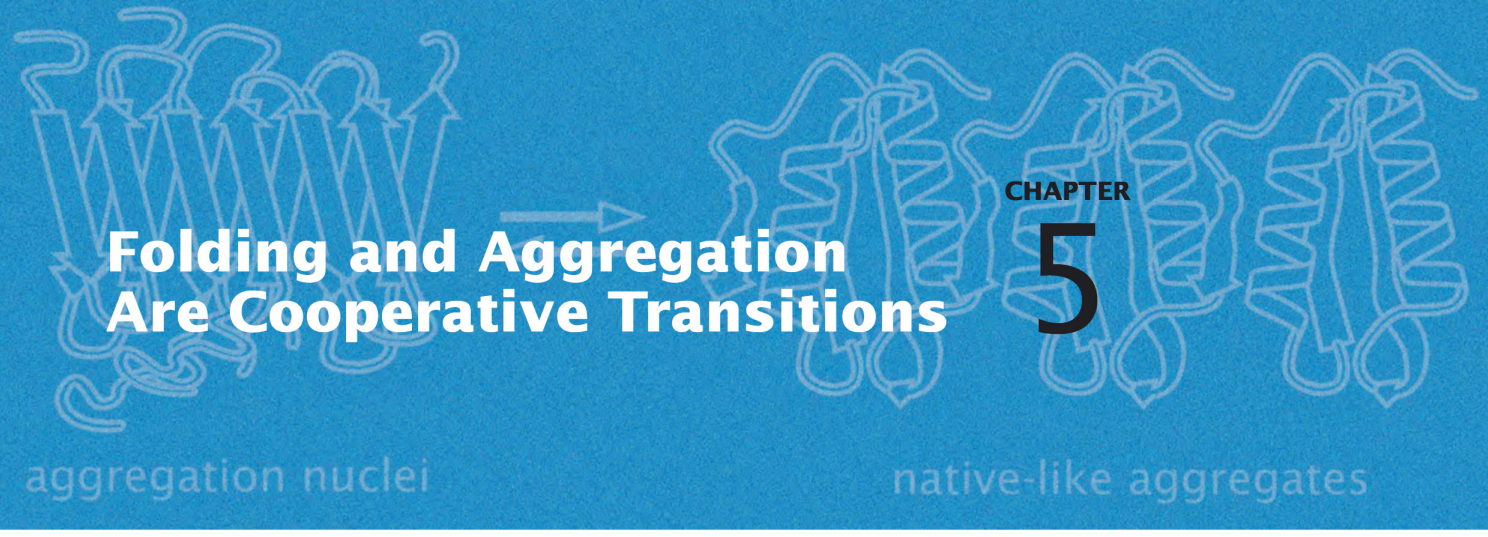

CHAPTER 5

Folding and Aggregation Are Cooperative Transitions

PROTEINS CAN UNDERGO SHARP TRANSITIONS IN THEIR STRUCTURES OR PROPERTIES

Proteins can undergo cooperative changes. In a cooperative change, a small perturbation leads to a large consequence. One example was already described in Chapter 4: the binding of a ligand drives allostery or large conformational changes in a protein. Here, we consider two other examples. One is the folding process. In test tube experiments, a protein can be driven from a denatured state to its native structure sometimes by a very small change in temperature or denaturant concentration. Another example is protein aggregation, where a small increase in a protein's concentration, under the right conditions, can drive individual proteins into large-scale multimolecular association in the form of aggregates, precipitates, crystals, or amyloid fibrils.

What is cooperativity? Consider three different types of change: **(1) A gradual change**, such as when the density of water changes only a little bit when you heat it by a few degrees around room temperature. **(2) A cooperative change**, as when the density of water undergoes a large and sharp change at its boiling point, from liquid density to gas density. Cooperative transitions are also called *two-state* or *all-or-none*. At its boiling point, you will find clusters of water in one of two states: either liquid or steam. **(3) A large continuous "noncooperative" change**, such as near a *critical point*. When water is heated at its critical point, some clusters of water are dense like a liquid, other clusters have low density like steam, and still other clusters have densities that are intermediate between those of water and steam. A noncooperative transition—also called a *higher-order transition* in macroscopic systems—is not two-state. How would you know if a transition has two-state cooperativity or not? Next, we address this question in three steps, for different types of transitions. What are the molecular structures of clusters of molecules? What experiments can tell the difference? And how can you recognize the distinction using energy landscapes of models?

What molecular structures define a cooperative transition? Consider a series of beakers, each containing protein, in which there are

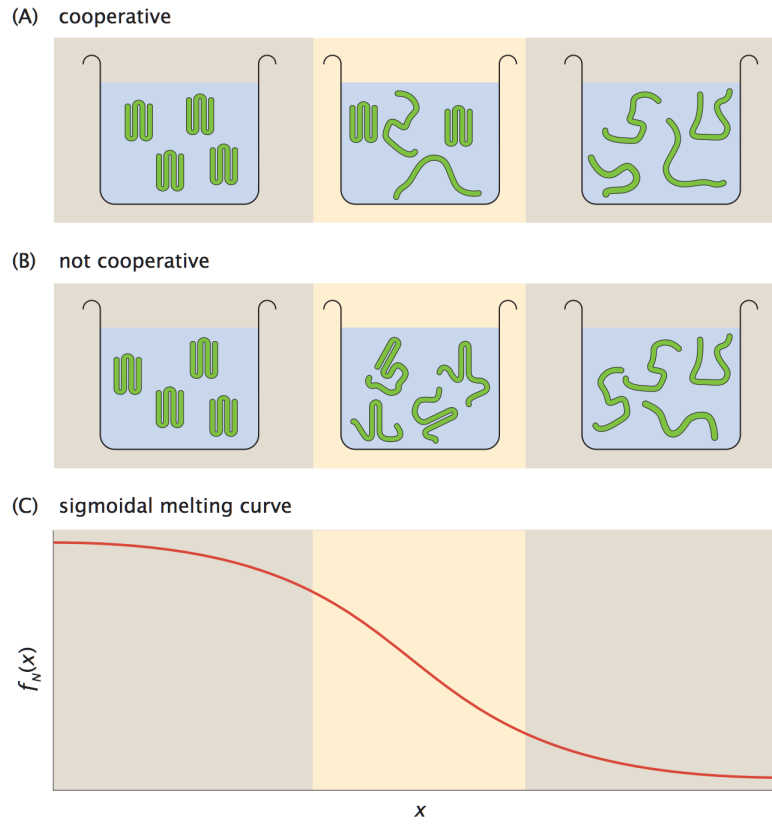

Figure 5.1 Is protein folding cooperative? You can tell by checking for intermediates. (A) From left to right, increasing denaturing agent leads from native to denatured proteins. In a two-state cooperative transition, the middle beaker is a mix of native plus denatured proteins. There are no intermediate structures. (B) In a noncooperative transition, the middle beaker has structures that are intermediate between native and denatured. (C) The denaturation profile (fraction of native protein versus x, denaturant concentration in this case) cannot distinguish between (A) and (B), because it is an average over all the molecules in each beaker.

increasing concentrations of denaturant from one beaker to the next (Figure 5.1). The first beaker has no denaturant, the second beaker has a small amount of denaturant, and the third beaker has a lot of denaturant. In the no-denaturant beaker, all the protein molecules are folded into their native structures. In the high-denaturant beaker, all the protein molecules are denatured. The middle beaker has an intermediate denaturant concentration that is at the midpoint of the denaturation transition. What is the state of protein folding in the middle beaker? Figure 5.1A and B show two possibilities: (A) Half of the protein molecules have fully native structures and half of the proteins are fully denatured. Very few of the molecules have structures that are intermediate between native (N) and denatured (D). This transition is *cooperative* or *two-state*. It is characterized by having intermediate populations that are zero or small. (B) If, instead, at the denaturation midpoint, you see a significant population of protein molecules having an intermediate structure between native and denatured, then the transition is called noncooperative or *multistate* because you would need to invoke more than just the two states (N and D) to model it. You call the third state (and others) *intermediate states*. The existence of intermediates—or not—defines the nature of the transition.

What experiments would tell you if a transition is cooperative or noncooperative? Just observing a sigmoidal shape in a denaturation curve is not sufficient. A denaturation curve just tells you the *average* state of the system in each beaker (for example, the average fraction of native contacts, without distinguishing whether any two contacts occur within one chain or in two different chains). Figure 5.1C

indicates that such an averaged denaturation curve may have the same sigmoidal shape no matter whether the underlying transition is cooperative or noncooperative. Rather, to learn the nature of the transition, you need to measure the individual subpopulations themselves. To know whether your system is cooperative or not, you need more than just the *population-average* state of the system at a given value of denaturing agent x; you need to know the *distribution of populations*. To prove whether or not your system is cooperative requires some direct measurement of the distinct populations, in this case N, D, and intermediates (I) in each of the beakers. Methods such as mass spectrometry can often distinguish among some subpopulations.

