DNA Structure and Properties

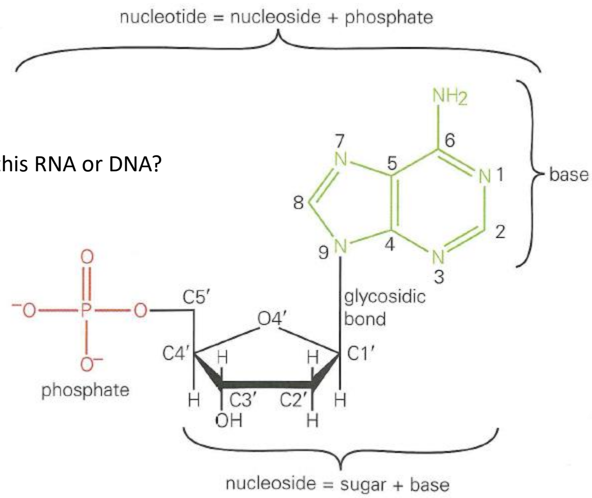
Biochemistry Boot Camp 2017
Session #6
Nick Fitzkee
nfitzkee@chemistry.msstate.edu

DNA

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

Nucleic Acid Structure

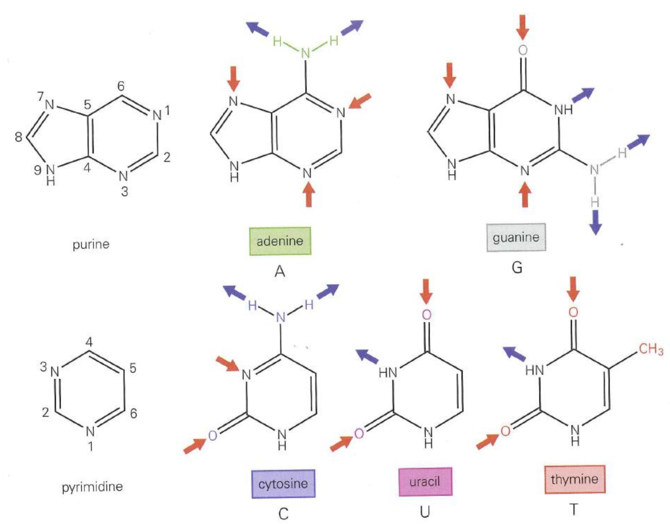
Question: Is this RNA or DNA?



Molecules of Life, pp. 15

3

Nucleic Acid Bases



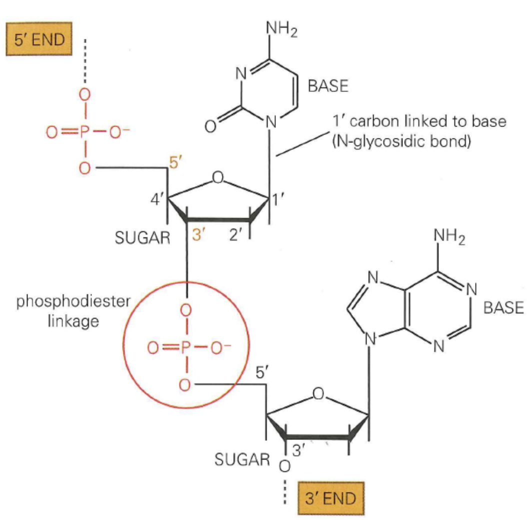
Molecules of Life, pp. 20

4

Nomenclature

	Base	Nucleoside	Nucleotide	Nucleic Acid
Purine	Adenine	Adenosine	Adenylate	RNA
		Deoxyadenosine	Deoxyadenylate	DNA
	Guanine	Guanosine	Guanylate	RNA
		Deoxyguanosine	Deoxyguanylate	DNA
Pyrimidines	Cytosine	Cytidine	Cytidylate	RNA
		Deoxycytidine	Deoxycytidylate	DNA
	Thymine	Thymidine	Thymidylate	
		Deoxythymidine	Deoxythymidylate	DNA
	Uracil	Uridine	Uridylate	RNA

