Chapter 3: Simulations of Molecular Ensembles

**3.3 Molecular Dynamics**

Main Idea: Moreover, since time-independent Hamiltonians are necessarily invariant to time reversal, a single phase point completely determines a full trajectory. As a result, phase space trajectories cannot cross themselves (since there would then be two different points leading away (in both time directions) from a single point of intersection).

Note that the complete set of all harmonic oscillator trajectories, which would completely fill the corresponding two-dimensional phase space, is composed of concentric ovals (concentric circles if we were to choose the momentum metric to be (mk) −1/2 times the position metric). Thus, as required, these (periodic) trajectories do not cross one another.

Thus, given a set of initial positions and momenta, and a means for computing the forces acting on each particle at any instant (and thereby deriving the acceleration), we have a formalism for ‘simulating’ the true phase-space trajectory.

In general, initial positions are determined by what a chemist thinks is ‘reasonable’ – a common technique is to build the system of interest and then energy minimize it partially (since one is interested in dynamical properties, there is no point in looking for an absolute minimum) using molecular mechanics. As for initial momenta, these are usually assigned randomly to each particle subject to a temperature constraint. The relationship between temperature and momentum is given by **Equation 3.18.**

*Time-independent Hamiltonians are necessarily invariant to time reversal; a single phase point completely determines a full trajectory.*

However, if we approximate Eqs. (3.10) and (3.12) as (this approximation, Euler’s, being exact in the limit of t → 0) we are offered a prescription for simulating a phase space trajectory. [Note that we have switched from the scalar notation of the one-dimensional harmonic oscillator example to a more general vector notation.

For systems more complicated than the harmonic oscillator, it is almost never possible to write down analytical expressions for the position and momentum components of the phase space trajectory as a function of time.

Certain aspects of this phase space trajectory merit attention. We noted above that a phase space trajectory cannot cross itself. However, it can be periodic, which is to say it can trace out the same path again and again; the harmonic oscillator example is periodic.

One interesting property of a phase point that has not yet been emphasized is that, since it is defined by the positions and momenta of all particles, it determines the location of the next phase point in the absence of outside forces acting upon the system. The word ‘next’ is used loosely, since the trajectory is a continuous curve of phase points (i.e., between any two points can be found another point) – a more rigorous statement is that the forward trajectory is completely determined by the initial phase point.

Chapter 3: Simulations of Molecular Ensembles

**3.2 Phase Space and Trajectories**

Main Idea: Over time, a dynamical system maps out a ‘trajectory’ in phase space. The trajectory is the curve formed by the phase points the system passes through.

The key to making this evaluation more tractable is to recognize that phase space is, for the most part, a wasteland. That is, there are enormous volumes characterized by energies that are far too high to be of any importance, e.g., regions where the positional coordinates of two different particles are such that they are substantially closer than van der Waals contact.

If we start a system at some ‘reasonable’ (i.e., low-energy) phase point, its energy-conserving evolution over time (i.e., its trajectory) seems likely to sample relevant regions of phase space.

*Over time, a dynamical system maps out a ‘trajectory’ in phase space. The trajectory is the curve formed by the phase points the system passes through.*

The ‘**ergodic hypothesis**’ assumes Eq. (3.9) to be valid and independent of choice of t0. It has been proven for a hard-sphere gas that Eqs. (3.5) and (3.9) are indeed equivalent (Ford 1973). No such proof is available for more realistic systems, but a large body of empirical evidence suggests that the ergodic hypothesis is valid in most molecular simulations.

From a mathematical standpoint, Eq. (3.6) shows that a high-energy phase point has a near-zero probability, and thus the integrand of Eq. (3.5) will also be near-zero (as long as property A does not go to infinity with increasing energy). As the integral of zero is zero, such a phase point contributes almost nothing to the property expectation value, and simply represents a waste of computational resources. So, what is needed in the evaluation of Eqs. (3.5) and (3.7) is some prescription for picking important (i.e., high-probability) points.

Chapter 3: Simulations of Molecular Ensembles

**3.1 Relationship Between MM Optima and Real Systems**

Main Idea: The point to be re-emphasized here is that the vast majority of experimental techniques measure molecular properties as averages- either time averages or ensemble averages or, most typically, both. Thus, we seek computational techniques capable of accurately reproducing these aspects of molecular behavior.

