21. (8 points) This problem is a small generalization of Example 4 in Lect. 23.2.
Consider a lattice polymer (cubic lattice) with immediate reversal forbidden, as in example 4, but now assume that there is an energy penalty Δ to bend the polymer to the 90° configurations.

(a) Show that the probability that the (n+1)st link continues in the direction of the nth link is $r \equiv \dfrac{1}{1+4e^{-\beta\Delta}}$.

(a) (2 points) At temperature T the probability of each state n is proportional to $e^{-\beta E_n}$. Thus, if the weight of the "forward" state is proportional to 1, the weight of each of the four 90° configurations is proportional to $e^{-\beta\Delta}$. Thus, $P_{forward} \equiv r = \dfrac{1}{Z}$ with $Z = 1 + 4e^{-\beta\Delta}$.

(b) Show that the mean-square end-to-end length is $\langle R^2 \rangle = Na^2\left(\dfrac{1+r}{1-r}\right) + O(1)$. Feel free to assume that the polymer is long enough so that $r^N << 1$.

(b) (4 points) The argument parallels the class development. $\vec{R} = \sum_{n=1}^{N} \vec{a}_n$, where $\vec{a}_n$ has magnitude a but is arbitrary in direction. $\langle R^2 \rangle = \left\langle \sum_{m=1}^{N} \vec{a}_m \cdot \sum_{n=1}^{N} \vec{a}_n \right\rangle = Na^2 + 2(N-1)\langle \vec{a}_n \cdot \vec{a}_{n+1} \rangle + 2(N-2)\langle \vec{a}_n \cdot \vec{a}_{n+2} \rangle + \ldots$
It is, again, only the all-"forward" configurations which contribute to calculating the averages $\langle \vec{a}_n \cdot \vec{a}_{n+k} \rangle$. (See argument in class notes.) The probability of this configuration is r^k. Thus,
$\langle R^2 \rangle = Na^2 + 2(N-1)r + 2(N-2)r^2 + \ldots$ The "subtracted" contributions are small relative to the O(N) contribution so long as $r^N << 1$. Thus, $\langle R^2 \rangle = Na^2\left[1 + 2r + 2r^2 + \ldots\right] + O(1) = Na^2\left[\dfrac{1+r}{1-r}\right] + O(1)$.

(c) Show that at low temperatures the Kuhn length $\ell_K \equiv a\left(\dfrac{1+r}{1-r}\right) \underset{T \to 0}{\to} \dfrac{a}{2} e^{\Delta/k_B T}$, i.e., it increases exponentially.

(c) (2 ponts) When temperature is low, $e^{-\beta\Delta}$ is very small, so $r \approx 1 - 4e^{-\beta\Delta}$ is close to 1 and
$\ell_K \equiv a\left(\dfrac{1+r}{1-r}\right) \approx a\left(\dfrac{2}{4e^{-\beta\Delta}}\right) = \dfrac{a}{2} e^{\beta\Delta}$.

22. (8 points) Radius of gyration of a freely jointed polymer.
Consider a freely-jointed homopolymer of step-length a and contour length Na in three dimensions. The N+1 monomers are at positions $\vec{R}_n, n = 0,1,\ldots,N$, with $\vec{R}_1 = \vec{R}_0 + \vec{a}_1, \vec{R}_2 = \vec{R}_0 + \vec{a}_1 + \vec{a}_2, \ldots$

(a) Show that the center of mass (C of M) is at $\vec{R}_{CofM} = \vec{R}_0 + \sum_{k=1}^{N} \vec{a}_k - \dfrac{1}{(N+1)} \sum_{k=1}^{N} k\vec{a}_k$.

(a) (3 points) The positions of the (N+1) monomers are:

$\vec{R}_0 = \vec{R}_0$

$\vec{R}_1 = \vec{R}_0 + \vec{a}_1$

$\vec{R}_2 = \vec{R}_0 + \vec{a}_1 + \vec{a}_2$

....