Stable and Unstable States Are Represented on Energy Landscapes

The nature of a transition is evident from an *energy landscape*. To describe landscapes, we first describe *order parameters*. For protein folding, let's define a variable ξ. An order parameter ξ is simply a progress coordinate that tells you the state of the system, in this case from fully denatured to fully folded. $\xi = 0$ means that all the proteins are fully denatured and $\xi = 1$ means that all the proteins are fully folded. An example of an order parameter for folding is the chain's radius of gyration. A native state is compact (has small radius), and a denatured state is expanded (large radius). So, the radius of gyration distinguishes the two states of interest. Or your order parameter might be an experimental quantity that distinguishes the native from denatured populations. A commonly used order parameter is the fraction of contacts in the chain that are native-like. A contact is a close interaction, typically closer than about 4.5 Å, between any pair of atoms belonging to two residues.

Figure 5.2 shows the general relationship between the population $p(\xi)$ of molecules in state ξ and its corresponding free energy $\Delta G(\xi)$:

$$p(\xi) \propto e^{-\Delta G(\xi)/RT}. \tag{5.1}$$

States (ξ) that have the largest populations have the most negative free energies.

Figure 5.3 shows the shapes of two different energy landscapes: (A) a cooperative, two-state transition and (B) a gradual change. The boiling of water is a cooperative process. At exactly the midpoint of boiling, half the water molecules are in low-density steam and half are in high-density liquid water. This is manifested as two dominant populations, or, equivalently, two minima in the free energy. Interestingly, the folding of many small single-domain proteins is cooperative: at the folding midpoint, the molecules in the beaker are either fully folded or unfolded, and very few of the protein molecules are in some intermediate state of folding; see Figure 5.3A. In contrast, Figure 5.3B shows a conceptual alternative of what would be expected if protein folding were a gradual change. This would have a single population, and thus a single minimum in free energy. Here we explore physical models of cooperative and noncooperative and gradual changes in proteins, starting from a simple, classical model, for helix–coil transition.

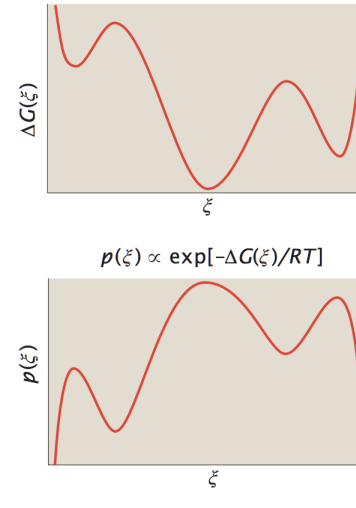

Figure 5.2 Relating populations $p(\xi)$ to free energies, $\Delta G(\xi)$.

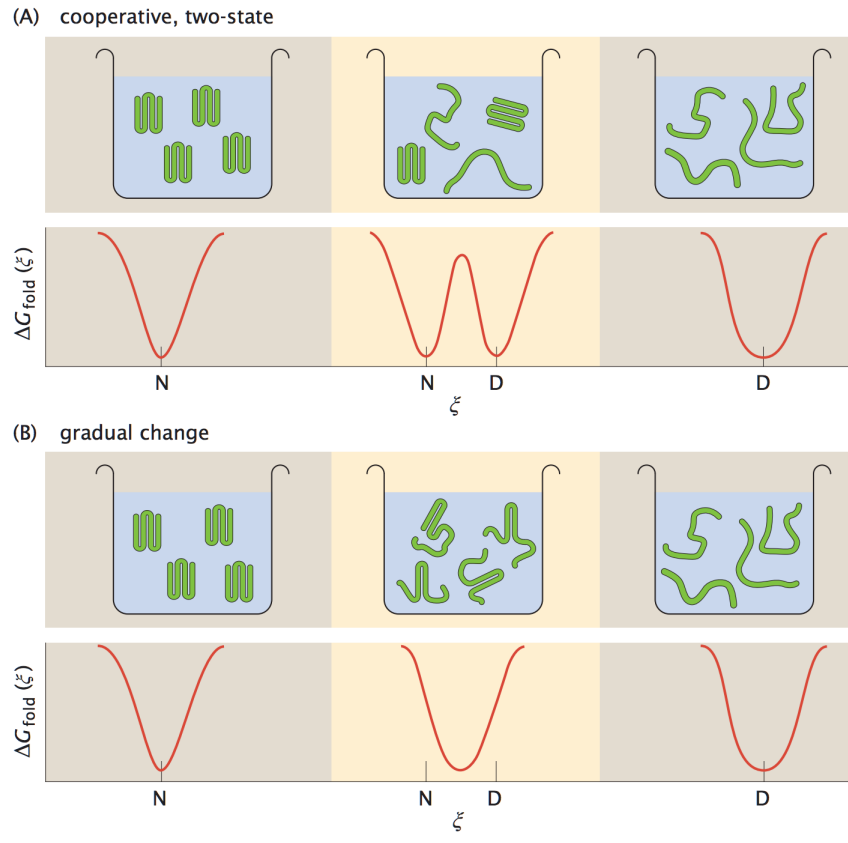

Figure 5.3 How can you tell if a model transition is cooperative? Its energy landscape has a barrier. (A) A two-state transition has multiple free-energy minima and a barrier between those states. (B) For comparison, a gradual change has a free-energy minimum that shifts continually from N, to midway to D, to D, with changing external conditions.

PROTEINS AND PEPTIDES CAN UNDERGO A COOPERATIVE HELIX–COIL TRANSITION

Before considering protein folding, let's look at a simple type of large change called the *helix–coil transition.* Some types of polymers have two experimentally distinguishable states: a *coil* state, which is a large ensemble of disordered conformations, and a *helical* state composed essentially of a single conformation in which the chain spirals in a helix shape (Figure 5.4). Changing the solvent conditions or the temperature can cause such polymers to undergo a sharp transition between the coil state and the helix. What drives this conformational change?

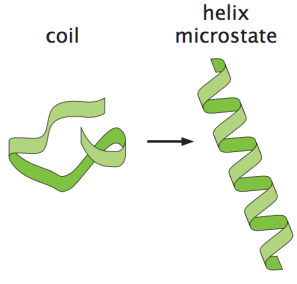

Figure 5.4 Some proteins undergo a transition from coil to helix. A *coil* state is a large ensemble of disordered conformations (many microstates). The helix is a relatively unique conformation (single microstate).

Let's express the conformation of a polypeptide as a one-dimensional binary string of characters: "c" for every unit (for example every residue) that is in a coil conformation and "h" for every residue in a helical configuration. For example, one particular conformation of a 16-mer chain can be expressed as

$$\text{ccchhhhhhhcccccc.} \tag{5.2}$$

Now, our goal is to make a physical model of the helix–coil transition from which we can compute experimentally observable properties. Often, it is easiest to measure properties that are averages over all the molecules in a container. Our aim is to compute averages over probability distributions. Different molecules will have $n = 1, 2, 3, \ldots$ helical units. We want to know the average helicity $\langle n \rangle$, where *average*

is defined to be

$$\langle n \rangle = \sum_{n=0}^{N} np(n), \tag{5.3}$$

for a chain having N total units. Now, in order to compute $\langle n \rangle$, we need to know the probability distribution $p(n)$. It is given by

$$p(n) = \frac{\omega(n)}{\sum_{n=0}^{N} \omega(n)}, \tag{5.4}$$

where $\omega(n)$ is the *statistical weight* that accounts for the population of chains that have n units in a helical conformation. A statistical weight is a relative population; it does not have to be normalized to sum to one, as probabilities must. Equation 5.4 shows how probabilities (normalized quantities) are related to statistical weights. A crucial quantity in the denominator of Equation 5.4 is the *partition function* Q,

$$Q = \sum_{n=0}^{N} \omega(n), \tag{5.5}$$

which is the sum of statistical weights over all the possible states of the system, where *state n* in this case refers to the set of conformations having n helical turns. So, our objective in modeling is to find a way to compute $\omega(n)$, the statistical weight for state n of the system.

Now, consider the three levels of statistical weights: (1) We need the statistical weight *for each* h *or* c *residue* in the chain. (2) We need the statistical weight *for each microstate of a whole chain*, that is, for one particular sequence of h and c units. (3) We need to sum the statistical weights over *whatever microstates compose the chain macrostate* of interest. Macrostates usually refer to some experimentally observable state. For example, a denatured state is a macrostate collection of microstates.

