UNIVERSITÀ DEGLI STUDI DI MILANO

SCUOLA DI DOTTORATO IN SCIENZE E TECNOLOGIE CHIMICHE

DIPARTIMENTO DI SCIENZE FARMACEUTICHE

CORSO DI DOTTORATO IN CHIMICA DEL FARMACO

CICLO XXVIII

DEVELOPMENT OF ALTERNATIVE METHODS TO IN VIVO AND IN VITRO ASSAYS FOR THE DETERMINATION OF SKIN PERMEABILITY TO CHEMICAL COMPOUNDS

SETTORE CHIM/09 FARMACEUTICO TECNOLOGICO APPLICATIVO

Dott. PAOLO ROCCO

Matricola: R10088

Tutor: Prof.ssa PAOLA MINGHETTI
Co-tutor: Prof. ALESSANDRO PEDRETTI

Coordinatore del dottorato: Prof. MARCO DE AMICI

ANNO ACCADEMICO 2014/2015

Summary

Acknowledgements

Thanks to Prof. Alessandro Pedretti, whose knowledge of Molecular Dynamics, VEGA ZZ, NAMD and programming in general made it possible. Thank you for guiding me.

Thanks to Prof. Paola Minghetti for giving me the opportunity to work in a great team and do what I love.

Thanks to Prof. Giulio Vistoli for essential advice on computational chemistry, Molecular Dynamics and for getting me started.

Thanks to Prof. Francesco Cilurzo for useful hints and advice on the experimental and historical aspects of the work.

Thanks to Dr. Umberto Musazzi for the useful discussions about skin permeation.

Thanks to Angelica, Matteo, Alessandra, Sivia, Chiara and all the people in the labs, for encouragement, talks, dinners, and making the experience much more enjoyable.

Thanks to my family and friends for support and motivation.

1 Introduction and aim of the work

A better understanding of the barrier function of the skin is relevant to a wide range of applications including, in particular, transdermal delivery of drugs and risk assessment of hazardous exposure to chemicals. The regulation of transdermal permeation is generally ascribed to the stratum corneum (SC), the outer layer of the skin [1].

The macroscopic structure of the stratum corneum was initially described according to the bricks and mortar model proposed by Michaels [2], where corneocytes are arranged like bricks in a lipid phase (the mortar phase), arranged in multilayers [3]. The lipids are predominantly in a continuous gel phase with fluid regions [4], organized in an orthorhombic conformation at low temperature, and undergoing a phase transition to hexagonal conformation at about 60 °C [5]. At the molecular level, the major components of the SC lipid matrix are ceramides (CER), free fatty acids (FFA) - ceramide 2 and FFA 24:0 being the most abundant - and cholesterol (CHOL) in a 1:1:1 molar ratio [6]. Selective inhibition of any one of these components is sufficient to compromise the barrier function of the skin [7].

Potts & Guy	$logKp(cm/s) = 0.71 \log K_{o/w} - 0.0061\,MW - 6.3$	$r^2 = 0.676$
Mitragotri	$Kp(cm/s) = 5.6\,10^{-6}\,K_{o/w} \cdot exp(-0.46\,MR^2)$	$r^2 = 0.698$

Table 1-1. Predictive equations for skin permeability. *Kp* = permeability coefficient expressed in cm/s; $K_{o/w}$ = octanol/water partition coefficient; *MW* = molecular weight; MR = molecular radius; r^2 = square of the correlation coefficient calculated on Flynn's set.

A large number of permeability studies have appeared in the last 60 years, addressing the prediction of skin permeability with mathematical models, in particular semi-empirical or mechanistic models, the most cited of which being the equations by Potts and Guy [8] and Mitragotri [9], respectively (Table 1-1). The correlation were focused on “Flynn’s set”, a set of 97 molecules of known Kp [10].

Although the discovery of the link between physicochemical properties of a given permeant and its ability to permeate the skin can be traced back to the works of Stephen Rothman in the 1930s and 1940s [11], the first modern mathematical models of percutaneous absorption were derived by Higuchi, in the 1960s. In his works, the Author describes the rate of release of transdermal drugs in terms of drug concentrations and diffusion coefficients according to Fick’s first law [12][13], laying the foundation of steady-state (time independent) models of skin permeability.

Assuming that the skin barrier is a homogenous membrane, and defining the permeability coefficient Kp as the steady state flux across the skin (J_{ss}) divided by the concentration gradient across the skin (ΔC), Fick's first law may be expressed as [14]:

$$Kp = \frac{J_{ss}}{\Delta C} = \frac{KD}{h} \quad \textbf{Equation 1-1}$$

where D is the solute's diffusion coefficient, K its partition coefficient between the vehicle and the skin, and h the skin thickness.

Equation 1-1 is expressed for the entire skin, but different degrees of simplification are possible. In principle, a mathematical model of skin permeability should consider the contribution of the various physiological regions: the stratum corneum, the viable epidermis, the dermis and appendages, such as hair follicles and glands. Appendages are a possible pathway for hydrophilic solutes, but it is unlikely that they may explain their transdermal transport [15], and are often ignored. The overall *Kp* could then be seen as deriving from the three different contribution in series, as proposed by Scheuplein and Blank [16][17]:

$$\frac{1}{K_{p.skin}} = \frac{1}{K_{p.sc}} + \frac{1}{K_{p.ve}} + \frac{1}{K_{p.der}}$$

where $K_{\text{p.skin}}$, $K_{\text{p.sc}}$, $K_{\text{p.ve}}$, $K_{\text{p.der}}$ are the permeability coefficients of the skin, stratum corneum, viable epidermis and dermis respectively.

In most works, the SC is considered the rate limiting barrier [15], even though focusing on the SC barrier only leads to unrealistic values of *Kp* for extremely lipophilic compounds, the percutaneous absorption of which is controlled by the hydrophilic viable epidermis [2].

Equation 1-1 may be seen as the basis for the development of the semi-empirical model proposed by Potts and Guy. Express the skin/vehicle partition coefficient *K* as a function of the octanol/water partition coefficient *P* [18]:

$$K = P^b$$

where b varies from about 0.4 to 0.9. Express the diffusion coefficient D as a function of molecular volume MV [8]:

$$D = D_0 e^{-\beta \cdot MV}$$

where D_0 and β are constants. Then, approximating the molecular volume MV of the permeant by its molecular weight MW, Equation 1-1, expressed as a logarithm, takes the form [10]:

$$\log K_p = -6.3 + 0.71 \log P - 0.0061 MW$$

where K_p is expressed in cm/s and the coefficients are obtained through linear regression of Flynn's data set excluding three compounds [10]. Since hydrophilic solutes are under-represented in Flynn's set, this model fail to provide satisfactory predictions for the K_p of hydrophilic solutes [19]. A thorough evaluation of semi-empirical methods is found in the review by Lian [1].

An example of (steady-state) mechanistic model is the equation proposed by Mitragotri [9], based on Scaled Particle Theory, in which the partition and diffusion coefficients in the SC are calculated from the work required to form cavities in the SC, with the assumption that the solute moves in a stationary frame of lipids [9].

Diffusion across a non-homogeneous, anisotropic medium like the SC is a function of its position inside the membrane, at least in the transversal direction (depth z), lateral (perpendicular to z direction) and transverse

components of the diffusion coefficient are also expected to be different, as an effect of anisotropy. Accounting for the anisotropy of the SC and for the average direction of diffusion, Scaled Particle Theory yields an average diffusion coefficient *D* [9]:

$$D = D_0 e^{-A \cdot r^2}$$

where *r* is solute molecular radius and *A* is a constant calculated to be ~ 0.4 for a dipalmitoylphosphatidylcholine (DPPC) bilayer [9]. The Mitragotri equation then takes the form [9]:

$$K_p = 5.6 \times 10^{-6} P^{0.7} e^{-0.46 r^2}$$

with the same notation as above. If $MW^{1/3}$ is used as an estimation for *r*, once the Mitragotri equation is put in logarithm form it may seem similar to the Potts and Guy equation. However, the numerical values are derived from the theory employed and not from linear regression.

A deeper insight on mathematical models as well as an account of other model based methods for the estimation of SC diffusion coefficients are provided in recent reviews [15][20][21]. In particular, the work by Mitragotri and colleagues [15] contains a thorough discussion on the theoretical estimation of partition and diffusion coefficients. The estimation of partition and diffusion coefficients may also be achieved through Molecular Dynamics (MD) simulations.

Both empirical and mechanistic models require a method of estimating essential parameters, such as diffusion and partition coefficients, which can be determined experimentally, theoretically or through computer simulations, in particular Molecular Dynamics simulations which are gaining increasing attention as a tool in skin permeability studies. Since even the longest MD simulations are still orders of magnitude shorter than the time scales of complex biological events, such as permeation [22], the data gathered from them as to be interpreted on the basis theoretical skin permeability models.

Molecular Dynamics can be a useful tool to perform permeability studies, with the added benefit of a greater understanding of the permeation process at a molecular level, and of the role of different part of the SC in determining permeability. Since the time scales of MD simulations are still orders of magnitude lower than the time scales of the most complex biological events, the atomistic details obtained with simulations have to be included in the framework of macroscopic mathematical models [22].

The use of MD simulations to study solute diffusion through lipid bilayer membranes was developed in the 1990s for phospholipid bilayers, starting with early studies on passive diffusion [23][24][25] to systematic studies of permeation by Marrink and Berendsen, which took into account diffusion and solubility of penetrants into the membrane [26][27], and extended to the study of SC permeation by Das and colleagues [7][28]. The approach used here, though, differs from that of the previous works. To allow for the screening of more than 80 molecules, a more indirect method has been attempted, not contemplating free energy calculations. Instead, only the diffusion coefficient and the conformational space explored by the solute in

the different microenvironments inside the SC have been directly derived from the MD simulations. All the other relevant parameters were estimated from those data.

This work focuses on the role of Molecular Dynamics, as a tool in skin permeability studies. In Chapter 2, an outline of Molecular Dynamics is given. In Chapter 3, we look at structural MD studies on lipid matrices, and we analyze how the methods based on MD simulations, developed for phospholipid bilayers, have been applied to estimate the permeability, diffusion and partition coefficients in the Stratum Corneum. Chapter 4 is about the computational details. The set-up of the lipid matrix model and the set of permeants is described, along with the simulation parameters and a description of production runs. Chapter 5 contains the theoretical background to the molecular properties and physical quantities discussed in Chapter 6. In Chapter 6, the results from Steered Molecular Dynamics (SMD) simulations are given. The goal was to investigate the dynamic behaviour of a permeant when crossing an ideal SC lipid matrix composed by a heterogeneous mixture of long-chain ceramides (CERs), free fatty acids (FFAs), and cholesterol (CHOL) in a 1:1:1 molar ratio. The data derived from the simulations are correlated to the experimental permeation coefficients contained in the reduced set (the intersection of Flynn's and the Fully Validated set by Vecchia and Bunge [81]), yielding encouraging predictive models which confirm the fruitfulness of MD simulations to analyze complex systems such as the SC at a molecular level.

2 Molecular Dynamics

Molecules possess several degrees of freedom (vibrational, rotational, etc.) which allow their existence in different conformational states. Conformational states dynamically evolve in time, forming the so-called conformational space of the molecule. The molecule exploring a conformational space is not limited only to the lowest energy conformations, but it can move among different equilibrium states, which are related to the kinetic energy of the atoms and so to the temperature of the system. The path followed in conformational space depend on inter-and intra-molecular interactions, which depend largely on the dynamic behaviour of atoms and on experimental conditions.

Molecular Dynamics is a computational technique that allows to follow the evolution of a chemical system (a group of atoms or a group of molecules) using theoretic physics methods of different sophistication, from *ab-initio* or semi-empirical quantum chemistry methods to Molecular Mechanics (also called *force-field* methods) [29]. This chapter explores the theory behind MD, in particular classical MD which is the most relevant for our study.

Molecular Dynamics is based on Newton's equation of motion, applied to the movement of particles. In the Molecular Mechanics (MM) approximation, MD simulations are described by deterministic classical physics, since the evolution of the system is determined by a set of initial conditions (coordinates of the particles, and the forces acting upon them). Modern MD, though, can take a stochastic approach, by the insertion of random forces along the ones described by Newton's equation.

2.1 Molecular Mechanics

In the MM approximation, it is understated that cumulative forces can be used to describe molecular geometries and energies [30]. A molecule is considered as a set of balls (atoms) and springs (bond) the conformation of which is determined by the force acting on any atom, calculated as the cumulative effect of bonded and non-bonded interactions.

In MM a potential energy hypersurface is constructed from a set of experimental data (e.g. crystal structure geometries, IR and Raman spectroscopy, heats of formation) and more recently even from quantum mechanical calculations. The functional form and the sets of parameters derived from experimental data used to calculate the potential energy $U(r_1, r_2, ...r_n)$ as a function of atomic spatial coordinates form the so-called *force-field*.

In the case of an isolated system composed of n particles, the force acting on the i^{th} particle is related to the system's potential energy by the following equation:

$$F_i = m_i \frac{d^2 r_i}{dt^2} = -\nabla\left[U\left(r_1, r_2, ...r_n\right)\right] \quad \textbf{Equation 2-1}$$

where m_i and r_i are the mass and position of the i^{th} particle, respectively and U is the potential energy and ∇ is the *gradient* mathematical operator. From this equation, the position of any atom can be determined and, and so its speed and acceleration. Knowledge of the potential energy function U, determined from the position of each particle, allows to calculate the force acting on each particle at any given *time step* (time is discretized). In turn, the

force so calculated is used to determine the particle's position at the subsequent time step, and so on to the end of the simulated time period.

A fundamental assumption of MM is that the potential energy of a system can be expressed as the sum of the potential energy terms deriving from simple interactions, usually divided in bonded and non-bonded interactions.

2.1.1 Bonded interactions

Potential energy terms derived from bonded interactions include: bond length potential, bond angle potential and torsional potential (note that the term 'potential' is improperly used here to designate a potential energy function).

The **bond length potential** U_r can be described by the harmonic potential, which is adequate for small deviation from reference values:

$$U_{harmonic}(r) = S_h(r - \bar{r})^2$$

where r is the bond length, $\bar{r}$ is the reference value, S_h is a constant related to the spring constant; or by the Morse potential, valid for greater bond deformations:

$$U_{Morse}(r) = D\{1 - \exp[S_M(r - \bar{r})]\}^2$$

where S_M and D represent the potential energy well width and depth (i.e. the dissociation energy), respectively. Opposite to the harmonic potential, the Morse potential has correct asymptotic behaviour: $U_{Morse}(r) \to D$ as $r \to \infty$ and $U_{Morse}(r) \to \infty$ as $r \to 0$.

Other potential forms exist, which take into consideration cubic and quartic terms. Bond length reference values are derived from X-ray crystal structures (which yield average values) or *ab-initio* quantum chemical methods (which yield equilibrium values), such as Hartree-Fock/Self Consistent Field.

The **bond angle potential** U_θ is usually described by a harmonic potential:

$$U_{harmonic}(\theta) = K_h(\theta - \overline{\theta})^2$$

or by a trigonometric potential

$$U_{trigonometric}(\theta) = K_t(\cos\theta - \cos\overline{\theta})^2$$

where θ is the bond angle, $\overline{\theta}$ is the reference value for the bond angle, K_h and K_t are constants. The trigonometric potential has the advantage of being bounded and easier to implement (e.g. easier to differentiate) [30].

The **torsional potential** U_φ is modelled by a simple periodic function:

$$U_{torsional}(\varphi) = V[1 - \cos(n\varphi)]$$

where φ is the torsion angle, n is an integer denoting the periodicity of the torsional barrier and V is a constant dependent on the barrier height.

The potential energy for bonded interactions can then be expressed as:

$$U_{bonded} = \sum_{bonds} U_r + \sum_{angles} U_\theta + \sum_{torsions} U_\varphi$$

2.1.2 Non-bonded interactions

Potential energy terms derived from bonded interactions include the Lennard-Jones potential and the Coulomb potential.

In MD, **6-12 Lennard Jones potential** is generally used to model cohesive van der Waals interactions, whit the addition of a short range repulsive term. For a pair of atoms at distance r, it takes the form:

$$U_{LJ}(r) = \alpha\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]$$

where α is an adimensional constant, ε is the depth of the potential well, σ is the distance at which the inter-particle potential is zero. The negative term in the expression above, corresponds to the attractive van der Waals interaction, which arises from the interaction between spontaneously oscillating electric dipole, and it can only be explained with Quantum

Mechanics [31]. The positive (repulsive) term acts at very short distances, originating from Pauli's exclusion principle.

The Coulomb potential is used to describe ionic interactions between fully or partially charged particles. For an atom pair *{i,j}* at distance *r*, it takes the form:

$$U_{Coulomb} = K_{Coulomb} \frac{q_i q_j}{\varepsilon \cdot r}$$

where q_i is the electric charge on atom *i*, ε is the dielectric constant and $K_{Coulomb} = 1/4\pi\varepsilon_0$ (ε_0 being the permittivity of a vacuum).

The potential energy for non-bonded interactions can then be expressed as:

$$U_{non-bonded} = \sum_{atom\_pairs} U_{LJ} + \sum_{atom\_pairs} U_{Coulomb}$$

2.2 Integration of the equations of motion: the Verlet algorithm

The commonly used methods for integrating Equation 2-1 with respect to time belong to the Stőrmer/Verlet/leapfrog class of algorithms. These methods, with respect to other integration methods, such as Runge-Kutta have the advantage of stability and they are symplectic integrators. Symplectic integrators have the property of "almost" preserving the value of the Hamiltonian, i.e. energy (the total energy is not preserved exactly). This

naturally lends to the study of the microcanonical or (NVE) ensemble, in which the total number of particles, *N*, the volume *V* and the total energy, *E* are held constant. If instead the canonical or (NVT) ensemble, where Temperature *T*, and not energy is constant, is to be studied, either the Verlet algorithm must be modified to include a thermostat, such as Nosė-Hoover thermostat, or Langevin's equation, instead of Newton's, must be used.

The integration is achieved with a discretization of time in timesteps Δ*t*. Then, the continuous variables *x(t)* and *v(t)*, position and velocity at time *t* respectively, are approximated by values at time intervals *n*Δt, where *n* is a positive integer. We only describe here a common variation of the Verlet algorithm, known as the leapfrog algorithm, so called because the velocity is defined a half steps $v\left(t+\frac{\Delta t}{2}\right)$ while the position is defined at whole steps $x(t+\Delta t)$. For a particle of mass m, on which a cumulative force *F* is acting, the velocity update is

$$v\left(t+\frac{\Delta t}{2}\right)=v\left(t-\frac{\Delta t}{2}\right)+\frac{F}{m}\Delta t$$

and the position update is:

$$x(t+\Delta t)=x(t)+v\left(t-\frac{\Delta t}{2}\right)+\frac{F}{m}\Delta t$$

In this equation, it is assumed that the acceleration *F/m* is constant during the timestep *Δt,* which must then assume very small values, which translated to timestep of 1 fs (femtosecond) for most systems. Indeed, the time step is

set around a value between 1/100 and 1/20 of the fastest period of oscillation of the system simulated. The optimal value is in the range of 0.2 fs (femtosecond), although the use of special algorithms (e.g. SHAKE) slow down the transitions, allowing longer time steps (1-2 fs). Since Equation 2-2 is applied recursively, the use of too large time steps (> 10 fs) can introduce recursive errors in the evaluation of the energy of the system and prevents a correct evaluation of the forces and accelerations. In this way, all non-linear energy terms are not properly estimated and consequently the system gradually acquires energy until its explosion.

2.2.1 Constraints dynamics

In the MD simulations, it is sometimes necessary to apply constraints that restrict the degrees of freedom of the system. Usually some torsion angles or atom position are restricted, for example to exploit data coming from NMR or X-ray studies. One of the methods employed is named SHAKE that tethers the distance between pairs of atoms:

$$r_{ij}^2 - d_{ij}^2 = 0$$

where the term r_{ij} represents the instantaneous separation between two atoms (i and j) and d_{ij} is the reference value of the constraint. In SHAKE dynamics, the energy function includes a loop-free constraint, and then a displacement vector is added to satisfy the constraints imposed to the system. The position of the *i-th* atom at time $t + dt$ is:

$$r_i(t + \Delta t) = r_i'(t + \Delta t) + \Delta r_i(t)$$

where $r'_i\,(t + dt)$ is the position vector of the i^{th} atom after an iteration without constraints, while $\Delta r_i(t)$ is the displacement that satisfies the constraint calculates as:

$$\Delta r_i(t) = \frac{2(\Delta t)^2}{m_i} \sum_j \lambda_{ij} r_{ij}(t)$$

where the λ_{ij} are Lagrange multipliers.

Another type of constraint commonly used in MD simulations of biomacromolecules (nucleic acids and proteins), which allows the study of experimentally resolved structures, is the *atom fixing*. It consists in blocking the coordinates of a subset of atoms to avoid excessive changes in the experimentally validated structure. This approach, however, induces strong energy gradients around the fixed atoms so, other kind of atomic constraints are usually preferred, such as *harmonic constraints,* where atoms are not fixed, but harmonically constrained to fixed points. A recent development is this technique is Steered Molecular Dynamics (SMD), in which the center of mass of a group of atoms is harmonically constrained to a moving point.

2.3 Choice and importance of simulation parameters

The choice of simulation parameters is a fundamental step to obtain reliable results. It is generally impossible to determine *a priori* which conditions are the best ones to simulate the behaviour of a given system properly, so different combinations of parameters should be tried in order to select the best conditions, evaluating all the possible situations, taking into consideration the available computational power.

It is important to analyze the main parameters required by each program for molecular dynamics and their influence on the evolving simulation besides the required computational time. We briefly discuss temperature.

2.3.1 Temperature

Usually, MD simulations are performed at an absolute temperature between 300 and 310 K, as these are the temperatures of experimental assays on biological matrices. More problematic is the choice of the conditions with which the system is brought to the desired temperature: at the beginning, the atoms are stationary, then normally distributed pseudo-random velocities are assigned to them. The average value E_k of the kinetic energy is related to the temperature of the system T as:

$$\langle E_k \rangle = \frac{1}{2} \left\langle \sum_{i=1}^{3N} m_i v_i^2 \right\rangle = \frac{N_F k_b T}{2}$$

where N_F are the degrees of freedom of the system (*3N* or *3N-3*) and k_b is Boltzmann's constant.

So if we assign to a "cold" system at the beginning of the simulation a temperature of 300 K this can lead to excessively high velocities that can make the integration algorithm unstable. For a correct simulation at constant temperature heating, from 0 K to the target temperature is so required. This period simulates the transfer of energy from the environment to the system, allowing atoms to accelerate, and, only from a certain step onwards, the temperature is kept constant. Once the system has been heated to the desired temperature, it undergoes an equilibration phase, to allow for the redistribution of the energy in order to stabilize the global structure. To simulate the system under thermostatic conditions, the energy is exchanged with the external environment, such as an external heat bath at constant temperature, for the duration of the calculation.

2.4 Stochastic Dynamics: fluctuation-dissipation relations and Langevin dynamics

In a deterministic system, described by Newton's equation of motion:

$$F_i = m_i \frac{d^2 r_i}{dt^2} = -\nabla[U(r_1, r_2, ...r_n)]$$

a time reversal of all the variables, at the end of a time period leads to the observation of the reverse trajectory in the same time period. The fact that Newton's equation of motion is time reversal symmetric is in contrast with the common experience that all but the simplest processes are irreversible.

This irreversibility is linked to a quantity that always grows in the universe, entropy, which has been given different explanations in the past, for example by Boltzmann and Gibbs. A more modern view is that Newton's equation of motion is capable of explaining irreversibility, if it's complemented by the acceptance of uncertainty in the way systems behave, due to interactions with the environment. So modified, Newton's equation leads to a quantity that always grows with time, like Entropy, and that's uncertainty itself. When this uncertainty is added to Newton's equation of motion, symmetry is broken, and the deterministic laws of evolution of the system must be replaced by rules for the evolution of the probability that the system takes a certain configuration [32]. Such rules define a *stochastic process.* A general equation describing the evolution of probability density functions *P* which are continuous in space and in time, is the Fokker-Plank equation (*FPE*):

$$\frac{\partial P(x,t)}{\partial t} = -\frac{\partial (M_1(x,t)P(x,t)}{\partial x} + \frac{1}{2}\frac{\partial^2 (M_2(x,t)P(x,t))}{\partial x^2}$$

where M_1 and M_2 are the first and second Kramers-Moyal coefficients (we omit their definition).

If Newton's equation is supplemented with environmental forces, of which some are random, we obtain an equation describing a stochastic process, such as Langevin equation, which describes the evolution of the velocity v of a particle. The Langevin equation can be put in the form:

$$\dot{v} = -\gamma v + b\xi(t) \qquad \textbf{Equation 2-2}$$

where γ is the damping coefficient, b is a constant, $\xi(t)$ has the property $\langle \xi(t) \rangle = 0$ and it has no autocorrelation in time. It can be shown [32] that Equation 2-2 is equivalent to the FPE is we take $M_1 = -\gamma v$ and $M_2 = b^2$. The FPE equivalent of Langevin equation is:

$$\frac{\partial P(v,t)}{\partial t} = -\frac{\partial(\gamma v P(v,t)}{\partial v} + \frac{b^2}{2}\frac{\partial^2 P(v,t)}{\partial v^2}$$

Since the stationary solution of this equation must be the Maxwell-Boltzmann distribution $P(v) \propto \exp(-mv^2 / 2k_b T)$, comparing it with the analytical solution, we get $b^2 = 2k_b T\gamma / m$ [32]. Since this relates a quantity, b, characterizing the fluctuations and quantity, γ, characterizing the dissipation, it is called a *fluctuation-dissipation* relation.

2.4.1 Langevin dynamics

Equation 2-2 forms the basis for Langevin dynamics. Langevin dynamics is an approximate method that allows us to eliminate unimportant degrees of freedom that are substituted by the medium effects (friction) and stochastic forces (random forces). Langevin equation for the i^{th} particle, in the presence of an external force $F_i(t)$ can be written as:

$$m_i \frac{d^2 r_i(t)}{dt^2} = F_i(t) - m_i \gamma v_i(t) + R_i(t)$$

where m_i is the mass, γ is the damping constant or collision frequncy, $R_i(t)$ is the random force, which oscillate around zero:

$$\langle R_i(t) \rangle = 0$$

The effects of random forces and frictional dispersion maintain the thermal balance, allowing the energy of the system to oscillate without varying excessively.

The covariance of the diagonal matrix $\boldsymbol{R}$ of the random forces R_i is dependent on γ:

$$\langle R(t)R(t')^T \rangle = 2\gamma\gamma_B TM\delta(t-t')$$

where k_B is the Boltzmann constant, T is temperature in Kelvin, $\delta(t-t')$ is Dirac delta function, and M is the diagonal matrix of dimension 3N of the masses m_i (each repeated three times). So, the damping constant γ controls not only the magnitude of the friction coefficients $\beta_i=m_i\ \gamma$, but also the variance of the random forces. For large values of γ, the kinetic energy is quickly damped by the collisions of the solvent and the system follows diffusive Brownian dynamics.

2.5 Common Force Fields in lipids simulations

The interaction potential can be computed by several approaches of different complexity also including high-level *ab-initio* methods. However, for large systems, such as a lipid bilayer, the interaction between particles (i.e. atoms or group of atoms) is described by the molecular mechanics approximation, where the possibility of covalent bond formation and

breaking is excluded. In the molecular mechanics "ball and spring" approximation, atoms are considered as spheres linked together by springs. The force acting on every particle, and so the trajectory, is calculated from a potential energy function expressed a set of parameters. These parameters define the so-called force field, and are derived from different sources, such as spectroscopy or quantum mechanical calculations. The potential energy function comprises bonded (stretching, bending and dihedral torsions) and pair wise non-bonded interactions (e.g. van der Waals interactions and electrostatic interaction between charged atoms) [33].

Force fields can be classified as atomistic (all atoms, united atoms) or coarse grained, where groups of atoms are grouped into pseudo-particles. In the simulation of lipid bilayers, both approaches have been used. In particular, as an all atoms force field, the most widely used is CHARMM [34], and as a united atoms force field, GROMOS [35][36].

In CHARMM force field, all atoms are described explicitly. CHARMM parameters for lipids, optimized on the condensed phase properties of alkanes [37], were introduced in CHARMM22 [38], and are periodically updated [39].

GROMOS employs a united atoms approach, and is optimized on the properties of pentadecane [79]. It represents each of non-polar CH, CH_2 and CH_3 as a single particle, i.e. it represents aliphatic hydrogens implicitly. Two main versions of GROMOS exist, the so called original GROMOS, and the most widely used Berger modification [40], which employs the Ryckaert-Bellemans potential for the description of torsions of the hydrocarbon chains.

When a much larger time scale than possible with atomistic simulations is required, a simplified coarse-grained (CG) model can be used. CG models surfaced in the 1990s, through the work of Smit and coworkers [41], allowing a speed-up of 3–4 orders of magnitude compared to atomistic simulations [42]. A typical application of coarse-graining is in the study of the self-assembly of lipids [43]. In CG models, groups of atoms are represented by a single bead. This allows a drastic reduction of the degrees of freedom in the simulation. The set of data against which CG force fields are optimized can be experimental, as in the case of the Martini force field, developed by Marrink and co-workers for the coarse grained simulations of lipid bilayers [44], or atomistic simulation data [45]. The use of the Martini model allows MD simulations of several microseconds, opening the door to the study of long time scale events such as gel phase formation [46] and hexagonal phase formation [41] in lipid bilayers.

When force fields are extended to lipids, the main focus in parameterization is the reproduction of experimentally accessible properties like electron density profile, area per lipid, order parameters, and membrane thickness. This can be often achieved, but there are exceptions. As an example, MD simulations of the spontaneous insertion of aliphatic alcohols into phosphatidylcholine bilayers with the Berger force field for the lipids (and the GROMOS force field for the alcohols) brought partition coefficients for the long-chain alcohols in good agreement with experimental data, but an overestimated partition coefficient for ethanol [79]. Similar results were obtained with CHARMM27 [47].

A thorough evaluation of different force fields for lipid bilayers simulations can be found in [79][48][49]. Several reviews and discussions

have dealt with coarse grained models of bilayers [50][42]. A review of coarse-grain modeling of lipid bilayers appeared in [51].

3 Molecular Dynamics as a tool for theoretical skin permeability studies. An overview.

In this chapter, we review the evolution of Molecular Dynamics simulations of the Stratum Corneum, addressing both structural and permeability studies. Both fields have evolved from simulations studies on generic lipid and phospholipid bilayers [7][48]. So this will be our starting point.

Kox and coworkers pioneered the field of structural studies, using MD simulations to study the first realistic monolayer of amphipathic molecules with atomistic resolution: the system consisted of 90 molecules composed of 7 repeating CH_2 units, with the first unit (the "head group") constrained in a plane to simulate the interaction between polar head groups [52]. Two years later, van den Ploeg and coworkers studied a bilayer of 2×16 decanoate molecules, applying periodic boundary conditions in two dimensions [53]. In the late 1980s, the work of Egberts and Berendsen on a system of sodium decanoate/decanol/water [54] led the way to the study of lipid mixtures. From that point on, the number of MD studies on bilayers began to grow exponentially.

The state of the art in bilayers simulations up to 1997, from the pioneering studies on dipalmitoylphosphatidylcholine (DPPC) and dimyristoylphosphatidylcholine (DMPC) in the liquid crystalline state to the studies on phospholipid bilayers which express a gel phase under 60°C (characteristic of the SC) is described in an extensive review by Tieleman,

Marrink and Berendsen [55]. More recent studies on phospholipid bilayers are covered in later reviews [50][42][48][49].

3.1 Structural studies on the stratum corneum lipid matrix

Compared to the number of MD simulations performed on phospholipid bilayers, there are few simulation studies on ceramide bilayers. This may be in part due to the fact that neither CHARMM nor GROMOS are optimized specifically for ceramides [56].

The first molecular dynamics simulation intended to mimic stratum corneum lipids was performed by Höltje and co-workers [57], to explore the effects of cholesterol. The results yielded by two systems, a FFA bilayer (stearic acid/palmitic acid) and a FFA/CHOL mixture at a 1:1 molar ratio are compared. The pure FFA bilayer is constituted by highly ordered lipid chains tilted relative to the bilayer normal, yielding a gel phase at 303 K. The presence of cholesterol increases the available free volume by decreasing the conformational order of the hydrocarbon chains. This effect is also confirmed for ternary mixtures of CER:CHOL:FFA [28]. The GROMOS87 force field is used for lipids, with Ryckaert-Bellemans potentials for the CH_2 and CH_3 groups, and the SPC model [58] for water.

The use of simplified models, possibly coupled with Coarse-Grained force fields, has been used to simulate phenomena requiring longer time scales to be observed, such as the self-assembly of a lipid bilayer [43].

Structural MD studies of realistic ceramide bilayers have begun to appear in the last ten years, and most are reviewed in [33]. It is important however to highlight that these include the first MD study of a ceramide bilayer by Pandit and Scott [59], the first study of a bilayer of asymmetric ceramides by Notman and co-workers [60] and the first systematic MD study on bilayers composed of ceramides, FFA and cholesterol at different molar ratios [7] by Das, Noro and Olmsted, along with other studies on mixtures by the same group [61][62]. Other works of note are a study on the effect of ceramide fatty acid chain length on the structure of bilayers [63], and a study on the structure and phase behavior of CER NS and CER NP bilayers [56]; a study on stacked bilayers of synthetic CER EOS (methyl-branched), CER AP, C22:0 FFA and CHOL in a 23:10:33:33 molar ratio in water [64]; a study on a mixture of CER NS (24:0), lignoceric acid and CHOL in 1:1:1 molar ratio, undergoing the addition of small quantities of oleic acid [65].

Besides the few structural simulation studies on the stratum corneum lipid matrix, a number of papers on MD simulations of ceramide molecules in different contexts [66][67] as well as simulations of sphingomyelin bilayers [68] have been published in recent years, but are outside the scope of this thesis.

3.1.1 Permeation simulation studies on the stratum corneum

The first thorough application of the inhomogeneous solubility-diffusion theory to solute permeation of the stratum corneum has been carried out by Das and colleagues [**Errore. Il segnalibro non è definito.**], although free energy calculations in ceramide bilayers had been performed before by

Notman and colleagues [69] relative to the study of the permeability enhancing mechanism of DMSO.

