BS3008: Computer Aided Drug Discovery

Kevin Fo

Table of contents

Preface

This is a Quarto book website (authored in the form of a website) that I (i.e., Kevin) has authored for the SBS module *BS3008: Computer Aided Drug Discovery*. As of the time of writing, BS3008 is a three **a**cademic **u**nit (i.e., **AU**) module available to SBS students as a core module.

BS3008 taught by professor Mu Yuguang.

Part I

PART 1 : LECTURES

1 Theoretical Foundations of BS3008

This week's (i.e., week 1) lecture aims to provide an introduction to BS3008's course contents by explaining its various theoretical aspects.

1.1 Discplines in BS3008

BS3008 covers numerous disciplines, including the following. A brief explanation on what each discipline is is also provided for each of the disciplines:

1. **Chemoinformatics**

   This discipline deals with similarities and differences between chemical compounds.

2. **Bioinformatics**

   This disciplines applies informatics tools (e.g., Python coding) to Biological molecules and data.

3. **Theoretical Chemistry (i.e., Quantum Chemistry)**

   This discipline provides the theoretical foundations needed to understand the course's contents.

4. **Computational Chemistry and Biology**

   This discipline not only encompasses theoretical chemistry, but also molecular mechanics, minimization, simulations, and conformational analysis.

5. **Molecular Modelling**

   This discipline uses all of the above disciplines to represent and manipulate the structures of molecules.

Hence, BS3008 primarily focuses on molecular modelling (with emphasis on theoretical chemistry for the theoretical component of the course).

1.1.1 What is Molecular Modelling?

According to Tamar Schlick, molecular modelling is:

> "...the science and art of studying molecular structure and function through model building and computation."
>
> – Tamar Schlick

"Computation" in this sense refer to practices such as:

1. ab initio and semi-empirical quantum mechanics
2. Molecular mechanics
3. Monte Carlo simulations
4. Molecular dynamics
5. Free energy and solvation methods
6. Structure / activity relationships (i.e., SAR analyses)
7. Chemical / biochemical information and databases

It is important to understand that while "model building" can be as simple as using plastic or metal rods to depict molecules' structures, it can also be as sophisticated as an interactive, animated color graphics and lasers.

Nonetheless, the computational tools used in molecular modelling is just as, if not more complex than Biological systems. However, the concepts in molecular modelling must be carefully applied and one must also be wary of molecular modeling's strengths and weaknesses.

1.1.2 Important Databases and Tools

Professor Mu lists some important molecular modelling tools in this chapter - click on their hyperlinks to access them:

1. **PDB**

   This is a database with numerous entries on proteins' information.

2. **PDBBinding**

   This is another database that provides entries on the binding affinity for all biomolecular complexes.

3. **ZINK DOCK**

4. **Autodock Zina**

   This is an open-source program for performing molecular docking.

1.1.3 Molecular Mechanics in Molecular Modelling

Molecular modelling started with the idea that molecular geometry, energy, and other molecular properties could be calculated from models (that are influenced by basic forces).

A **molecule** - hence - is a system of particles (i.e., atoms) connected by springs (i.e., bonds). This molecule is free to rotate, vibrate, and adopt a favorable conformation in space as a result of the inter- *and* intramolecular forces acting upon it.

1.2 Structure, Topology, Motion, Functions, and Potential Energy

This section aims to illustrate how different energy functions can influence the behavior of particles in a system.

1.2.1 Case #1

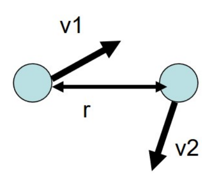

Figure 1.1: A Simple System to Consider

The potential and kinetic energy E_p and E_k respectively in this system is given by:

$$E_p = E_p(\vec{x}) = \sum_i f_i(x, y, z) \tag{1.1}$$

$$E_k = \frac{1}{2} \cdot (m_1 v_1^2 + m_2 v_2^2) \tag{1.2}$$

Where $\vec{x}$ is the system and $f_i(x, y, z)$ a function that calculates the potential energy for each particle in the system (i.e., each atom). Hence, we can say that:

$$E_{tot} = E_p + E_k \tag{1.3}$$

Where E_{tot} is the total energy of the system.

