Nucleic Acids
Why do I care?
Nucleic Acids
Why do I care?
Proteins do everything, right?
Nucleic Acids
Why do I care?

Proteins do everything, right?

revolutions at the turn of the century

opportunities for the 21st century
In the beginning…

DNA

\downarrow

mRNA

\downarrow

Protein
In the beginning…

DNA → mRNA → Protein
In the beginning…

DNA → mRNA → Protein

Archival information storage
In the beginning...

DNA → mRNA → Protein

Archival information storage

Transient information storage
In the beginning...

DNA → mRNA → Protein

Archival information storage

Transient information storage

Catalysis, structure, regulation, et al.
Chicken & Egg?

DNA → mRNA → Protein

Archival information storage

Transient information storage

Catalysis, structure, regulation, et al.
RNA can do everything

DNA  →  RNA  →  Protein

Archival information storage

Transient information storage  
Catalysis!

Catalysis, structure, regulation, et al.

1980-2000
RNA - primordial molecule

DNA

mRNA

Protein

So we accepted that RNA was probably the first, primitive do-everything biomolecule.

But proteins came along to supplant everything and make the world, evolutionarily, what it is today. All hail the protein!
Seems simple...

DNA

mRNA

Protein

AUGCGGAGT...

MetXxxXxx...
Seems simple?

DNA → pre-mRNA → mRNA → Protein

AUGCGGAGT… → AUGUUUCGUGAGC… → MetZzzXxxWww…
Even more complicated…

DNA ➞ pre-mRNA ➞ mRNA ➞ Protein

Exon  Intron  Exon  Intron  Exon

Splicing

AUGCGGAGU…

MetXxxXxx…

pre-mRNA ➞ mRNA ➞ Protein

DNA ➞ pre-mRNA ➞ mRNA ➞ Protein

Exon  Intron  Exon  Intron  Exon

Splicing

AUGCGGAGU…

MetXxxXxx…
Reality…

DNA ➔ pre-mRNA ➔ mRNA ➔ Protein
Reality...
and our perception of it

DNA → pre-mRNA → mRNA → Protein

Exon
Intron
Junk
AUGCGGAGT...

MetXxxXxx...
Reality…
and our perception of it
Reality...
and our perception of it

Genome project goals

Identify and characterize the proteins. What are their structures? What do they do? How do they interact?
20th Century View

Kinases
Polymerases
Receptors
Hydrogenases
Oxygenases
Proteases
20th Century View

- Kinases
- Oxygenases
- Receptors
- Proteases
- Hydrogenases
- Polymerases
- Nucleic acids
21st Century News

Genome project

Number of protein-encoding genes in the human genome: 25,000
21st Century News

Genome project

Number of protein-encoding genes in the human genome: 25,000

Number of protein-encoding genes in the 1,000 cell C elegans genome: 19,500
21st Century News

Genome project

Number of protein-encoding genes in the human genome: 25,000

Number of protein-encoding genes in the 1,000 cell *C elegans* genome: 19,500

Number of protein-encoding genes in the corn genome: 40,000
Genome project

Number of protein-encoding genes in the human genome: 25,000
Genome project

Number of protein-encoding genes in the human genome: 25,000

Number of different proteins: ≫25,000
21st Century News

Genome project

Number of protein-encoding genes in the human genome: 25,000

Number of different proteins: ≫25,000

How? RNA editing and alternative splicing
Genome project

Number of protein-encoding genes in the human genome: 25,000

Number of different proteins: \( \gg 25,000 \)

How? RNA editing and alternative splicing

Up to 3/4 of all human genes are subject to alternative editing
21st Century News

Genome project

Number of protein-encoding genes in the human genome: 25,000

Number of different proteins: »25,000

How? RNA editing and alternative splicing

Up to 3/4 of all human genes are subject to alternative editing

The prevalence of alternative editing appears to increase with an organism’s complexity
Genome project

How? RNA editing and alternative splicing

Up to 3/4 of all human genes are subject to alternative editing

The prevalence of alternative editing appears to increase with an organism’s complexity
21st Century News

Genome project

How? RNA editing and alternative splicing

Up to 3/4 of all human genes are subject to alternative editing

The prevalence of alternative editing appears to increase with an organism’s complexity

Humans have the highest number of introns per gene of any organism
21st Century News

Genome project

How? RNA editing and alternative splicing

Up to 3/4 of all human genes are subject to alternative editing

The prevalence of alternative editing appears to increase with an organism’s complexity

Humans have the highest number of introns per gene of any organism

At least 15% of the gene mutations that produce genetic diseases and cancers do so by effecting pre-mRNA editing
21st Century News

Genome project

How? RNA editing and alternative splicing

Up to 3/4 of all human genes are subject to alternative editing

The prevalence of alternative editing appears to increase with an organism’s complexity

Humans have the highest number of introns per gene of any organism

At least 15% of the gene mutations that produce genetic diseases and cancers do so by effecting pre-mRNA editing
21st Century News

Intron

Exon

DNA

mRNA

Protein

pre-mRNA

Alternative Splicing

1980's

AUGCGGAGT...

MetXxxXxx...
21st Century News

Instructions and chemistry for alternative editing

AUGCGGAGT...

MetXxxXxx...