$\vec{R}_N = \vec{R}_0 + \sum_{k=1}^{N} \vec{a}_k$

Thus, adding, $\sum_{n=0}^{N}\vec{R}_n = (N+1)\vec{R}_N - \left[\vec{a}_1 + 2\vec{a}_2 + \ldots + N\vec{a}_N\right]$. Dividing by (N+1) gives the result claimed.

(b) The radius of gyration R$_G$ is defined by $R_G^2 = \frac{1}{(N+1)}\sum_{n=0}^{N}\left\langle\left(\vec{R}_n - \vec{R}_{CofM}\right)^2\right\rangle$, i.e., R$_G^2$ is the average over all polymer configurations of the mean-square distance between a monomer and the C of M. R$_G$ like the rms en-to-end length $\left\langle R^2\right\rangle$ is a measure of the size of a random-coil polymer.

Show for the model described above that $R_G^2 = \frac{N(N+2)a^2}{6(N+1)}$.

Note: For large N, this means that $R_G^2 \xrightarrow[N\text{ large}]{} \frac{Na^2}{6} = \frac{1}{6}\left\langle R^2\right\rangle$, where $\left\langle R^2\right\rangle$ is the mean end-to-end length, as calculated in class. This relation between the radius of gyration and the end-to-end length turns out to be general for all ideal Gaussian polymers (but I am not asking you to prove that here).

(b) (5 points) Calculate:

$$\left(\vec{R}_n - \vec{R}_{CofM}\right) = -\sum_{k=n+1}^{N}\vec{a}_k + \frac{1}{(N+1)}\sum_{k=1}^{N}n\vec{a}_k$$

$$\left\langle\left(\vec{R}_n - \vec{R}_{CofM}\right)^2\right\rangle = a^2(N-n) + \frac{a^2}{(N+1)^2}\sum_{k=1}^{N}k^2 - \frac{2a^2}{(N+1)}\sum_{k=n+1}^{N}k .$$

Now, use the common summation formulae,

$\sum_{k=1}^{n}k = \frac{n(n+1)}{2}$ and $\sum_{k=1}^{n}k^2 = \frac{n(n+1)(2n+1)}{6}$, (can you derive them?)

to obtain $\frac{1}{a^2}\left\langle\left(\vec{R}_n - \vec{R}_{CofM}\right)^2\right\rangle = (N-n) + \frac{N(2N+1)}{6(N+1)} - \frac{1}{(N+1)}\left[N(N+1) - n(n+1)\right]$.

Now, sum over n to obtain,

$\frac{1}{a^2}\sum_{n=0}^{N}\left\langle\left(\vec{R}_n - \vec{R}_{CofM}\right)^2\right\rangle = \frac{N(N+1)}{2} + \frac{N(2N+1)}{6} - \left[N(N+1) - \frac{N(2N+1)}{6} - \frac{N}{2}\right] = \frac{N(N+2)}{6}$, so finally,

$$R_G^2 \equiv \frac{1}{(N+1)}\sum_{n=0}^{N}\left\langle\left(\vec{R}_n - \vec{R}_{CofM}\right)^2\right\rangle \boxed{= \frac{N(N+2)a^2}{6(N+1)}}.$$

Comment: Note as a check that this formula works for the dimer (N=1), for which the C of M is the midpoint of the (single) bond, so $R_G^2 \equiv \frac{1}{(1+1)}\sum_{n=0}^{1}\left\langle\left(\vec{R}_n - \vec{R}_{CofM}\right)^2\right\rangle = \frac{a^2}{2}\left(\frac{1}{2^2} + \frac{1}{2^2}\right) = \frac{a^2}{4}$.