(1) Statistical weights for individual h or c units. To begin, we define the statistical weight of each single coil unit, c, to be 1. We are free to choose this one statistical weight arbitrarily because only ratios of statistical weights matter for computing probabilities. Next, we assign a statistical weight of s to each helical residue, h. You can think of s as the equilibrium constant for converting a c unit to an h unit. s depends on the chemical nature of the unit and on the chemical and thermal environment of the chain. The environmental conditions of solvent and temperature sometimes cause an amino acid to prefer the helical conformation. If $s > 1$, it means that you are considering a situation in which helix is more favorable than coil. Or, if the conditions lead to $s < 1$, it means the residue favors the coil state. $s = 1$ means that h and c are equally populated. In reality, different amino acids will have different values of s, but for our purposes here of capturing the essential ideas in the simplest possible model, we take all amino acids to have the same values of s.

(2) Statistical weights for whole chains of combined h and c units. To construct Q, you can reason with the rules of probability, as described in Box 5.1.

Box 5.1 A Useful Aside about Probabilities and Averaging

Recall the two main rules of probability: (1) If states A and B are mutually exclusive and if you want to compute the statistical weight for seeing either state A OR B, then you add: $\omega(\text{A OR B}) = \omega(\text{A}) + \omega(\text{B})$. (2) If states A and B are independent, and if you want to compute the statistical weight for seeing both states A AND B, then you multiply: $\omega(\text{A AND B}) = \omega(\text{A})\omega(\text{B})$. We will add or multiply statistical weights and probabilities accordingly.

Here is a math shortcut that is useful for computing averages. We will have partition functions that take the form

$$Q = 1 + s + s^2 + s^3 + \ldots \tag{5.6}$$

In such expressions, the term s^n is the statistical weight of a helix having n turns, and the corresponding probability is

$$p(n) = \frac{s^n}{Q} \tag{5.7}$$

Suppose you want to compute the average number of helical turns, $\langle n \rangle$. Then combining Equations 5.6 and 5.7 with Equation 5.3 gives

$$\langle n \rangle = \sum_n np(n) = \frac{s + 2s^2 + 3s^3 + \ldots}{1 + s + s^2 + s^3 + \ldots}$$

$$= \frac{s}{Q}\frac{dQ}{ds} = \frac{d\ln Q}{d\ln s}. \tag{5.8}$$

The second line of Equation 5.8 shows that this average can also be obtained by taking a derivative of $Q(s)$ (because $sds^n/ds = ns^n$ and because $d\ln Q = (1/Q)dQ$.

Now we apply this reasoning to perform calculations for different models. First, in order to see the nature of cooperativity, let's start with a model that, by definition, has no cooperativity.

A first model to try: independent units

Consider a model in which each c and h unit is independent of the others. In that case, the partition function over all the microstates will be

$$Q = (1 + s)^N, \tag{5.9}$$

because each unit in a chain can be either c or h, and these are mutually exclusive options for a unit. This accounts for the term $(1 + s)$. Then, you raise $(1 + s)$ to the power N because there are N units in the chain, and because you are seeking the statistical weight that unit 1 and unit 2 and unit 3 ... are each in a particular state. Substituting Equation 5.9 into Equation 5.8 for average fractional helicity gives

$$\frac{\langle n \rangle}{N} = \frac{s}{NQ}\frac{dQ}{ds} = \frac{s(1 + s)^{N-1}}{(1 + s)^N} = \frac{s}{1 + s}. \tag{5.10}$$

Now, suppose you perform a series of experiments in which the solvent or temperature are systematically changed. Each experiment corresponds to a different value of s, the helical propensity. Equation 5.10

for this independent-units model predicts that the average helicity changes gradually as a function of s, not sigmoidally (Figure 5.5).

This independent-units model does not predict a sigmoidal shape for $\langle n \rangle / N$ versus s and does not predict that this function (helicity per unit) gets steeper with N, both of which are observed in experiments. So, next we consider a model that better captures helix–coil cooperativity.

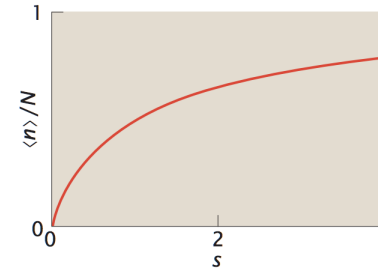

Figure 5.5 If units are independent, the average helicity changes only gradually with helical propensity s. In the independent-units model, if s were caused to vary by changing temperature or solvent, the average helicity $\langle n \rangle / N$ versus s would change only gradually. And the slope of this curve of helicity per monomer would not depend on chain length N (because monomers are independent). But experiments show that transitions become sharper as N gets larger.

A second model: two-state model

Now, let's try a different model. In this model, the chain is either all coil (cccc…c), or all helix (hhhh…h), and no other microstate is populated. This model is maximally cooperative, having only two states. If any one unit is in the h state, all units are in that state. Its partition function is

$$Q = 1 + s^N. \tag{5.11}$$

The reasoning here is that you have either all coil, with statistical weight 1^N OR all helix, with weight s^N, and these two chain states are mutually exclusive. Substituting Equation 5.11 into the helicity equation, Equation 5.8, gives

$$\frac{\langle n \rangle}{N} = \frac{s^N}{1 + s^N}. \tag{5.12}$$

This model predicts that $\langle n \rangle / N$ versus s is a sigmoidal function and gets steeper with increasing N (Figure 5.6). These features are qualitatively correct, but quantitatively, this transition is too steep. In experiments, such steepness is found only for chains having $N > 1000$.

So, now consider a third model, due originally to John Schellman, that is intermediate between these two extremes. In the following sections, we use the term "two-state" not just in the previous sense of zero intermediate population, but also for situations where the intermediate populations are relatively small. In the Schellman model, the formation of a helix involves two aspects: initiation and propagation.

The Schellman Model Describes the Helix–Coil Transition as Nucleation Followed by Growth

The Schellman model [1] defines a helix as a stretch of uninterrupted h's. To keep the math simple, suppose that there is, at most, one helical segment (of any length n) in the chain. This *single-helix approximation* is often valid for peptides in solution that are shorter than about 20 amino acids long. Again, the statistical weight of c is 1 and that of h is s. But we now introduce another parameter σ. Whereas s is a *propagation parameter*, for converting one c to one h, σ is a *nucleation parameter*, the equilibrium constant for initiating a helix. That is, every "first" h in a helix is assigned a statistical weight σs. Every subsequent h in a helix is assigned a weight s. Box 5.2 shows how you assign statistical weights to the different chain configurations.

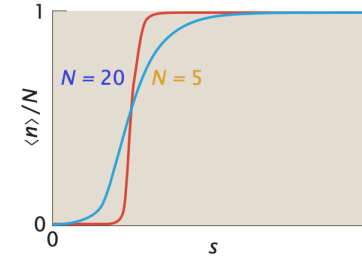

Figure 5.6 The two-state model predicts a sharp increase in helicity with s. But the predicted dependence on N is much steeper than seen in experiments.

Box 5.2 Examples of Statistical Weights

Chain configuration	Statistical weight
ccccccccc	1
cccccchccc	σs
cccchhccc	σs^2
chhcchhcc	$\sigma^2 s^4$
chhhhhhhh	σs^8

You can compute the statistical weight ω_n of all the conformations having a single helix containing n helical amino acids in a row, anywhere in the chain, as

$$\omega_n = (N - n + 1)\sigma s^n, \tag{5.13}$$

relative to a chain that is configured with all its residues in the coil state. The factor of $(N - n + 1)$ in Equation 5.13 counts the number of possible starting locations where an n-mer stretch of helical units ($n > 0$) can begin in the N-mer chain.

Now, to convert ω_n to a probability $p(n)$, you divide by the sum over all helix lengths, Q_1 (where the subscript 1 indicates a single helix, to distinguish from helix-bundle models below):

$$\begin{aligned} Q_1 &= 1 + \sum_{n=1}^{N} \omega_n \\ &= 1 + \sigma N s + \sigma(N-1)s^2 + \sigma(N-2)s^3 + \ldots + \sigma s^N, \end{aligned} \tag{5.14}$$

where the leading term, 1, is the statistical weight for the coil state. So, the probability that a chain has n helical residues is

$$p(n) = \frac{\omega_n}{Q_1} = \frac{(N - n + 1)\sigma s^n}{Q_1}, \tag{5.15}$$

and the probability of having an all-coil chain is $p(0) = 1/Q_1$.

You can compute the average helicity as

$$\langle n \rangle = Q_1^{-1}[N\sigma s + 2(N-1)\sigma s^2 + 3(N-2)\sigma s^3 + \cdots + N\sigma s^N]. \tag{5.16}$$

Here is how you use the Schellman helix–coil model. You are given the maximum helix length N, the nucleation parameter σ, and the propagation parameter s. Now use Equation 5.14 to compute the partition function Q_1. Then use Equation 5.15 to compute any population $p(n)$ of interest to you, including the population of the fully helical molecule, $p(N)$. To compare with experiments, you want the average helicity $\langle n \rangle$ from Equation 5.16. (Using standard statistical mechanical expressions, you can also take a second derivative of Equation 5.16 to compute the variance in helix length.)