Nucleic Acids Are Also Polymers

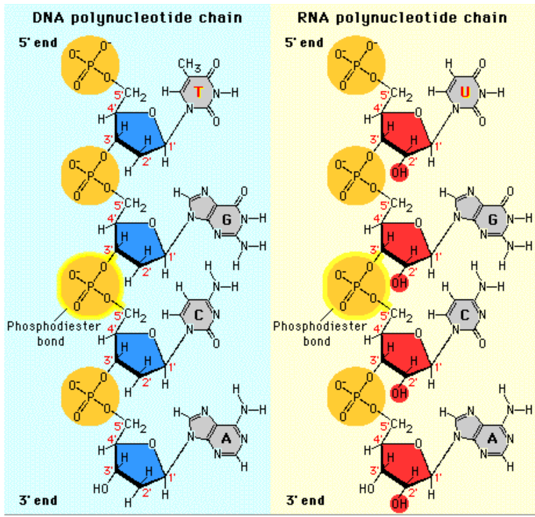


DNA & RNA Polymerase: Build up DNA and RNA from nucleoside triphosphates (5′ → 3′ synthesis)

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

Molecules of Life, pp. 21

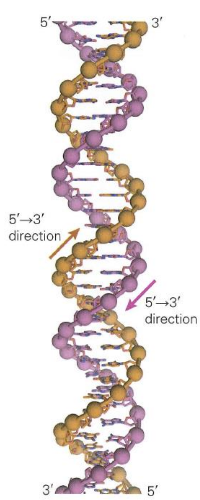
DNA vs RNA

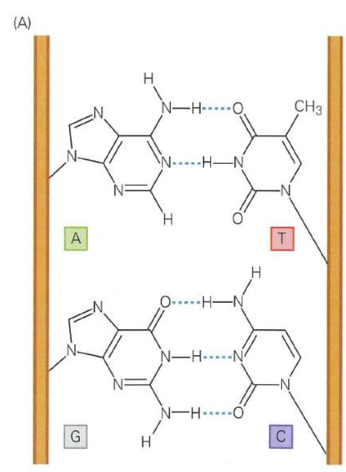


- DNA less reactive
- RNA is easily attacked by enzymes

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

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





Molecules of Life, pp. 23

8

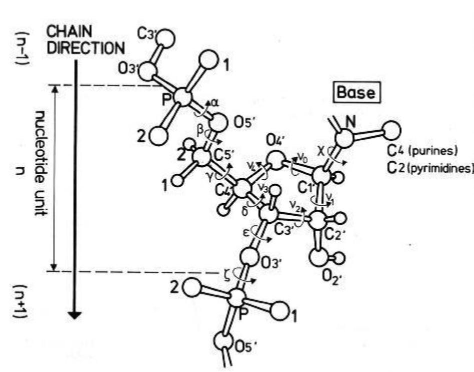
Reverse Complement

- 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

Simplification: Sugar Pucker

- ν angles are related, so sugar ring can be simplified
- Think "chair" and "boat" forms of cyclohexane

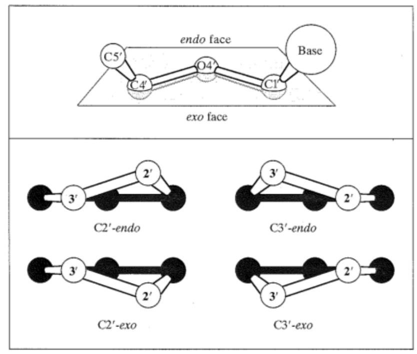


Figure 1.38 Sugar conformations of nucleic acids. The pucker of the sugar ring in RNA and DNA is defined relative to the plane formed by the C1′-carbon, C4′-carbon, and O4′-oxygen of the five-member ring. The *endo* face lies above the plane, toward the nucleobase, while the *exo* face lies below the plane.

van Holde, *et al*. *Principles of Physical Biochemistry.*

11

Nucleic Acid Primary Structure

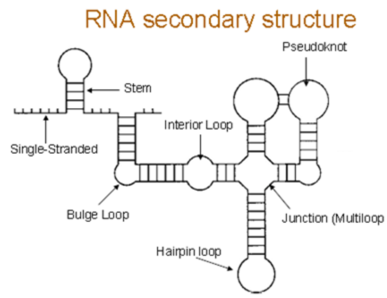
- **Just like proteins:** the sequence of bases