In the work by Das and colleagues, Molecular Dynamics and the inhomogeneous solubility-diffusion theory are applied to water permeation. The water molecule is constrained at a fixed distance (z) from the bilayer mid-plane. The simulations are performed in the extended ensemble at constant temperature (350K) and constant pressure with the GROMACS MD package [70][71]. As force field, a modified version of united atom OPLS [72] is used for the lipids and SPC [58] for water. The bilayer consists of 128 Ceramide molecules and 5250 water molecules. Data from different simulations are used to calculate $\Delta G(z)$, and the diffusion coefficient in the *xy* plane and in the direction of the bilayer normal ($D_{\perp}(z)$ and $D_{\parallel}(z)$, respectively). Then permeability is calculated from $\Delta G(z)$ and $D_{\parallel}(z)$ as outlined above. The simulations are then repeated with bilayers consisting of ternary mixtures of Ceramide:Cholesterol:Free Fatty Acid in different proportions at 300K and 350K. It is found that $D_{\perp}(z)$ falls by a factor of 1000 in the ordered tail region (while $D_{\parallel}(z)$ falls by a factor of 10). This suggests that the ordered hydrocarbon tails define "channels" perpendicular to the bilayer surface. Moreover, there is a larger local density at the mid plane, where the interdigitation of the asymmetrical tails occurs, compared with a DPPC bilayer, where the diffusivity at the mid plane is the same as in bulk water. The permeability calculated for the 2:2:1 CER:CHOL:FFA bilayer at 300K is about 30 times lower than the experimental value. Moreover, the calculation of the diffusion path according to Fick's diffusion model, from the calculated $D_{\parallel}(z)$ and lag-time, yield an apparent path length that is 250 times larger than the bilayer thickness [28].

4 Computational details.

In this chapter the details about the purely computation aspects of the work are given. From the structure of the lipid matrix model and its set-up, and the set-up of the 80 permeants, to the protocol followed during the Steered Molecular Dynamics production runs, and the choice of simulation parameters, the force field used, and the software packages employed for all those tasks.

4.1 Set-up of the lipid matrix model

An ideal SC lipid matrix model has been generated according to the model proposed by Iwai et al. [3]. It is composed by a heterogeneous mixture of long-chain ceramides (CERs), free fatty acids (FFAs), and cholesterol (CHOL) in a 1:1:1 molar ratio [6], which proved to be the most stable upon equilibration. Starting with a basic unit composed of one molecule each of CER 2 in extended conformation (Figure 4-1a), 24:0 FFA (Figure 4-1b) and CHOL (Figure 4-1c), a minimization by NAMD 2 and subsequent optimization by MOPAC 2012 [73] are performed followed by a 1 ns molecular dynamics at 300 K. 4 basic units (4 CER 2, 4 FFA, 4 CHOL) are assembled to form a monolayer unit. In each monolayer unit, one CER 2 was replaced by one CER 1 (thus giving a 25% of CER 1 in the ceramide component of the membrane), and the system undergoes a 1 ns MD simulation after that an energy minimization was performed. From this monolayer unit, membrane models of increasing sizes is built, up to a bilayer consisting of 868 lipid molecules (Figure 4-2). At this point, water molecules

are added to the model to account for solvation effects in the donor and acceptor phases, and a 10 ns MD simulation is performed (Figure 4-2**2**) to allow to equilibrate the bilayer as confirmed by the corresponding MSD profile (Figure 4-3).

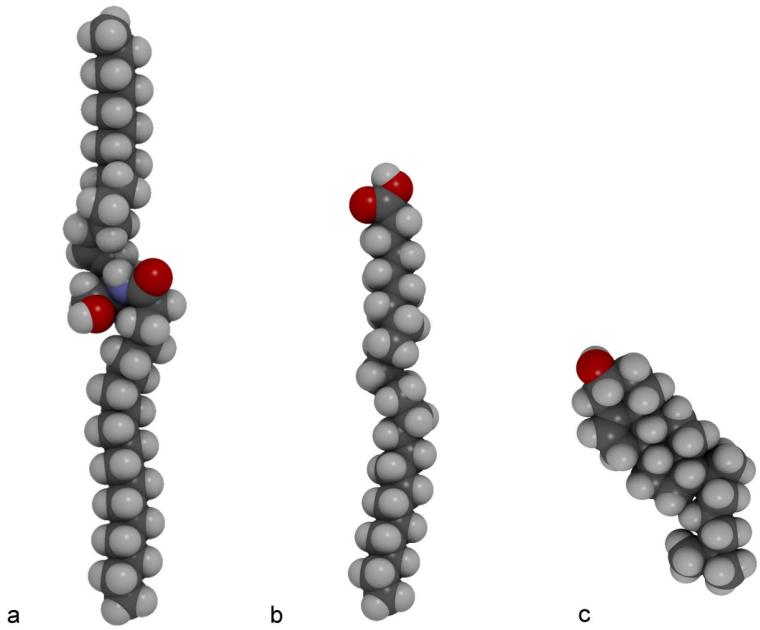

Figura 4-1. a) Ceramide 2 (NS) in extended conformation. **b)** 24:0 FFA (lignoceric acid). **c)** Cholesterol.

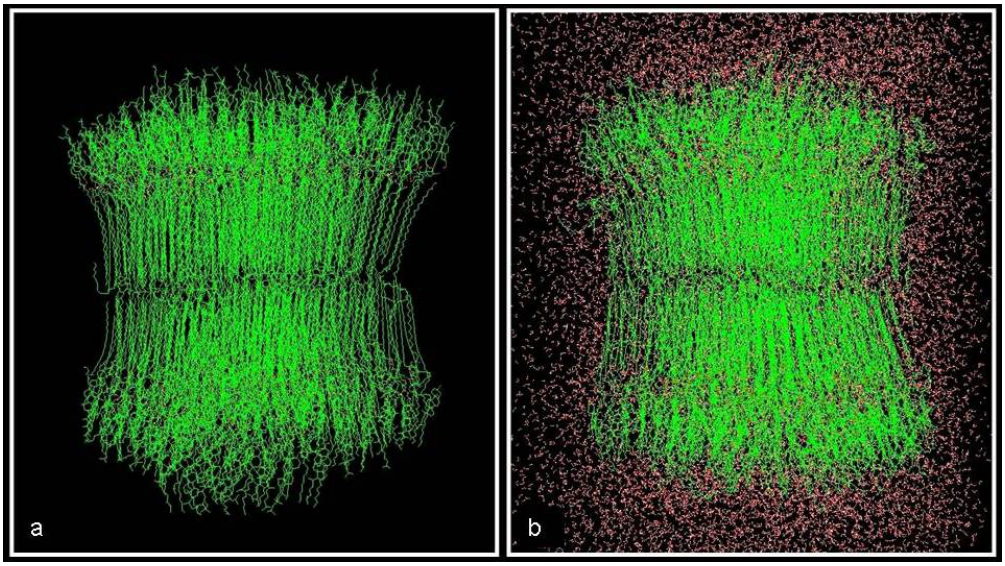

Figure 4-2. The final 868 molecules bilayer before (a) and after (b) a 10 ns equilibration molecular dynamics in presence of explicit solvent (water). Apolar hydrogen have been removed from picture. Images obtained with VEGA ZZ [78].

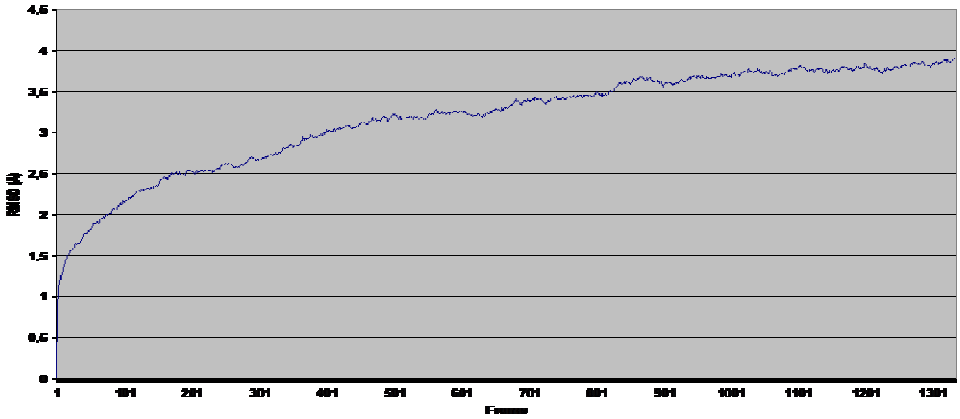

Figure 4-3. Change in Root mean square deviation (RMSD) during the 10ns equilibration MD simulation.

4.2 Set-up of the permeants

The below described SMD simulations were performed on the reduced of 80 permeants set as defined in Appendix A. This allowed us to avoid data in Flynn's set that do not meet the Fully validated set validation criteria (Appendix A), still being able to compare our results with existing models. In detail, the 80 molecules were simulated in their neutral form since it is involved in permeation processes. The conformational profile was explored by a quenched Monte Carlo procedure which produced 1000 conformers by randomly rotating the flexible torsions. For each considered permeant, the so obtained lowest energy conformer was further optimized by PM7 semi-empirical calculations (also to derive more precise atomic charges) and underwent SMD simulation.

4.3 Steered MD (SMD) simulations

As a preamble it should be emphasized that the limited time period under investigation does not allow the spontaneous transport of molecules across the membrane to be simulated. Instead, steered molecular dynamics (SMD) simulations are used (paragraph 5.2), where solute molecules are dragged from water into and through the bilayer at a constant velocity in the direction perpendicular to the bilayer surface. SMD runs were performed on the 80 permeants with the following characteristics: (a) periodic boundary conditions (108 Å x 108 Å x 162 Å) were applied to stabilize the simulation space; (b) Newton's equation was integrated using the r-RESPA method (every 4 fs for long-range electrostatic forces, 2 fs for short-range non bonded

forces, and 1 fs for bonded forces); (c) the long-range electrostatic potential was computed by the Particle Mesh Ewald summation method (108 × 108 × 160 grid points) (d) the temperature was maintained at 300 ± 10 K by Langevin's algorithm; (e) Lennard-Jones (L-J) interactions were calculated with a cut-off of 10 Å and the pair list was updated every 20 iterations; (e) a frame was memorized every 10 ps, thus generating 1000 frames; and (f) no constraints were imposed to the systems. The simulations were carried out in two phases: an initial period of heating from 0 K to 300 K over 300000 iterations (300 ps, i.e. 1 K/ps) and the monitored phase of 2.5 ns. During this time, the solute molecule was forced to cover a distance of 50 Å at a speed of 0.02 Å/ps by applying a harmonic constraint force equal to 5 Kcal/mol/Å^2.

Though the trajectory imposed on the solute in the SMD simulation doesn't necessarily follow the same tortuous path as in the physical system [9], the solute experiences all the microenvironments that can be encountered in a real permeation process. In detail, the trajectory imposed on the solute in the SMD simulation doesn't follow the route through the least dense region. Instead, the solute is made to follow the transversal route through the polar headgroups. This allows for an investigation of the behaviour of the permeants along the whole double layer, through all the possible microenviroments, without prejudice about which one is the path of least resistance, thus exploring the entire conformational and physicochemical property space [74] as the molecule moves through the bilayer.

As physicochemical properties, polar surface area (PSA), surface, and lipophilicity (Virtual logP) [75] are considered, averaged in the different zones of the bilayer. The automatic linear regression script in VEGA ZZ is

used to find a correlation between the physicochemical properties and the permeation coefficient for the set of solutes. Other descriptors that can be derived directly from the SMD simulations, such as the force acting on the solute molecule, have proved of little value. Improvements for the estimation of the molecule behaviour, such as Free Energy calculations [26] couldn't be applied so far. As a matter of fact, a Free Energy calculation from a SMD using Jarzynski equality [76] would require multiple trajectories for every solute, and that is unfeasible for a screening of 80 molecules.

4.4 Simulation parameters

Molecular dynamics simulations are performed using NAMD 2 [77] and VEGA ZZ [78] software packages, on consumer grade desktop computers. The force fields used are CHARMM 36 [39] for the lipids, and the SPC potential [58] for water. In CHARMM force field, all atoms are described explicitly. CHARMM parameters for lipids, optimized on the condensed phase properties of alkanes [79], were introduced in CHARMM22 [80], and are periodically updated [**Errore. Il segnalibro non è definito.**].

All lipid molecules were in non-ionized form, as well as all solute molecules. A correction has been applied to the experimental Kp of molecules that are partially ionized at experimental conditions, based on the estimate of the non-ionized fraction [81].

5 Theoretical background

In this chapter, the theoretical background to our work is presented in detail. First, an account of the inhomogeneous solubility-diffusion model, developed by Marrink and Berendsen for studying membrane permeation is given. Then an explanation is given of the molecular properties and quantities used in Chapter 6.

The inhomogeneous solubility-diffusion model represents an important approach to the problem, but it is based either on equilibrium MD simulations or on multiple non equilibrium MD simulations. Neither of these two cases applies to our work, since it was our goal to explore the different regions of the lipid matrix and to screen a large number of permeants. Either of this two goals requires multiple simulations: multiple equilibrium simulations in different regions or multiple SMD (non-equilibrium) simulations needed for calculating Free Energy differences from Jarzinsky equality (Paragraph 5.2.1).

An alternative method, which was explored in our work, was to use SMD simulations to explore the conformational space and property space of the permeants [82]. SMD simulations lead to the exploration of a reduced space, with respect to a conformational search, since MD takes naturally into account the environment. So only the conformational and property subspaces explored by the molecules in the lipid matrix are considered.

5.1 Molecular Dynamics simulations of solute permeation: the inhomogeneous solubility-diffusion model

The use of MD simulations to study solute diffusion through lipid bilayer membranes was developed in the 1990s for phospholipid bilayers, starting with early studies on passive diffusion [23][24][25] to systematic studies of permeation by Marrink and Berendsen, which took into account diffusion and solubility of penetrants into the membrane [26][27], and the statistical mechanics studies of MD trajectories by Xiang [83]. A thorough review of the first works on solute permeation of biomembranes can be found in [55]. More recent works has been reviewed in [84] and [14].

The inhomogeneous solubility-diffusion model, developed by Marrink and Berendsen, accounts for the non-homogeneity of biomembranes. Though the model is rigorously derived in [26], a simplified derivation is proposed here, with the same *a priori* assumptions, to highlight its physical meaning and applicability. If the membrane is homogeneous, the total resistance R (for our purpose, defined as the inverse of the permeability coefficient Kp) opposed by the bilayer to solute permeation is (Equation 1):

$$R = \frac{1}{K_p} = \frac{h}{KD}$$

In a non-homogeneous membrane, as K, D, and therefore R are a function of depth z, the total resistance R can be expressed as the sum of the infinitesimal resistances dR:

$$dR = \frac{dz}{D(z)K(z)}$$

over the bilayer thickness *h*, yielding [84][33]:

$$R = \int_0^h \frac{dz}{D(z)K(z)} \quad \textbf{Equation 5-1}$$

where *K(z)* is the partition coefficient from water into the membrane at depth *z*, *D(z)* is the diffusion coefficient at depth *z*.

The integral in Equation 5-1 assumes that the permeant moves, predominantly, along the direction of the bilayer normal, z, thus going through all the barriers of resistance *dR* 'in series'. It is rigorously valid for plasma membranes, where a solute undergoing passive permeation has to reach the inside of the cell from the outside. On the other hand, the different topology of the SC allows the permeant to follow a tortuous path [15][28], which will be a combination of motion along the bilayer normal and perpendicular to it. One of the models proposed for lipids arrangement assumes that trans-bilayer steps are not required during the permeation process, and a continuous path without interruptions along the whole SC depth is allowed [15]. In this case, the permeant could avoid the zones of higher resistance along the bilayer normal, which will thus make a negligible contribution to the integral in Equation 5-1. As a consequence, the total resistance *R* will be lower, and the permeation coefficient will be higher than calculated through Equation 5-1, as observed by Das and colleagues for a water molecule [28].

In general, the calculation of the total resistance R in the case of the SC will require a generalization of Equation 5-1, accounting for the importance of lateral diffusion [9], and for the effective path followed by the solute. An expression for R, for example, could take the form of a line integral over the solute's effective path L' in the SC lipid matrix:

$$R = \int_{L'} \frac{dL}{D(L)K(L)} \quad \textbf{Equation 5-2}$$

where $D(L)$ and $K(L)$ are the path-dependent partition and diffusion coefficients at point L along the solute's path, respectively. Note that the relevant quantities in Equation 5-2, $D(L)$ in particular, lacks a clear definition in this formulation. Further refinements of the equation, though, are outside the scope of this work. A review of relevant papers about the effects of SC topology on effective path length calculation is contained in [15].

5.1.1 Estimation of diffusion and partition coefficients with Molecular Dynamics

Molecular Dynamics performed on bilayers naturally accounts for the anisotropic and non-homogeneous structure of biomembranes, where the partition coefficient K(z) and the diffusion coefficient D(z) are a function of depth z inside the bilayer. The diffusion coefficient will have different values in the direction or parallel (z) or perpendicular (xy) to the bilayer normal.

The local diffusion coefficient D(z) can be calculated in different ways from Molecular Dynamics trajectories, the simplest method being from the

mean square displacement (MSD). For diffusion in the z direction (but the reasoning can be extended to calculation in the xy plane) [86]:

$$D(z) = \lim_{t\to\infty} \frac{\left\langle (z(t) - z(0))^2 \right\rangle}{2t} \quad \textbf{Equation 5-3}$$

where the average $\langle\ \rangle$ is, theoretically, over different system replicas, but it is replaced, for our purposes, with the time average, a procedure that would be rigorous only if the ergodic hypothesis could be verified (which is seldom the case). The limit of this approach is that, as the solute diffuse it goes through regions with different diffusion coefficient, so that Equation 5-3 is applicable only to short time diffusion periods [86].

In this work, both D_z and D_{xy} were calculated from the MSD (in the z direction and in the xy plane, respectively) obtained from the simulations, using the algorithm proposed by Cameron Abrams [85].

A more general approach relates D(z) to the force fluctuations. In the case of a solute constrained at depth z by a force $F(z,t)$, the diffusion coefficient can then be expressed in terms of the time integration of the autocorrelation function of the random forces $\Delta F(z,t)$ [86]:

$$D(z) = (RT)^2 \Big/ \int_0^\infty \left\langle \Delta F(z,t) \Delta F(z,0) \right\rangle dt$$

where the random forces $\Delta F(z,t)$ are calculated as the deviation of the instantaneous force $F(z,t)$ acting on the constrained molecules from the average force $\langle F(z,t)\rangle$:

$$\Delta F(z,t) = F(z,t) - \langle F(z,t)\rangle .$$

The partition coefficient *K*(z) can be calculated from the free energies of solute partitioning from water to depth *z* of the bilayer [14]:

$$K(z) = \exp\left(-\Delta G(z)/RT\right)$$

where

$$\Delta G = \int_{outside}^{z} \langle F(z')\rangle_t \, dz'$$

where $\langle F(z')\rangle_t$ is the constraint force at position z' averaged both over time and over different system replicas, and so a large number of different simulations are required for free energy calculations.

Once the partition and diffusion coefficients are estimated using the above methods, the permeability coefficient *Kp* can be calculated from Equation 2 [83][84][86]. The model, initially applied to the study of membrane permeation by water [26], was later applied by the same Authors to ammonia, and oxygen molecules [27][87]. It has become widely used as a model for biomembrane permeation studies, as in the case of small molecules [88], β-blocker drugs [89] and valproic acid [90].

5.2 Force, Work and Free Energy differences

In a constant velocity Steered Molecular Dynamics simulation, the center of mass of the tagged atoms (in our case the center of mass of the permeant) is harmonically constrained (with force constant *k*) to move at constant velocity *v* in the direction $\vec{n}$. The situation is equivalent to having a "dummy carrier" moving at velocity *v* in the direction $\vec{n}$, bound to the center of mass of the tagged atom with a spring of force constant *k*. So the potential applied to the moving atoms is:

$$U(\vec{r}_1, \vec{r}_2, ..., t) = \frac{1}{2} k \left[vt - (\vec{R}(t) - \vec{R}(0)) \cdot \vec{n} \right]^2$$

where $\vec{R}(t)$ is the position of the center of mass at time *t*, and $\vec{R}(0)$ is the position of the center of mass (at the start of the simulation) as defined by the input PDB file. The force applied to the center of mass of the tagged atoms is then:

$$\vec{F} = -\vec{\nabla} U$$

and it is calculated through Hooke's law, from the force constant *k* and the elongation of the spring: $vt - (\vec{R}(t) - \vec{R}(0)) \cdot \vec{n}$.

In the course of our simulations, value of the force was stored every 3 ps, for a total of 833 force values during a 2.5 ns simulation. The work performed by this force could be calculated as well. Unfortunately, the system not being in equilibrium, the calculated work has little significance,

since it cannot be related to Free Energy changes. The calculation of Free Energy requires performing multiple (usually 10) simulations for each system and the applying Jarzinsky equality. Since our work consisted in the screening of more than 90 molecules (initially) the use of multiple simulations was unfeasible.

Force values though can bring qualitative insight on the behavior of the permeant in the different molecular environments. A typical progress of Force vs Position is shown in Figure 5-1.

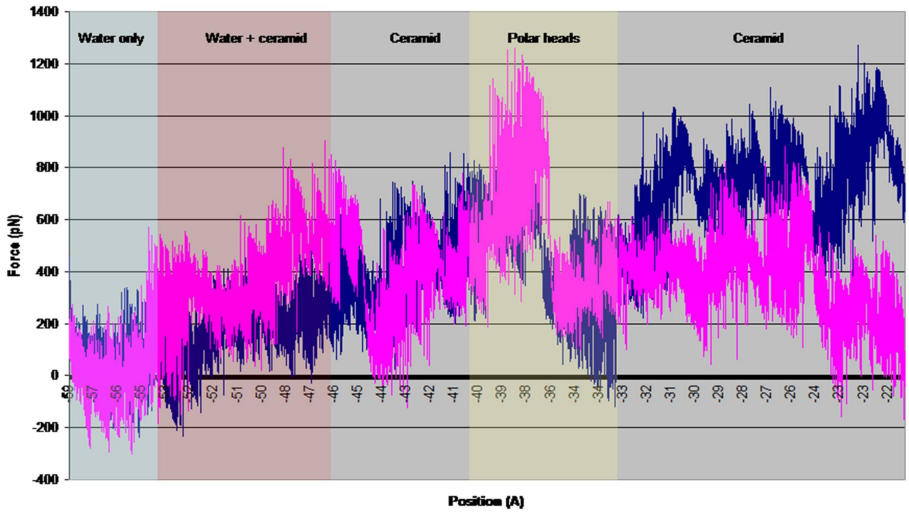

Figure 5-1. Force vs Position for Sucrose (blue) and Estradiol (magenta).

In Figure 5-1, it is clear that the hydrophilic molecule encounters more resistance in the ceramide phases (more exactly ceramide + cholesterol and ceramide + free fatty acids phases), while the hydrophobic molecule encounters a resistance peak in the polar heads region. Since the path imposed on the permeant during the *in silico* simulation, though, is not

necessarily the path followed by the molecule *in vitro*, these "obvious" energy barriers are not real, and could not be correlated to permeability coefficients.

5.2.1 Free energy calculations from multiple trajectories

For three molecules, namely pentanoic acid, progesterone and styrene, multiple trajectories were obtained. The goal was to study the effect of a (admittedly large) perturbation in initial conditions. The perturbation only affected the initial x-y position of the permeant's center of mass (COM). In the Trajectory 1 the permeant's COM is placed at x = 0, y = 0 at the beginning of the simulation. Trajectory 2 replicates the same initial position as Trajectory 1. In the other trajectories, it is displaced by 1 Å at a time along x and along y. The starting points for trajectories 2 to 5 are so (in Å):

- trajectory 3: x = 0, y = 1
- trajectory 4: x = 1, y = 0
- trajectory 5: x = 1, y = 1

Even though the perturbation seems large, it should be noted that the permeants undergoes an erratic movement in water at the beginning of the simulation, for 10-15 Å. So the point of insertion into the SC is quite unpredictable. This is particularly true for small molecules, whose low linear moment makes the movement in water highly erratic. Though the trajectory starting at (0, 0) is taken as a reference, it has no special meaning, as it can lead to unusable pathways at the edges of the simulation box, as in the case of 4-bromophenol (Figure 5-2), for which the trajectory starting (0, 0) clearly

ends in a region outside the model membrane (which goes approximately from -34 Å to +32 Å in the x direction, and from -35 Å to +35 Å in the y direction).

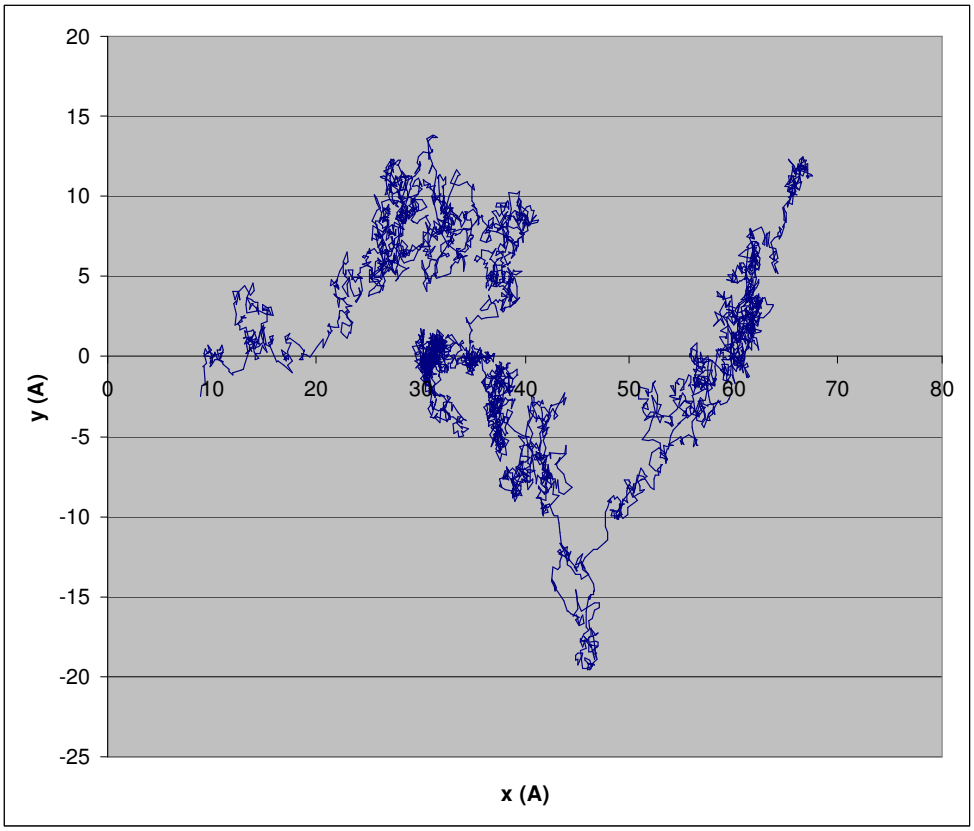

Figure 5-2. x (Å) vs. y (Å) plot for the trajectory starting at (0,0) for 4-nitrophenol.

In Figures 5-3, 5-4, 5-5, the x-y plots of the trajectories for pentanoic acid, progesterone and styrene, respectively are depicted. In the case of progesterone and styrene, trajectories 3 and 5 (yellow and purple lines) are clearly outside the model membrane, and have to be omitted from further analysis.

The availability of multiple trajectories, allows us to use Jarzinsky equality for the calculation of Free Energy differences (in our case Helmholtz Free Energy), from the work done by the SMD force in the different simulations. Jarzinsky equality states that:

$$e^{-\Delta F / kT} = \left\langle e^{-W / kT} \right\rangle$$

where ΔF is the Free Energy difference associated with the trajectory, W is the work done by the SMD force for the same process, k is Boltzmann's constant, T the absolute temperature. The average is taken over an ensemble, in our case over the different trajectories.

In Figures 5-6, 5-7 and 5-8, the work W is depicted for the different trajectories for the 3 molecules under study, as well as the Free Energy difference ΔF. From the plots, it is clear that in most cases the work calculated from SMD simulations is higher than the Free Energy difference, confirming that the simulations are away from equilibrium. Work calculated in our simulations cannot be used as such to estimate Free Energy.

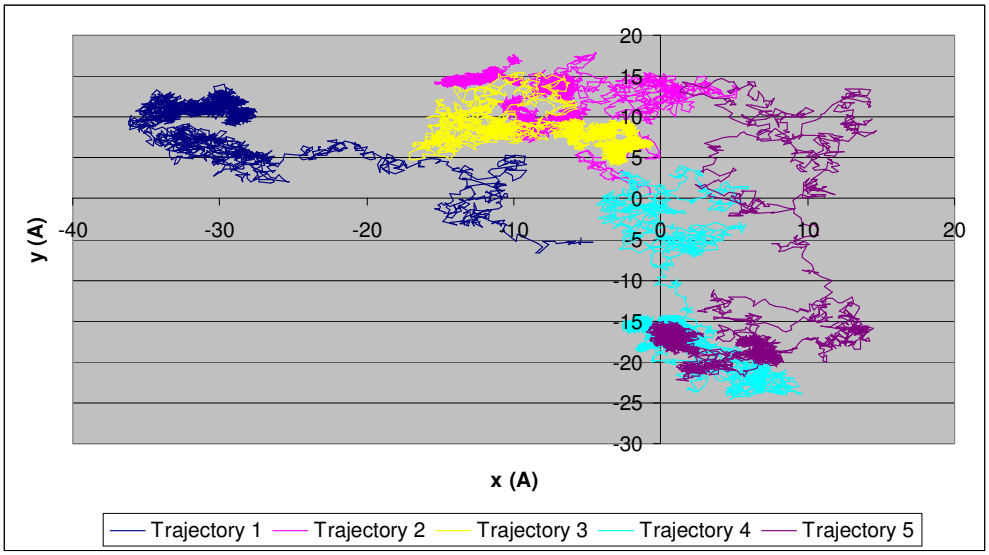

Figure 5-3. x (Å) vs. y (Å) plot for the 5 trajectories for pentanoic acid.

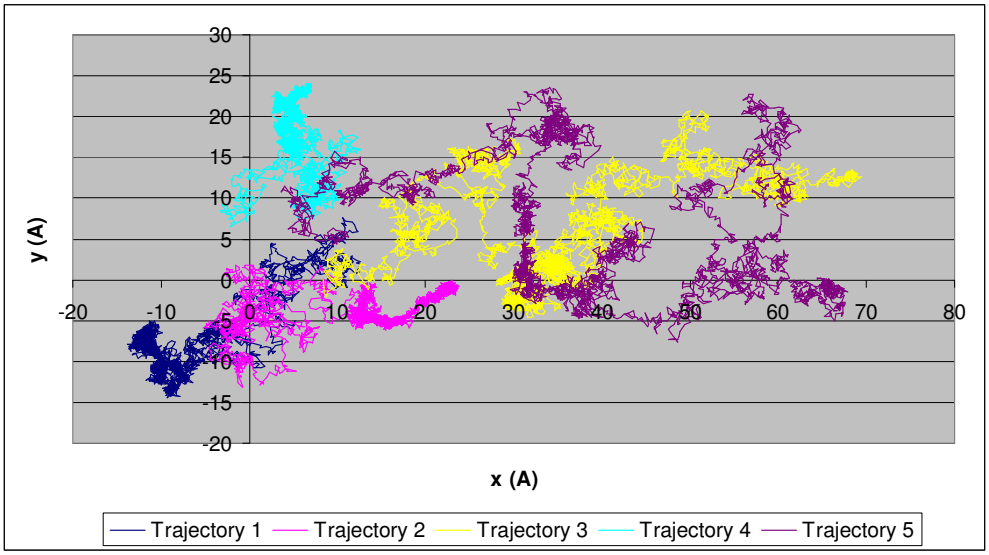

Figure 5-4. x (Å) vs. y (Å) plot for the 5 trajectories for progesterone.

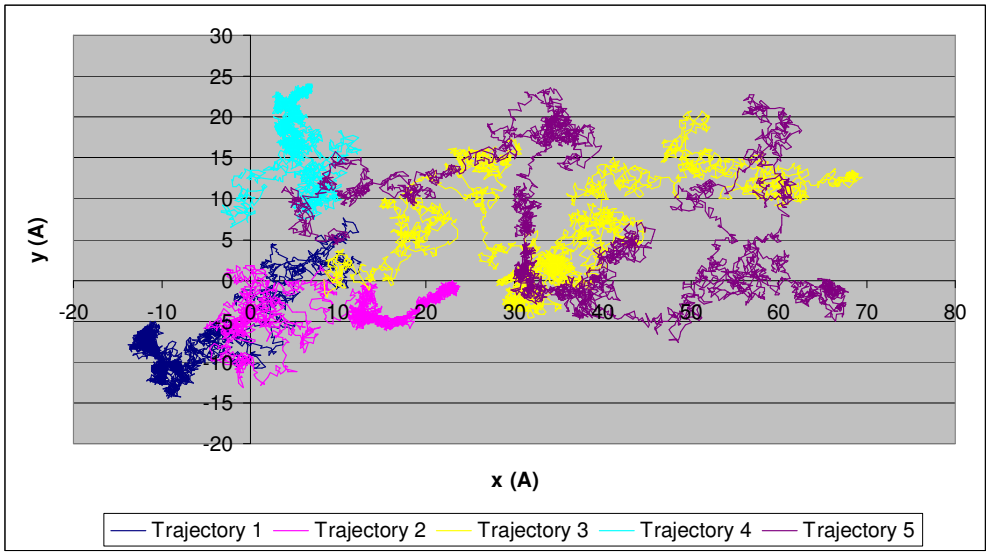

Figure 5-5. x (Å) vs. y (Å) plot for the 5 trajectories for styrene.