Since r represents the distance between both molecules, therefore $E_p(r) = 0$.

$$\vec{F} = \frac{\partial E_p(\vec{x})}{\partial \vec{x}} \tag{1.4}$$

The force[1] $\vec{F}$ on the system is denoted via the above equation.

Since $E_p(r) = 0$ in the first figure, it follows that $F(r) = 0$.

1.2.2 Case #2

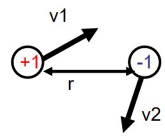

Figure 1.2: A System with Two Charged Molecules

Here, the charge V_{ele} and the potential energy $E_p(r)$ is given via the following equations:

$$V_{ele}(r) = \frac{1}{4\pi\epsilon_0} \cdot \frac{q_1}{r} \tag{1.5}$$

$$E_p(r) = \frac{1}{4\pi\epsilon_0} \cdot \frac{q_1 q_2}{r} \tag{1.6}$$

Some important considerations to think about include the variables and the parameters of the system.

1.2.3 Case #3

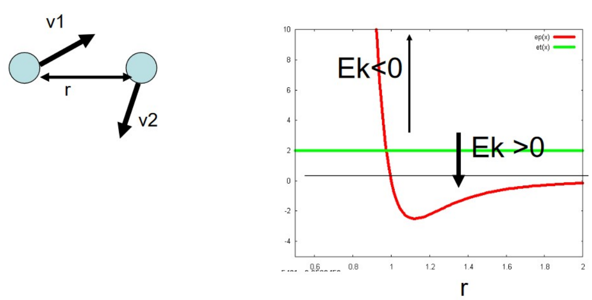

Figure 1.3: A System with Two Molecules and their Energy Graph

[1]Note that $\vec{F}$ is caused by a change in potential energy, not by potential energy itself!

In this system, we let:

$$E_p(r) = 4\epsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^{6} \right] \tag{1.7}$$

In this system, we also note that $E_p(r) = 0$ when $r = \sigma$ and that $E_p(\sqrt[6]{2}\sigma) = -\epsilon$.

Do also consider the variables and parameters in this system (and whether or not both particles in this system can move freely).

1.2.4 Case #4

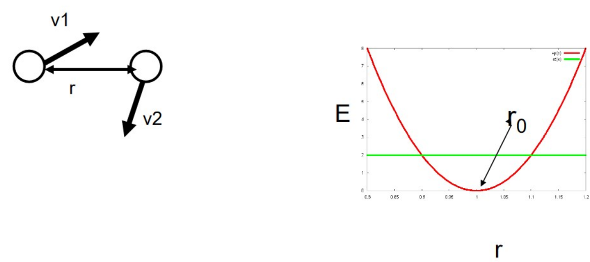

Figure 1.4: A System with Two Molecules and their Energy Graph

Here, we define the system's potential energy E_p as:

$$E_p(r) = \frac{1}{2}k(r - r_0)^2 \tag{1.8}$$

While the energy graph for this system appears to be that of bonding, it is still important to consider the variables and the parameters of the system.

We can also further decompose the above equation to its spatial components x, y, and z and say that:

$$E_p(r) = \frac{1}{2}k(x - x_0)^2 + \frac{1}{2}k(y - y_0)^2 + \frac{1}{2}k(z - z_0)^2 \tag{1.9}$$

In this sub-case, the system appears to be a lattice. However, are the particles still movable?

1.3 Professor Mu's Current Works

As of the time of writing, professor Mu's lab is currently focused on the following topics:

1. Amyloidogenic protein / peptide aggregation and misfolding
2. DNA-DNA, DNA-ions, and DNA-protein interactions
3. Drug-protein interaction and drug candidate screening
4. Peptide-membrane interactions.

For more information on professor Mu's current research topics, do visit his lab's homepage.

2 Force Fields

A **force field** implies that a molecule's atoms are a collection of different matter interacting with one another via *forces described by empirical energy functions.* This is unlike quantum mechanical calculations: the electrons and atoms' nuclei are not explicitly included in such calculations.

