

# Calculating DNA Properties

Biochemistry Boot Camp 2022

Session #4

Chris Johnson

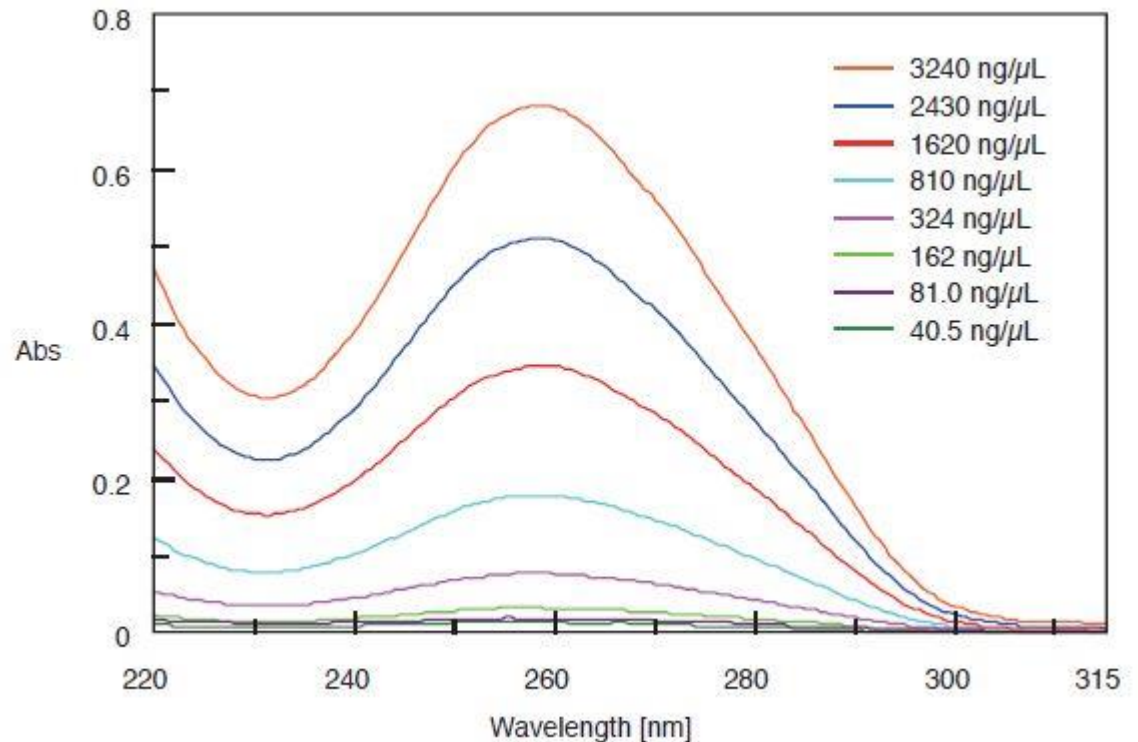
[cn.johnson@chemistry.msstate.edu](mailto:cn.johnson@chemistry.msstate.edu)

# Nucleic Acid Extinction Coefficient

**DNA Concentrations:** Often measured in  $\mu\text{g}/\text{mL}$  (or the equivalent  $\text{ng}/\mu\text{L}$ ) instead of M, mM, etc. Also sequence isn't exactly known in many cases.

**Rule of Thumb:** For double-stranded, plasmid DNA, the extinction coefficient at 260 nm is

$$0.020 (\mu\text{g}/\text{mL})^{-1} \text{cm}^{-1}$$



# DNA vs. Protein Absorbance

**DNA Concentrations:** At 260 nm, double-stranded DNA has an extinction coefficient of

$$0.020 (\mu\text{g/mL})^{-1} \text{ cm}^{-1}$$

**Protein Concentrations:** At 280 nm, the GB3 protein has an extinction coefficient (in equivalent units) of

$$0.0016 (\mu\text{g/mL})^{-1} \text{ cm}^{-1}$$

Which is more sensitive?

What are the implications?

