

УДК 541.64

ЭФФЕКТЫ ЗАМОРОЖЕННОГО БЕСПОРЯДКА ДЛЯ ЧЕТЫРЕХБУКВЕННОЙ РНК

А.Л. ЦАТУРЯН

Ереванский государственный университет, Ереван, Армения

e-mail: haykcaturyan@gmail.com

(Поступила в редакцию 14 октября 2016 г.)

Представлена модель одонитевой РНК со случайной четырехбуквенной последовательностью. С использованием метода ограниченного отжига, ранее примененного для случайной цепи РНК с двухбуквенным алфавитом, рассчитаны такие термодинамические величины, как свободная энергия, энтропия, теплоемкость и т. д. Показано, что при соответствующем выборе параметров модель РНК с четырехбуквенным алфавитом демонстрирует как тепловую, так и холодную денатурацию РНК.

1. Введение

Молекула РНК является одним из важнейших гетерополимеров, который встречается во всех живых организмах [1]. У нее много функций таких, как хранение, обработка и передача генетической информации. Предполагается, что этот биополимер может играть ключевую роль в пребиотической эволюции [2]. Цепь РНК состоит из сахарных колец, связанных с фосфатными группами. Каждое сахарное кольцо связано с одним из четырех видов азотистого основания. Фосфатные группы соединяют 5' углерод одного сахарного кольца с 3' углеродом другого, образуя сахарофосфатный остов. РНК состоит из четырех видов нуклеотидов: урацила (U), аденина (A), цитозина (C) и гуанина (G). Все четыре нуклеотида могут образовывать водородные связи между комплементарными парами AU и GC. Эти пары нуклеотидов образуют вторичную структуру (рис.1).

Одним из самых интересных свойств этой макромолекулы является способность к самоорганизации (или фолдингу) РНК [3] и одной из популярных гипотез является гипотеза иерархического фолдинга [4]. Предполагается, что вначале образуется стабильная вторичная структура, а потом образуется третичная структура, которая незначительно меняет вторичную структуру. Кроме того, при образовании вторичной структуры пренебрегают так называемыми псевдоузлами, поскольку предполагается, что псевдоузлы порождают лишь небольшие

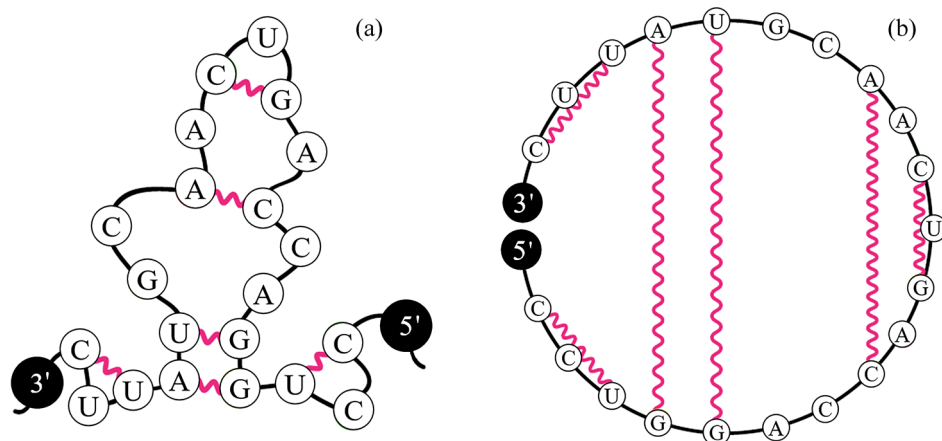


Рис.1. (a) Вторичная структура РНК и (b) ее схематическое круговое представление.

масштабные перестановки во вторичной структуре. Пренебрежение псевдоузлами видно на рис.1b, где хорды в окрестности не пересекаются. Это упрощение позволяет вычислить статистическую сумму для заданной последовательности РНК с помощью рекурсивных алгоритмов [5].

Другим интересным явлением, характеризующим как нуклеиновые кислоты, так и белки, является холодовая денатурация [6]. Причем, если тепловая денатурация является прямым физическим следствием возрастания энтропии, то денатурация при понижении температуры обычно понимается в терминах гидрофобных взаимодействий [7–10].

Целью настоящей работы является исследование тепловой и холодовой денатурации одонитевой РНК со случайной последовательностью нуклеотидов.

