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

Question: Is this RNA or DNA?

Nucleic Acid Bases

Molecules of Life, pp. 15

Molecules of Life, pp. 20
Nomenclature

<table>
<thead>
<tr>
<th>Base</th>
<th>Nucleoside</th>
<th>Nucleotide</th>
<th>Nucleic Acid</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Purine</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adenine</td>
<td>Adenosine</td>
<td>Adenylate</td>
<td>RNA</td>
</tr>
<tr>
<td></td>
<td>Deoxy adenosine</td>
<td>Deoxyadenylate</td>
<td>DNA</td>
</tr>
<tr>
<td>Guanine</td>
<td>Guanosine</td>
<td>Guanylate</td>
<td>RNA</td>
</tr>
<tr>
<td></td>
<td>Deoxyguanosine</td>
<td>Deoxyguanylate</td>
<td>DNA</td>
</tr>
<tr>
<td><strong>Pyrimidines</strong></td>
<td>Cytosine</td>
<td>Cytidine</td>
<td>Cytidylate</td>
</tr>
<tr>
<td></td>
<td>Deoxy cytidine</td>
<td>Deoxycytidylate</td>
<td>DNA</td>
</tr>
<tr>
<td>Thymine</td>
<td>Thymidine</td>
<td>Thymidylate</td>
<td></td>
</tr>
<tr>
<td></td>
<td>deoxy thymidine</td>
<td>Deoxythymidylate</td>
<td>DNA</td>
</tr>
<tr>
<td>Uracil</td>
<td>Uridine</td>
<td>Uridylate</td>
<td>RNA</td>
</tr>
</tbody>
</table>

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

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

Molecules of Life, pp. 23
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 AUCCGCTT?

Structure in DNA

- Bases are planar

- Torsion angles are shown
  - Much more complex than proteins

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

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

\[ 5'-dAdGdTdTdTdTdTdTdTdTdTdTdTAdCdCdCdC-3' \quad \text{(DNA)} \]

\[ \text{AGTTCACCC} \]

\[ 5'-\text{AGUUCACCC-3'} \quad \text{(RNA)} \]

Secondary Structure

• Base pairing motifs

Tertiary Structure

A Form DNA  B Form DNA  Z Form DNA

Source: Steven Carr, www.mun.ca

Tertiary Structure

<table>
<thead>
<tr>
<th>Structure Type</th>
<th>Average Torsion Angles for Nucleic Acid Helices (in °)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Alpha</td>
</tr>
<tr>
<td>A-DNA (fibres)</td>
<td>-50</td>
</tr>
<tr>
<td>GGCCGGCC</td>
<td>-75</td>
</tr>
<tr>
<td>B-DNA (fibres)</td>
<td>-41</td>
</tr>
<tr>
<td>CGCGAATTCGCG</td>
<td>-63</td>
</tr>
<tr>
<td>Z-DNA (C residues)</td>
<td>-137</td>
</tr>
<tr>
<td>Z-DNA (G residues)</td>
<td>47</td>
</tr>
<tr>
<td>DNA-RNA decamer</td>
<td>-69</td>
</tr>
<tr>
<td>A-RNA</td>
<td>-68</td>
</tr>
</tbody>
</table>

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.)


Think and Discuss

What are the major differences between DNA and protein structures? What are the similarities?
Review of Intro Biology

• Parts of a eukaryotic animal cell

• Has a nucleus where DNA is stored

• Membrane-bound organelles


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

It’s Crowded in There!

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

Central Dogma

• DNA $\rightarrow$ mRNA
  “Transcription”
  – Synthesized RNA
  Polymerase
  – RNA formed from 5’ to 3’

• mRNA $\rightarrow$ 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.

Genetic Code

Different Reading Frames

reading frame: 123

first reading frame
second reading frame
third reading frame
Think and Discuss

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

Biochemistry Research Flow Chart

Choose a System → Available Protein? → Yes → Purify from Source → Do experiments

No → Determine DNA Sequence → Express in Bacteria → Good Expression?

Yes → Purify Protein

No → Try Another Expression System
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


---

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

\[ 0.020 \text{ (μg/mL)}^{-1} \text{ cm}^{-1} \]

Source: www.jascoinc.com

![Graph showing absorbance vs. wavelength for different DNA concentrations](image)

DNA vs. Protein Absorbance

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

\[ 0.020 \text{ (μg/mL)}^{-1} \text{ cm}^{-1} \]

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

\[ 0.0016 \text{ (μg/mL)}^{-1} \text{ cm}^{-1} \]

Which is more sensitive?

What are the implications?
Other Values for Long, Randomized Sequences

• Single-Stranded RNA: 0.025 (µg/mL)⁻¹ cm⁻¹

• Single-Stranded DNA: 0.030 (µg/mL)⁻¹ cm⁻¹

• For a pure nucleic acid, the 260/280 nm ratio should be approximately 1.8-2.0

Nucleic Acids – Smaller Molecules

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

Source: www.jascoinc.com
Calculating Reverse Complement


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

• Entrez works for DNA sequences, too (reference sequences start with NM_)

• PDB also houses a number of RNA/DNA structures in addition to proteins

Putting it Together:
SDSC Biology Workbench

• Site: http://workbench.sdsc.edu/

• Exercise: Create an account, try to examine some of the tools. What looks familiar?
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?

\[ \Delta \bar{G}^0 = ? \]
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:** GCCCTGAAGGTCAGGTCCCCC
  — 14 G-C = 56 °C
  — 7 A-T = 14 °C
  — Prediction is 70
Predicting Melting Temperatures

- **IDT 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](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

Why is this Useful?

- Site-Directed Mutagenesis

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

http://www.genomics.agilent.com/primerDesignProgram.jsp
Example: General Approach

Check this box

http://biophysics.idtdna.com/
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/
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
  – 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 into 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