# Other Values for Long, Randomized Sequences

- Single-Stranded RNA:  $0.025 (\mu\text{g}/\text{mL})^{-1} \text{ cm}^{-1}$
- Single-Stranded DNA:  $0.030 (\mu\text{g}/\text{mL})^{-1} \text{ cm}^{-1}$
- For a pure nucleic acid, the 260/280 nm ratio should be approximately 1.8-2.0

# Nucleic Acids – Smaller Molecules

## OligoAnalyzer

Instructions | Definitions | Feedback

Sequence 5' MOD ▾ INTERNAL ▾ 3' MOD ▾ MIXED BASES ▾

CGA AGA ACA GGA AGC GGA ATT TAA AGA AG

Bases 29 CLEAR SEQUENCE

[Try the new batch mode here](#)

Parameter sets SpecSheet (Default) ▾

Target type DNA ▾

Oligo Conc 0.25  $\mu\text{M}$

Na<sup>+</sup> Conc 50 mM

Mg<sup>++</sup> Conc 0 mM

dNTPs Conc 0 mM

**ANALYZE**

HAIRPIN

SELF-DIMER

HETERO-DIMER

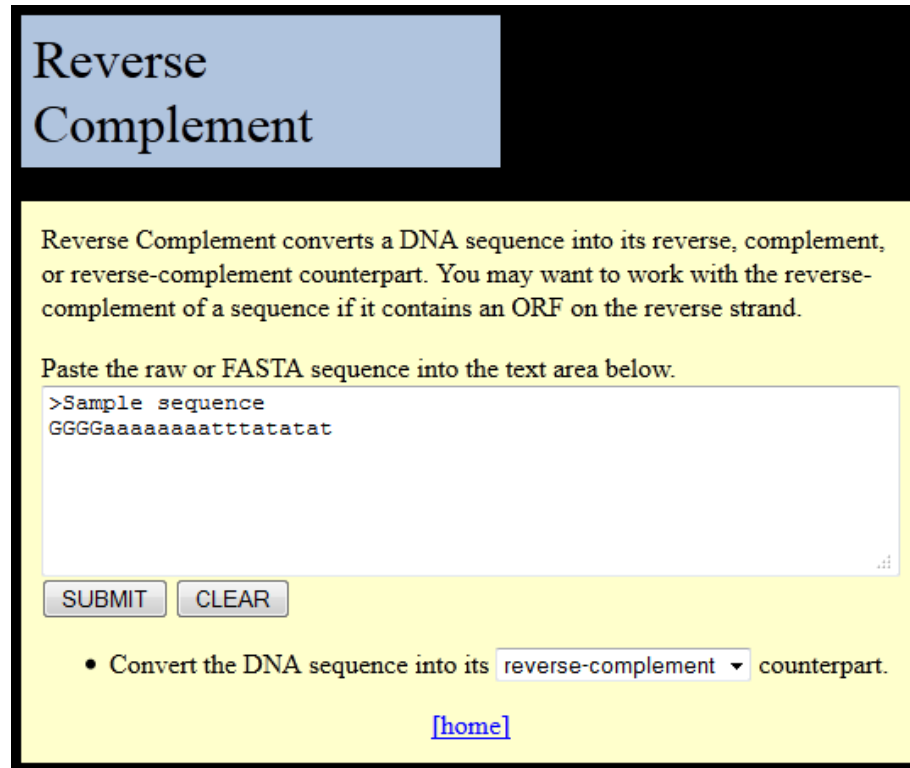
NCBI BLAST

TM MISMATCH

**ADD TO ORDER**

- IDT DNA Analyzer (extinction coefficient, T<sub>m</sub>):  
<https://www.idtdna.com/pages/tools/oligoanalyzer>  
– Need to log in, create an account (free)

# Calculating Reverse Complement



The screenshot shows a web interface for calculating the reverse complement of a DNA sequence. The title is "Reverse Complement". Below the title, there is a descriptive paragraph: "Reverse Complement converts a DNA sequence into its reverse, complement, or reverse-complement counterpart. You may want to work with the reverse-complement of a sequence if it contains an ORF on the reverse strand." Below this is a text area with the instruction "Paste the raw or FASTA sequence into the text area below." The text area contains the sample sequence: ">Sample sequence\nGGGGaaaaaaaaatttatatat". Below the text area are two buttons: "SUBMIT" and "CLEAR". Below the buttons is a radio button option: "Convert the DNA sequence into its reverse-complement counterpart." The "reverse-complement" option is selected. At the bottom, there is a link labeled "[home]".

