Estimating Polar Interaction Energies in the Presence of Water

David Grierson

EPCC–SS94–13

Estimating Polar Interaction Energies in the Presence of Water

David Grierson

EPCC–SS94–13

September 1994

Abstract

The dielectric of aqueous solutions has often been ignored or oversimplified when modelling chemical systems with charges present. This, however, can cause the model to be incorrect or to produce non-realistic results.

The aims of this project were therefore to model a packing of water molecules, thus allowing a method of calculating the internal energy of the molecular system by means of a task farm performing a Monte Carlo simulation upon the energies of the molecules.

The overall aim was to find a transferrable potential energy function for water with charges to allow simpler molecular modelling in solution.

1 Introduction

Water is without doubt the most abundant, the most accessible, and the most studied of all chemical compounds. Water plays a prominent role in most creation myths and has a symbolic purifying or regenerating significance in many great religions even to the present day. Thales of Miletus, who is generally regarded as the initiator of the Greek classical tradition of philosophy, *ca.* 585BC, considered water to be the sole fundamental principle in nature. His celebrated dictum maintains: "It is water that, in taking different forms, constitutes the earth, atmosphere, sky, mountains, gods and men, beasts and birds, grass and trees, and animals down to worms flies and ants. All of these are but different forms of water!" Though this may sound quaint or even perverse to modern ears, we should reflect that some marine invertibrates are indeed 96–97% water, and the human embryo during its first month is 93% water by weight [Greenwood and Earnshaw, 1993].

It is because of these large quantities of water present in the animal *and* human bodies that drug, chemical, and environmental groups are especially interested in reactions which take place in the presence of water, be it natural protein synthesis, the diffusion of an ion into an active site [Getzoff *et al.*, 1992], or fertiliser effects upon brain tissue.

In the development of new drugs and chemicals, chemical companies are becoming more and more reliant upon the use of computer systems to design and model molecules and molecular systems. Computers are allowing the shape and structure of molecules to be visualised and simulated in a virtual environment defined by the programmer making it easier to understand why a reaction is taking place in a particular manner.

Since the majority of the reactions which which are of particular interest take place in aqueous solution it would be desirable to describe the solution accurately and effectively, yet at the same time quickly enough to allow molecular modelling to take place in the solution defined. At the moment these molecular models are not completely accurate as the water in the solution is ignored and the interactions of the reactants are somewhat over simplified.

1.1 Project Aims

During the modelling of systems in aqueous solution, such as inter-molecular interactions between enzymes and substrates the dielectric of the water surrounding the two molecules is often ignored, or given some function to try to imitate its behaviour with distance. Quite often this is simply expressed as the inverse of the radius $\epsilon = \frac{1}{r}$. Clearly this is not going to produce realistic results for the interactions if the reaction is at all dependant upon the dielectric of the surrounding system.

This project aimed to allow a simple method of molecular modelling for solution systems containing charged particles by modelling the electrostatic properties of a solution effectively and efficiently using a Monte Carlo simulation upon the individual atoms of the water molecules surrounding the charges, in order to see if it is possible to find a transferrable potential energy function for the dielectric of water.

If a potential energy function is found it would allow systems such as the one above [Getzoff *et al.*, 1992] and possibly more complicated problems, e.g. large molecular diffusions, to be described in a more complete manner.

This report provides some background to the nature of the problem, it then discusses how the calculation of the energies was done and the various problems which were encountered when trying to simulate the packing of a water cluster and the calculation of the energies themselves.

2 The Dielectric Constant

The potential, or voltage of an ion of charge $Z_i e$, at a distance of r, in a vacuum is given by the equation:

	ϵ_r
Water	78.54
Ammonia	16.9
	22.4 (at $-33°C$)
Ethanol	24.30
Benzene	2.274

Table 1: Dielectrics of various solvents at 298 K

$$\phi_i = \frac{(z_i e/4\pi\epsilon_0)}{(1/r)} \tag{1}$$

where ϵ_0 is the permattivity of a vacuum. In a solution there are modifications which need to be made; the solvent decreases the strength of the potential, if the electric permattivity is ϵ, then the potential at distance r is:

$$\phi_i = \frac{(z_i e/4\pi\epsilon)}{(1/r)} \tag{2}$$

The permattivity is usually expressed in terms of the relative permattivity or *dielectric constant* ϵ_r through $\epsilon = \epsilon_r\epsilon_0$. Some values are shown in table 1. [Atkins, 1987] The dielectrics listed in table 1 are due solely to the solvent present and have nothing to do with ionic charges in the solvent cluster which is the area of interest in this project.

