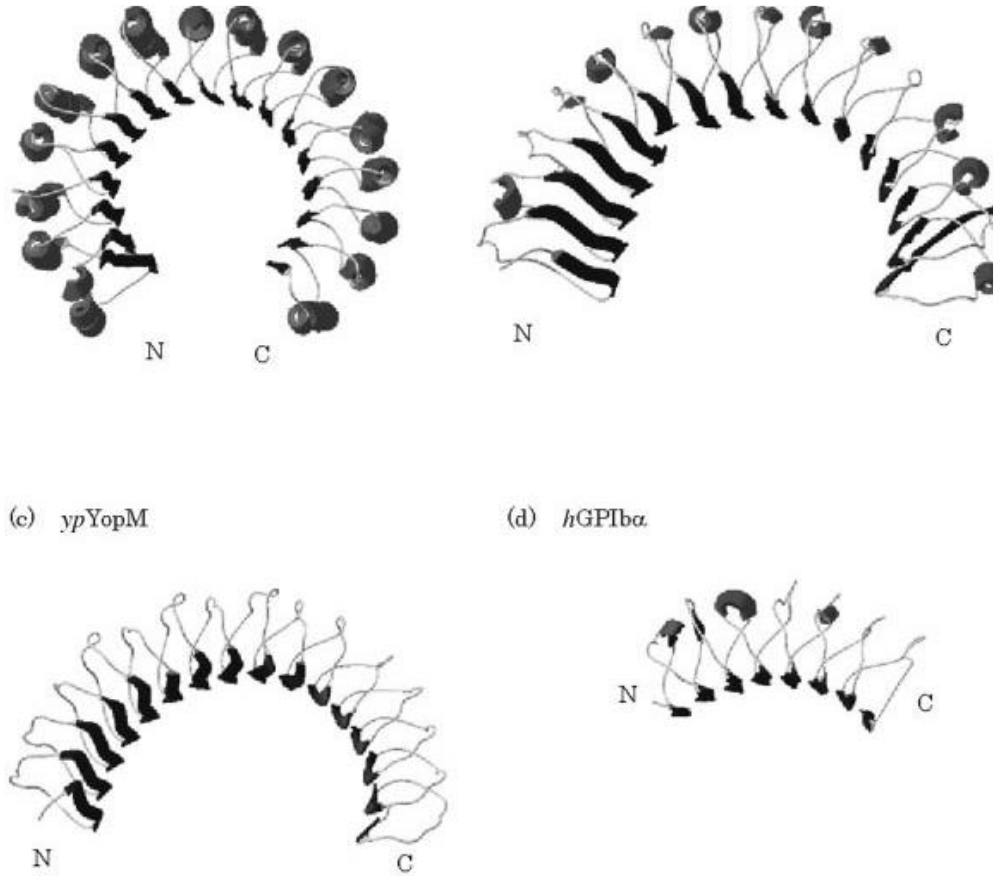


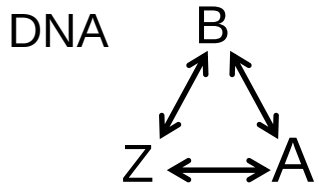
LRR-containing proteins are present in over 2000 proteins from viruses to eukaryotes.
Most LRRs are 20–30 amino acids long, and the repeat number ranges from 2 to 42. The known structures of 14 LRR proteins, each containing 4–17 repeats, have revealed that the LRR domains fold into a horseshoe (or arc) shape with a parallel β-sheet on the concave face and with various secondary structures, including α-helix, 3_{10}-helix, and pII helix on the convex face.
Conserved segments have an 11-residue stretch, LxxLxLxxNxL, or a 12-residue stretch, LxxLxLxxCxxL, in which “L” is Val, Leu, or Ile, “N” is Asn, Thr, Ser, or Cys, and “C” is Cys or Ser.

Lecture 2. 2022
Nucleic Acid Duplex Stability
why do we care?

Duplex nucleic acid stability. The Helix-Coil transition.

Nomenclature: duplex = double strand.
Duplex ⟷ single-strand transition is a helix ⟷ coil transition, keeping in mind that

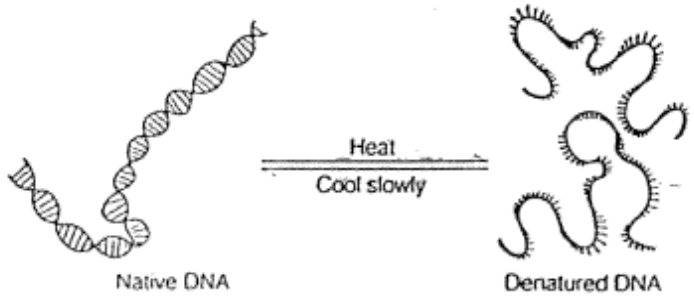
a) Single-stranded nucleic acids form helices and
b) Most single strands are not random coils



