Computational Modeling of Protein-Ligand Interactions

Steven R. Gwaltney
Department of Chemistry
Mississippi State University
Mississippi State, MS 39762
“Every attempt to refer chemical questions to mathematical doctrines must be considered, now and always, profoundly irrational, as being contrary to the nature of the phenomena. . . . but if the employment of mathematical analysis should ever become so preponderant in chemistry (an aberration which is happily almost impossible) it would occasion vast and rapid retrogradation, by substituting vague conceptions for positive ideas, and an easy algebraic verbiage for a laborious investigation of facts.”
“The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble.”
Why the Change?

Quantum Mechanics

- Postulated by Schrödinger in 1926
- Time dependent version $\frac{\hbar}{i} \frac{\partial}{\partial t} = H$
- Time independent version $H \Psi = E \Psi$
- Partial differential equations
- No exact solutions for real systems
Approximate

- We can’t solve the Schrödinger equation for molecules.
- The trick is to choose appropriate approximations – tradeoff of time versus accuracy
- “The right answer for the right reason”
Theory’s Family Tree

Theoretical Chemistry

- Electronic Structure Theory
- Dynamics
- Statistical Mechanics
- Ab Initio
- Quantum Dynamics
- Molecular Dynamics
- Density Functional Theory
- Semiempirical Quantum Chemistry
The Three Main Branches

■ **Electronic Structure Theory**
  - Uses the time independent Schrödinger equation to describe the molecule’s electron configuration
    ▪ Can calculate energies, geometries, vibrational frequencies, dipole moments, NMR spectra, etc.

■ **Dynamics**
  - Studies how the system changes over time
    ▪ Uses either quantum mechanics or Newtonian mechanics

■ **Statistical Mechanics**
  - Studies the average behavior of complex ensembles
    ▪ Often used for liquids, polymer melts, similar systems
Theory’s Family Tree

Theoretical Chemistry

- Electronic Structure Theory
  - Semiemperical
  - Density Functional Theory
  - Ab Initio

- Dynamics
- Statistical Mechanics

- Quantum Dynamics
- Molecular Dynamics
The Dynamics Siblings

- Quantum Dynamics uses time dependent Schrödinger equation
  - Can only handle up to four degrees of freedom

- Classical Dynamics moves atoms by \( F=ma \)
  - Describe systems of several thousand atoms
  - Uses molecular mechanics force fields
Molecular Mechanics

- Describes bond lengths and bond angles as springs
- Also includes terms for out of plane bends, torsions, electrostatics, hydrogen bonds, and van der Waals interactions
- Very fast
- Parameters chosen to fit certain classes of molecules
- Can’t break bonds
An Example
$S_{N2}$ Reaction

Reactant → Transition State → Product
Theory’s Family Tree

Theoretical Chemistry

Electronic Structure Theory
- Semiempirical
- Density Functional Theory
- Ab Initio

Dynamics
- Quantum Dynamics
- Molecular Dynamics

Statistical Mechanics
Semiemperical Methods

- Molecular Hamiltonian consists of 4 terms:
  - Kinetic energy of the electrons
  - Nuclear-nuclear repulsion
  - Electron-nuclear attraction
  - Electron-electron repulsion

- Semiemperical methods throw out most of the two-electron integrals and parameterize the rest of the terms.
  - Different parameters for different properties

- Speed advantage is diminishing.

- Importance of methods is decreasing.
Ab Initio Methods

- No experimental data used to fit results
- Simplest method is Hartree-Fock
  - Electrons move in the average electric field produced by the other electrons
  - Origin of the molecular orbital picture
  - Formally scales as system size to the fourth, in practice much cheaper
  - Neglects the instantaneous correlation of electron motions
Correlated Methods

- Add in missing correlation energy
- Equations look like either a large system of nonlinear equations (CC) or a large eigenvalue/eigenvector problem (CI)
- Best methods are very accurate and very costly
  - Errors as low as 0.2 kcal/mol for atomization energies and 0.004 Å for bond lengths
  - Cost scales as system size to the seventh power
  - Limited to less than 20 atoms
- We know how to converge to the exact solution
Density Functional Theory

- Describe system via electron density (3 variables) instead of wave function (3n variables)
- Existence proof for exact form
- Practical methods use a few parameters and fit to experimental data
- Errors of around 3 kcal/mol for atomization energies
DFT Continued

- Solved self consistently
- Formally scale as system size to the fourth, but linear scaling versions have been developed
- Can handle up to a couple hundred atoms
- Rapidly becoming the workhorse method of computational chemistry
DFT, Part 3

Form of functional

\[ E[\sigma] = T_s[\sigma] + E_J[\sigma] + E_{xc}[\sigma] \]

No one knows how to get the exact \( E_{xc}[\sigma] \).
- Instead, approximations must be used.

