

DNA Structure and Properties

Biochemistry Boot Camp
Session #6
Dinesh Yadav
dky12@msstate.edu

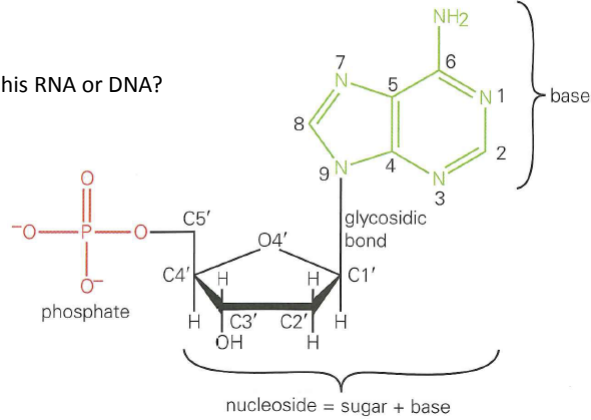
DNA

- DNA- a polymer of deoxyribonucleotides
- Found in chromosomes, mitochondria and chloroplasts
- Carries the genetic information

Nucleic Acid Structure

nucleotide = nucleoside + phosphate

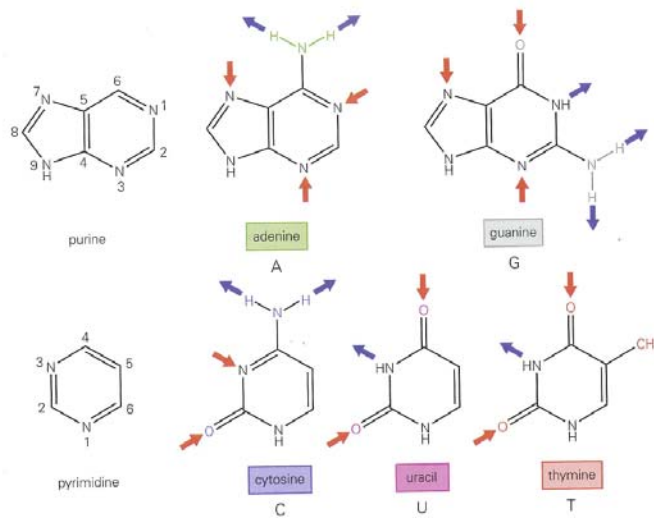
Question: Is this RNA or DNA?



Molecules of Life, pp. 15

3

Nucleic Acid Bases



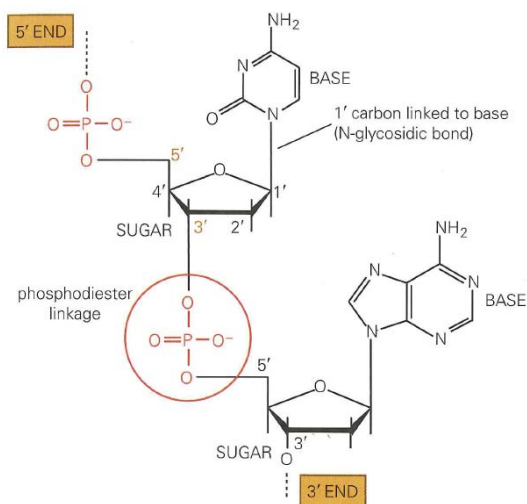
Molecules of Life, pp. 20

4

Nomenclature

	Base	Nucleoside	Nucleotide	Nucleic Acid
Purine	Adenine	Adenosine	Adenylate	RNA
		Deoxy adenosine	Deoxyadenylate	DNA
	Guanine	Guanosine	Guanylate	RNA
		Deoxyguanosine	Deoxyguanylate	DNA
Pyrimidines	Cytosine	cytidine	Cytidylate	RNA
		Deoxy cytidine	Deoxycytidylate	DNA
	Thymine	thymidine	Thymidylate	
		deoxy thymidine	Deoxythymidylate	DNA
Uracil	uridine	Uridylate	RNA	