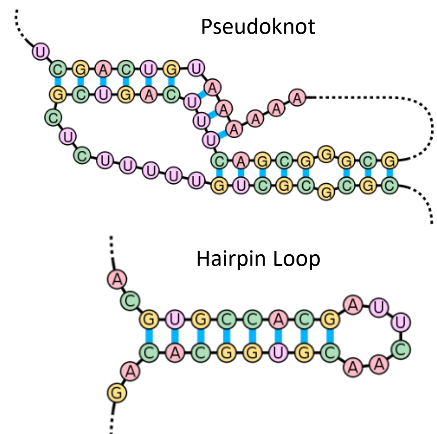
5'-dAdGdTdTdCdAdCdCdC-3' (DNA)
AGTTCACCC

5'-AGUUCACCC-3' (RNA)

12

Secondary Structure



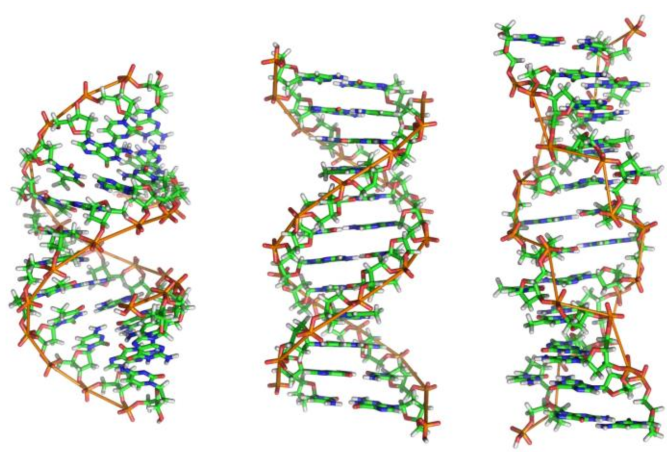


- Base pairing motifs

Source: Wikipedia, “RNA Secondary Structure,” “Nucleic Acid Secondary Structure”

13

Tertiary Structure



A Form DNA B Form DNA Z Form DNA

Source: Steven Carr, www.mun.ca

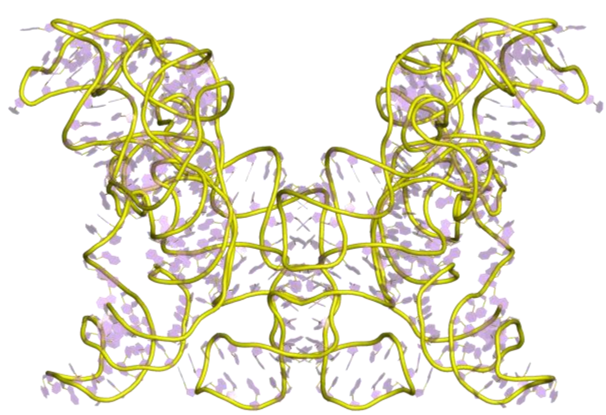
14

Tertiary Structure

	Average Torsion Angles for Nucleic Acid Helices (in °)						
Structure Type	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.*

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.

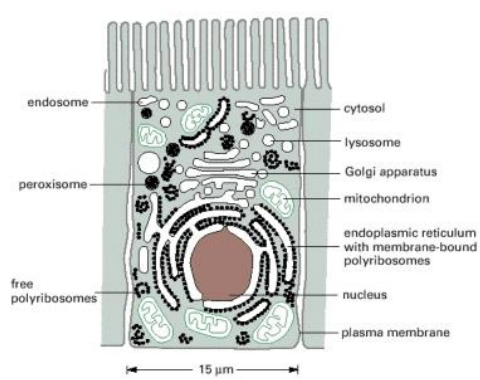
Think and Discuss

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

17

Review of Intro Biology

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

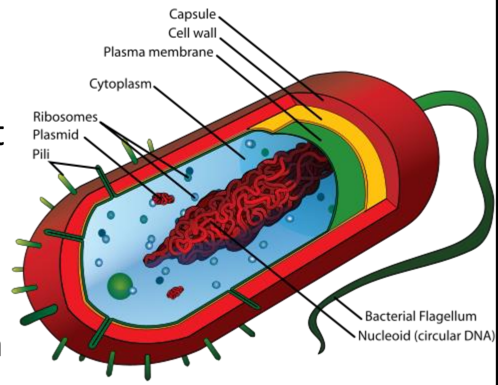


Alberts, *et al*. *Molecular Biology of the Cell,* 4th Edition.