- Bioinformatics.org Calculator (no-frills):  
[http://bioinformatics.org/sms/rev\\_comp.html](http://bioinformatics.org/sms/rev_comp.html)

# DNA Translation Tool

- **Site:**

<http://web.expasy.org/translate/>

- **Input:** DNA or RNA sequence (5' → 3' orientation)

- **Output:** All six possible translation frames

# Other Databases

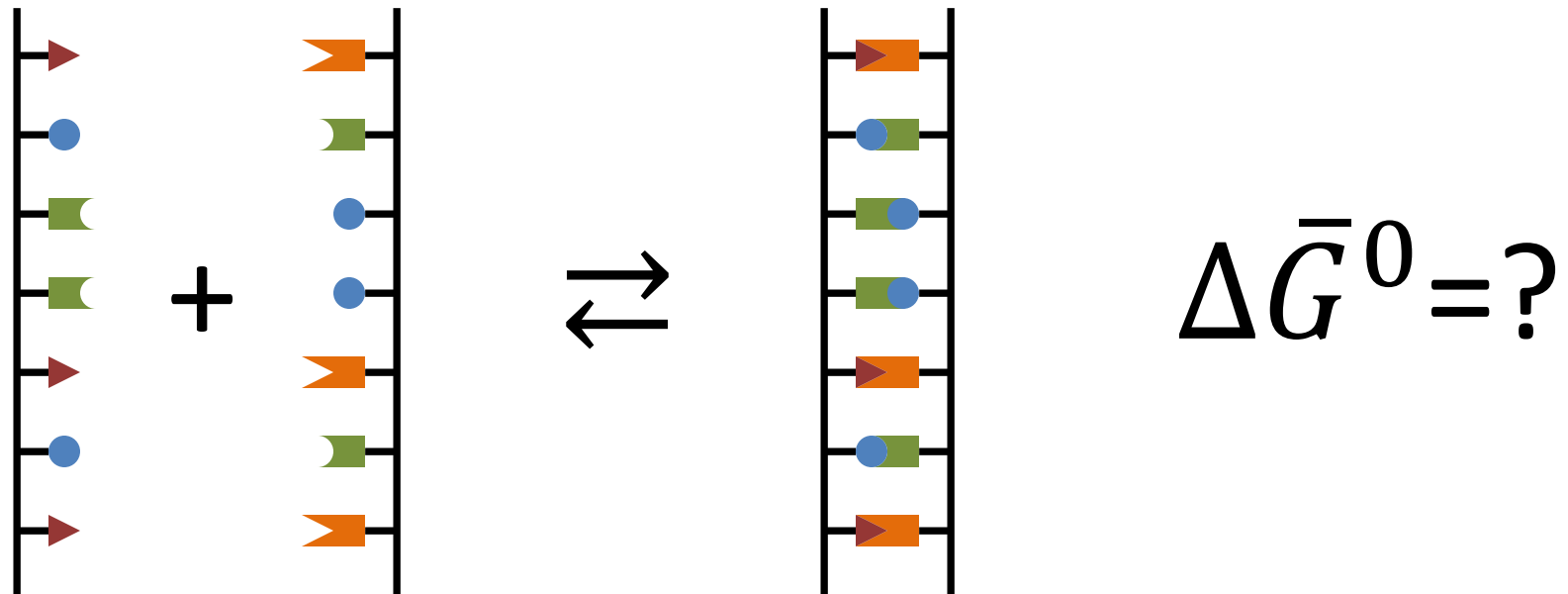
- NCBI Databases work for DNA sequences, too (reference sequences start with NM\_)
- PDB also houses a number of RNA/DNA structures in addition to proteins



## *Think And Discuss*

How can these databases be used to make your lab work easier? What are some practical examples?

# DNA “Melting”



- Two strands come together:
  - How much work can be done?
  - Which side of the reaction does temperature favor?

# Thermal Melts

- Adding heat favors highly random systems, DNA will separate at high temperature
  - Secondary and tertiary structure is lost, primary is maintained
- What will affect the melting temperature?

