Biophysical Model Building

- **Step 1:** Come up with a hypothesis about how a system works
  - How many binding sites?
  - Is there cooperativity?

- **Step 2:** Translate the qualitative hypotheses into an observable mathematical form with *parameters*
  - Example parameters: K, tau, N
  - Parameters may not be known

- **Step 3:** Design an experiment that that can produce observables from step 2; perform the experiment
  - *Optimize* the parameters to make the fit look as good as possible

- **Step 4:** Assess the fit – Is the agreement convincing?

Example: Single Site Binding

- **Situation:** You are studying a novel DNA-binding protein
- **Hypothesis:** $P + D \leftrightarrow PD$
  - Parameters: K
  - Implicitly, we assume that N = 1, no cooperativity
- **Experiment:** Collect a binding curve (dialysis)
  - Optimize K for the best fit
- **Assess:** How good is our fit? Use statistics!

Example: Single Site Binding

- **Si**
  pr
- **H**
  -
  -
- **Ex**
  -
- **A**

What about the Thermodynamics?

- **Free energies:**

$$\Delta\bar{G}^0 = -RT\ln K$$

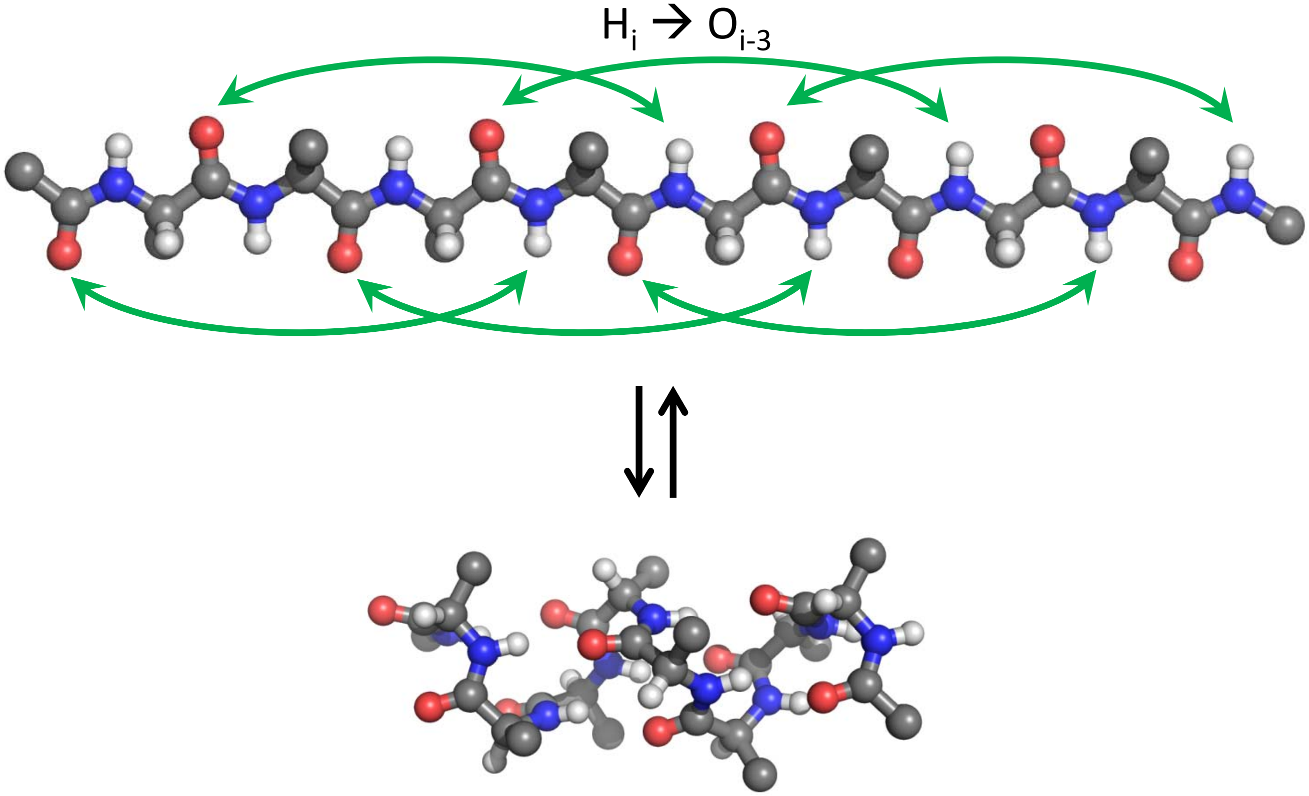
- **Enthalpies:** Van’t Hoff

$$\ln\frac{K_2}{K_1} = -\frac{\Delta\bar{H}^0}{R}\left(\frac{1}{T_2}-\frac{1}{T_1}\right)$$