2.1 Calculation of the Potential Energy of a Molecular System

For this project we are interested in the electrostatic energy of the system which can be expressed as the sum of the interactions between every pair of atoms in every pair of molecule in the system. To be able to calculate the energy it is necessary to build a cluster of molecules.

2.1.1 Packing of Molecules

The water was packed round the solute particle using a triangular method. If there are three molecules arranged together within a certain distance then it is possible to pack a fourth molecule at a calculated distance from them, forming a tetrahedron.

This was done using vector arithmetic; to find the position of the sphere packed against molecules i, j and k the method was as follows:

1. Find the vector which is normal to the base plane (i.e. at right angles to the base plane),

$$u_{ijk} = \frac{u_{ij} \times u_{ik}}{\sin\omega_{ijk}} \tag{3}$$

2. Find the vector in the direction of the basepoint,

$$u_{tb} = u_{ijk} \times u_{ij} \tag{4}$$

3. Find the basepoint,

$$b_{ijk} = t_{ij} + u_{tb}.\frac{[u_{ik}.(t_{ik} - t_{ij})]}{\sin\omega_{ijk}} \tag{5}$$

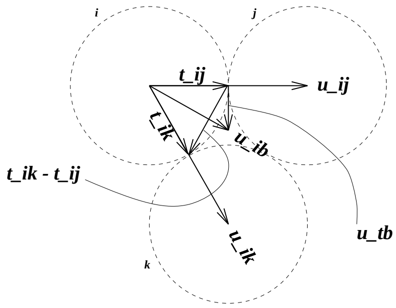


Figure 1: Three water molecules used to calculate a fourth.

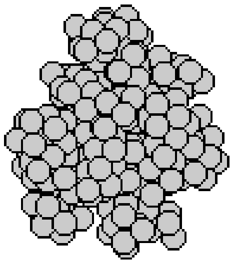


Figure 2: A 250 molecule cluster.

4. Place a probe above and below b_{ijk} at a distance of:

$$d = \sqrt{2r^2 - |u_{ib}|^2} \tag{6}$$

Therefore starting with three touching molecules (figure 1) we can place a third solvent molecule upon these in the position calculated using the above method. This then gives us a tetrahedron of four molecules all touching; each face of this can then be used as three touching molecules to calculate the next position. From this a recursive method can be used to pack the solvent around the solute.

When the fourth molecule is calculated from the three specified a tetrahedron of molecules formed, thus the cluster is not completely spherical, figure 2.1.1 shows a packing of 250 molecules which clearly is

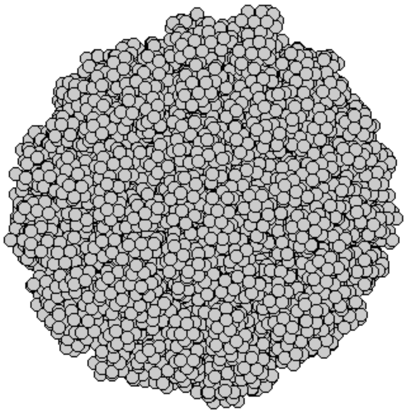


Figure 3: A 10000 molecule cluster.

not spherical but more terahedral; however for larger numbers of molecules the arms of the tetrahedrons twist back in upon the origin of themselves to form a spherical cluster, for example figure 2.1.1 which is 10000 molecules packed together.

2.1.2 Time Factors

The method described above, in section 2.1, produces a packing of 1000 water molecules around a solute particle in approximately 5 minutes and a packing of 10000 molecules in approximately 1 hour. The method of packing is approximately $O(n^2/2)$ although this has yet to be proven.