2. РНК со случайной последовательностью

Рассмотрим модель, пренебрегающую стэкинг-взаимодействием и энтропией петель и состоящую из четырех видов оснований $\{U, A, C, G\}$. Для описания нуклеотида введены две переменные изинговского типа h_i и g_i , т. е. каждый нуклеотид может быть описан так, как представлено в табл.1.

Табл.1. Описание нуклеотидов четырех типов

	$h_i = +1$	$h_i = -1$
$g_i = +1$	C	G
$g_i = -1$	U	A

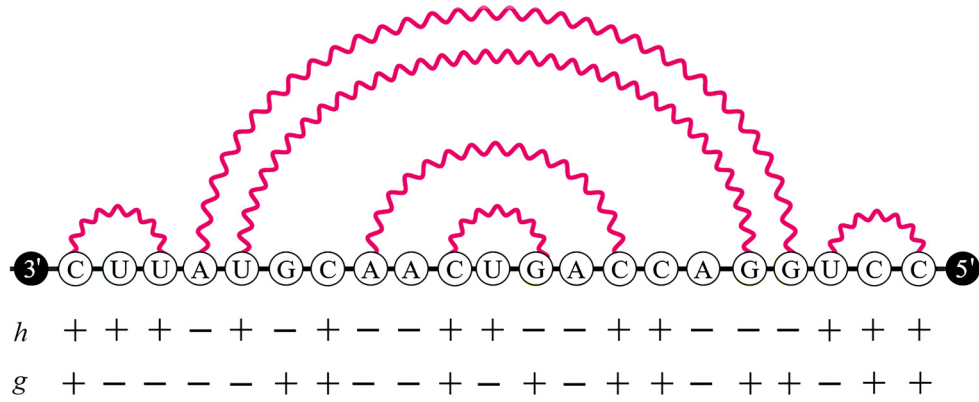


Рис.2. Последовательность вложенных друг в друга и независимых пар оснований для вторичной структуры, показанной на рис.1.

Обозначим каждую последовательность РНК как (h, g) . На рис.2 представлена последовательность вложенных друг в друга и независимых пар оснований (h, g) для вторичной структуры, изображенной рис.1. Последовательности генерируются с помощью двух независимых случайных величин h и g . Для этой конкретной последовательности использовались две равные вероятности $p_h = 0.5$ и $p_g = 0.5$. Здесь молекула изображена в виде диаграммы непересекающихся дуг (волнистые линии), представляющих собой водородные связи. Кроме того, определим матрицу пар оснований S , которая полностью определяет вторичную структуру РНК, как симметричную матрицу $N \times N$ с компонентами $s_{i,j}$ равными единице, если (i, j) образовали связь, и равной нулю в противном случае. Гамильтониан для конфигурации последовательности (h, g) записывается в виде суммы по неповторяющимся парам оснований

$$H(S, h, g) = \sum_{(i,j) \in S} \epsilon_{i,j} = \sum_{1 \leq i < j \leq N} s_{i,j} \epsilon_{i,j}, \quad (1)$$

где

$$\epsilon_{i,j} = \epsilon_0 + \epsilon_1 h_i h_j + \epsilon_2 g_i g_j. \quad (2)$$

В отличие от элементов матрицы пар оснований $s_{i,j}$ переменные h_i и g_i заморожены и могут принимать только значения $+1$ или -1 . Они не могут измениться, чтобы уменьшить суммарную энергию системы. Константы $\epsilon_0 < 0$, $\epsilon_1 > 0$ и $\epsilon_2 < 0$ позволяют различать по энергиям не уотсон–криковские взаимодействия в отличие от рассматриваемых ранее решеточных моделей [11–13]. При таком выборе можно определить четыре константы взаимодействия

$$\epsilon_{GC} = \epsilon_0 - \epsilon_1 + \epsilon_2, \quad \text{для GC и UA}, \quad (3)$$

$$\epsilon_{GA} = \epsilon_0 + \epsilon_1 - \epsilon_2, \quad \text{для GA и CU}, \quad (4)$$

$$\epsilon_{GU} = \epsilon_0 - \epsilon_1 - \epsilon_2, \quad \text{для GU и CA}, \quad (5)$$

$$\epsilon_{GG} = \epsilon_0 + \epsilon_1 + \epsilon_2, \quad \text{для GG, CC, AA и UU}. \quad (6)$$

В нашем случае имеется дополнительное ограничение $\epsilon_1 > |\epsilon_2|$, так что все канонические пары оснований, включая пару GU, имеют меньшую энергию, чем ϵ_{GG} .