One has to accept the *flexibility of PEG as an intrinsic characteristic* of the molecule, and any attempt to understand its other properties must account for its structureless nature.

If a *molecule is quite small* and is characterized by fairly ‘stiff’ molecular coordinates, then its ‘well’ on the PES will be ‘narrow’ and ‘deep’ and the range of structures it samples will all be fairly close to the minimum-energy structure.

*The point to be re-emphasized here is that the vast majority of experimental techniques measure molecular properties as averages- either time averages or ensemble averages or, most typically, both. Thus, we seek computational techniques capable of accurately reproducing these aspects of molecular behavior.*

The relative probability of clustering about any given minimum is a function of the temperature and some particular thermodynamic variable characterizing the system (i.e. Helmholtz free energy), that variable depending on what experimental conditions are being held constant (i.e. temperature and volume). Those variables being held constant define the ‘**ensemble**’.

Chapter 2: Molecular Mechanics

**2.5 Menagerie of Modern Force Fields**

Main Idea: In the context of molecular dynamics simulations of proteins, the term "force field" refers to the combination of a mathematical formula and associated parameters that are used to describe the energy of the protein as a function of its atomic coordinates.

*Table 2.1* contains an alphabetic listing of force fields which for the most part continue to be in use today. *Page 50* also goes on to describe the sometimes confusing naming system given to all available force fields.

**How do I know I can trust the results?** The process of testing the utility of a force field for molecules other than those over which it was parameterized is known as ‘**validation**’.

*In the context of molecular dynamics simulations of proteins, the term "force field" refers to the combination of a mathematical formula and associated parameters that are used to describe the energy of the protein as a function of its atomic coordinates.*

**How do I pick the best force field for my problem?** One should pick the force field that has been previously shown to be most effective for the most closely related problem one can find. That demonstration of effectiveness may have taken place within the process of parameterization or by post-development validation.

Chapter 2: Molecular Mechanics

**2.4 Geometry Optimization**

Main Idea: One priority in force-field development is to adopt reasonably simple functional forms so as to facilitate geometry optimization. In the case of geometry optimization, we are working with far fewer variables (the geometric degrees of freedom) and have, at least with a force field, analytic expressions for how the energy depends on the variables.

What we really want to do at any given point in the multi-dimensional case is move not in the direction of a *single* coordinate, but rather in the direction of the greatest downward slope in the energy with respect to *all* coordinates. This direction is the opposite of the gradient vector **g**, which is defined on page 43.

When one does not know how the energy depends on the geometric coordinates of our molecule, to **optimize the geometry** all one can do is keep trying different geometries until we are sure we have found the one with the *lowest possible energy*. **See details of process on page 42.**

The condition for a minimum is that all coordinate second derivatives (i.e. all diagonal elements of the Hessian matrix) be positive. In order to verify the nature of a located stationary point, it is necessary to compute an accurate Hessian matrix and inspect its eigenvalues.

*One priority in force-field development is to adopt reasonably simple functional forms so as to facilitate geometry optimization.*

In the most favorable cases, FMM methods scale linearly with system size.

In aperiodic systems, another important contribution has been the development of a so-called ‘**Fast Multipole Moment**’ **(FMM)** method, which takes advantage of the significant cancellations in charge-charge interactions between widely separated regions in space, and the increasing degree to which those interactions can be approximated by highly truncated multipole-multipole interactions.

The bottom line for all **transition state (TS)** structure location methods is that they work best when the chemist can provide a reasonably good initial guess for the structure, and they tend to be considerably more sensitive to the availability of a good Hessian matrix, since finding the TS essentially amounts to distinguishing between different local curvatures on the PES.

Chapter 2: Molecular Mechanics

**2.3 Force-field Energies and Thermodynamics**

Main Idea: One measure of the accuracy of a force field can be its ability to predict heats of formation. Key point that must be noted is that strain energies for two different molecules cannot be meaningfully compared unless the zero of energy is identical.

Reviewing equations 2.4 and 2.9, the minimum value for the energy is zero (assuming positive force constants and sensible behavior for odd power terms). An energy of zero is obtained when the bond length or angle adopts its equilibrium value. Thus, a ‘**strain-free**’ molecule is one in which every coordinate adopts its **equilibrium** value.

Neglecting non-bonded terms, we see that the raw force-field energy can be called the ‘**strain energy**’, since it represents the positive deviation from a hypothetical strain-free system.

