May 2016

Protein Adsorption using a Lattice Toy Model

Ari J. Weiland

Macalester College, math1man@gmail.com

MACALESTER COLLEGE

Protein Adsorption using a Lattice Toy Model

Abstract

Protein adsorption is an important subfield of Biophysics particularly relevant in medical science. Using a computational simulation with a basic but configurable two-dimensional square lattice model of approximate amino acid interactions, I investigated the entropic effects of protein adsorption on a weakly attractive surface. These simulations allow for a precise calculation of the partition functions of these complex systems, from which I can then analyze other thermodynamic properties.

Keywords

protein, folding, adsorption, lattice

Abstract

Protein adsorption is an important subfield of Biophysics particularly relevant in medical science. Using a computational simulation with a basic but configurable two-dimensional square lattice model of approximate amino acid interactions, I investigated the entropic effects of protein adsorption on a weakly attractive surface. These simulations allow for a precise calculation of the partition functions of these complex systems, from which I can then analyze other thermodynamic properties.

1

1   Introduction

Protein folding is one of the key focuses of the field of Biophysics. Protein folding is the study of what shape a protein sequence will take, and how it takes that shape. Protein adsorption, the process by which a protein adheres to a surface, is a less-explored but incredibly interesting aspect of the study of protein folding, with applications such as the design of medical tools and implants, and studies of immune responses, blood clotting, and strokes. As with much of the current research pertaining to protein folding, adsorption is not all that well understood and experiments that seek to model it tend to be computationally intensive.

Lattice models, first notably pioneered in the 1980s [1], offer a simplified way to represent, fold, and analyze proteins by describing conformations on a two- or three-dimensional lattice. These models are very configurable, and usually also approximate interactions between amino acids by categorizing them as one of a few different types of interactions, each with a single defined energy. These models usually focus on entropic interactions between amino acids and water, the so-called hydrophobic interaction [1], which is currently believed to be the primary driving force behind protein folding.

In this research, I attempted to understand what factors impact the likelihood of a protein to adsorb to a surface. In addition, I considered an alternate model of protein adsorption where a protein is represented as a point particle and compared this alternate model to see if this further simplification might

2

still be a meaningful interpretation.

2 Methodology

2.1 Enumerative Algorithm

In order to model these simplified lattice proteins, I developed a fully enumerative algorithm that counts all the possible conformations a polypeptide sequence could take, and the energy associated with each state. This allows me to exactly determine the density of states for a particular system. I developed versions for enumerating the density of states in systems with and without a surface, and in two or three dimension, each of which behave slightly differently.

I programmed the algorithm from scratch in Java. Without a surface, the algorithm generates every state by dropping the first amino acid at the origin, then generating conformations for every possible position that each consecutive amino acid can take, where it must be directly (not diagonally) adjacent to the previous amino acid. To save computation time, the algorithm only generates one of the four or six positions for the second amino acid, as each of these will generate a subset of symmetrical but rotated conformations. This symmetrical factor is then multiplied into the density of states after the counting completes.

When using a surface, it is defined as an infinite line or plane at the coordinates where $y = 0$. The algorithm otherwise behaves as described above,

3

except that the first amino acid is placed at a point above the surface ($y > 0$). All starting points where the polypeptide sequence may be able to come into contact with the surface are tested, though only resulting conformations with at least one contact with the surface are actually counted. Also, because the existence of the surface breaks the rotational symmetry of the system, all possible positions of the second amino acid are tested.

Because the number of conformations is exponentially related to the length of the sequence, my algorithm has exponential time complexity, making it impractical to run on sequences much larger than 26 in two dimensions without a surface, 21 with a surface, 18 without a surface in three dimension, or 14 with a surface.

To test my algorithm, I first made sure that it generated the appropriate number of conformations in three dimensions for a sequence of a given length as listed in Table 1 from [2]. I also calculated specific heat curves from the densities of states and compared them to similar graphs from the literature. Figure 1 demonstrates one of these comparisons to Figure 11c from [2].

2.2   Lattice Interactions

The interactions included in a lattice model, along with the shape and dimension, define that model. I included the hydrophobic and hydrophilic interactions. This method classifies every amino acid as either hydrophobic (H) or hydrophilic (P), and establishes effective interaction energies between them. As per [2], I defined my energy scale in terms of the hydrophobic-

4

hydrophobic interaction by defining $\epsilon_{HH} = -1$. Note that the energy is negative, indicating a favorable interaction. From [3], I also chose to define a much weaker interaction $\epsilon_{HP} = -1/7$ between hydrophobic and hydrophilic amino acids. The interaction strength between hydrophilic amino acids is effectively 0.

Because the surface is not inherently biological, I defined the interaction arbitrarily compared to the other interactions. Similar to [4] and [5], I focused on a surface which interacts equally with both hydrophobic and hydrophilic amino acids. Though I explored various strengths for the interaction ranging from 0 and 0.01 to 1, I focused on an interaction strength of $\epsilon_S = -0.08$, similar to the value of $-1/12$ used in [4] and [5].