To perform an accurate calculation there must be a sufficient packing of water between the two charges to cause any masking effects necessary. Therefore, ideally the cluster size used would be about 1000 molecules or larger as this would surround both point charges completely, but both the packing calculation time and the actual time which the simulation would take would be so large that it would make the acquisition of results both physically and computationally time consuming.

2.2 Monte Carlo Simulation of Water Electrostatics

A Monte Carlo simulation is carried out by stepping through the elements in a physical system, applying a random perturbation to them, then accpeting or rejecting the perturbed state according to a random stochastic update rule. In physics and chemistry, the rule applied is an exponential function of the relative energies of the original and perturbed states, equation 7. This is a poweful and general technique, since the update rule can be proven to converge on a thermodynamically meaningful result for many systems. If the system has few enough free parameters it is often the best way for asessing the internal energy of a system, but for more complex/realistic systems the simulation time becomes unacceptably long.

$$P_{\text{accept}} = \begin{array}{ll} 1 & \text{if } \delta E \leq 0 \\ e^{-\delta E/kT} & \text{if } \delta E > 0 \end{array} \tag{7}$$

We are interested in the electrostatic energy of the system which can be expressed as the sum of the interactions between every pair of atoms in every pair of molecules in the cluster;

$$E = \sum_{i,j} \frac{q_i q_j}{\epsilon_0 r_{ij}} \tag{8}$$

where i, j represents every pair of atoms in the system which are not part of the same molecule.

To carry out a simulation of the electrostatic properties of water, the following updating method was used:

- For each set of water clusters, m_N, generated:
- Measure the total energy, ΣE, of the system.
- For each of T timesteps:
  - for each molecule, m, in turn,
    1. The energy of the old conformation of atoms was calculated (this would have been changed by any updates already carried out).
    2. A new conformation of atoms in the molecule, m', was generated.
    3. The energy of the new conformation was calculated.
    4. The update rule is applied to choose between the new and old conformations.

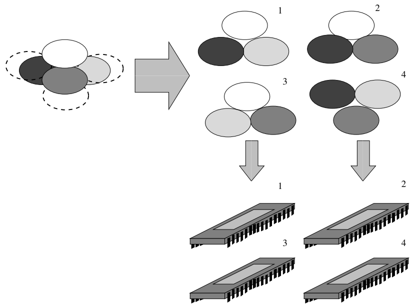


Figure 4: Distribution of triplets of molecules onto separate processors.

5. The next molecule for updating is selected.

- calculate the total energy for the system at the end of each timestep.

By observing the evolution of the energy through time, it is possible to determine whether a stable value for the energy is converged upon: the physical significance of that energy can be assessed by observing the temperature dependence of the final energy calculated.

3 Parallelism

3.1 The Need for Parallelism

For an accurate result to be obtained from the Monte Carlo simulation it is necessary to perform the simulation over a large number of timesteps and a large number of times. It would therefore be desirable to perform multiple tasks at the same time independantly of one another to allow the simulation to be done several times faster.

There would be two ways of parallelising the code:

1. Parallelise the packing algorithm.
2. Parallelise the simulation.

3.2 Parallelism of the Packing Algorithm

When a molecule is packed against a group of three molecules, we are getting a tetrahedron of four molecules we are then getting four triangular faces produced, each of which could be packed against. This is an obvious area which could be parallelised by instead of serially doing each calculation, simultaneously calculating the new positions corresponding to each, figure 4.

The main disadvantage of using a method like the one described above to pack a sphere of molecules is the communications cost.

Every time a new molecule position is calculated using three others, the data about the three component molecules will have to get sent to a processor using calculation time with communications. After the position of the new molecule has been calculated it has to get tested to check that it doesn't overlap or clash with any existing molecule positions. The communications involved in only a few thousand molecules would completely hinder the calculation to the extent of slowing it down.

