CS 4414. Homework Assignment 1. Group work, strictly follow the guidelines in the class syllabus. Submit only one report per group, indicating who did what. PDF format only.

1 Motivation

Most biological processes occur in an aqueous environment; thus an accurate qualitative and quantitative description of aqueous solvation of molecules is very important in understanding such phenomena. The feasibility of a process (a chemical reaction, binding etc.) is usually determined via thermodynamic observables such as free energy, entropy etc. In 1920 Max Born proposed a simple equation to calculate the free energy of solvating (transferring an ion from gas phase to aqueous phase) a simple spherical ion of charge q and ionic radius R_i:

$$\Delta G = -166\left(1-\frac{1}{\epsilon}\right)\frac{q^2}{R_i} \tag{1}$$

where, ϵ is the dielectric constant of water ~ 80 at room temperature. According to the formula, the solvation energy has quadratic dependence on ion charge *i.e.* $\Delta G \propto q^2$, implying that ions of opposite sign but same ionic radii would have identical solvation energies. That is $\Delta\Delta G = \Delta G(q,R) - \Delta G(-q,R) = 0$. The experimental reality, however, is very different (see Fig 1).

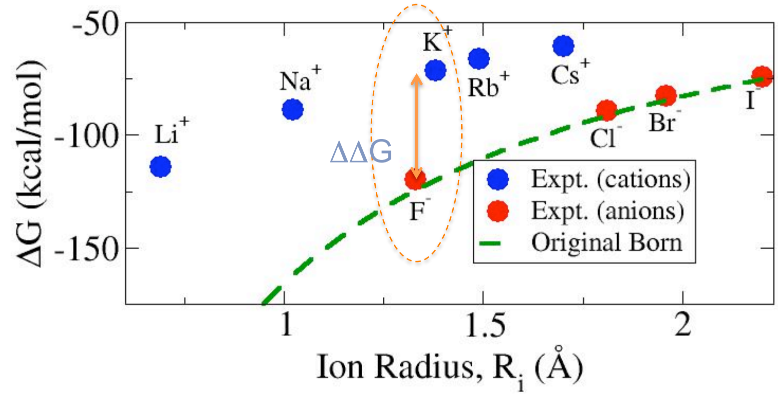

Figure 1: The solvation energy of simple ions. $\Delta\Delta G$ in the figure corresponds to the difference in the solvation energies of the pair K^+ and F^-. Notice how these two ions have similar ionic radii but very different values of ΔG

2 Model

In order to resolve the discrepancy with experiment, a group of physicists proposed a simple correction to the Born equation:

$$\Delta G(q,R_i) = -166\left(1-\frac{1}{\epsilon}\right)\frac{q^2}{R_i+R_s}\frac{\ln\left(\frac{e^{-\beta E^{2P}(+)}+e^{-\beta E^{2P}(-)}}{2}\right)}{\ln\left(\frac{e^{-\beta E^{SPD}(+)}+e^{-\beta E^{SPD}(-)}}{2}\right)} \tag{2}$$

where $E^{2P}(+) = \frac{332qq_O}{\epsilon_W}\left(\frac{1}{R_i+R_w} - \frac{1}{R_i+R_w+R^z_{OH}}\right)$, $E^{2P}(-) = \frac{332qq_O}{\epsilon_W}\left(\frac{1}{R_i+R_w} - \frac{1}{R_i+R_w-R^z_{OH}}\right)$, $E^{SPD}(+) = \frac{332qq_O R^z_{OH}}{\epsilon_W(R_i+R_w)^2}$ and $E^{SPD}(-) = -\frac{332qq_O R^z_{OH}}{\epsilon_W(R_i+R_w)^2}$.

Here SPD and 2P and two different types of water model. For this assignment, you do not need to worry too much about what each parameter means, but here is a description anyway, to give you some idea. The parameter $R^z_{OH} = 0.58588\mathring{A}$ is the projected length of the hydrogen arm of a three point water model along its angle bisector, $R_w = 1.4\mathring{A}$ is the radius of water molecule and the charge of oxygen, $q_0 - = -0.834$ (atomic units). $\beta = 1/0.6(kcal/mol)^{-1}$ is defined as the Boltzmann factor. $\epsilon_W = 10$ is the effective permittivity in the vicinity of the ion-water interface. $R_s = 0.52\AA$ corresponds to a shift in the dielectric boundary a.k.a the boundary between ion and water.