18

Review of Intro Biology

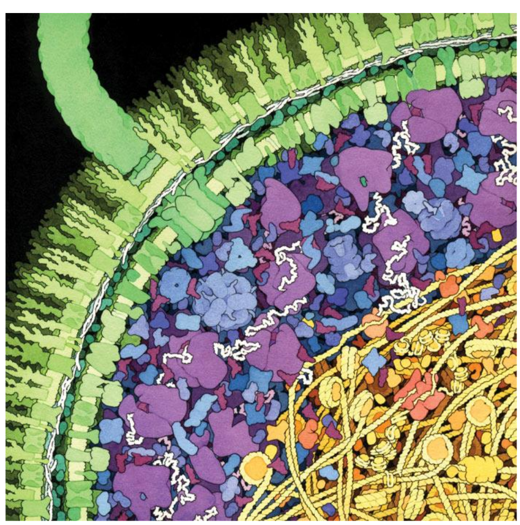
- Parts of a prokaryotic bacterial cell
- No nucleus: DNA is not linear but circular (no ends)
- No organelles, but ribosomes, etc. exist in the cytoplasm



Source: Wikipedia, "Bacterial Cell Structure."

19

It's Crowded in There!



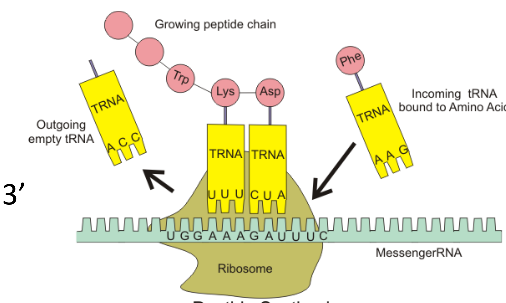
Source: Goodsell, D. http://mgl.sripps.edu/people/goodsell/illustration/public/

20

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



Peptide Synthesis

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

Source: Wikipedia, "Genetic Code"