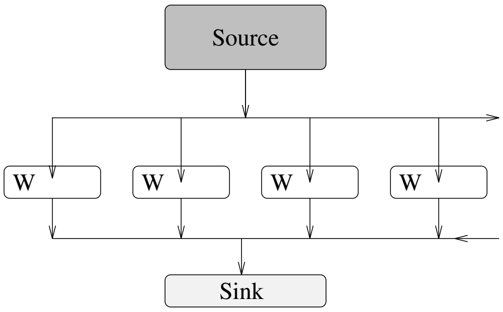


Figure 5: A Task Farm.

3.3 Task Farm

The Monte Carlo simulation requires a large number of different sets of data to perform simulations on to get a value which is more meaningful than a single simulation, due to the random factors involved. However each simulation is not dependant upon any of the other simulations running; this allows a parallel method of performing the simulations using a task farm.

A task farm, figure 5, is made up of three different constituent programs:

1. The **Source** sends each worker a task which is made up of:
   - a seed for the random number generator,
   - a size to make the cluster,
   - a distance to place the two charged particles apart,
   - a number of iterations to perform the Monte Carlo simulation over,
   - and a temperature to use in the update rule,
2. The **Workers** which, after receiving their task from the Source, each:
   - builds a cluster to the specified size using the method described in section 2.1,
   - performs a Monte Carlo simulation over the specified number of iterations, using the temperature specified in the update rule and the seed in the random number generator, on the cluster of water molecules which it packed, then,
   - sends the total energies calculated in the simulation along with the information which the Source sent to the Worker to the sink.
3. The **Sink**:
   - collects the results from each of the Workers and writes them to a datafile, for viewing using a visualiser, and also writes parts of the data structure to a text file.

3.3.1 Task Farm Implimentation

The task farm was implimented using the Parallel Utilities Library task farm functions (PUL-TF). These are routines specifically written to allow task farms to be written with ease in either C or fortran without any need to worry about how the message passing between processors or between workstations is being done.

Quite simply any message being sent by the Source will go to either a single Worker or all of the Workers depending upon the function used and the flags specified and any message sent by a Worker will go to the sink.

As this project is a task farm performing a Monte Carlo simulation so each individual task sent out will take a considerable amount of time and thus there is no real dependance upon the communications capabilities of the network used, whether it be an ethernet or an internal grid of transputers.

3.4 Problems Encountered

The only limit upon the type of processor used is the amount of memory located upon each processor. The simulation itself is not heavily memory intensive but to be able to perform the simulation it is necessary to have a cluster built. However, the packing algorithm is implimented in a recurrsive function; most compilers are able to optimise *tail recurrsive* functions into a loop instead, however with this algorithm there are four recurrsive calls one after another — causing the compiler to be unable to optimise the code in this manner.

For this reason the program was unable to be run upon the *T800* transputers available in the Summer Scholarship Programme machine, and meaning that the only real option was using clusters of sun workstations which have virtual memory.

4 Results — theoretical and actual

The simulation itself is very heavily CPU consumptive (90% CPU usage) and also needs to be performed over a large number of iterations for a large number of times, so workstations are not always the best option if other users are trying to work on a machine which is being used as the user notices a definate slow-down of their machine. For this reason the results produced were done to show that the cluster model was accurate enough and the simulation producing reasonable results for a small number of iterations, so that the actual simulation could be performed at some time in the future.

4.1 Simplification

Each iteration of the simulation itself is very time consuming, between 15–16 seconds for one molecule calculating it's energy contribution and then flipping and calculating the new energy contribution. This may not seem like very long but for 250 molecules it means that one iteration for all of the molecules is likely to take about an hour. Ideally you would perform the simulation over a large number of iterations, 100–200 to allow time for the conformations to arrange into a reasonable energy state.

4.1.1 Speed Up

The distance between two molecules is calculated using simple pythagorean mathematics, equation 9:

$$distance = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2} \tag{9}$$

Since the energy contribution of a molecule is equal to the sum of the interactions of every pair of atoms in every molecule, equation 8, thus for every molecule you are performing a square root for every other pair of molecules, a total of n^3 calculations, causing cosiderable slowdown.