Для каждой последовательности (h, g) введем случайные замороженные переменные h_i и g_i , распределенные как

$$P(h) = \prod_{i=1}^N \rho(h_i), \quad P(g) = \prod_{i=1}^N \rho(g_i), \quad (7)$$

где

$$\rho(h_i) = p_h \delta(h_i - 1) + (1 - p_h) \delta(h_i + 1), \quad (8)$$

$$\rho(g_i) = p_g \delta(g_i - 1) + (1 - p_g) \delta(g_i + 1). \quad (9)$$

Так как h и g взаимно независимы, то выполняются условия нормировки

$$\sum_{\{h\}} P(h) = 1, \quad \sum_{\{g\}} P(g) = 1. \quad (10)$$

В этом случае свободная энергия $f(h, g)$ РНК с заданной замороженной последовательностью имеет вид

$$\beta f(h, g) = -\frac{1}{N} \ln Z_N(h, g). \quad (11)$$

Таким образом, свободная энергия РНК зависит от последовательности. Однако согласно свойству самоусредняемости, в термодинамическом пределе свободная энергия оценивается как [14]

$$\beta f = -\frac{1}{N} \overline{\ln Z_N(h, g)}, \quad (12)$$

где $\overline{}$ – это усреднение по функциям распределения (7).

3. Метод ограниченного отжига

Основная идея заключается в том, чтобы вместо усреднения логарифма функции распределения в формуле (12) получить гораздо легко вычисляемое так называемое отожденное среднее, которое определяется как

$$\beta f^a = -\frac{1}{N} \ln \overline{Z_N(h, g)} = -\frac{1}{N} \ln Z_N^a. \quad (13)$$

Другой подход, получивший название метода ограниченного отжига, был

первоначально введен в работе [15] и успешно применен для модели Изинга. Вместо того, чтобы фиксировать значение каждой переменной h_i и g_i методом неопределенных множителей Лагранжа, фиксируются некоторые экстенсивные макропараметры, зависящие от большого количества переменных h_i и g_i . С помощью этих ограничений можно свести к минимуму разницу между точным значением замороженной средней и величиной, полученной этим методом. Таким образом, получается достаточно хорошая оценка свободной энергии (12).

Согласно работам [13,15], строится макроскопический параметр для ограничения отжига системы

$$\alpha_{[h,g]} = \frac{1}{N} \sum_{i=1}^N [[h,g]_i - (2p_{[h,g]} - 1)], \quad (14)$$

где

$$\overline{[h,g]} = \sum_{\{[h,g]\}} P([h,g])[h,g] = \sum_{[h,g]=\pm 1} \rho([h,g])[h,g] = 2p_{[h,g]} - 1. \quad (15)$$

Статистическая сумма имеет вид

$$Z_N(h,g) = \sum_{\{S\}} e^{-\beta \epsilon_0 |S|} \prod_{k < l}^N \exp(-\beta \epsilon_1 s_{k,l} h_k h_l) \prod_{m < n}^N \exp(-\beta \epsilon_2 s_{m,n} g_m g_n). \quad (16)$$

Полученная методом ограниченного отжига статистическая сумма будет имеет вид

$$Z_N^{\text{ca}}(\mu_h, \mu_g) = \exp[N\mu_h(2p_h - 1)] \exp[N\mu_g(2p_g - 1)] \times \sum_{\{S\}} e^{-\beta \epsilon_0 |S|} \prod (\mu_h, \mu_g), \quad (17)$$

где

$$\prod (\mu_h, \mu_g) = \sum_{\{h,g\}} \prod_{i=1}^N \rho(h_i) e^{-\mu_h h_i} \rho(g_j) e^{-\mu_g g_j} \times \prod_{k < l}^N \exp(-\beta \epsilon_1 s_{k,l} h_k h_l) \prod_{m < n}^N \exp(-\beta \epsilon_2 s_{m,n} g_m g_n).$$