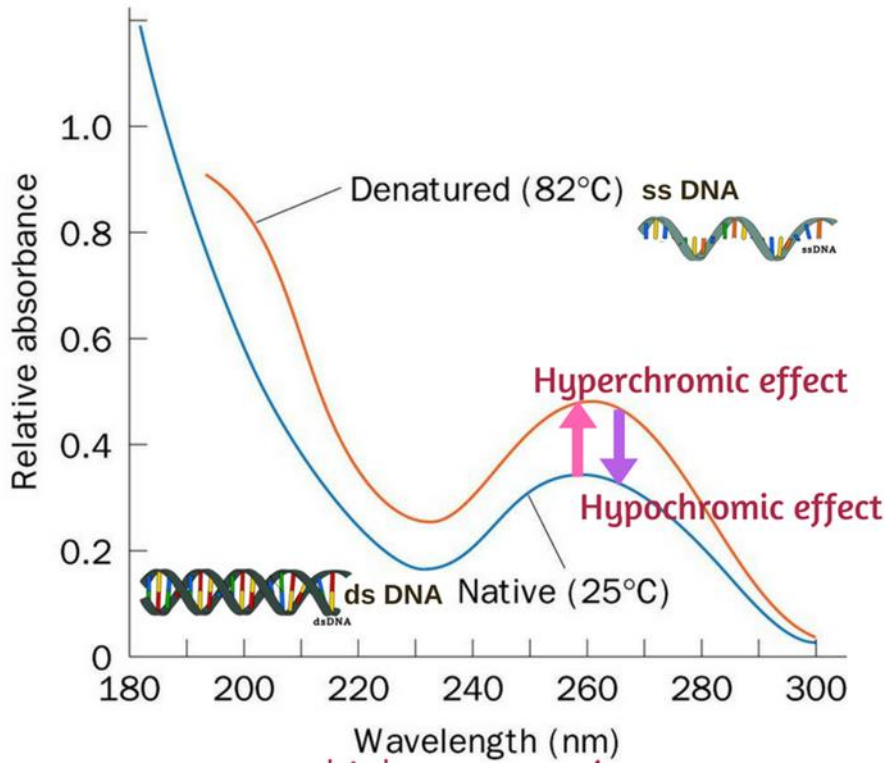
Experimentally, at 37° C in 0.2 M KCl, duplex B-form DNA is stabilized relative to single-strands by ~ 1 kcal/mol base pairs.

The simplest variable to use to denature duplex nucleic acids is temperature.

The transition is broadly characterized by an enthalpy ($\Delta H°$) of +6-11 kcal/mol bp, which reflects the unstacking of bases.
The entropy ($\Delta S°$) of denaturation is ~25 eu per base pair.



Stacked bases of nucleic acids have a lower absorbance than unstacked bases, due to interactions between the electric dipole transition moments of individual bases with those of their neighbors, and thus absorbance is sensitive to the relative orientation of the bases. This effect of increased absorption upon denaturation to the random coil form is called **"hyperchromicity"**, and is most pronounced in the helix↔coil transition, but also present for single-stranded polynucleotides (which retain base stacking in solution).



What is the relative absorbance of duplex B-form DNA and A-form RNA?

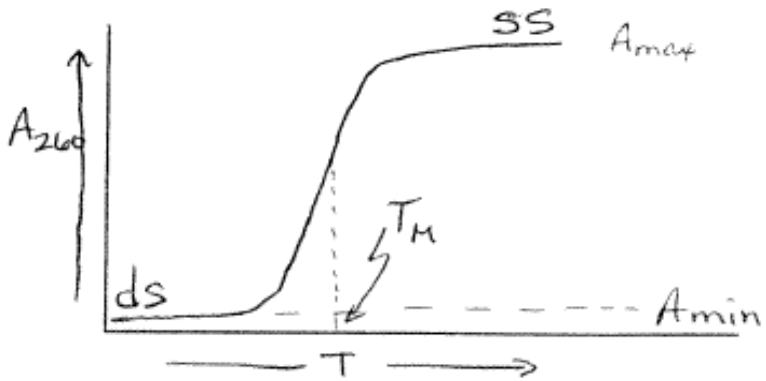
A word about base stacking.

- Stacking occurs through electronic interactions between bases.
- Stacking is enthalpically driven ($\Delta H° < 0$)
- Entropy of stacking is unfavorable.
- If stacking were hydrophobic, then it would be accompanied by a release of water and an increase in entropy, which is not observed. Stacking is not hydrophobic.

Base stacking is the driving force for nucleic acid folding/duplex formation

Characteristics of a spectroscopic "melting" curve.

The absorbance spectrum of a polynucleotide (or oligonucleotide) retains the features of the individual nucleotides; i.e. the absorbance maximum at neutral pH is near 260 nm. Changes in absorbance are therefore monitored at this wavelength.

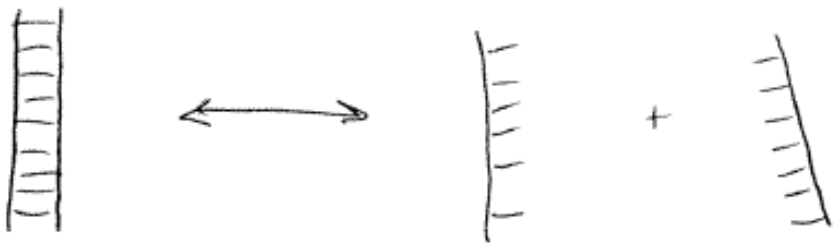


$$\%ss = \frac{A_T - A_{min}}{A_{max} - A_{min}} \times 100$$

The melting curve has a transition midpoint (T_M), which is the temperature at which the duplex is one-half denatured, and the free energy of the transition (ΔG°) is zero. (recall that $\Delta G^\circ = -RT(\ln K_{(eq)})$. Since $K_{(eq)}=[den]/[nat]=1$ at the T_M, $\Delta G^\circ = 0$.)

The width of the transition is an indication of the cooperativity of the transition.

For short DNA fragments (≤200 bp), melting is all or none. At the T_M, ½ of the molecules are fully denatured and ½ are fully duplex.



For DNA duplexes ≤200 b.p., the denaturation of the ends becomes a thermodynamically significant event, so that the melting temperature decreases as the length decreases.
For short oligonucleotides, the T_M is also concentration dependent (mass action, higher concentration favors the duplex).