23. (10 points) Entropic Elasticity in one dimension. (see Lect. 25.2)

Consider a 1D polymer model with N steps of length a. Whether each step n is to the right (+) or to the left (-) is controlled by variables $\left\{\sigma_n = \pm 1\right\}_{n=1}^{N}$, so that the overall displacement $X = a\sum_{n=1}^{N}\sigma_n$. The end of this polymer is attached to an agency (perhaps, a very long spring) which exerts a constant force to the right of magnitude |f|. The energy E stored in this external "spring" is

$E = E(\{\sigma_n\}) = -|f|X = fX = (\text{force})\times(\text{distance})$, where f is the force exerted by the polymer and points to the left (negative). Note that positive energy is stored when the polmeric length X decreases.

(a) Calculate the Helmholtz free energy F of the whole system, including the external agency.

Hint: $Z = \sum_{n} e^{-\beta E_n} = \sum_{\{\sigma\}} e^{-\beta f X(\{\sigma\})}$.

(a) (2 points) $Z=\sum_n e^{-\beta E_n}=\sum_n e^{-\beta f X_n}=\sum_{\{\sigma_k\}} e^{-\frac{fa}{k_BT}\sum_{k=1}^{N}\sigma_k}=\left(\sum_\sigma e^{-\frac{fa\sigma}{k_BT}}\right)^N=\left(e^{\frac{fa}{k_BT}}+e^{-\frac{fa}{k_BT}}\right)^N=\boxed{\left(2\cosh\frac{fa}{k_BT}\right)^N.}$

(b) In this formulation, the polymeric displacement X is not fixed but becomes a statistical variable. Show generally that $\langle X\rangle=-\frac{\partial}{\partial(\beta f)}\ln Z$ and $\left\langle\left(X-\langle X\rangle\right)^2\right\rangle=\frac{\partial^2}{\partial(\beta f)^2}\ln Z$.

(b) (2 points) $\langle X\rangle=\frac{1}{Z}\sum_n X_n e^{-\beta E_n}=\frac{1}{Z}\sum_n X_n e^{-\beta f X_n}=\boxed{-\frac{\partial}{\partial(\beta f)}\ln Z}$ and similarly

$$\boxed{\frac{\partial^2}{\partial(\beta f)^2}\ln Z}=-\frac{\partial}{\partial(\beta f)}\langle X\rangle=-\frac{\partial}{\partial(\beta f)}\left[\frac{1}{Z}\sum_n X_n e^{-\beta f X_n}\right]=\frac{1}{Z}\sum_n X_n^2 e^{-\beta f X_n}-\sum_n X_n e^{-\beta f X_n}\frac{\partial}{\partial(\beta f)}\left(\frac{1}{Z}\right)$$

QED

$$=\frac{1}{Z}\sum_n X_n^2 e^{-\beta f X_n}+\left[\frac{1}{Z}\sum_n X_n e^{-\beta f X_n}\right]\frac{1}{Z}\frac{\partial Z}{\partial(\beta f)}=\left\langle X^2\right\rangle-\langle X\rangle^2=\boxed{\left\langle\left(X-\langle X\rangle\right)^2\right\rangle}$$

(c) Using the results of (a) and (b), calculate $\langle X\rangle$ as a function of the applied force f and invert to find f as a function of $\langle X\rangle$.

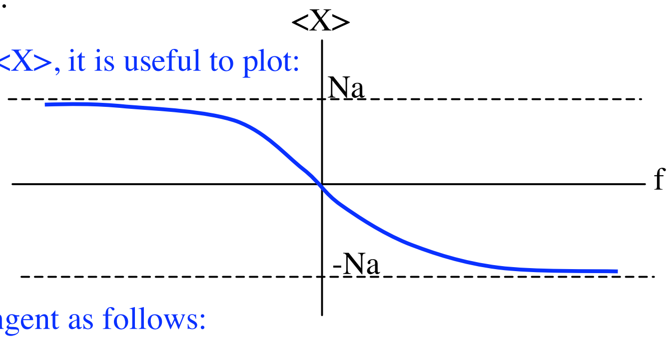
Hint: The result is $f=-\frac{k_BT}{a}\tanh^{-1}\left(\frac{\langle X\rangle}{Na}\right)$.