Force fields provide a fast computational method that works for small and big molecules alike (and even complex molecular systems).

2.1 Typical Force Fields

A typical force field $U(\{r_{ij}\})$ has the following formula:

$$U(\{r_{ij}\}) = \sum_j \frac{k_j^l}{2}(l_j - l_j^0)^2 + \sum_j \frac{k_j^\delta}{2}(\delta_j - \delta_j^0)^2 + \sum_{torsions} \frac{V_n}{2}(1 + \cos(n\phi - \gamma)) \tag{2.1}$$

$$+ \sum_{i,j=1}^{N} \frac{q_i q_j}{r_{ij}} + \sum_{i,j=1}^{N} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] \tag{2.2}$$

Or in more layman terms:

$$\text{Force field} = \text{bond stretching} + \text{valence angle bending} \tag{2.3}$$

$$+ \text{torsions} + \text{Electrostatic charges} \tag{2.4}$$

$$+ \text{van der Waals forces} \tag{2.5}$$

Both bond stretching and valence angle bending refer to *intramolecular forces.* This is in contrast to electrostatic charges and van der Waals forces: *intra- and intermolecular bonding.*

2.1.1 Bond Stretching

The Morse potential $E(l)$ and the Harmonic potential a are:

$$E(l) = D_e\{1 - \exp[-a(l - l_0)]\}^2 \quad (2.6)$$

$$a = \omega\sqrt{\frac{\mu}{2D_e}} \quad (2.7)$$

D_e in the above equations represent the depth of the potential energy minimum:

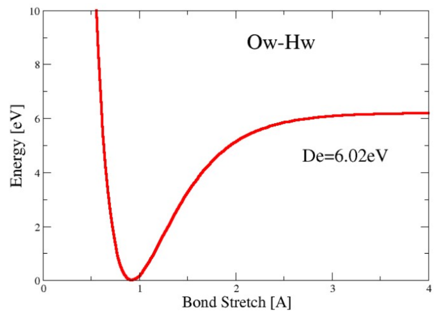

Figure 2.1: A Potential Energy Graph Between an Oxygen Atom and a Hydrogen Atom

ω represents the bond vibration frequency, μ the reduced mass, and l_0 the reference bond length[1]

2.1.1.1 At Room Temperature (i.e., 298 K)

In such a case, the thermal kinetic energy of the system (i.e., molecule) falls within $\frac{1}{2}k_B T$ and $300K$. An approximation for the thermal kinetic energy $E_{thermal}$ is about *0.3 kcal / mol*. We can also say that:

$$E(l) = \frac{1}{2}k(l - l_0)^2 \quad (2.8)$$

$$k = 2D_e a^2 = \mu\omega^2 \quad (2.9)$$

At room temperature, the energy potential of a molecule can be described via the following graph:

This graph is also called a **Hooke's spring**.

[1] This is 0.923 $\mathring{A}$ in this scenario.

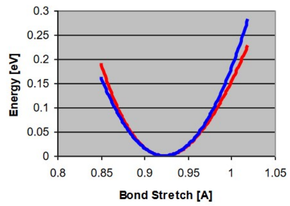

Figure 2.2: A Potential Energy Well at Room Temperature

2.1.2 Angle Bending

Bond	l_0(A)	k^l (kcal /mol A^2)	Angle	θ_0	k^θ (kcal /mol rad^2)
Hw-Ow	0.96	553	Hw-Ow-Hw	104	100
C=O	1.2	570	N-C=O	123	80
C-Ca	1.4	469	C-Ca-Cb	120	63

Figure 2.3: Information on Several Bond Angles

Because of a covalent bond's directionality, its bond angles do not change that much.

$$E(\theta) = \frac{k}{2}(\theta - \theta_0)^2 \tag{2.10}$$

Therefore, Hooke's law is often used to calculate the harmonic potential energy of a certain type of bond angle.