22

Different Reading Frames

```
reading frame:                 123
                               |||
                               acttacccggacta
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

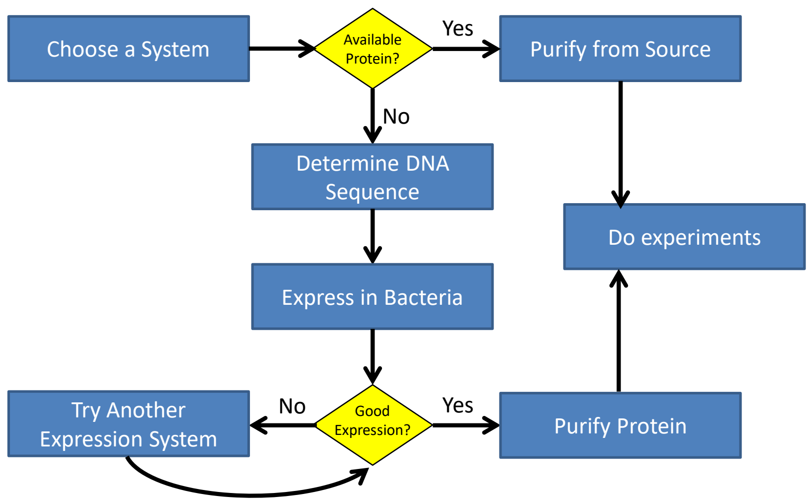
23

Think and Discuss

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

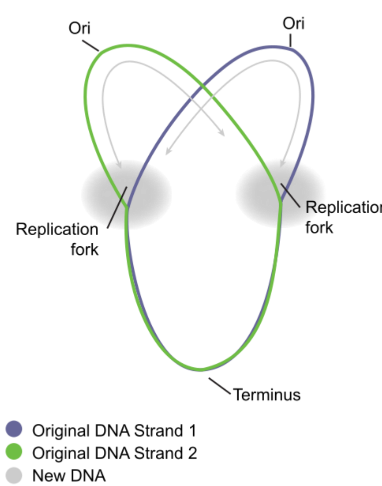
24

Biochemistry Research Flow Chart



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



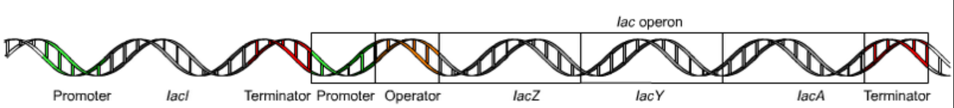
Source: Wikipedia, "Circular Bacterial Chromosome"

The Lactose (lac) Operon

- **Idea:** Bacteria only want to produce proteins if they are needed
- Why metabolize lactose (hard) when glucose (easy) is available?
- **Operon:** A set of genes (proteins) under the control of other genes in the cell

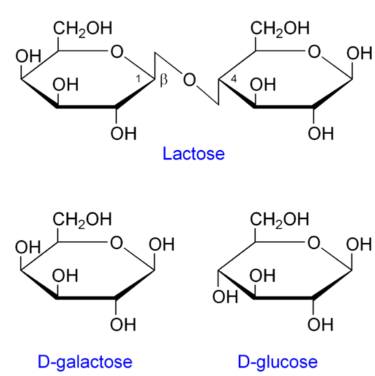
27

The Lactose (lac) Operon



Proteins:

- **lacI** (lac repressor): binds at operator when no lac present; prevents binding of RNA polymerase at promoter
- **lacZ** (β-galactosidase): converts Lac in to Gal and Glc by hydrolyzing glycosidic linkage
- **lacY** (β-galactoside permease): Pumps Lac into the cell

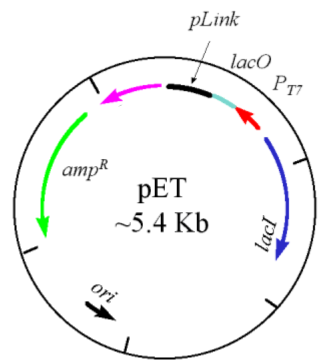


Source: Wikipedia, "Lac Operon"

28

Bacterial Expression Vectors

- pET Plasmid Genes
  - Origin of replication
  - Lac repressor (lacI)