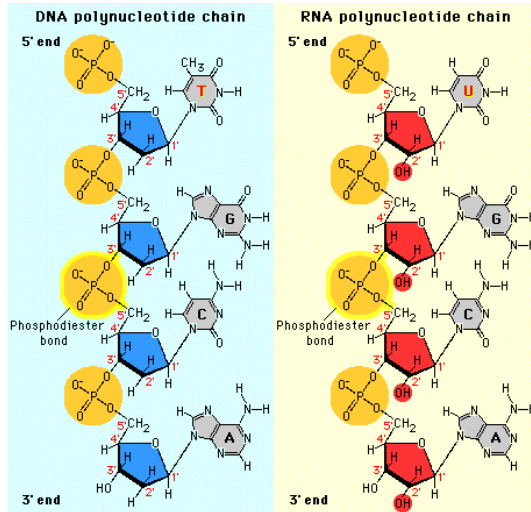
Nucleic Acids Are Also Polymers



DNA & RNA Polymerase: Build up DNA and RNA from nucleoside triphosphates (5' → 3' synthesis)

Convention: RNA/DNA typically is read from 5' to 3' direction (e.g. 5'-ATTGCAAC-3')

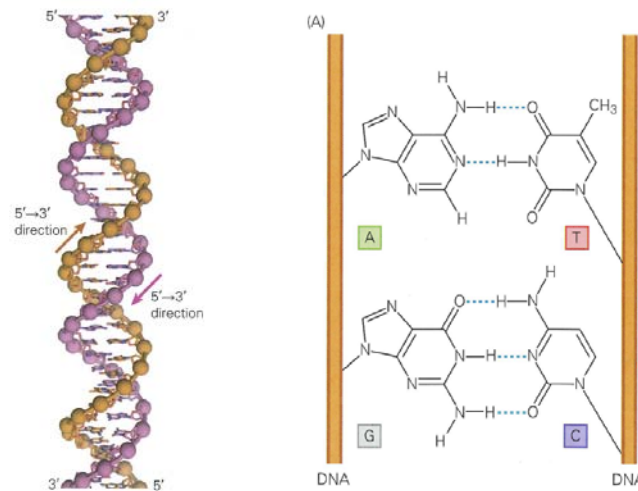
DNA vs RNA



- DNA less reactive
- RNA is easily attacked by enzymes

Science, www.phschool.com (Accessed on June 02, 2014)

Watson-Crick Base Pairing in an (Antiparallel) Double Helix



Molecules of Life, pp. 23

8

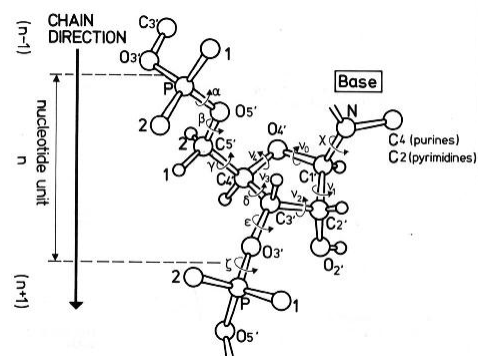
Reverse Complement

- Watson-Crick base pairing
 - A pairs with T (or U in RNA)
 - G pairs with C
- RNA can “hybridize” with DNA, forming mixed strands
- **Example:** What’s the reverse complement to AUCCGCCTT?

9

Structure in DNA

- Bases are planar
- Torsion angles are shown
 - Much more complex than proteins



Saenger, W. *Principles of Nucleic Acid Structure*.

10

Nucleic Acid Primary Structure

- **Just like proteins:** the sequence of bases

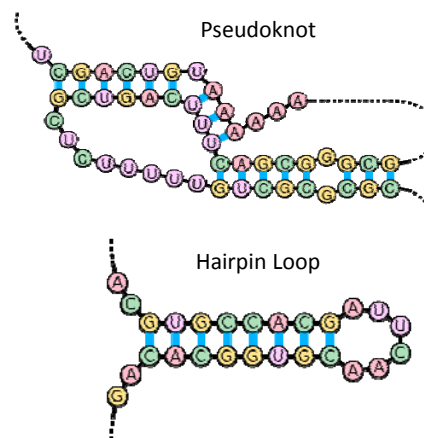
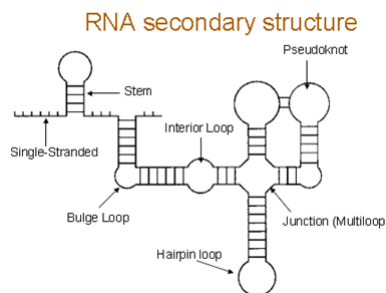
5'-dAdGdTdTdCdAdCdCdC-3' (DNA)

