile, butyronitrile). In these molecules (ε_N,σ_N) et (ε_C,σ_C) are taken equal to the values obtained for acetonitrile whereas CH_2 and CH_3 Lennard Jones parameters are extracted from the AUA4 databank. Harmonic constant, partial charges and stable conformation are obtained from quantum DFT calculations.

DFT electrostatic surface potential are fitted to partial atomic charges using the simple Mulliken population analysis which splits equally the electronic distribution according to the Van der Waals radius or using the MEP method which mimic the electrostatic potential surface with a least square method. As shown below, the MEP analysis gives the best results but does not pass the genericity test. Each atom bears a partial charge.

All quantum calculated conformations and dipolar moment agree with experimental data [GOL 96] and harmonic constants agree with literature reference values [GOL 96, UNG 00].

Parameter regression requires for each optimization cycle (two are used) 16 Gibbs Ensemble Monte Carlo simulations to compute the gradients varying $[(\varepsilon_N+\delta\varepsilon_N,\sigma_N)\ ;\ (\varepsilon_C,\sigma_C)]$, $[(\varepsilon_N,\sigma_N+\delta\sigma_N)\ ;\ (\varepsilon_C,\sigma_C)]$, $[(\varepsilon_N,\sigma_N)\ ;\ (\varepsilon_C,\sigma_C+\delta\sigma_C)]$, $[(\varepsilon_N,\sigma_N)\ ;\ (\varepsilon_C+\delta\varepsilon_C,\sigma_C)]$ for each of the 4 temperatures considered.

Each simulation takes 20 hours on a Linux Pentium IV, 1.9 GHz with 512 Mo RDRAM. Equilibration period requires 10^6 configurations and the production period ranges from 2.3 to $4.5.10^6$ configurations.

Results of the optimization of the Lennard Jones parameters are:

	ε_C/k_B (K)	ε_N/k_B (K)	σ_C (A°)	σ_N (A°)
Charges MEP optimization	**50.677**	**65.470**	**3.5043**	**3.3077**
Charges Mulliken optimization	**95.52**	**162.41**	**3.2183**	**3.5638**

Table 1. Optimal (ε_N,σ_N) and (ε_C,σ_C) parameter values for the nitrile group. Mulliken Charges.

As indicated below (figure10), the set of MEP parameters gives the best results for the acetonitrile with a mean standard deviation over all reference values of 1.9% and a very good estimate of critical point With an underestimation of the vapor densities, an overestimation of liquid densities at elevated temperature and a poor estimation of the critical point, the Mulliken set gives an error of 3.1%.

But MEP generic character is poor. Whereas Mulliken one is excellent for propionitrile and butyronitrile vapor pressure predictions (figure 11).

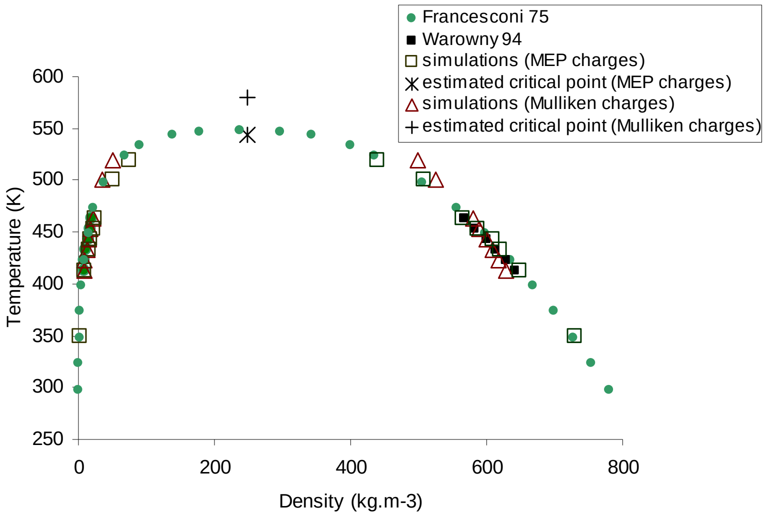


Figure 10. Acetonitrile experimental and simulated vapor – liquid equilibrium.