  - RNA Pol promoter (P_{T7})
  - Lac Operator (lacO)
  - Polylinker – where your DNA sequence goes (pLink)
  - Ampicillin resistance (amp^R)
- Is this plasmid persistent?

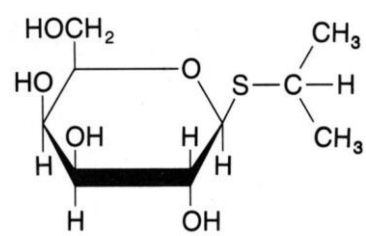


Source: Mike Blaber, BCH5425 Course Notes

29

Inducible Expression

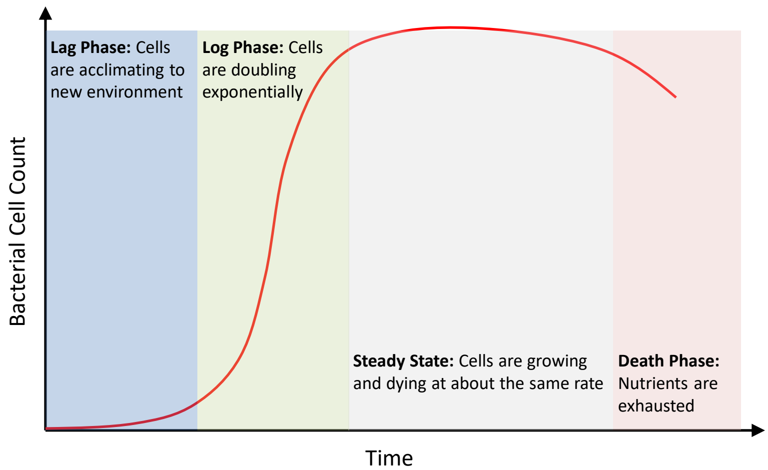
- **IPTG:** Turns on protein expression without being hydrolyzed
- Protein expression can be switched on when desired



Isopropyl-β-D-thiogalactoside (IPTG)

30

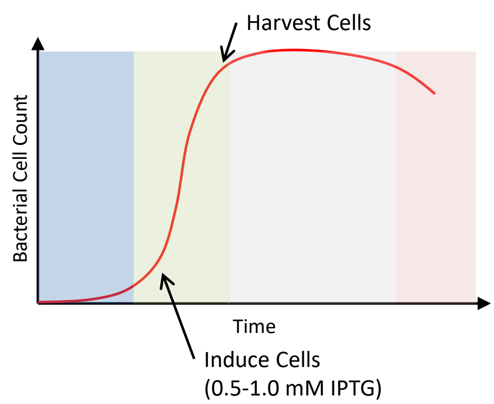
When Should I Induce?



31

When Should I Induce?

- Protein expression is greatest during log phase
- Inducing at lag phase may unnecessarily cripple your cells
- Typically, induce at an OD_{600} of 0.5-0.6
- Always follow your lab's protocols!



32

Think and Discuss

Why is Ampicillin resistance necessary for the function of the pET vector system?

Summary

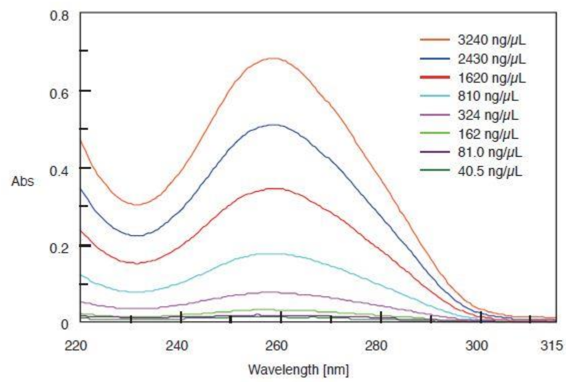
- DNA structure is as varied as protein structure, and nucleic acids can catalyze chemical reactions (“ribozymes”)
- Bacterial and animal cells store and process DNA slightly differently, although both use similar ribosomes and the same genetic code
- Modern molecular biology allows us to express virtually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

Nucleic Acid Extinction Coefficient

