

Research Galore!!!

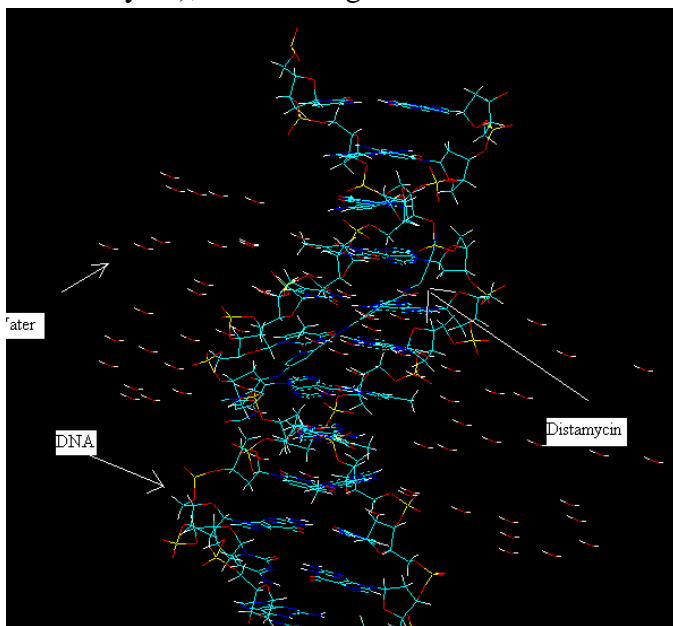
“At least two publications a year” is the old “convention of wisdom” that you hear, and it is typically followed by warnings that “of course, that is the bare minimum!” If you want to get tenure, you must publish papers (the “publish or perish” philosophy). It’s an important thing; publications are how universities gain their notoriety, and it is a source of income since publications can be used to attract grant money. Of course, MOST faculty LOVE research, and are willing to put up with teaching requirements for the academic freedom to pursue the research of their choice at most universities.

I, myself, am a freak. Yes, I do really enjoy research, but my FIRST love is teaching. This is why DSU and I are such a good match; while DSU encourages research, the main focus here is teaching, just like my philosophy. And, fortunately, you don’t have to do much research to get tenure here; I will be up for tenure soon, and I expect I will get it, even though I have (to date) had no publications while at DSU (save a few books on GED and AP test preparation). However, I have had several “pet” research projects on the back burner, and while I try to encourage my students to find projects of interest to them personally, I have at the same time kept a few things alive that are of interest to me.

This is not an article about my teaching. Instead, it is a report on very exciting recent advancements in my research, and just might include one or two opportunities for you as well.

DNA Sequence Specific Binders and the Bifurcated Hydrogen Bond

Let’s begin with a little history. Many years ago (I am thinking around 1989 give or take a year), I was sitting in an auditorium at Boston College listening to a seminar



given by Peter Dervan, from California. A very talented biochemist, he had a problem he was trying to resolve. He had a small organic molecule, Distamycin (an anti-tumor agent), that bound to DNA in the minor groove, but only to the specific sequence AAATT. The question was obvious; how could this small molecule recognize the base pairs in the DNA? The problem was further complicated, because Distamycin could only form four hydrogen bonds, so the particularly perplexing question is how can four hydrogen bonds lead

to a recognition of five base pairs?

Although he was a brilliant biochemist, he was a poor theoretical chemist. He took a program and “drew in” Distamycin in the DNA minor groove, and asked the

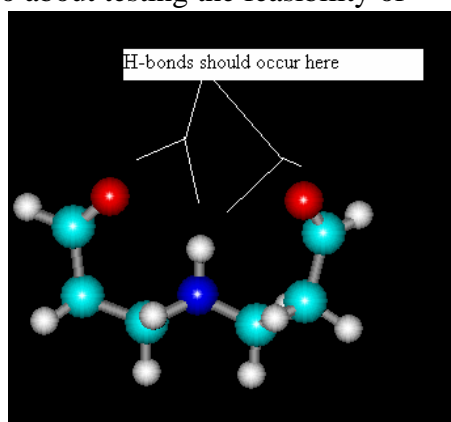
question, “where are the hydrogen bonds?” The program diligently determined where the hydrogen bonds existed, and did something that Peter Dervan did not expect; it drew not one, but TWO hydrogen bonds for each hydrogen in the Distamycin. This led him to resurrect a concept that is even older than I am, the “bifurcated hydrogen bond.” The four hydrogens in Distamycin are recognizing five base pairs, he said, because each hydrogen forms two hydrogen bonds.

