TKJ4205/KJ8902 Molecular Modelling
Exam 12.12.2013, 9.00-13.00

Norges Teknisk-Naturvitenskapelige Universitet (NTNU)
Code A: All printed and hand-written texts are allowed. All calculators are allowed.
Contact: Assoc. Prof. Titus van Erp, Dep. of Chemistry, cell phone: 98048297
NB: The tasks are not weighted equally. The weight for each task is given in parenthesis after the task number. The weights sum up to 100.

Exercise 1 (15)

What is the variational theorem and why is it useful in quantum chemistry? Which of the following quantum chemical methods are variational: Hartree-Fock, the configuration interaction (CI) method, Møller-Plesset second-order perturbation theory (MP2), and Kohn-Sham density functional theory?
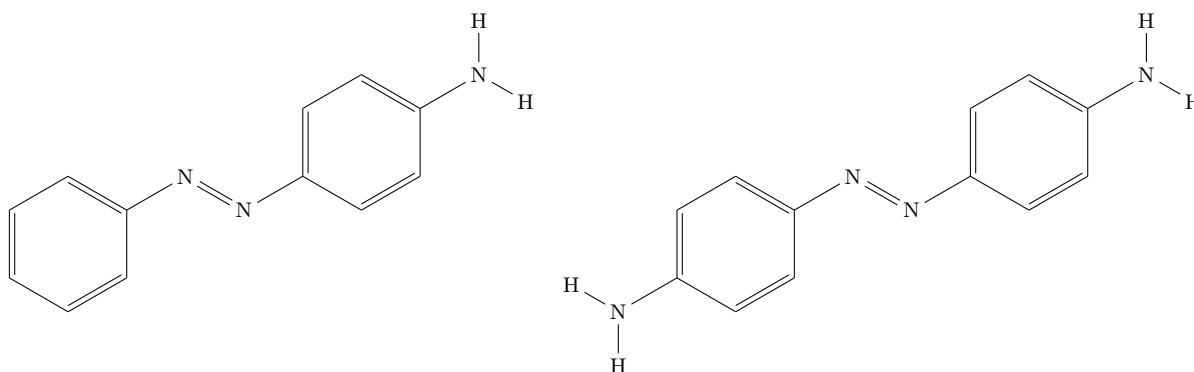
Exercise 2 (10)

Why is hydrogen bonding difficult to model by a standard force field like

$$V = \sum_{I=1}^{N} \sum_{J=I+1}^{N} \frac{q_I q_J}{4\pi\varepsilon_0 R_{IJ}} + 4\epsilon_{IJ} \left(\left(\frac{\sigma_{IJ}}{R_{IJ}} \right)^{12} - \left(\frac{\sigma_{IJ}}{R_{IJ}} \right)^{6} \right)$$

Suggest two approaches to improve the description of hydrogen bonding in force fields.

Exercise 3 (10)

The two azobenzenes in the figure below

have slightly different colours. In an experiment, your fellow organic chemist does not know which of the two azobenzenes he/she has synthesized, but can you suggest a computational strategy to distinguish between the two molecules based on colour? In some cases, changing solvent may shift the colour of an azobenzene. Suggest different methods to include solvent effects in the calculations.

Exercise 4 (15)

What is the major difference between solid state calculations and quantum chemical calculations on molecular systems? Suggest a method to study catalysis on a surface. Finally, when is relativistic effects important in quantum chemistry, and suggest one approach where relativistic effects may be included approximatively.

Exercise 5 (50: 6, 8, 8, 10, 10, 8)

A complex molecule that is emerged in a solvent has two stable configurations. This molecule consists of two molecular groups that are rigid, but the distance between these two molecular groups can vary. To a very good approximation this complex molecule can be described by a simple model of a diatomic molecule $A - B$ where the "atoms" A and "B" refer to the two rigid molecular groups. Based on DFT calculations, the potential energy as function of separation distance d of the center of masses was obtained. It is shown in Fig. 1. It clearly shows that there are two stable configurations, a contracted state and an extended state.
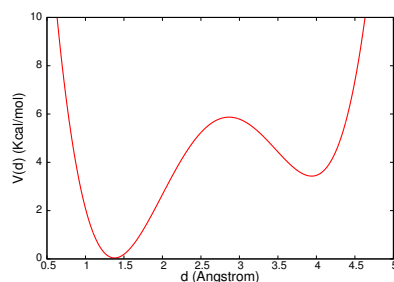