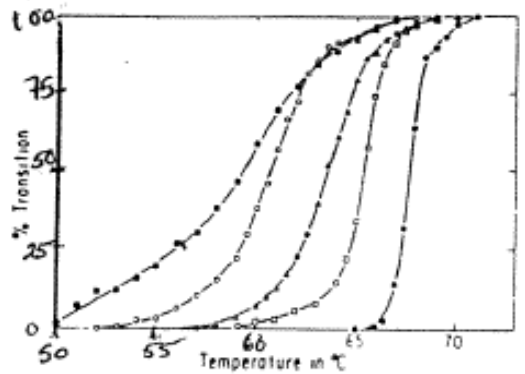


Fig. 6. Normalized melting curves for unsonicated calf thymus DNA (●) and various fractionated fragments (Fig. 5) in 2.4 *M* TEA: (□) $\bar{M}_w$ = 78,000, (△) $\bar{M}_w$ = 48,000, (○) $\bar{M}_w$ = 40,000, (■) $\bar{M}_w$ = 30,000.

Mw = 78,000 = 118 bp
Mw = 48,000 = 72 bp
Mw = 40,000 = 60 bp
Mw = 30,000 = 45 bp

We want to express the transition from duplex to single strand in thermodynamic terms which will provide a means to understand the reaction. First, define an expression for this transition, based on the experimental parameters.

The equilibrium expression:

[*S*] + [*S*] ↔ [D] where [*S*] is *S*trand concentration and [D] is duplex concentration

leads to conservation of mass: $S_T = [S] + 2[D]$ in terms of total strand concentration.

$K = \frac{[D]}{[S]^2}$ by mass action is the equilibrium constant

$f = \frac{2[D]}{S_T}$ the fraction of single strands in a duplex

Combine these three relationships to get an expression for f that does not contain [D] and [*S*], using the quadratic equation:

$$f = \frac{4KS_T + 1 - (1+8KS_T)^{1/2}}{4KS_T}$$

Since we are measuring the temperature dependence of the equilibrium, define the change in f with K:

$$\frac{\partial f}{\partial \ln K} = 1 - f - (1+8KS_T)^{-1/2}$$

When f = 0.5, this is the midpoint of the melting curve (the melting temperature T_M), and $S_T = 4[D]$

Replace this expression in $\{S_T = [S] + 2[D]\}$ to give $[S] = 2[D]$ And replace that result in the expression for K, to give $K = 1/(4[D])$

You can see that when f = 0.5, $KS_T = 1$, so that $\left.\frac{\partial f}{\partial \ln K}\right|_{f=0.5} = \frac{1}{6}$

Where do the thermodynamic parameters enter?

Any method that allows you to measure the equilibrium constant as a function of temperature will also allow you to obtain the enthalpy of the reaction, as van't Hoff described

$$\frac{\partial \ln K}{\partial (T)} = \frac{\Delta H^\circ}{RT^2}$$

where R is the gas constant (1.987 cal/mol-K) and T is temperature in K.

$$\frac{\partial f}{\partial (T)_{f=0.5}} = \frac{\Delta H^\circ}{6RT^2}$$

Which says that at the midpoint of the transition, the slope is a measure of the change in enthalpy.

The two-state melting curve shown is then fit by the expression (recalling that $\Delta G^o = \Delta H^o - T\Delta S^o$) using nonlinear least-squares with six independent variables:

$$K = \exp[-\Delta H^o/RT + \Delta S^o/R]$$

For a bimolecular reaction, the T_M will vary with S_T. To show this, recall

$$\Delta G^o = -RT\ln(K)$$

At T_M, $KS_T = 1$,

$$-RT_M\ln(K) = RT_M\ln(S_T) = \Delta H^\circ - T_M \Delta S^\circ$$

To finally give

$$\ln(S_T) = \frac{\Delta H^\circ}{R}\frac{1}{T_M} - \frac{\Delta S^\circ}{R}$$

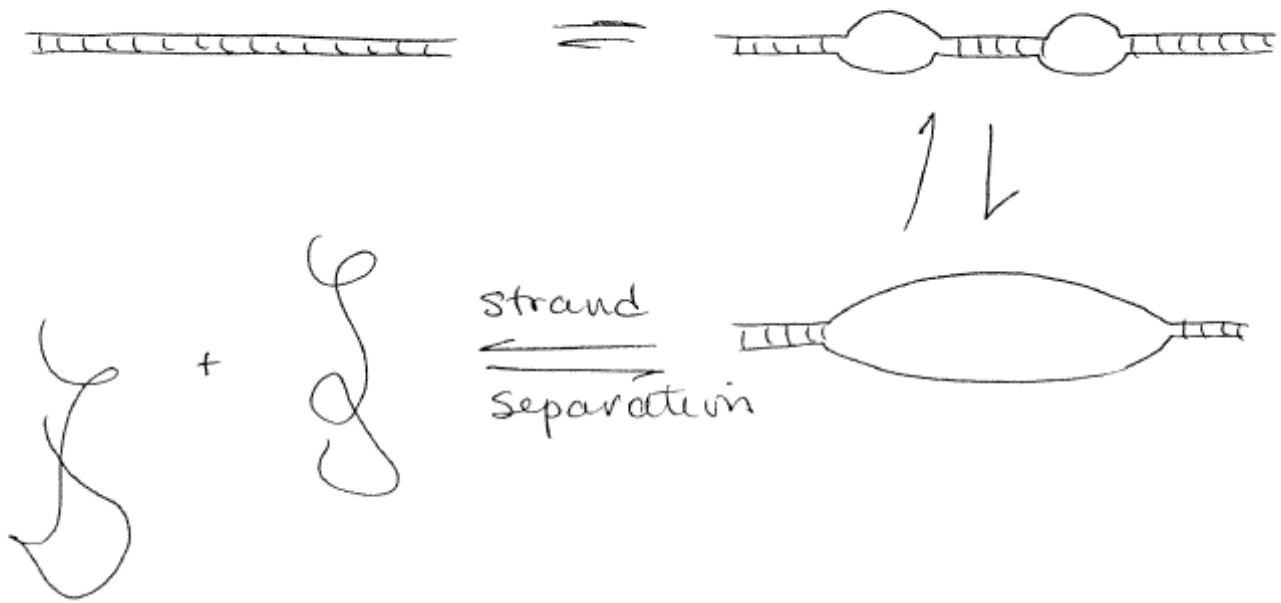
A plot of $\ln(S_T)$ *vs* $1/T_M$ has a slope of $R/\Delta H^\circ$ and an intercept of $\Delta S^\circ/\Delta H^\circ$.