AGTTCACCC

5'-AGUUCACCC-3' (RNA)

11

Secondary Structure

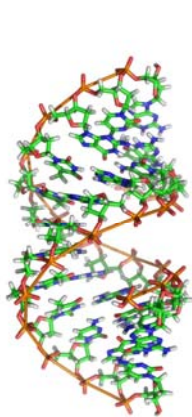


- Base pairing motifs

Source: Wikipedia, "RNA Secondary Structure," "Nucleic Acid Secondary Structure"

12

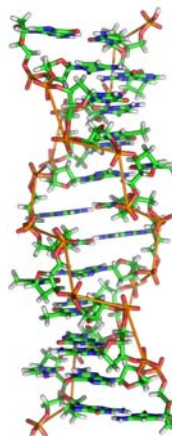
Tertiary Structure



A Form DNA



B Form DNA



Z Form DNA

Source: Steven Carr, www.mun.ca

13

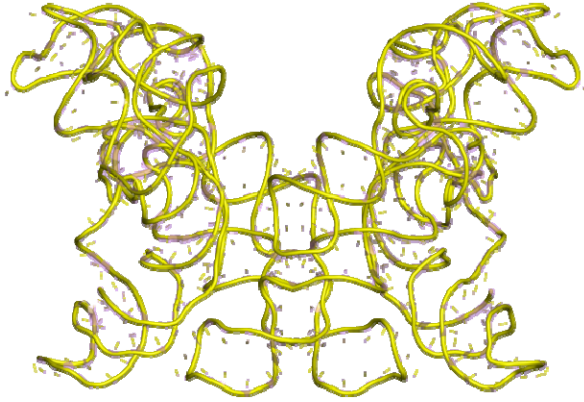
Tertiary Structure

Structure Type	Average Torsion Angles for Nucleic Acid Helices (in °)						
	Alpha	Beta	Gamma	Delta	Epsilon	Zeta	Chi
A-DNA (fibres)	-50	172	41	79	-146	-78	-154
GGCCGGCC	-75	185	56	91	-166	-75	-149
B-DNA (fibres)	-41	136	38	139	-133	-157	-102
CGCGAATTCGCG	-63	171	54	123	-169	-108	-117
Z-DNA (C residues)	-137	-139	56	138	-95	80	-159
Z-DNA (G residues)	47	179	-169	99	-104	-69	68
DNA-RNA decamer	-69	175	55	82	-151	-75	-162
A-RNA	-68	178	54	82	-153	-71	-158

Blackburn and Galt. *Nucleic acids in chemistry and biology.*

14

Tertiary & Quaternary Structure



Ribozyme: An RNA capable of catalyzing a chemical reaction

The ribosome contains a significant amount of RNA as well as proteins

Macromolecules can perform incredibly diverse structures!
(And we haven't even mentioned lipids and sugars.)

Wikipedia, "Group I Catalytic Intron." Accessed 8/23/2012.

15

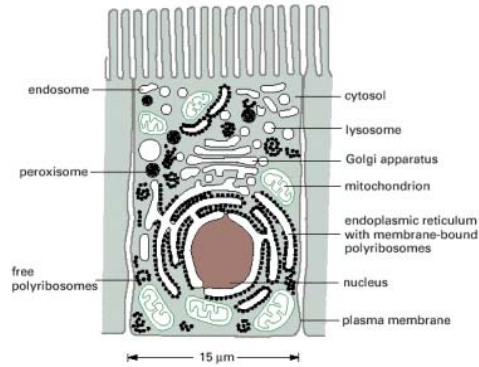
Think and Discuss

What are the major differences between DNA and protein structures? What are the similarities?