When a molecule contains one or more of these atom types, the force field cannot compute a molecular heat of formation.

*One measure of the accuracy of a force field can be its ability to predict heats of formation.*

For some atom types, thermodynamic data may be lacking to assign a reference heat of formation.

What is necessary to compute a **heat of formation**, is to define the heat of formation of each hypothetical unstrained atom type. The molecular heat of formation can then be computed as the sum of the heats of formation of all the atoms types plus the strain energy.

Thus, a typical *force-field energy calculation will report* any or all of (1) a **strain energy**, which is the energetic consequence of the deviation of the internal molecular coordinates from their equilibrium values, (2) a **force-field energy**, which is the sum of the strain energy and the non-bonded interaction energies, and (3) a **heat of formation**, which is the sum of the force-field energy and the reference heats of formation for the constituent atom types. **See Figure 2.8 page 41.**

Chapter 2: Molecular Mechanics

**2.2 Potential Energy Functional Forms**

Main Idea: Vibrational spectroscopy reveals that, for small displacements from equilibrium, energy variations associated with bond angle deformation are as well modeled by polynomial expansions as are various associated with bond stretching.

In the absence of opportunities for bonding interactions, **Pauli repulsion** (or ‘exchange repulsion’) causes the energy of the system to rise rapidly with decreasing bond length. The force associated with this interaction is referred to as ‘dispersion’, ‘**London**’ force, or the ‘attractive van der Waals’ force. Bonds between heteroatoms and hydrogen atoms are amongst the most polar found in non-ionic systems. This polarity is largely responsible for the well-known phenomenon of **hydrogen bonding**, which is a favorable interaction between a hydrogen and a heteroatom to which it is not formally bonded.

How to construct a molecular PES is to take the sum of energies from chemically intuitive functional forms that depend on internal coordinates and on atomic properties.

When working with force fields a bond is *defined* to be a vector connecting two atoms.

*Vibrational spectroscopy reveals that, for small displacements from equilibrium, energy variations associated with bond angle deformation are as well modeled by polynomial expansions as are various associated with bond stretching.*

The barriers of rotational potential are primarily associated with *steric* interactions between eclipsing atoms/groups. An example of zero potential energy is the mutual approach of two noble gases, which at infinite separation have no interaction between them (ideally).

A **force field** is nothing but a collection of functional forms and associated constants. With that collection in hand, the energy of a given molecule (whose atomic connectivity must in general be specified) can be evaluated by computing the energy associated with every defined type of interaction occurring in the molecule.

**Step 1 of parameterization** is to assemble the experimental data. For molecular mechanics, these data consist of structural data, energetic data, and possibly data on molecular electric moments.

**Step 2 of parameterization** is to define a ‘penalty function’, that is a function that provides a measure of how much deviation there is between our predicted values and our experimental values. Our goal with then be to select force-field parameters that minimize the penalty function.

Chapter 2: Molecular Mechanics

**2.1 History and Fundamental Assumptions**

Main Idea: To a computational chemist, the PES is a surface that can be generated point by point by use of some computational method which determines a molecular energy for each point’s structure.

The **allowed transitions** from ground state (v=0) to the first vibrationally excited state (v=1) are monitored by absorption spectroscopy; the typical photon energy for the excitation falls within the infrared region of light.

Vibrational spectroscopy measures the energy separations between different vibrational levels, which are quantized.

The spacings between the various vibrational energy levels depend on the potential energy associated with bond stretching.

Energetically unfavorable non-bonded, non-angle-bending interactions have come to be called ‘**steric effects**’.

*To a computational chemist, the PES is a surface that can be generated point by point by use of some computational method which determines a molecular energy for each point’s structure.*

Any classical mechanical bond can deviate from its equilibrium bond length due to other energies of interaction having to be considered (repulsive van der Waals interactions) as well as in regards to bond angles (transferable force constants and optimal values).

‘**Transferability**’ is the phenomenon that states that the force constants and equilibrium bond lengths are largely the same from one molecule to the next. This allowed for a reasonable approximation to be made that molecular **enthalpies** could be determined as a sum of bond enthalpies.

More sensitive experimental apparati are needed to observe other allowed absorptions (or emissions) between more highly excited vibrational states or **forbidden transitions** (between states differing by more than 1 vibrational quantum number).