# Predicting Melting Temperatures

- To calculate  $T_m$ , add 4 °C for each G-C pair, and 2 °C for each A-T
  - Not terribly accurate
- **Example:** GCCCTGAAGGTCAAGTCCCCC
  - 14 G-C = 56 °C
  - 7 A-T = 14 °C
  - Prediction is 70

# Predicting Melting Temperatures

- IDT OligoAnalyzer:  
<https://www.idtdna.com/pages/tools/oligoanalyzer>
- **Input:** Your DNA sequence of interest, salt concentration
- **Output:**  $T_m$ , extinction coefficient, %GC content

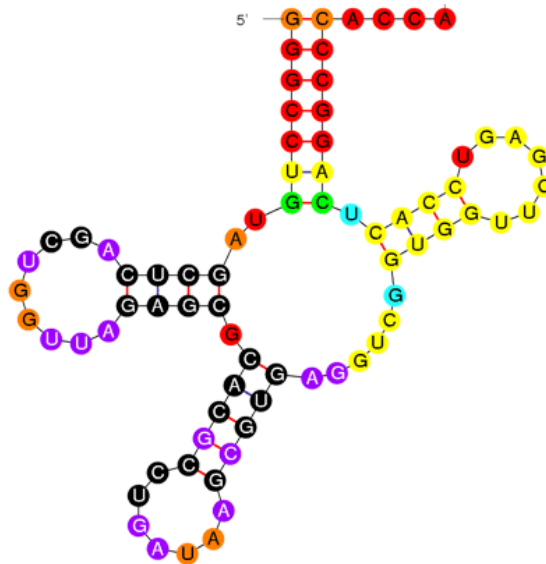
# Predicting Secondary Structure

- mfold Web Server:

<http://mfold.rna.albany.edu/?q=mfold>

- **Input:** RNA/DNA sequence

- **Output:**



# Example: HIV TAR RNA

- Trans-Activation Response Element – Binds with a protein (Tat) to promote viral transcription
- Sequence:

GGGUCUCUCUGGUUAGACCAGAUCUGAGCCUGGGAGCUCUCU  
GGCUAACUAGGGAACCCAC

# Why is this Useful?

- Site-Directed Mutagenesis
- Good Primers:
  - $T_m > 78\text{ }^\circ\text{C}$  (2 mM  $\text{MgCl}_2$ , 50 mM NaCl)
  - GC content  $> 40\%$
  - No secondary structure (< 50 bp)
  - End with G or C



