

Problem Solving Handout 1: Protein helix to coil transition

Dr Benjamin J. Morgan¹ and Dr Andrew R. McCluskey^{1,2}

¹*Department of Chemistry, University of Bath, email: b.j.morgan@bath.ac.uk*

²*Diamond Light Source, email: andrew.mccluskey@diamond.ac.uk*

January 10, 2020

The focus of this exercise is to give you experience of applying programming to an abstract chemical problem. This exercise will involve a substantial amount of problem solving, and will be similar in style to the examination at the end of the module. This handout provides necessary background to help you solve the problem that is detailed in the Problem Solving Exercise 1 sheet.

Read through this material completely before the session on X-DATE-GOES-HERE-X.

1 Background

Proteins are polymers molecules of biochemical importance, the specific biochemistry of a particular protein is heavily influenced by their three dimensional structure. The particular monomers which make up protein molecules are known as amino acids. These individual amino acids are able to form secondary structures through complex hydrogen bonding networks, these are secondary to the covalently bonded amino acids. The most common secondary structures are α -helices and β -sheets, examples of these are shown in Figure 1. Under certain conditions, these secondary structures can undergo collapse into a random coil. The unwinding of an α -helix into a random coil is a **cooperative transition**, in which the helix becomes increasingly susceptible to structure changes once the process has begun. Statistical thermodynamics can be used to develop a model that describes the probability of this transition while accounting for the cooperatively of the occurrence.

2 Small protein

In order to evaluate the fraction of amino acids present in a random coil or an α -helix it is necessary to set up the partition function for the various states of

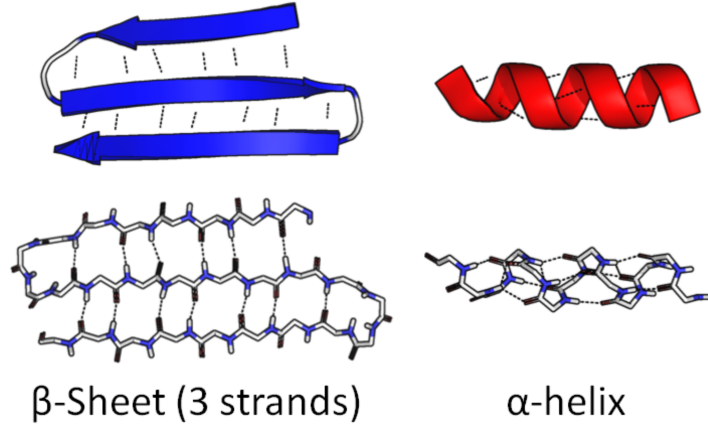


Figure 1: Examples of the secondary structures formed by amino acids, the hydrogen bonding is represented by dashed black lines. Figure by Thomas Shafee, distributed under a CC BY-SA 4.0 license.

the molecule. Considering first a short protein consisting of four amino acids, each of which can either be in a coil, c , or a helix, h . When the configuration is $hhhh$, the partition coefficient will be termed, q_0 , while if the configuration is $cccc$ the partition coefficient will be termed q_4 . The total partition function (q) is therefore,

$$q = C(4,0)q_0 + C(4,1)q_1 + C(4,2)q_2 + C(4,3)q_3 + C(4,4)q_4, \quad (1)$$

where, $C(n,i)$ gives the number of permutations of i random coil amino acids in n total amino acids. This can be found with the following relationship,

$$C(n,i) = \frac{n!}{(n-i)!i!}, \quad (2)$$

where, ! indicates that the *factorial* is being taken. Equation 1 may be rewritten as,

$$q = q_0 \left(C(4,0) + \frac{C(4,1)q_1}{q_0} + \frac{C(4,2)q_2}{q_0} + \frac{C(4,3)q_3}{q_0} + \frac{C(4,4)q_4}{q_0} \right), \quad (3)$$

from which, we can suppose that each partition function differs from q_0 only by the energy of each conformation relative to $hhhh$, therefore,

$$\frac{q_i}{q_0} = \exp \left(\frac{-(\varepsilon_i - \varepsilon_0)}{kT} \right), \quad (4)$$

where, $\varepsilon_i - \varepsilon_0$ is the transition energy from 0 to i , k is the Boltzmann constant, and T is the temperature.

3 Non-cooperative model

In order to generalise this for any number of amino acids, we will first consider the case where the α -helix to random coil transitions are non-cooperative. This

means that the energy associated with changing one h amino acid to a c amino acid is the same, regardless of how many h or c amino acids are already present. The energy change between $c^i h^{4-i}$ and $c^{i+1} h^{3-i}$ is the same, γ for any value of i , which implies that $\varepsilon_i - \varepsilon_0 = i\gamma$. By bringing together Equations 3, 4, and the above assumption, we can write the total partition function as,

$$q = q_0 \left[C(4, 0) + C(4, 1) \exp\left(\frac{-\gamma}{kT}\right) + C(4, 2) \exp\left(\frac{-2\gamma}{kT}\right) + C(4, 3) \exp\left(\frac{-3\gamma}{kT}\right) + C(4, 4) \exp\left(\frac{-4\gamma}{kT}\right) \right], \quad (5)$$

$$q = q_0(C(4, 0) + C(4, 1)s + C(4, 2)s^2 + C(4, 3)s^3 + C(4, 4)s^4),$$

where,

$$s = \exp\left(\frac{-\gamma}{kT}\right), \quad (6)$$

which is known as the stability parameter. The term within the parenthesis on the last line of Equation 5 has the form of the binomial expansion of $(1 + s)^4$.

The binomial expansion of $(1 + x)^n$ is,

$$(1 + x)^n = \sum_{i=0}^n C(n, i)x^i. \quad (7)$$

Therefore it is possible to rewrite Equation 5 as,

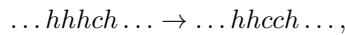
$$\frac{q}{q_0} = \sum_{i=0}^4 C(4, i)s^i. \quad (8)$$

This can be generalised to a protein molecule of any number of amino acids,

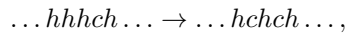
$$\frac{q}{q_0} = \sum_{i=0}^n C(n, i)s^i. \quad (9)$$

4 Cooperative model

The application of a cooperative model is more complex, as it depends on modelling how neighbours facilitates each other's conformational change. The *zipper* model only allows for the conversion of h to c if an amino acid adjacent has already undergone the change. So the following,



is an allowed transition. However,



is not. Of course, for this to be possible, the first conversion must not have this rule applied (or else the collapse would be impossible). Cooperativity is

included in the zipper model by the assuming that the *nucleation step*, the first transition, is less favourable than the remaining conversions. Therefore, for the nucleation step, the stability parameter is multiplied by some parameter σ , where $\sigma \ll 1$. Each subsequent, *propagation* step has a stability parameter of s . This leads to a functional description of the partition function,

$$q = 1 + \sigma(n+1) \sum_{i=1}^n s^i - \sigma \sum_{i=1}^n n s^i \quad (10)$$

The fraction, p_i , of molecules that has i amino acids in the coil conformation can be found as,

$$p_i = \frac{(n-i+1)\sigma s^i}{q}, \quad (11)$$

and the mean value of i is found as $\langle i \rangle = \sum_i i p_i$.