So there are the thermodynamic parameters for the denaturation of a self-complementary duplex. By convention, these are the standard state parameters, and typically given for the reaction at 37 °C (310 K), since $\Delta G^\circ = \Delta H^\circ(1 - T/T_M)$.

For high MW DNA, melting is partial, in the sense that at the T_M each long molecule is 50% duplex and 50% single-stranded. Intermediates in the melting process can be observed for long duplex DNAs, where melting occurs in segments of the duplex ("cooperative units") that are shorter than the DNA molecule itself.

For high MW DNA, the T_M is
independent of MW and independent of DNA concentration

The cooperative units vary from between 10 - 100 b.p.

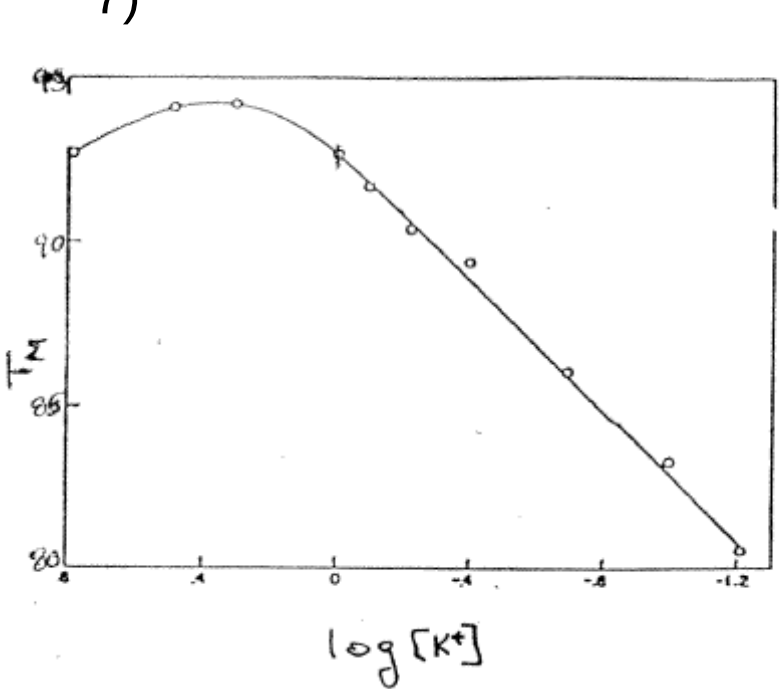


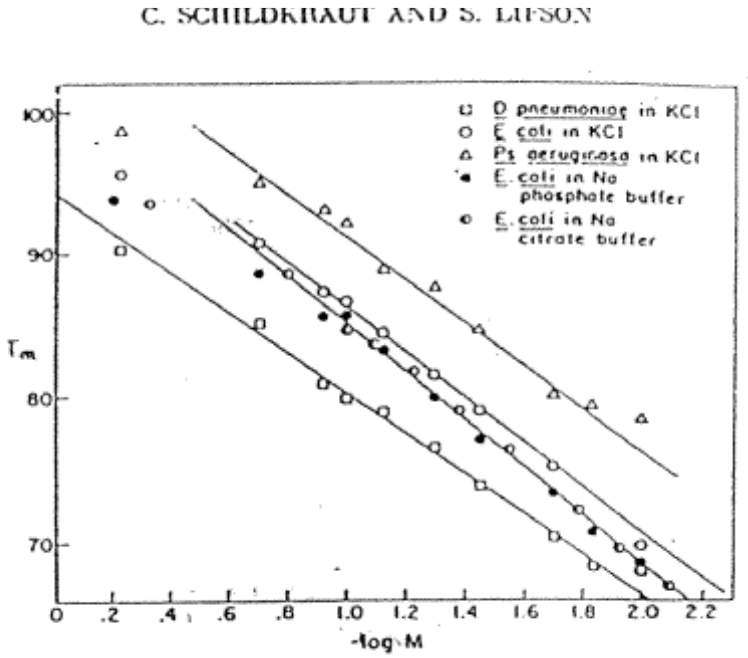
Factors affecting duplex nucleic acid stability

1. Base composition (GC content) and sequence (for oligonucleotides)
2. Molecular weight
3. Monovalent cation concentration
4. Multivalent cation concentration
5. Temperature
6. pH
7. Tertiary structure
8. Non-aqueous solvents
9. Proteins and other ligands that bind preferentially to either duplex or single strand
10. Anions at > 1M concentrations

Monovalent ion concentrations

- Below 1 M MX, there is no effect of anions.
- For Cations, there is a strong effect on duplex stability, since the cations electrostatically interact with the phosphates
- Experimentally for DNA duplexes, the relation is $dT_M/d(\log[M^+]) = 19°$ C (pH 7)





Brief comment on the electrostatic problem

Duplex DNA, assuming B form and 10.4 bp/turn, has 2 PO_4^-/3.4 Å.

Duplex RNA is A-form, and 11-12 bp/turn, has 2 PO_4^-/2.6 Å

The B-form duplex has ~10 bp/turn, and 34 Å/turn. Thus base pairs stack 3.4 Å apart.
The A-form duplex has ~11-12 bp/turn, and 28 Å/turn. Thus base pairs stack ~2.6 Å apart.