Figure 1: Intra-molecular potential energy as function of separation distance.

i) Since the A and B groups are rigid the configurational space of a single molecule is six-dimensional. Each configuration can be given by the vector $(x_A, y_A, z_A, x_B, y_B, z_B)$. To get a better understanding of the system in a lower dimensional space, you want to study the free energy as function of separation distance $A(d)$ of this molecule. Can you give the relation between the probability distribution function $P(d)$ and the free energy $A(d)$?

ii) Suppose that the temperature is sufficiently high so that a NVT molecular dynamics can easily explore all configurations (both extended and contracted state will be sampled). In that case, how can you determine numerically $P(d)$ from a very long MD simulation? What possibilities do you have if the temperature is too low to efficiently sample configuration space (hence, the system will be trapped for very long times in either the extended or contracted state)

iii) Show that for an isolated molecule in the gas phase that is free to rotate and has no interaction with other molecules, the free energy $A(d)$ must be given by

$$A(d) = V(d) - k_B T \ln d^2 + \text{constant} \tag{1}$$

where k_B is the Boltzmann constant and T the temperature in Kelvin. To explain this expression think about the relation of free energy and relative probability and about the orientations that a diatomic molecule can take in 3D. What units must the constant have in Eq. 1?

iv) You now want to study the dynamics of this transition in the solvent at a high temperature. At this temperature the transitions occur more frequently and can be studied by long brute force MD trajectories. You want to determine the following correlation function

$$C(t) = \frac{\langle \theta(2.9 - d(0))\theta(d(t) - 2.9)\rangle}{\langle \theta(2.9 - d(0))\rangle} \tag{6}$$

Here, θ is the Heaviside step-function: $\theta(x) = 1$ if $x > 0$, $\theta(x) = 0$ if $x \leq 0$. The brackets $\langle \ldots \rangle$ denote an NVT ensemble average over the initial conditions (at $t = 0$). The 2.9 Å is the top of the potential curve in Fig. 1 and you define any state with a distance d smaller than this value to be compact, and with a distance d larger than this value to be extended. Show that Eq. 6 is equivalent to

$$C(t) = \langle \theta(d(t) - 2.9)\rangle_{d(0) < 2.9 \text{ Å}} \tag{7}$$

where the subscript means that you only consider configurations with an initial $A - B$ distance smaller than 2.9 Å. Can you argue, based on Eq. 7, that $C(t)$ is in fact nothing else as the conditional probability that the molecule is in the extended state at time t given that it was in the contracted state at time 0.

v) You want to calculate $C(t)$ for $t = 1$ ps. In other words, you are interested in the probability that the molecule is in the extended state if it was in the contracted state 1 ps earlier. To do that you do the following simulation.

(a) You run a MC simulation of 1000 steps to generate initial configurations with $d < 2.9$ Å. Each of these configurations is saved to the hard disk.

(b) For each configuration point, we generate random velocities according to the correct temperature distribution

(c) Starting from each configuration point with corresponding velocities, we run a 1000 step MD trajectory with a time step of 1 fs.

(d) For each trajectory we sample the value 1 if the end point is in the extended state and 0 otherwise.

So in the end we obtain 1000 data points, each data point is either 0 or 1, and the average gives this conditional probability or $C(1\text{ps})$.

The 1000 measurements are grouped in 5 blocks of 200. The average of each block is calculated. These are: 0.17, 0.07, 0.11, 0.17, and 0.08. Calculate $C(1\text{ps})$ together with its actual statistical statistical error (absolute and relative error) based on the block analysis. We may assume the averages of each block is uncorrelated.

vi) What would have been the expected error in the case that each of the 1000 trajectories was uncorrelated? Could you think of a simple way to reduce the correlations between the trajectories and by that get a more efficient algorithm? Estimate the increase in efficiency.