16

Review of Intro Biology

- Parts of a eukaryotic animal cell
- Has a nucleus where DNA is stored
- Membrane-bound organelles

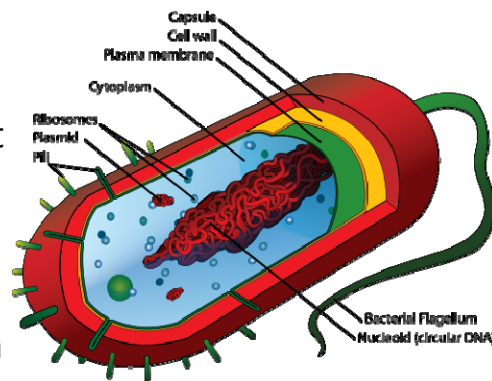


Alberts, et al. *Molecular Biology of the Cell*, 4th Edition.

17

Review of Intro Biology

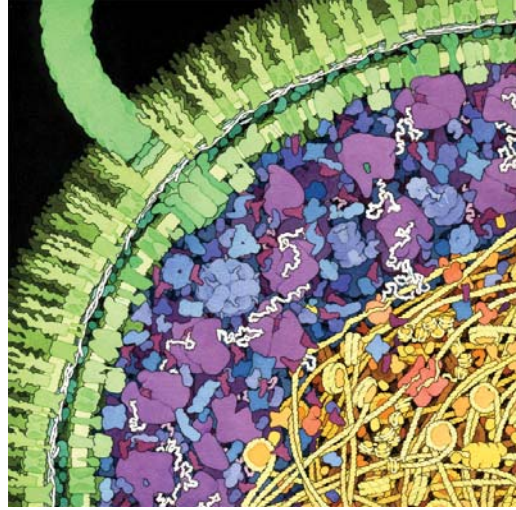
- Parts of a prokaryotic bacterial cell
- No nucleus: DNA is not linear but circular (no ends)
- No organelles, but ribosomes, etc. exist in the cytoplasm



Source: Wikipedia, "Bacterial Cell Structure."

18

It's Crowded in There!

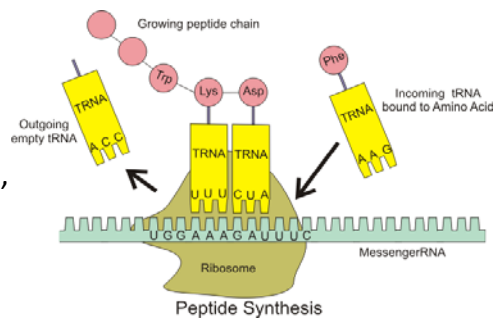


Source: Goodsell, D. <http://mgl.sripps.edu/people/goodsell/illustration/public/>

19

Central Dogma

- DNA → mRNA
“Transcription”
 - Synthesized RNA Polymerase
 - RNA formed from 5' to 3'
- mRNA → Protein
“Translation”
 - Synthesized by ribosome
 - New proteins formed from NT to CT



Trick: Reading the DNA in the “standard way”, one can easily identify the codons for peptide synthesis.

Source: Wikipedia, “Ribosome”

Genetic Code

nonpolar polar basic acidic (stop codon)

Standard genetic code

1st base	2nd base				3rd base
	U	C	A	G	
U	UUU (Phe/F) Phenylalanine	UCU	UAU (Tyr/Y) Tyrosine	UGU (Cys/C) Cysteine	U
	UUC	UCC (Ser/S) Serine	UAC	UGC	C
	UUA	UCA	UAA Stop (Ochre)	UGA Stop (Opal)	A
	UUG	UCG	UAG Stop (Amber)	UGG (Trp/W) Tryptophan	G
C	CUU (Leu/L) Leucine	CCU	CAU (His/H) Histidine	CGU	U
	CUC	CCC (Pro/P) Proline	CAC	CGC (Arg/R) Arginine	C
	CUA	CCA	CAA (Gln/Q) Glutamine	CGA	A
	CUG	CCG	CAG	CGG	G
A	AUU (Ile/I) Isoleucine	ACU	AAU (Asn/N) Asparagine	AGU (Ser/S) Serine	U
	AUC	ACC (Thr/T) Threonine	AAC	AGC	C
	AUA	ACA	AAA (Lys/K) Lysine	AGA (Arg/R) Arginine	A
	AUG ^[A] (Met/M) Methionine	ACG	AAG	AGG	G
G	GUU (Val/V) Valine	GCU	GAU (Asp/D) Aspartic acid	GGU (Gly/G) Glycine	U
	GUC	GCC (Ala/A) Alanine	GAC	GGC	C
	GUA	GCA	GAA (Glu/E) Glutamic acid	GGA	A
	GUG	GCG	GAG	GGG	G

