

Lattice protein models

Marc R. Roussel
Department of Chemistry and Biochemistry
University of Lethbridge

March 5, 2009

1 Model and assumptions

The ideas developed in the last few lectures can be applied to more complicated systems. In particular, the monomers don't all have to be identical. In these notes, we will develop a lattice model of proteins. We will see that many known thermal properties of proteins are captured, at least qualitatively, by this model.

Ordinary proteins are made up of one of the 20 standard amino acids. These amino acids differ by an R group attached to the central carbon. The R groups corresponding to the 20 naturally occurring amino acids vary greatly in properties. In particular, some are polar and some non-polar. In our last lecture, we saw that there is the energy involved in making a contact between a polymer and the solvent (water in this case) is

$$\Delta\epsilon = \epsilon_{12} - \frac{1}{2}(\epsilon_{11} + \epsilon_{22}),$$

where ϵ_{12} is the energy released when a new contact is made between the solvent and a segment of the polymer, ϵ_{11} is the energy of a solvent-solvent interaction, and ϵ_{22} is the energy of a solute-solute interaction. In the case of a protein, There should be different values of ϵ_{12} for each amino acid. For a start, we can distinguish two cases, namely that of a hydrophilic amino acid, and a hydrophobic amino acid. A hydrophilic amino acid would have $\Delta\epsilon \lesssim 0$, which a hydrophobic amino acid would have $\Delta\epsilon \gg 0$. On energetic grounds alone, during protein folding, the tendency would be for hydrophobic amino acids to become buried inside the protein to avoid unfavorable contacts with the solvent, while hydrophilic amino acids remain on the surface in contact with the solvent. A number of biochemical studies have confirmed that hydrophobic burial is a key factor in controlling the folding of many protein domains,¹ although clearly it is not the only factor of importance.

We are going to look at a highly simplified model of protein folding. As hinted above, we will start with a lattice model, similar to the one we studied for simple polymers. We are going to confine ourselves to a square lattice in two dimensions. Surprisingly, this is

¹A protein domain is a part of a protein that folds independently from the rest of the protein, i.e. if we cut a domain out of the protein, it will still fold the same way.

enough to discuss the effect of hydrophobic burial on the thermodynamics of proteins. Only nearest-neighbor contacts on this lattice will be considered. The amino acids will be labeled A, B, C, ... starting at the amino-terminal end.² We could look at a detailed model including all 20 amino acids, but from the discussion above, you may have concluded that the hydrophilic/hydrophobic distinction is, at least in the first instance, the most important one. We will therefore reduce the 20-amino acid genetic code to a binary code where we only concern ourselves with whether a given amino acid is hydrophilic or hydrophobic. Specifically, we will encode a protein sequence using a string of 1's and 0's, where 1 represents a hydrophobic amino acid and 0 a hydrophilic amino acid. This is of course a radical approximation given that there are degrees of hydrophobicity or hydrophilicity, according to the actual value of $\Delta\epsilon$ for a given amino acid. We will moreover assume that $\Delta\epsilon \approx 0$ for typical hydrophilic amino acids. Thus, in what follows, we use $\Delta\epsilon$ to represent the average (positive) contact energy for a hydrophobic-solvent contact. There could be yet another contact energy between a hydrophobic and a hydrophilic amino acid, but we won't run into these in the very simple proteins we will treat here.

2 Thermodynamics of a lattice tetramer

2.1 Conformations of a tetramer on a lattice

Suppose that we study a four-amino acid protein in the two-dimensional lattice with the binary sequence 1001, i.e. the first and last amino acids are hydrophobic. Figure 1 shows the five possible conformations of a tetramer on a two-dimensional lattice. Solvent molecules occupy any sites not occupied by amino acids. Any other tetramer you could draw would just be a rotation of one of these five shapes, which would not be a different conformation. Note that tetramers 2b and 2c are distinguishable because the two ends of the protein (amino- and carboxy-terminal) are distinguishable, even if the tetramer has a mirror-symmetric sequence (e.g. LeuSerSerLeu).