A possible explanation in the poor MEP predictions lies in the charge values computed for the propionitrile and n-butyronitrile [HAD 04]: the least square fitting of the quantum calculated electrostatic potential surface has in that case led to unphysical values with positive nitrogen atomic charge, in contradiction with the well known electronegativity character of this atom.Also, the MEP nitrile σ_C parameter value which is too elevated compared to other σ_C values associated with other carbonated chemical groups of the AUA4 force field (table 3). Correctly, the σ_C value with the Mulliken distribution follows the decreasing trend as the carbonated chemical group size decreases.

So, two criteria for generic set of Lennard Jones parameters able to model the Van der Waals interaction: (i) physically meaningful values of the Lennard-Jones σ and ε parameters, (ii) physically meaningful set of atomic charges representing the electrostatic potential surface of the molecule. The importance of representation of the electrostatic potential surface has also been acknowledged in COSMO approaches for the computation of physical properties which had recently won the first industrial fluid properties simulation challenge [SAN, 04; CAS 04].

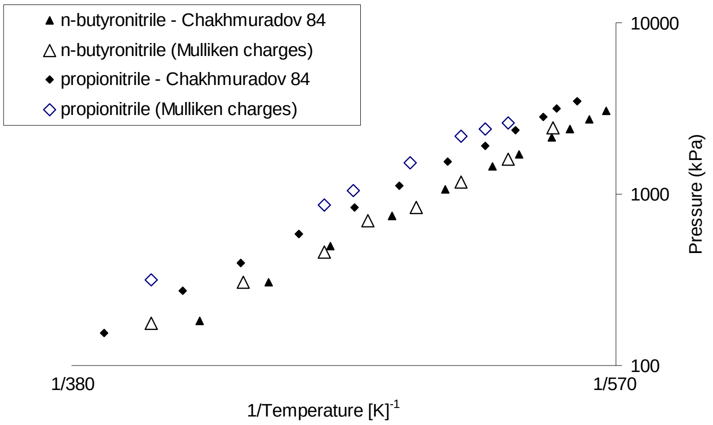


Figure 11. Predicted saturated vapor pressure – propionitrile and butyronitrile. Mulliken charges.

AUA4 Chemical group	$-CH_3$	$=CH_2$	$\equiv CH$	-C MEP charges	-C Mulliken charges
σ_C (Å)	**3.6072**	**3.4612**	**3.3625**	**3.5043**	**3.2183**

Table 2. Comparison of σ_C parameters for various chemical groups in the AUA4 force field.

Critical points in figure 10 are obtained using Ising method [FRE 96]. Direct simulations near the critical point are difficult to converge as shown on figure 12 for the densities versus configurations for a simple Lennard Jones fluid: fluctuations increases and boxes interchange as the reduced temperature nears 1. Experimental observation of a similar phenomenon is well known and demonstrate that the molecular simulation is indeed a numerical experiment that can not only compute accurate physico-chemical data but also behave as an efficient sensor of system behavior at the molecular scale.