Source: Wikipedia, "Genetic Code"

21

Different Reading Frames

```

reading frame:           123
                        |||
                        acttaccgggacta
first reading frame      T Y P G L
second reading frame    L T R D
third reading frame     L P G T
    
```

Source: <http://www.ncbi.nlm.nih.gov/Class/MLACourse/Original8Hour/Genetics/readingframe.html>

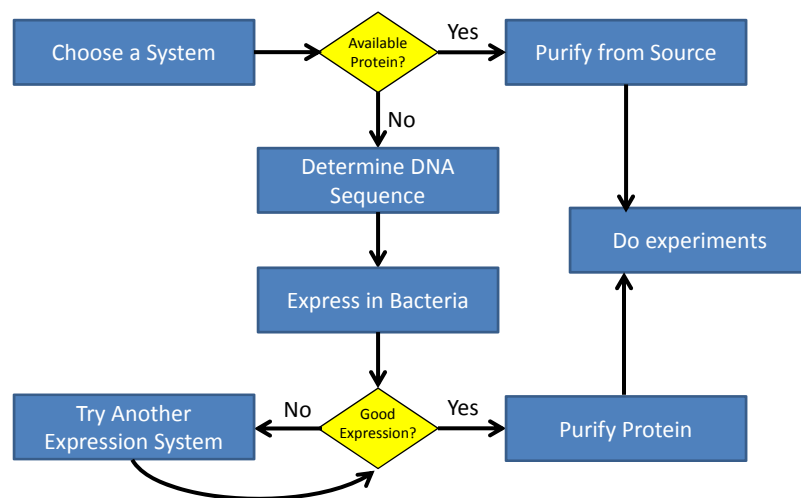
22

Think and Discuss

Our biochemistry experiments are normally done in aqueous buffer. Is this a good model for the inside of a cell?

23

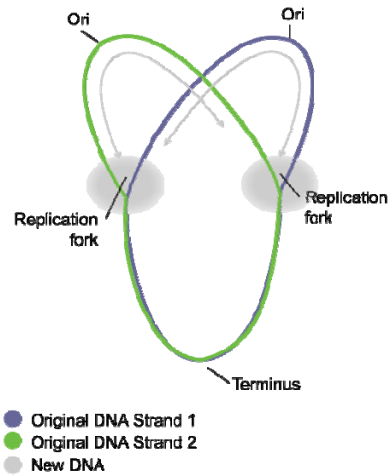
Biochemistry Research Flow Chart



24

Bacterial DNA: Features

- Chromosome is *circular*
- Replication starts at the *origin of replication* (Ori, TTATCCACA)
- **Plasmid:** Any circular DNA in the bacterial cell can be replicated if it has an Ori



Source: Wikipedia, "Circular Bacterial Chromosome"

25

Summary

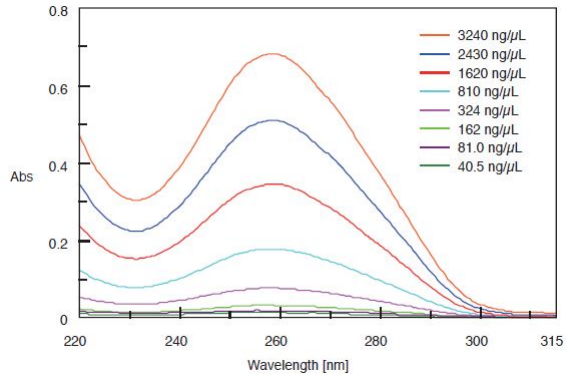
- DNA structure is as varied as protein structure, and nucleic acids can catalyze chemical reactions ("ribozymes")
- Bacterial and animal cells store and process DNA slightly differently, although both use similar ribosomes and the same genetic code
- Modern molecular biology allows us to express virtually any gene using bacterial expression systems

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

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



Source: www.jascoinc.com

27

DNA vs. Protein Absorbance

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

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

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

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

Which is more sensitive?