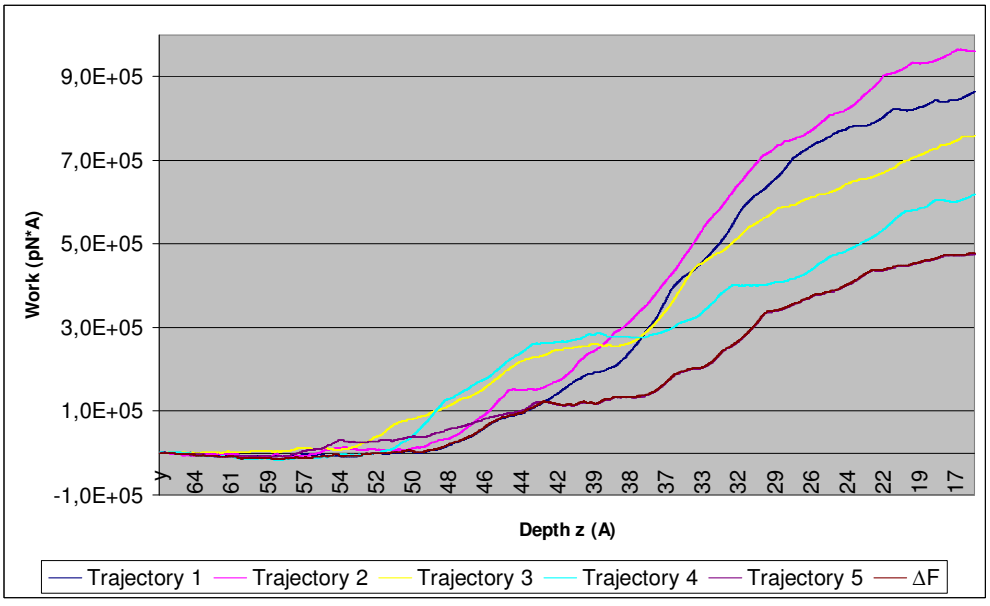

Figure 5-6. Work and Free Energy difference plot for pentanoic acid.

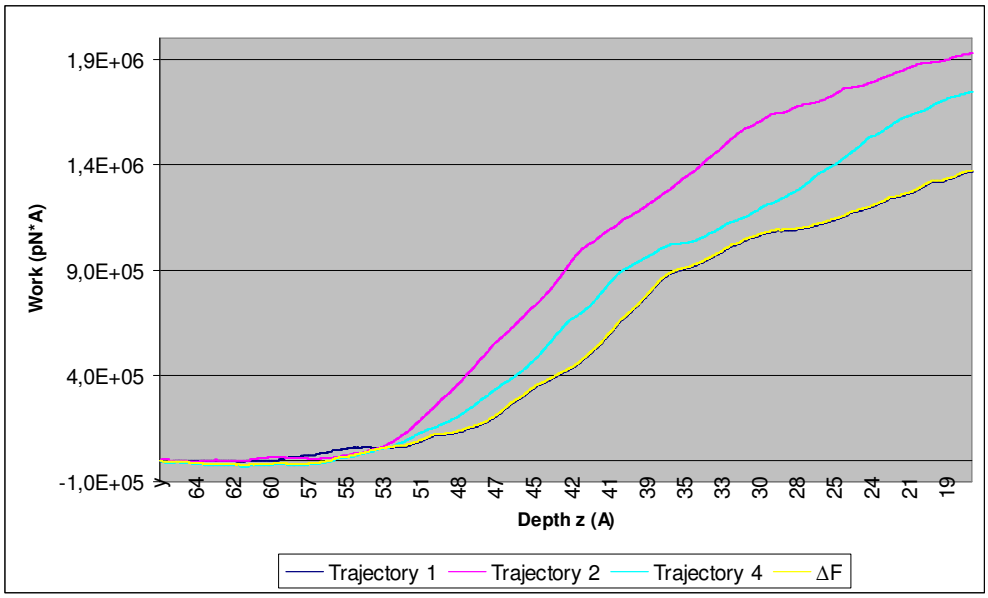

Figure 5-7. Work and Free Energy difference plot for progesterone.

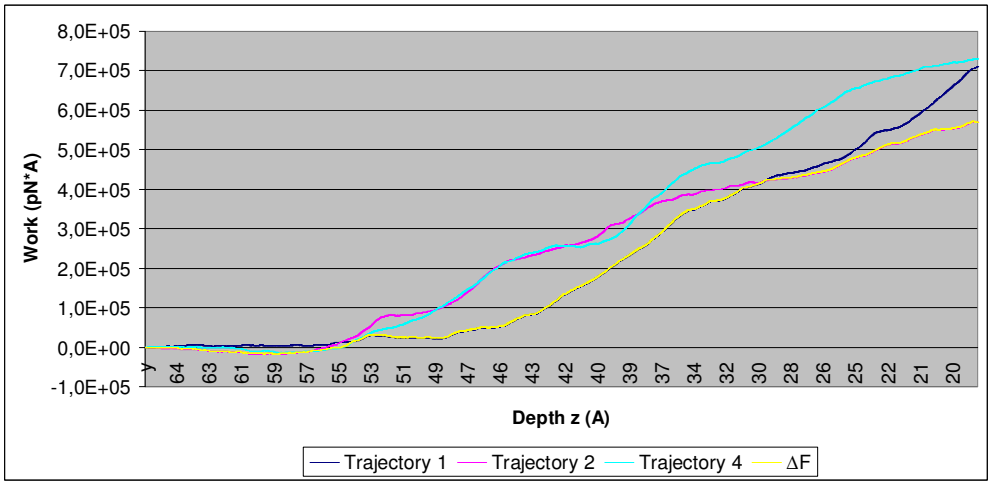

Figure 5-8. Work and Free Energy difference plot for styrene.

5.3 Molecular Dynamics simulations of solute permeation: exploration of conformational and property space

The concept of a property space, as the space of conformer-dependent properties experienced by a molecule when exploring its conformational space, was introduced by Bernard Testa and colleagues in [91], and later applied to the exploration of lipophilicity space in [82]. Fundamental in the development of the concept of property space were the assumptions that for a conformation-dependent physicochemical property:

1. a flexible molecule will assume different values dependent on the conformation;
2. the average value of a property, particularly a weighted average (and also its range), is more indicative than any conformer-specific value.

MD simulations represent a viable technique for the exploration of the conformational space of a molecule in a specific, and explicitly described, environment. In our particular case, and SMD simulation spanning the length of a SC monolayer, allowed the exploration of the conformational spaces of the 80 permeants in the reduced set (Appendix A). If conformation-dependent physicochemical properties are calculated at regular intervals, this results in a concurrent exploration of property space in the same environment.

Different molecular properties can be calculated during a SMD simulation. The "static" properties are calculated for example on optimized geometries in vacuum with *ab-initio* or *semi-empirical* quantum mechanics calculations. Molecular Dynamics allows for the determination of

"dynamical" properties, i.e. the relevant properties are calculated for the set of molecular geometries encountered in the actual molecular environment.

5.3.1 Physicochemical properties

The physicochemical properties considered were conformation-dependent properties, in particular *virtual log P* (calculated by a *MLP* approach) [75], polar superficial area (*PSA*), lipole, surface (calculated with a probe of diameter 1.40 Å), ovality and frontal area. Even properties which have little dependence on conformation, like molecular volume (*MV*) or no dependence, like molecular weight (*MW*) were considered. As the partition coefficient of a molecule plays a key role among physicochemical properties, a detailed account is given on the method used to estimate its value.

5.3.2 MLP and virtual log P

"The MLP defines the influence of all lipophilic fragmental contributions of a molecule on its environment. The MLP value at a point in space is generated as the result of the intermolecular interactions between all fragments in the molecule and the solvent system, at that given point" [75].

In practice, MLP in a given point in space k is expressed (and calculated) as the product of the lipophilic constant of a fragment i and a distance function $fct(d_{ik})$, where d_{ik} is the distance between fragment i and point space k, integrated over all the fragments in the molecule:

$$MLP_k = \sum_{i=1}^{N} f_i fct(d_{ik})$$

where *N* is the total number of fragments in the molecule. In particular, the MLP used is based on the atomic lipophilic system of Broto and Moreau and on an exponentially decaying distance function:

$$MLP_k = \sum_{i=1}^{N} f_i e^{-d_{ik}/2}$$

To obtain a quantity that can be correlated with experimental log P, the MLP_k values so obtained are integrated, either by summation over all MLP values on the surface ($\sum MLP$) or by summation of the positive MLP values ($\sum MLP^+$, representing the "lipophilic" part of the molecule) and the negative MLP values ($\sum MLP^-$, representing the "hydrophilic" part of the molecule) separately. These three parameters, $\sum MLP$, $\sum MLP^+$ and $\sum MLP^-$ are dependent on the tridimensional structure of the molecules, and the log P generated by the MLP through linear correlation with experimental log P values, called the "virtual log P" is dependent on molecular conformation [75].

5.3.3 Frontal area

One the quantities that showed some correlation to the permeation coefficient, was the Frontal area, i.e. the projection of the molecule in the *xy* plane, calculated as *Volume/z-dimension*. A frontal area so calculated gives a

good approximation in the case of cylindrically molecule, but it gives an underestimated measure in the case of a spherically shaped molecule. In the latter case, the z-dimension is the sphere diameter, so:

$$Estimated\ \ frontal\ area\,(sphere) = \frac{4}{3}\pi R^3 \Big/ 2R = \frac{2}{3}\pi R^2$$

which is *2/3* of the true frontal area. A better estimation of the frontal area is obtained by multiplying the value of *Volume/z-dimension* by the quantity *3/(2*ovality)*, which has value *3/2* for a sphere and approximately 1 for a cylinder.

5.3.4 Scoring functions

A different approach consisted in the evaluation of the SC-permeant interaction energy through a scoring function. Five different scoring functions were tested, of which four are variations of the MLP Interaction Score [92]: MLP_{InS}, MLP_{InS2}, MLP_{InS3}, MLP_{InSF} and CHARMM.

"The **MLP Interaction Score** (MLP_{InS}) is computed using the atomic fragmental system proposed by Broto and Moreau and a distance function that defines how the score decrease with increasing distance between interacting atoms. The equation to compute the interaction score is:

$$MLP_{InS} = \sum_p \sum_m - \frac{(f_a - f_b)}{f(r_{ab})}$$

where f_a and f_b are the lipophilicity increments for a pair of atoms and r_{ab} is the distance between them. The first sum is over all the ligand's atoms, the second sum is over all receptor's atoms.

The basic assumption in the calculation of the MLP_{InS}, which encodes the contributions of the various intermolecular forces measured experimentally in partition coefficients, is that the score is favourable (i.e. negative) when both increments have the same sign (as denoted by the negative sign in the equation), or unfavourable (repulsive forces) when the score has a positive sign. When the atomic parameters are both positive, MLP_{InS} encodes hydrophobic interactions and dispersion forces, the importance of which is well recognized in docking simulations, and it accounts for polar interactions, in particular H-bonds and electrostatic forces when the atom parameters are both negative" [92].

The tested MLP Interaction Scoring functions differ in the calculation of hydrophobic interaction:

- in MLP_{InS}, hydrophobic interaction is calculated as product of the Broto's and Moreau's atomic constants divided by the distance between the interacting atom pair;
- in MLP_{InS2}, as above, but the distance between interacting atom pairs is considered as a square value;
- MLP_{InS3}, as above, but the distance between interacting atom pairs is considered as a cube value;
- MLP_{InSF}, as above, but the distance is evaluated by Fermi's equation.

In the CHARMM scoring function, Van der Waals interactions are estimated by CHARMM 22 force field.

5.4 On the confidence interval of the correlation coefficient r

Given the uncertainty in the experimental data contained in Flynn's data set and even in the Fully validated data set (Paragraph 5.4.1), it is important to estimate a confidence interval for r or r^2 and to know if there is a theoretical maximum correlation coefficient for a set of data with given uncertainties.

In determining the confidence interval of r, where r is seen as an estimate of a true value ρ, a first analytical method applied the approach devised by R.A. Fisher [93], using the Fisher $r \rightarrow Z$ transform [94][95], which is defined as:

$$Z = \frac{1}{2}[ln(1+r) - ln(1-r)] \qquad \textbf{Equation 5-4}$$

The transformed variable Z has a different behaviour than r. It tends to a normal distribution as the number of data (n) becomes larger, and its variance σ_Z^2 is independent on the value of r:

$$\sigma_Z^2 = \frac{1}{n-3} \qquad \textbf{Equation 5-5}$$

To get a confidence interval for *r*, the value of *r* is transformed into *Z*, a confidence interval is calculated for *Z* and it is transformed in the confidence interval of *r*, using the reverse transform:

$$r = \frac{e^{2Z} - 1}{e^{2Z} + 1} \qquad \textbf{Equation 5-6}$$

Results obtained through Fisher $r \rightarrow Z$ transform are independent of the uncertainty of experimental data and require the *r* value is known *a priori*. Even though the r^2 mean of a correlative analysis based on a dataset characterized by a known standard deviation of the errors, ε, is roughly equal to $1 - \varepsilon$, a numerical simulation was performed to better investigate the correlation between the confidence interval of *r* and the uncertainty of experimental data. As described below, such a simulation involved the introduction of increasing random errors into the experimental *pKp* data to analyse their influence on the correlation coefficient. Moreover, by repeating many times the numerical simulation, it is possible to derive both the *mean* r^2 and the corresponding a confidence interval as defined by the corresponding minimum and maximum of r^2 values. Interestingly, these numerically derived results will be compared to those calculated by using the Fisher transform. Both methods are applied to the 80 compounds in the reduced set and compared in Table 5-1 (r^2 values are given).

In the simulation, y_i are the experimental *pKp* (= *-logKp*) for the 80 compounds in the data set:

$$y_i = -logKp_i \qquad i = 1,2 \ldots n$$

We suppose to have a perfect estimator φ for y_i, so that given a set of variables $\{x_{ij}\}$, where x_{ij} represents the j^{th} molecular property of the i^{th} molecule, the equation:

$$y_i = \varphi(x_{ij}) \qquad i = 1,2 \ldots n, j = 1, 2 \ldots n$$

Equation 5-7

brings a correlation coefficient $r = 1$.

For every y_i, we now introduce an error, $\varepsilon \cdot c_i \cdot y_i$, where the c_i values are normally distributed pseudo-random numbers with zero average and unitary standard deviation, and ε corresponds to the standard deviation of the errors, normalized by y_i. Equation 5-7 becomes:

$$y_i (1 \pm \varepsilon c_i) = \varphi(x_{ij}) \qquad i = 1, 2 \ldots n, j = 1, 2 \ldots n$$

Equation 5-8

which has a correlation coefficient $r < 1$. For increasing values of ε (from 0.1 to 0.5), we repeated the simulation 99 times by applying Equation 4-6 to the same set $\{y_i, x_{ij}\}$, each time changing the set $\{c_i\}$. Since the so calculated correlation coefficients (r_k) are not normally distributed, we apply Fisher $r \rightarrow Z$ transform to obtain 99 Z_k values, for which the standard deviation is calculated:

$$S_z = \frac{1}{99} \sum_{k=1}^{99} (Z_k - \overline{Z})^2 .$$

Once the confidence interval for *Z*, has been calculated as $(\overline{Z} - S_z, \overline{Z} + S_z)$, the reverse transform is used to get a confidence interval for *r* (Figure 5-9). Results of the simulation, performed on the 80 compounds in the reduced data set are listed in Table 5-1.

The normally distributed pseudo-random numbers are obtained starting from random numbers and applying the Box-Muller transform. Given a set of pairs of random numbers $\{u_i, u_j\}$ in the interval (0, 1), the set $\{c_i, c_j\}$, where c_i, c_j are defined by:

$$c_i = \sqrt{-2lnu_i}\cos(2\pi o_j)$$
$$c_j = \sqrt{-2lnu_i}\sin(2\pi i_j)$$

The Box-Muller transform allow to obtain a set of normally distributed pseudo-random numbers, with zero average and unitary standard deviation.

ε	**0.10**	**0.15**	**0.20**	**0.25**	**0.30**	**0.35**	**0.40**	**0.45**	**0.50**
r^2 mean simulated	0.95	0.88	0.81	0.74	0.66	0.59	0.53	0.47	0.42
r^2 lower simulated	0.93	0.86	0.77	0.68	0.58	0.50	0.43	0.36	0.31
r^2 higher simulated	0.96	0.91	0.85	0.79	0.73	0.68	0.62	0.57	0.53
r^2 lower calculated	0.93	0.86	0.77	0.69	0.60	0.52	0.45	0.39	0.34
r^2 higher calculated	0.96	0.91	0.85	0.79	0.72	0.66	0.60	0.55	0.51

Table 5-1. Mean and confidence interval for r^2 obtained from numerical simulation (simulated), corresponding to different values of standard deviation of *pKp* (ε = Std. dev. of *pKp* / average *pKp*). Mean and Confidence interval for r^2 (calculated) obtained from Equations 5 and 6 corresponding to the simulated r^2 mean value

Results in Table 5-1 show how the correlation worsens when the uncertainties on experimental data increase even considering an ideally perfect estimator of for y_i. Mean r^2 values decrease as an S-shaped function

as ε increases (Figure 5-9). As expected and to a first approximation mean r^2 mean values can be estimated as

$$r^2 = 1 - \varepsilon$$

which implies that even a perfect estimator for *pKp* cannot explain all the variability associated in the experimental data, given their uncertainties.

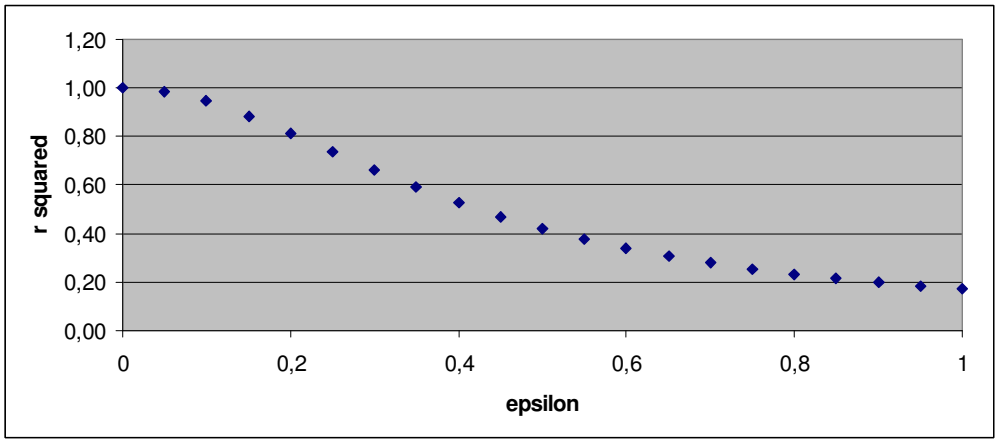

Figure 5-9. Mean r^2 values vs. different values of standard deviation of *pKp* (*epsilon* = Std. dev. of *pKp* / average *pKp*).

5.4.1 Uncertainty in experimental data and repercussions of the correlation coefficient

To get an idea of the typical uncertainties of the *pKp* values in the existing datasets, we consider compounds for which multiple (more than 2) measurements are included in the FV data set (Figure 5-2), and calculate the standard deviation of *pKp* normalized by the *pKp* average. This term can then be compared to ε from Equation 5-8. As it can be seen from Table 5-2, values

of ε of 0.25 or greater are possible. Results from the simulation are shown in Table 5-3.

Corticosterone (n=4)	**pKp**	**t (°C)**
Average	3.67	29.25
St. Dev.	0.94	6.55
St. Dev./Average	**0.26**	0.22

Phenol (n=3)	**pKp**	**t (°C)**
Average	2.53	28.00
St. Dev.	1.12	7.94
St. Dev./Average	**0.44**	0.28

Estradiol (n=5)	**pKp**	**t (°C)**
Average	2.52	31.00
St. Dev.	0.58	4.00
St. Dev./Average	**0.23**	0.13

Ethanol (n=3)	**pKp**	**t (°C)**
Average	3.37	25.67
St. Dev.	0.24	4.04
St. Dev./Average	0.07	0.16

Mannitol (n=4)	**pKp**	**t (°C)**
Average	4.09	31.50
St. Dev.	0.12	5.20
St. Dev./Average	0.03	0.16

Water (n=8)	**pKp**	**t (°C)**
Average	2.96	29.88
St. Dev.	0.20	2.10
St. Dev./Average	0.07	0.07

Salicylic acid (n=3)	**pKp**	**t (°C)**
Average	1.72	34.67
St. Dev.	0.43	4.04
St. Dev./Average	**0.25**	0.12

Octanol (n=3)	**pKp**	**t (°C)**
Average	1.26	25.67
St. Dev.	0.04	4.04
St. Dev./Average	0.03	0.16

Table 5-2. Average and standard deviation for *pKp* and temperature (*t*, °C) based on data included in the Fully Validated data set and published in [81]**.**

ε	**0.05**	**0.1**	**0.15**	**0.2**	**0.25**	**0.3**	**0.4**
***r* mean**	0.993	0.973	0.94	0.90	0.86	0.82	0.73
***r* low**	0.992	0.968	0.93	0.88	0.83	0.77	0.66
***r* high**	0.995	0.979	0.96	0.93	0.89	0.86	0.79
r^2 **mean**	0.986	0.948	0.89	0.82	0.74	0.67	0.53
r^2 **low**	0.984	0.936	0.87	0.78	0.69	0.60	0.44
r^2 **high**	0.989	0.959	0.91	0.86	0.80	0.74	0.63

Table 5-3. Mean and Confidence interval for *r* and r^2 obtained through numerical simulation, corresponding to different values of standard deviation of *pKp*. ε = (Std. dev. of *pKp*) / (average *pKp*).

There is a correlation (r^2 = 0,64) between the normalized standard deviation of *pKp* and the normalized standard deviation of experimental temperatures, suggesting that the effect of temperature should be taken into consideration for a better correlation of *pKp* to the molecular properties, either in the form of a correction factor or performing the MD simulations at the same temperature as the experiment.

5.5 On the linear filtering of differential and exponential quantities.

When dealing with quantities with a rich higher harmonics content, such as position and force data derived from Molecular Dynamics simulations, a certain degree of filtering seems necessary. A word of caution, though, is due if linear filtering is to be applied to quantities, such as position, which are

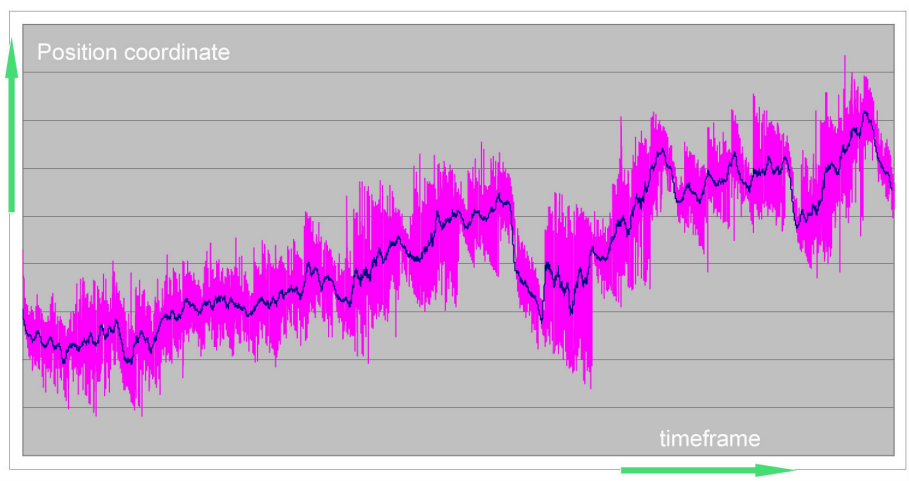

Figure 5-10. Position coordinate of the center of mass of a molecule in a SMD simulation (magenta) and a 37 period MA (blue).

later manipulated to obtain differential quantities, such as velocities (and so even the Mean Square Deviation, MSD, used to derive the Diffusion coefficient) or exponential quantities such as kinetic energy.

The hypothesis when applying filters for the purpose of reducing the casual error in calculation/measurement, is that the error is indeed casual, with zero average. A typical signal from an SMD simulation, for example a position coordinate (but it could be a linear momentum or a force), with 37 period Moving Average (MA) applied to it as a filter is shown in Figure 5-10.

The MA is defined as:

$$MA_p(x_i) = \frac{1}{p} \sum_{j=i-(p-1)/2}^{i+(p-1)/2} x_j$$

The value x_i can be thought of as a true value $x_{i,TRUE}$ plus a casual error ε_i, which can be positive or negative:

$$x_i = x_{i,TRUE} + \varepsilon_i$$

Since ε_i, $\sum_1^N \varepsilon_i \to 0$ as $N \to \infty$ then, if N is sufficiently large:

$$\sum x_i = \sum x_{i,TRUE} + \sum \varepsilon_i \approx \sum x_{i,TRUE}$$

and so, from the definition of *MA*:

$$MA(x_i) = MA(x_{i,TRUE}) + MA(\varepsilon_i) \approx MA(x_{i,TRUE})$$

which allows as to take the *MA* of x_i as an estimate of the true value $x_{i,TRUE}$. Then, the casual error ε_i can be calculated as:

$$\varepsilon_i = x_i - x_{i,TRUE}$$

as depicted in Figure 5-11.

The average error $\frac{1}{N}\sum_{1}^{N}\varepsilon_i$ clearly goes to zero as the number of timeframes increases (Figure 5-12).

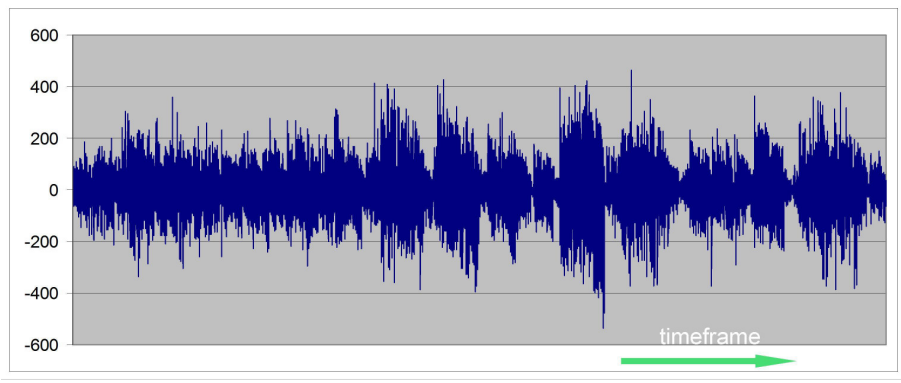

Figure 5-11. Estimation of the casual error ε_i in the determination of the position coordinate x_i of the center of mass of a molecule in a SMD simulation.

If we take the square of the coordinate from the SMD simulation x_i^2, as in the case of calculating kinetic energy from the linear momentum, and the

square of the *MA* of x_i as an estimate of its true value $x^2_{i,TRUE}$, clearly the error $\varepsilon^*_i = x^2_i - [MA(x_i)]^2$ shown in Figure 5-13, is not casual and it does not have in general zero average as the number of timeframe increases (Figure 5-14).

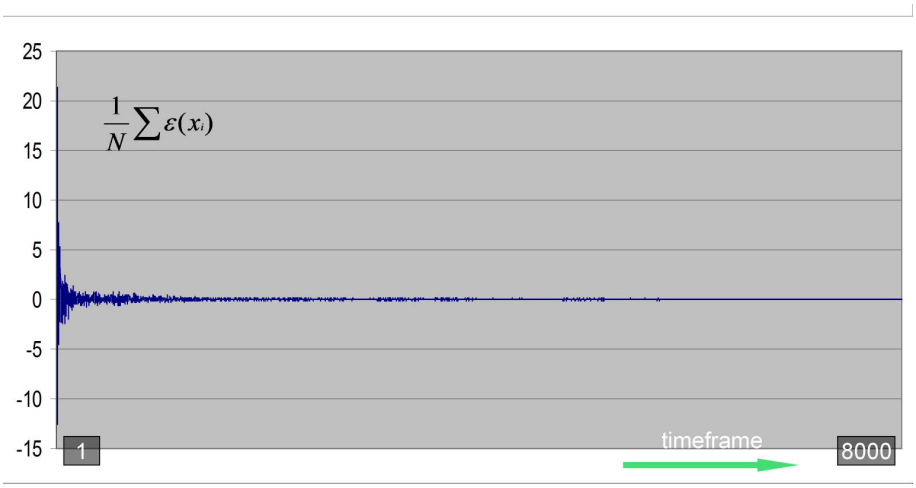

Figure 5-12. Average of the casual error ε_i in the determination of the position coordinate x_i of the center of mass of a molecule in a SMD simulation.

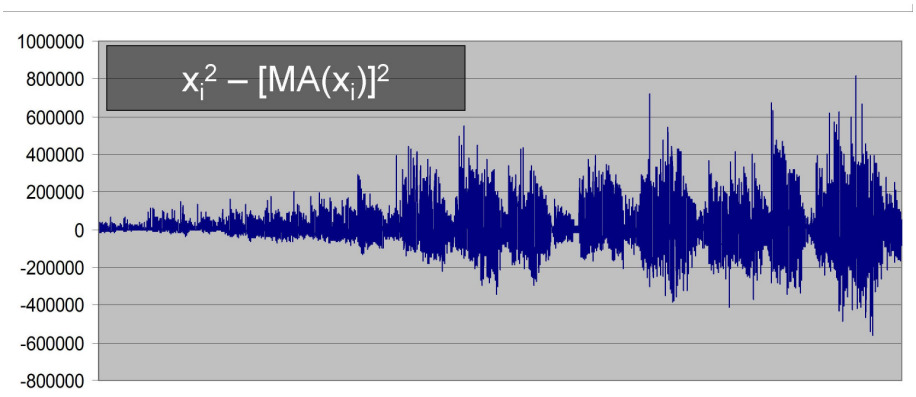

Figure 5-13. Error $\varepsilon^*_i = x^2_i - [MA(x_i)]^2$ in the determination of the square of coordinate x_i in a SMD simulation.

Indeed, as $x_i = x_{i,TRUE} + \varepsilon_i$ then:

$$x_i^2 = x_{i,TRUE}^2 + 2\varepsilon_i x_{i,TRUE} + \varepsilon_i^2$$

where ε_i is the casual error. If we take $[MA(x_i)]^2$ as an estimate of $x_{i,TRUE}^2$, then:

$$\varepsilon_i^* = x_i^2 - [MA(x_i)]^2 = 2\varepsilon_i [MA(x_i)] + \varepsilon_i^2$$

and the average of ε_i^*:

$$\frac{1}{N}\sum_1^N \varepsilon_i^* = \frac{2}{N}\sum_1^N \varepsilon_i MA(x_i) + \frac{1}{N}\sum_1^N \varepsilon_i^2$$

clearly does not tend to zero as *N* increases (Figure 5-14).

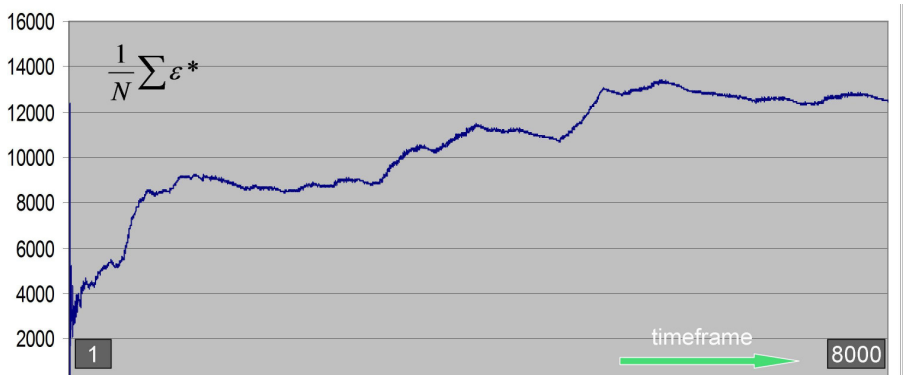

Figure 5-14. Average of error $\varepsilon_i^* = x_i^2 - [MA(x_i)]^2$ in the determination of the square of coordinate x_i in a SMD simulation.

In the case of a differential quantity $\Delta x_i = x_{i+1} - x_i$, such as velocity if x_i is a position coordinate (Figure 5-15), if we filter the original coordinates x_i and x_{i+1} with a *MA* of period *p*, then we get:

$$\Delta x_i = MA_p(x_{i+1}) - MA_p(x_i) = x_{i+(p+1)/2} - x_{i-(p-1)/2}$$

Equation 5-9

So we are simply replacing a difference between adjacent *x* values, with a difference between *x* values at the extremes of the *MA* period. This only means a loss of information instead of filtering. Since *MA* is a linear operator, applying the *MA* to Δx_i instead of applying it to x_i and x_{i+1} brings the same result.

If we apply a multiple pass moving average, i.e. we apply a MA of period p to Δx_i as calculated in Equation 5-9, we get:

$$\begin{aligned} &MA_p(\Delta x_i) = MA_p(MA_p(x_{i+1}) - MA_p(x_i)) = \\ &= MA_p(MA_p(x_{i+1})) - MA_p(MA_p(x_i)) = \\ &= MA_p(x_{i+(p+1)/2}) - MA_p(x_{i-(p-1)/2}) \end{aligned}$$

So we are replacing a difference between adjacent *x* values, with a difference between moving averages centered at the extremes of the first *MA* period.

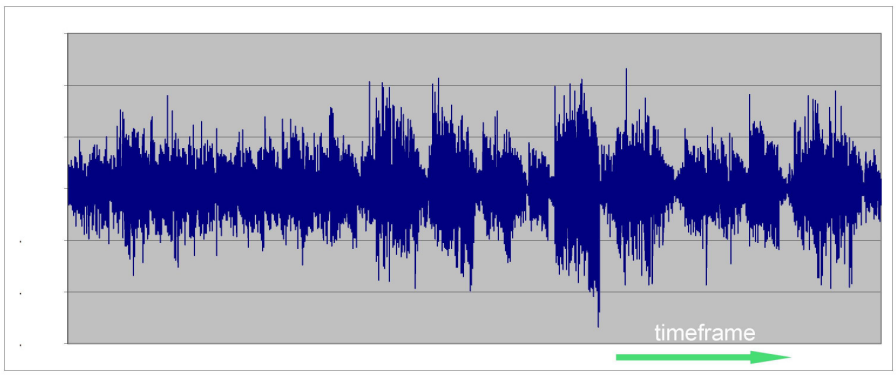

Figure 5-15. Differential quantity $\Delta x_i = x_{i+1} - x_i$ derived from coordinate x_i.

Other linear filters, such as Savitsky-Golay filters, bring different results, but there is always a loss (albeit smaller) of information around the values to be subtracted, and their replacement with a linear combination of neighboring values. A Savitzky-Golay filter has the general form:

$$SG(x_i) = \frac{1}{N}\sum_{j=1}^{m} c_j x_j$$

where integer coefficients c_j have different values based on the width and polynomial degree of the filter. For example, in a quadratic filter of width 5, the coefficients are: -3, 12, 17, 12, -3.