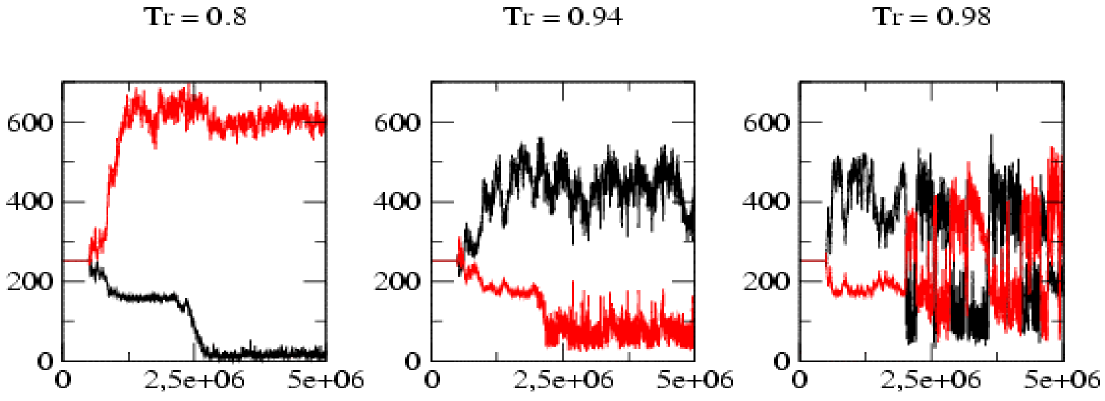


Figure 12. Density versus configuration number in the vicinity of the critical point for a Lennard Jones fluid.

3.2.8. CONCLUSIONS

Molecular modeling is an emerging discipline for the study of energetic interaction phenomena. A molecular simulation performs numerical experiments that enable to get accurate physico-chemical data provided sampling and energy force field issues are addressed carefully. Still computer demanding, molecular modeling tools will likely not be used “on line” or be incorporated in process simulators. But, rather like computer fluid dynamics tools, they should be used in parallel with existing efficient simulation tools in order to provide information at the molecular scale on energetic interaction phenomena and increase the knowledge of processes that must manufacture ever more demanding end products.

REFERENCES

[ALL 87] M.P. Allen and D.J. Tildesley, *Computer Simulation of Liquids*, Oxford University Publications, UK, 1987. [new edition in 2000].

[BOU 02a] Bourrasseau E., P. Ungerer, A. Boutin, A. Fuchs. *Monte Carlo simulation of branched alkanes and long chain n-alkanes with ansotropic united atoms intermolecular potential*. Mol. Sim., 2002, 28(4), 317.

[BOU 02b] Bourrasseau E., Ungerer P., Boutin A., *Prediction of Equilibrium Properties of Cyclic Alkanes by Monte Carlo Simulations – New Anisotropic United Atoms Intermolecular Potential – New Transfer Bias Method*. J. Phys. Chem. B, 2002, 106, 5483.

[CAS 04] Case F., Chaka A., Friend D.G., Frurip D., Golab J., Johnson R., Moore J., Mountain R.D., Olson J., Schiller M. and Storer J., *The first industrial fluid properties simulation challenge*, Fluid Phase Equilibria, 2004, 217(1), 1. and http://www.cstl.nist.gov/FluidSimulationChallenge/

[CHA 84] Chakhmuradov C.G., Guseinov S.O., Iz. Vys. Uc. Zav., 65-69, 1984 (in russian)

[CHE 02] Chen C.C., P.M. Mathias, *Applied Thermodynamics for Process Modelling*, AIChE Journal, 2002, 48 (2), 194.

[DEL 00] Delhommelle J., C. Tschirwitz, P. Ungerer, G. Granucci, P. Millie, D. Pattou, A.H. Fuchs, *Derivation of an Optimized Potential Model for Phase Equilibria (OPPE) for Sulfides and Thiols*, J. Phys. Chem. B, 2000, 104, 4745.

[DEL 99] Delhommelle J., G. Granucci, V. Brenner, P. Millie, A. Boutin, A.H. Fuchs, *A new method for deriving atomic charges and dipoles for alkanes: investigation of transferability and geometry dependence*, Mol. Phys., 1999, 97(10), 1117.

[DEP 02] De Pablo J.J., F.A. Escobedo, *Perspective: Molecular Simulations in Chemical Engineering: Present and Future*, AIChE Journal, 2002, 48 (12), 2716.

[ERR 99] Errington J. R., A. Z. Panagiotopoulos, *A New Potential Model for the n-Alkanes Homologous Series,* J. Phys. Chem. B, 1999, 103, 6314.

[FRA 75] Francesconi A.Z., Franck E. U., Lentz H., *Die PVT-daten des Acetonitrils bis 450°C und 2500 bar,* Ber. Bunsen-Ges. Phys. Chem. 1975, 79, 897 (in German).