What are the implications?

28

Other Values for Long, Randomized Sequences

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

29

Nucleic Acids – Smaller Molecules

The screenshot shows the IDT Biophysics website interface for the 'UV Spectrum of DNA' tool. At the top, there is a navigation bar with links for 'DNA Thermodynamics', 'UV Spectrum', 'Publications', 'Tool Help', and 'Contact Us'. Below this, the title 'UV Spectrum of DNA' is displayed. A text input field labeled 'Sequences:' is followed by a '# Bases: 0' indicator. Below the input field, there are three radio button options: 'Single stranded DNA' (selected), 'Duplex DNA', and 'Apply Cavaluzzi-Borer Correction' (checked). At the bottom of the form, there are 'Calculate' and 'Reset' buttons.

- IDT DNA Calculator:
<http://biophysics.idtdna.com/UVSpectrum.html>

Source: www.jascoinc.com

30

Calculating Reverse Complement

Reverse Complement

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.

Paste the raw or FASTA sequence into the text area below:

```
>Sample sequence
GGGGaaaaaaaaatttatatat
```

SUBMIT CLEAR

• Convert the DNA sequence into its reverse-complement counterpart.

[\[home\]](#)

- Bioinformatics.org Calculator (no-frills):
http://bioinformatics.org/sms/rev_comp.html

Source: www.jascoinc.com

31

DNA Translation Tool

- **Site:**
<http://web.expasy.org/translate/>
- **Input:** DNA or RNA sequence (5' → 3' orientation)
- **Output:** All six possible translation frames

32

Other Databases

- Entrez works for DNA sequences, too (reference sequences start with NM_)
- PDB also houses a number of RNA/DNA structures in addition to proteins

33

Putting it Together:
SDSC Biology Workbench

SDSC
SAN DIEGO SUPERCOMPUTER CENTER

Biology WorkBench

- **Site:** <http://workbench.sdsc.edu/>
- **Exercise:** Create an account, try to examine some of the tools. What looks familiar?

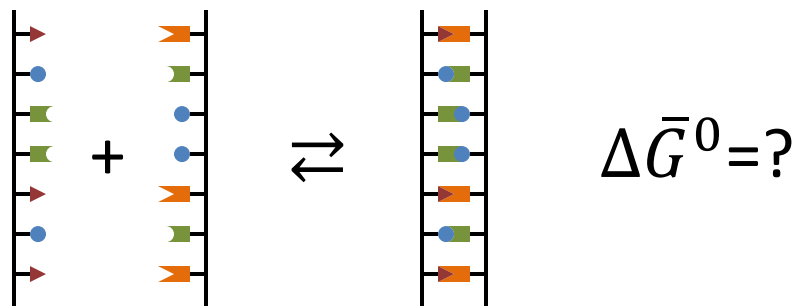
34

Think And Discuss

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

35

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?

37

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

38

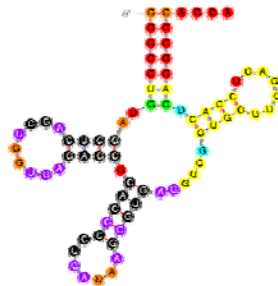
Predicting Melting Temperatures

- IDT OligoAnalyzer:
<http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/>
- **Input:** Your DNA sequence of interest, salt concentration
- **Output:** T_m , extinction coefficient, %GC content

39

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

41

Why is this Useful?

- Site-Directed Mutagenesis
- Good Primers:
 - $T_m > 78\text{ }^\circ\text{C}$ (2 mM 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

42

Example: Designing Primers

Agilent Technologies | **Genomics** | English | United States | Contact Us | Logout | My Account | Quick Order | Cart (0)

Home | Products & Services | Applications | Resources & Support | Search products or part numbers

QuikChange Primer Design

Overview | General Info | Links | BioCalculators | QuikChange Primer Design

[? Help](#)

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.