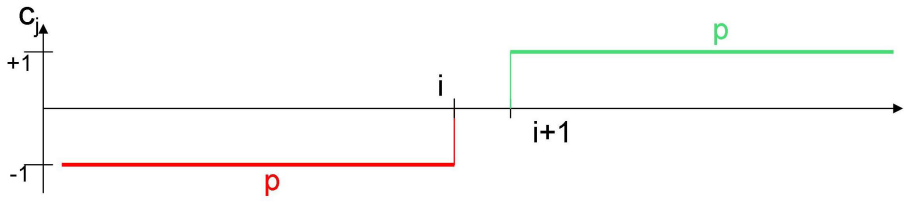

Figure 5-16. Coefficients c_j of a Multiple (Double) Pass Moving Average of period p applied to the differential quantity $\Delta x_i = x_{i+1} - x_i$.

In the case of the moving average, the coefficients of the linear combination are uniformly distributed (Figure 5-16), while in the case of Savitsky-Golay filters values far from the original coordinates to be subtracted have the lowest weight (Figure 5-17).

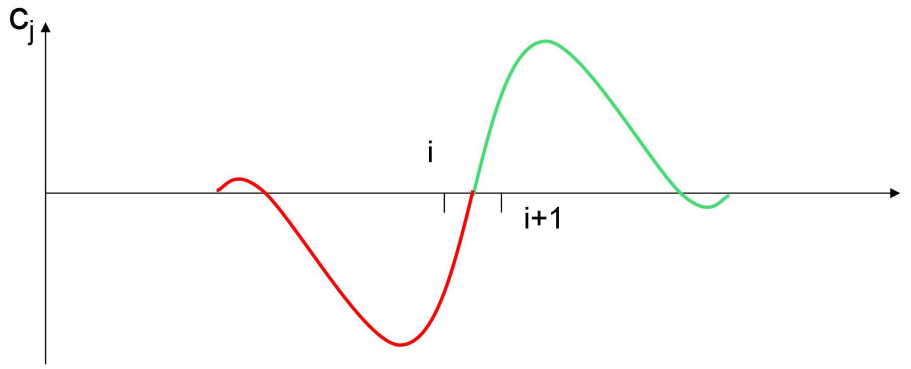

Figure 5-17. Qualitative trend of the coefficients c_j of a Savitzky-Golay filter applied to the differential quantity $\Delta x_i = x_{i+1} - x_i$.

An advantage of Savitzky-Golay filters over *MA*s in the case of differentials though, is that the filter can be applied to the signal to calculate directly the 1st derivative, thus avoiding the loss of information.

The loss of information could be resolved in different ways, for example using non-linear filters, or applying the 1st derivative Savizky-Golay filter. We chose the simplest method of taking the median of a value (be it position, or a molecular property) inside the relevant Stratum Corneum region.

6 Predictive analyses

In this chapter, the results obtained in the search of a predictive equation of skin permeability to chemicals are shown. The chapter starts with the preliminary explorations of semi-empirical equations, based on the Potts and Guy equation, without the use of Molecular Dynamics, which can be refined to give results at the limit of what is possible given the uncertainty in experimental data (Paragraph 5.4) at the risk of overfitting the available data (n = 80). Then, results obtained from MD simulations are given. Two different path have been attempted:

1. the use of position and force data from SMD simulations. The use of force data to calculate work and then Free Energy differences had to be discarded, because our simulations are to far from equilibrium (Paragraph 5.2.1) to yield correct Free Energy values. Moreover, since our work consist of a screening of a large number of molecules performing multiple simulations (from 5 to 10 at least) proved unfeasible. Position data, on the other hand were used to calculate Diffusion Coefficients.
2. the use of MD simulations as a mean to explore the conformational space and property space of the permeants, in a real-world scenario, i.e. on a reduced conformational and property subspaces, that the molecules are likely to explore during the process of Stratum Corneum permeation.

In practice, in the best predictive equations, we mixed molecular properties from step 2 with diffusion coefficients from step 1.

All correlations in the present chapter (except for the Potts and Guy equation) are performed on the reduced data set defined in Appendix A.

6.1 Predictive models based on Potts & Guy equation

The most cited semi-empirical model of permeability is the one by Potts and Guy [8], whose equation (correlated on Flynn's set of 94 permeants) is:

$$logKp\ (cm/s) = -6.3 + 0.71\ log\ K_{o/w} - 0.0061\ MW \qquad r^2 = 0.68$$

where Kp is the permeability coefficient expressed in cm/s, $K_{o/w}$ the octanol/water partition coefficient, MW the molecular weight and r the square of the correlation coefficient calculated on Flynn's set.

As pointed out in the previous chapter, a temperature correction factor may be added to a predictive equation, to account for the variability in experimental conditions both in Flynn's [10] and in the Fully Validated [81] sets. Introducing the effect of temperature on reducing the activation energy for transport of penetrants through the SC, Vecchia and Bunge [81] proposed the modified equation:

$$logKp = a + b \cdot log\ K_{o/w} + c \cdot MW/T \qquad \textbf{Equation 6-1}$$

where T is the absolute temperature, and a, b, c the coefficient to be determined empirically. It is notable that Equation 6-1 has the same degrees of freedom as the Potts & Guy equation. Applying Equation 6-1 to the reduced set (since not all experimental temperatures are known for Flynn's set), we get the following result, not significantly better than Potts & Guy's:

$$logKp\ (cm/s) = -6.22 + 0.62\ log\ K_{o/w} - 1.72\ MW/T$$

$$r^2 = 0.69,\ q^2 = 0.66,\ n = 80$$

where q^2 is the square of correlation coefficient obtained with the Leave One Out (LOO) algorithm, and it is a measure of the predictive power of the equation.

Surprisingly, replacing the absolute temperature T with temperature t in °C, r^2 and q^2 improve:

$$logKp\ (cm/s) = -5.97 + 0.64\ log\ K_{o/w} - 0.21\ MW/t \quad \textbf{Equation 6-2}$$

$$r^2 = 0.75,\ q^2 = 0.70,\ n = 80$$

Since this result relies only on a change in unit of measure, it seems suspect of overfitting of data, and it should be tested on larger data sets, such as the full Fully Validated data set. However, since the dependence of logKp on t is non-linear, and given that this is a peculiar change in unit of measure: $T\ (K) = t\ (°C) + 300$, this is not impossible from a strictly mathematical point of view. Predicted *logKp* values from Equation 6-2 vs experimental

logKp values are graphically represented in Figure 6-1. Leaving out the value for Oubain (experimental *logKp* = -9.67), we get $r^2 = 0.77$, $q^2 = 0.75$.

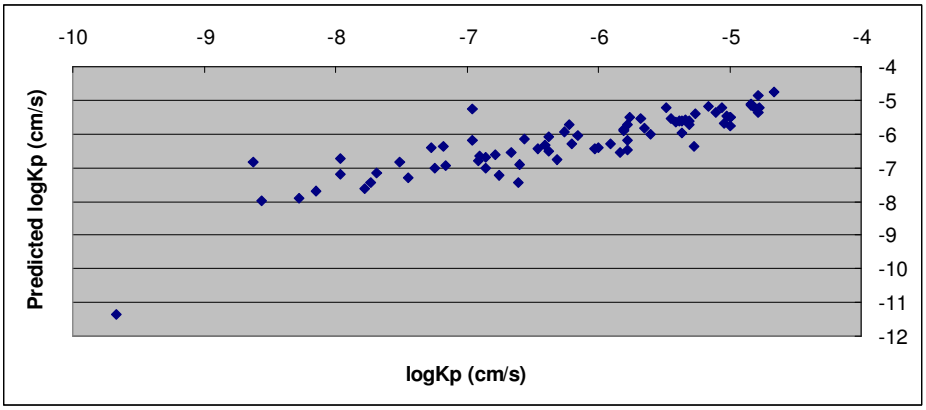

Figure 6-1. Predicted *logKp* values from Eq. 7-2 vs experimental *logKp* values from the common set between Flynn's and the Fully Validated sets.

Results improve to $r^2 = 0.80$, $q^2 = 0.74$ if pH corrected values from the Fully validated set are used. If Molecular Volume (*MV*), as calculated at the water/bilayer interface during MD simulations, replaces Molecular Weight (*MW*), we get another improvement:

$$logKp\ (cm/s) = -5.88 + 0.71\ log\ K_{o/w} - 0.24_5\ MV/t \quad \textbf{Equation 6-3}$$

$r^2 = 0.81$, $q^2 = 0.76$, $n = 80$

Predicted *logKp* values from Equation 6-3 vs experimental *logKp* values are graphically represented in Figure 6-2.

Validation of Eq. 6-3 brings to the results listed in Table 6-1. Equation 3 gives better results than all the equations considered in the previous chapter.

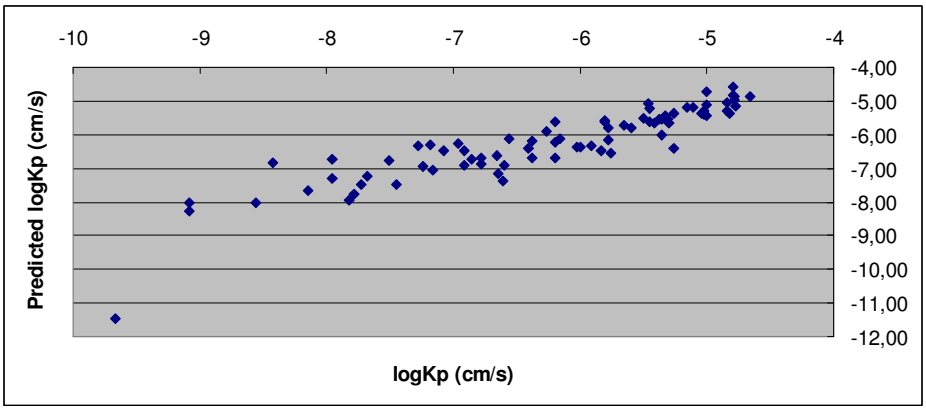

Figure 6-2. Predicted *logKp* values from Eq. 7-3 vs experimental *logKp* values from the common set between Flynn's and the Fully Validated sets.

	r^2 training set	Adjusted r^2 training set	r^2 test set	Std. Dev. of Errors training set
Eq. 6-3 (n = 80)	0,82±0,03	0,80±0,04	0,82±0,05	0,51±0,06

Table 6-1. Model validation for Eq. 7-3. Dimension of the test set = 27. Dimension of the training set = 53. Validation performed on 20 tests.

From the considerations on the confidence interval of the correlation coefficient r made in Appendix A, we are clearly reaching the limits of a significant r^2 value for n = 80, even with a very limited use of Molecular Dynamics (calculation of the Molecular Volume at the water/bilayer

interface). Applying Eq. 6-2 and Eq. 6-3, as well as a Potts & Guy type equation to a larger set would be essential to reduce the probability of overfitting and to reduce the confidence interval of correlation coefficient r, thus improving the significance of the calculated r^2 value. The role of Molecular Dynamics simulations clearly seems more oriented to understand the mechanics of skin permeation, more than to develop improved predictive equations. Though, the capability of MD to directly calculate quantities of physical significance, forms a protection from theoretical artifacts and from data overfitting.

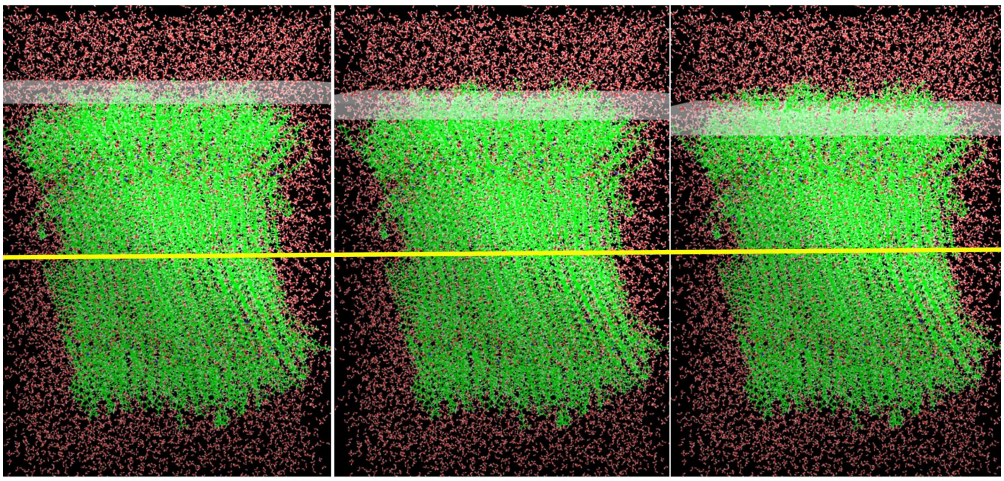

Figure 6-3. Regions 1, 3 and 5 as defined in Table 5-3 (semi-transparent plane). Yellow line represents depht $z = 0$.

6.2 Results from MD simulations

Even though results obtained with semiempirical equations are promising, they do not offer better insight into the mechanics of permeation the original Potts and Guy equation. Molecular Dynamics, on the other hand can help gain a better understanding of the process under study.

One way in which we exploited this characteristic of MD was to divide the lipid matrix model in 18 partially overlapping regions, to study the importance of any region in the explanation of experimental data variability.

The model membrane was ideally subdivided into 18 regions along the z direction, i.e. the bilayer normal (Figures 6-3, 6-4). The thickness of every region had to be chosen as a compromise: a region should not be so large that calculated quantities (permeability and diffusion coefficients) vary significantly, but it should be large enough to allow for a significant diffusion coefficient to be calculated from the RMSD. A thickness of 4 Å was chosen somehow arbitrarily, based on those considerations. Since the simulations span, at least from z = 60 Å (measured from the membrane midplane) to z = 20 Å, the correlations were performed on 4 Å frames starting from z = 58 Å, and sliding the frames by 2 Å at a time to z = 21 Å.

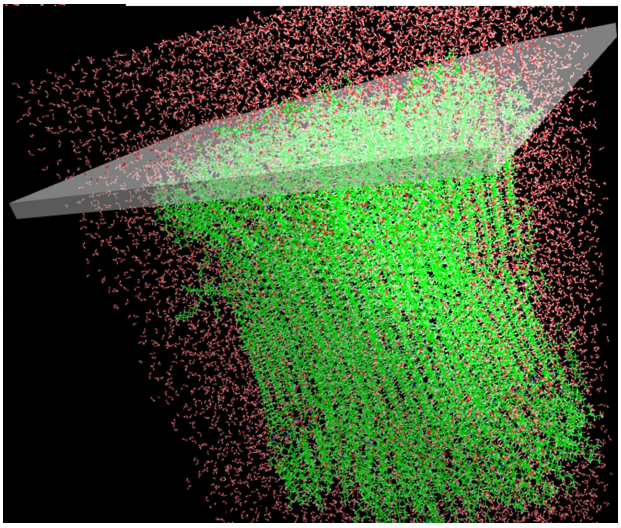

Figure 6-4. Tridimensional view of the model membrane intersected by the plane defining regions.

6.2.1 Diffusion coefficients

Both the diffusion coefficient in the *xy* plane, D_{xy}, and along the *z* axis, D_z, were tested, although the latter was affected by a high bias introduced by steering in the *z*-direction. Results confirmed that the values obtained for D_z from MSD calculations have little physical significance. Even if the SMD force is applied only in the *z* direction, it will also affect the calculation of the diffusion coefficient in the *xy* plane, since it will affect the linear momentum of the permeant and the frequency of collisions with SC molecules. Though not entirely unbiased, D_{xy} is less affected the by the forced movement in the *z*-direction, and maintains a physical significance showing a better correlation coefficient than D_z when applied to the reduced data set, at least in some regions (Table 6-2). It must be noted that the correlation coefficient is highly dependent on the region where the diffusion coefficient is calculated.

Quantity	Region	r^2
log D_{xy}	3	0.20
log D_{xy}	2	0.12
log D_{xy}	5	0.11
log D_{xy}	4	0.10
log D_z	14	0.09

Table 6-2. Contribution of *log* D_{xy} vs *log* D_z to the explanation of the variability of the common data set.

6.2.2 Physicochemical properties

The correlation of different physicochemical properties of the permeant, calculated in the different regions of the SC, with skin permeation have been explored. The analysis starts with virtual log P, as the partition coefficient, not surprisingly, plays a pivotal role in many predictive equations, starting from Potts and Guy's.

6.2.2.1 Virtual log P

The values of virtual log P have been calculated for the 80 compounds in the 18 regions of the stratum corneum, and then correlated to the permeability coefficient. The value of r^2 for these correlations ranges from 0,52 to 0,53 in the 18 regions, and q^2 is about 0,50. It is clear that the role of virtual log P is unaffected by the region in which it is calculated.

Even if the above values confirm the importance of lipophilicity in the determination of skin permeability, and the suitability of the calculated virtual log P as a parameter to be used in predictive equations, the r^2 and q^2 values obtained above are too low for expecting a predictive value of this quantity alone. So a predictive equation of the form of Potts and Guy's, with or without temperature correction has been explored. The best equations turned out to be:

$$pK_p = 3.26 - 0.63\ virtual\ log\ P + 0.0026\ MW \quad \textbf{Equation 6-4}$$

$$r^2 = 0.61\ q^2 = 0.59$$

without temperature correction, and

$$pK_p = 2.97 - 0.63\ virtual\ log\ P + 0.11\ MW/t \quad \textbf{Equation 6-5}$$

$$r^2 = 0.68\ q^2 = 0.65$$

with temperature correction.

Both equations bring inferior results to their counterpart using experimental log P values. As we will show in Paragraph 6.3, the predictive power of virtual log P will be fully exploited only when it is paired with the diffusion coefficient.

6.2.3 Other physicochemical properties

Physicochemical properties different from the virtual log P bring no improvement to the correlation, when used in isolation. Some of them like molecular weight or molecular volume greatly improve the predictive power of equations when paired to the virtual log P, but are of little value when used alone. The results obtained correlating the different molecular properties in the 18 different regions of the lipid matrix model are presented here. Results are shown in Table 6-3.

Property	Region	r^2
PSA	any	0.43
Estimated frontal area	2,3	0.32
MW	any	0.27
Molecular volume	any	0.26
Lipole (Broto & Moreau)	8,9	0.26
Ovality	any	0,16

Table 6-6-3. Contribution of different molecular properties to the explanation of the variability of the common data set.

6.2.4 Scoring functions

The approach of evaluating the SC-permeant interaction energy through a scoring function brought interesting results, similar, in the case of MLP_{Ins} to the results obtained through the use of Virtual log P, not surprisingly. Of the five scoring functions, only CHARMM brought clearly inferior results, with MLP scoring functions bringing the best correlations, in particular at the

water/SC interface (Region 2) in the order MLP_{InSF} > MLP_{InS3} > MLP_{InS2} > MLP_{InS} (Table 6-3).

Scoring function	SC region	r^2
MLP_{InSF}	2	0.59
MLP_{InS3}	2	0.56
MLP_{InS2}	2	0.53
MLP_{InS}	2	0.51
MLP_{InS2}	3	0.51
MLP_{InS}	3	0.50
MLP_{InS3}	3	0.50
MLP_{InSF}	3	0.46
CHARMM	2	0.18

Table 6-6-4. SC-permeant Interaction Score in different regions of the SC.

6.3 Predictive equations based on MD

Running the correlation analysis using the data from the reduced set (n = 80), the best equation obtained of the form:

$$pK_p = \alpha + \beta log(PD/MV)$$

is:

$$pKp = 1.44 - 0.30\log(P_{Zone3}D_{Zone3} / MV_{Zone3}) \quad \textbf{Equation 6-6}$$

r^2 values for all the considered regions are shown in Table 6-5. Even considering the uncertainty on r^2 from Table 6-5, the obtained results show that calculations performed in region 3 (the region at the water/bilayer interface) can explain 70% of the variance in *pKp* values, even using a non-optimized equation. Calculations on region 2 (which is overlapping) give similar results, while r^2 decreases rapidly for other regions, with minima at 9, 14, 16 regions. These results suggest that the water/bilayer interface is the primary target to be taken into consideration when studying the behaviour of penetrants.

A comparison with Potts & Guy on the same set (r^2 = 0.66), shows that Equation 5-1, does not represent such a significant improvement to justify the use of MD simulations. Its importance lies mostly in highlighting that there are regions where the correlation is clearly superior to that of neighbouring (and even overlapping) regions, a remarkable result. So if regions 2 and 3 seems to bring similar results, the correlation clearly drops in region 4, and subsequent regions.

To conclude, though, that the region at the water/SC interface plays a major role in determining skin permeation would be erroneous. The result is equally likely to emerge from the limits of the method employed. The best results are originating from the first part of the simulation, when the trajectories from different simulations are better correlated to one another then at the end of the simulation. It must be noted, however, that the first Region 1 (water) does not correspond to the beginning of the simulations. At the beginning of the simulation, the center of mass of the permeant is positioned at z = 69 Å, but the first 10 Å of erratic movement of the permeant in water are omitted from the analysis. At the end of this 10 Å walk, the

trajectories from different simulation already show a degree of divergence (see Appendix D).

Region	Region position (Å)	r^2	Region	Region position (Å)	r^2
1	59-55	0.64	10	41-37	0.53
2	57-53	0.69	11	39-35	0.58
3	**55-51**	**0.71**	12	37-33	0.59
4	53-49	0.61	13	35-31	0.56
5	51-47	0.62	14	33-29	0.48
6	49-45	0.58	15	31-27	0.53
7	47-43	0.55	16	29-25	0.47
8	45-41	0.54	17	27-23	0.54
9	43-39	0.47	18	25-21	0.51

Table 6-5. r^2 values obtained running the regression analysis on 18 different regions of the *in silico* membrane.

An improvement on Equation 6-6 can be obtained introducing a temperature related term, to account for the different temperatures at which experimental measurements of *Kp* were performed.

One of the advantages of Molecular Dynamics is also the possibility of performing the simulations at the same temperature as the experiment. This procedure has been avoided so far due to the added computational cost of equilibrating the membrane at different temperatures, but it will be explored in the near future.

The dependence of the diffusion coefficient D on absolute temperature T, for small molecules in rubbery polymers, can be expressed as [81]:

$$D = D_0 e^{-\gamma MV/T}$$

where D_0 is a constant for a given molecule, MV is the molecular volume and γ is a constant. In logarithm form:

$$\log D = \log D_0 - \gamma MV/T$$

So, as already pointed out by Vecchia and Bunge [81] a term proportional to MV/T (or MW/T where MW is molecular weight) can be used as a temperature correction factor. In our case, the simulations were performed at 300K, and the ratio between D_T and D_0, the diffusion coefficient at temperature T and 300K, respectively is:

$$\frac{D_T}{D_{300}} = e^{-\gamma MV(1/T - 1/300)}$$

So the term $\log(PD_T)$ can be expressed as:

$$\log(PD_T) = \log(PD_{300}) + \gamma MV\left(\frac{1}{300} - \frac{1}{T}\right)$$

Equation 6-6 becomes:

$$pkp = -1.2062 - 0.3081\log(PD_{300}) + 1.2355\log MV + 23.0409MV\left(\frac{1}{300} - \frac{1}{T}\right)$$

Equation 6-7

where P, D_{300} and MV are calculated in region 3 (Figure 6-5).

	r^2 training set	Adjusted r^2 training set	r^2 test set	Std Dev of Errors training set
Potts & Guy (n = 80)	0,69±0,03	0,66±0,04	0,71±0,06	0,66±0,05
Vecchia & Bunge (n = 80)	0,69±0,02	0,67±0,02	0,73±0,06	0,65±0,04
Equation 6-6 (n = 80)	0,70±0,05	0,68±0,05	0,72±0,09	0,64±0,05
Equation 6-7 (n = 80)	0,79±0,03	0,77±0,04	0,75±0,08	0,56±0,04

Table 6-6. Model validation for both equations. Dimension of the test set = 27. Dimension of the training set = 53. Validation performed on 20 tests.

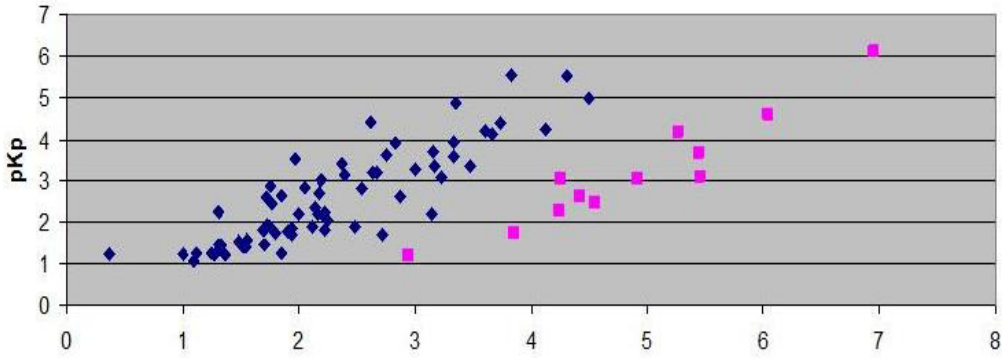

Figure 6-5. Predicted *logKp* values from Eq. 6-7 vs experimental *logKp* values from the common set between Flynn's and the Fully Validated sets. Pink dots represent *pKp* values determined at 310K. Blue dots represent *pKp* values determined around 300K (see Appendix A).

Results for region 3 are shown in Table 6-6, as compared with the Potts & Guy model and the Vecchia & Bunge equation of the form:

$$logkp = a' + b' logP + c' logMW/T$$

Temperature correction to the Potts & Guy equation in the form MW/T brings no significant improvement when regression analysis is performed on the FV set. Temperature correction to Equation 6-6 brings a significant improvement on the training set r^2, adjusted r^2 and standard deviation of the errors, at the expense of adding one degree of freedom, but a more limited improvement to the r^2 of the test set.

To further assess the predictive power of the equation devised in this chapter, the cross-correlation between the different quantities has been considered. A cross-correlation matrix for key values used in the equations is shown in table 6-7. Though some of quantities obviously show a high correlation, quantities appearing in the same equation always have a low correlation, the highest value being r^2 = 0.14 between log(PD)_Region 3 and log MV (also evaluated in Region 3)

	Virtual log P	**Log $D_{xy\ Region\ 3}$**	**Log P*$D_{xy\ Region\ 3}$**	**MW**	**logMV**
Virtual log P	1				
Log $D_{xy\ zone\ 3}$	< 0.001	1			
Log P*$D_{xy\ zone\ 3}$	0.80	0.21	1		
MW	0.03	0.20	0.23	1	
logMV	< 0.001	0.26	0.14	0.91	1

Tabella 6-7. Cross correlation matrix for quantities appearing in the equations devised in this chapter. Listed are r^2 values for any pair.

Conclusions

Molecular Dynamics is now an established tool providing information on the possible impact of chemical modifications on skin penetration by a given compound. The availability of such a tool, in conjunction with existing mathematical models of skin permeability, can help rationalize the design of new drugs to manage cutaneous pathologies, as well as the lead optimization of a molecule to be administrated by transdermal route.

Although Molecular Dynamics does not yet allow the spontaneous permeation of a solute through a membrane to be observed, its characteristics seem to complement existing macroscopic models. For example, the parameters obtained from simulations naturally take into account the anisotropy and non-homogeneity of biomembranes.

One of the strengths of MD is that, when the simulations are well planned, it may represent a "statistical mechanics laboratory", and the analysis of MD trajectories can employ the results from statistical mechanics to calculate relevant quantities. Steered Molecular Dynamics (SMD), in particular, truly is applied Non Equilibrium Statistical Mechanics (NESM), a field that gained momentum in the last decades, with the development of fluctuation-dissipation relations, and other important results, such as Jarzinsky equality. The two fields of SMD and NESM complement each other well and are destined to bring other exciting developments.

On the other hand, one of the difficulties in obtaining meaningful results with numerical simulations, as opposed to a purely theoretical approach, lies

in the complexity of planning the *in silico* experiments, including the choice of simulation parameters. This is particularly true in the case of the screening of a large number of systems: weaknesses in the experimental design often become apparent only after doing a great amount of calculations, and force the *in silico* experimenter to make corrections to the protocol and start again.

In this work, after building a stable and equilibrated model for the stratum corneum (SC) lipid bilayer, SMD simulations were performed to study the penetration of 80 permeants of known permeability coefficient through the SC. MD has been utilized before by other Authors for studying permeation of chemicals through plasma membranes, notably the work by Marrink and Berendsen, whose approach was later applied to the SC by Das, Noro and Olmsted. A screening of a large number of molecules with MD simulations, though has never been attempted to our knowledge. The goal was to assess the possibility of using MD to assist in the development of predictive equations of skin permeability, more than gaining new insight in the mechanics of SC permeation. The approach developed (for plasma membranes) by Marrink and Berendsen for the prediction of permeability couldn't be applied, though, because of the unfeasibility of Free Energy calculations from SMD simulations in the screening of 80 permeants, due to the large amount of trajectories required.

Instead, SMD has been used in an indirect way as a mean to explore the conformational and property space of the permeants in the different microenvironments of the stratum corneum. SMD was also used in a direct way to calculate averaged diffusion coefficients in different regions of the stratum corneum. What proved to be the best option, was a mixed approach where the averaged diffusion coefficients in the plane parallel to the bilayer

surface were calculated in different regions of the SC, and then used together with the physicochemical properties in the correlation equation.

This approach has allowed us to correlate the permeability coefficient to averaged physicochemical properties, improving on existing semi-empirical methods (employing the same quantities determined experimentally). The obtained equation compares well with the Potts and Guy equation. Further improvement of the correlation coefficient seems difficult, considering the heterogeneity of the experimental data sets and the experimental error involved in the measurement of the permeability coefficient.

Further study of SMD data may still allow us to gain a better understanding of the physical process. Performing multiple SMD simulations for a small number of permeants, or designing a new protocol using only equilibrium simulations in specific regions of the lipid matrix model, will allow us to make free energy calculations. So, the work done here is just a first step in the direction of actually employing Molecular Dynamics in developing viable predictive equations of skin permeability.

As a final point, our work has shown that, as stated above, the uncertainty of experimental values of the permeability coefficient is too high to expect significant improvements in the predictive power of equations. The Fully Validated set represents a great advance in reducing the bias of Flynn's set and making the variability of experimental conditions explicit. We feel, though, that this is not enough. As an effect of Regulation (EC) No 1223/2009 on cosmetic products, testing of finished cosmetic products and

ingredients on animals is prohibited in the European Union, as well as the marketing of cosmetic products which have been tested on animals. If governmental Authorities feel that the development of alternative methods to *in vivo* and *in vitro* assays for the determination of skin permeability to chemical compounds is worth consideration and funding, then developing a research project for measuring skin permeability coefficients of a large set of molecules in controlled and reproducible conditions would be an important step in the right direction.

Appendix A. The reduced data set

The reduced data set (Table A.1) is the intersection of Flynn's set and the Vecchia and Bunge's Fully Validated set. It is comprised of 80 compounds, and it is used in this work to eliminate non validated data (according to the five validation criteria exposed below) from Flynn's set, while retaining the ability to compare results with methods optimized on Flynn's set.

Fully validated set validation criteria.

"Data in the FV database were required to meet five criteria:

(a) the temperature must be known and be between 20 and 40 °C,

(b) more than 10% of the penetrating compound must be in nonionized form,

(c) a valid log K_{ow} (Editor's note: octanol/water partition coefficient, logP in this work) [...] must represent the the penetrating molecule (usually the nonionized compound),

(d) the measurement must have been determined at a steady state,

(e) the donor and receptor fluid do not compromise (more than water does) the barrier of the skin. Steady-state permeability coefficients require either constant vehicle concentration and sink conditions in the receptor or adjustment of the data to account for changing vehicle and/or receptor concentrations" [81].

Name	Mass	pKp	LogP	T (K)	t (°C)
2,3-butanediol	90.12	4.4	-0.92	303	30
2,4,6-trichlorophenol	197.45	1.23	3.69	298	25
2,4-dichlorophenol	163	1.22	3.08	298	25
2-butanone	72.11	2.35	0.28	303	30
2-chlorophenol	128.56	1.48	2.15	298	25
2-naphtol	144.17	1.55	2.84	298	25
3,4-xylenol	122.16	1.44	2.35	298	25
3-nitrophenol	139.11	2.25	2	298	25
4-bromophenol	173.01	1.44	2.59	298	25
4-chlorophenol	128.56	1.26	2.39	298	25
4-choro-m-cresol?	142.58	1.44	3.1	298	25
4-ethylphenol	122.16	1.46	2.4	298	25
4-nitrophenol	139.11	2.25	1.96	298	25
aldosterone	360.44	5.52	1.08	299	26
amobarbital	226.27	2.64	1.96	303	30
atropine	289.37	4.86	1.81	303	30
barbital	184.19	3.95	0.65	303	30
benzyl alcohol	108.13	2.22	1.1	298	25
buatnoic acid	88.1	3	0.79	300	27
butobarbital	212.25	3.72	1.65	303	30
cellosolve (2-ethoxy ethanol)	90.12	3.6	-0.54	303	30
chloroxylenol	156.61	1.23	3.39	298	25
chlorpheniramine	274.79	2.64	3.38	303	30
cortexolone	346.46	4.12	2.52	299	26
cortexone (deoxycorticosterone)	330.46	3.35	2.88	299	26
corticosterone	346.46	4.22	1.94	298	25
cortisone	360.44	5	1.42	299	26
diethylcarbamazine	199.29	3.89	0.1	303	30
ephedrine	165.23	2.2	1.03	303	30
estradiol	272.38	3.52	2.69	299	26
estriol	288.38	4.4	2.47	299	26
estrone	270.37	2.44	2.76	299	26
ethanol	46.07	3.1	-0.31	298	25
ethylether	74.12	1.8	0.83	303	30
fentanyl	336.47	1.94	4.37	303	30
heptanoic acid	130.18	1.7	2.5	300	27
hexanoic acid	116.16	1.85	1.9	300	27
hydrocortisone	362.46	5.52	1.53	299	26
[hydrocortisone-21-yl] hemipimelate	504.61	2.64	3.26	310	37
[hydrocortisone-21-yl] hemisuccinate	462.53	3.09	2.11	310	37
[hydrocortisone-21-yl] hexanoate	460.6	1.74	4.48	310	37

Name	Mass	pKp	LogP	T (K)	t (°C)
[hydrocortisone-21-yl]-hydroxy-hexanoate	476.6	3.04	2.79	310	37
[hydrocortisone-21-yl]-N,N-dimethyl-succinamate	491.64	4.17	2.03	310	37
[hydrocortisone-21-yl]-octanoate	488.66	1.21	5.49	310	37
[hydrocortisone-21-yl]-pimelamate	519.67	3.05	2.31	310	37
[hydrocortisone-21-yl]-propionate	418.52	2.47	3	310	37
[hydrocortisone-21-yl]-succinamate	463.59	4.59	1.43	310	37
hydroxyprogesterone	332.48	3.22	3.17	299	26
isoquinoline	129.16	1.77	2.03	303	30
m-cresol	108.14	1.82	1.96	298	25
methanol	32.04	3.3	-0.77	303	30
methyl-[hydrocortisone-21-yl]-pimelate	518.64	2.27	3.7	310	37
methyl-[hydrocortisone-21-yl]-succinate	476.56	3.68	2.58	310	37
methyl-4-hydroxybenzoate	152.15	2.04	1.96	298	25
naproxen	230.26	3.12	3.18	310	37
n-butanol	74.12	2.6	0.88	298	25
n-decanol	158.28	1.1	4	298	25
n-heptanol	116.2	1.46	2.72	298	25
n-hexanol	102.17	1.89	2.03	298	25
nicotine	162.23	1.7	1.17	303	30
n-nonanol	144.25	1.22	3.62	298	25
n-octanol	130.23	1.28	2.97	298	25
n-pentanol	88.15	2.22	1.56	298	25
n-propanol	60.1	2.85	0.25	298	25
o-cresol	108.14	1.8	1.95	298	25
octanoic acid	144.21	1.6	3	300	27
ouabain	584.65	6.11	-2	303	30
p-cresol	108.14	1.75	1.95	298	25
pentanoic acid	102.13	2.7	1.3	300	27
phenobarbital	232.24	3.35	1.47	303	30
phenol	94.11	2.09	1.46	298	25
pregnenolone	316.48	2.82	3.13	299	26
progesterone	314.46	2.82	3.77	299	26
resorcinol	110.11	3.62	0.8	298	25
salicylic acid	138.12	1.89	2.26	303	30
sufentanyl	386.55	1.9	4.59	310	37
testosterone	288.42	3.4	3.31	299	26
thymol	150.22	1.28	3.34	298	25

Table A.1. The reduced data set. *logKp* is the logarithm of the permeation coefficient *Kp*, where *Kp* is expressed in cm/h. *log P* is the logarithm of the octanol/water partition coefficient. *T* and *t* are the experimental temperatures in K and °C, respectively [81].