[FRE 96] D. Frenkel, B. Smit , *Understanding Molecular Simulation. From Algorithms to Applications*, Academic Press, San Diego, USA, 1996. [new edition in 2002].

[GOL 96] Goldstein E., Buyong MA, Lii J.H., Allinger N.L., *Molecular Mechanics Calculations (MM3) on Nitriles and Alkynes.* J. Phys. Org. Chem., 1996, 9, 191.

[HAD 03] HADJ-KALI M.K., GERBAUD V., JOULIA X., LAGACHE M., BOUTIN A., UNGERER P., MIJOULE C., DUFAURE C. *Prediction of liquid-vapor equilibrium by molecular simulation in the Gibbs ensemble: application to nitriles.* Comput. Aided Chem. Engin., 14, 653-658, proceedings of ICheap 6 Ed.: Pierucci S. ISBN 88-900775-2-2 (Pise, Italy, 06/08-11/03), 2003

[HAD 04] HADJ-KALI M.K., *Application de la simulation moléculaire pour le calcul des équilibres liquide vapeur des nitriles et pour la prédiction des azéotropes*, Thèse de Doctorat en Génie des Procédés, Institut National Polytechnique de Toulouse, Toulouse, France.

[HLO 00] Hloucha M., Sum A.K., Sandler S.I., *Computer Simulation of Acetonitrile and Methanol with ab-initio Based Pair Potential.*J. Chem. Phys., 2000, 113 (13), 5401.

[JON 94] Jonsdottir S.O., Rasmussen K. Fredenslund A. *UNIQUAC parameters determined by Molecular Mechanics*. Fluid Phase Equilibria , 1994,100, 121.

[JOR 84] JORGENSEN W. L., J. D. MADURA AND C. J. SWENSON. *Optimized Intermolecular Potential Functions for Liquid Hydrocarbons.* J. Am. Chem. Soc., 1984, 106, 6638.

[KAR 70] M. Karplus and R.N. Porter, *Atoms and Molecules*, Benjamin, New York, USA, 1970.

[KRA 85] KRATZKE H., MULLER S. Thermodynamic properties of acetonitrile 2. (P, ρ, T) of saturated and compressed liquid acetonitrile. J. Chem. Thermodynamics, 1985, 17, 151.

[LEA 96] A.R.Leach, *Molecular Modelling, Principles and Applications*, Longmann, Harlow, UK, 1996.

[MAR 98] MARTIN M.G., AND J.I. SIEPMANN. *Transferable Potentials for Phase Equilibria. 1. United Atom Description of n-Alkanes*. J. Phys. Chem. B, 1998, 102, 2569.

[MCQ 76] D.A. Mc Quarry, *Statistical thermodynamics*, Harper and Collins Publishers, New York, USA, 1976.

[NAT 98] NATH S.K., F.A. ESCOBEDO AND J.J. DE PABLO. *On the Simulation of Vapor-Liquid Equilibrium for Alkanes*. J. Chem. Phys. 1998, 108, 9905.

[SAN 03] SANDLER S.I.. *Quantum mechanics: a new tool for engineering thermodynamics*, Fluid Phase Equilibria, 2003, 210 (2), 147.

[SUM 99] SUM A.K. and SANDLER S.I. *A novel approach to phase equilibria predictions using ab-initio methods*. Ind. Chem. Eng. Res. , 1999, 38, 2849.

[TOX 90] TOXVAERD S.. *Equations of state of alkanes I*. J. Chem. Phys. 1990, 93, 4290.

[TOX 97] TOXVAERD S.. *Equations of state of alkanes II*. J. Chem. Phys. 1997, 107, 5197.

[UNG 00] UNGERER P., BEAUVAIS C., DELHOMMELLE J., BOUTIN A., ROUSSEAU B. and FUCHS A.H... *Optimisation of the anisotropic united atoms intermolecular potentiel for n-alkanes*. J. Chem. Phys., 2000, 112 (12), 5499.

[WAR 94] WAROWNY W., *Volumetric and phase behavior of acetonitrile et temperatures from 363K to 463K,* J. Chem. Eng. Data, 1994, 39, 275.