Можно показать, что

$$Z_N^{\text{ca}}(\mu_h, \mu_g) = \Xi(\mu_h, \mu_g) Z_N^{\text{hom}}(\mu_h, \mu_g), \quad (18)$$

где

$$\Xi(\mu_h, \mu_g) = e^{N\mu_h(2p_h - 1)} e^{N\mu_g(2p_g - 1)} \Omega_h^N(\mu_h) \Omega_g^N(\mu_g).$$

Здесь величина $Z_N^{\text{hom}} = \sum_{\{S\}} [\omega^{\text{ca}}]^{|S|}$ является статистической суммой гомополимера при ограниченном отжиге со статистическим весом $\omega^{\text{ca}} = e^{-\beta \epsilon^{\text{ca}}}$ и эффективная энергия взаимодействия пары нуклеотидов

$$\epsilon^{\text{ca}}(\mu_h, \mu_g) = \epsilon_0 + \overline{\epsilon}_h(\mu_h) + \overline{\epsilon}_g(\mu_g). \quad (19)$$

Максимизация $f^{\text{ca}}(\mu_h, \mu_g) = -\frac{1}{N\beta} \ln Z_N^{\text{ca}}(\mu_h, \mu_g)$ по параметрам μ_h и μ_g дает значения $\widetilde{\mu}_h$ и $\widetilde{\mu}_g$, зависящие от $\beta \in_0$, $\beta \in_1$ и $\beta \in_2$, для которых $f^{\text{ca}}(\widetilde{\mu}_h, \widetilde{\mu}_g) \leq \overline{f(h, g)}$.

Тогда отоженная с ограничением свободная энергия будет равна

$$f^{\text{ca}}(\mu_h, \mu_g) = -\mu_h(2p_h - 1) - \mu_g(2p_g - 1) - \ln \Omega_h(\mu_h) - \ln \Omega_g(\mu_g) - \ln \left(1 + 2\sqrt{\omega^{\text{ca}}(\mu_h, \mu_g)} \right), \quad (20)$$

где $\widetilde{\mu}_h$ и $\widetilde{\mu}_g$ определяются уравнениями

$$1 - 2p_{[h,g]} = \frac{\partial \ln \Omega_{[h,g]}}{\partial \mu_{[h,g]}} - \frac{2\sqrt{\omega^{\text{ca}}}}{1 + 2\sqrt{\omega^{\text{ca}}}} \left(\frac{\partial \ln \Omega_{[h,g]}}{\partial \mu_{[h,g]}} - \frac{1}{2} \frac{\partial \ln \Upsilon_{[h,g]}}{\partial \mu_{[h,g]}} \right) \quad (21)$$

для h и g , соответственно. Степень спиральности, определяемая как доля спаренных нуклеотидов во вторичной структуре, оценивается как

$$\theta^{\text{ca}}(\mu_h, \mu_g) = \frac{2\sqrt{\omega^{\text{ca}}(\mu_h, \mu_g)}}{1 + 2\sqrt{\omega^{\text{ca}}(\mu_h, \mu_g)}}. \quad (22)$$

Удельная теплоемкость определяется соотношением

$$C_V^{\text{ca}}(\widetilde{\mu}_h, \widetilde{\mu}_g) = -T \frac{\partial^2 f^{\text{ca}}(\widetilde{\mu}_h, \widetilde{\mu}_g)}{\partial T^2}. \quad (23)$$

4. Результаты и их обсуждение

Поведение четырехбуквенной модели определяется значением параметра

$$\Lambda \equiv \left| \frac{\epsilon_1}{\epsilon_0 + \epsilon_2} \right|. \quad (24)$$

Для простоты возьмем $\epsilon_0 = -1$, создавая привлекательный фон взаимодействия, а также рассмотрим возможные значения отношения $\Lambda \equiv \left| \frac{\epsilon_1}{\epsilon_0 + \epsilon_2} \right|$. Так при $\Lambda > 1$

проявляется эффективная конкуренция между GC и парами, образуемыми одноименными нукелотидами, и наблюдается холодовая денатурация. В этом случае вторичная структура РНК при низких температурах ослабевает за счет замороженного беспорядка в первичной структуре. При неконкурентном режиме

($\Lambda < 1$) холодная денатурация не наблюдается, и мы имеем только высокотемпературное плавление, что видно на рис.3с. При этом удельная теплоемкость всегда имеет два пика (рис.3а). Критическое значение $\Lambda = 1$ является точкой, при которой низкотемпературный пик становится доминирующим и для которого $\theta_0 \neq 1$. При $\Lambda > 1$ пик растет, а степень спиральности принимает свое минимальное значение $\theta_0 = 0$ в области низких температур (рис.3b и d). При этом имеет место как тепловая, так и холодная денатурация. В обоих сценариях в пределе высоких температур степень спиральности асимптотически стремится к значению $\theta_\infty \equiv \lim_{T \rightarrow \infty} \theta^{\text{hom}} = 2/3$, что соответствует высокотемпературной расплавленной фазе.