Appendix B. Flynn's set

Name	Mass	logKp (cm/h)	logKp (cm/s)	LogP
2,3-butanediol	90.12	-4.40	-7.96	-0.92
2,4,6-trichlorophenol	197.45	-1.23	-4.79	3.69
2,4-dichlorophenol	163	-1.22	-4.78	3.08
2-butanone	72.11	-2.35	-5.91	0.28
2-chlorophenol	128.56	-1.48	-5.04	2.15
2-naphtol	144.17	-1.55	-5.11	2.84
3,4-xylenol	122.16	-1.44	-5.00	2.35
3-nitrophenol	139.11	-2.25	-5.81	2.00
4-bromophenol	173.01	-1.44	-5.00	2.59
4-chlorophenol	128.56	-1.44	-5.00	2.39
4-choro-m-cresol	142.58	-1.26	-4.82	3.10
4-ethylphenol	122.16	-1.46	-5.02	2.40
4-nitrophenol	139.11	-2.25	-5.81	1.96
aldosterone	360.44	-5.52	-9.08	1.08
amobarbital	226.27	-2.64	-6.20	1.96
atropine	289.37	-5.07	-8.63	1.81
barbital	184.19	-3.95	-7.51	0.65
benzyl alcohol	108.13	-2.22	-5.78	1.10
buatnoic acid	88.1	-3.00	-6.56	0.79
butobarbital	212.25	-3.71	-7.27	1.65
cellosolve (2-ethoxy ethanol)	90.12	-3.60	-7.16	-0.54
chloroxylenol	156.61	-1.28	-4.84	3.39
chlorpheniramine	274.79	-2.66	-6.22	3.38
codeine	299.36	-4.31	-7.87	0.89
cortexolone	346.46	-4.13	-7.69	2.52
cortexone (deoxycorticosterone)	330.46	-3.35	-6.91	2.88
corticosterone	346.46	-4.22	-7.78	1.94
cortisone	360.44	-5.00	-8.56	1.42
diethylcarbamazine	199.29	-3.89	-7.45	0.10
digitoxin	764.94	-4.89	-8.45	1.86
ephedrine	165.23	-2.22	-5.78	1.03
estradiol (1)	272.38	-3.52	-7.08	2.69
estradiol (2)	272.38	-2.28	-5.84	2.69
estriol	288.38	-4.40	-7.96	2.47
estrone	270.37	-2.44	-6.00	2.76
ethanol	46.07	-3.10	-6.66	-0.31

Name	Mass	logKp (cm/h)	logKp (cm/s)	LogP
ethylbenzene	106.17	0.08	-3.48	3.15
ethylether	74.12	-1.80	-5.36	0.83
etorphine	411.53	-2.44	-6.00	1.86
fentanyl (1)	336.47	-2.25	-5.81	4.37
fentanyl (2)	336.47	-2.00	-5.56	4.37
fluocinonide	494.52	-2.77	-6.33	3.19
heptanoic acid	130.18	-1.70	-5.26	2.50
hexanoic acid	116.16	-1.85	-5.41	1.90
hydrocortisone (1)	362.46	-5.52	-9.08	1.53
hydrocortisone (2)	362.46	-3.93	-7.49	1.53
[hydrocortisone-21-yl]-octanoate	488.66	-1.21	-4.77	5.49
[hydrocortisone-21-yl]-propionate	418.52	-2.47	-6.03	3.00
[hydrocortisone-21-yl] hemipimelate	504.61	-2.75	-6.31	3.26
[hydrocortisone-21-yl] hemisuccinate	462.53	-3.20	-6.76	2.11
[hydrocortisone-21-yl] hexanoate	460.6	-1.75	-5.31	4.48
[hydrocortisone-21-yl]-hydroxy-hexanoate	476.6	-3.04	-6.60	2.79
[hydrocortisone-21-yl]-N,N-dimethyl-succinamate	491.64	-4.17	-7.73	2.03
[hydrocortisone-21-yl]-pimelamate	519.67	-3.05	-6.61	2.31
[hydrocortisone-21-yl]-succinamate	463.59	-4.59	-8.15	1.43
hydromorphone	285.3	-4.82	-8.38	1.25
hydroxypregnenolone	332.48	-3.22	-6.78	3.00
hydroxyprogesterone	332.48	-3.22	-6.78	2.74
isoquinoline	129.16	-1.78	-5.34	2.03
m-cresol	108.14	-1.82	-5.38	1.96
meperidine	247.33	-2.43	-5.99	2.72
methanol	32.04	-3.30	-6.86	-0.77
methyl-[hydrocortisone-21-yl]-pimelate	518.64	-2.27	-5.83	3.70
methyl-[hydrocortisone-21-yl]-succinate	476.56	-3.68	-7.24	2.58
methyl-4-hydroxybenzoate	152.15	-2.04	-5.60	1.96
morphine	285.34	-5.03	-8.59	0.62
naproxen	230.26	-3.40	-6.96	3.18
n-butanol	74.12	-2.60	-6.16	0.88
n-decanol	158.28	-1.10	-4.66	4.00
n-heptanol	116.2	-1.50	-5.06	2.72
n-hexanol	102.17	-1.89	-5.45	2.03
nicotine	162.23	-1.71	-5.27	1.17
nitroglycerine	227.09	-1.96	-5.52	2.00
N-nitrosodiethanolamine	134.13	-5.22	-8.78	-1.58
n-nonanol	144.25	-1.22	-4.78	3.62
n-octanol	130.23	-1.28	-4.84	2.97
n-pentanol	88.15	-2.22	-5.78	1.56
n-propanol	60.1	-2.85	-6.41	0.25
o-cresol	108.14	-1.80	-5.36	1.95

Name	Mass	Log-Kp (cm/h)	Log-Kp (cm/s)	LogP
octanoic acid	144.21	-1.60	-5.16	3.00
ouabain	584.65	-6.11	-9.67	-2.00
p-cresol	108.14	-1.75	-5.31	1.95
pentanoic acid	102.13	-2.70	-6.26	1.30
phenobarbital	232.24	-3.34	-6.90	1.47
phenol	94.11	-2.09	-5.65	1.46
pregnenolone	316.48	-2.82	-6.38	3.13
progesterone	314.46	-2.82	-6.38	3.77
resorcinol	110.11	-3.62	-7.18	0.80
salicylic acid	138.12	-2.20	-5.76	2.26
scopolamine	303.35	-4.30	-7.86	1.24
styrene	104.15	-0.19	-3.75	2.95
sucrose	342.3	-5.28	-8.84	-2.25
sufentanyl	386.55	-1.92	-5.48	4.59
testosterone	288.42	-3.40	-6.96	3.31
thymol	150.22	-1.28	-4.84	3.34
toluene	92.14	0.00	-3.56	2.75
water	18.01	-3.30	-6.86	-1.38

Table B.1. Flynn's data set. *logKp* (cm/h) is the logarithm of the permeation coefficient *Kp*, where *Kp* is expressed in cm/h. *logKp* (cm/s) is the logarithm of the permeation coefficient *Kp*, where *Kp* is expressed in cm/s. *log P* is the logarithm of the octanol/water partition coefficient (note the different notation with respect to Flynn's original publication)

Appendix C. Graphs and trajectories

x-y trajectories of the permeant's Center of Mass (*COM*) are represented here. Sparser region correspond to movement in water and at the water/SC interface, while denser regions correspond to deeper layers of the SC. x-y trajectories allow to follow the evolution of the simulation, in particular it allows to spot simulations where the permeant has crossed simulation box boundaries.

2-butanone

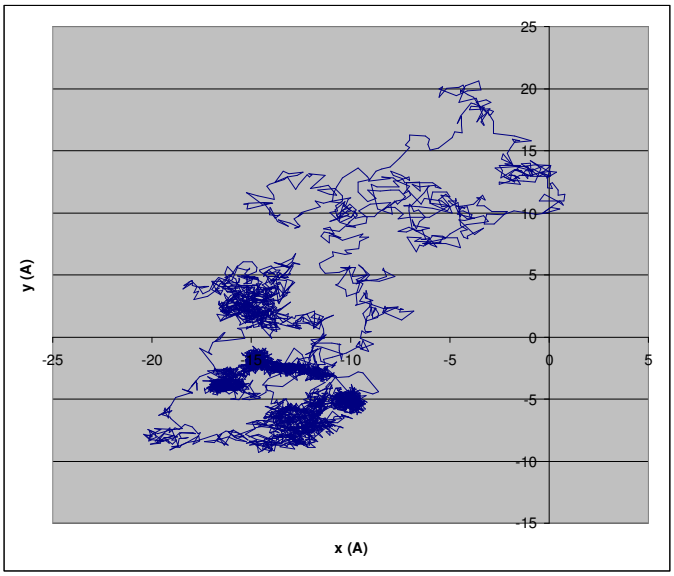

Figure C.1. x(Å) vs y(Å) coordinate of permeant's center of mass.

2-chlorophenol

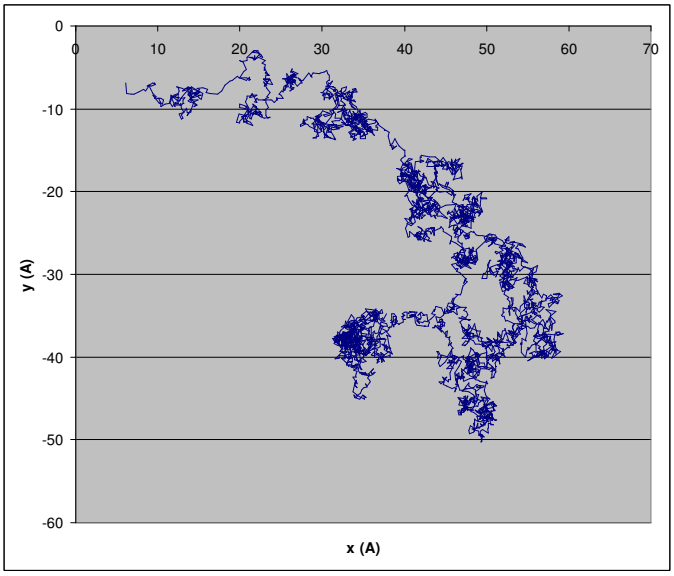

Figure C.2. x(Å) vs y(Å) coordinate of permeant's center of mass.

2-naphtol

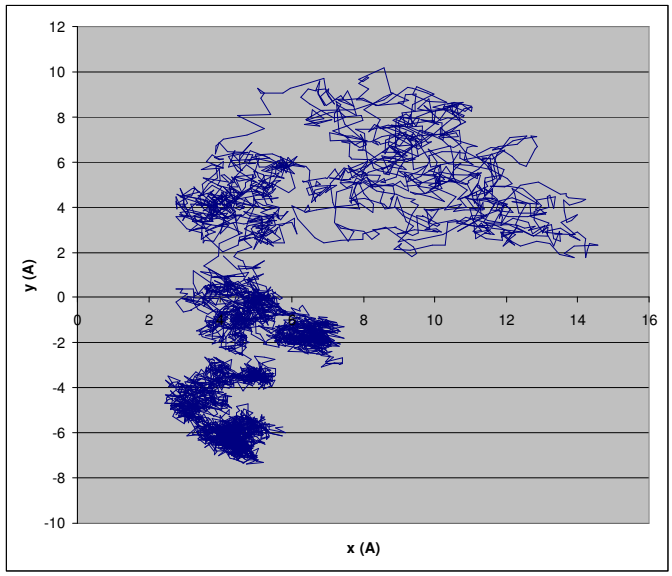

Figure C.3. x(Å) vs y(Å) coordinate of permeant's center of mass.

3-nitrophenol

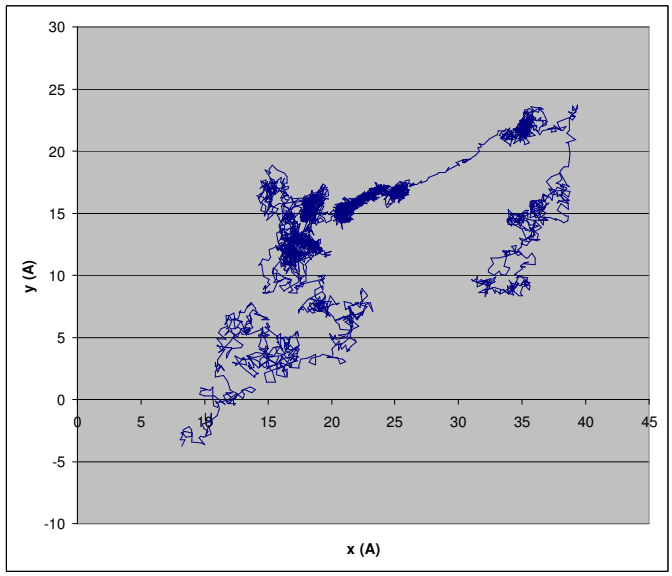

Figure C.4. x(Å) vs y(Å) coordinate of permeant's center of mass.

4-bromophenol

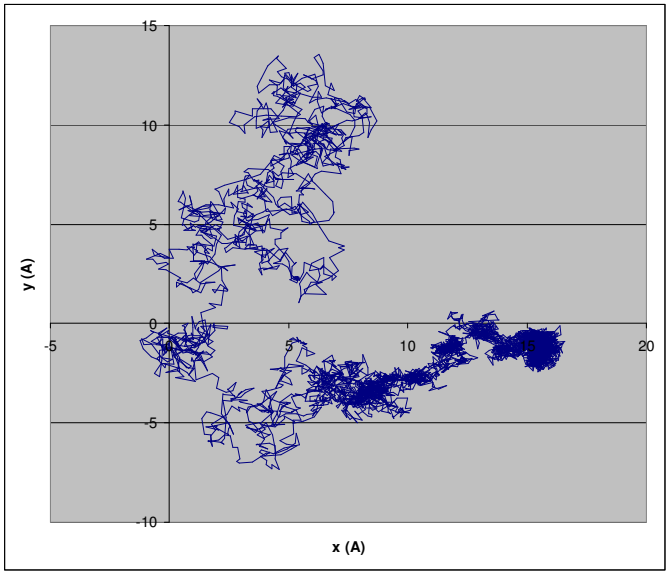

Figure C.5. x(Å) vs y(Å) coordinate of permeant's center of mass.

4-chlorocresol

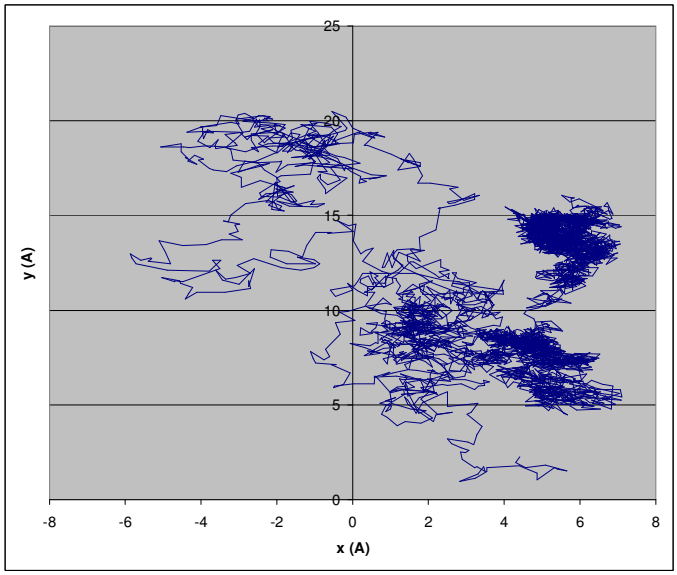

Figure C.6. x(Å) vs y(Å) coordinate of permeant's center of mass.

4-chlorophenol

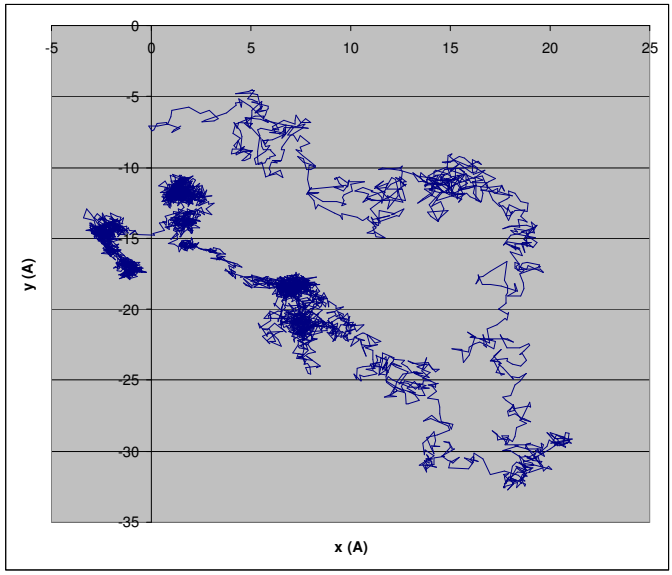

Figure C.7. x(Å) vs y(Å) coordinate of permeant's center of mass.

4-ethylphenol

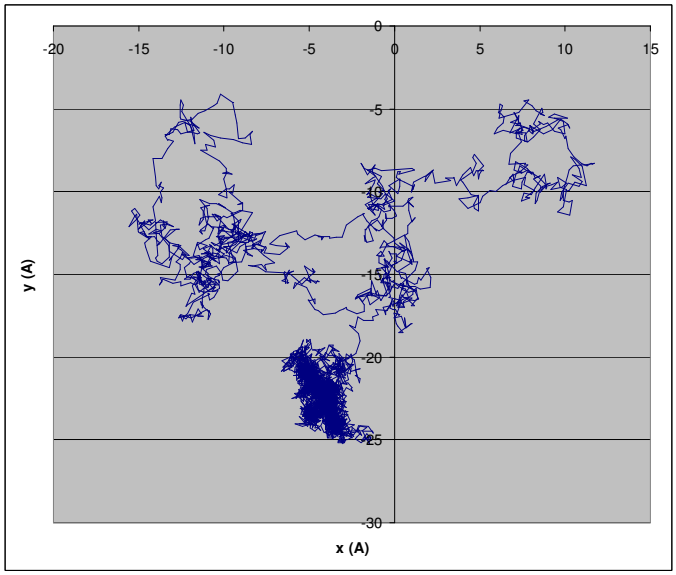

Figure C.8. x(Å) vs y(Å) coordinate of permeant's center of mass.

4-nitrophenol

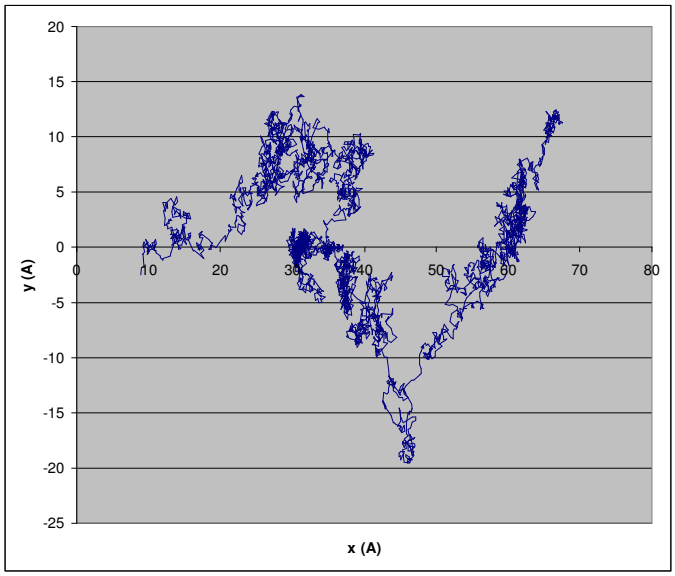

Figure C.9. x(Å) vs y(Å) coordinate of permeant's center of mass.

Aldosterone

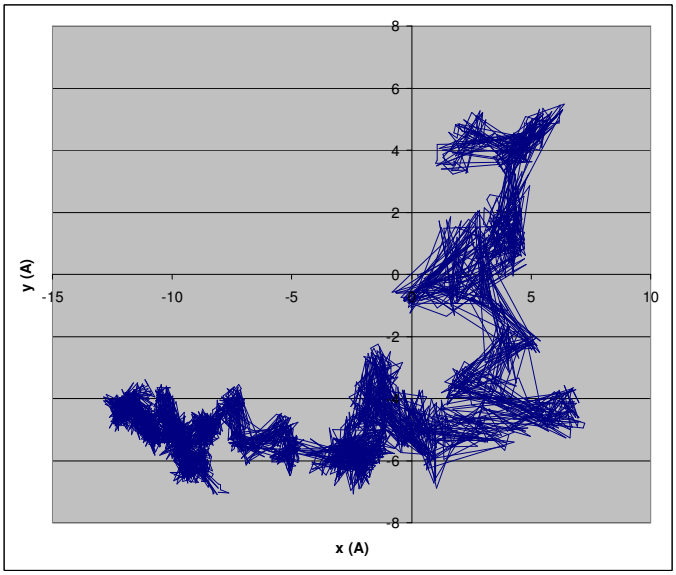

Figure C.10. x(Å) vs y(Å) coordinate of permeant's center of mass.

Amobarbital

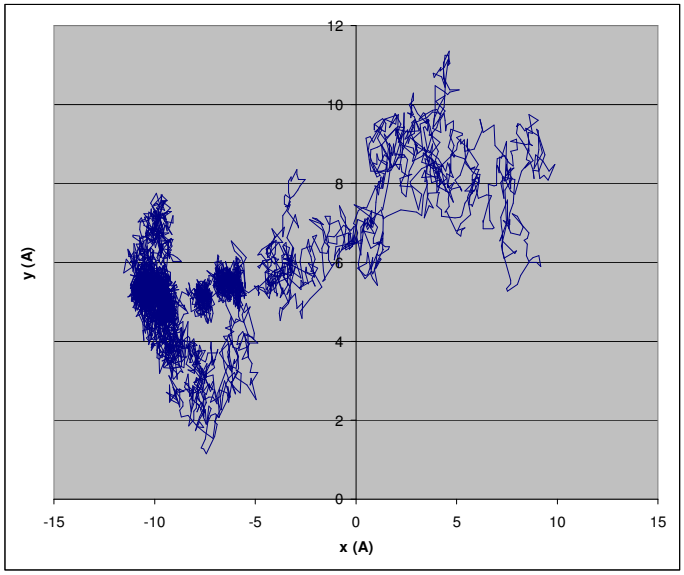

Figure C.11. x(Å) vs y(Å) coordinate of permeant's center of mass.

Atropine

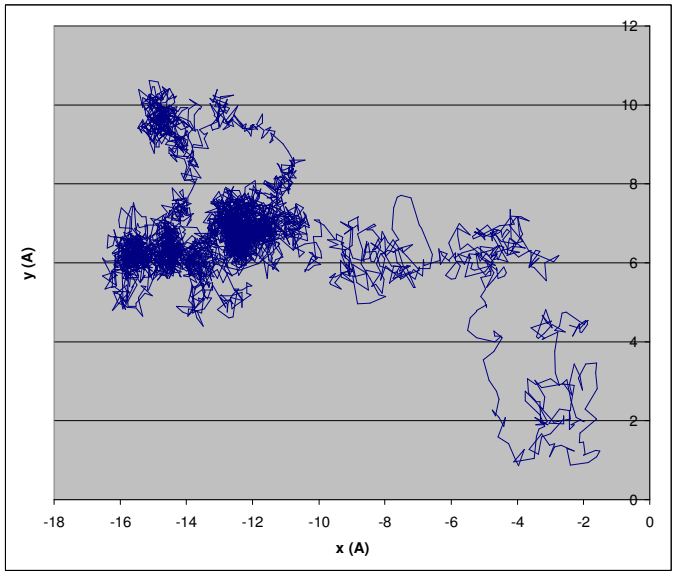

Figure C.12. x(Å) vs y(Å) coordinate of permeant's center of mass.

Barbital

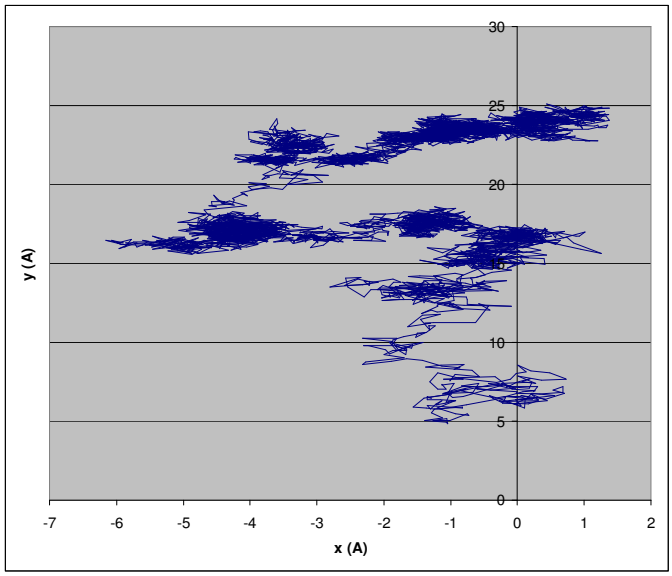

Figure C.13. x(Å) vs y(Å) coordinate of permeant's center of mass.

Benzyl alcohol

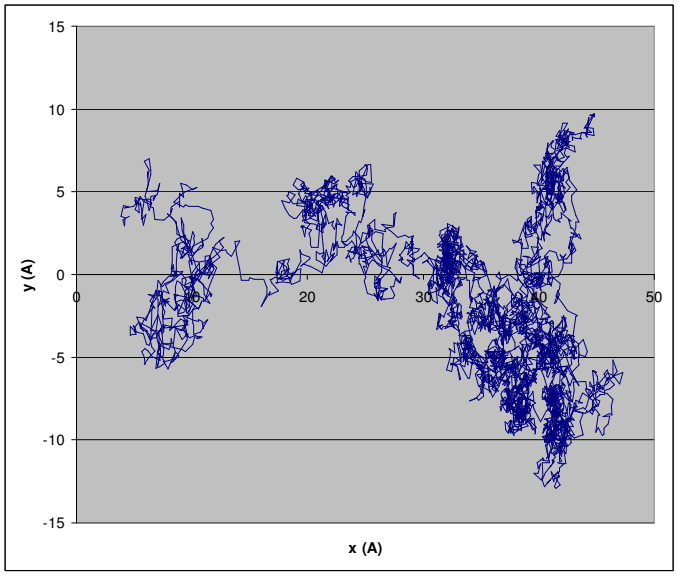

Figure C.14. x(Å) vs y(Å) coordinate of permeant's center of mass.

2,3-butanediol

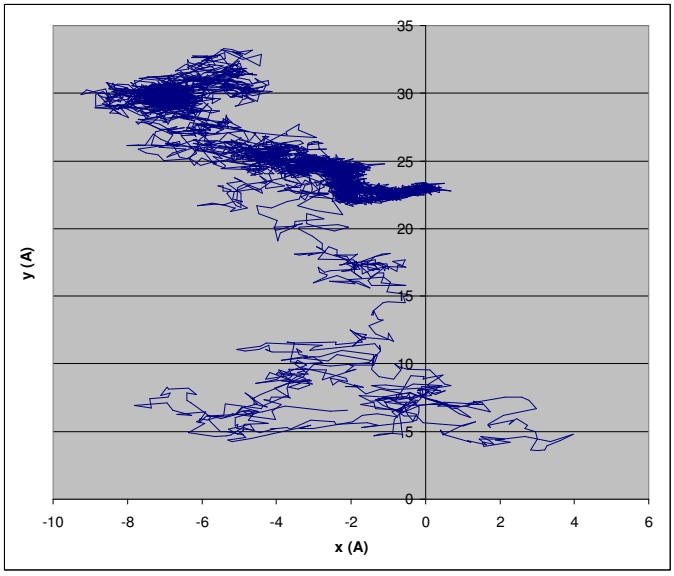

Figure C.15. x(Å) vs y(Å) coordinate of permeant's center of mass.

Butanoic Acid

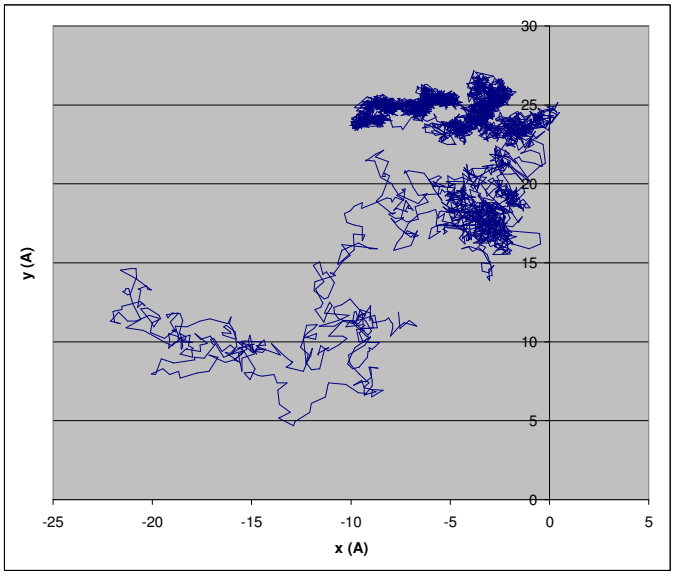

Figure C.16. x(Å) vs y(Å) coordinate of permeant's center of mass.

Butobarbital

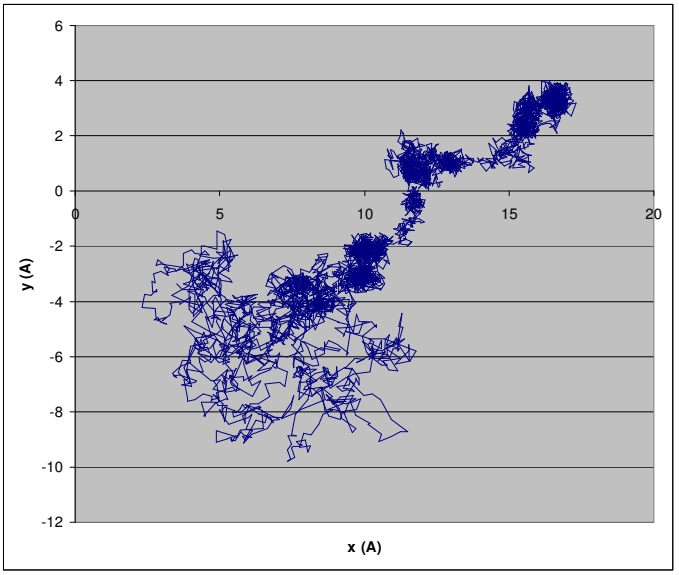

Figure C.17. x(Å) vs y(Å) coordinate of permeant's center of mass.

Cellosolve

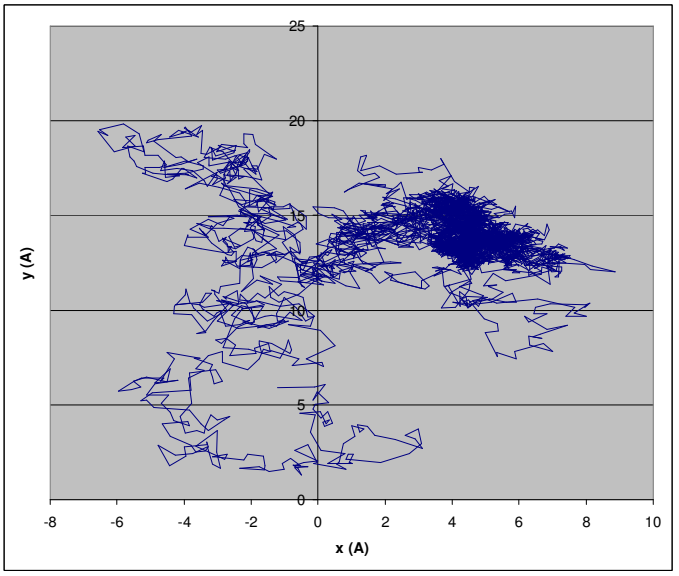

Figure C.18. x(Å) vs y(Å) coordinate of permeant's center of mass.