2.1.3 Torison Terms

The **torsional energy** is defined between every quartet of atoms - it depends on the dihedral angle ϕ made by two planes (and also incorporating the first and last three terms in the torsion).

$$E(\psi) = \sum_{n=0}^{N} \frac{V_n}{2}[1 + \cos(n\psi - \psi_0)] \tag{2.11}$$

Torsional motions are typically hundreds of times less stiff than bond stretching motions.

Torsion terms also mimic bonding characteristics and neighboring atoms' and their side groups' steric hindrances about the chain axis.

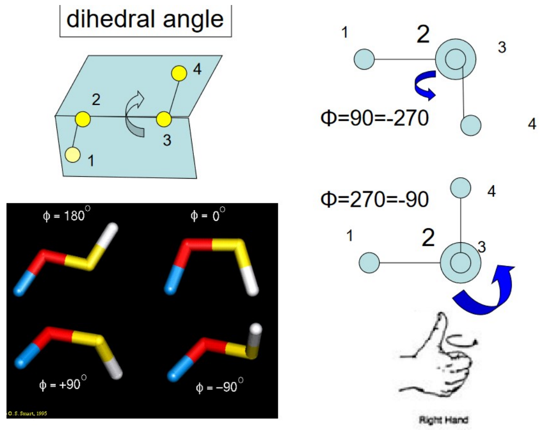

Figure 2.4: Professor Mu's Slides on Torsion Terms and Dihedral Angles

2.1.3.1 Exammple Torsional Terms for Ethane

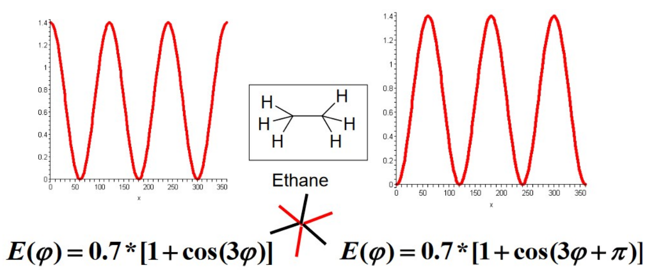

Figure 2.5: Example Torsional Term for an Ethane Molecule

Note that the y-axis of the above graphic is in kcal / mol.

2.1.4 Non-Bonded Interactions

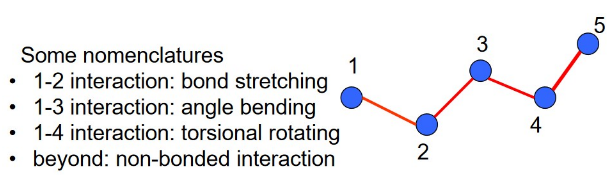

Figure 2.6: Nomenclature for Non-Bonded Interactions

The kind of bonded interactions depend on the bonding relationship between atoms. Such energy functions in this scenario describe the total interactions between atoms and cannot be further decomposed.

2.1.5 Electrostatic Interactions

This is a group in its own right - the other group of non-bonded terms is van der Waals interactions.

$$E_{ele} = \sum_{i>j} \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}} \tag{2.12}$$

In a simple model consisting of two water molecules, their partial charges are $q_O = -0.834$ and $q_H = 0.417$.

2.1.5.1 Calculating Partial Charges

A molecule's electrostatic potential can be measured - it can also be determined from molecular wavefunctions (from quantum mechanics):

$$R = \sum_{i=1}^{N_{points}} (\phi_i^{calc} - \phi_i^0)^2 \tag{2.13}$$

The goal is to find a set partial charge from which the calculated potentials are closest to the reference ones.

2.1.6 van der Waals Interactions

They arise from a balance between attractive and repulsive forces.

The attractive force is due to dispersion forces and is equivalent to $\frac{1}{r^6}$. The repulsive force originates from quantum mechanics and can be understood using Pauli's exclusion principle.

2.1.6.1 Lennard-Jones Potential

The **Lennard-Jones 12-6 function** $4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right]$ has two adjustable parameters:

1. Collision diameter σ
2. Well depth σ

The $\frac{1}{r^{12}}$ is questionable at times, but also allows for rapid computations.

2.1.6.2 Combination Rules

A way to approximate parameters is needed to calculate the van der Waals interactions between different kinds of atoms.