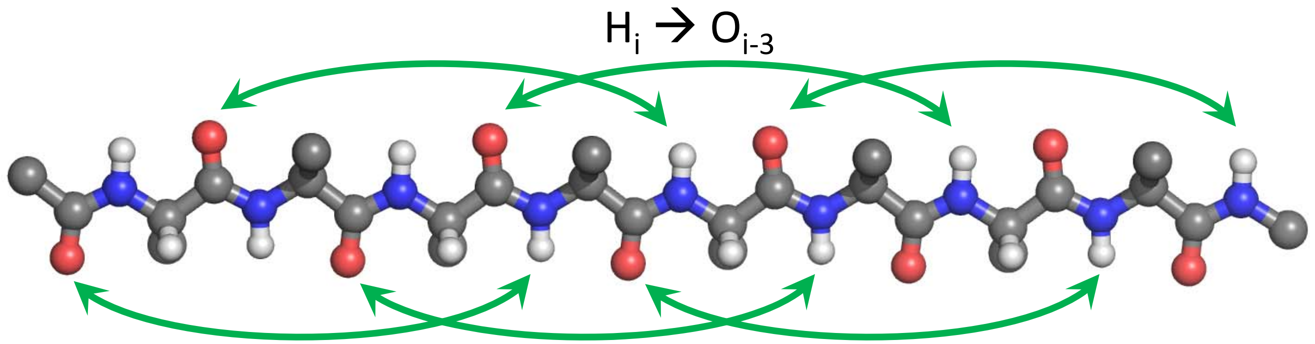
- Once the appropriate thermodynamic variables are known, one can predict K at any T, P, etc.

Helix-Coil Theory

- **Question:** How does an α-helix fold?



Helix-Coil Theory: Considerations



- **Entropic cost:** 3 pairs of ϕ, ψ torsions must be "fixed" before 1 hydrogen bond is formed
- **Energetic benefit:** forming H-bond is favorable once torsions are "fixed"
- **End effects:** No H-bonds for the end residues

Helix-Coil Theory: Assumptions

- **Assumption 1:** Each residue can exist in one of two conformational states: *h* or *c*
  - *h* (helix) is one conformation, *c* (coil) represents many *energetically equivalent* conformations
  - We can "enumerate" confirmations with a series of *h*'s and *c*'s
  - For 8 residues:

hhhhcccc *ccchhccc*

hchchchc *hhhhhhhh*

Helix-Coil Theory: Assumptions

- **Assumption 2:** For any sequence we consider, assume it's flanked by an infinite number of "coil" residues
  - This allows us to ignore end effects

ccchhccc → *...cccccchhcccccc...*

hhhhhhhh → *...ccchhhhhhhhccc...*

hhhhcccc → *...ccchhhhcccccccc...*

Helix-Coil Theory: Assumptions

- **Assumption 3:** Assume some conformations are not observed
  - *hch*, *hcch*: helical kinks, but no energetic benefit

ccchhccc → possible

hhhhhhhh → possible

hchhhccc → *not* possible

hhhcchhh → *not* possible

hhccchhh → possible

Helix-Coil Theory: Assumptions

- **Assumption 4a:** Individual residues are in equilibrium
  - If a new helical residue tries to form at the end of an existing helix:

$$\text{hhh}ccccc \rightarrow hhhhcccc \qquad K_{eq} = s$$

Helix-Coil Theory: Assumptions

- **Assumption 4b:** Individual residues are in equilibrium
  - If a new helical residue tries to form “from scratch”:

$$cccccccc \rightarrow ccchcccc \qquad K_{eq} = \sigma s$$

  - σ represents entropic cost of forming a helical ϕ, ψ values ($\sigma \ll 1$)

Helix-Coil Theory: Assumptions

- **Assumption 5:** Let the “unfolded” (i.e. all-coil) state be our reference state
  - Statistical weight of …cccccccc… = 1

Helix-Coil Theory: So What?