A veritable plethora of exchange-correlation functionals exist.
- Often difficult to tell which one works best
- No way to converge to the exact answer
A Note On Basis Sets

- The wave function (or density) is expanded in terms of Gaussian-shaped orbitals centered on each atom.
- Sets of standard basis sets exist.
  - These vary primarily by the number of basis functions on each atom.
- Bigger basis sets equal:
  - Better answers
  - Longer calculations
$S_{\text{N}2}$ Revisited

A quantum treatment can break the bond.
“Usually, a poison has a specific molecule with which it interacts and it is that interaction that causes the toxicity.”

Russell Carr
Organophosphate Insecticides

- Very heavily used, especially in agricultural areas
- Acts by reacting with the active site of the enzyme acetylcholinesterase
- Acute exposure to OP agents can lead to vomiting, muscle twitches, convulsions, and even death.
- Closely related to nerve gasses, both in structure and in mode of action
Chlorpyrifos
The neurotransmitter acetylcholine (ACh) is the primary signal carrier in cholinergic nerve/nerve and nerve/muscle junctions. Acetylcholinesterase (AChE) breaks down ACh, causing the nerve signal to terminate. AChE exists in vivo as a membrane bound monomer, a dimer, and a tetramer.
The chemical structure of the toxicant before it enters your body is often well known.

- However, in vivo is the parent or a metabolite the active species?

The structure of a protein is much harder to determine.

No general method exists to go from the sequence to the tertiary structure of a protein.

- Nobel Prize is waiting!
The two primary ways of experimentally determining the structure of a protein are X-ray crystallography and NMR studies.

Journals require authors to submit solved structures to a central repository, the Protein Data Bank (PDB).

Structures from the PDB are available free of charge.
Mouse AChE

Tetramer with 17,000 non-hydrogen atoms
Single Monomer

547 amino acids, 4,300 non-hydrogen atoms
What Do We Want to Know?

- Once we have structures, we need to decide what information we want to learn.
- This determines what methods we should use for our calculations.
A Little Physical Chemistry

- $E+S \overset{K_A}{\underset{?}{\rightleftharpoons}} ES \overset{k_p}{\rightarrow} EP$

- $K_A$ is the equilibrium constant for enzyme/substrate association
  - $K_A = e^{-\frac{G_b}{RT}}$

- $k_p$ is the rate of product formation
  - $k_p = Ae^{-\frac{E_a}{RT}}$
Reaction Diagram

Need three points

E+S

ES

Transition state

E_a

E+S

?G_b

Transition state

E_a

EP
The Problem

1. Enzymes are too big to study with quantum mechanics.
2. Molecular mechanics can’t break bonds.
3. How do we bridge the gap?
Combine the Two

“For every problem there is a solution which is simple, obvious, and wrong”

Albert Einstein
QM/MM

Problems

- How do you define the border?
- How do you couple the two regions together?
Can we cut out a piece of the enzyme?
- The piece must be small enough to calculate.
- The piece must be able to describe the chemistry.
AChE Active Site

6 amino acids, 42 non-hydrogen atoms
The Role of the Rest

- Active site is 6 out of 547 amino acids.
- The rest of a protein serves to hold the active site and the substrate in an optimal configuration.
- It also provides a polarized environment, allosteric interactions, and gross conformational changes.
A Bigger Piece

26 amino acids, 214 non-hydrogen atoms
So, what do we do?

- Use linear scaling DFT calculations to calculate a “chunk” of the enzyme
- Big basis set in the middle - small basis set at the edge
Not Quite so Simple

1. The multiple minimum problem
2. How does the substrate fit it?
3. Where are the waters?
Use MD simulations to provide initial geometries for DFT studies

- Easy to add water molecules to the simulation
  - Can then put them into the DFT calculations in the right places
- Allow the enzyme to relax in the presence of the substrate
- Can give us multiple starting structures if multiple important structures exist
One Final Quote

“In theory, there is no difference between theory and practice; in practice, there is.”

Chuck Reid