For certain values of the parameters σ and s, the Schellman model predicts two-state cooperativity. Figure 5.7 shows the distributions of helical lengths for three different values of s. In this example, the intermediate states i are much less populated than either the all-coil state

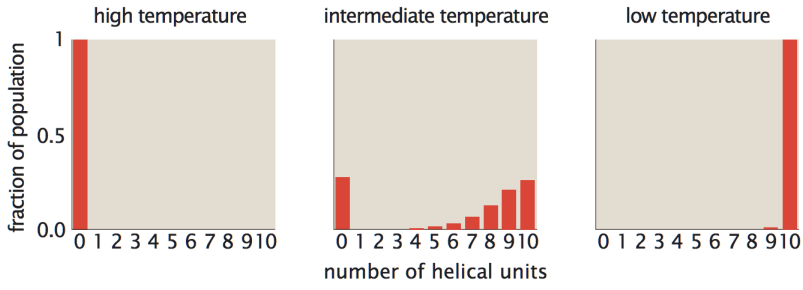

Figure 5.7 Example of a helix–coil transition that is cooperative, having limited populations of intermediate states. Calculations using Equation 5.15 are for $N = 10$, $\sigma = 10^{-4}$, and various values of s. The chain is all-coil under conditions of high temperature ($s = 0.02$) and all-helix at low temperature ($s = 200$). At a middle temperature ($s = 2.5$) (for which Equation 5.14 gives $Q_1 = 3.64$, for example) all the intermediate states are less populated than the coil ($i = 0$) or helical ($i = 10$) states.

or the all-helical state, that is, $p_1 > p_i < p_N$ at the transition midpoint. Two-state behavior occurs when nucleation is difficult (that is, when $\sigma \ll 1$), and when propagation is favored (that is, when $s > 1$). In this case, the coil is favored by its large conformational entropy. The helix is favored by its low energy, due to its multiple hydrogen-bonded units along the chain. The states intermediate between full helix and full coil are less favorable on both counts. In short, when nucleation is difficult, you will find either coil molecules or long helices, but not short helices. It is entropically costly to form the first helical turn, but it becomes energetically favorable to do so if the chain forms a sufficiently large number of helical turns.

Helices Can Be Denatured by Heating

In the last section, we expressed the helix–coil equilibrium as a function of the two parameters σ and s. However, often we will be interested instead in how the helix–coil populations depend on experimentally controllable variables such as temperature or denaturant concentration. The helix–coil equilibrium depends on temperature through $s = s(T)$ and $\sigma = \sigma(T)$. You can express the temperature dependence of the equilibrium coefficients σ and s in terms of their corresponding enthalpy ΔH and entropy ΔS. Doing so gives useful insights into the microscopic physical bases for σ and s. $s(T)$ can be expressed in terms of a free energy $\Delta G(T) = -RT \ln K = \Delta H - T\Delta S$ for the formation of one helical turn:

$$s(T) = e^{-\Delta G/RT} = e^{-\Delta H/RT + \Delta S/R}, \tag{5.17}$$

where ΔH ($= \epsilon_{hb}$, where $\epsilon_{hb} < 0$) is the enthalpy decrease upon forming one helical contact relative to the coil conformation, by forming a hydrogen bond. $\Delta S(T) = -R\ln(z-1)$ is the entropy decrease for extending the helix by one more residue, assuming that each residue has access to z local conformational (or isomeric) states (for example, we could make a simple estimate of $z = 3$ for helical, β-strand, and coiled states). In other words, the entropy of a nonhelical amino acid is $S = R\ln(z-1)$, while that of a helical amino acid is $S = R\ln 1 = 0$, hence the change $\Delta S = -R\ln(z-1)$ accompanying the transition from coil to helix, for each amino acid. R is the gas constant and T is the absolute temperature. The enthalpy change for extending the helix by one residue is favorable (negative) and the entropy change is unfavorable (negative). So, the temperature dependence of the propagation parameter is given by

$$s(T) = \frac{e^{-\epsilon_{hb}/RT}}{z-1}. \tag{5.18}$$

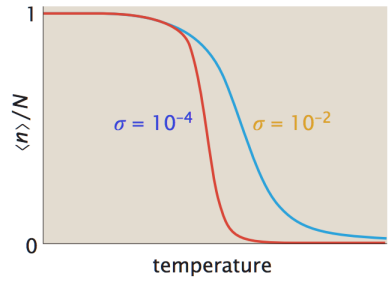

Figure 5.8 The Schellman helix–coil model predicts a sharp thermal transition. This calculation is for a model of $N = 10$ helical bonds, assuming $z = 5$ possible directions for each bond, an energy of $\epsilon_{hb}/R = -500\,K$ for each helical bond, and two different values of σ. The average number of helical residues, $\langle n \rangle$, is calculated from Equation 5.16. The smaller the value of σ, the sharper the transition.

Similarly, nucleation can also be expressed in terms of an enthalpy ϵ_{nuc}:

$$\sigma(T) = \frac{e^{-\epsilon_{nuc}/RT}}{(z-1)^3}. \tag{5.19}$$

Equations 5.18 and 5.19 express $\sigma(T)$ and $s(T)$ in terms of a model of energy and entropy components. These local terms describe how nucleation is entropically disfavored relative to propagation. The cost of nucleation involves the entropy of fixing the conformation in order to form a hydrogen bond between the C=O group of amino acid i and the N–H group of amino acid $i+4$. The first of these four residues can be oriented in any direction; then the next three are fixed, hence the factor $(z-1)^3$. In contrast, the cost of propagation involves the entropy of fixing only one residue. Because of the simplicity of our model, these factors are only approximate, but they capture the physics that nucleation is entropically less favorable than propagation.

Figure 5.8 shows that $\langle n \rangle$ is a sigmoidal function of temperature T: the helix melts out with increasing temperature. It also shows that the transition becomes sharper as σ becomes smaller (that is, when the energetic cost of nucleation is very high).

Figure 5.9 shows that this simple helix–coil theory fits experimental measurements of average helicity as a function of temperature or as a function of denaturant (urea) concentration.

What values of σ and s should you use? Different amino acids have different helical propensities, s. (The value of σ, which ranges from 10^{-2} to 10^{-4} in different models, is thought to be not very dependent on amino acid type). Table 5.1 gives one compilation of experimental values. In general, it is found that alanine has a high propensity to form a helix, while proline is a helix breaker, for example. Sometimes, to explore matters of principle, you use a single value for σ and a single value for s for all amino acids, as we have done earlier. Other times, you may prefer to account for the different s values of the different individual amino acids within a peptide or to go beyond our single-helix approximation. For advanced helix–coil modeling, see Appendix 5A.

PROTEIN FOLDING COOPERATIVITY ARISES FROM SECONDARY AND TERTIARY INTERACTIONS

Now, how should we understand cooperativity in a more complex process such as protein folding? Small proteins are found to fold

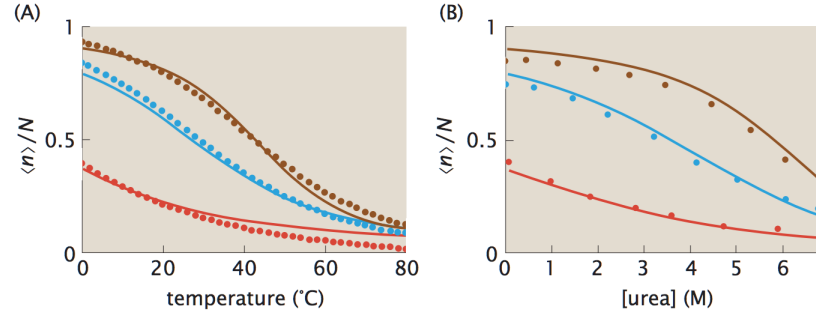

Figure 5.9 The helix–coil model can model helix denaturation by temperature or urea, for chain lengths $N = 50$ (*brown*), $N = 26$ (*blue*), and $N = 14$ (*red*): (A) thermal denaturation data; (B) urea denaturation. For computational details, see [2]. (A, adapted from JM Scholtz, H Qian, EJ York, et al. *Biopolymers*, 31:1463–1470, 1991; B, adapted from JM Scholtz, D Barrick, EJ York, et al. *Proc Natl Acad Sci U S A*, 92:185–186, 1995. Copyright (1995) National Academy of Sciences, USA.)

with two-state cooperativity, that is, with negligible populations of intermediates. Helix–coil theories alone are not sufficient to explain protein-folding cooperativity. For one thing, β-sheet proteins fold cooperatively too. For another thing, most helices in folded proteins are short, yet helix–coil theory says that short peptides should not form stable helices. What are we missing?