Sitting in the audience, as a graduate student, and listening to this, I immediately had two words ring through my head, and the first one was “bull.” See, in a hydrogen bond, electrons are attracted to the hydrogen from a nearby oxygen (or nitrogen or fluorine), but, if you attract the electrons from one oxygen, the charge of these electrons should repel the electrons from another nearby oxygen. The program did exactly what it was programmed to do; it looked at the distance between the hydrogen and the oxygens, and found that two of them were the correct distance for a hydrogen bond, but it was not programmed to consider the actual location of the electrons (or the repulsive forces between them). Immediately, I picked this up as a research project.

It didn't take too long (maybe a few years) for me to have an inspiration; the recognition did not come from the attractive forces, but rather, the repulsive forces! There was too much repulsion between the Distamycin and the DNA if you had guanine or cytosine in the sequence. I wrote a mathematical model to predict DNA recognition by Distamycin, and lo and behold, the model worked beautifully, not only for Distamycin, but for ANY of the sequence specific binders I put in (both molecules that bind to “AT” rich regions of DNA, as well as those that bind to “GC” rich regions). I wrote a paper explaining the results, but never submitted it for publication.

See, Peter Dervan had one thing that I did not; name recognition. His model was already very popular, so to come out with a competitive would draw little attention. I needed one more thing; I needed the theoretical background to demonstrate that the bifurcated hydrogen bond, the centerpiece of his model, was not reasonable.

There are a couple of different ways to go about testing the feasibility of bifurcated hydrogen bonds; theoretical and experimental. What I needed was a “sample molecule” that could, indeed, form bifurcated hydrogen bonds but did not have to. Such a molecule, then, would form the bifurcated hydrogen bond if, indeed, it were more stable than single hydrogen bonds, otherwise, it should only form a single hydrogen bond. The molecule is simple and obvious; dipropanal amine. The very real problem, then, is determining if the bifurcated hydrogen bond is possible for this molecule.



This is where the research went into its cocoon. I tried a few minor approaches for the next few years, then would put the research on a shelf for a few years. It seemed like I could never quite get the break I needed in figuring out how to calculate the bifurcated hydrogen bond. I figured I'd need to run quantum mechanical calculations, which calculates the most likely distribution of electrons in a molecule, but how should I analyze the results?

This year I had a break. In teaching Inorganic Chemistry, one of the subjects was the hydrogen bond. The textbook, and ensuing lecture, led to an idea that I had not considered before. So, as soon as I returned to my office, I started some calculations. The data I've collected from these calculations on the nature of the hydrogen bond, and the bifurcated hydrogen bond, will lead to a paper that I am hoping to submit before the end of the summer. This paper, then, will be used as the foundation to present my alternative model for sequence recognition by Distamycin (the second paper), and will open up the opportunity for experiments to be performed on the actual compound (dipropanal amine), which should lead to a third paper.

Pivot Minimization Algorithm

Suppose you are in a war. You have a certain number of resources (missiles, troops, etc.), a certain number of targets (refineries, enemy troops, command posts, etc.), and a certain percentage that your resource will be destroyed on its way to any given target based on the defenses of the enemy. What is the best plan to maximize injury on your enemy while minimizing loss to yourself?

OK, you don't like death and destruction. How about this; you own a factory which produces a certain number of products. Each product sells for a different price; the lower the price, the more you will sell, but the less you make per sale. What's more, the more you make, the less each unit will cost, so the number of units of each product will influence both the manufacturing cost and the sales price. Because of limited resources, if you make too much of one product, you cannot make as much of the other products. So how many of each product should you produce, and how much should you sell them for, so as to maximize your profit?