3 The Assignment

Your task is to streamline the transition from math, Eq. 2, to numerical implementation of it. This is a very realistic scientific computing problem, not just a text book assignment, so, not every little detail is specified. Use your common sense to fill in the gaps, if any. Questions like "how do I make Mathematica do..." you need to tackle on your own.

1. (**10 points**) Use Equation 2 and a reasonable programming language (C, C++ preferred) to generate a plot of solvation energy, ΔG vs. charge for both positive and negative ions such that $|q| \in [1 : 500]$ for $R_i = 1$ Å . Does your code work in the whole range? Why?
2. (**10 points**) Now use Mathematica and repeat 1. What do you infer? Provide a plot of the function in Equation 2 generated with Mathematica for $R_i = 1$ Å . Provide another plot(use any software) to show the discrepancy (relative error as a function of q) between the results using Mathematica and those obtained with a C/C++ code in problem 1.
Use root mean square percentage error, RMSPE to quantify the over-all discrepancy. Use Mathematica results as reference.
(RMSPE=$\sqrt{\sum_i \left(\frac{\hat{y}_i - y_i}{y_i}\right)^2 /n}$, the sum extends over all of your data points)
3. (**10 points**) How would you modify your C++ code from problem 1 to get rid of the error and obtain a well-behaved function $\Delta G(q, R)$ for $|q| \in [1 : 500]$, $1.0 \leq R_i \leq 3.0$ Å ? Your solution should give no more that 10% RMSPE relative to the Mathematica results. Show the key part of your code and explain the main idea.

4. (**10 points**) Download the table ("Cl.dat") with the experimental data of $\Delta G(q)$ of Cl sized ions ($R_i = 1.81\AA$) with different charges, $q \in [-1 : 1]$. Use the C++ code you wrote in 1 to compute $\Delta G(q)$ at $R_i = 1.81$ and compare with the experiment. Pay special attention to the limit of small q $i.e. q \to 0$, including $q = 0$. Plot and discuss your findings.

5. (**10 points**) What computation-ready functional form for $\Delta G(q, R_i)$ in Equation 2 do you propose so that the entire range of $|q| \in [0 : 500]$, $1.0 \leq R_i \leq 3.0$ Å is covered within, say, 10 % error relative to the exact result (Eq. 2) at each point? The new function can be piece-wise continuous, as in $\Delta G(q, R_i) = ...$ if $|q| < \alpha$, $\Delta G(q, R_i) = ...$ if $\alpha \leq |q| \leq \beta$, $\Delta G(q, R_i) = ...$ if $|q| > \beta$. You need to fill in the blanks and determine α, β

Extra credit(30 points): Provide a single, simple differentiable analytic function that decently approximates Equation 2 in the entire range of q and R_i specified in 5, giving correct asymptotic behavior in small as well as large q regimes, and covering the intermidiate values at least qualitatively correctly. Motivation: a simple function like this can make a large difference where computational speed is key, as in molecular dynamics simulations. It can also simplify derivations and improve stability of the algorithms, especially if the latter involve higher derivatives. Our research group may actually use a good solution in practice.

4 Instructions and further details

a) For your convenience I have put all the constant parameters for this model in a file "constants.dat". "Cl.dat" contains the solvation energy of Cl sized ions, the first column contains the charge, q of the ion and the second column has the corresponding solvation energies, ΔG.

b) All units of energy e.g. $E^{2P}(\pm)$, $E^{SPD}(\pm)$, ΔG, $\Delta\Delta G$ are expressed in kcal/mol, charge e.g. q, q_O are in atomic units (e), all lengths e.g. R_i, R_w, R_s, R^z_{OH} are expressed in $\AA$. You don't have to worry about any conversions, but make sure to label the axes of all your plots!