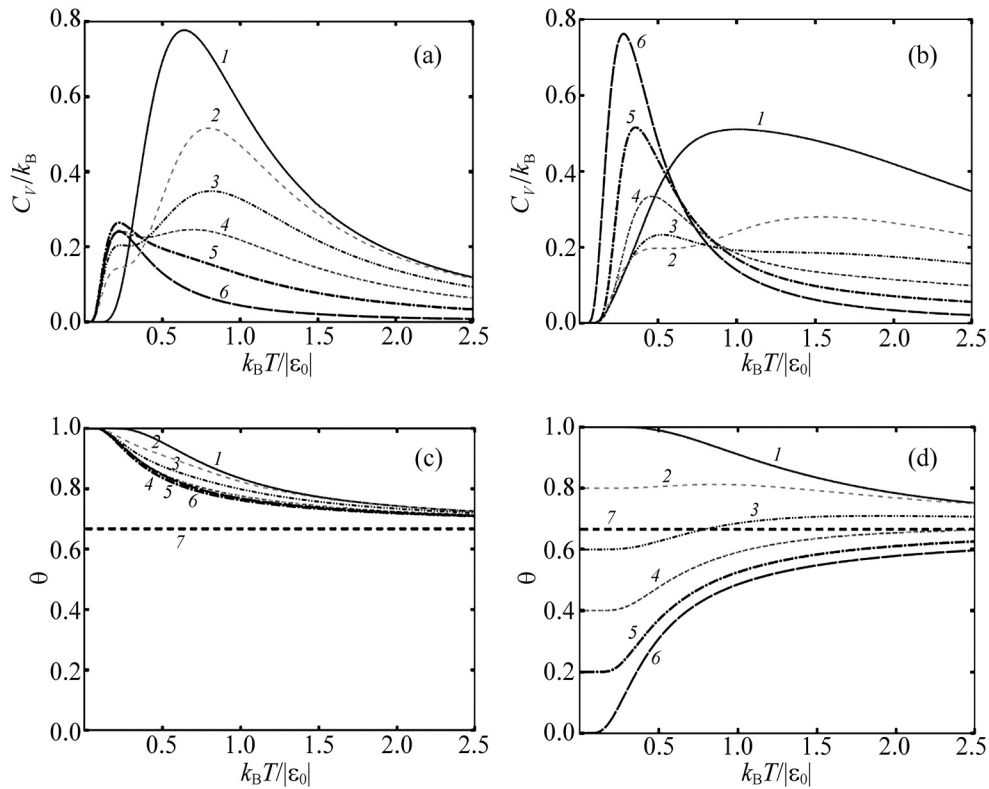


Рис.3. (a,b) Удельная теплоемкость и (c,d) степень спиральности для разных вероятностей $0.5 \leq (p_h, p_g) \leq 1.0$ в случае ограниченного отжига: 1 – $p_{[h,g]} = 0.5$, 2 – $p_{[h,g]} = 0.6$, 3 – $p_{[h,g]} = 0.7$, 4 – $p_{[h,g]} = 0.8$, 5 – $p_{[h,g]} = 0.9$, 6 – $p_{[h,g]} = 1.0$.

Наличие двух пиков наблюдалось также для двухбуквенной РНК [13]. Подобное поведение сопровождается низкотемпературным плавлением вторичной структуры РНК, описываемым степенью спиральности системы. Это имеет место, когда есть конкуренция между образованием GC и GG пар. Таким образом,

холодовая денатурация РНК определяется аномальным поведением статистических весов конкурирующих пар оснований.