DNA Concentrations: Often measured in µg/mL (or the equivalent ng/µL) 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 $(\mu g/mL)^{-1}\ cm^{-1}$



Source: www.jascoinc.com

35

DNA vs. Protein Absorbance

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

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

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

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

Which is more sensitive?

What are the implications?

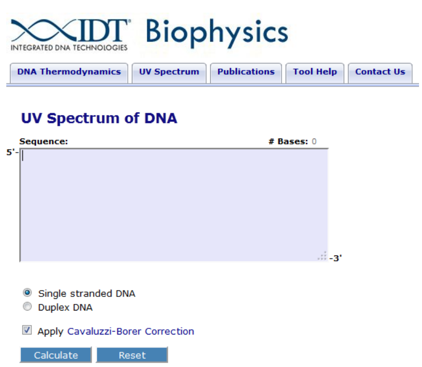
36

Other Values for Long, Randomized Sequences

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

37

Nucleic Acids – Smaller Molecules

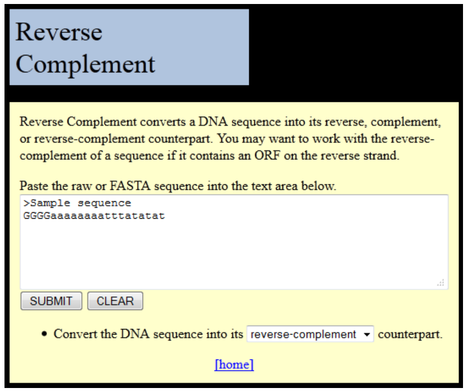


- IDT DNA Calculator: http://biophysics.idtdna.com/UVSpectrum.html

Source: www.jascoinc.com

38

Calculating Reverse Complement



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

Source: www.jascoinc.com

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

41

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?

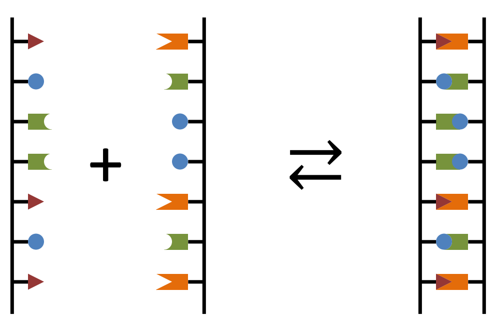
42

Think And Discuss

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

43

DNA "Melting"



$\Delta\bar{G}^0 = ?$

- 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?

45

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

46

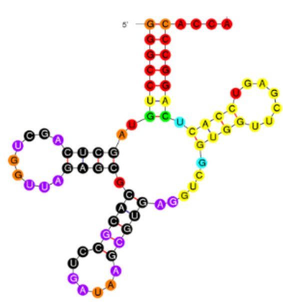