The Nobel Prize in Chemistry 1993

Kary B. Mullis, Michael Smith

The Nobel Prize in Chemistry 1993

Nobel Prize Award Ceremony

Kary B. Mullis

Michael Smith



Kary B. Mullis



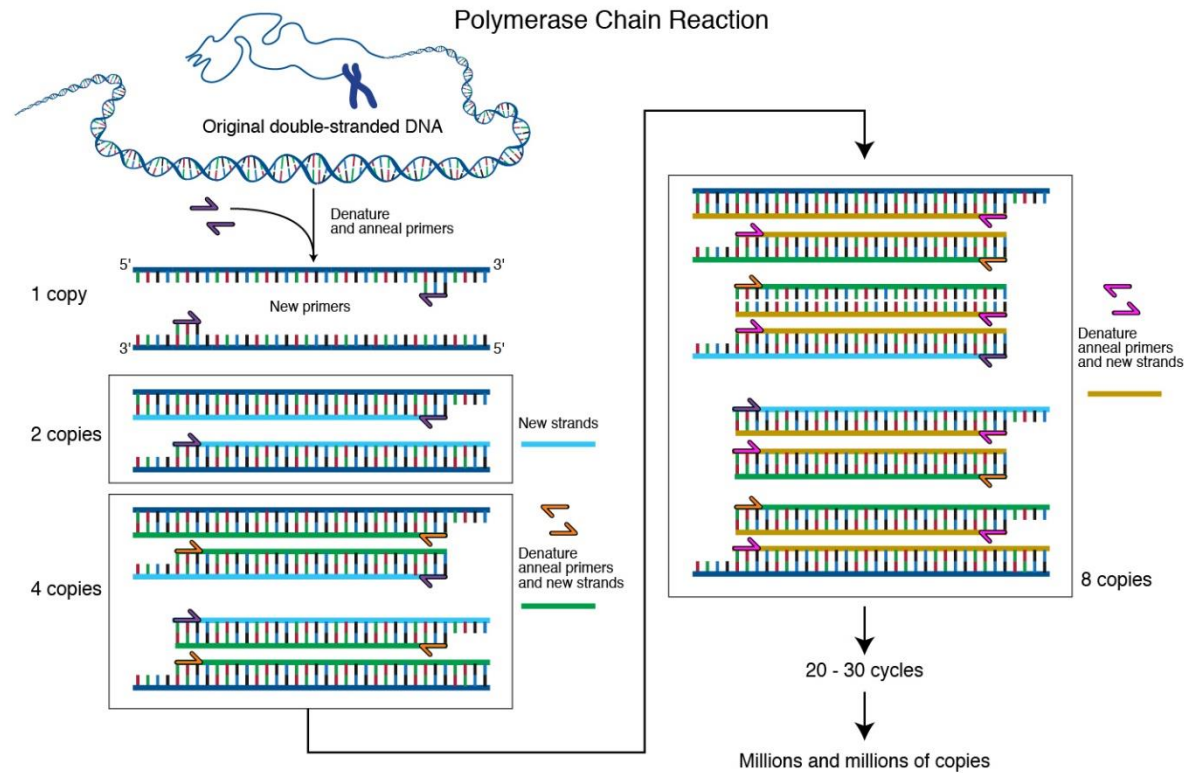
Michael Smith

The Nobel Prize in Chemistry 1993 was awarded "for contributions to the developments of methods within DNA-based chemistry" jointly with one half to Kary B. Mullis "for his invention of the polymerase chain reaction (PCR) method" and with one half to Michael Smith "for his fundamental contributions to the establishment of oligonucleotide-based, site-directed mutagenesis and its development for protein studies".

Photos: Copyright © The Nobel Foundation




# Polymerase Chain Reaction (PCR)



- **Key consideration:** Temperature for primer annealing (computational tools)!

# Example: Designing Primers



The screenshot shows the Agilent Technologies Genomics website. The header includes the Agilent Technologies logo, the word "Genomics", and navigation links for "English", "United States", "Contact Us", "Logout", "My Account", "Quick Order", and "Cart (0)". Below the header is a navigation bar with "Home", "Products & Services", "Applications", and "Resources & Support". A search bar is located on the right side of the navigation bar. The main content area features a breadcrumb trail: "Overview | General Info | Links | BioCalculators | QuikChange Primer Design". The title "QuikChange Primer Design" is displayed in a large, bold, red font. Below the title is a blue button labeled "? Help". The text below the button reads: "The QuikChange® Primer Design Program supports mutagenic primer design for your QuikChange mutagenesis experiments. Using primer design guidelines described in QuikChange manuals, this program calculates/designs the appropriate primer sequences with the optimal melting temperature. Read [Help](#) for more information about the program." Below this text is a blue link labeled "Expand Help" with a plus icon. The first step of the instructions is: "1. It is recommended that you clear this form prior to loading each new sequence:"



1. It is recommended that you clear this form prior to loading each new sequence:

Clear Input

2. Select QuikChange® mutagenesis kit that you are using:

QuikChange® II

Help in choosing a mutagenesis kit

3. Find your DNA sequence by pressing

Load a file from your hard-drive:

No file selected

Browse...

Or, paste plain text or FASTA-formatted DNA sequence in the box below:

atagatatagacagacatacatagacatacatagaggaggatacacag  
acatacatagacaga

4. Load it.

Upload Now

- or -

Upload Translated

or specify a DNA region to translate from

to

Upload Translated Region

5. Select up to seven nucleotides that you want to change

DNA: change nucleotide(s) to: Site 1: g Site 2: - Site 3: - Site 4: - Site 5: - Site 6: - Site 7: -

or

Delete a region between two checked nucleotides (Note: two checked nucleotides will not be deleted)

or

Insert between two checked nucleotides

- Grid of 65 checkboxes for nucleotide positions 1-65. Position 24 is checked.

Finally, obtain your designed primer sequences.

Design Primers

Clear this form to load a new sequence:

Clear Input



Insert between two checked nucleotides

- |                               |                               |                               |                               |                               |  |                               |                               |                               |
|-------------------------------|-------------------------------|-------------------------------|-------------------------------|-------------------------------|--|-------------------------------|-------------------------------|-------------------------------|
| <input type="checkbox"/> 1 a  | <input type="checkbox"/> 2 t  | <input type="checkbox"/> 3 a  | <input type="checkbox"/> 4 g  | <input type="checkbox"/> 5 a  | <input type="checkbox"/> 6 t             | <input type="checkbox"/> 7 a  | <input type="checkbox"/> 8 t  | <input type="checkbox"/> 9 a  |
| <input type="checkbox"/> 10 g | <input type="checkbox"/> 11 a | <input type="checkbox"/> 12 c | <input type="checkbox"/> 13 a | <input type="checkbox"/> 14 g | <input type="checkbox"/> 15 a            | <input type="checkbox"/> 16 c | <input type="checkbox"/> 17 a | <input type="checkbox"/> 18 t |
| <input type="checkbox"/> 19 a | <input type="checkbox"/> 20 c | <input type="checkbox"/> 21 a | <input type="checkbox"/> 22 t | <input type="checkbox"/> 23 c | <input checked="" type="checkbox"/> 24 a | <input type="checkbox"/> 25 g | <input type="checkbox"/> 26 a | <input type="checkbox"/> 27 c |
| <input type="checkbox"/> 28 a | <input type="checkbox"/> 29 t | <input type="checkbox"/> 30 a | <input type="checkbox"/> 31 t | <input type="checkbox"/> 32 a | <input type="checkbox"/> 33 c            | <input type="checkbox"/> 34 a | <input type="checkbox"/> 35 g | <input type="checkbox"/> 36 g |
| <input type="checkbox"/> 37 a | <input type="checkbox"/> 38 g | <input type="checkbox"/> 39 g | <input type="checkbox"/> 40 a | <input type="checkbox"/> 41 g | <input type="checkbox"/> 42 g            | <input type="checkbox"/> 43 a | <input type="checkbox"/> 44 t | <input type="checkbox"/> 45 a |
| <input type="checkbox"/> 46 c | <input type="checkbox"/> 47 a | <input type="checkbox"/> 48 c | <input type="checkbox"/> 49 a | <input type="checkbox"/> 50 g | <input type="checkbox"/> 51 a            | <input type="checkbox"/> 52 c | <input type="checkbox"/> 53 a | <input type="checkbox"/> 54 t |
| <input type="checkbox"/> 55 c | <input type="checkbox"/> 56 a | <input type="checkbox"/> 57 t | <input type="checkbox"/> 58 a | <input type="checkbox"/> 59 c | <input type="checkbox"/> 60 a            | <input type="checkbox"/> 61 g | <input type="checkbox"/> 62 c | <input type="checkbox"/> 63 a |
| <input type="checkbox"/> 64 g | <input type="checkbox"/> 65 a |                               |                               |                               |  |                               |                               |                               |

Finally, obtain your designed primer sequences.

Design Primers

Clear this form to load a new sequence:

Clear Input

### Primer sequences:

Primer Name	Primer Sequence (5' to 3')
a24c_	5'-atcctcctcctgtatattgctggatgtatgtctgtctatatc-3' 5'-gatatagacagacatacatccgacatatacaggaggaggat-3'

### Oligonucleotide information:

Primer Name	Length (nt.)	Tm	Duplex Energy at 68 °C	Energy Cost of Mismatches
a24c_	41	78.60°C	-45.55 kcal/mole	3.05%
a24c_	41	78.60°C	-49.02 kcal/mole	3.83%

### Primer-template duplexes:

Primer Name	Primer-Template Duplex
a24c_	<pre> gatatagacagacatacatcagacatatacaggaggaggatatac       3'-ctatatctgtctgtatgttaggctgtatattgtcctcctccta-5' 5'-gatatagacagacatacatccgacatatacaggaggaggat-3'       ???ctatatctgtctgtatgttagtctgtatattgtcctcctcctatgt </pre>



# General Primer Design Principles

- **PCR Steps:** Denature (95 °C), anneal (60 °C), extend (70 °C)
- **Considerations:**
  - **Melting Temperature:** Should be 52-58 °C
  - **GC Content:** 40-60%
  - **Length:** ~30 bp (but longer can be okay)
  - **Secondary Structure:** Avoid if possible
- Lots of software exists (some costs \$\$\$). For more information (some trial and error here):
  - <https://goo.gl/4EwMG3> (Life Technologies)
  - [http://www.premierbiosoft.com/tech\\_notes/PCR\\_Primer\\_Design.html](http://www.premierbiosoft.com/tech_notes/PCR_Primer_Design.html)

# Example: General Approach

5' 3'

## Exact and Single Base Mismatch DNA Thermodynamics

Primary Sequence: 5' to 3'; Target Sequence: 3' to 5'