Protein collapse theories alone are also not sufficient to explain protein-folding cooperativity. The hydrophobic residues in a protein cause the protein to collapse into a compact structure in water, and this collapse process is abrupt. However, lattice-model studies show that polymer collapse leads to noncooperative transitions, not cooperative ones. Figure 3.9 in Chapter 3 shows that the 6-mer HP model has a significant population of intermediate structures at the midpoint of the denaturation transition. This lack of cooperativity is not because 6-mer chains are too short—longer chains, too, undergo noncooperative collapse.

So, if we can't explain the two-state nature of protein folding by either helix–coil processes alone or collapse alone, what does explain it? Folding cooperativity appears to be due to a combination of secondary and tertiary interactions [3]. When two helices are forming in protein folding, each helix forms with some cooperativity on its own, but, in addition, the packing of the two helices next to each other helps stabilize the two-helix pairing even more. A protein helix is commonly *amphipathic*, meaning that it has a stripe of hydrophobic residues along one side. The hydrophobic stripes of adjacent helices often face each other when those secondary structures pack together.

To illustrate how tertiary interactions between secondary structures can contribute stability and cooperativity, let's focus on helix-bundle proteins. In native helix-bundle proteins, multiple helices (usually three or more) are aligned and packed against each other, side by side, like a bundle of rods. As the helices form cooperatively, they also bundle together, contributing even more stability. Helices help each other to form. Here is a simple model that illustrates the idea.

Table 5.1 Helical propensities of amino acids

Residue	*s*
Ala	1.54
Arg^+	1.10
Leu	0.92
Lys^+	0.78
Glu^0	0.63
Met	0.60
Gln	0.53
Glu^-	0.43
Ile	0.42
Tyr	0.37–0.50
His^0	0.36
Ser	0.36
Cys	0.33
Asn	0.29
Asp^-	0.29
Asp^0	0.29
Trp	0.29–0.36
Phe	0.28
Val	0.22
Thr	0.13
His^+	0.06
Gly	0.05
Pro	≈0.001

Data measured at 273 K (from A Chakrabartty, T Kortemme, and RL Baldwin. *Protein Sci*, 3:843–852, 1994.)

Helix–Helix Interactions Contribute to Folding Cooperativity in Helix-Bundle Proteins

First, let's model a two-helix bundle assembly. Then, we will model a three-helix bundle protein. The two-helix chain has N monomers. Figure 5.10 shows the most important conformations for our two-helix model. We divide the chain conformations into classes: **(i) Coil state:** The whole chain is fully denatured, having no structure. **(ii) Single-helix state:** The chain is Schellman-like: it can contain a single stretch of helix anywhere and of any length (up to the maximum of N residues long). To keep the math simple, let's approximate each helical turn as having four monomers (rather than 3.6 per turn in α-helices). **(iii) Two adjacent helices are zipped up, but only partially.** The chain has two helices side by side. The two helices have exactly the same number $m = 1, 2, 3, \ldots, M$ of helical turns. We neglect unmatched helix lengths or helices in wrong locations, because their populations will be much smaller. Connecting the two helices is a loop of coil segments. (So, we must have $M < N/2$.) The nonhelical remainder of each chain is coil. **(iv) Native state: Two adjacent helices are fully zipped up together, $m = M$.**

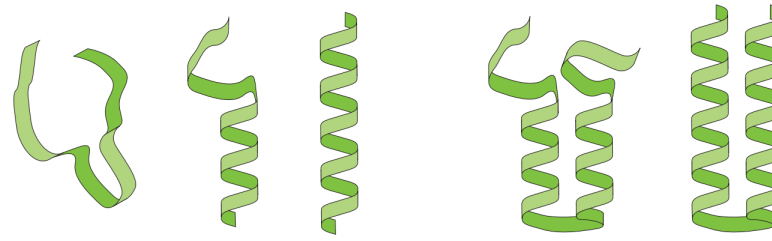

Figure 5.10 Classifying the two-helix-bundle conformations. (A) Three states of the single helix–coil state: No helix (that is, random coil), partially zipped-up helix, and fully zipped helix. (B) Two states of the two-helix-bundle states: partially formed helices having $m = 1, 2, 3, \ldots$ turns each, or native fully formed two-helix bundle having $m = M$ turns of each helix.

Now, we want to calculate the populations of the various protein conformations. Let's construct the statistical weights for these states. First, notice that we have already modeled all the states involving any single helix or coil: the set of conformations on the left side of Figure 5.10. The partition function for these states is just Q_1, given by Equation 5.14 for a chain of length N.

Second, let's count the two-helix-bundle states, shown on the right side of Figure 5.10. For a given zipping state m of the two-helix bundle, the statistical weight is

$$\omega_m = \sigma^2 s^{8m} r^m. \tag{5.20}$$

Here, σ^2 accounts for nucleating both helices, s is the statistical weight for the formation of each helical turn (giving a total of $8m = (4$ monomers per turn of each helix) $\times$ (2 helices) $\times m$ turns in each helix). And r (> 1) is an equilibrium constant that expresses how much extra stabilization results from each one of the m direct contacts between the adjacent helices. Helix–helix contacts are often hydrophobic and packing interactions. Equation 5.20 is the counterpart of Equation 5.13 for the single-helical chain.

The contribution to the overall partition function from the two-helix bundle conformers of various lengths m (the counterpart of Equation 5.14) is

$$\begin{aligned} Q_2 &= \sigma^2 \sum_{m=1}^{M} s^{8m} r^m \\ &= \sigma^2 \left(s^8 r + s^{16} r^2 + \cdots + s^{8M} r^M \right). \end{aligned} \tag{5.21}$$

To compute the population of any of these states, put the statistical weight of that particular state in the numerator, and the sum $Q_{2\text{hb}} = Q_1 + Q_2$ in the denominator. For example, $p(2) = \sigma^2 s^{16} r^2 / (Q_1 + Q_2)$. Then, you can compute averages, such as the average length of the two-helix bundle, using the definition of average, $\langle m \rangle = \sum_{m=0}^{M} m p(m)$.

You can express r in terms of a free energy ϵ_{hh} of forming a helix–helix contact interaction:

$$r = e^{-\epsilon_{\text{hh}}/RT}. \tag{5.22}$$

When there are no helix–helix interactions, $\epsilon_{\text{hh}} = 0$, resulting in $r = 1$. When helix–helix interactions are stabilizing, $\epsilon_{\text{hh}} < 0$ and you have $r > 1$; this is called *positive cooperativity*. In this model, protein folding cooperativity can arise from both the helix–coil contributions (expressed in terms of σ and s) and the helix–helix attractions, expressed by $r > 1$.

Here is how you use this two-helix-bundle model. You are given N (the maximum number of hydrogen bonds that could be formed if the whole chain were a helix); σ and s (the helix–coil nucleation and propagation parameters); r (the equilibrium constant for each favorable helix–helix interaction); and M, the number of amino acids of the maximal-length helix when the protein is in the two-helix-bundle state. (Or, you can begin with ϵ_{hb}, ϵ_{nuc}, and ϵ_{hh}.) Then, you compute the partition function Q_{2hb}, the populations of the states of interest in the model, and averages and variances. Box 5.3 shows how you can generalize this to handle three-helix-bundle proteins, using Q_1 and Q_2, which you have already calculated.

Box 5.3 Three-helix-bundle proteins

Using the same logic as before, you can model the folding of a three-helix-bundle protein. Suppose the dominant three-helix-bundle states are those in which each of the three helices has the same number of turns, m. And suppose that the helices are perfectly adjacent, so there are $3m$ pairwise helix–helix contacts: between helices 1–2, 2–3, and 1–3. Again, take σ as the nucleation constant for each helix, s as the propagation equilibrium constant, and r as the pairwise helix–helix equilibrium constant. Within this simple model, the partition function for the three-helix-bundle states will be

$$Q_3 = \sigma^3 \sum_{m=1}^{M} s^{12m} r^{3m}, \tag{5.23}$$

where $M(< N/3)$ is the maximum number of turns of each helix in the native structure. To compute the population of any state of the three-helix-bundle protein, put the statistical weight of the state you're interested in into the numerator, and put the sum $Q_1 + 3Q_2 + Q_3$ into the denominator. Figure 5.11 shows how this type of simplified model captures the folding–unfolding cooperativity in a three-helix bundle induced by temperature and urea.