Predicting Melting Temperatures

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

47

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:

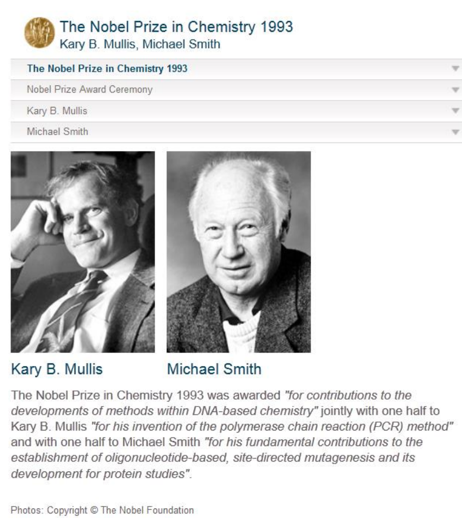
GGGUCUCUCUGGUUAGACCAGAUCUGAGCCUGGGAGCUCUCU
GGCUAACUAGGGAACCCAC

49

Why is this Useful?

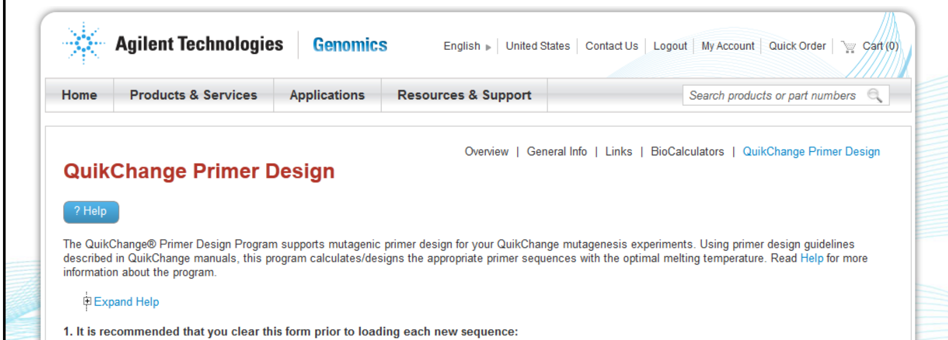
- Site-Directed Mutagenesis

- Good Primers:
  - T_m > 78 °C (2 mM $MgCl_2$, 50 mM NaCl)
  - GC content > 40%
  - No secondary structure (< 50 bp)
  - End with G or C



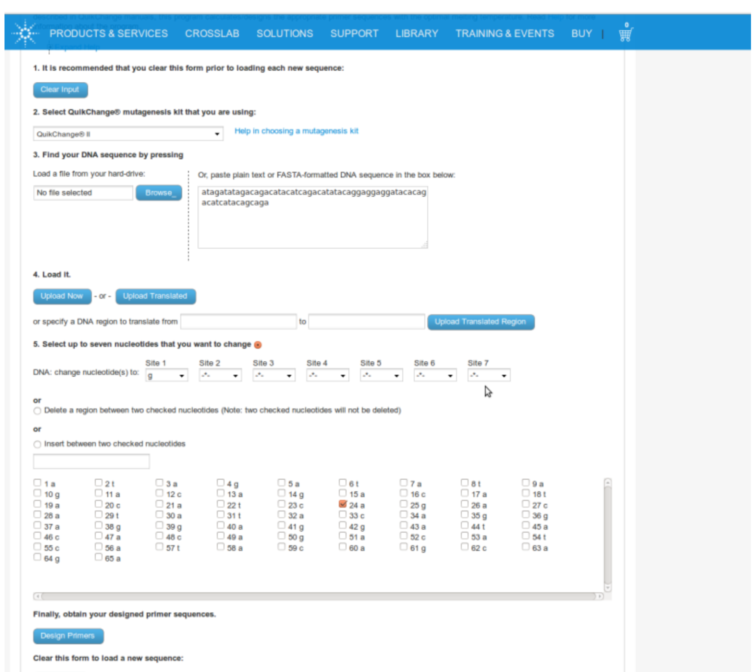
50

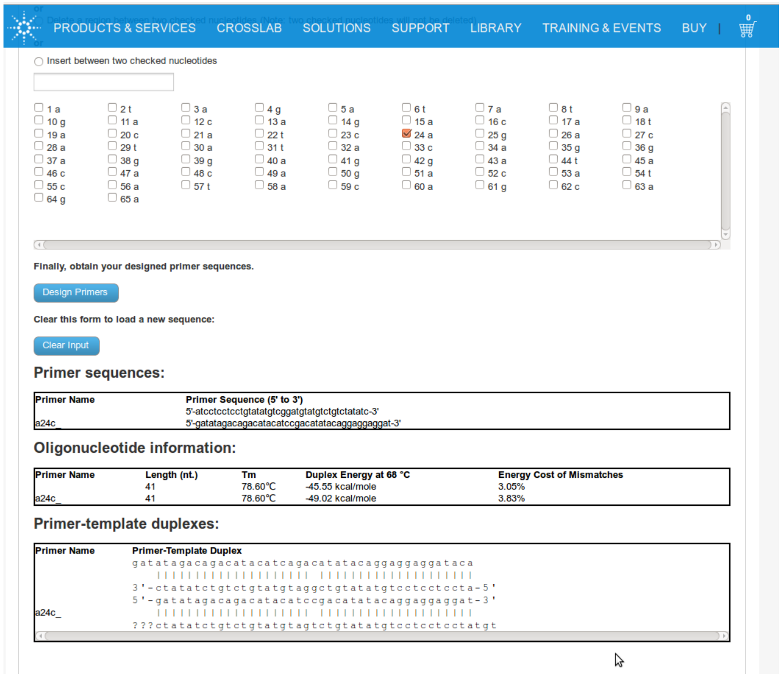
Example: Designing Primers



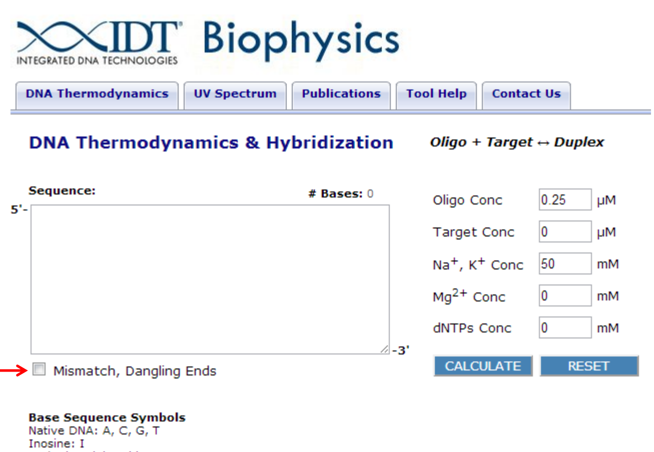
http://www.genomics.agilent.com/primerDesignProgram.jsp

51





Example: General Approach



Check this box`

http://biophysics.idtdna.com/

54

General Primer Design Principles

- **PCR Steps:** Denature (95 °C), anneal (60 °C), extend (70 °C)

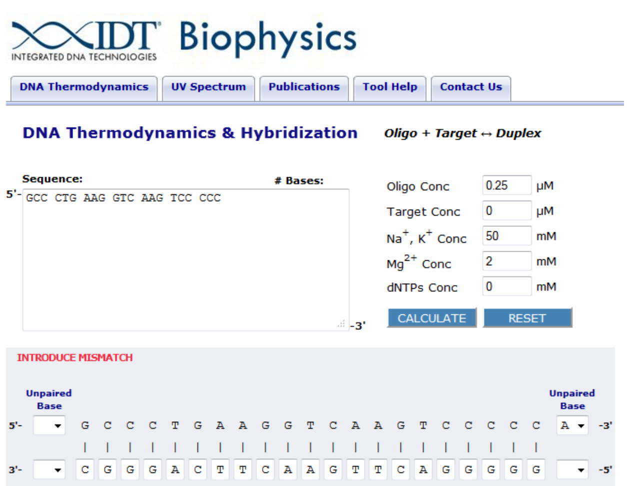
- Considerations:
  - **Melting Temperature:** Should be 52-58 °C
  - **GC Content:** 40-60%
  - **Length:** ~30 bp (but longer can be okay)
  - **Secondary Structure:** Avoid if possible

- 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



http://biophysics.idtdna.com/

56

Think And Discuss

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

57

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