Hmm...well, how about an example from chemistry. You have a protein that wants to find its folding configuration with the lowest possible energy. A protein is essentially a long chain of atoms; if you put atoms too close, they will repel one another, but if you have them too far apart, they cannot attract one another. Each type of atom on the chain has a different attractive force with every other atom. How do you fold the chain so you have the lowest possible energy?

There are literally thousands (or even millions or more) of examples like this, and what they all have in common is a very complex mathematical model that must be minimized. I know I know, the second example is not minimization, but rather maximization, but, think about it; change the sign of the function, and the maximization problem becomes a minimization problem. The problem is made more complicated because all of these functions have multiple minima; that is, somewhere in the equation, there is more than one minimum that can be found mathematically, but you are seeking only the LOWEST minimum you can (the "Global Minimum," rather than a "local minimum"). Applications range from chemistry, physics and biology, to micro and macroeconomics, military applications, medicine, ecology, insurance, and just about any other application where you are trying to get the "most out of" some given problem.

How do you solve this? Well, if you plot the function, you can just see where the minimum is. The problem is, most of these problems are not one- or two-dimensional, but rather N dimensional, where N can be in the thousands. So you can't plot it, and, if you ask a computer to find the minimum, the computer has no eyes anyway.

Talk with a math instructor, and they might tell you to take the first derivative to find the extrema and inflection points, then take the second derivative to determine to separate the minima from the maxima and inflection points, and then compare the minima you find to see which is lowest. Well, in an N-dimensional problem, you have to take N simultaneous derivatives, and sort through, potentially, a great number of local minima. These problems could have thousands of parameters and thousands of local minima to go through. In short, they are just not conducive to the “brute force” (or, as my math friends might remind me, the “proper”) approach.

While at Purdue, I came up with a pretty good idea for a minimization algorithm, now known as the “Pivot Method.” The concept is easy; sample the phase space randomly for the function; then sample the regions around the lowest energy samples more carefully. I used the terms “probe” for where we are sampling phase space, and “pivots” for the lower energy “probes” (that is, the areas for further investigation). This yielded excellent results; my algorithm was anywhere from 3 to 7 times more efficient than any other algorithm we could locate on various test functions. But there were two problems.

First, there were too many parameters to choose. How many locations do we sample initially (aka how many probes do we choose)? How do we go about sampling the area around the pivot probes? The second major problem was one of time; as I started a new “teaching gig” in Ohio shortly after this paper was published, I just never had the time to go back and continue to develop the method. Recently, I found a paper that referenced the pivot method, and had one criticism; not enough work has been done to demonstrate its efficiency.

So, I picked up the old program and started working on it. I’ve modified it to try to find the optimal choice for the variable parameters, and to run more realistic functions rather than JUST the standard test functions (better known as the “Lennard-Jones Clusters” problem). Right now, I’m re-running the test functions to try to maximize the parameters. At the same time, I have another group of students re-writing the algorithm into C (rather than Fortran) so we can run it on other machines.

This should result in at least two papers; the first is a report on the maximized choices for parameters (and to remind everybody about the method since it’s been a decade since the last publication), and at least one more on results of the Lennard-Jones clusters.

Here’s where YOU come in!

As always, I love including students in my research whenever possible. Both of these topics do still need a lot of work to prepare the paper, ranging from running the simulations to library work. Regardless of your background, if you find either of these interesting, and would like to spend a little time on it over the summer, whether it be here in Madison or from home (wherever that may be), please feel free to talk with me about the opportunities that this provides. It’s great for a resume; imagine how it would look to say “yes, and I am an author in a paper published by the American Chemical Society.” I’m afraid I cannot offer funds as an incentive; it would have to be strictly voluntary, and as such, that means you can spend as much time as you like, quit at any time, or spend as little as three hours a week. Let me know if you are interested!