(c) (2 points) $\langle X\rangle=-\frac{\partial}{\partial(\beta f)}\ln Z=-\frac{\partial}{\partial(\beta f)}N\ln\left(2\cosh(\beta fa)\right)=-Na\tanh\left(\frac{fa}{k_BT}\right)$, so $\boxed{f=-\frac{k_BT}{a}\tanh^{-1}\left(\frac{\langle X\rangle}{Na}\right).}$

(d) By expanding your result in (c), show that $f=-\frac{k_BT}{\left\langle X^2\right\rangle}\langle X\rangle$ for small <X>, i.e., at low force, and

$f=\frac{k_BT}{2a}\ln\left(1-\frac{\langle X\rangle}{Na}\right)$, as $\langle X\rangle\to Na^-$, i.e., at large force.

(d) (2 points) To visualize the relations between f and <X>, it is useful to plot:



It is also useful to re-express the inverse hyperbolic tangent as follows:

Let $y=\tanh z=\frac{e^z-e^{-z}}{e^z+e^{-z}}=\frac{e^{2z}-1}{e^{2z}+1}$.

Now solve for z (=$\tanh^{-1}$y): $e^{2z}=\frac{1+y}{1-y}\quad\Rightarrow\quad z=\frac{1}{2}\ln\left(\frac{1+y}{1-y}\right)\equiv\tanh^{-1}y$.

Note from either of these forms that, for small y, $\tanh^{-1}y=y+O\left(y^3\right)$.

Thus, for small <X>, we have from (c),

$$f=-\frac{k_BT}{a}\left(\frac{\langle X\rangle}{Na}+O\left(\left(\frac{\langle X\rangle}{Na}\right)^3\right)\right)=-\frac{k_BT\langle X\rangle}{Na^2}+\text{small}=\boxed{-\frac{k_BT\langle X\rangle}{\left\langle X^2\right\rangle}+\text{small}.}$$

When f get large (and negative!), then <X> gets pulled out towards its maximum value, Na. In this region, it is convenient to use the logarithmic expression,

$$f = -\frac{k_BT}{2a}\ln\left(\frac{1+\frac{\langle X\rangle}{Na}}{1-\frac{\langle X\rangle}{Na}}\right) \approx -\frac{k_BT}{2a}\ln\left(\frac{2}{1-\frac{\langle X\rangle}{Na}}\right) = \frac{k_BT}{2a}\ln\left(\frac{1-\frac{\langle X\rangle}{Na}}{2}\right) \approx \boxed{\frac{k_BT}{2a}\ln\left(1-\frac{\langle X\rangle}{Na}\right),}$$

where at the last step I have dropped the ln2 because it is small relative to $\left|\ln\left(1-\frac{\langle X\rangle}{Na}\right)\right|$ when <X>is near its upper limit.

(e) Use your result in (b) to calculate the variance σ_X of X and show that it goes as $\sqrt{N}$, i.e., that the fluctuations of X become negligible at large N.

(e) (2 points) Calculate: $\sigma_X^2 \equiv \left\langle \left(X-\langle X\rangle\right)^2\right\rangle = -\frac{\partial}{\partial(\beta f)}\langle X\rangle = Na\frac{\partial}{\partial(\beta f)}\tanh(\beta fa) = \frac{Na^2}{\cosh^2(\beta fa)}$,