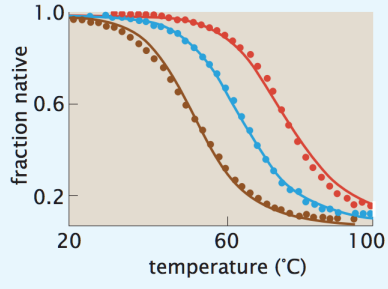

Figure 5.11 The three-helix-bundle model illustrates denaturation by temperature and urea. The thermal denaturation of alpha 3C protein is shown for three concentrations of urea: 2.0 M (*brown*), 1.5 M (*blue*), and 1.0 M (*red*). (Data from JW Bryson, JR Desjarlais, TM Handel, and WF Degrado. *Protein Sci*, 7:1404– 1414, 1998; figure from K Ghosh and KA Dill. *J Am Chem Soc*, 131:2306–2312, 2009. Reprinted with permission from American Chemical Society.)

In summary, the negligible populations of intermediate states that are observed in the folding of small proteins can be explained as the product of two types of cooperativity: the individual helix–coil process is

cooperative, and helix–helix packing interactions are further stabilizing. The helix zipping factor s combines with the helix–helix interaction factor r to cause the two-helix-bundle state to be more stable than any partially zipped or partially packed states. Whenever $r > 1$, it means that helix–helix interactions help stabilize the bundle states. Three-helix bundles are more stable than two-helix bundles in this model because of the factor of r^3 in the three-helix bundles compared with the factor of only r in the two-helix bundles.

This model is quite simplified. For one thing, we have only enumerated here the "dominant states," not all the possible conformations. Some additional classes of conformations could be included, at the cost of extra mathematical complexity. When we model the two-helix states, we have considered only those in which the helices are packed perfectly together, because the r^m factor says that nonbundled pairs of helices will be less populated. And we have left out the combinatoric factor that would allow two-helix-bundle states to form in non-native locations in the sequence, because those terms are small. The point of simple models such as this is to capture general principles, not the fine details.

PROTEINS CAN ASSEMBLE COOPERATIVELY INTO AGGREGATES, FIBRILS, OR CRYSTALS

Now, let's consider another cooperative process. Proteins can often associate into multiprotein assemblies (Figure 5.12) as a sharp function of protein concentration. Examples include the formation of protein complexes or assemblies, or aggregation, precipitation, crystallization, amyloidogenesis, or fibrillization. For instance, proteins can *crystallize* into symmetrical repeating patterns of native molecules packed next to each other. Crystallization is important because it enables determination of protein structures by X-ray diffraction or scattering. Multiprotein assembly also occurs in *inclusion bodies*: in biotechnology, organisms are often engineered to overproduce a particular protein of interest. Overproduction causes the protein to accumulate in high concentration, often driving it to aggregate into structures called inclusion bodies, which are large collections of either misfolded or native proteins. Let's look at some general properties of protein aggregation before we consider a specific model for fibrillization.

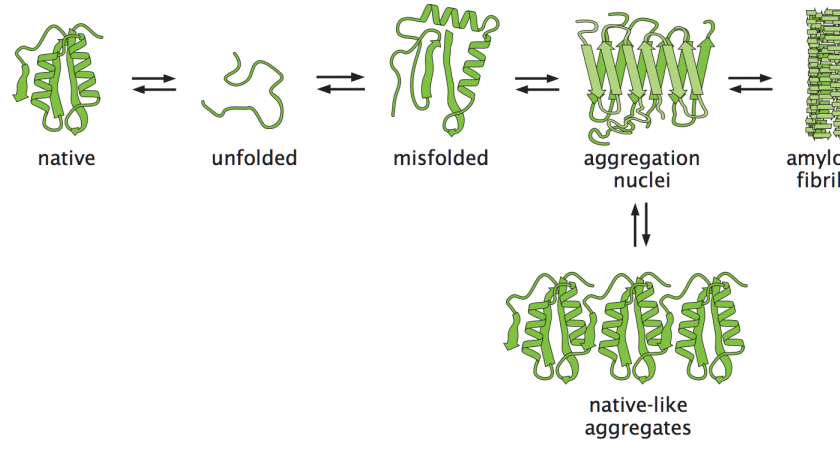

Figure 5.12 Proteins can adopt various states—native, unfolded, misfolded, aggregated, crystalline, or fibrillar, for example.

Attractive Interactions Can Drive Proteins to Aggregate

Protein aggregation processes are not yet well understood. Some are irreversible, so they cannot be studied by equilibrium methods. However, a few guiding principles are known. First, charged proteins repel each other. Proteins have the strongest tendency to associate, aggregate or crystallize when they have no net charge, which occurs when the pH of the solution is equal to the isoelectric point of the protein, called the *pI* (Figure 5.13A) [4].

Second, adding salt favors the aggregation of proteins that have substantial net charge. If two protein molecules have a net charge of the same sign, adding salt shields the charges, weakening the repulsion, and promoting protein processes such as aggregation, fibrillization and crystallization. For example, Figure 5.13B shows that lysozyme solubility decreases as NaCl concentration is increased, meaning that aggregation increases [5].

Third, protein molecules can stick to each other through hydrophobic interactions. Aggregation can occur among proteins that have hydrophobic "sticky" patches, or when interior hydrophobic residues are exposed. For example, heating proteins often unfolds them, at least partially, favoring aggregated states.

Below, we focus on one simple model of just one type of aggregation, namely, fibril formation, chosen because the model is quantitative, agrees with available experiments, and gives a few basic insights.

Amyloid Peptides Can Assemble to Form Fibrils

Some types of proteins, when put into solution at high concentration, will form fibrillar aggregates. A fibril is like dry spaghetti in a package: the many individual chain molecules are stretched out fairly straight, lined up, and packed closely together. As a function of the protein concentration, fibril formation results from a sharp transition: at low concentrations, most of the proteins are free and

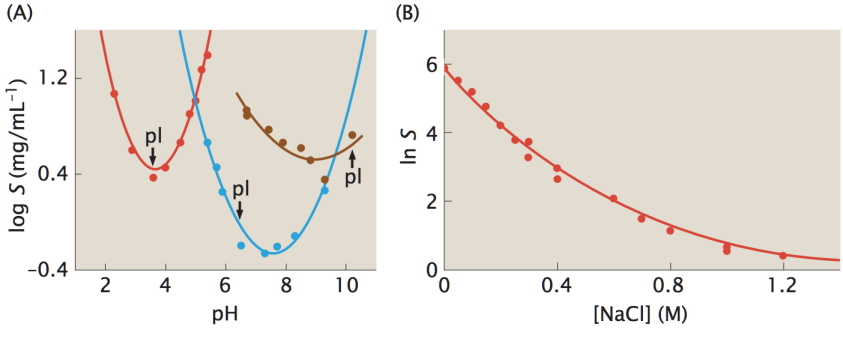

Figure 5.13 A protein's solubility depends on pH and salt concentration. (A) Solubility is minimal near the isoelectric pH. The *red curve* shows the solubility *S* of RNAse Sa (pI = 3.5); the *blue curve* shows the solubility of a mutant of RNAse Sa with pI = 6.4; and the *brown curve* shows a mutant with pI = 10.2. Note that in the case of the wild-type protein, the minimum occurs exactly at pI = pH, where it is slightly shifted in the two mutants, due to other effects (such as dipolar interactions) that perturb the solubility. (B) Proteins are less soluble in high salt concentrations. Adding salt shields net charges on proteins, reducing the repulsions between the proteins, and so reducing protein solubility and facilitating aggregation. This is called *salting out*. In other cases typically not involving charged proteins, increasing salt can increase the solubility, called *salting in*. Data shown are for lysozyme (A, from KL Shaw, GR Grimsley, GI Yakovlev, et al. *Protein Sci*, 10:1206–1215, 2001. B, from E Ruckenstein and IL Shulgin. *Adv Coll Interface Sci*, 123–126:97–103, 2006. With permission from Elsevier.).

independent and monomeric in solution. However, above a particular concentration, the molecules assemble into aggregates. Fibril formation appears to be quite general; it is observed for many different peptides and proteins. Fibrils and soluble oligomers are observed in folding diseases such as Alzheimer's, Parkinson's, Huntington's, and prion diseases. In some cases, a protein that has a normally stable native structure can be perturbed to expose a hydrophobic surface, leading the protein to stick to others, forming aggregates. In other cases, peptides that do not have stable folded structures can aggregate. For example, Aβ is a peptide of 40–42 amino acids that is implicated in Alzheimer's disease. α-synuclein has 140 amino acids, with a highly charged and unstructured 44-residue C-terminus that is implicated in Parkinson's disease. And extended glutamine sequences (for example, more than 40) at the N-terminal segment of the huntingtin protein can lead to aggregation into plaques in Huntington's disease.