We can now assign energies to our five conformations. Conformations 2a–d all have identical energies because they each have six (three per hydrophobic amino acid) contacts between hydrophobic amino acids and the solvent. Thus, for these tetramers,

$$\epsilon_2 = 6\Delta\epsilon$$

with degeneracy $g_2 = 4$. On the other hand, conformation 1 only has four nearest-neighbor hydrophobic-solvent contacts. Its energy is therefore

$$\epsilon_1 = 4\Delta\epsilon.$$

Since $\Delta\epsilon > 0$, $\epsilon_1 < \epsilon_2$. We therefore have a unique ground-state fold. The ground-state fold is often the active form of a protein. We will assume that this is the case here.

²We don't need to know much biochemistry here, but for your information, proteins are synthesized by making a peptide bond between an a carboxylic acid and an amine group on two on two amino acids. Each protein therefore has one end with a free amine group and another with a free carboxylic acid. Synthesis of amino acids starts at the amino-terminal end, so if we're going to list amino acids in order, this is the end it makes most sense to start from.

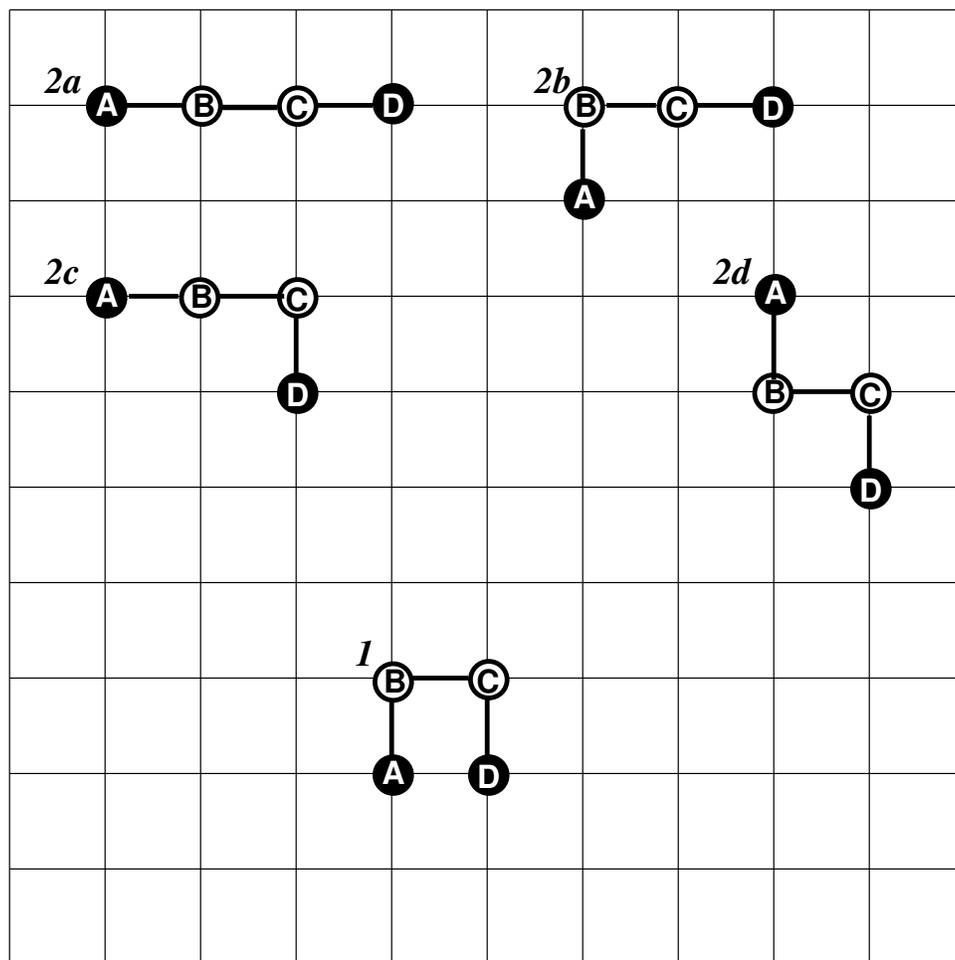


Figure 1: Possible conformations of a two-dimensional lattice tetramer. This particular tetramer has the binary sequence 1001, symbolized by black circles for the hydrophobic amino acids, and white circles for the hydrophilic amino acids. Solvent molecules are assumed to fill any lattice sites not occupied by amino acids.