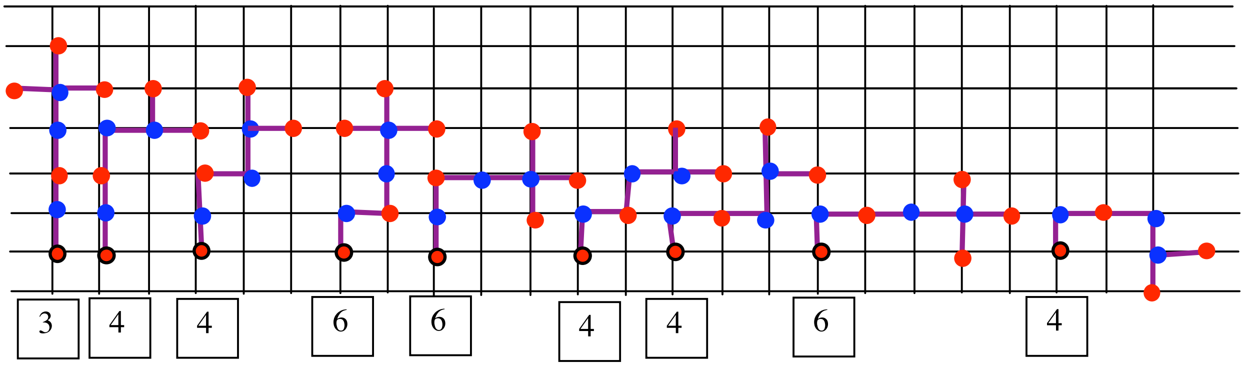
so, clearly $\boxed{\sigma_X = \frac{a}{\cosh(\beta fa)}\sqrt{N}\ .}$

The next two problems deal with the folding of a simple HP model of a lattice heteropolymer. Assume that the lattice is a two-dimensional square lattice and that the only interactions present are (i) a repulsive interaction of energy $\varepsilon > 0$ for every WH nearest-neighbor contact, (ii) an attractive interaction of energy $-\varepsilon$ for every WP nearest-neighbor contact, and (iii) an attractive interaction of energy $-\Delta$ $(\Delta > 0)$ for every HH nearest-neighbor contact.

24. (12 points) Consider the hexamer "protein" with the (ordered) sequence HPHPPH.
(a) Show that this protein has a ground ("native") state which is unique in shape (up to rotations, reflections, translations).
(b) Find the partition function and calculate as a function of temperature T the probability P_{native} that the protein will be found in its native state. What are the low- and high-temperature limits?
Hints:
The total number of self-avoiding hexamer configurations is 4x71=4x(81-10). Can you show this?
You will find that only five distinct energies occur. What are they? Label them $E_0 < E_1 < ... < E_4$ and write your answer in terms of these energies.

(10 points) There is no way to do this problem except direct enumeration of the distinct microstates and their energies. If you start at the left end and set the initial H on a given site of the lattice (this eliminates the translational degeneracy) and then set the first bond (H-to-P) along a given direction (this eliminates the overall 4-fold rotational degeneracy, then you have only four additional bonds to lay down on the lattice. Each bond has 3 possible directions, so your overall choices are $3^4 = 81$. However, this calculation is an overcount, since some of these configurations put two or more of the 6 monomers on the same lattice site. A little work will show you that there are 10 of these forbidden configurations. We simply count the remaining 71 configurations:

The easiest are those with no nearest-neighbor contacts:
All of these have 8 WH contacts and 6 WP contacts, so they all have energy $E_3 = 2\varepsilon$.
There are 41 such microstates:
Note: In the box underneath each microstate, I show the number of configurations of this type (including reflected versions).



Every time there is a nearest-neighbor contact, two water contacts are lost. There are a maximum of two nearest-neighbor contacts possible, each of which can be H-H (which gives an interaction energy -Δ) or H-P or P-P (which in this model do not contribute to the energy.