2.3 Sequences

Though at different stages of my research I ran simulations on various polypeptide sequences with lengths ranging from 13 to 26, I focused the majority of my final data collection on one particular sequence. Both [2] and [3] recommended focusing on "designing" sequences: sequences with a single lowest-energy conformation (up to symmetry). For efficiency, I restricted myself to gather data in two dimensions, and chose to focus primarily on the sequence $S_{21} = $ PHPHPPHPHPPHPPHPHPPHP from [3]. This sequence of length 21 is a designing sequence with 8 H amino acids and 13 P amino acids. A length of 21 was ideal because it was right at the limit of my computational power. Figure 2 shows the lowest-energy conformation without a surface,

5

and figure 3 shows the lowest-energy conformation with a weakly attractive surface.

3 Results and Discussion

3.1 Calculating Adsorption

The probability of adsorption can be easily calculated from the densities of states. Let a solution have some volume V, and some attractive surface area A. Regardless of the actual geometry of the system, if the lattice is extended to the entire solution, then the probability of a protein adsorbing is

$$\mathcal{P}(S) = \frac{c_S \mathcal{Z}_S}{c_B \mathcal{Z}_B + c_S \mathcal{Z}_S} \tag{1}$$

where $\mathcal{Z}$ are the partition functions on the surface and in bulk, respectively, and the c coefficients are the number of lattice locations for the protein to be either on the surface or in bulk. Since the partition function gives the probability of being in any state, it is appropriate to use here.

Calculating the c coefficients is fairly straightforward. The number of lattice locations on the surface is proportional to the surface area, and therefore given by

$$c_s = \frac{A}{d^2} \tag{2}$$

6

where d is the approximate diameter of an amino acid. Similarly,

$$c_b = \frac{V}{d^3}.$$
(3)

However, in most realistic systems, the volume will be much greater than the surface area, so $c_B \mathcal{Z}_B \gg c_S \mathcal{Z}_S$ in the denominator, and we can simplify equation 1 to

$$\mathcal{P}(S) = d\frac{A}{V}\frac{\mathcal{Z}_S}{\mathcal{Z}_B},$$
(4)

the approximate probability of a protein adsorbing in a solution with the appropriate surface area and volume.

If the solution contains more than one protein in it, we can calculate the number of proteins adsorbed by multiplying equation 4 by the total number of proteins. We can express the number of proteins as a concentration $[P]$ times the volume, which gives us the number adsorbed as

$$N_{ads} = [P] * d * A\frac{\mathcal{Z}_S}{\mathcal{Z}_B}.$$
(5)

To completely remove the dependency on the system geometry, I can divide both sides of the equation by the surface area, giving the number adsorbed per unit area

$$N_{ads/A} = [P] * d * \frac{\mathcal{Z}_S}{\mathcal{Z}_B},$$
(6)

7

a function completely independent of the geometry of the system.

3.2 Adsorption Data

Using equation 6, I was able to calculate adsorption per unit area for an arbitrarily defined system. I defined my concentration arbitrarily as $[P] = 10^{-9}$ mol/L, a rough approximation based on blood protein concentrations, and calculated the approximate diameter of an amino acid as $d = 5*10^{-7}$ mm, based on the average volume of an amino acid. I ran simulations for surface interaction energy ϵ_S values on the interval $[0, 0.2\epsilon_{HH}]$ at 0.01 increments, as well as $\epsilon_S = 0.3$ and $\epsilon_S = 1.0$, the latter being an unrealistic but potentially interesting scenario.

Figure 4, the most basic resulting curve, shows the number of proteins adsorbed versus surface interaction energy. Note that even when the surface interaction energy is nonexistent, there is still a small amount of adsorption, which corresponds to the random likelihood that at any given time a protein might happen to brush up against the surface. Also note that the curve appears to have a nice exponential shape. This will be emphasized later. Many of the following curves are modifications of this curve to emphasize different features.

Figure 5 shows the same curve with the random adsorption factor subtracted out. This figure emphasizes the meaningful adsorption as a function of interaction energy. In contrast, figure 6 shows the curve with the random adsorption factor divided out. This curve emphasizes the general underlying

8

behavior and is only dependent on the particular sequence and the definition of the hydrophobic interaction strengths. Figure 7 shows the log of the previous curve. This figure emphasizes the strongly exponential relationship between adsorption and interaction energy, given the linear nature of the curve. This exponential relationship is the primary result of this section.

Figure 8 shows a contour of the log of adsorption ratio across multiple temperature values as well as for varying interaction energies. All of the constant temperature contours are approximately linear, reinforcing the previously described exponential relationship. This curve also shows the temperature versus adsorption relationship. In the low temperature limit, these curves approach infinity as it becomes infinitely unfavorable to not be on the surface, while at the high temperature limit the log of the relationship approaches zero, meaning that the ratio approaches one. This is because in the high temperature limit, the entropic effects dominate energetic effects and the only adsorption is random baseline adsorption.