Chloroxylenol

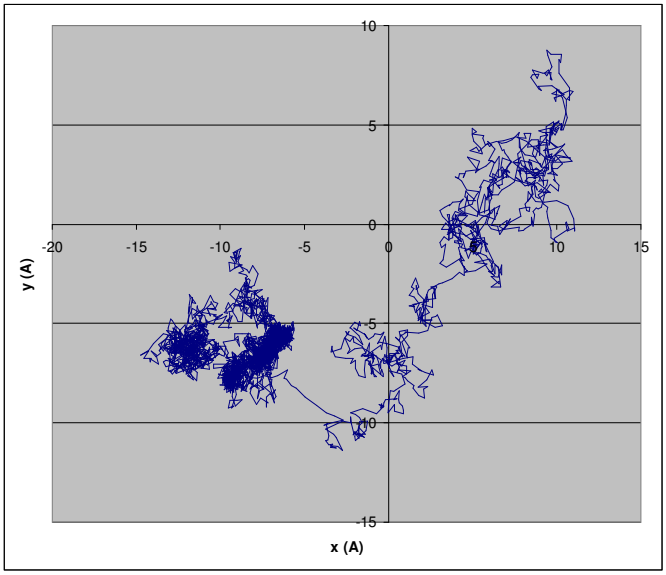

Figure C.19. x(Å) vs y(Å) coordinate of permeant's center of mass.

Chlorpheniramine

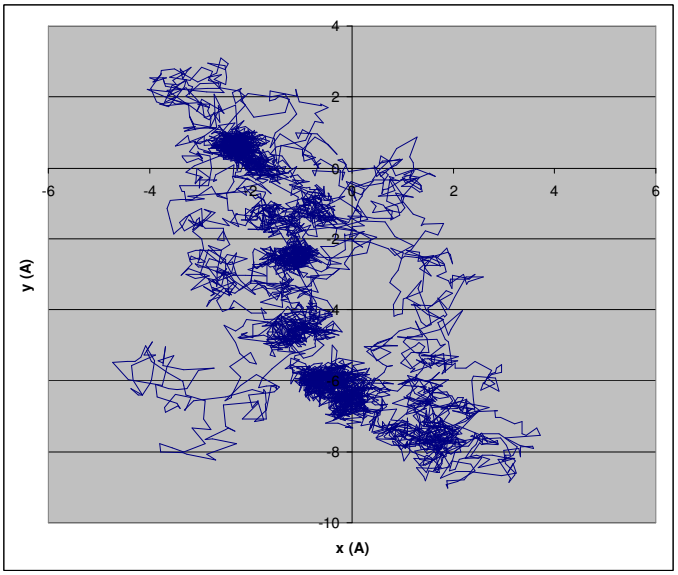

Figure C.20. x(Å) vs y(Å) coordinate of permeant's center of mass.

Cortexolone

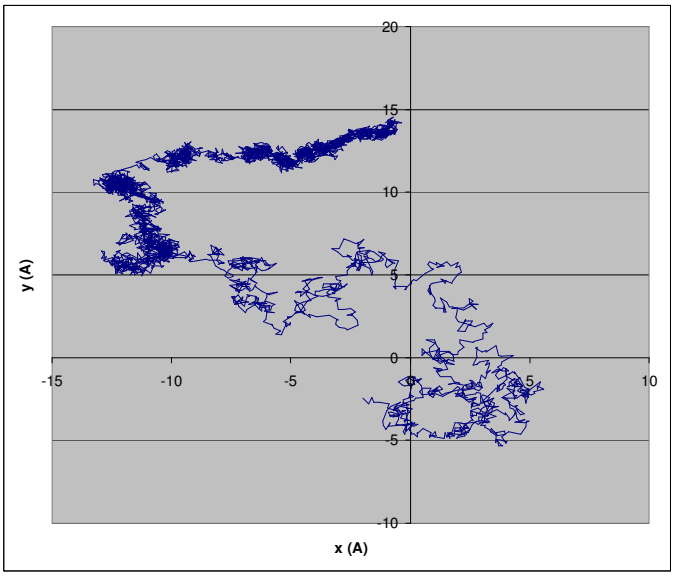

Figure C.21. x(Å) vs y(Å) coordinate of permeant's center of mass.

Cortexone

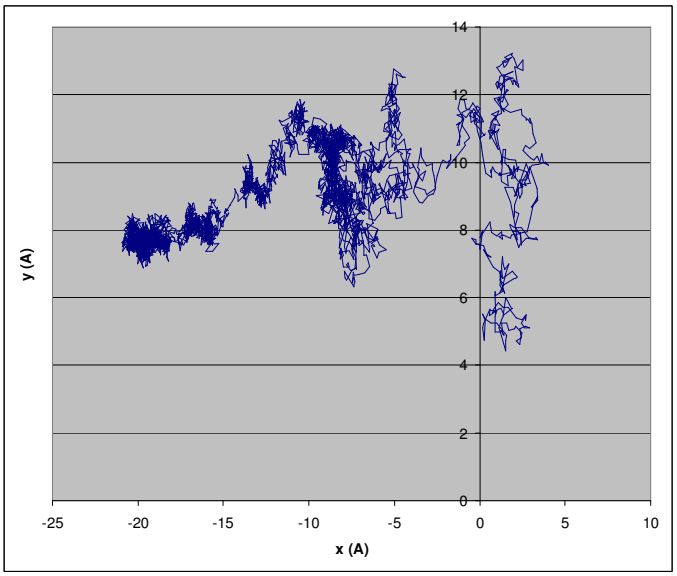

Figure C.22. x(Å) vs y(Å) coordinate of permeant's center of mass.

Corticosterone

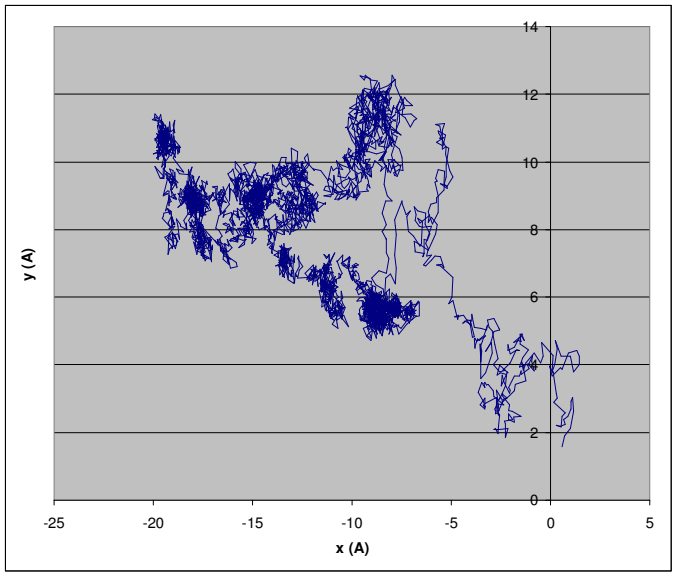

Figure C.23. x(Å) vs y(Å) coordinate of permeant's center of mass.

Cortisone

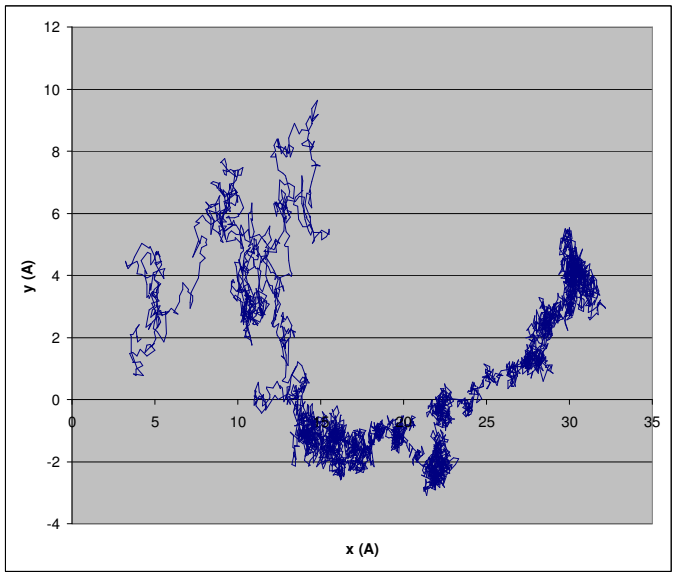

Figure C.24. x(Å) vs y(Å) coordinate of permeant's center of mass.

2,4-dichlorophenol

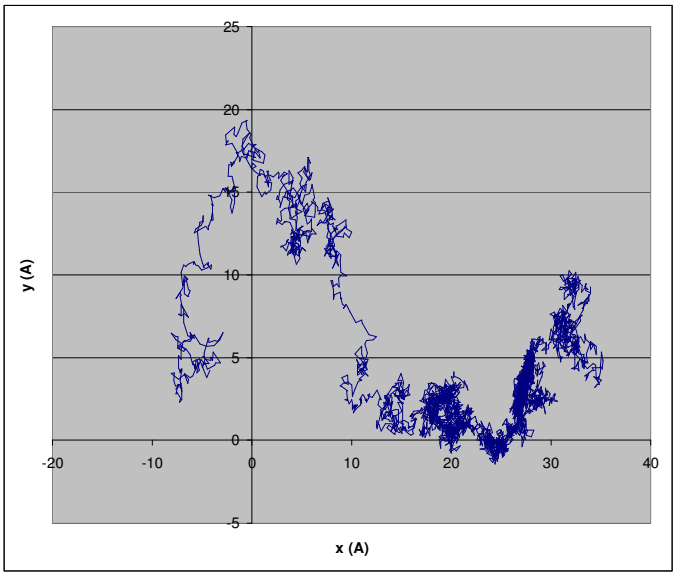

Figure C.25. x(Å) vs y(Å) coordinate of permeant's center of mass.

Diethylcarbamazine

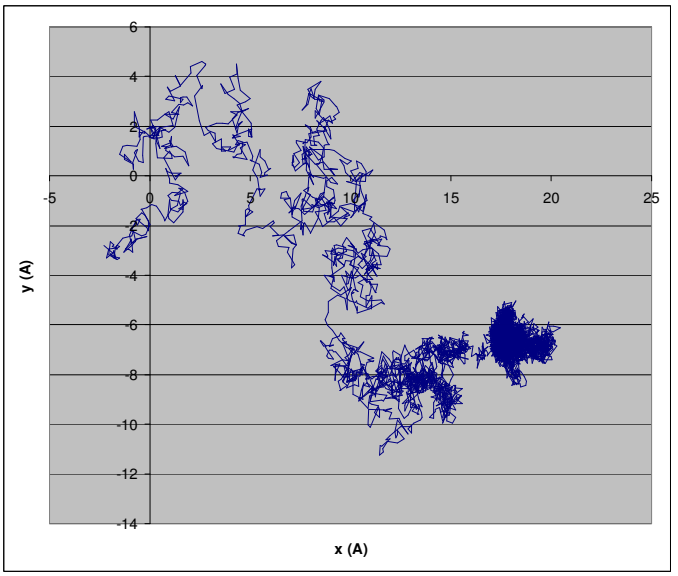

Figure C.26. x(Å) vs y(Å) coordinate of permeant's center of mass.

Ephedrine

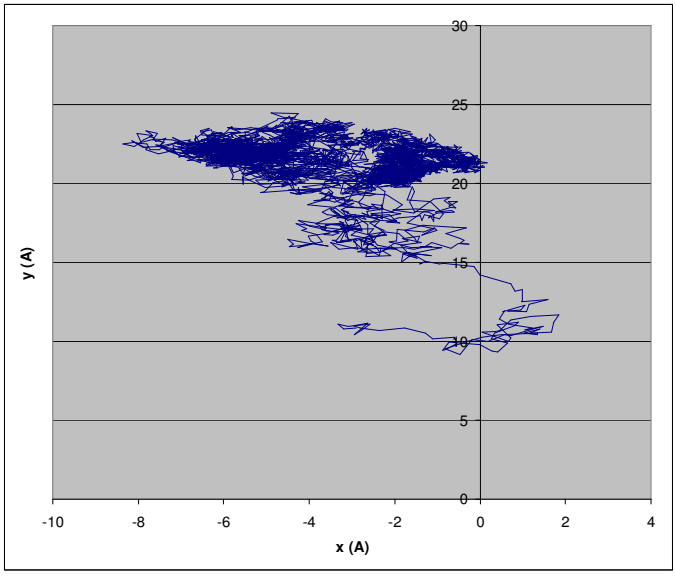

Figure C.27. x(Å) vs y(Å) coordinate of permeant's center of mass.

Estradiol

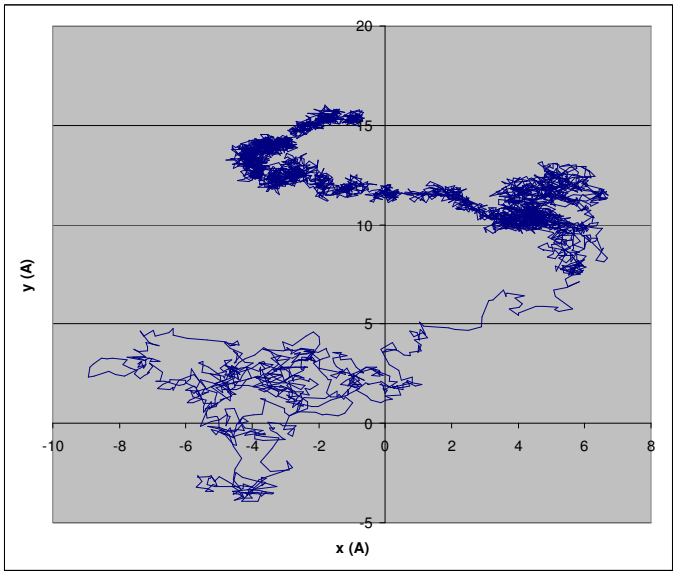

Figure C.28. x(Å) vs y(Å) coordinate of permeant's center of mass.

Estriol

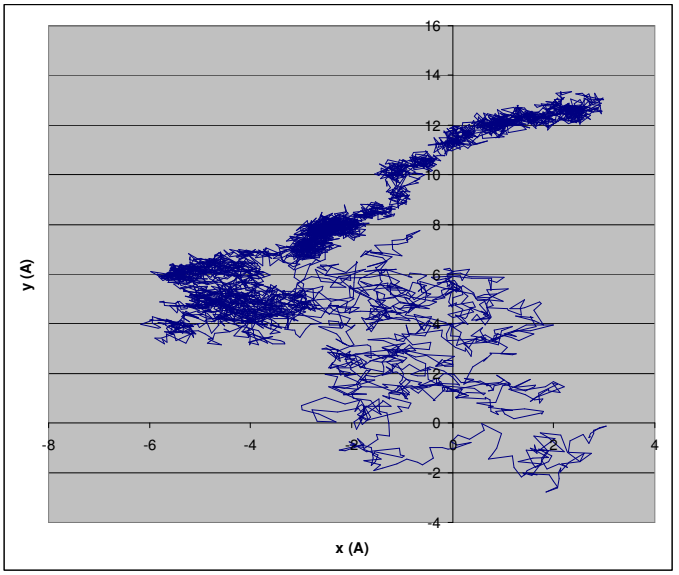

Figure C.29. x(Å) vs y(Å) coordinate of permeant's center of mass.

Estrone

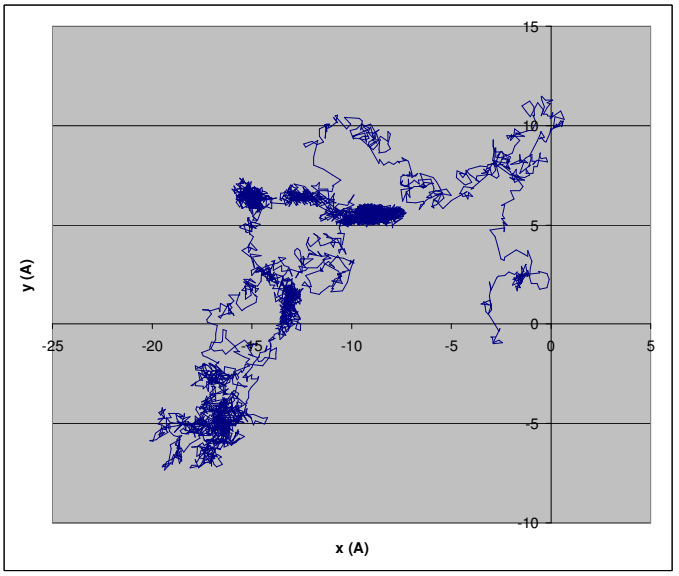

Figure C.30. x(Å) vs y(Å) coordinate of permeant's center of mass.

Ethanol

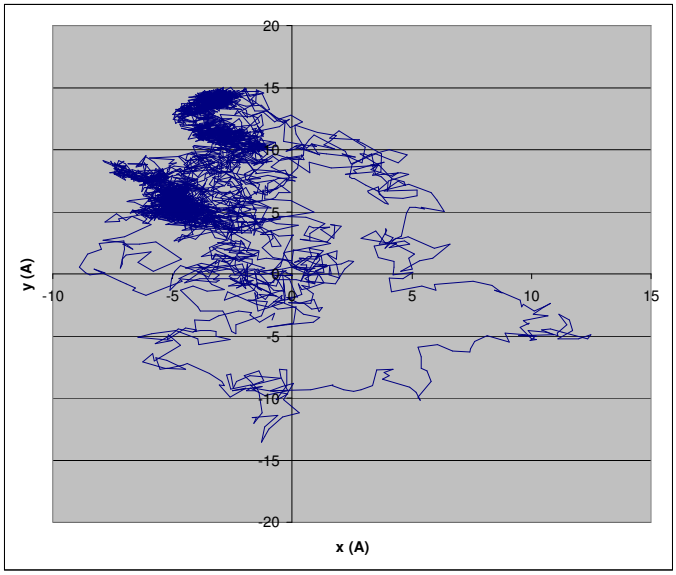

Figure C.31. x(Å) vs y(Å) coordinate of permeant's center of mass.

Ethylether

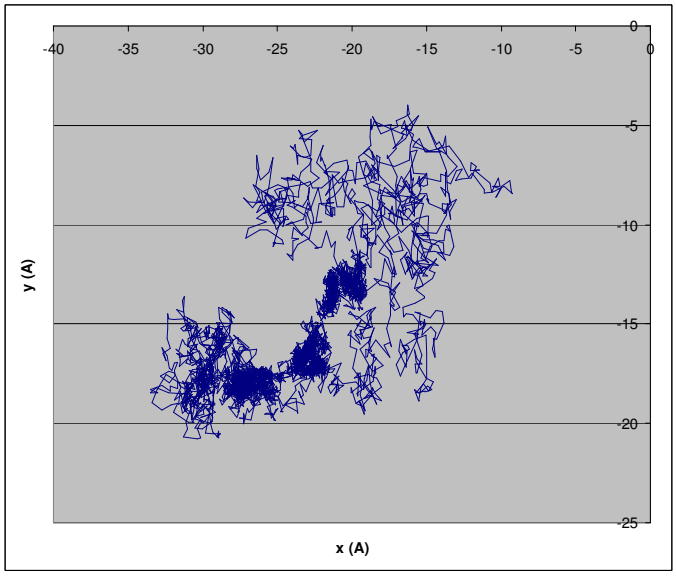

Figure C.32. x(Å) vs y(Å) coordinate of permeant's center of mass.

Fentanyl

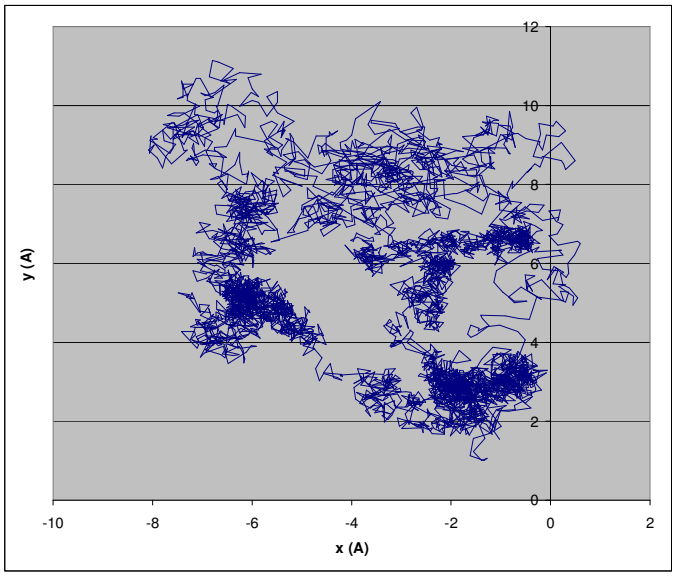

Figure C.33. x(Å) vs y(Å) coordinate of permeant's center of mass.

[hydrocortisone-21-yl]-hemipimelate

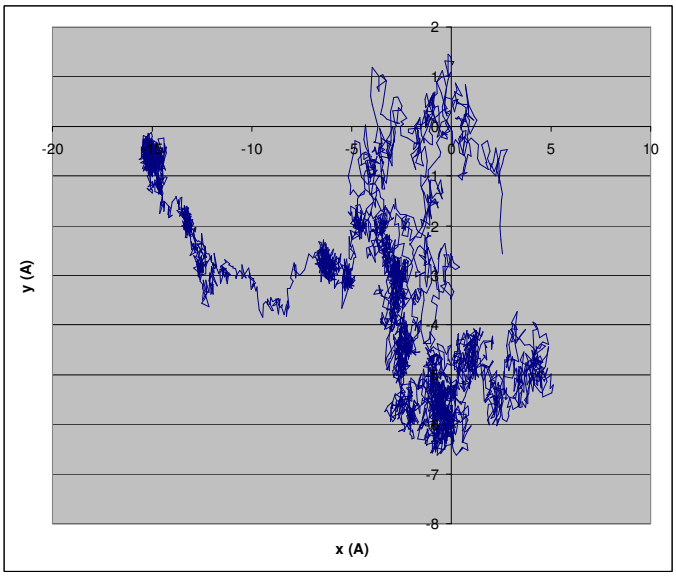

Figure C.34. x(Å) vs y(Å) coordinate of permeant's center of mass.

[hydrocortisone-21-yl]-hemisuccinate

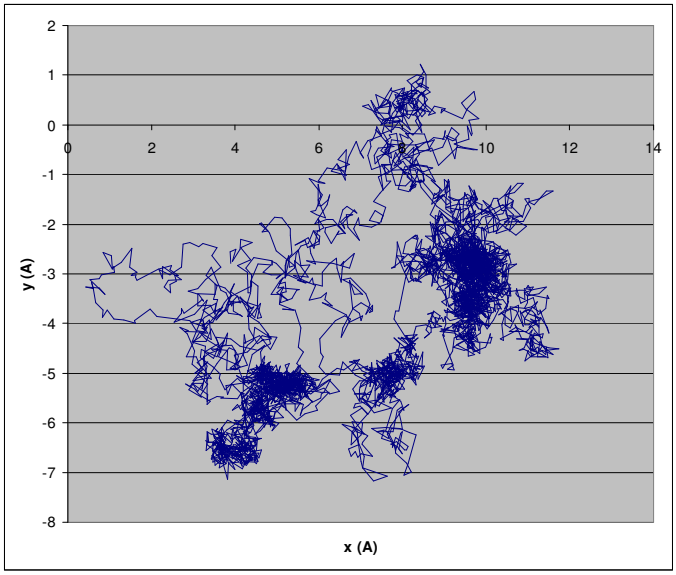

Figure C.35. x(Å) vs y(Å) coordinate of permeant's center of mass.

[hydrocortisone-21-yl]-hexanoate

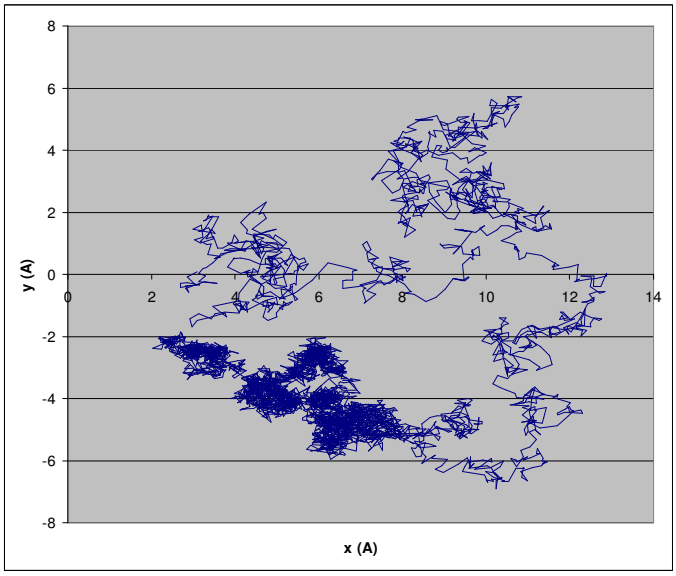

Figure C.36. x(Å) vs y(Å) coordinate of permeant's center of mass.

[hydrocortisone-21-yl]-hydroxy-hexanoate

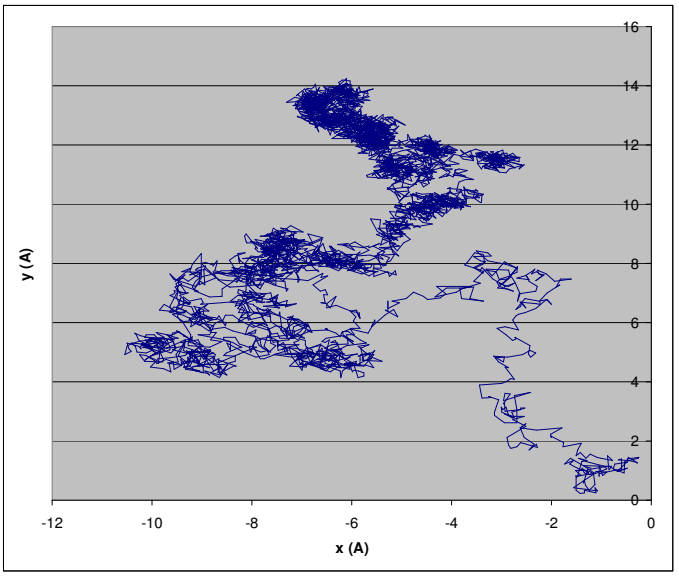

Figure C.37. x(Å) vs y(Å) coordinate of permeant's center of mass.

[hydrocortisone-21-yl]-N,N-dimethylsuccinamate

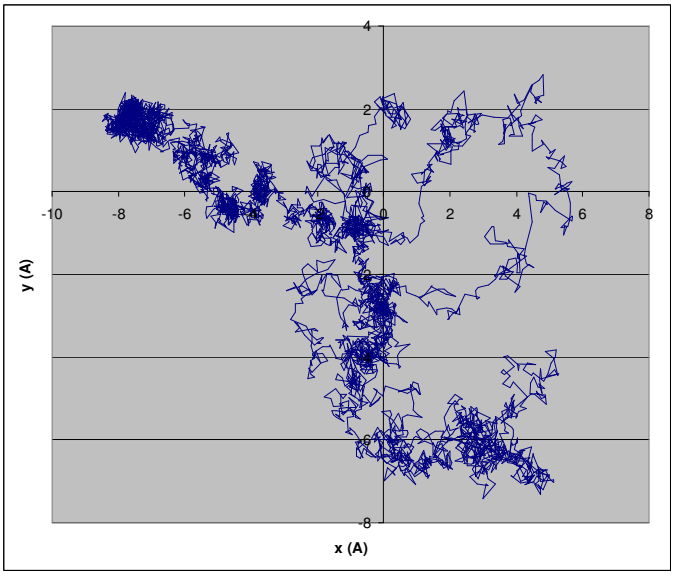

Figure C.38. x(Å) vs y(Å) coordinate of permeant's center of mass.

[hydrocortisone-21-yl]-octanoate

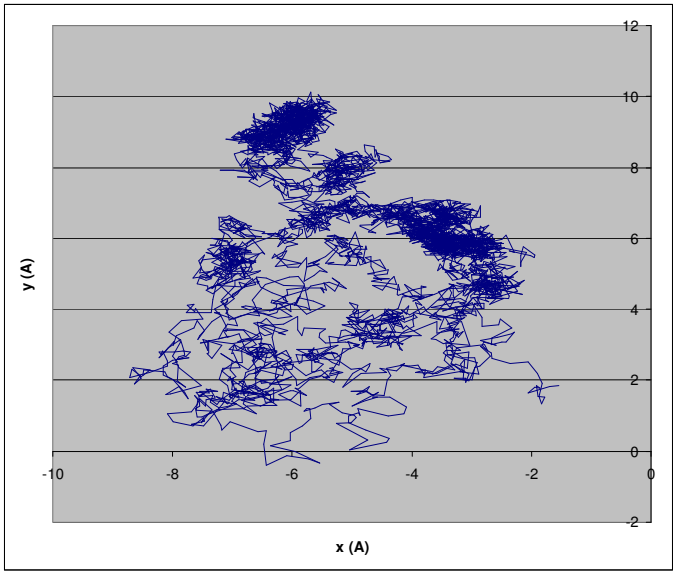

Figure C.39. x(Å) vs y(Å) coordinate of permeant's center of mass.

[hydrocortisone-21-yl]-pimelamate

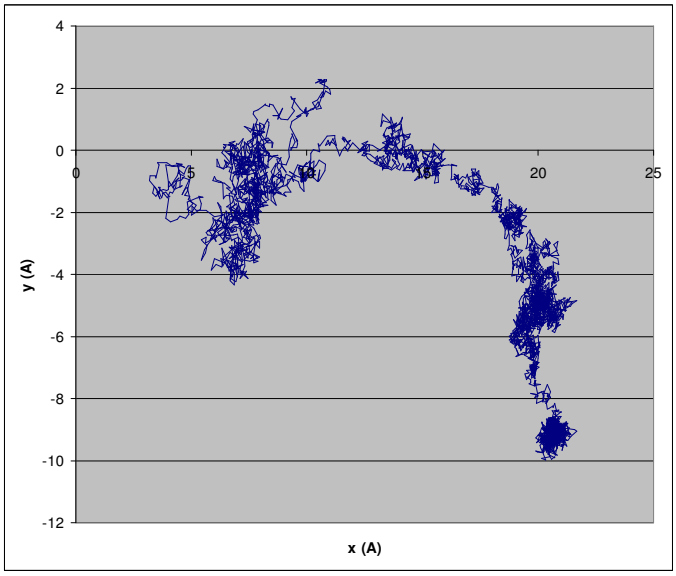

Figure C.40. x(Å) vs y(Å) coordinate of permeant's center of mass.

[hydrocortisone-21-yl]-propionate

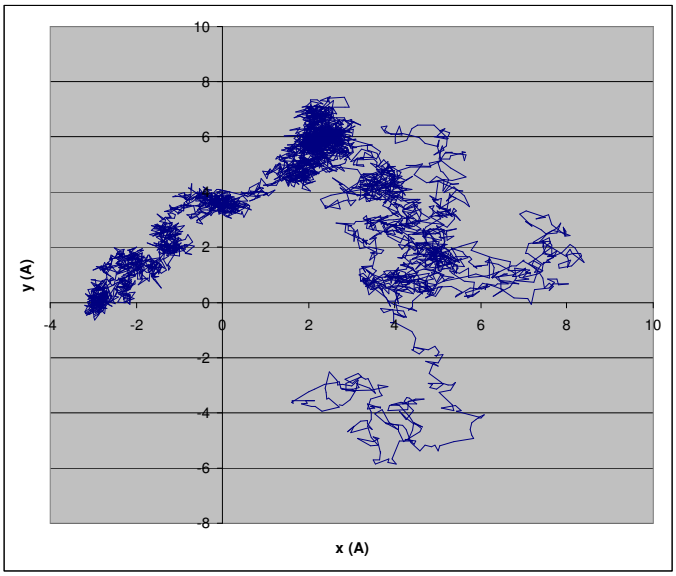

Figure C.41. x(Å) vs y(Å) coordinate of permeant's center of mass.

[hydrocortisone-21-yl]-succinamate

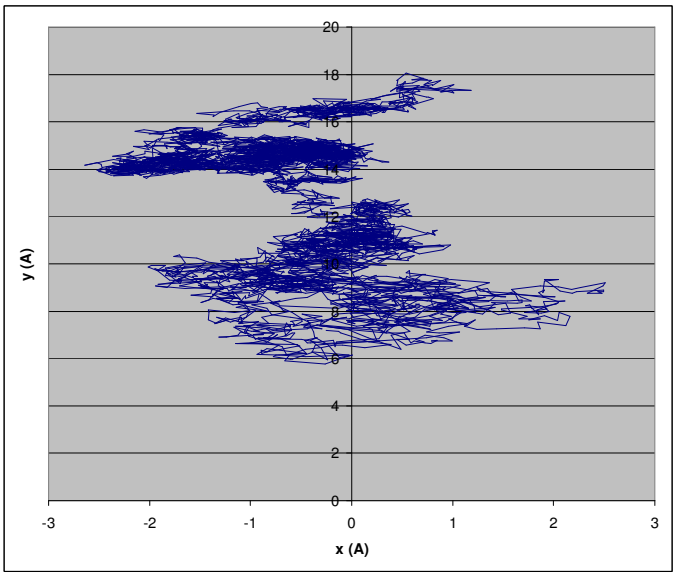

Figure C.42. x(Å) vs y(Å) coordinate of permeant's center of mass.

Heptaonic acid

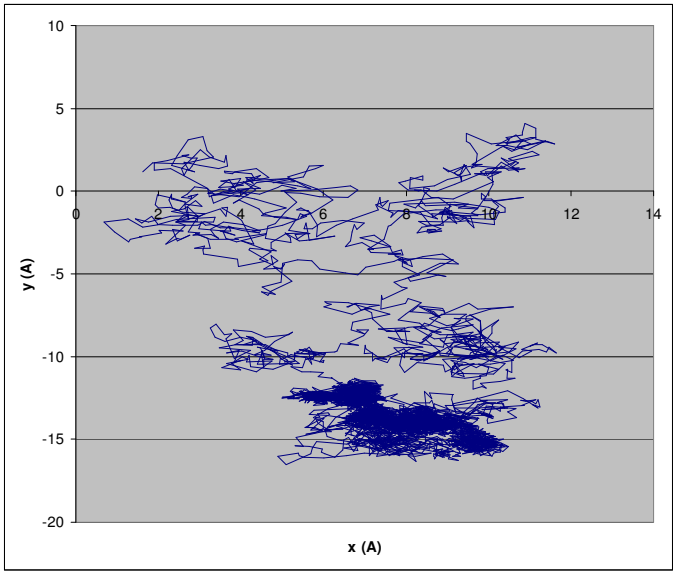

Figure C.43. x(Å) vs y(Å) coordinate of permeant's center of mass.