Before we go too much further, let's talk briefly about one aspect of this model that isn't very satisfactory from the point of view of real proteins. Suppose that amino acid C was hydrophobic. In all five of the conformations, this amino acid has two solvent contacts. This model would therefore predict that all of these folds are equivalent with respect to a hydrophobic amino acid C. However, hydrophobicity in amino acids is associated with the R group. By suitable internal rotations, in conformation 1, it might be possible to bring C's R group into the U-shaped pocket we have in this fold. Our system for counting hydrophobic contacts on a lattice is therefore a little naïve. We have to remember however that we are working with a *model* which we intend to use to discuss the effect of hydrophobicity on protein thermal properties, and *not* one intended to allow us to predict the fold of any particular protein. Our model is adequate for the former application, but clearly not for the latter.

2.2 Partition function and denaturation curve

The molecular partition function is

$$q = e^{-\epsilon_1/(kT)} + g_2 e^{-\epsilon_2/(kT)} = e^{-4\Delta\epsilon/(kT)} + 4e^{-6\Delta\epsilon/(kT)}.$$

Biochemists are often interested in **denaturation curves**. This is a plot of the fraction of the protein in an active conformation vs some parameter (temperature, ionic strength, concentration of a denaturing agent,³ etc.). Assuming that the ground-state fold is the active conformation, the fraction of molecules in the active conformation in a large ensemble corresponds to p_1 , the probability that a particular molecule is in the ground state. We will just consider the effect of temperature here, but you can probably imagine how we could incorporate other effects in our simple model.

The probability p_1 is

$$p_1 = e^{-4\Delta\epsilon/(kT)}/q.$$

This depends only on the variable $u = kT/\Delta\epsilon$. We could therefore study this question in a very general way by using this variable. This would tell us how p_1 depends on T without having to pick a value for $\Delta\epsilon$. To be definite though, I'm going to do my calculations with $\Delta\epsilon = 5$ kJ/mol. If you look at the definition of $\Delta\epsilon$ and consider that a hydrogen bond typically contributes somewhere between 5 and 15 kJ/mol to the binding energy of a system, you will see that this is a reasonable value to use. In any event, different values of $\Delta\epsilon$ would only shift the temperatures at which various features are observed, not the qualitative behavior.⁴

Figure 2 shows a denaturation curve for our tetramer. The sigmoidal shape of this curve is fairly typical of such curves, except that similar curves for real proteins tend to have a much steeper profile. Of course, we're looking at a particularly small model with just two hydrophobic amino acids at either end of a short chain, and our model has a number of other unrealistic aspects. Nevertheless, it's nice to see that our model can reproduce one well-known property of proteins.

³A denaturing agent is a chemical that promotes unfolding of a protein.

⁴Technically, changing $\Delta\epsilon$ corresponds to a stretch of the T axis.