[Expand Help](#)

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

<http://www.genomics.agilent.com/primerDesignProgram.jsp>

43

PRODUCTS & SERVICES | CROSSLAB | SOLUTIONS | SUPPORT | LIBRARY | TRAINING & EVENTS | BUY

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: [Browse](#) | No file selected
Or, paste plain text or FASTA-formatted DNA sequence in the box below:
`atagatatagacagacatcacatcacatcacaggaggatcacag
accatatacagcaga`
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 Site 2 Site 3 Site 4 Site 5 Site 6 Site 7
g ^ ^ ^ ^ ^ ^ ^
or
 Delete a region between two checked nucleotides (Note: two checked nucleotides will not be deleted)
or
 Insert between two checked nucleotides

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

Finally, obtain your designed primer sequences.
[Design Primers](#)
Clear this form to load a new sequence:
[Clear Input](#)

PRODUCTS & SERVICES CROSSLAB SOLUTIONS SUPPORT LIBRARY TRAINING & EVENTS BUY

Insert between two checked nucleotides

1 a 21 3 a 4 g 5 a 6 t 7 a 8 t 9 a
 10 g 11 a 12 c 13 a 14 g 15 a 16 c 17 a 18 t
 19 a 20 c 21 a 22 t 23 c 24 a 25 g 26 a 27 c
 28 a 29 t 30 a 31 t 32 a 33 c 34 a 35 g 36 g
 37 a 38 g 39 g 40 a 41 g 42 g 43 a 44 t 45 a
 46 c 47 a 48 c 49 a 50 g 51 a 52 c 53 a 54 t
 55 c 56 a 57 t 58 a 59 c 60 a 61 g 62 c 63 a
 64 g 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')
s24c	5'-atctctctctgtatagtcggatgtagtctgtctatctc-3' 5'-gatataagacagacatcacatccagacatatacaggaggagat-3'


Oligonucleotide information:

Primer Name	Length (nt)	Tm	Duplex Energy at 68 °C	Energy Cost of Mismatches
s24c	41	78.60°C	-45.55 kcal/mole	3.05%
s24c	41	78.60°C	-49.02 kcal/mole	3.83%

Primer-template duplexes:

Primer Name	Primer-Template Duplex
s24c	gatataagacagacatcacatccagacatatacaggaggagataca 3'-ctatatctgtctgtatgttaggtgtatatgtctctctccta-5' 5'-gatataagacagacatcacatccagacatatacaggaggagat-3' ???ctatatctgtctgtatgttagtctgtatatgtctctctcctatgt

Example: General Approach


Biophysics
INTEGRATED DNA TECHNOLOGIES

[DNA Thermodynamics](#)
[UV Spectrum](#)
[Publications](#)
[Tool Help](#)
[Contact Us](#)

DNA Thermodynamics & Hybridization *Oligo + Target ↔ Duplex*

Sequence: # Bases: 0
 5'- -3'

Mismatch, Dangling Ends

Oligo Conc: μM
 Target Conc: μM
 Na⁺, K⁺ Conc: mM
 Mg²⁺ Conc: mM
 dNTPs Conc: mM

Base Sequence Symbols
 Native DNA: A, C, G, T
 Inosine: I
 Locked nucleic acids: +A, +C, +G, +T

Check this box' →

<http://biophysics.idtdna.com/>
46

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

Example: General Approach

IDT[®] Biophysics
INTEGRATED DNA TECHNOLOGIES

[DNA Thermodynamics](#) [UV Spectrum](#) [Publications](#) [Tool Help](#) [Contact Us](#)

DNA Thermodynamics & Hybridization *Oligo + Target ↔ Duplex*

Sequence: # Bases:

Oligo Conc: μ M
 Target Conc: μ M
 Na⁺, K⁺ Conc: mM
 Mg²⁺ Conc: mM
 dNTPs Conc: mM

INTRODUCE MISMATCH

Unpaired Base																							Unpaired Base	
5'-	G	C	C	C	T	G	A	A	G	G	T	C	A	A	G	T	C	C	C	C	C	A	-3'	
3'-	C	G	G	G	A	C	T	T	C	A	A	G	T	T	C	A	G	G	G	G	G	-5'		
	-1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	

<http://biophysics.idtdna.com/>

48

Think And Discuss

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

49

Example: Sequence Analysis of SH3 Mutants

- Step 1: Design Primers
– 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