- Make a table (for N = 3):

State	# Helical Residues	Weight
...*ccc*...	0	$\frac{[ccc]}{[ccc]} = 1$
...*hcc*...	1	$\frac{[hcc]}{[ccc]} = ?$
...*chc*...	1	$\frac{[chc]}{[ccc]} = ?$
...*cch*...	1	$\frac{[cch]}{[ccc]} = ?$
...*hhc*...	2	$\frac{[hhc]}{[ccc]} = ?$
...*hch*...	2	$\frac{[hch]}{[ccc]} = ?$
...*chh*...	2	$\frac{[chh]}{[ccc]} = ?$
...*hhh*...	3	$\frac{[hhh]}{[ccc]} = ?$

Helix-Coil Theory: So What?

- Make a table (for N = 3):

State	# Helical Residues	Weight
…ccc…	0	$\frac{[ccc]}{[ccc]} = 1$
…hcc…	1	$\frac{[hcc]}{[ccc]} = \sigma s$
…chc…	1	$\frac{[chc]}{[ccc]} = \sigma s$
…cch…	1	$\frac{[cch]}{[ccc]} = \sigma s$
…hhc…	2	$\frac{[hhc]}{[ccc]} = ?$
…hch…	2	$\frac{[hch]}{[ccc]} = ?$
…chh…	2	$\frac{[chh]}{[ccc]} = ?$
…hhh…	3	$\frac{[hhh]}{[ccc]} = ?$

Helix-Coil Theory: So What?

- Make a table (for N = 3):

State	# Helical Residues	Weight
…*ccc*…	0	$\frac{[ccc]}{[ccc]} = 1$
…*hcc*…	1	$\frac{[hcc]}{[ccc]} = \sigma s$
…*chc*…	1	$\frac{[chc]}{[ccc]} = \sigma s$
…*cch*…	1	$\frac{[cch]}{[ccc]} = \sigma s$
…*hhc*…	2	$\frac{[hhc]}{[ccc]} = \sigma s^2$
…*hch*…	2	$\frac{[hch]}{[ccc]} = 0$
…*chh*…	2	$\frac{[chh]}{[ccc]} = \sigma s^2$
…*hhh*…	3	$\frac{[hhh]}{[ccc]} = ?$

Disallowed state!

Helix-Coil Theory: So What?

- Make a table (for N = 3):

State	# Helical Residues	Weight
…ccc…	0	$\frac{[ccc]}{[ccc]} = 1$
…hcc…	1	$\frac{[hcc]}{[ccc]} = \sigma s$
…chc…	1	$\frac{[chc]}{[ccc]} = \sigma s$
…cch…	1	$\frac{[cch]}{[ccc]} = \sigma s$
…hhc…	2	$\frac{[hhc]}{[ccc]} = \sigma s^2$
…hch…	2	$\frac{[hch]}{[ccc]} = 0$
…chh…	2	$\frac{[chh]}{[ccc]} = \sigma s^2$
…hhh…	3	$\frac{[hhh]}{[ccc]} = \sigma s^3$

Helix-Coil Theory: Experiment

- Partition function:

$$Z = \sum w_i$$

$$Z = 1 + 3\sigma s + 2\sigma s^2 + \sigma s^3$$

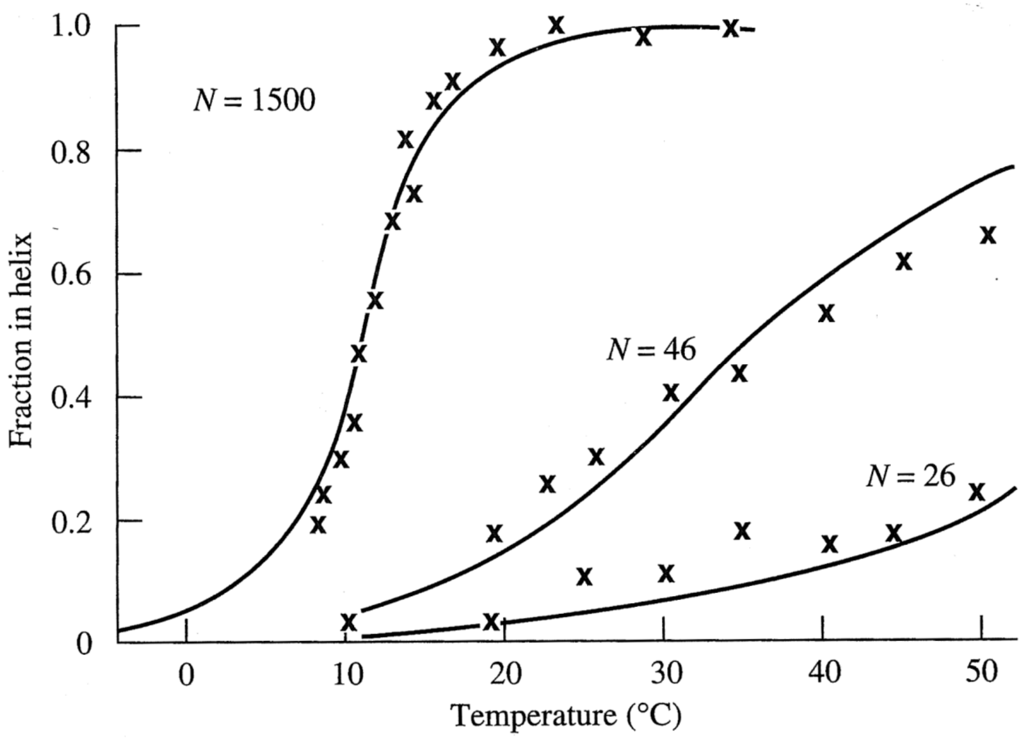
- Weighted average # helical residues:

$$\langle \text{helix} \rangle = \frac{\sum n_i w_i}{Z}$$

- <helix> is measurable!

State	# Helical Residues	Weight
…ccc…	0	$\frac{[ccc]}{[ccc]} = 1$
…hcc…	1	$\frac{[hcc]}{[ccc]} = \sigma s$
…chc…	1	$\frac{[chc]}{[ccc]} = \sigma s$
…cch…	1	$\frac{[cch]}{[ccc]} = \sigma s$
…hhc…	2	$\frac{[hhc]}{[ccc]} = \sigma s^2$
…hch…	2	$\frac{[hch]}{[ccc]} = 0$
…chh…	2	$\frac{[chh]}{[ccc]} = \sigma s^2$
…hhh…	3	$\frac{[hhh]}{[ccc]} = \sigma s^3$

Helix-Coil Theory: Experiment



◀ **FIGURE 11.7**

Temperature dependence of the fraction of monomer units in the helix form for poly-γ-benzyl-L-glutamate (in a 7:3 mixture of dichloracetic acid and 1,2-dichloroethane as the solvent) with polymer lengths of 26, 46, and 1500 monomer units. In this solvent the helix is favored at high temperatures. (Data from B. H. Zimm and J. K. Bragg, 1959, *J. Chem. Phys.* 31:526; B. H. Zimm, P. Doty, and K. Iso, 1959, *Proc. Natl. Acad. Sci. USA* 45:1601.)

Tinoco, p. 175.

Helix-Coil Theory: Takeaways

- Typical parameters
  - $s \approx 1$ (sometimes favorable, sometimes unfavorable depending on the solvent → hydrogen bond strength?
  - $\sigma \ll 1$ (typically on the order 0.001) → entropy is difficult to overcome

- Cooperativity comes from small σ relative to s

- It’s hard to break a helix in the middle
  - Helix typically “frays” at the ends

Helix-Coil Theory: DNA?