Thermodynamically, there is not a 1:1 ratio of phosphate to M^+ ion. Instead, there are 0.88 M^+ bound per phosphate for duplex DNA.
And 'bound' does not mean stable electrostatic binding, for these ions are loosely associated with the nucleic acid and float around it.

Single strand DNA has 1 phosphate per 3.4 Å, but there are only 0.70 M^+ 'bound' to each phosphate (thermodynamically).

Therefore:

duplex $\leftrightarrow$ single strand + xM^+

where ions are released upon melting of the duplex. For DNA, x = 0.88 – 0.70 = 0.18 M^+ per phosphate.
Increasing the monovalent ion concentration favors the duplex (by Le Chatelier's principle of mass action), and so the T_M increases.

4. Multivalent ions

- 'Bind' with high affinity to DNA and RNA, and affect the T_M at lower concentrations than do monovalent cations.
- For DNA duplexes, the most important multivalent cations are spermidine, spermine, and putrescine.
- For RNAs, the most important ion is Mg^{2+}.

Spermidine $NH_3^+CH_2CH_2NH_2^+CH_2CH_2CH_2CH_2NH_3^+$

Spermine $NH_3^+CH_2CH_2CH_2NH_2^+CH_2CH_2CH_2NH_2^+CH_2CH_2CH_2NH_3^+$

Polyelectrolyte properties of DNA (RNA)

- Highly charged polyanion
- One negative charge per phosphate (its pKa ~ 2)
- Associates with cations in solution
- In a solution of duplex DNA + KCl, about 75% of the phosphate charge is neutralized by K+ 'bound' to phosphate backbone. This is NOT site-specific binding; the cation retains its water of hydration.

5. Temperature

Increased temperature destabilizes the duplex ($\Delta H^\circ > 0$ for melting)

6. pH

Protonating or deprotonating the bases can disrupt Watson-Crick pairing. Duplex stability is pH-independent between pH 6-9.5.

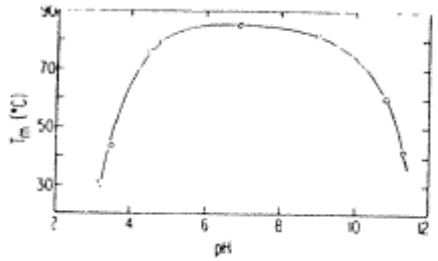


Fig. 4. Dependence of T_m (calf thymus DNA, 0.12M Na^+) on pH.

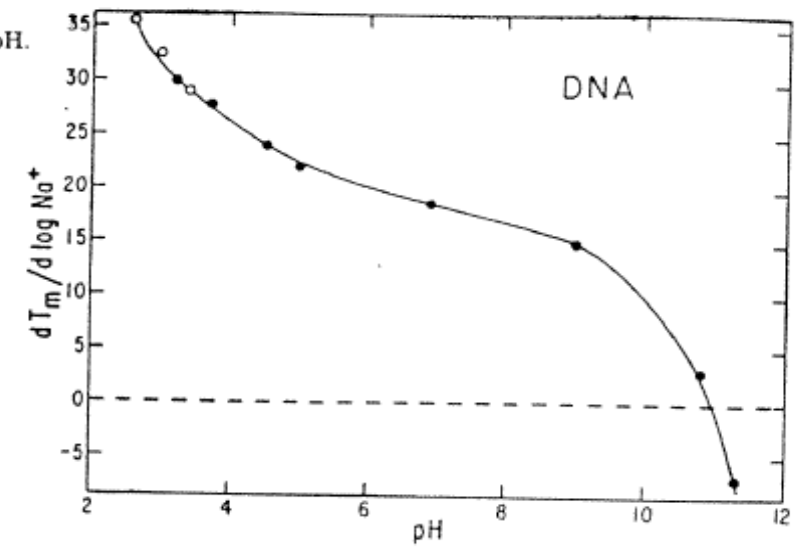


Fig. 1. Variation of dT_m/d log Na^+ with pH. (●), data of Record;[6] (○), obtained by interpolation of the data of Dore et al.[21]

CAN YOU PREDICT THE THERMODYNAMIC STABILITY OF A DNA/RNA DUPLEX?

YES

For example, how do you design DNA oligonucleotides for PCR, site-directed mutagenesis, or sequencing?

There are tables of “Nearest Neighbor” energies that are used by software packages from IDT (for example).

How do they work?

5′**AA**
TT5′

Table 1: Nearest-Neighbor Thermodynamic Parameters for Watson–Crick Base Pair Formation in 1 M NaCl[a]

propagation sequence	ΔH° (kcal/mol)	ΔS° (eu)	ΔG°_{37} (kcal/mol)
AA/TT	−7.9 ± 0.2	−22.2 ± 0.8	−1.00 ± 0.01
AT/TA	−7.2 ± 0.7	−20.4 ± 2.4	−0.88 ± 0.04
TA/AT	−7.2 ± 0.9	−21.3 ± 2.4	−0.58 ± 0.06
CA/GT	−8.5 ± 0.6	−22.7 ± 2.0	−1.45 ± 0.06
GT/CA	−8.4 ± 0.5	−22.4 ± 2.0	−1.44 ± 0.04
CT/GA	−7.8 ± 0.6	−21.0 ± 2.0	−1.28 ± 0.03
GA/CT	−8.2 ± 0.6	−22.2 ± 1.7	−1.30 ± 0.03
CG/GC	−10.6 ± 0.6	−27.2 ± 2.6	−2.17 ± 0.05
GC/CG	−9.8 ± 0.4	−24.4 ± 2.0	−2.24 ± 0.03
GG/CC	−8.0 ± 0.9	−19.9 ± 1.8	−1.84 ± 0.04
init. w/term. G–C[b]	0.1 ± 1.1	−2.8 ± 0.2	0.98 ± 0.05
init. w/term. A–T[b]	2.3 ± 1.3	4.1 ± 0.2	1.03 ± 0.05
symmetry correction	0	−1.4	0.4

[a] Errors are resampling standard deviations (see text). [b] See text for how to apply the initiation parameters.