3.3 Point Particle Approximation

After investigating the relationships between adsorption, temperature, and interaction strength, I then investigated a more directly practical application: investigating whether treating proteins with a complex internal structure as point particles was a valid experimental simplification. To define this point particle approximation, I defined a single state for the particle adsorbed, and a single state for the particle in solution.

9

I calculated the internal energy as the average energy of the possible conformation states. Initially I plotted adsorption curves of this new model as compared to the precise model in figure 9, but the graphs did not show anything meaningful aside from a convergence of the models at high temperatures and a divergence of the models at low temperatures. This behavior of follows from the theory, because at high temperature the internal states tend towards being equally probable, which is equivalent to them being indistinguishable and thus equivalent to a single, average state. The divergence of the models is explained similarly by the inverse argument.

To get a more meaningful perspective on the comparison of these models, I plotted their ratio in figure 10. The curve clearly demonstrates the convergence at high temperatures, where the ratio quickly approaches one, while at low temperatures it diverges very significantly. The curve also reveals more of the actual relationship to the divergence region, namely that it peaks and then approaches 4 as temperature goes to zero. These two features turn out to be consistent across most interaction strengths for this particular polypeptide sequence, though, the zero temperature limit can vary for other sequences. The height of the peak also varies seemingly exponentially with interaction strength, though is not incredibly significant since even a small deviation away from a value of 1 indicates a fairly significant deviation in adsorption.

The zero temperature limit turns out to correspond to the ratio of the number lowest energy states in bulk versus on the surface. As temperature

10

approaches zero, the lowest energy Boltzmann factor overpowers all the others, while the average energy also approaches the lowest energy. Thus in the ratio, the Boltzmann factor terms for both models cancel out, leaving the ratio of multiplicities. In the point model, this ratio is just $1/1$, leaving just the multiplicity ratio for the precise model. In the case of my sequence, that ratio is $8/2 = 4$.

This comparison would indicate that at reasonably high temperatures, a point particle approximation would be a reasonable simplification when studying protein adsorption, while at lower temperatures, the model breaks down. However, quantifying what "reasonably high" actually meant posed difficulties. Because the temperature scale is defined relative to the strength of the hydrophobic interaction (as is typical in the literature [2][3][4][5]), I first had to establish a value for that interaction.

The problem arises from the fact that the most commonly used value for this is $\epsilon_{HH} = \alpha k_B T$, where α is a constant that is usually given on the order of magnitude of 1. However, because of the inherent temperature dependence, defining my temperature scale as $k_B T / \epsilon_{HH}$ reduces to α^{-1}. Thus the scale is essentially a function of this constant, whose meaningful range on the graph could fall anywhere between the "high temperature" convergence or the "low temperature" divergence. Of course, in reality that definition of the hydrophobic interaction is likely an oversimplification, but in order to treat it more precisely would still require much more careful calculations.

11

4   Summary and Conclusions

All of these conclusions generally follow logically from normal thermodynamic behavior. The data indicates that adsorption is exponentially related to the surface interaction strength, while the relationship to temperature seems to follow a relationship similar to an inverse relationship. Of course, the definition of temperature used, which appears to be standard in this area of research, is confusing and not ideal, and more work both regarding adsorption and protein folding in general needs to be done to treat the temperature variable and the hydrophobic interaction more rigorously.

My interpretations of the point particle model also generally follow normal thermodynamic behavior. As one might predict, at high temperature the point particle model approaches the precise model, and I hypothesize that the model will be valid in this case. Unfortunately, due to the temperature consideration, I was unable to get a sense of what exactly qualifies as high temperature. Again, more work is needed to treat the temperature variable more rigorously.

In addition, future work should investigate more different polypeptide sequences, especially sequences of varying lengths (notably longer and more realistic). Explorations of simulations in three dimensions would also be interesting. I hypothesize that most of these conclusions should hold true independent of the sequence (or the dimension), but without large amounts of data on various sequences, I cannot verify this hypothesis. Unfortunately, my

12

enumerative model is relatively slow, and takes exponential time so cannot handle longer sequences or three dimensions. Supercomputers could improve the time, but a better solution would be to investigate statistical simulations that have simpler time complexities. I explored the Wang-Landau simulation[4][5] method on small molecules, but could not get it match up with my other data, and often it would not even give data consistent with itself. More work is needed on this and other statistical simulations.

13

References

[1] Lau, Kit Fun, and Ken A. Dill. "A Lattice Statistical Mechanics Model of the Conformational and Sequence Spaces of Proteins." *Macromolecules* 22.10 (1989): 3986-997. Web.

[2] Schiemann, Reinhard, Michael Bachmann, and Wolfhard Janke. "Exact Sequence Analysis for Three-dimensional Hydrophobic-polar Lattice Proteins." *The Journal of Chemical Physics J. Chem. Phys.* 122.11 (2005): 114705. Web. 5 Feb. 2016.