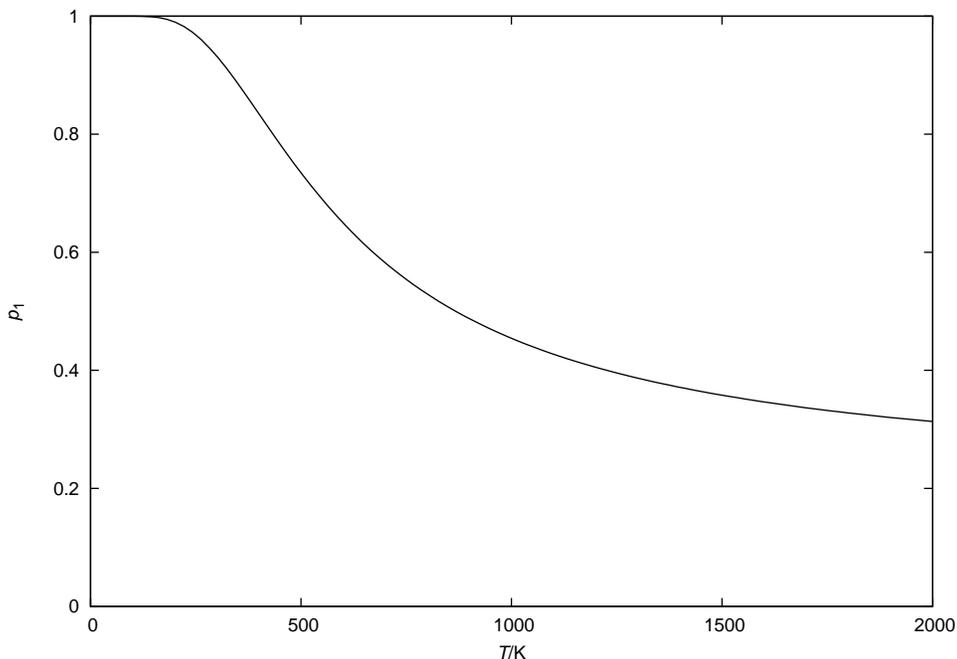


Figure 2: Denaturation curve for the 1001 tetramer with $\Delta\epsilon = 5$ kJ/mol.

2.3 Internal energy and heat capacity

Recall that the molar internal energy can be calculated by

$$U = RT^2 \left. \frac{\partial \ln q}{\partial T} \right|_V.$$

Evaluating the derivative and using the fact that $R = kN_A$, we get

$$U = \frac{4\Delta\epsilon_m (e^{-4\Delta\epsilon/(kT)} + 6e^{-6\Delta\epsilon/(kT)})}{e^{-4\Delta\epsilon/(kT)} + 4e^{-6\Delta\epsilon/(kT)}},$$

where $\Delta\epsilon_m$ is the molar value of $\Delta\epsilon$. Figure 3 shows the internal energy vs temperature. The energy is $4\Delta\epsilon$ at low temperature, and increases toward the high-temperature limit $(28/5)\Delta\epsilon$ (the average energy of the five conformations).

We can also easily get $C_{V,m}$ by differentiating the internal energy with respect to T :

$$C_{V,m} = \left. \frac{\partial U}{\partial T} \right|_V = \frac{16(\Delta\epsilon_m)^2 e^{2\Delta\epsilon/(kT)}}{RT^2 (e^{2\Delta\epsilon/(kT)} + 4)^2}.$$

Figure 4 shows the heat capacity as a function of T : The heat capacity is essentially zero at low temperatures. It increases sharply near $T = 200$ K as the denatured states become accessible, goes through a maximum, then slowly decreases toward zero. A sharp increase in the heat capacity is characteristic of a phase transition, and the event seen here is sometimes