The experimental observation was that thermodynamic stability of a given base pair depends on its sequence **and** the flanking base pairs (hence the nearest neighbor).

So: Synthesize DNA and RNA oligos, and measure their thermodynamic stability. [How many oligos?]

Measure their enthalpy and calculate the entropy and free energy. Use 1 M NaCl. pH 7.0

RNA experimental thermodynamic parameters for calculations of Nearest-Neighbor Duplex stabilities

Sequence	ΔG^{o}_{37} (kcal/mol)	ΔH^{o} (kcal/mol)	ΔS^{o} [b](eu)
5′AA3′ 3′UU5′	-0.93 (0.03)	-6.82 (0.79)	-19.0 (2.5)
5′AU3′ 3′UA5′	-1.10 (0.08)	-9.38 (1.68)	-26.7 (5.2)
5′UA3′ 3′AU5′	-1.33 (0.09)	-7.69 (2.02)	-20.5 (6.3)
5′CU3′ 3′GA5′	-2.08 (0.06)	-10.48 (1.24)	-27.1 (3.8)
5′CA3′ 3′GU5′	-2.11 (0.07)	-10.44 (1.28)	-26.9 (3.9)
5′GU3′ 3′CA5′	-2.24 (0.06)	-10.40 (1.23)	-29.5 (3.9)
5′GA3′ 3′CU5′	-2.35 (0.06)	-12.44 (1.20)	-32.5 (3.7)
5′CG3′ 3′GC5′	-2.36 (0.09)	-10.64 (1.65)	-26.7 (5.0)
5′GG3′ 3′CC5′	-3.26 (0.07)	-13.39 (1.24)	-32.7 (3.8)
5′GC3′ 3′CG5′	-3.42 (0.08)	-14.88 (1.58)	-36.9 (4.9)
initiation[c]	+4.09 (0.22)	+3.61 (4.12)	-1.5 (12.7)
Per Terminal AU[d]	+0.45 (0.04)	+3.72 (0.83)	+10.5 (2.6)
Symmetry (self-complementary)	+0.43	0	-1.4
Symmetry (Non-self-complementary)	0	0	0

X	Y	XG YU	XU YG
A	U	(-0.5)	-0.7
C	G	-1.5	-1.5
G	C	-1.3	-1.9
U	A	-0.7	(-0.5)
G	U	(-0.5)	(-0.5)
U	G	-0.6	(-0.5)

Numbers in parentheses are uncertainties from experimental data in 1 M NaCl, pH 7.

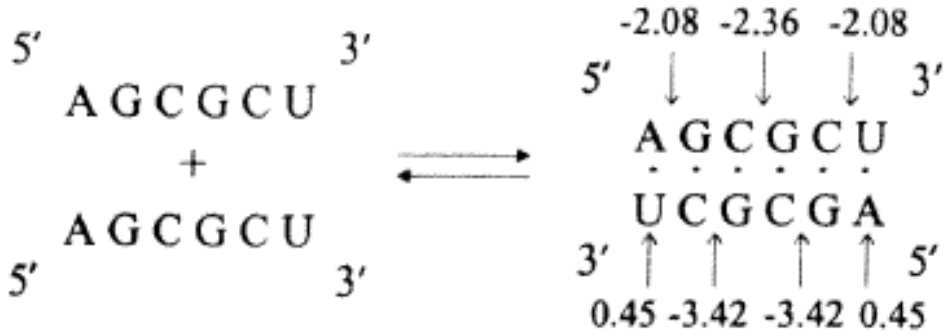
[b] The entropy is calculated from the experimental ΔG°_{37} and ΔH° data.

[c] Initiation of duplex formation is entropically unfavorable. This term includes potential terminal GC effects.

[d] The thermodynamic penalty for each terminal AU pair[13].

AN RNA EXAMPLE

(a)



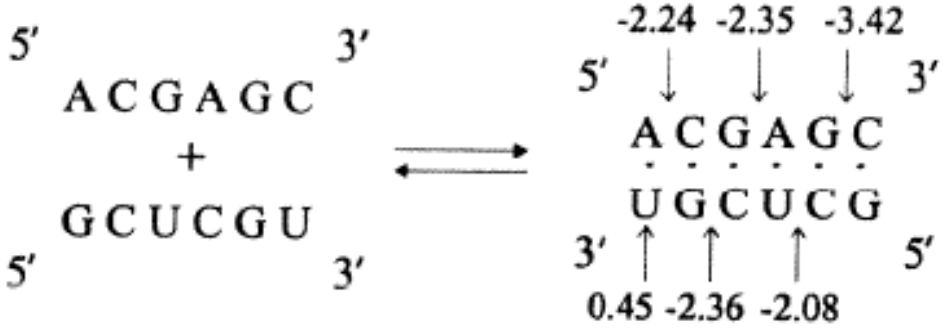
Strands are self-complementary

$$\Delta G^{\circ}_{TOT} = \Delta G^{\circ}_{INIT} + \Delta G^{\circ}_{SYM} + \sum \Delta G^{\circ}_{NN} + 2\Delta G^{\circ}_{TERM-AU}$$

$$= 4.09 + 0.43 + (-13.36) + 2 \times 0.45$$

$$= -7.94 \text{ kcal / mol}$$

(b)