Hexanoic acid

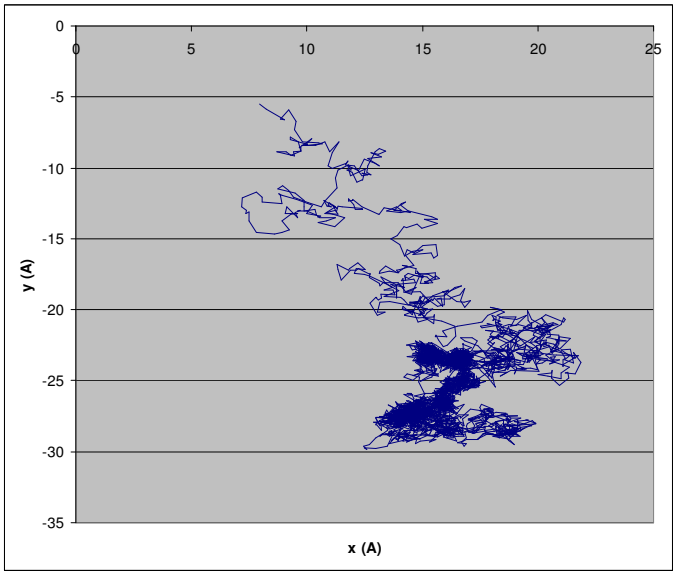

Figure C.44. x(Å) vs y(Å) coordinate of permeant's center of mass.

Hydrocortisone

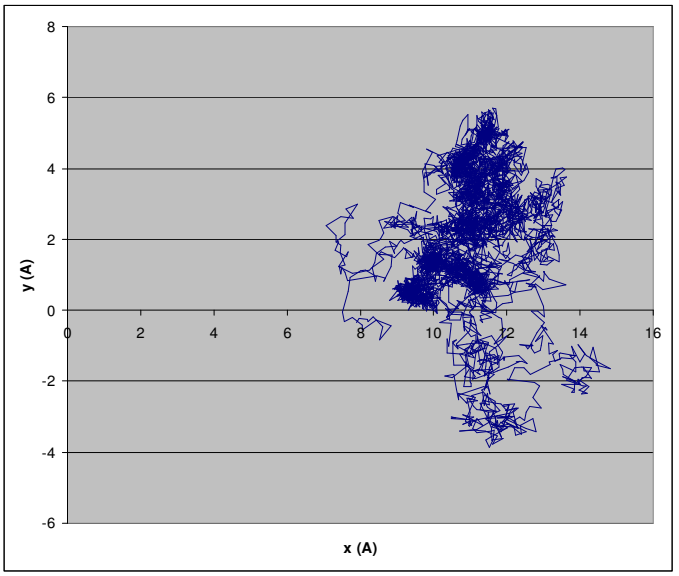

Figure C.45. x(Å) vs y(Å) coordinate of permeant's center of mass.

Hydroxypregnenolone

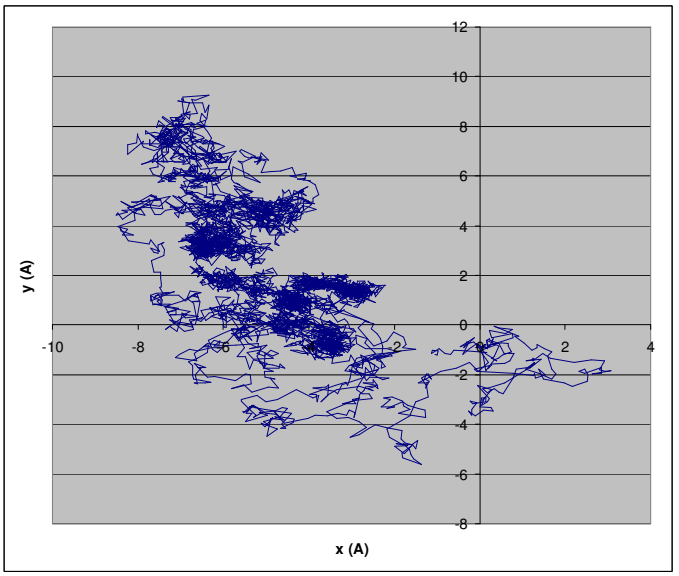

Figure C.46. x(Å) vs y(Å) coordinate of permeant's center of mass.

Hydroxyprogesterone

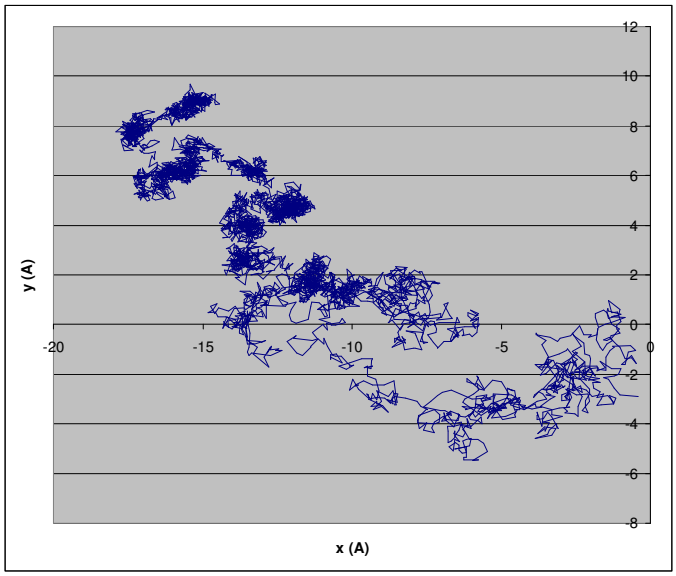

Figure C.47. x(Å) vs y(Å) coordinate of permeant's center of mass.

Isoquinoline

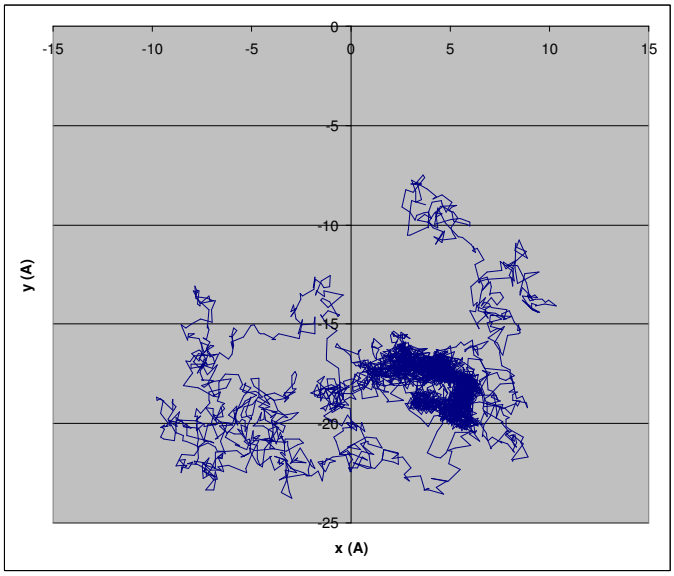

Figure C.48. x(Å) vs y(Å) coordinate of permeant's center of mass.

m-Cresol

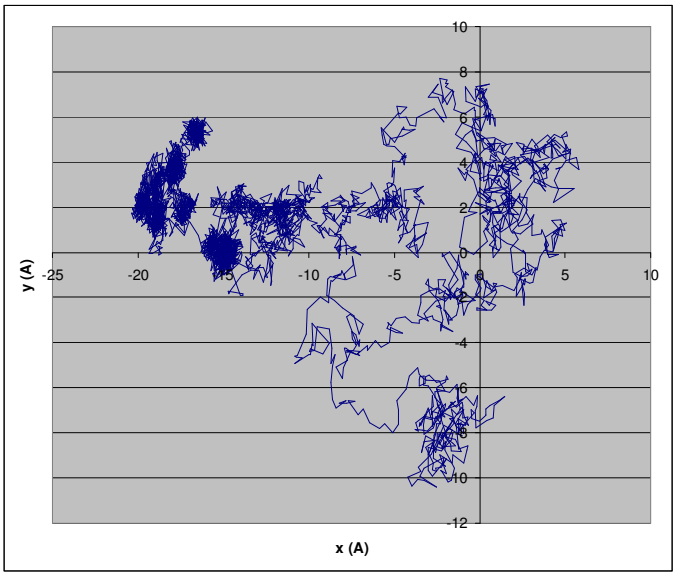

Figure C.49. x(Å) vs y(Å) coordinate of permeant's center of mass.

Methanol

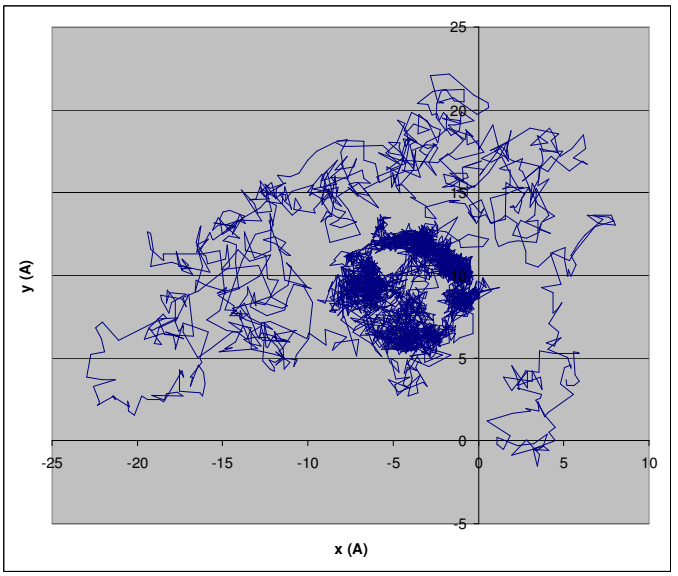

Figure C.50. x(Å) vs y(Å) coordinate of permeant's center of mass.

Methyl-4-hydroxybenzoate

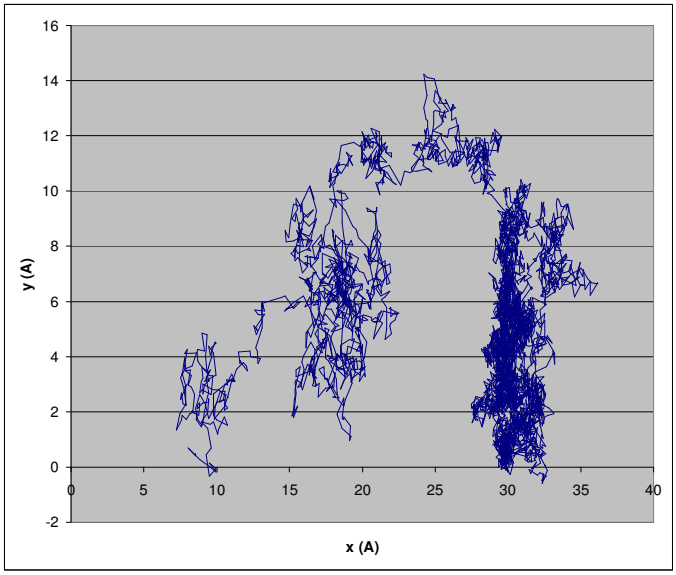

Figure C.51. x(Å) vs y(Å) coordinate of permeant's center of mass.

Methyl-[hydrocortisone-21-yl]-pimelate

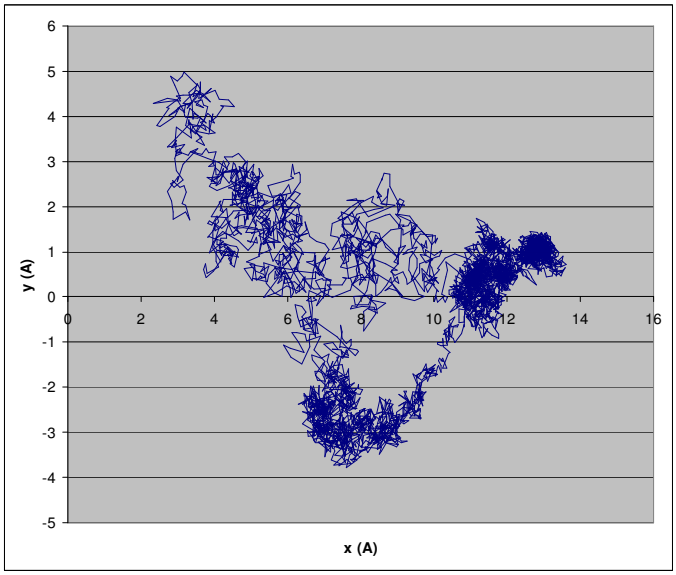

Figure C.52. x(Å) vs y(Å) coordinate of permeant's center of mass.

Methyl-[hydrocortisone-21-yl]-succinate

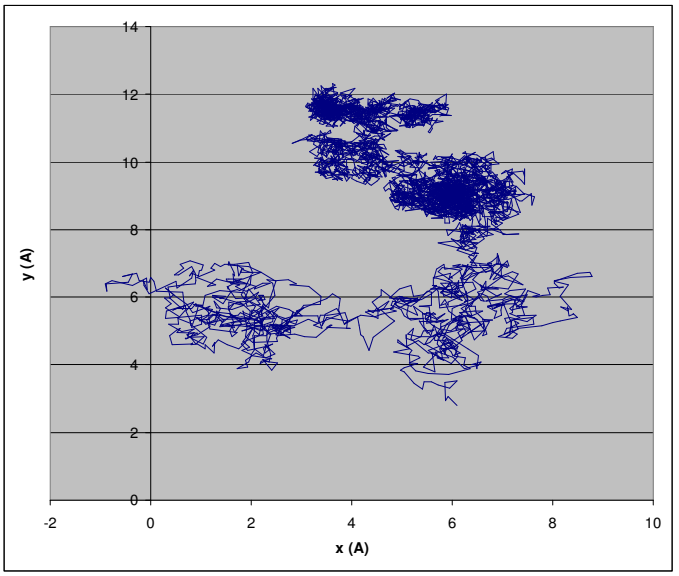

Figure C.53. x(Å) vs y(Å) coordinate of permeant's center of mass.

Naproxene

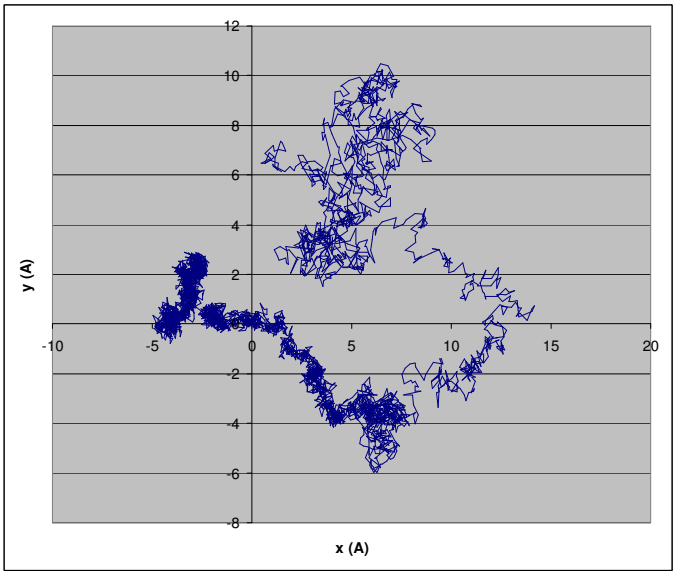

Figure C.54. x(Å) vs y(Å) coordinate of permeant's center of mass.

n-Butanol

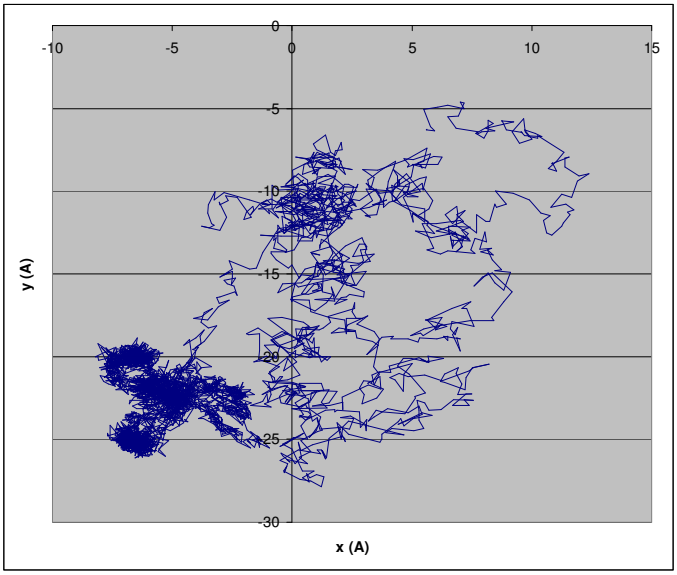

Figure C.55. x(Å) vs y(Å) coordinate of permeant's center of mass.

n-Decanol

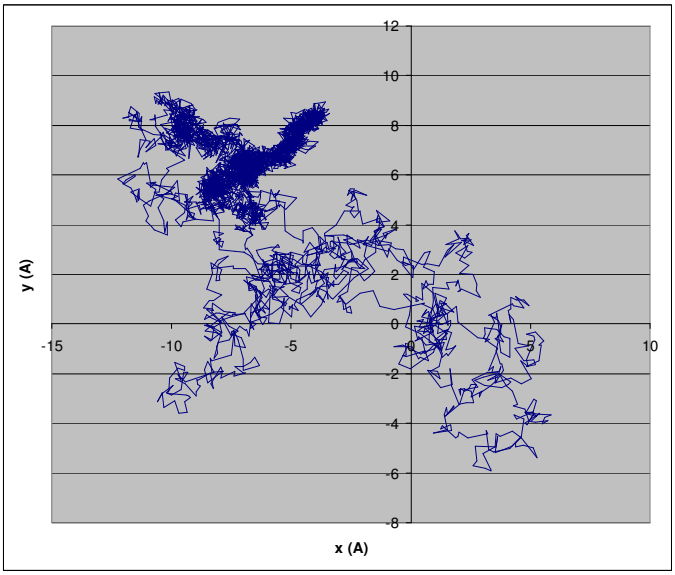

Figure C.56. x(Å) vs y(Å) coordinate of permeant's center of mass.

n-Heptanol

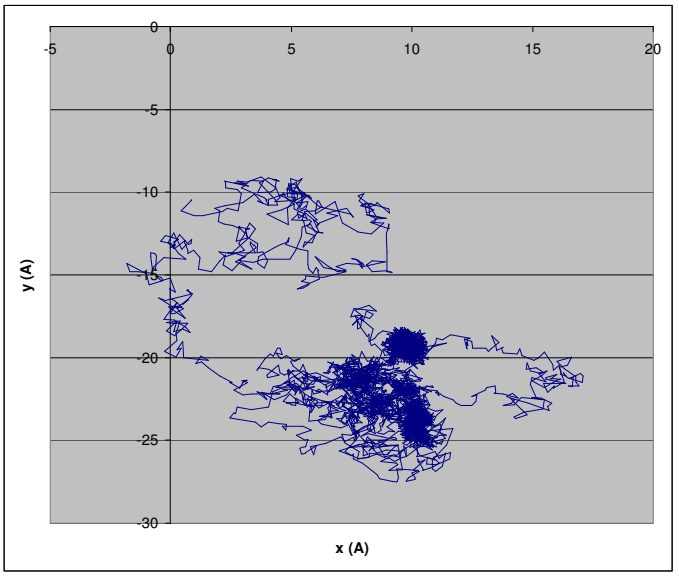

Figure C.57. x(Å) vs y(Å) coordinate of permeant's center of mass.

n-Hexanol

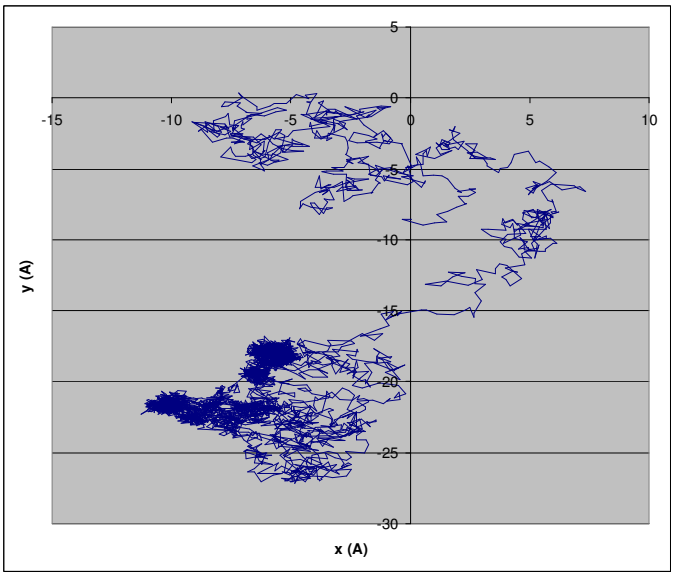

Figure C.58. x(Å) vs y(Å) coordinate of permeant's center of mass.

n-Nonanol

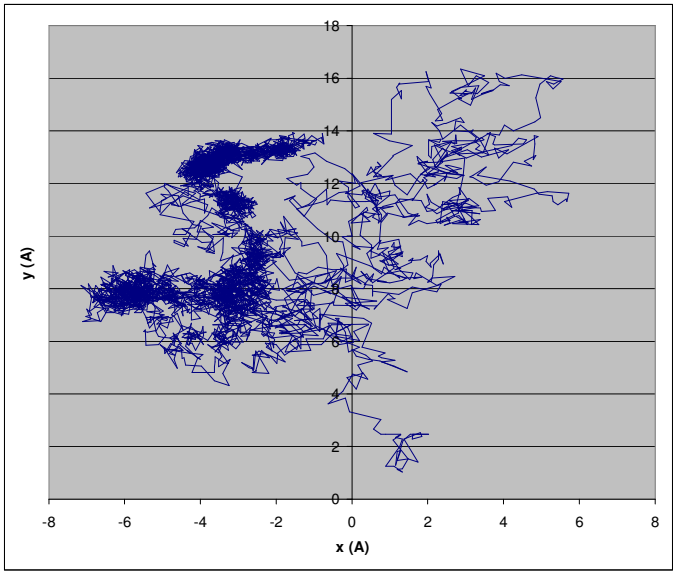

Figure C.59. x(Å) vs y(Å) coordinate of permeant's center of mass.

n-Octanol

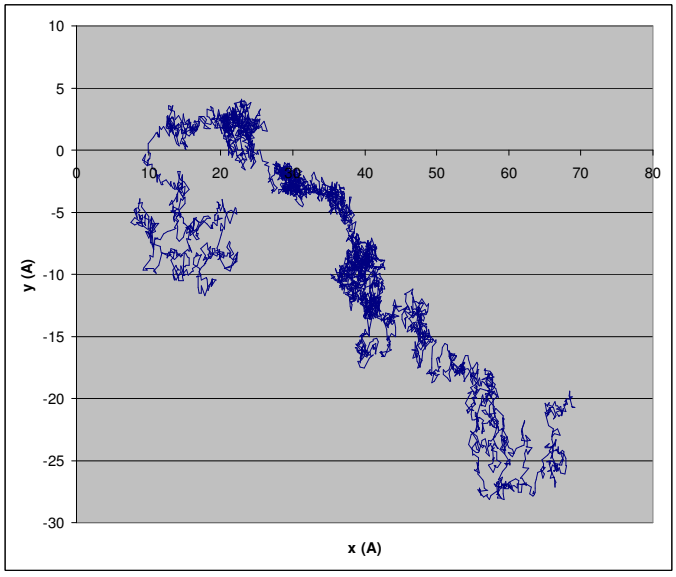

Figure C.60. x(Å) vs y(Å) coordinate of permeant's center of mass.

n-Pentanol

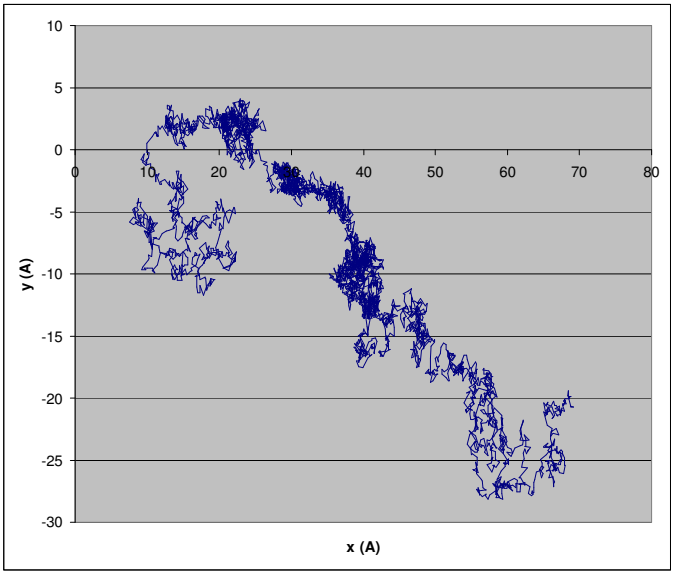

Figure C.61. x(Å) vs y(Å) coordinate of permeant's center of mass.

n-Propanol

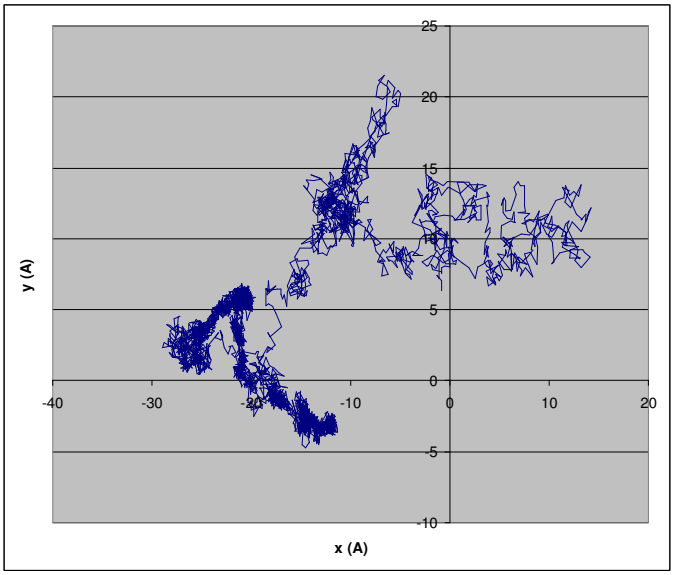

Figure C.62. x(Å) vs y(Å) coordinate of permeant's center of mass.

Nicotine

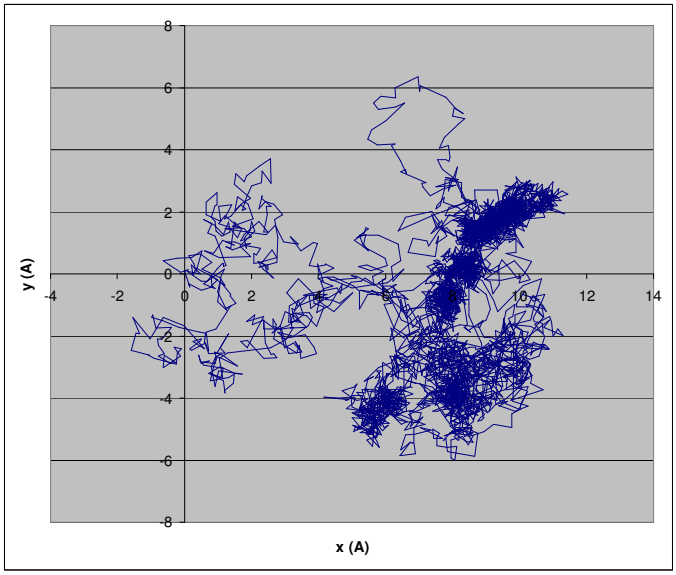

Figure C.63. x(Å) vs y(Å) coordinate of permeant's center of mass.

N-nitrosodiethanolamine

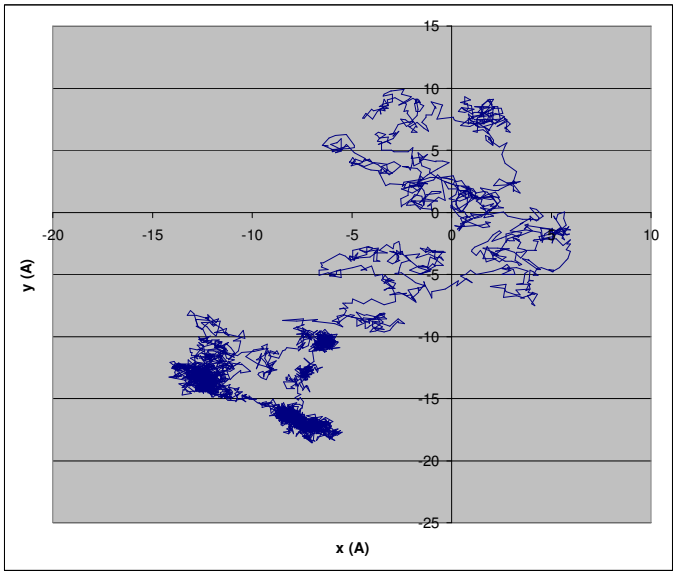

Figure C.64. x(Å) vs y(Å) coordinate of permeant's center of mass.

Octaonic acid

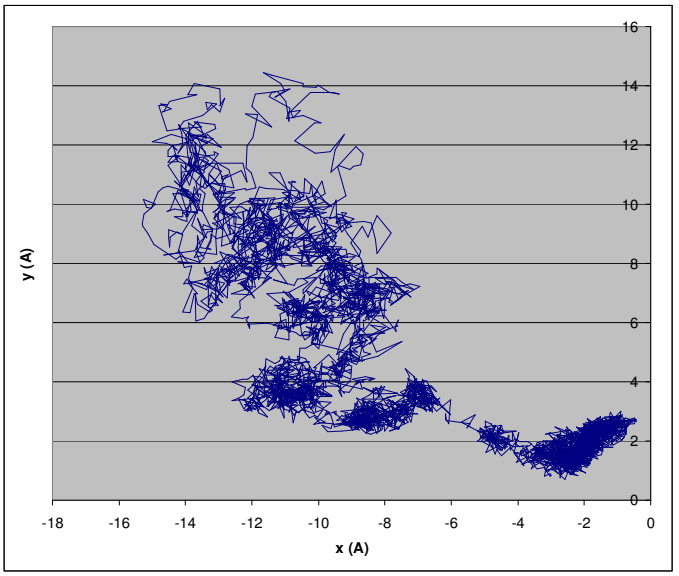

Figure C.65. x(Å) vs y(Å) coordinate of permeant's center of mass.

Ouabain

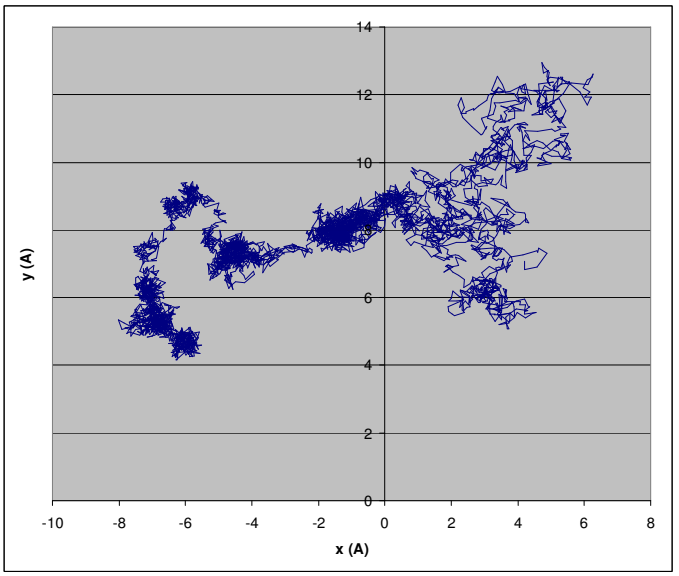

Figure C.66. x(Å) vs y(Å) coordinate of permeant's center of mass.

p-Cresol

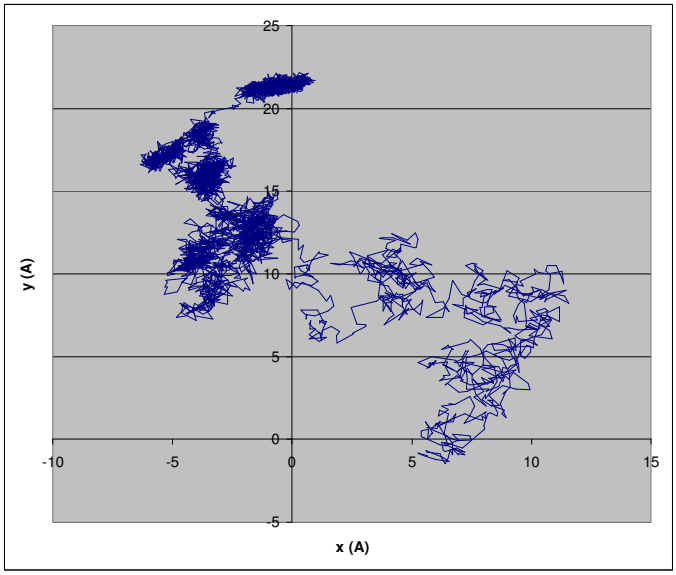

Figure C.67. x(Å) vs y(Å) coordinate of permeant's center of mass.

Pentanoic acid

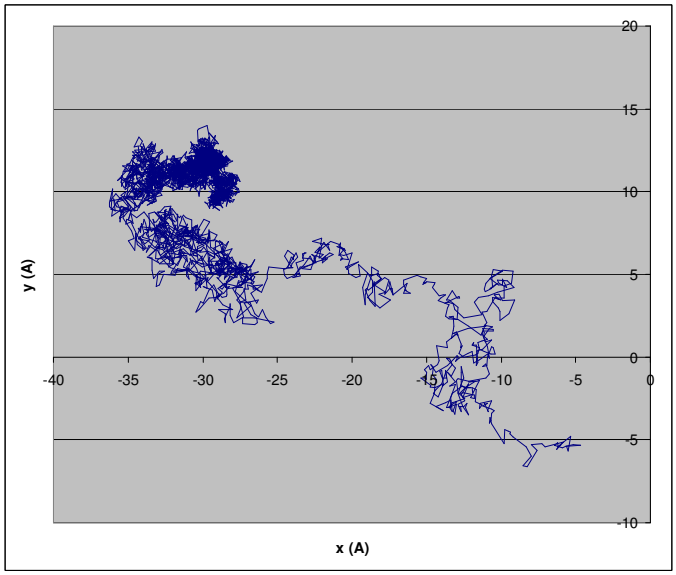

Figure C.68. x(Å) vs y(Å) coordinate of permeant's center of mass.