Figure 5.14 shows a simple model for the cooperativity of the monomer-to-fibril equilibrium [6]. We are interested in computing the concentration distribution of the different fibril sizes. Let $[A_1]$ represent the concentration of monomeric chains in solution (in units of the number of fibrils per unit volume). $[A_2]$ is the concentration of fibrils containing two protein chains, and $[A_m]$ is the concentration of fibrils containing m protein chains. Let's express our independent variable as $x = [A_1]$ to simplify using standard notation for binding polynomials (see Chapter 4). We now express $[A_m]$ in terms of the concentration x of the isolated chains, as follows.

First, express the concentration of an m-mer fibril as the binding equilibrium of adding one chain to an existing fibril $(m-1)$-mer:

$$K = \frac{[A_m]}{[A_{m-1}]x}, \tag{5.24}$$

where K is the binding constant for adding one protein to a growing fibril. Rearranging Equation 5.24 gives

$$[A_m] = (Kx)[A_{m-1}]. \tag{5.25}$$

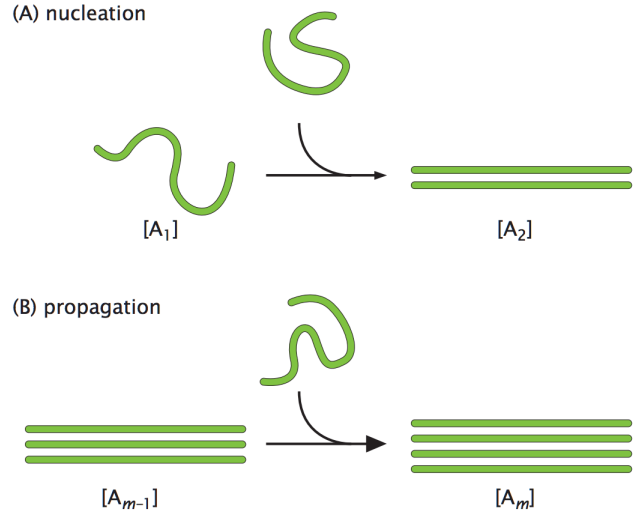

Figure 5.14 Fibril formation can be modeled as two steps: nucleation and propagation. (A) In *nucleation*, the first two chains come together to start the fibril. (B) In *propagation*, additional chains add to the growing fiber.

Equation 5.25 describes the *fibril propagation* step, for growing a fibril from size $m-1$ to size m; see Figure 5.14B. That is, you multiply by a factor of Kx for every chain that adds to the growing fibril. Now, in the same spirit as in helix–coil theory, we also define a *fibril nucleation* step. Let the nucleation equilibrium quantity δ account for intiation when the second protein adds onto the first to begin the formation of the fibril:

$$[A_2] = (Kx)[A_1]\delta = \delta Kx^2. \tag{5.26}$$

Here, (Kx) is the propagation parameter and δ is the nucleation parameter (similar to s and σ in helix–coil theory). $\delta \ll 1$ means that initiating fibril formation is difficult. The difference between our fibrillization model and helix–coil theory is that fibril formation depends on *protein concentration*, whereas helix–coil theory, which describes only the conformations of a single isolated molecule, does not. The cooperativity we model here is a matter of binding equilibrium, not conformational equilibrium.

In order to obtain the concentration $[A_m]$ in terms of x, the concentration of free monomeric protein, you multiply by $(Kx)^{m-1}$ because you have propagated by adding $m-1$ chains, and you multiply by δ because you initiated the fibril to get to the dimer. So, you have

$$[A_m] = \delta x(Kx)^{m-1}. \tag{5.27}$$

You can use Equation 5.27 to compute the relative concentrations of all the species, if you know the value of x. But, before going further, let's switch to another common way of expressing the relative concentrations. Our quantities above, $[A_m]$, are numbers of *m-mer fibrils per unit volume*. Now, we want to count protein molecules, not fibrils. We want c_m, the numbers of *protein molecules in m-mer fibrils per unit volume*. To convert, use $c_m = (m$ proteins per fibril$) \times$ (concentration of fibrils): $c_m = m[A_m] = (\delta/K)m(Kx)^m$. c_1 is the concentration of free protein molecules, c_2 is the concentration of protein molecules in 2-mer fibrils, and so on.

The total concentration of protein in solution is a sum over state concentrations:

$$\begin{aligned} c_{\text{tot}} &= c_1 + c_2 + c_3 + \ldots \\ &= c_{\text{monomer}} + c_{\text{fibril}} \\ &= x + \frac{\delta}{K}\sum_{m=2}^{\infty} m(Kx)^m. \end{aligned} \tag{5.28}$$

You can see that Equation 5.28 resembles the helix–coil partition function in Equation 5.14. Both are sums of statistical weights over all the possible states accessible to the system. Both have a nucleation equilibrium constant (σ for helix–coil processes and δ for monomer–fibril assembly). Both have a propagation constant (s for helix–coil processes and Kx for monomer–fibril assembly). Both monomer–fibril assembly and helix–coil transitions are two-state transitions if nucleation is difficult. In both processes, intermediate states are not populated. For helix–coil processes, if σ is small, then the chain will be a coil or a long helix, but not a short helix. For monomer–fibril processes, if δ is small (nucleation is difficult), then most of the proteins will be either in the form of monomers or in big fibrils, but not in small fibrils. A key factor that governs the balance between monomers and fibrils is Kx: if either

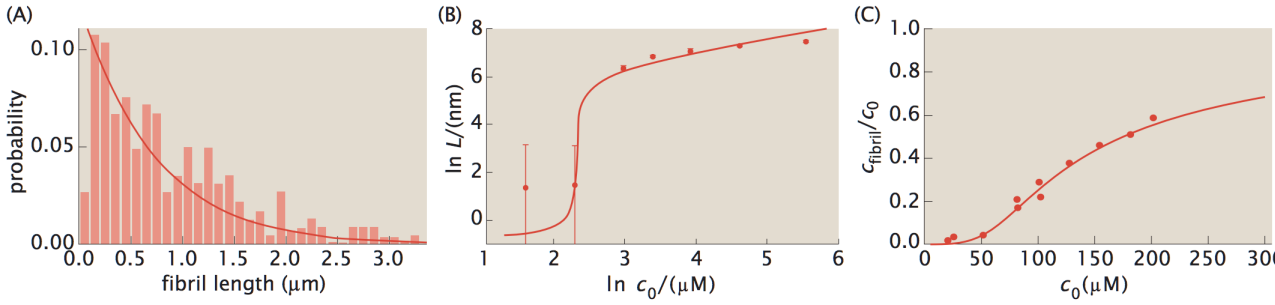

Figure 5.15 Amyloid fibril growth. (A) The length distribution of fibrils in amyloid. The bars show the experimental distribution of lengths of the amyloid peptide α-synuclein, for 30 μM bulk concentration. (B) Amyloid fibrils grow longer sharply as a function of peptide concentration. (C) The concentration of fibrils increases nonlinearly with the concentration of α-synuclein. For computational details, see [6]. (A and B, data from ME van Raaij, J van Gestel, IMJ Segers-Nolten, SW de Leeuw, and V Subramaniam. *Biophys J*, 95:450–458, 2008; C, data from E Terzi, G Hölzemann, and J Seelig. *J Mol Biol*, 252:633–642, 1995.)

K is large or the protein concentration x is large, the system will be mostly fibrillar. Increasing the protein concentration in solution leads to a transition from the monomeric to fibrillar state.

Here are three main conclusions from this model (details are given in Appendix 5B): (1) For a given protein concentration, the distribution $[A_m]$ of fibril lengths m is exponentially decreasing (Figure 5.15A). (2) Adding more protein into the solution leads to a cooperative increase in the *average lengths* of fibrils (see Figure 5.15B). (3) Adding more protein into the solution increases the *concentration* of fibrils, nonlinearly (see Figure 5.15C).

Cooperativity is a common feature of biological mechanisms. For example, viral capsids appear to assemble in all-or-none fashion from many protein molecules at a time. Also, the forces and velocities of molecular motors—which are proteins that move along protein tracks to perform molecular transport or to create forces and flows—are enhanced when multiple motor molecules work cooperatively together. And in *chemotaxis*—the process by which cells move toward food—receptor proteins assemble cooperatively in the membrane to amplify the cell's detection of food signals. In addition, the assembly of protein machines from their component proteins may be cooperative, but little is yet known.

SUMMARY

We have considered various cooperative processes in proteins. Some peptides and proteins undergo a sharp helix–coil transition, from a large ensemble of denatured conformations to a single helical conformation. The classical helix–coil model involves nucleation and propagation. Nucleation is unfavorable, but propagation is slightly favorable, so the chain forms either no helix or long helices. Protein folding in general is also cooperative. Both the secondary and tertiary structures can contribute to stability and cooperativity. In addition, we explored how proteins at high concentrations can associate to form fibrillar assemblies.