[3] Lee, Jae Hwan, Seung-Yeon Kim, and Julian Lee. "Study on Collapse and Folding Transitions of a Lattice Protein Using Exact Enumeration." *AIP Advances* 5.12 (2015): 127211. Web. 5 Feb. 2016.

[4] Li, Ying Wai. Unraveling Universal Thermodynamic and Structural Behavior of HP Model Protein Adsorption a Wang-Landau Study. Thesis. 2012. Print.

[5] Li, Ying Wai, Thomas Wüst, and David P. Landau. "Generic Folding and Transition Hierarchies for Surface Adsorption of Hydrophobic-polar Lattice Model Proteins." *Physical Review E Phys. Rev. E* 87.1 (2013). Web. 11 Jan. 2016.

14

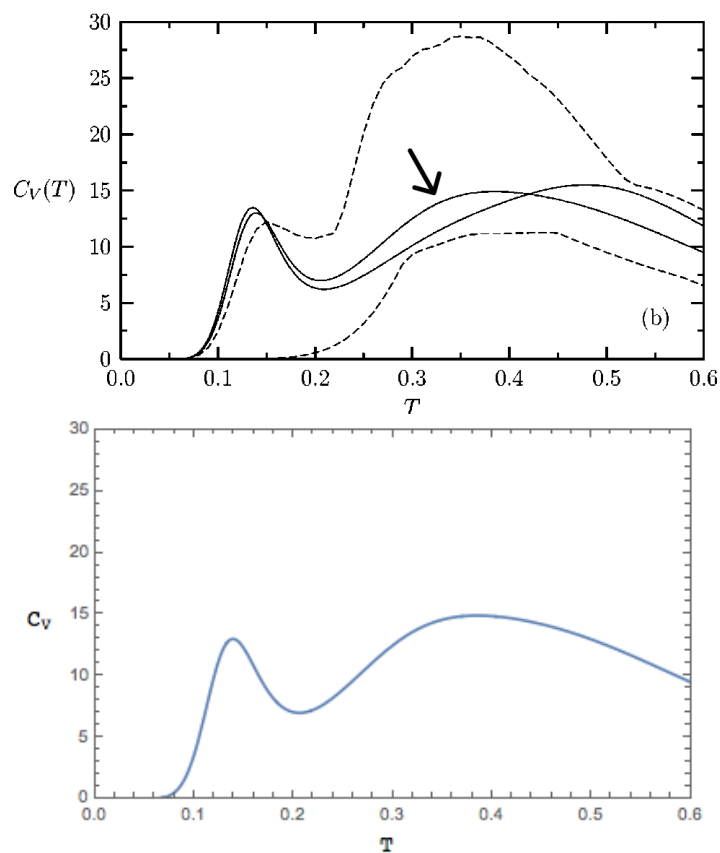# **Figures**

Validation Test

Figure 1: Comparison of my algorithm's results versus previously published results. The indicated curve (top) from [2] matches my curve (bottom). I could not get precise data on their curve, but visibly they appear identical.
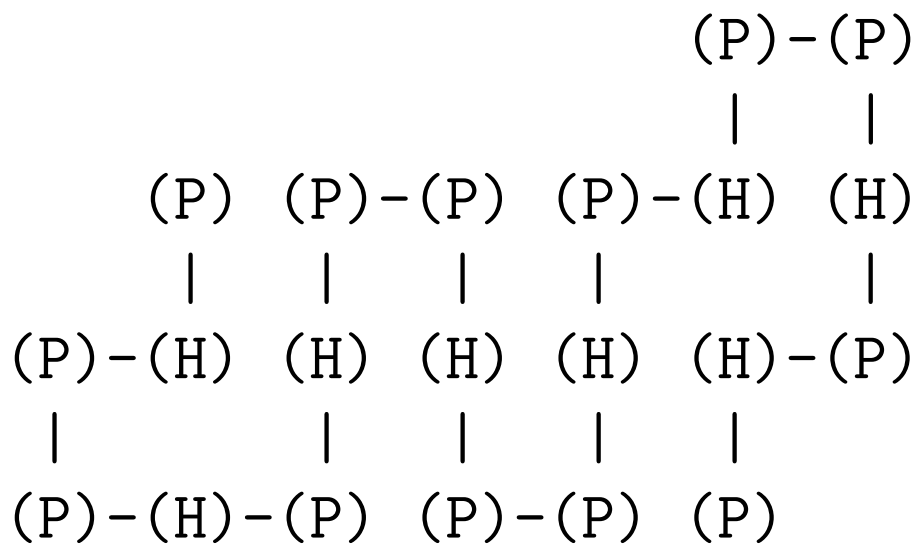
15

```
                                        (P)-(P)
                                         |    |
                    (P)  (P)-(P)  (P)-(H)  (H)
                     |    |    |    |         |
                (P)-(H)  (H)  (H)  (H)  (H)-(P)
                 |         |    |    |    |
                (P)-(H)-(P)  (P)-(P)  (P)
```