Phenobarbital

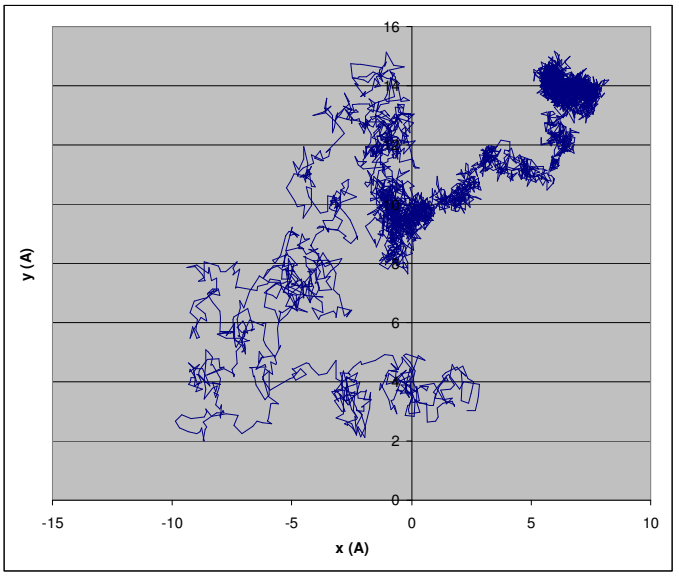

Figure C.69. x(Å) vs y(Å) coordinate of permeant's center of mass.

Phenol

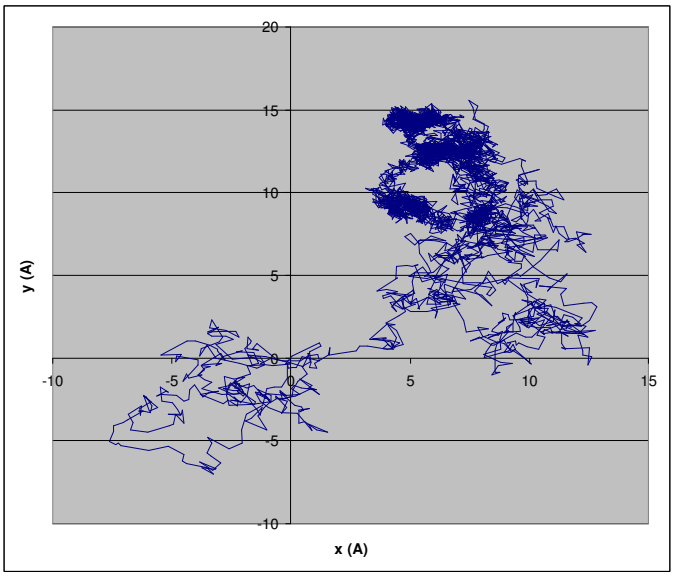

Figure C.70. x(Å) vs y(Å) coordinate of permeant's center of mass.

Pregnenolone

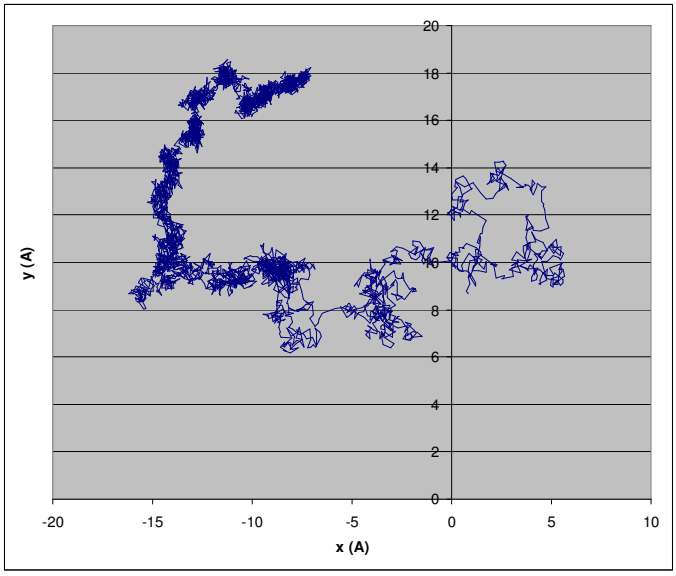

Figure C.71. x(Å) vs y(Å) coordinate of permeant's center of mass.

Progesterone

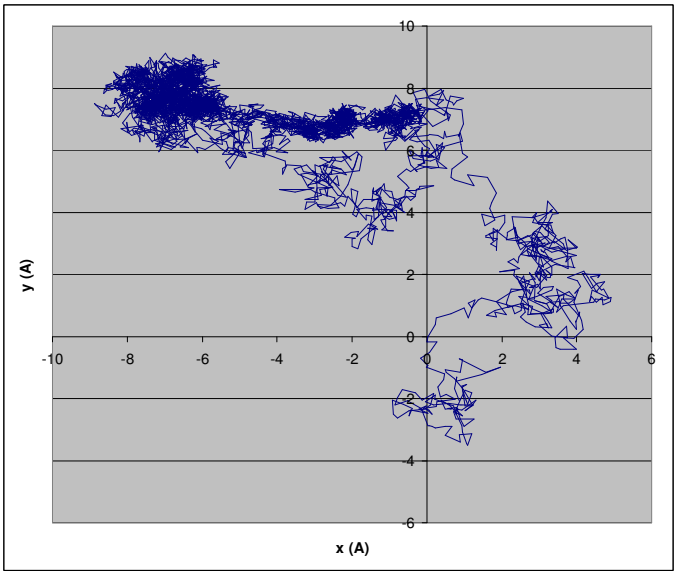

Figure C.72. x(Å) vs y(Å) coordinate of permeant's center of mass.

Resorcinol

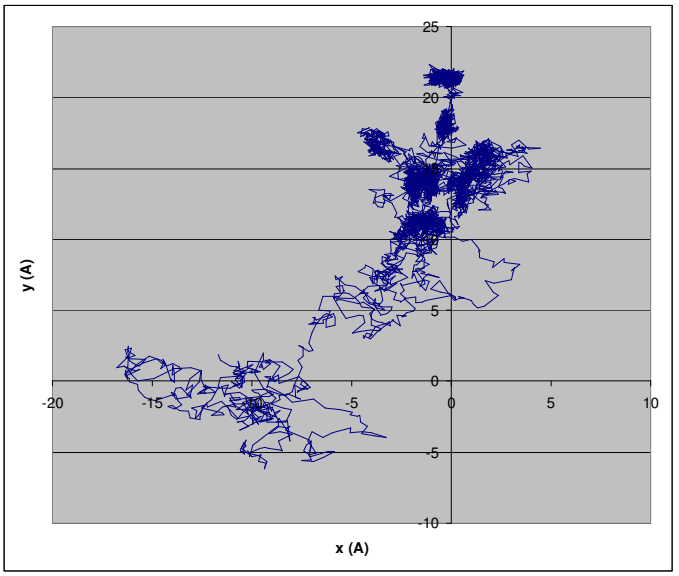

Figure C.73. x(Å) vs y(Å) coordinate of permeant's center of mass.

Salicylic acid

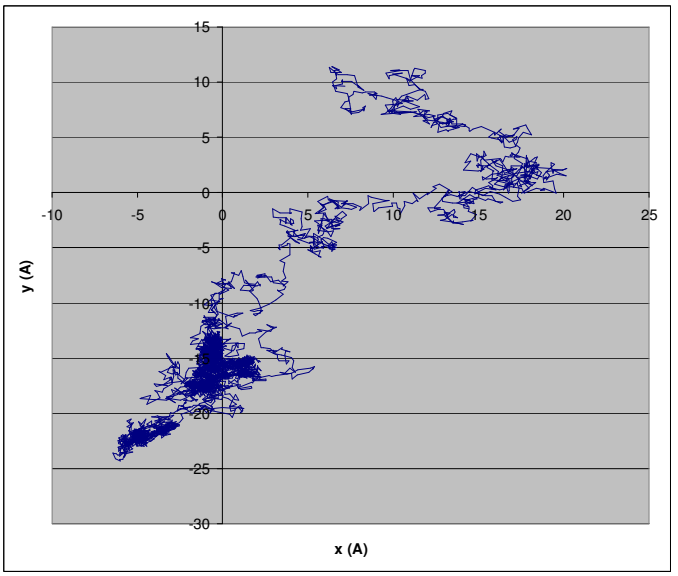

Figure C.74. x(Å) vs y(Å) coordinate of permeant's center of mass.

Sufentanyl

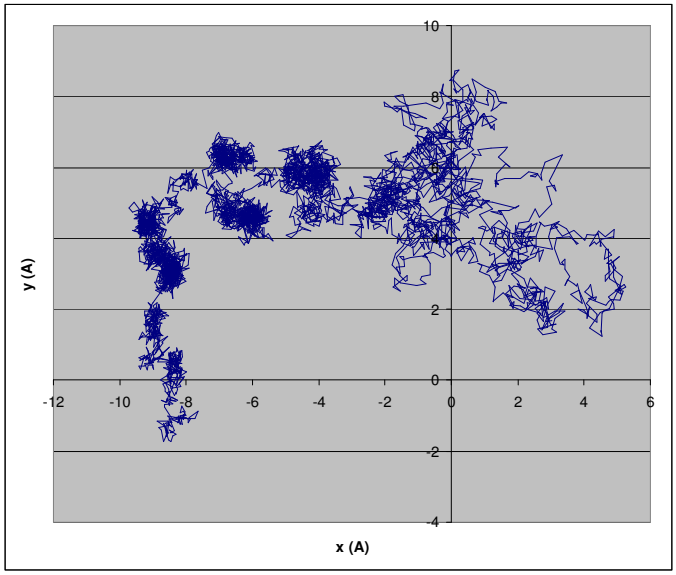

Figure C.75. x(Å) vs y(Å) coordinate of permeant's center of mass.

Testosterone

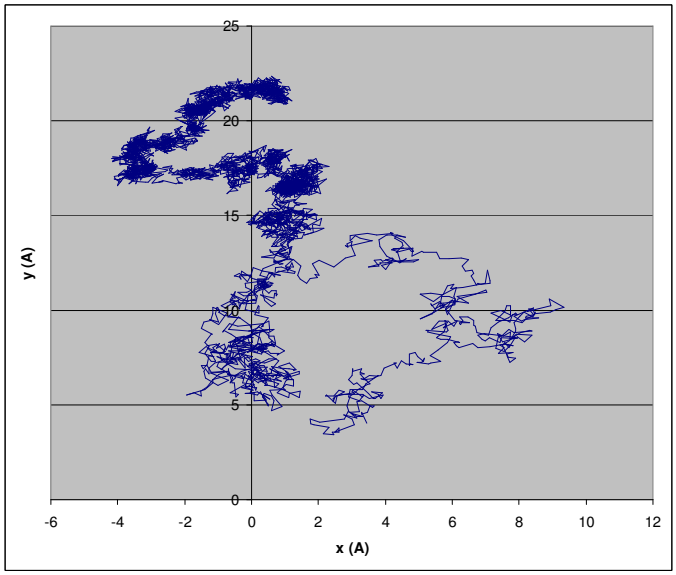

Figure C.76. x(Å) vs y(Å) coordinate of permeant's center of mass.

Thymol

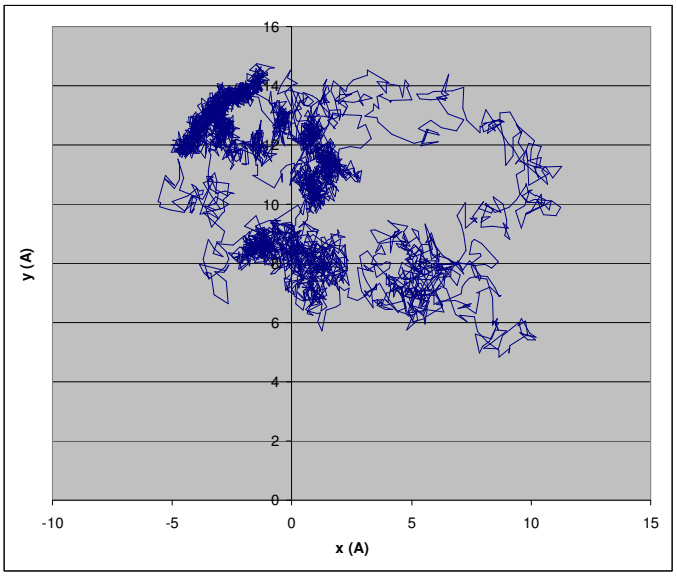

Figure C.77. x(Å) vs y(Å) coordinate of permeant's center of mass.

Toluene

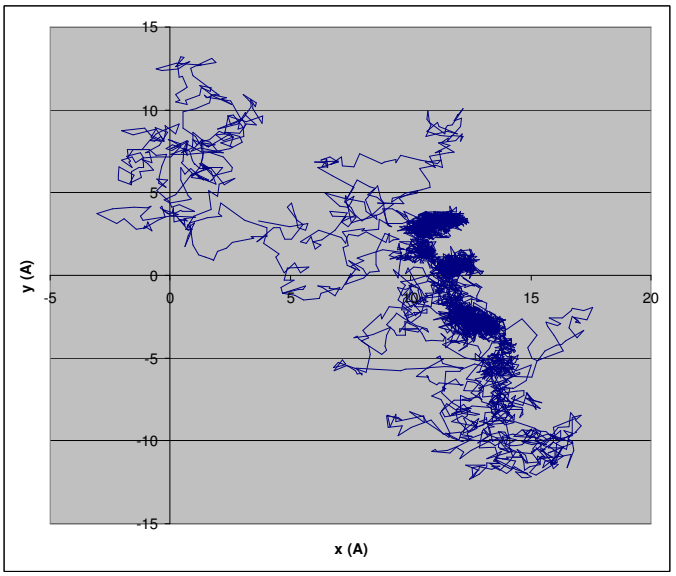

Figure C.78. x(Å) vs y(Å) coordinate of permeant's center of mass.

2,4,6-trichlorophenol

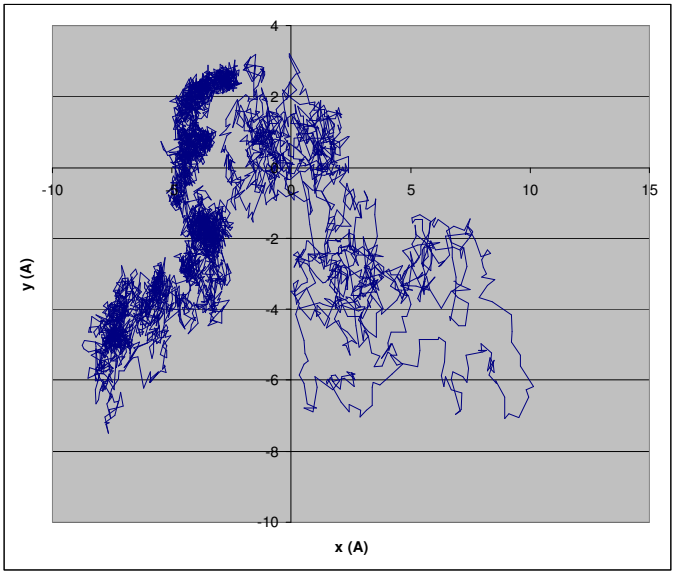

Figure C.79. x(Å) vs y(Å) coordinate of permeant's center of mass.

3,4-xylenol

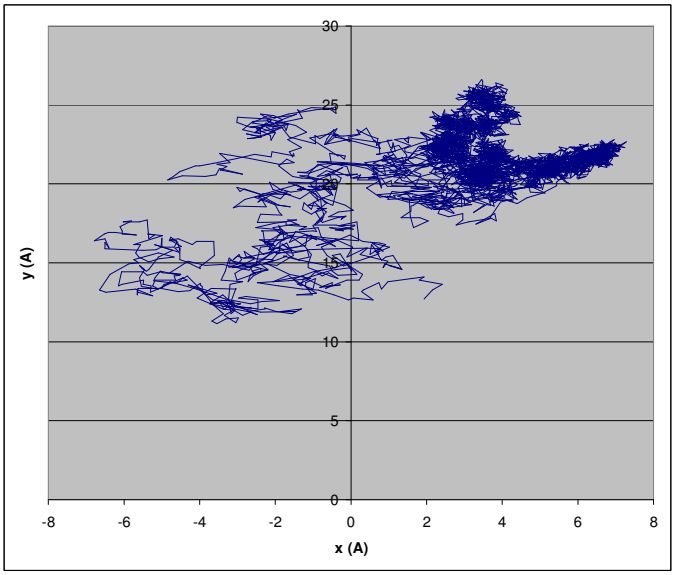

Figure C.80. x(Å) vs y(Å) coordinate of permeant's center of mass.

References

1 Lian G, Chen L, Han L. An evaluation of mathematical models for predicting skin permeability. J Pharm Sci 2008; 97(1): 584-98.

2 Michaels AS, Chandrasekaran SK, Shaw JE. Drug permeation through human skin: Theory and in vitro experimental measurement. AIChE J. 1975; 21: 985-96.

3 Iwai I, Han H, den Hollander L, Svensson S, Ofverstedt LG, Anwar J, Brewer J, Bloksgaard M, Laloeuf A, Nosek D, Masich S, Bagatolli LA, Skoglund U, Norlén L. The human skin barrier is organized as stacked bilayers of fully extended ceramides with cholesterol molecules associated with the ceramide sphingoid moiety. J Invest Dermatol. 2012; 132(9): 2215-25.

4 Bouwstra JA, Pilgram GSK, Ponec M. Does the Single Gel Phase Exist in Stratum Corneum ? J Invest Derm. 2002; 118: 897-898.

5 Bouwstra JA, Gooris GS, van der Speck JA, Bras W. Structural investigations of human stratum corneum by small-angle X-ray scattering. J Invest Dermatol. 1991; 97: 1005-1012.

6 Wertz P, Norle´n L (2003) ‘‘Confidence Intervals’’ for the ‘‘true’’ lipid compositions of the human skin barrier ? In: Forslind B, Lindberg M (eds.) Skin, Hair, and Nails Structure and Function. Marcel Dekker, New York, 85-106.

7 Das C, Olmsted PD, Noro MG. Simulation studies of stratum corneum lipid mixtures. Biophys J 2009; 97: 1941-51.

8 Potts RO, Guy RH. Predicting skin permeability. Pharm Res 1992; 9: 663-69.

9 Mitragotri S. A theoretical analysis of permeation of small hydrophobic solutes across the stratum corneum based on Scaled Particle Theory. J Pharm Sci. 2002; 91(3): 744-52.

10 Flynn GL. Physicochemical determinants of skin absorption. In Garrity TR, Henry CJ, Eds. Principles of route-to route extrapolation for risk assessment. New York: Elsevier 1990; 93-127.

11 Rothman S. The principles of percutaneous absorption. J Lab Clin Med 1943; 28: 1305-21.

12 Higuchi T. Physical chemical analysis of percutaneous absorption process from creams and ointments. J Soc Cosmet Chem 1960; 11: 85-97.

13 Higuchi T. Rate of release of medicaments from ointment bases containing drugs in suspension. J Pharm Sci 1961; 50874-875.

14 Orsi M, Essex JW. Passive permeation across lipid bilayers: a literature review. In Biggin PC, Sansom MSP, Eds. Molecular simulations and biomembranes: from biophysics to function. London: RSC Publishing 2010; 76-90.

15 Mitragotri S, Anissimov YG, Bunge AL, Frasch HF, Guy RH, Hadgraft J, Kasting GB, Lane ME, Roberts MS. Mathematical models of skin permeability: an overview. Int J Pharm 2011; 418: 115129.

16 Scheuplein R, Blank I. Permeability of the skin. Physiol Rev 1971; 51: 702-47.

17 Jepps OG, Dancik Y, Anissimov YG, Roberts MS. Modeling the human skin barrier — Towards a better understanding dermal absorption. Adv Drug Deliv Rev 2013; 65(2): 152-68.

18 Vecchia BE, Bunge A. Partitioning of Chemicals into the Skin: Results and Predictions. In Hadgraft J, Guy RH, Eds. Transdermal Drug Delivery. New York: Marcel Dekker 2002; 143-98.

19 Chen L, Han L, Lian G. Recent advances in predicting skin permeability of hydrophilic solutes. Adv Drug Deliv Rev 2013; 65(2): 295-305.
20 Naegel A, Heisig M, Wittum G. Detailed modeling of skin penetration--an overview. Adv Drug Deliv Rev 2013; 65(2): 191-207.
21 Frasch HF, Barbero AM. Application of numerical methods for diffusion-based modeling of skin permeation. Adv Drug Deliv Rev. 2013; 65(2): 208-20.
22 Scott HL. Modeling the lipid component of membranes. Curr Opin Struct Biol 2002; 12(4): 495-502.
23 Bassolino-Klimas D, Alper HE, Stouch TR. Solute diffusion in lipid bilayer membranes: an atomic level study by molecular dynamics simulation. Biochemistry. 1993; 32: 12624-37.
24 Bassolino-Klimas D, Alper HE, Stouch TR. Mechanism of solute diffusion through lipid bilayer membranes by molecular dynamics simulation. J Am Chem Soc. 1995; 117: 4118-29.
25 Alper HE, Stouch TR. Orientation and diffusion of a drug analogue in biomembranes: Molecular dynamics simulations. J Phys Chem. 1995; 99: 5724-31.
26 Marrink SJ, Berendsen HJC, Simulation of water transport through a lipid membrane, J Phys Chem. 1994; 98: 4155-4168.
27 Marrink SJ, Berendsen HJC. Permeation process of small molecules across lipid membranes studied by molecular dynamics simulations, J Phys Chem. 1996; 100: 16729-38.
28 Das C, Olmsted PD, Noro MG. Water permeation through stratum corneum lipid bilayers from atomistic simulations. Soft Matter. 2009; 5: 4549-555.
29 Allinger NL. Calculation of molecular structure and energy by force-field methods. Adv Phys Org Chem, 1976, 13:1-85.
30 Schlick T. Molecular Modeling and Simulation. An Interdisciplinary Guide. 1st ed. Springer, New York, 2010.
31 Waigh TA. The Physics of Living Processes. A Mesoscopic Approach, 1st ed. Wiley, New York, 2014.
32 Spinney R, Ford I. Fluctuation Relations: a Pedagogical Overview in Klages R, Just W, Jarzinsky C (eds.). Nonequilibrium Statistical Physics of Small Systems. Fluctuation Relations and Beyond. Wiley-VCH, Weinheim, 2013.
33 Notman R, Anwar J. Breaching the skin barrier--insights from molecular simulation of model membranes. Adv Drug Deliv Rev 2013; 65(2): 237-50.
34 Brooks BR, Brooks III CL, Mackerell AD, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, and Karplus M. CHARMM: The Biomolecular simulation Program. J Comp Chem 2009; 30: 1545-615.
35 Hermans J, Berendsen HJC, van Gunsteren WF, Postma JPM. A consistent empirical potential for water–protein interactions. Biopolymers. 1984; 23: 1513-1518.
36 Chandrasekhar I, Kastenholz M, Lins RD, Oostenbrink C, Schuler LD, Tieleman DP, van Gunsteren WF. A consistent potential energy parameter set for lipids: dipalmitoylphosphatidylcholine as a benchmark of the GROMOS96 45A3 force field. Eur Biophys J. 2003; 32(1): 67-77.
37 Siu SW, Vácha R, Jungwirth P, Böckmann RA. Biomolecular simulations of membranes: physical properties from different force fields. J Chem Phys 2008; 128(12): 125103.

38 Feller SE, Yin D, Pastor RW, MacKerell AD Jr. Molecular dynamics simulation of unsaturated lipid bilayers at low hydration: parameterization and comparison with diffraction studies. Biophys J 1997; 73: 2269-79.
39 Klauda JB, R. M. Venable RM, J. A. Freites JA, O'Connor JW, Tobias DJ, Mondragon-Ramires C, Vorobyov I, MacKerell AD, and R. W. Pastor RW. Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. J Phys Chem 2010; 114: 7830-43.
40 Berger O, Edholm O, Jahnig F. Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature. Biophys J. 1997; 72: 2002-–13.
41 Marrink SJ, Mark AE. Molecular view of hexagonal phase formation in phospholipid bilayers. Biophys J. 2005; 12(87): 3894-900.
42 Pandit SA, Scott HL. Multiscale simulations of heterogeneous model membranes, Biochim Biophys Acta. 2009; 1788: 136-48.
43 Hadley KR, McCabe C. A Simulation Study of the Self-Assembly of Coarse-Grained Skin Lipids. Soft Matter. 2012; 8(17): 4802-14.
44 Marrink SJ, de Vries AH, Mark AE. Coarse grained model for semi-quantitative lipid simulations. J Phys Chem B. 2004; 108: 750-60.
45 Izvekov S, Voth GA. A multiscale coarse-graining method for biomolecular systems. J Phys Chem B. 2005; 109(7): 2469-73.
46 Marrink SJ, Risselada J, Mark AE. Simulation of gel phase formation and melting in lipid bilayers using a coarse grained model. Chem Phys Lipids. 2005; 135: 223-44.
47 MacKerell, Jr. AD, Banavali N, Foloppe N. Development and current status of the CHARMM force field for nucleic acids. Biopolymers. 2001; 56 (4): 257-65.
48 Lyubartsev AP, Rabinovich AL. Recent Development in Computer Simulations of Lipid Bilayers. Soft Matter 2011; 7: 25-39.
49 Rabinovich AL, Lyubartsev AP. Computer Simulation of Lipid Membranes: Methodology and Achievements. Polymer Science Ser C. 2013; 55(1): 162-80.
50 Marrink SJ, de Vries AH, Tieleman DP. Lipids on the move: simulations of membrane pores, domains, stalks and curves. Biochim Biophys Acta. 2009; 1788: 149-68.
51 Orsi M, Sanderson W, Essex JW. Coarse-grain modelling of lipid bilayers: a literature review. In Hicks MG, C. Kettner C, Eds. Molecular Interactions - Bringing Chemistry to Life. Frankfurt: Beilstein-Institut 2007; pp. 185-205.
52 Kox J, Michels JPJ, Wiegel FW. Simulation of a lipid monolayer using molecular dynamics. Nature. 1980; 287: 317-19.
53 van der Ploeg P, Berendsen HJC. Molecular dynamics simulation of a bilayer membrane. J Chem Phys 1982; 76: 3271-76.
54 Egberts E, Berendsen HJC, Molecular dynamics simulation of a smectic liquid crystal with atomic detail. J Chem Phys 1988; 89: 3718-32.
55 Tieleman DP, Marrink SJ, Berendsen HJC. A computer perspective of membranes: molecular dynamics studies of lipid bilayer systems. Biochim Biophys Acta. 1997; 1331: 235-70.
56 Guo S, Moore TC, Iacovella CR, Strickland LA, McCabe C. Simulation study of the structure and phase behavior of ceramide bilayers and the role of lipid head group chemistry. J Chem Theory Comput. 2013; 9(11): 5116-26.
57 Höltje M, Förster T, Brandt B, Engels T, von Rybinski W, Höltje HD. Molecular dynamics simulations of stratum corneum lipid models: fatty acids and cholesterol. Biochim Biophys Acta. 2011; 1511: 156–67.

58 Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J. Interaction Models for Water in Relation to Protein Hydration. In Pullman B, Ed. Intermolecular Forces. Dordrecht: Reidel 1981; pp. 331–42.
59 Pandit SA, Scott HL. Molecular dynamics simulation of a ceramide bilayer. J Chem Phys. 2006; 124: 14708.
60 Notman R, den Otter RW, Noro MG, Briels WJ, Anwar J. The Permeability Enhancing Mechanism of DMSO in Ceramide Bilayers Simulated by Molecular Dynamics. Biophys J. 2007; 93: 2056–68.
61 Das C, Noro MG, Olmsted PD. Fast cholesterol flip-flop and lack of swelling in skin lipid multilayers. Soft Matter. 2014; 10(37): 7346-52.
62 Das C, Noro MG, Olmsted PD. Lamellar and inverse micellar structures of skin lipids: effect of templating. Phys Rev Lett 4. 2013; 111(14): 148101.
63 Imai Y1, Liu X, Yamagishi J, Mori K, Neya S, Hoshino T. Computational analysis of water residence on ceramide and sphingomyelin bilayer membranes. J Mol Graph Model. 2010; 29(3): 461-9.
64 Engelbrecht T, Hauss T, Suess K, Vogel A, Roark M, Feller SE, Neubert RHH, Dobner B. Characterisation of a new ceramide EOS species: synthesis and investigation of the thermotropic phase behaviour and influence on the bilayer architecture of stratum corneum lipid model membranes. Soft Matter. 2011; 7: 8998-9011.
65 Hoopes MI, Noro MG, Longo ML, Faller R. Bilayer structure and lipid dynamics in a model stratum corneum with oleic acid. J Phys Chem B. 2011; 115(12): 3164-71.
66 Metcalf R, Pandit SA. Mixing properties of sphingomyelin ceramide bilayers: a simulation study. J Phys Chem B. 2012; 116(15): 4500-9.
67 Bennett WF, Tieleman DP. Molecular simulation of rapid translocation of cholesterol, diacylglycerol, and ceramide in model raft and nonraft membranes. J Lipid Res. 2012; 53(3): 421-9.
68 Chiu SW, Vasudevan S, Jakobsson E, Mashl RJ, Scott HL Structure of sphingomyelin bilayers: a simulation study. Biophys J. 2003; 85(6): 3624-35.
69 Notman R, Anwar J, Briels WJ, Noro MG, den Otter WK. Simulations of skin barrier function: free energies of hydrophobic and hydrophilic transmembrane pores in ceramide bilayers. Biophys J. 2008; 95(10): 4763-71.
70 Berendsen HJC, van der Spoel D, van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. Comp Phys Comm. 1995; 91: 43-56.
71 Lindahl E, Hess B, van der Spoel D. GROMACS 3.0: A package for molecular simulation and trajectory analysis. J Mol Mod 2001; 7: 306-17.
72 Jorgensen WL, Maxwell DS, Tirado-Rives J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. J Am Chem Soc. 1996; 118(45): 11225-36.
73 Stewart JJP. Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters, J. Mol. Model. 2013; 19: 1-32.
74 Vistoli G, Pedretti A, Villa L, Testa B. Solvent constraints on the property space of acetylcholine. I. Isotropic solvents. J Med Chem. 2005; 48(6): 1759-67.
75 Gaillard P, P.A. Carrupt, B. Testa, A. Boudon Molecular Lipophilicity Potential, a tool in 3D QSAR: Method and applications. J Comput Aided Mol Des. 1994; 8(2): 83-96.
76 Jarzynski C. Nonequilibrium Equality for Free Energy Differences. Phys Rev Lett. 1997; 78(14): 2690-93.

77 Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K. Scalable molecular dynamics with NAMD. J Comput Chem. 2005; 26: 1781-802.

78 Pedretti A, Villa L., Vistoli G. VEGA: a versatile program to convert, handle and visualize molecular structure on Windows-based PCs. J Mol Graph. 2002; 21: 47-9.

79 Siu SW, Vácha R, Jungwirth P, Böckmann RA. Biomolecular simulations of membranes: physical properties from different force fields. J Chem Phys. 2008; 128(12): 125103.

80 Feller SE, Yin D, Pastor RW, MacKerell AD Jr. Molecular dynamics simulation of unsaturated lipid bilayers at low hydration: parameterization and comparison with diffraction studies. Biophys J. 1997; 73: 2269-79.

81 Vecchia BE, Bunge AL (2003). Skin Absorption Databases and Predictive Equations. In: Guy RH, Hadgraft J (eds) Transdermal Drug Delivery, 2nd ed. CRC Press, New York, pp 100-6.

82 Vistoli G, Pedretti A, Testa B. Partition coefficient and molecular flexibility: the concept of lipophilicity space. Chem Biodivers. 2009; 6(8):1152-69.

83 Xiang TX, Anderson BD. Molecular distributions in interphases: statistical mechanical theory combined with molecular dynamics simulation of a model lipid bilayer, Biophys J. 1994; 66: 561-72.

84 Xiang TX, Anderson BD. Liposomal drug transport: a molecular perspective from molecular dynamics simulations in lipid bilayers. Adv Drug Deliv Rev. 2006; 58(12-13): 1357-78.

85 Abrams C. Molecular Simulations. Case Study 3: Dynamical Properties: The Self-Diffusion Coefficient available at http://www.pages.drexel.edu/~cfa22/msim/node29.html. Last retrieved on 10/27/2015.

86 Berendsen HJC, Marrink SJ. Molecular dynamics of water transport through membranes: Water from solvent to solute. Pure and Appl Chem. 1993; 65: 2513-20.

87 Marrink SJ, Jähnig F, Berendsen HJC. Proton transfer across transient single-file water pores in a lipid membrane studied by molecular dynamics simulations. Biophys J. 1996; 71: 632-47.

88 Bemporad D, Essex JW, Luttmann C. Computer Simulation of Small Molecule Permeation across a Lipid Bilayer: Dependence on Bilayer Properties and Solute Volume, Size, and Cross-Sectional Area. Biophys J. 2004; 87: 1-13.

89 Bemporad D, Essex JW, Luttmann C. Behaviour of small solutes and large drugs in a lipid bilayer from computer simulations. Biochim Biophys Acta. 2005; 1718: 1-21.

90 Ulander J, Haymet ADJ. Permeation across hydrated DPPC lipid bilayers: simulation of the titrable amphiphilic drug valproic acid. Biophys J. 2003; 85: 3475-84.

91 Testa B, Vistoli G, Pedretti A, Bojarski AJ. Atomic diversity, molecular diversity, and chemical diversity: the concept of chemodiversity. Chem Biodivers. 2009 Aug;6(8):1145-51.

92 Vistoli G, Pedretti A, Mazzolari A, Testa B. In silico prediction of human carboxylesterase-1 (hCES1) metabolism combining docking analyses and MD simulations. Bioorg. Med. Chem. 2010; 18: 320-19.

93 Fisher RA. Statistical Methods for Research Workers, 5th ed. Oliver and Boyd, Edimburgh, 1936; pp 160-197.

94 Fisher, RA. Frequency distribution of the values of the correlation coefficient in samples of an indefinitely large population. Biometrika. 1915; 10 (4): 507–21.

95 Fisher, RA. On the "probable error" of a coefficient of correlation deduced from a small sample. Metron. 1921; 1: 3–32.