Strands are non self-complementary

$$\Delta G^{\circ}_{TOT} = \Delta G^{\circ}_{INIT} + \Delta G^{\circ}_{SYM} + \sum \Delta G^{\circ}_{NN} + \Delta G^{\circ}_{TERM-AU}$$

$$= 4.09 + 0 + (-12.45) + 0.45$$

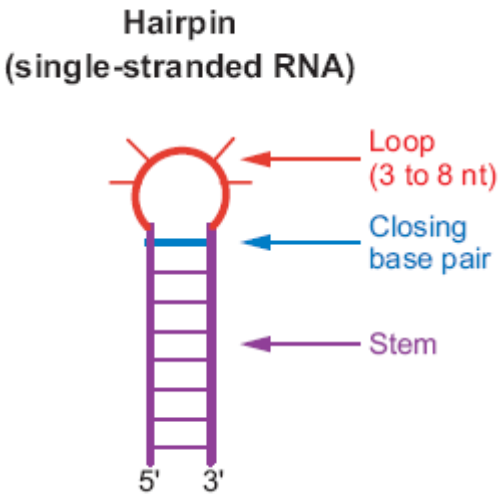
$$= -7.91 \text{ kcal / mol}$$

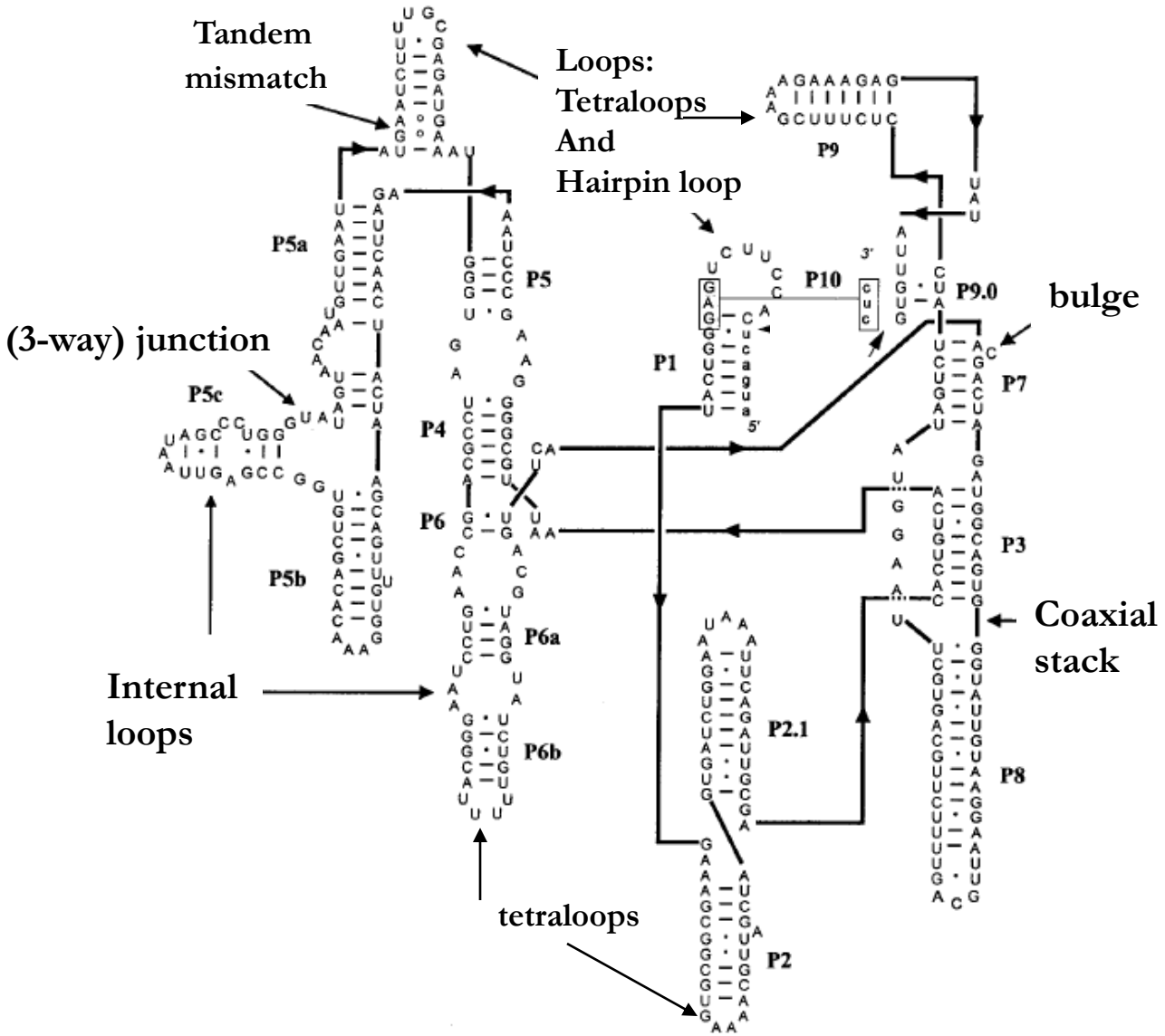
Table 1: Free Energy and T_M Data for Oligonucle