APPENDIX 5A: ADVANCED HELIX–COIL THEORIES

Various advanced helix–coil theories allow you to account for the specific sequences of amino acids in the chain and allow you to treat

multiple stretches of helix within a chain. Examples are the models of Zimm and Bragg [7], Lifson and Roig [8], and Agadir [9]. Here is the basic idea of the Zimm–Bragg model. Suppose you want to allow for multiple helices in a chain of different s values for each monomer type.

First, define the *generator matrix*,

$$\mathbf{G} = \begin{bmatrix} 1 & \sigma s \\ 1 & s \end{bmatrix}. \tag{5.A.1}$$

The partition function will be

$$Q_N = \begin{bmatrix} 1 & \sigma s \end{bmatrix} \mathbf{G}^{N-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix}. \tag{5.A.2}$$

$$\left(\text{Try it for } N = 2 : Q_2 = \begin{bmatrix} 1 & \sigma s \end{bmatrix} \begin{bmatrix} 1 & \sigma s \\ 1 & s \end{bmatrix} = 1 + \sigma s + \sigma s + \sigma s^2, \text{ which gives}\right.$$

$$\left. Q_2 = \underbrace{1}_{\text{cc}} + \underbrace{2\sigma s}_{\text{ch+hc}} + \underbrace{\sigma s^2}_{\text{hh}}. \right)$$

Now, for a specific sequence, such as Ala-Trp-Gly, take the values of s for each amino acid from Table 5.1, and choose a value of σ. Then, replace the generic matrix $\mathbf{G}^3$ with the product $\mathbf{G}_{\text{Ala}}\mathbf{G}_{\text{Trp}}\mathbf{G}_{\text{Gly}}$. Each such matrix uses a fixed value of σ and the s that is appropriate for that particular amino acid.

APPENDIX 5B: AMYLOID AGGREGATION THEORY

Here are a few further details of the amyloid fibrillization model described in the text. You can interpret fibrillization equilibria by looking at which concentration terms are large and which are small in Equation 5.28, for particular values of δ, K, and c_{tot}. Note that our independent variable here for protein concentrations is c_{tot}, not x. c_{tot} is the independent variable because you control it; it is the total concentration of protein that you put into the solution. Once you fix the value of c_{tot}, that determines the concentration of free monomer, x. This poses a little computational obstacle: you need to express the protein concentration c_m as a function of c_{tot}, rather than expressing $c_m(x)$ as a function of x. You can do so by solving Equation 5.28 numerically. The solution to this equation gives you the populations of all the fibrillar species of different lengths, either as protein concentrations c or as fibril concentrations $[A]$. From those values, you can compute average quantities, such as described in the next section.

Most Fibrils Are Relatively Short

How many fibrils are long, and how many are short? What fraction $f(m)$ of all fibrils has m peptide molecules per fiber? The relative populations of fibrils of different sizes are given by

$$f(m) = \frac{[A_m]}{\sum_{j=1}^{\infty}[A_j]} = \frac{(Kx)^m}{\sum_{j=1}^{\infty}(Kx)^j} \approx (Kx)^{m-1}(1 - Kx), \tag{5.B.1}$$

where we have used Equation 5.27. The approximation on the right-hand side comes from using the summation relationship $\sum_{m=1}^{\infty} y^m =$

$y(1-y)^{-1}$, which is valid as long as $y < 1$. Equation 5.B.1 predicts that more fibrils are short, and fewer fibrils are long. The fibril fraction $f(m)$ decreases geometrically with m for a fixed peptide concentration. This prediction is compared with experiments in Figure 5.15A.

Here is an important subtle point. On the one hand, $[A_m]$ diminishes with m. On the other hand, the concentration $c_m = m[A_m]$ first increases at small m, reaches a peak, and then decreases at large m. So, fibrils have a well-defined average length, which we compute below.

Fibril *Lengths* Increase Sharply with Peptide Concentration

Adding peptide to the solution can cause a sharp jump in the average size of fibrils. How does the average fibril length change with peptide concentration? You can readily compute the average fibril length as

$$\langle m \rangle = \sum_{m=1}^{\infty} mf(m) = \frac{\sum_m m(Kx)^m}{\sum_m (Kx)^m} = (Kx)(1-Kx)^{-1}, \tag{5.B.2}$$

where we have used Equation 5.B.1 and the additional summation relationship $1 + 2y + 3y^2 + \ldots = (1-y)^{-2}$. An important conclusion is qualitatively expressed in Equation 5.B.2. As x increases to the point that $Kx \to 1$, Equation 5.B.2 predicts that the average fibril size becomes infinite: $\langle m \rangle \to \infty$. It indicates that fibrils undergo a sharp transition in their average lengths. In dilute peptide solutions, fibrils are short. Increasing the peptide concentration leads to a sharp increase in fibril lengths. However, computing the average lengths and transition point accurately requires some numerical computations, for reasons noted previously. Figure 5.15B shows the results of the full calculation, and compares the model with experimental data on α-synuclein [6]. It shows how fibrils undergo a sharp length transition with increasing peptide concentrations in solution.

Increasing the Concentration of Peptides Increases the Fibril Concentration

Now, how does the fibril concentration depend on the amount of protein in solution? At high protein concentrations, each added peptide chain goes into forming fibril. Recall from earlier that for $\delta \ll 1$, our model gives two states for c_m: monomers and long fibrils. So, we can assume that $\sum_{m=2}^{\infty} \approx \sum_{m=1}^{\infty}$ because we are neglecting only a very small population of short fibrils. This allows us to use the relationship $\sum_{m=1}^{\infty} my^m = y(1-y)^{-2}$, which is valid for $y < 1$, where $y = Kx$ in this case. This gives a closed-form expression for the fibril concentration:

$$c_{\text{fibril}} = \frac{x\delta}{(1-Kx)^2}. \tag{5.B.3}$$

Now use Equations 5.28 and 5.B.3 to compute the fraction of protein, f_{fibril}, that is in the form of fibrils:

$$f_{\text{fibril}} = \frac{c_{\text{fibril}}}{c_{\text{tot}}} = \frac{\delta Kx/(1-Kx)^2}{x + \delta Kx/(1-Kx)^2} = \frac{\delta K}{(1-Kx)^2 + \delta K}. \tag{5.B.4}$$

To interpret Equation 5.B.4, note that the quantity Kx ranges from 0 to 1. As $c_{\text{tot}} \to \infty$, $Kx \to 1$. In this limit of high peptide concentration, you can see that (1) $f_{\text{fibril}} \to 1$, meaning that any additional proteins added

to solution will go into forming fibrils, and (2) $x \to (1/K)$ (because $Kx \to 1$), meaning that the concentration of monomers reaches a saturation concentration that you can compute from the propagation constant.

REFERENCES

[1] J Schellman. The factors affecting the stability of hydrogen-bonded polypeptide structures in solution. *J Phys Chem*, 62:1485–1494, 1958.

[2] K Ghosh and KA Dill. Theory for protein folding cooperativity: Helix bundles. *J Am Chem Soc*, 131:2306–2312, 2009.

[3] T Bereau, M Bachmann, and M Deserno. Interplay between secondary and tertiary structure formation in protein folding cooperativity. *J Am Chem Soc*, 132:13129–13131, 2010.

[4] KL Shaw, GR Grimsley, GI Yakovlev, et al. The effect of net charge on the solubility, activity, and stability of ribonuclease Sa. *Protein Sci*, 10:1206–1215, 2001.

[5] E Ruckenstein and IL Shulgin. Effect of salts and organic additives on the solubility of proteins in aqueous solutions. *Adv Coll Interface Sci*, 123–126:97–103, 2006.

[6] JD Schmit, K Ghosh, and KA Dill. What drives amyloid molecules to assemble into oligomers and fibrils? *Biophys J*, 100:450–458, 2011.

[7] BH Zimm and JK Bragg. Theory of the phase transition between helix and random coil in polypeptide chains. *J Chem Phys*, 31:526–535, 1959.

[8] S Lifson and A Roig. On the theory of helix–coil transition in polypeptides. *J Chem Phys*, 34:1963–1974, 1961.

[9] V Muñoz and L Serrano. Development of the multiple sequence approximation within the Agadir model of α-helix formation. comparison with Zimm–Bragg and Lifson–Roig formalisms. *Biopolymers*, 41:495–509, 1997.

SUGGESTED READING

Cantor CR and Schimmel PA, *Biophysical Chemistry. Part III: The Behavior of Biological Macromolecules*. WH. Freeman, San Francisco, 1980.

Chiti F and Dobson CM, Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem*. 75:333–366, 2006.

Dill KA and Bromberg S, *Molecular Driving Forces: Statistical Thermodynamics in Biology, Chemistry, Physics, and Nanoscience*, 2nd ed. Garland Science, New York, 2011.

Flory PJ, *Statistical Mechanics of Chain Molecules*. Wiley-Interscience New York, 1969; new edition by Hanser Publishers, Munich, 1989.