- DNA is more complicated than protein
  - Different base pairs → different s values
  - Two strands involved → σ is concentration dependent
  - Individual strands can "shift" relative to one another

- Lots of people have studied this (see book)

Random Walk in 1D

- Molecule diffuses l (average distance) before colliding with another molecule
  - This is the "step size" per hop

- For one step, probability (weight) of traveling forward is p, probability of traveling backward is $q = (1 - p)$

Random Walk for 3 Steps

- Partition function:

$$Z = \sum w_i$$

$$Z = p^3 + 3p^2q + 3pq^2 + q^3$$

- Avg. # forward steps:

$$\langle \mathrm{m} \rangle = \frac{\sum m_i w_i}{Z}$$

- Mean displacement:

$$\langle \mathrm{d} \rangle = [\langle m \rangle - (N - \langle m \rangle)]l$$

$$\langle \mathrm{d} \rangle = [2\langle m \rangle - N]l$$

- When $p = \frac{1}{2}$, $\langle \mathrm{d} \rangle = 0$
  - *Why?*

Steps	# Forward Steps	Weight
fff	3	p^3
bff	2	p^2q
fbf	2	p^2q
ffb	2	p^2q
bbf	1	pq^2
bfb	1	pq^2
fbb	1	pq^2
bbb	0	q^3

Random Walk

- Random walk in 1-D can be solved in general for N steps (see p. 167-172)

- **Main point:** When $p = \frac{1}{2}$, the *mean-squared displacement* is:

$$\langle d^2 \rangle = N l^2$$

- This can be used to model the end-to-end distance of an unfolded chain with no self-avoidance (N = # of links)

What Have We Learned?

- Many biological systems are made up of discrete states
  - Bound → Free; Native → Unfolded; Helix → Coil

- If we can determine relative concentrations (*weights*), we can sum the weights to get a *partition function*

- Once we know weights and the partition function, we can calculate observable values
  - Mole fractions: $X_i = \frac{w_i}{Z}$
  - Weighted averages: $\langle x \rangle = \frac{\sum x_i w_i}{Z}$
  - Degree of binding: $\bar{\nu} = \frac{\sum n_i w_i}{Z}$

What Don’t We Know?

- Can we calculate weights directly from energy differences (i.e. $\Delta \bar{G}$)?
- Can we relate entropy to the number of possible states?
- What is the molecular significance of the partition function?

Generic Table of States

State	Degeneracy	Generic Energy mol^{-1}	Weight
1	g_1	E_1	g_1X_1
2	g_2	E_2	g_2X_2
3	g_3	E_3	g_3X_3
…	…	…	…
N	g_N	E_N	g_NX_N

- <u>Boltzmann Distribution</u> gives mole fractions:

$$X_i = e^{-E_i/RT}$$

- Partition function is defined like before:

$$Z = \sum g_iX_i = \sum g_ie^{-\Delta\bar{G}_i/RT}$$

Boltzmann Distribution: Implications

- If we know total molecules (N_{tot}), the number of molecules in state *i* (N_i):

$$N_i = N_{tot} w_i = N_{tot} g_i e^{-E_i/RT}$$

- The fraction of molecules is simply

$$f_i = N_i / N_{tot} = g_i e^{-E_i/RT}$$

- Total and average energy:

$$\mathrm{E_{tot}} = \sum N_i E_i \qquad \langle E \rangle = \frac{\sum E_i g_i e^{-E_i/RT}}{Z}$$

Work and Heat

- "Generic Energy" can be $\Delta\bar{G}$ (binding), but if E is internal energy, we write:

$$dE_{tot} = \underbrace{\sum N_i dE_i}_{\text{Work}} + \underbrace{\sum E_i dN_i}_{\text{Heat}}$$

- Derivation is pretty neat (p. 157-158)
  - – Adding heat changes distribution of states (dN)

Entropy

- If there are Ω possible ways of forming a particular system (e.g. Ω conformations of a protein with the same energy):

$$S = k_B \ln \Omega$$

- k_B is the "Boltzmann constant" ($k_B = \frac{R}{N_A} = 1.381 \times 10^{-23}$ J K^{-1})

Partition Function

- The partition function (Z) can be related to U and S:

$$U = Nk_BT^2\left(\frac{\partial \ln Z}{\partial T}\right)_V \qquad S = k_BN\ln Z + \frac{U}{T}$$

- Why is this important?
  - If we knew the partition function exactly (hard to do), we could <u>predict</u> equilibrium constants, etc.

Summary: Statistical Thermodynamics

- Tables of states → What fraction of molecules is in each state
  - Statistics pertains to counting: how many x in y

- Partition function is the key to all systems we've discussed

- We've covered small numbers of discrete states (with discrete energies)
  - Large systems often have many more states and energies look continuous