	sequences	T_M (°C) RR[e]	DD	DR	RD	$-\Delta G°_{37}$ (kcal/mol of strands) RR	DD	DR	RD
Group A									
(1)	5′-TCC CTC CTC TCC[d] 3′-AGG GAG GAG AGG	71.2	47.3	61.4	43.4	18.6	10.7	15.6	9.6
(2)	5′-CCT TCC CTT 3′-GGA AGG GAA	52.9	32.6	44.8	20.5	12.2	7.3	10.0	5.8
(3)	5′-TTC CCT TCC 3′-AAG GGA AGG	51.6	28.5	44.2	14.9	11.9	6.5	10.1	5.1
Group B									
(4)	5′-GCT CTC TGG C 3′-CGA GAG ACC G	62.4	43.7	50.9	40.8	15.7	9.5	11.2	8.9
(5)	5′-CTC GTA CCT TCC GGT CC 3′-GAG CAT GGA AGG CCAGG	74.5	60.6	64.8	56.0	21.9	15.7	17.0	12.7
(6)	5′-CTC GTA CCT TTC CGG TCC 3′-GAG CAT GGA AAG GCC AGG	74.2	61.7	65.2	56.8	21.3	16.9	18.1	14.6
(7)	5′-TAG TTA TCT CTA TCT 3′-ATC AAT AGA GAT AGA	52.7	42.1	45.4	34.9	12.4	9.6	10.5	7.5
Group C									
(8)	5′-GCA CAG CC 3′-CGT GTC GG	53.4	32.1	37.2	35.6	12.3	7.5	8.1	7.9
(9)	5′-GAG CTC CCA GGC 3′-CTC GAG GGT CCG	73.7	54.0	60.3	56.7	18.9	12.9	14.3	13.4
(10)	5′-GCC GAG GTC CAT GTC GTA CGC 3′-CGG CTC CAG GTA CAG CAT GCG	80.2	69.5	68.2	68.1	24.1	20.8	17.2	18.0
(11)	5′-TGT ACG TCA CAA CTA 3′-ACA TGC AGT GTT GAT	60.4	53.3	50.6	49.2	15.1	13.0	11.8	11.2
(12)	5′-TAT ACA AGT TAT CTA 3′-ATA TGT TCA ATA GAT	48.4	39.7	35.2	35.9	11.3	8.7	7.7	7.8
Group D									
(13)	5′-CGA CTA TGC AAA AAC 3′-GCT GAT ACG TTT TTG	54.2	50.4	39.0	47.3	13.8	13.1	8.7	11.3
(14)	5′-CGC AAA AAA AAA ACG C 3′-GCG TTT TTT TTT TGC G	51.2	54.1	28.7	50.2	13.1	14.2	5.9	13.0

RR is RNA:RNA DD is DNA:DNA DR upper strand DNA, lower strand RNA. RD upper strand RNA, lower strand DNA

Biological RNAs have duplexes, but they are short.





Experimental parameters for duplex regions are robust, and can be applied to predict stability of folded regions in algorithms such as Viennafold, unifold, or RNAFOLD.
The problem is the contribution of loops to the folding free energy, because those experimental values are soft:

Note that all the free energies of loops are positive and unfavorable

Table 1. Free energy parameters for loops

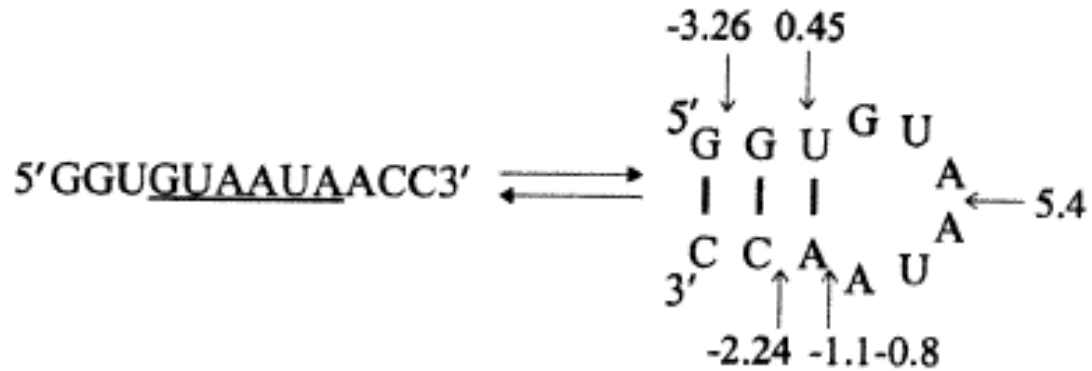
Size, nt	Internal	Bulge	Hairpin
1		+3.9	
2	+4.1	+3.1	
3	+4.5	+3.5	+4.5
4	+4.9	(+4.2)	+5.5
5	(+5.3)	+4.8	+4.9
6	+5.7	(+5.0)	(+5.1)
7	(+5.9)	(+5.2)	+5.2
8	(+6.0)	(+5.3)	(+5.5)
9	(+6.1)	(+5.4)	+5.8
10	(+6.3)	(+5.5)	(+5.9)

Free energy parameters in kcal/mol for RNA loops at 37°C in 1 M NaCl. For larger loop sizes, n, use $\Delta G^\circ(n) = \Delta G^\circ(n_{max}) + 1.75\,RT \ln(n/n_{max})$, where n_{max} is 6, 5, and 9 for internal, bulge, and hairpin loops, respectively. Parameters not derived from experimental measurements are listed in parentheses. Parameters for hairpin loops >3 nt and for internal loops assume additional stability is conferred by stacking of terminal mismatches at helix ends. Asymmetric internal

Calculating hairpin loop folding free energies includes the ΔG° (stem) and ΔG°(loop) based on its number of nucleotides, plus a free energy term from the mismatch flanking the loop-closing base pair, and an extra energy term if that first mismatch is a GA or UU.
These are empirical terms, not theoretical.

.

(c)



$$\Delta G^\circ_{TOT} = \Delta G^\circ_{INIT}(n = 6) + \Delta G^\circ(\text{First Mismatch}) + \Delta G^\circ(\text{Stem})$$

$$= 5.4 + (-1.1 - 0.8) + (-5.05) = -1.55 \text{ kcal/mol}$$

Using thermodynamic basis sets, we can predict the structures of any RNA sequence and compare their free energies. There are several software packages that incorporate the current data bases. (in pubmed, search for **RNA folding programs**; in Google, try **RNA secondary structure prediction**). RNAFOLD, Vienna package.

I recommend RNASTRUCTURE from Dave Mathews' lab.

These parameters have been shown to successfully predict about 80% of the correct structures. BUT: which 80%?

AND: How do you validate the predictions?

miRNAs care about duplex RNA stability