This slowdown was too excessive to make the program at all useful, therefore, it was decided to alter the code such that instead of calculating the distance between the two atoms, the distance squared would be used, effectively doubling the speed of the code.

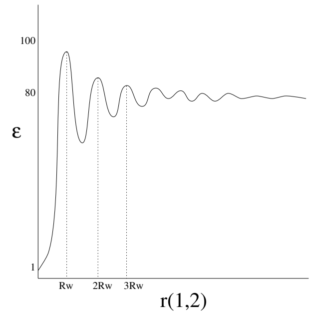


Figure 6: Expected graph of dielectric against distance between two charged particles

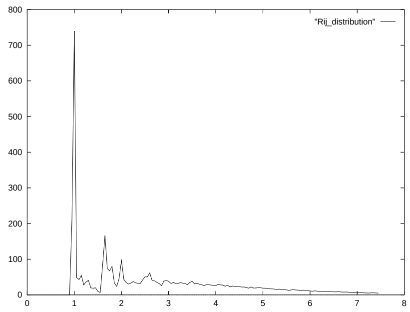


Figure 7: Distribution of molecules with varying distance R_{ij}

4.2 Expected results

The expected result for a graph of the variation of the dielectric with distance between point charges is shown in figure 4.1.1. There are peaks at R_w, $2R_w$, $3R_w$ etc. This is due to the solvation shells of water molecules packed around the two charged particles. If the charged particles are close together, i.e. $< R_w$, then there is insufficient room between them to be able to pack any water molecules, thus the dielectric increases with distance. However, once there is sufficient room, $> R_w$ the water molecules form solvation shells between the particles causing masking effects upon the dielectric. This is why the dielectric oscillates above and below the dielectric of water ($78.54J^{-1}C^2m^{-1}$) which it will tend to as the distance between the particles tends towards infinity and the become point charges in solution.

4.3 Actual results

Firstly while the packing algorithm was still being tested it was decided to plot a graph of distribution of molecules against distance between charged particles, figure 4.2. This also shows the peaks at R_w,

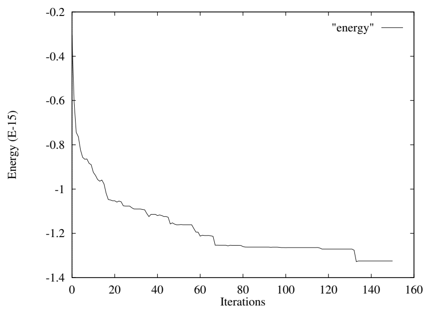


Figure 8: Energy minimisation of water for 250 molecules over 150 iterations

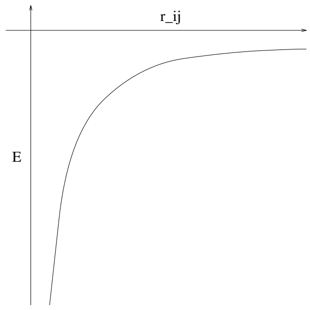


Figure 9: Total energy of two point charges in a vacuum

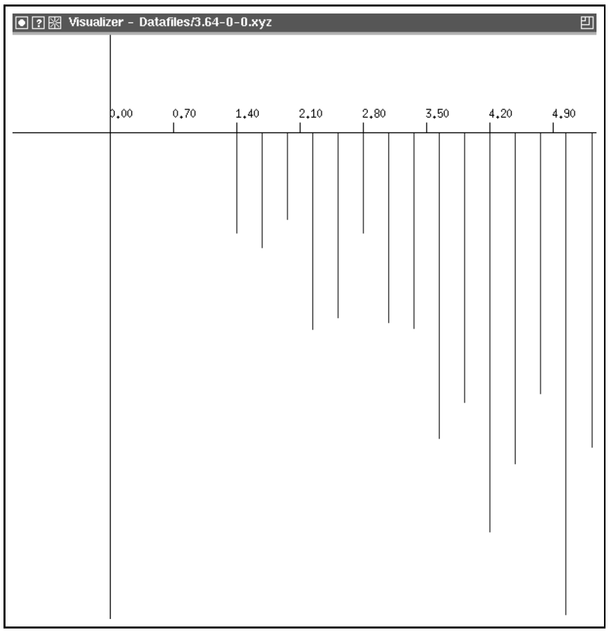


Figure 10: Actual graph obtained for two point charges surrounded by water at varying distances.