There are two methods covered in BS3008:

1. **Amber and Charmm**

$$\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j} \tag{2.14}$$
$$\sigma_{ij} = \frac{\sigma_i + \sigma_j}{2} \tag{2.15}$$

   The Lorentz-Berthelodt rules are used.

2. **OPLS[2] Force Fields**

$$\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j} \tag{2.16}$$
$$\sigma_{ij} = \sqrt{\sigma_i \sigma_j} \tag{2.17}$$

2.1.6.3 Parameters of van der Waals Forces

The **L**ennard **J**ohnson (i.e., **LJ**) parameters state the following:

1. Heat of vaporization
2. Density (i.e., molecular volume)
3. Partial Molar Volume
4. Crystal simulations

2.2 Common Empirical Force Fields

BS3008 lists several different force fields for one's own reference:

1. **Class I Force Fields**
   - CHARMM
   - CHARMm
   - AMBER

[2] This is short for **O**ptimized **P**otentials for **L**iquid **S**imulations

- OPLS / Schrodinger
- ECEPP (i.e., free energy force field)
- GROMOS

2. **Class II Force Field**

- CFF95
- MM3
- MMFF94
- UFF, DREIDING

2.2.1 On Class II Force Fields

$$
\begin{aligned}
&\sum_{bonds}\left[K_{b,2}(b-b_o)^2+K_{b,3}(b-b_o)^3+K_{b,4}(b-b_o)^4\right] \\
&+\sum_{angles}\left[K_{\theta,2}(\theta-\theta_o)^2+K_{\theta,3}(\theta-\theta_o)^3+K_{\theta,4}(\theta-\theta_o)^4\right] \\
&+\sum_{dihedrals}\left[K_{\phi,1}(1-\cos\phi)+K_{\phi,2}(1-\cos 2\phi)+K_{\phi,3}(1-\cos 3\phi)\right] \\
&+\sum_{impropers}K_{\chi}\chi^2 \\
&+\sum_{bonds}\sum_{bonds}K_{bb}(b-b_o)(b-b_o)+\sum_{angles}\sum_{angles}K_{\theta\theta}(\theta-\theta_o)(\theta-\theta_o) \\
&+\sum_{bonds}\sum_{angles}K_{b\theta}(b-b_o)(\theta-\theta_o) \\
&+\sum_{bonds}\sum_{dihedrals}(b-b_o)\left[K_{\phi,b1}\cos\phi+K_{\phi,b2}\cos 2\phi+K_{\phi,b3}\cos 3\phi\right] \\
&+\sum_{bonds}\sum_{dihedrals}(b-b_o)\left[K_{\phi,b1}\cos\phi+K_{\phi,b2}\cos 2\phi+K_{\phi,b3}\cos 3\phi\right] \\
&+\sum_{angles}\sum_{dihedrals}(\theta-\theta_o)\left[K_{\phi,\theta1}\cos\phi+K_{\phi,\theta2}\cos 2\phi+K_{\phi,\theta3}\cos 3\phi\right] \\
&+\sum_{angles}\sum_{angles}\sum_{dihedrals}(\theta-\theta_o)(\theta-\theta_o)\cos\phi
\end{aligned}
$$

The first three terms represent bond (b) stretch, angle bending (θ) and torsion (ϕ), while the fourth gives the energy contribution when atoms are distorted (χ) out of the plane formed by the atoms to which they are bonded. The next 6 terms represent the coupling between the intramolecular coordinates. For example $K_{bb'}$, $K_{\phi b}$, $K_{\phi b'}$ parameterise the coupling between adjacent bonds, and that between the torsion and central and peripheral bonds. The triple sum gives the energetic contribution of the coupling of torsion and the two bond angles around the interior bond. The final two terms are the intermolecular electrostatic and van der Waals energies. It is found that all the terms described above are necessary and may contribute equally to the energy and forces.

25

Figure 2.7: Equation and Summary of Class II Force Fields

The above image was taken off professor Mu's teaching slides.

3 Summary

In summary, this book has no content whatsoever.

```r
1 + 1
```

```
[1] 2
```