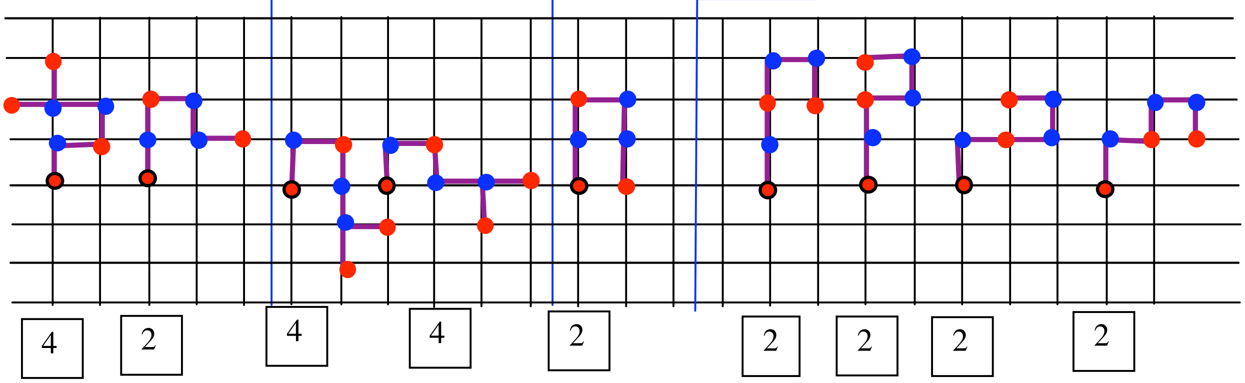
Thus, we can classify the remaining microstates:

8 WH contacts,
4 WP contacts,
no HH contacts,
energy $E_4 = 4\varepsilon$

7 WH contacts
5 WP contacts
no HH contacts
energy $E_3 = 2\varepsilon$

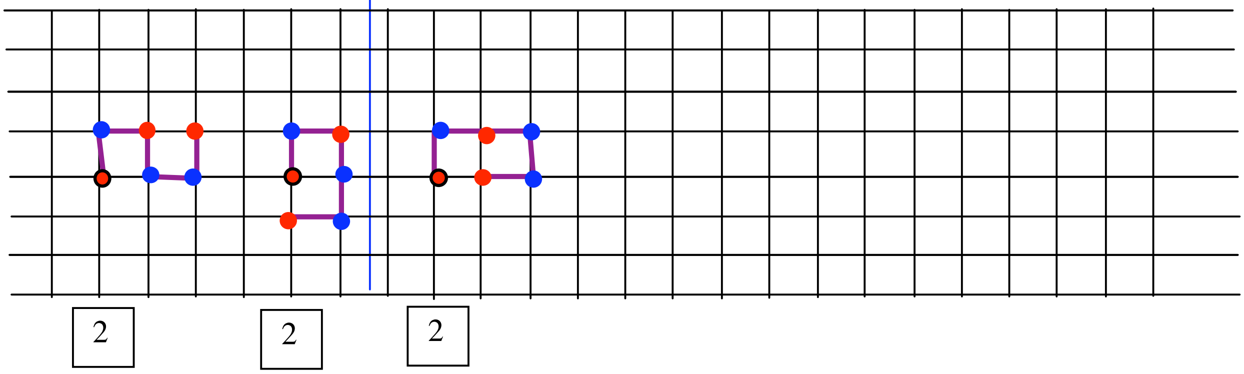
6 WH contacts
4 WP contacts
one HH contact
energy $E_2 = 2\varepsilon - \Delta$

6 WH contacts
6 WP contacts
one HH contact
energy $E_1 = -\Delta$



5 WH contacts
5WP contacts
one HH contact
energy $E_1 = -\Delta$

4 WH contacts
6 WP contacts
two HH contacts
energy $E_0 = -2\varepsilon - 2\Delta$



That's the hard part:

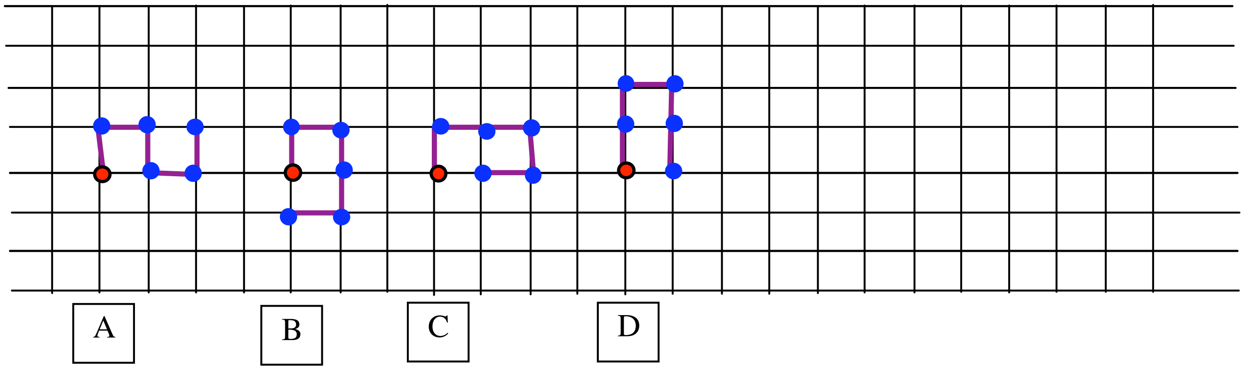
(a) At this point, it is easy to see that the unique ground state is the last one shown above.

(b) (2 points) By adding up contributions from all these 71 microstates, we find:

$$\boxed{P_{native}=\frac{1}{Z}2e^{\beta(2\varepsilon+2\Delta)}=\frac{2e^{\beta(2\varepsilon+2\Delta)}}{2e^{\beta(2\varepsilon+2\Delta)}+12e^{\beta\Delta}+2e^{\beta(-2\varepsilon+\Delta)}+49e^{-2\beta\varepsilon}+6e^{-4\beta\varepsilon}}}$$

At low temperature the native-state probability goes to unity (since the ground state is unique).
At high temperatures it goes to 2/71, since all states have the same probability.
25. Consider now other hexamers with different H,P sequences.
(a) There are only four "compact shapes," counting distinct orderings along the backbone (123456) but not rotations, translations, and reflections. What are they?

(a) The four compact shapes can be found by looking at the list from problem 24. They are:
I continue to show the first amino acid in red, just to indicate which is the beginning and which is the end.
Each of these states continues to have its reflected couterpart.



(b) Show that a noncompact shape can never be a UNIQUE ground state for any H,P sequence.

(b) The argument is simple. Every non-compact shape will have at least one residue which can be moved without changing the microstate energy. Such a state might be a ground state but it can never be a *unique* ground state.

(c) Consider the set of all possible H,P hexamer sequences (there are 2^6=64 of them).
Show that 57 of these do not "code" for a unique ground state. For the remaining 7 "codes", show the following results: Three of them code uniquely for 1 compact shape, three for another (which shapes are these?), the last codes for a third compact shape (which?), and the fourth compact shape is not the native state for any code.
Note: In part (c) ignore the water interactions (i.e., set ε=0).

(c) We have already seen from (b) that only compact shapes are candidates for unique ground states, so we need only consider the four shapes A, B, C, and D above. In the absence of water interactions, the only thing which can possibly make one of these compact shapes a ground state is the presence of one or two nearest-neighbor HH bonds.
A single HH bond is insufficient because
(i) internal bonds that leave one end free to be moved permit changes that do not change the energy, so the ground state is not unique, and
(ii) a single bond joining the two ends is possible for configurations B, C, and D provided that residues 1 and 6 are both H's. However, in this case configurations B, C, and D have the same energy, so the ground state is not unique.
It follows that any candidate for a unique ground state must have two internal HH bonds.
Below is a table showing which residues must be H in order to form these two bonds for each of the four compact structures. I have also filled in as "P" residues which must be P in order not to create degeneracy.
For example, for B: If the third residue is H, then the A and B structures will be degenerate, so B cannot be a unique ground state.