Figure 2: Lowest energy conformation of S_{21} without a surface. $E = -7.0$

16

```
                                (P)-(P)
                                 |   |
        (P)  (P)-(P)  (P)-(H)  (H)
         |    |    |    |         |
    (P)-(H)  (H)  (H)  (H)  (H)-(P)
     |             |    |    |    |
    (P)-(H)-(P)  (P)-(P)  (P)

    (S)  (S)  (S)  (S)  (S)  (S)  (S)
    -+---+---+---+---+---+---+-
```
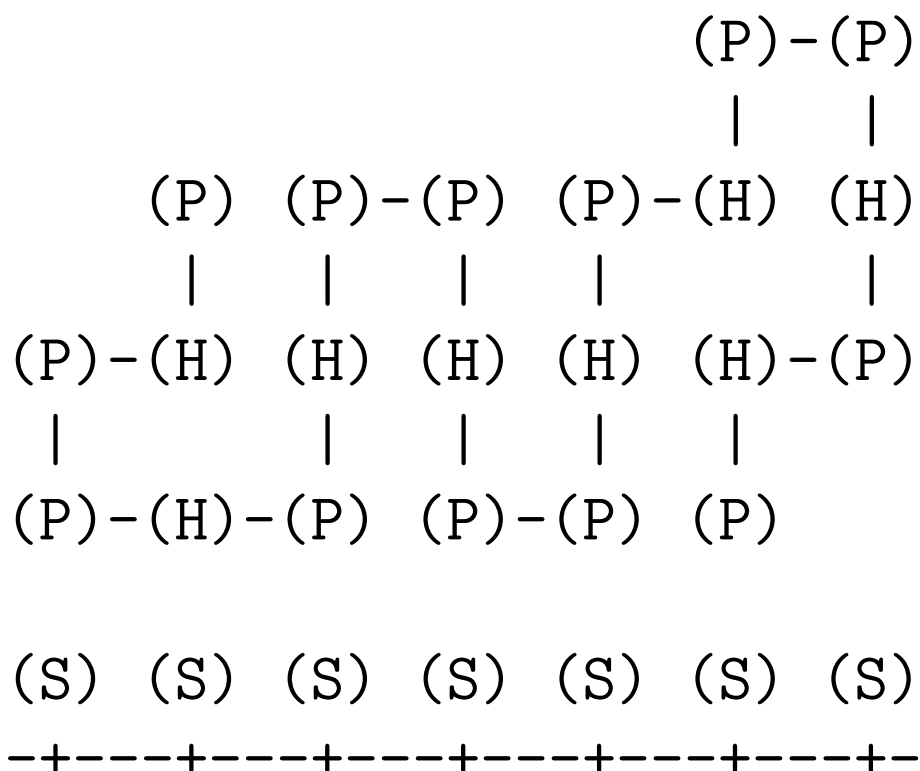
Figure 3: Lowest energy conformation of S_{21} on a weakly attractive surface. For surface interaction $\epsilon_S = 0.08$, $E = -7.48$

17

Number of Adsorbed Proteins
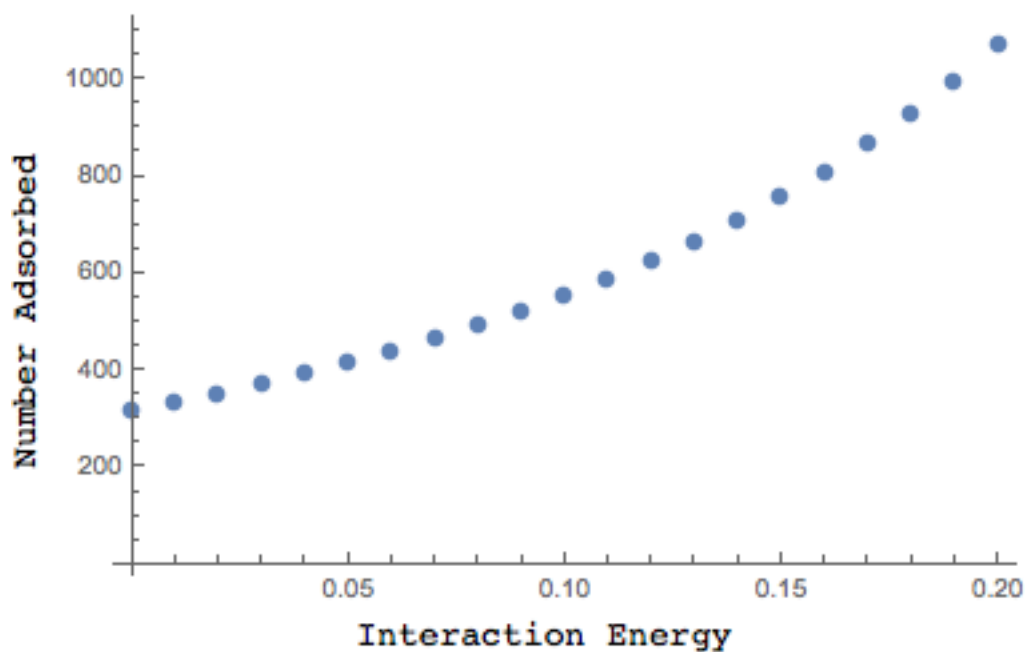
Figure 4: Number of adsorbed proteins at $[P] = 10^{-9}$ mol/L and $k_B T/\epsilon_{HH} = 0.5$. Note that even at surface interaction energy $\epsilon_S = 0$ there is a small amount of adsorption due to completely random kinetic factors.