Additional Target Base

5'- CGAAGAACAGGAAGCGGAATTTAAAGAAG  
|||||  
3'- GCTTCTTG  CCTTCGCCTTAAATTTCTTC

Additional Target Base

-3'  
-5'

ANALYZE  
HAIRPIN  
SELF-DIMER  
HETERO-DIMER  
NCBI BLAST  
TM MISMATCH  
ADD TO ORDER

Hybridization Temperature

Use Exact Complement  $T_m$   
 User Defined

Target Concentration

# Other Calculators: GeneRunner

- Download (free) from <http://generunner.net/>
- Analysis → Oligo brings up window to the right
- Suggested options shown to the right
  - SantaLucia temperature & salt recommended
  - Check your specific dNTP, DNA, Mg<sup>2+</sup> and salt concentrations

The screenshot shows the GeneRunner software interface for oligo analysis. The main window displays the sequence: CAACTACTATCGCTGTTGATGCTGAAAC. Below the sequence, the analysis results are shown, including Mol Wt: 8547.6, Tm: 62.8 (circled in red), Filter Tm: 55.2, %GC Tm: 53.4, GC+AT Tm: 80.0, nMol/A260: 3.8, ug/A260: 32.2, %GC: 42.9, dG: -33.6, dH: -202.7, dS: -527.0, and 3'-end dG: -5.5. The parameters section shows dNTP con (milli Mol): 0.60, DNA con (nano Mol): 50.00, Salt con (milli Mol): 50.00, Divalent con (milli Mol): 1.50, 3'-end len: 7, Base run >=: 4, Pal len >=: 8, Stem len >=: 3, and Guidance: [empty]. The TM Method section shows SantaLucia selected. The Salt Correction section shows SantaLucia selected. The Force Short Tm Calcs checkbox is unchecked. The Max Len: 35. The bottom section shows 1 of 4 oligos, with Tm: 42.3 and dG: -0.8. The bottom right section shows Hairpin loops selected, and the sequence analysis results: 5' CAACTACTA, 3' CAAAGTCGTAGTTGCTCGCT, STEM AT 1 IS 4 BP LONG, LOOP = 10.

# Other Calculators: Phusion Calculator

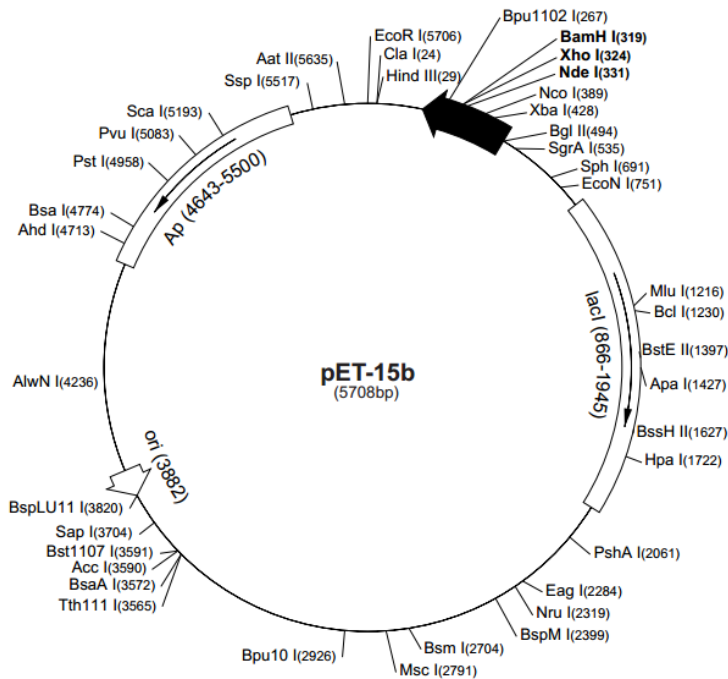
The screenshot shows the Phusion Calculator interface with the following details:

- Product Group:** Phusion
- Polymerase/Kit:** Phusion High-Fidelity DNA Polymerase (HF Buffer)
- Primer Concentration (nM):** 500 (with a "Reset concentration" button)
- Primer 1:** CAACTACTATCGCTGTTGATGCTGAAAC
- Primer 2:** GCATCAACAGCGATAGTAGTTGTTTCG
- Buttons:** "Switch to batch mode", "Clear", and "Use example input"
- Results:**
  - Anneal at: 67 °C (with a link "Why is this so high?")
  - Primer 1: 28 nt, 43% GC, Tm: 64°C
  - Primer 2: 28 nt, 46% GC, Tm: 66°C