$2R_w$, and also shows that it is tending towards a certain value. The reason that the graph has an overall decreasing nature is because the distribution was not normalised for the volume of the sphere — thus it would tend towards a certain value normally.

Then to check that the energy calculation was being done suitably well an energy minimisation of the atoms in a cluster of 250 molecules with no charged particles performed over 150 iterations. In this the probability of flipping into a higher energy conformation of atoms was always zero, i.e. whichever of the old and new molecules had a lower energy contribution was chosen to be the one to remain. Thus after a number of timesteps the dipoles in the water molecules would have arranged themselves into the lowest energy possible for that packing. A graph of how the total energy of the cluster varied with timesteps is shown in figure 8.

Due to the fact that the workstations used were departmental machines which were in use every day, it was difficult to find workstations not in use to allow the simulation to run - however it was managed to run a simulation which varied the distance between the two particles over the range $2R_w$–$8R_w$ and calculated 10 iterations at T = 0. Thus the graph produced was an energy minimisation for two charges at varying distance in the presence of water, figure 10.

Normally for two particles in a vacuum at varying distance the graph for the energy calculation looks like figure 9, with the lowest energy state being when the charges are touching one another. However as can be seed from the results of the simulation, figure 10 there are peaks and troughs up to a minimum energy

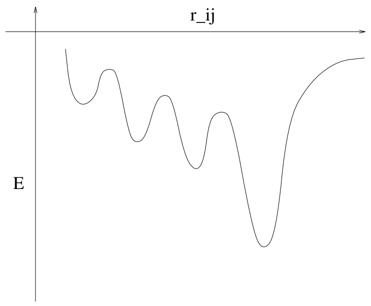


Figure 11: Expected graph for two point charges in water of energy vs. distance

value at about $6R_w$ and then an increase in the energy in the same way as the charges in a vacuum. A diagram of the expected graph for two charges in water is shown in figure 11.

This is due to the water forming solvation shells between the two particles. Thus they have a lower energy overall when there are two solvation shells between them.

5 Conclusions

This project has managed to form of clusters of water molecules packed around charged particles to an unlimited size, and also managed to simulate the energies of the atoms in the molecules in the presence of these charges. The project also used PUL-TF task farm routines to perform a Monte Carlo simulation upon multiple conformations of water at mulitple temperatures.

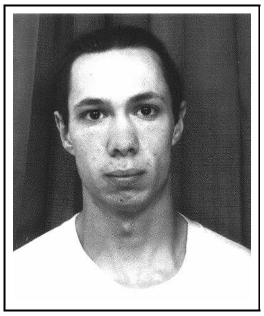
References

[Atkins, 1987] P. W. Atkins. *Physical Chemistry*. Oxford University Press, 3rd edition, 1987.

[Getzoff *et al.*, 1992] E. D. Getzoff, D. E. Cabelli, C.L. Fisher, H. E. Parge, M. S. Viezzoli, L. Banci, and R. A. Hallewell. Faster superoxide dismutase mutants designed by enhancing electrostatic guidance. *Nature*, vol. 358, 23 July 1992.

[Greenwood and Earnshaw, 1993] N. N. Greenwood and A. Earnshaw. *Chemistry of the Elements*. Pergamon Press, 1993.

[Lamport, 1986] Leslie Lamport. *LaTeX: A Document Preparation System (User Guide & Reference Manual)*. Addison-Wesley, 1986.



David Grierson is a BSc student of Chemistry at the University of Aberdeen. He became interested in computing science while using unix during his first year computing course. After graduation in 1995 he hopes to pursue a career in Computational Chemistry.
This project was supervised by Peter MacCallum (EPCC) and Felicity George (EPCC)