residue #	1	2	3	4	5	6	
A	H		H	H		H	(no code)
B	H		P	H		H	HPPHPH, HHPHPH, HPPHHH (“highly designable”)
C	H		H	P		H	HPHPPH, HPHPHH, HHHPPH (“highly designable”)
D	H	H	P	P	H	H	HHPPHH (unique code)

Note that A cannot be a unique ground state for ANY sequence, since any code with the A structure can form the B and C 2-bond states as well.
The codes which make each of the remaining ground states unique are shown above.
All other 57 codes have degenerate ground states.

Comment on water interactions:
Introducing water interactions makes the problem harder. Interestingly, however, the results are exactly the same as without the water interactions. But, the argument is harder.

The central problem is that now WP bonds lower energy, so that there may now be a trade-off between compact configurations which have more HH bonds and more-open configurations which have more WP bonds.

As before, only compact structures can be *unique* ground states. Thus, to search for codes with unique ground states, we need only examine them on the compact structures.

For any code, a particular compact structure *cannot* be a unique ground state, if I can show you another state (compact or not) with the same or lower energy.

To evaluate the energies, note that the number of WP bonds in one of the open configuration (no nearest-neighbor contacts) is: $WP = 2P + P_e$, where P is the total number of P residues and P_e is the number of P-end residues (1 or 6), if any. You lose water contacts whenever you form a PP or PH bond, so more generally, $WP = 2P + P_e - PH - 2PP$, where PH and HH are the numbers of PH and HH bonds,
Similarly, $WH = 2H + H_e - PH - 2HH$.
Thus, the total energy is $E = \varepsilon\left(WH - WP\right) - \Delta(HH) = 2\varepsilon\left(H - P\right) + \varepsilon\left(H_e - P_e\right) + 2\varepsilon\left(PP - HH\right) - \Delta(HH)$.
Since the first two terms on the right are fixed by the code, i.e., independent of configuration, it follows that the ground state requires minimizing the number of PP bonds and maximizing the number of HH bonds.

Now, every compact configuration has 2 bonds, each of which may be PP, HH, or HP, depending on the code.
Codes giving 2 PP’s, PP+PH cannot be *ground* states, since “opening” these cofigurations (i.e, breaking the bonds) decreases the energy.
Codes with 2PH or PP+HH cannot be *unique* ground states, since opening configurations leaves the energy unchanged.
Codes with 2 HH bonds were discussed in the regular part of the problem and give the old results.
Thus, the only potentially “new” situations here are codes providing HH+PH for one of the compact configurations.
If the PH bond can be broken without breaking the HH bond, this costs no energy, so such a configuration cannot be a unique ground state. It follows that the HH bond must be 1-6, which rules out configuration A.
The remaining possibilities are:

residue #	1	2	3	4	5	6
B	H		x	P		H
C	H		P	x		H
D	H	H	x	x	P	H
D	H	P	x	x	H	H

Putting an H in any one of the “x” positions allows the code to achieve a lower-energy state with two HH bonds by fitting into one of the patterns on the previous page, as does putting two H’s in the 2-5 positions.

The only remaining codes which might select for the B, C, or D configurations have the structures HPPPPH, HPPPHH, and HHPPPH. For the first of these B and C are degenerate; for the last two B, C and D are all degenerate. It follows that there is no code for which HH+HP can give a *unique* ground state.

Comment: These results illustrate two general ideas in a very simple context:
A. Only special “amino-acid” sequences lead to uniquely folded, native shapes.
B. Certain “native” shapes can be “designed for” by choosing the “right” “amino-acid” sequence, while others cannot. More extensive calculations suggest that it is common for some shapes to be easily “designable” (i.e., that there may be many sequences leading to the same or similar shapes), while other shapes are not.