- Some polymerases have their own calculator specific to buffers, recommendations
- Phusion Polymerase found at NEB website (<https://tmcalculator.neb.com/#!/main>)



# GB3 Protein and Expression Plasmid (pET-15b)



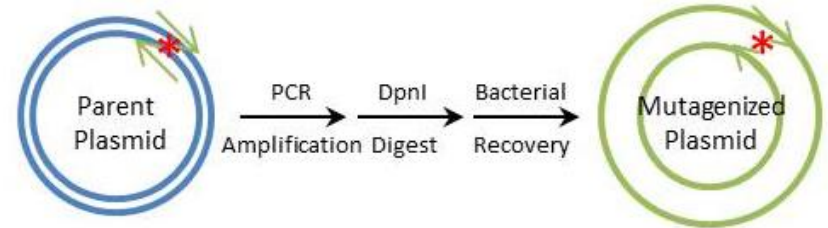
WT GB3 DNA/Protein Sequence:

```

(-6) L * E G D I H M Q Y K L V I N G K T L K 13
1   c t t t a a g a a g g a t a t a c a t a t g c a g t a c a a a t t a g t t a t c a a t g g t a a a c a t t g a a a   60
14  G E T T T K A V D A E T A E K A F K Q Y 33
61  g g c g a a a c a a c t a c t a a a g c t g t t g a t g c t g a a a c t c a g a a a a a g c t t t c a a c a a t a c   120
34  A N D N G V D G V W T Y D D A T K T F T 53
121 g c t a a c g a c a a c g g t a t t g a c g g t a t t g g a c t t a c g a c g a t g c g a c t a a g a c c t t t a c a   180
54  V T E * D P A A N K A
181 g t t a c t g a a t a g g a t c c g g c t g c t a a c a a a g c c   213
  
```

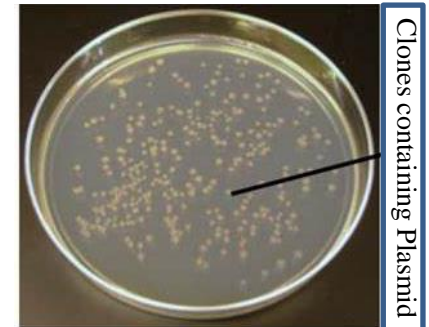
# Successful Mutagenesis Primers for GB3 K19I

Primer Name	Primer Sequence (5' to 3')
K19I-Fwd	CAACTACTATCGCTGTTGATGCTGAAAC
K19I-Rev	GCATCAACAGCGATAGTAGTTGTTTCGC



Description	Primer Duplex
K19I -Fwd/ Template	<p>5' -CAACTACTATCGCTGTTGATGCTGAAAC</p> <p>   </p> <p>3' -TAACTTCCGCTTTGTTGATGATTCGACAACACTACGACTTTGACGT</p>
K19I -Rev/ Template	<p>CATTGAAAGGCGAAACAACACTACTAAAGCTGTTGATGCTGAAACTGCAGAAA</p> <p>   </p> <p>CGCTTTGTTGATGATAGCGACAACACTACG</p>
K19I -Fwd/ K19I -Rev	<p>CAACTACTATCGCTGTTGATGCTGAAAC</p> <p>CGCTTTGTTGATGATAGCGACAACACTACG</p>

After Transformation:



- Annealing temperature used: 53 °C

## *Think And Discuss*

Compared to DNA, why is it harder to calculate melting temperature and dimerization for proteins?

# Example: Sequence Analysis of SH3 Mutants

- Step 1: Design Primers (for T22G)
  - Agilent Web Program (we'll do this)
- Step 2: Do experiments, get sequence of result
- Step 3: Check sequence to see if mutation was successful (we'll do this)

## *Think and Discuss*

What problems could arise when introducing new mutations in to a known sequence?

# Summary

- Advanced computational tools for nucleic acids depend on two things:
  - The simplicity of DNA primary structure (4 bases)
  - The regularity of Watson-Crick base pairing
- Combining DNA and protein tools makes it possible to perform very advanced sequence analysis