5. Заключение

Исследованы термодинамические свойства модели однострессовой РНК со случайной последовательностью нуклеотидов и с энергией связывания, обеспечивающей существование большей части пар оснований. Введены две группы статистически независимых переменных, описывающих четырехбуквенную последовательность и рассмотрена спин-стекольная модель с замороженным беспорядком по вторичной структуре. Статистическая независимость вышеупомянутых групп переменных, описывающих беспорядок в нуклеотидной последовательности, позволила аналитически рассчитать свободную энергию системы, используя метод отжига с ограничениями и двумя вариационными параметрами. Изучено поведение модели в области низких температур и выявлены условия, при которых в системе происходит как тепловая, так и холодная денатурация. Возможность получения аналитических оценок особенно важна при изучении свойств длинных цепей РНК, когда численные расчеты требуют значительных временных и вычислительных ресурсов.

Выражаю благодарность Е.Ш. Мамасакхлисову за постановку задачи и поддержку в работе.

ЛИТЕРАТУРА

1. **R.F. Gesteland, T.R. Cech, J.F. Atkins.** The RNA World. Cold Spring Harbor Laboratory Press, 2005.
2. **G. Joyce.** Nature, **418**, 214 (2002).
3. **R. Russell.** Biophysics of RNA Folding. New York, Springer Science, 2013.
4. **I. Tinoco, O.C. Uhlenbeck, M.D. Levine.** Nature, **230**, 362 (1971).
5. **J. McCaskill.** Biopolymers, **29**, 1105 (1990).
6. **P.L. Privalov.** Critical Reviews in Biochemistry and Molecular Biology, **25**, 281 (1990).
7. **P.J. Mikulecky, A.L. Feig.** J. American Chem. Soc., **124**, 890 (2002).
8. **P.J. Mikulecky, A.L. Feig.** Nucleic Acids Research, **32**, 3967 (2004).
9. **A.V. Finkelstein.** Protein Physics. Academic Press, 2002.
10. **F. Sedlmeier, D. Horinek, R.R. Netz.** J. Chem. Phys., **134**, 055105 (2011).
11. **R. Bundschuh, T. Hwa.** Phys. Rev. E, **65**, 031903 (2002).
12. **Г.Н. Айрапетян, А.Л. Цатурян, Ш.А. Тоноян, Е.Ш. Мамасакхлисов.** Изв. НАН Армении, Физика, **48**, 144 (2013).
13. **G.N. Hayrapetyan, F. Iannelli, J. Lekscha, V.F. Morozov, R.R. Netz, Y.S. Mamasakhlisov.** Phys. Rev. Lett., **113**, 068101 (2014).
14. **M. Mezard, G. Parisi, M. Virasoro.** Spin Glass Theory and Beyond. World Scientific, Singapore, 1988.
15. **M. Serva, G. Paladin.** Phys. Rev. Lett., **70**, 105 (1993).

ՍԱՌԵՑՎԱԾ ԱՆԿԱՐԳԱՎՈՐՎԱԾՈՒԹՅԱՆ ԷՖԵԿՏՆԵՐԸ
ՔԱՌԱՏԱՌ ՌՆԹ-Ի ԴԵՊՔՈՒՄ

Հ.Լ. ԾԱՏՈՒՐՅԱՆ

Ներկայացված է պատահական քառատառ հաջորդականությամբ միաթել ՌՆԹ-ի մոդելը: Կիրառելով նախկինում պատահական երկտառ հաջորդականությամբ ՌՆԹ-ի համար օգտագործված սահմանափակ թրծման մեթոդը՝ հաշվարկված են թերմոդինամիկական պարամետրերը՝ ազատ էներգիան, էնթալպիան, ջերմունակությունը, և այլն: Ցույց է տրված, որ պարամետրերի համապատասխան ընտրության դեպքում քառատառ ՌՆԹ-ի մոդելը դրսևորում է ինչպես ջերմային, այնպես էլ սառը դենատուրացիա:

QUENCHED DISORDER EFFECTS FOR THE FOUR-LETTER ALPHABET RNA

H.L. TSATURYAN

Model of single-stranded RNA with the random quenched four-letter alphabet sequence is presented. Using constrained annealing approach, previously applied for the two-letter alphabet RNA, thermodynamic quantities such as free energy, entropy, heat capacity etc. are obtained. It is shown that at the appropriate choice of parameters for the four-letter alphabet of RNA, the model demonstrates both hot and cold denaturation.