18
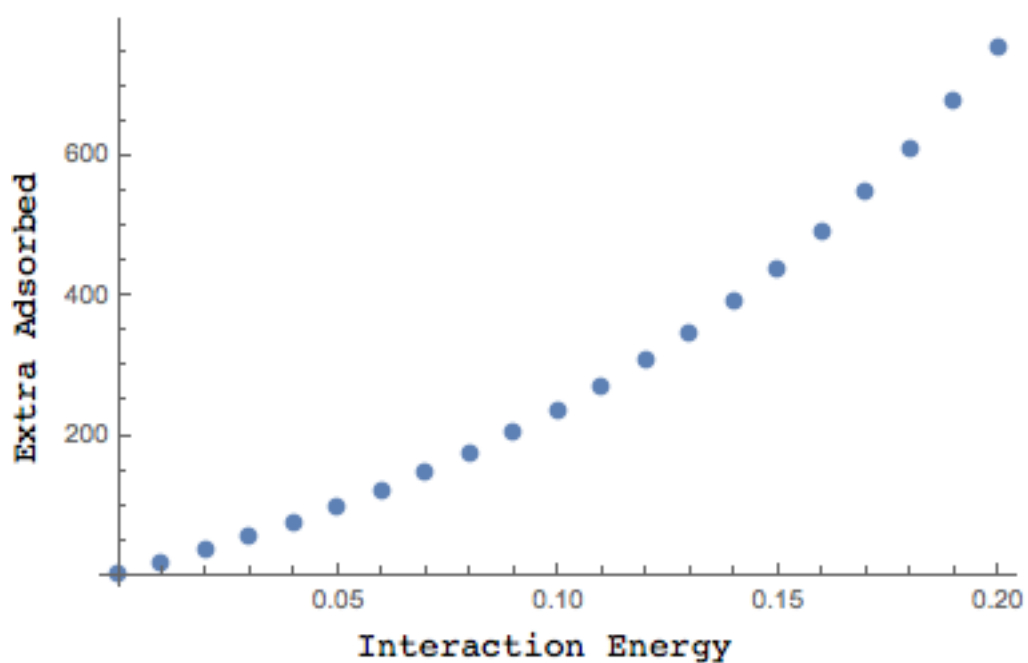
Figure 5: Number of adsorbed proteins more than the baseline at $[P] = 10^{-9}$ mol/L and $k_B T / \epsilon_{HH} = 0.5$. This graph zeroes the adsorption curve by subtracting out the random baseline adsorption.
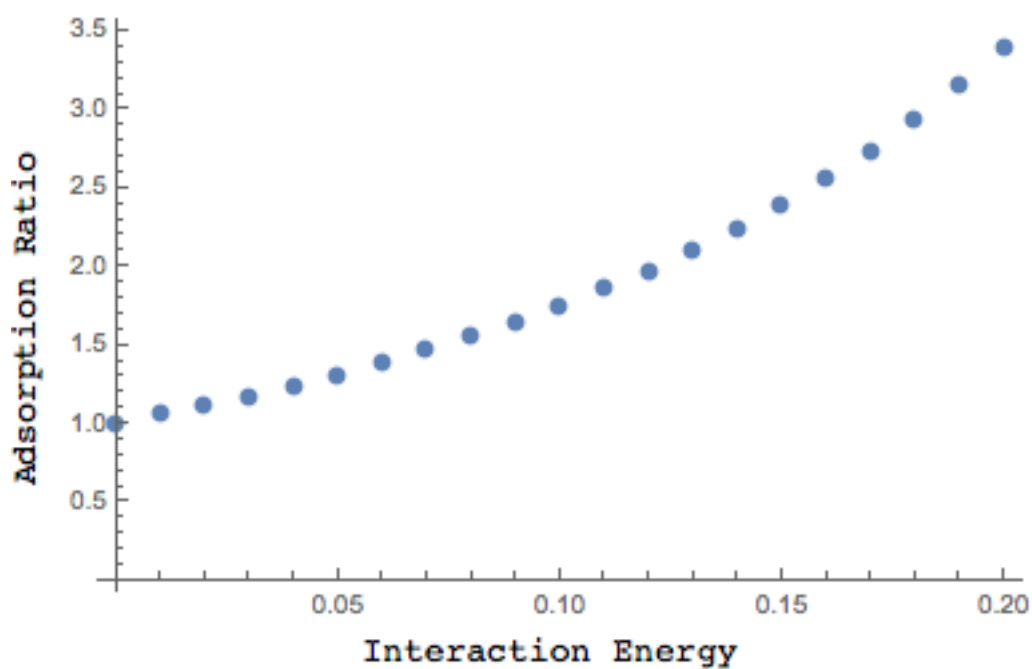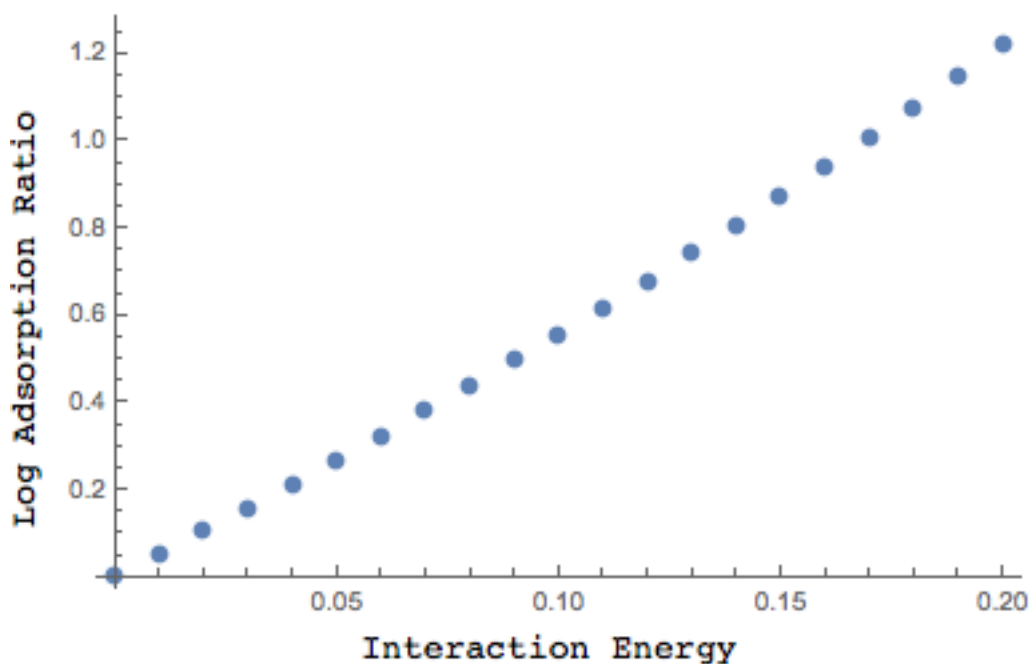
19