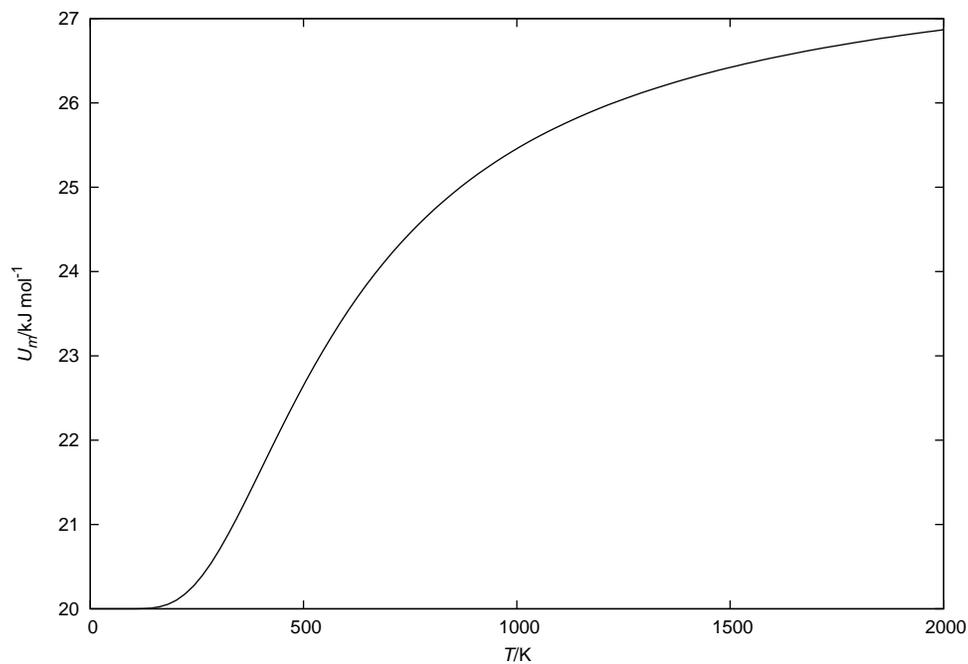


Figure 3: Molar internal energy of the solvated protein as a function of temperature if $\Delta\epsilon = 5$ kJ/mol.

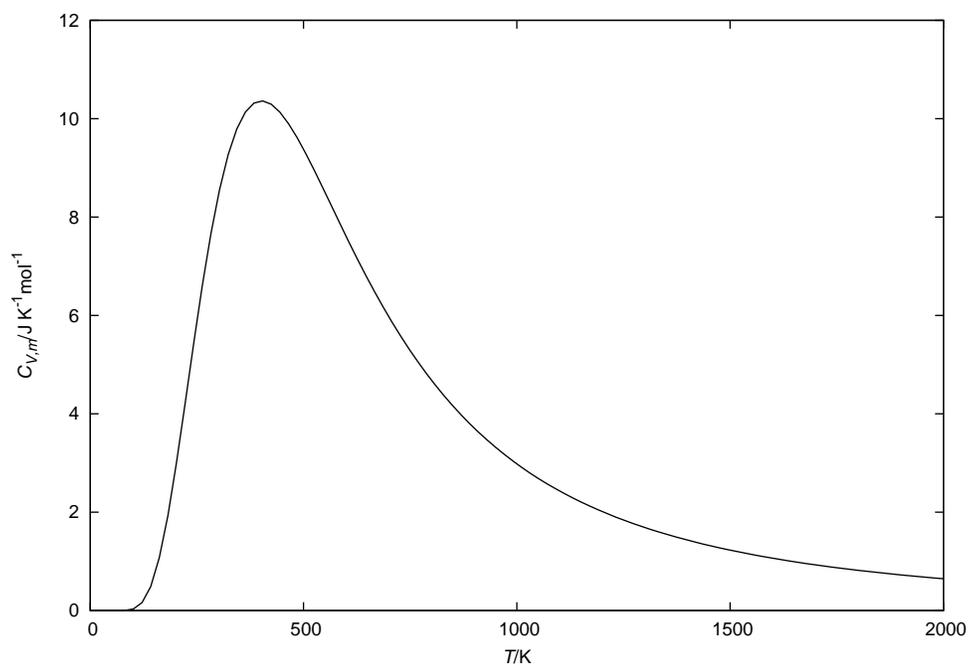


Figure 4: Molar constant-volume heat capacity of the solvated protein as a function of temperature with $\Delta\epsilon = 5$ kJ/mol.

described as “melting” of a protein. Heat capacity curves like the one seen in figure 4 are routinely observed during the thermal denaturation of proteins, notably by differential scanning calorimetry. Again our very simple model has reproduced an experimental property of proteins.