Ratio of Adsorbed Proteins

Figure 6: Ratio of the number of adsorbed proteins at $k_B T / \epsilon_{HH} = 0.5$ to the number adsorbed without a surface interaction. This graph is effectively independent of the concentration or number of proteins in solution.

20

Log of the Ratio of Adsorbed Proteins

Figure 7: The log of the ratio of the number of adsorbed proteins at $k_B T/\epsilon_{HH} = 0.5$ to the number adsorbed without a surface interaction. This graph demonstrates the exponential relationship between adsorption and surface interaction energy.
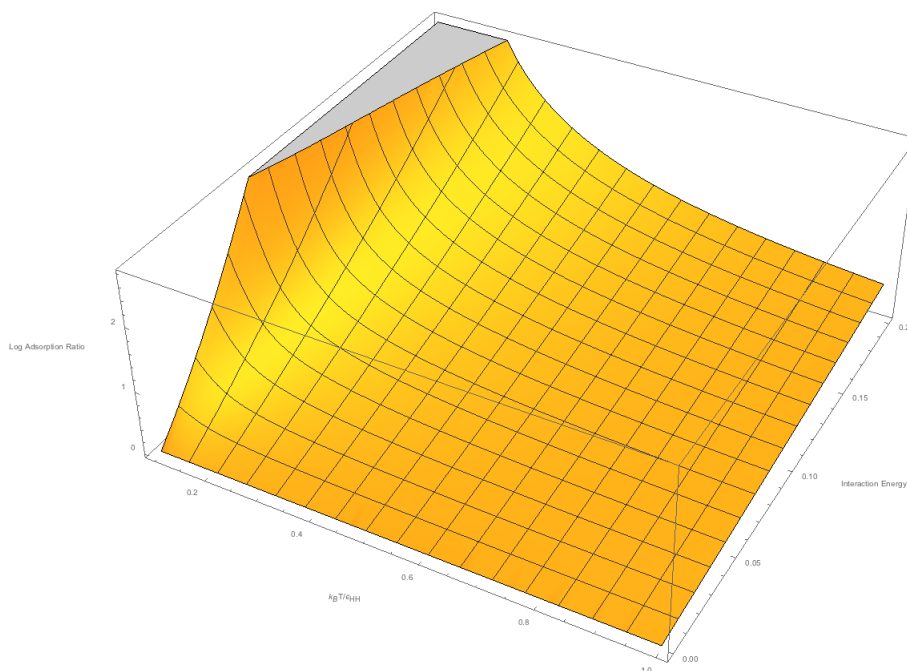
21

Contour of the Log of the Ratio of Adsorption

Figure 8: The log of the ratio of the number of adsorbed proteins at various temperatures. Note that all of the constant temperature contour lines are roughly linear, suggesting that the exponential relationship holds across temperatures.
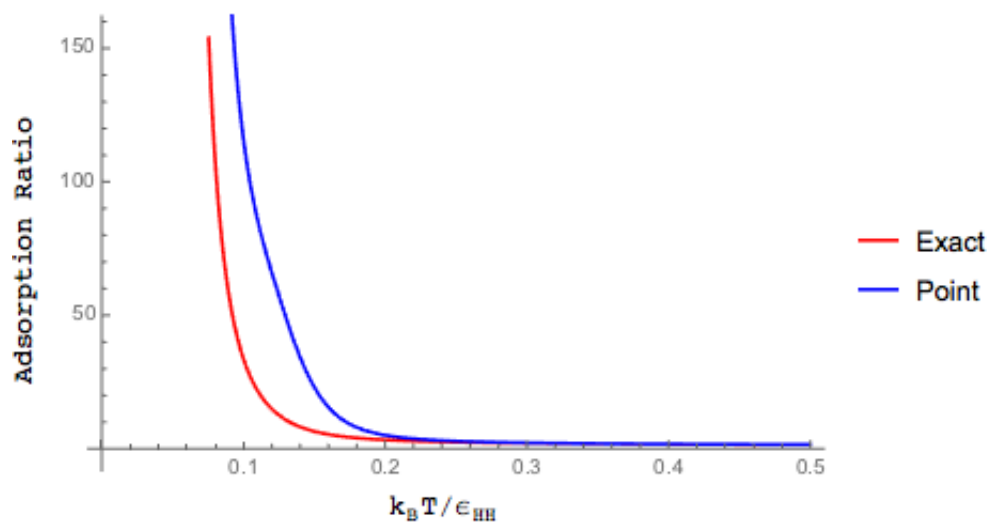
22

Figure 9: Comparison of the point particle approximation adsorption to the "exact" adsorption, at a surface interaction $\epsilon_S = 0.08\epsilon_{HH}$. The two models clearly diverge around $T = 0.2$, but it is hard to tell what is happening any more specifically.
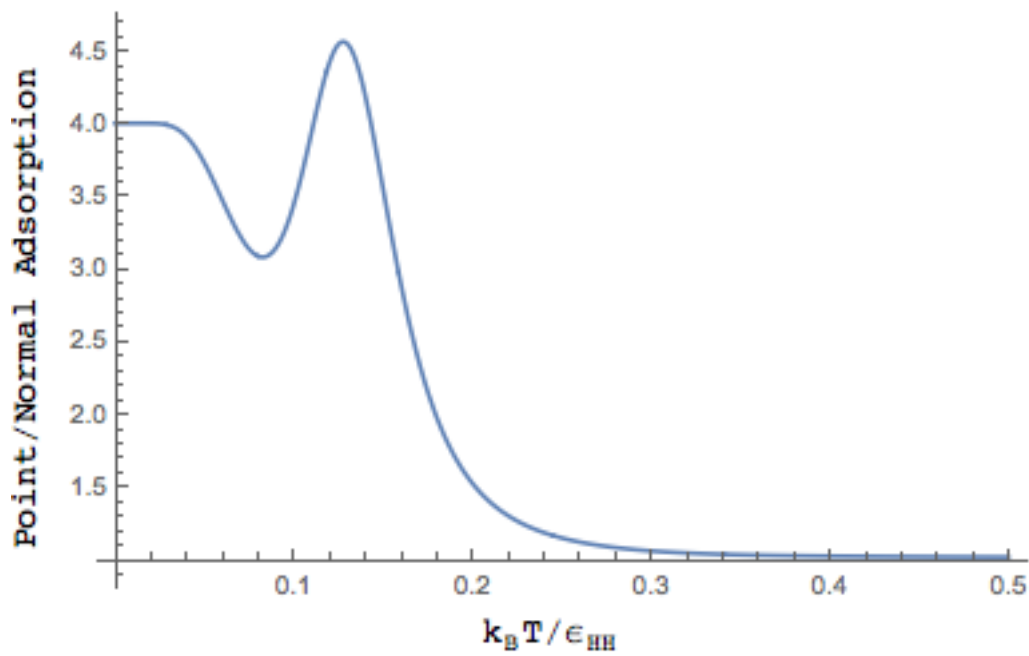
23

Point Particle Adsorption Comparison Ratio

Figure 10: Ratio of the point particle approximation adsorption to the standard adsorption, at a surface interaction $\epsilon_S = 0.08\epsilon_{HH}$. At high temperatures, the two models are almost identical as the ratio approaches 1, while at low temperatures, the models diverge significantly.

24