Exon

Intron

Exon

Intron

Exon

Intron

DNA

pre-mRNA

mRNA

Protein

Alternative Splicing

1980’s
Back in the lab...
Back in the lab...

With the advent of PCR, nucleic acids have been recognized as extremely powerful combinatorial tools in the test tube.
Back in the lab…

With the advent of PCR, nucleic acids have been recognized as extremely powerful combinatorial tools in the test tube.

Aptamers can be selected that bind to “your favorite molecule”

Can create riboswitches
Back in the lab...

With the advent of PCR, nucleic acids have been recognized as extremely powerful combinatorial tools in the test tube.

- Aptamers can be selected that bind to “your favorite molecule”
- Can create riboswitches

Ron Breaker:
“If it’s so easy for us, I’ll bet nature exploits this”
More gene regulation

DNA  mRNA  Protein

5' untranslated region (5'-UTR)

protein seq

RNA polymerase

regulatory proteins

regulatory elements

promoter
More gene regulation

Ron: search for genes with no known protein regulator and which have a highly conserved 5’ UTR
More gene regulation

Ron: search for genes with no known protein regulator and which have a highly conserved 5’ UTR

Does RNA from that conserved 5’ UTR bind the product or substrate of the encoded enzyme?
More gene regulation

Ron: search for genes with no known protein regulator and which have a highly conserved 5’ UTR

Does RNA from that conserved 5’ UTR bind the product or substrate of the encoded enzyme?

One week: a Nature paper! 2004
Project Encode (2007)
(More) rewriting of textbooks

June 2007, published in Nature

⭐ Some regions of DNA far from protein-coding genes (extreme “junk?”) are nevertheless highly conserved

⭐ Most of both strands of the DNA is transcribed (far beyond that required for protein-coding genes)
21st Century Opportunities

DNA → pre-mRNA → mRNA → Protein
21st Century Opportunities

DNA

pre-mRNA

mRNA

Protein

Alternative splicing

early ‘80s, but…
21st Century Opportunities

DNA → pre-mRNA → mRNA → Protein

Alternative splicing

RNA Editing

early ‘80s, but… late ‘80s

DNA

pre-mRNA

mRNA

Protein
21st Century Opportunities

DNA

RNA Editing

late '80s

pre-mRNA

Alternative splicing

early ‘80s, but…

mRNA

RNAi

late ‘90’s

Protein
21st Century Opportunities

DNA

RNA Editing
late '80s

Alternative splicing
early '80s, but...

pre-mRNA

micro RNA
late '90's

RNAi
late '90's

mRNA

Protein
21st Century Opportunities

DNA

RNA Editing

late '80s

pre-mRNA

Riboswitches

early '80s, but...

Alternative splicing

Riboswitches

early '00s

mRNA

RNAi

late '90's

Riboswitches

micro RNA

late '90's

Protein
21st Century Opportunities

DNA → pre-mRNA → mRNA → Protein

- Riboswitches:
  - early '80s, but...
  - early '00s
  - late '80s

- RNA Editing:
  - late '80s

- Alternative splicing:
  - early '80s

- micro RNA:
  - late '90s

- RNAi:
  - late '90s

- To be discovered...
  - 2010

"so 20th century"
Large Macromolecular Complexes
Ribosome
An RNA machine with protein cofactors
What stabilizes protein structures?
What stabilizes protein structures?

What *directs* protein structures?
The DNA Duplex
The DNA Duplex

What stabilizes the duplex?
The DNA Duplex

What stabilizes the duplex?

What *directs* duplex structure?
Which is more stable?
(which has a higher melting temperature?)

ACCGCCACCAGAAG
TGGCGGTGGCTTC

or

ACCGCCACCAGAAG
TGGCGGTGGCTTA
Which is more stable? (which has a higher melting temperature?)

ACCGCCACCGAAG
TGGCGGTGGCTTC

or

ACCGCCACCGAAG
TGGCGGTGGCTTA

51.6° C
Which is more stable?
(which has a higher melting temperature?)

ACCGCCACCGAAG
TGGCGGTGGCTTC

or

ACCGCCACCGAAG
TGGCGGTGGCTTA

51.6° C
52.5° C

Calculations from http://www.idtdna.com/analyze/Application/OligoAnalyzer/
DNA
A look at the Chemistry
DNA
A look at the Chemistry
DNA
A look at the Chemistry
DNA
A look at the Chemistry
DNA
A look at the Chemistry
DNA
A look at the Chemistry
DNA
A look at the Chemistry
DNA
A look at the Chemistry
DNA
A look at the Chemistry
DNA
A look at the Chemistry
DNA
A look at the Chemistry
DNA
A look at the Chemistry
DNA
A look at the Chemistry
DNA
A look at the Chemistry
DNA
A look at the Chemistry
What forces are important?
What forces are important?

Electrostatics
What forces are important?

Electrostatics

Hydrogen Bonding
What forces are important?

- Electrostatics
- Hydrogen Bonding
- Burial of Hydrophobic Surface & Stacking
Base Pairing
(Donors matched to Acceptors)
Base Pairing
(Donors matched to Acceptors)
Base Pairing
(Donors matched to Acceptors)
Base Pairing
(Donors matched to Acceptors)
Base Pairing
(Donors matched to Acceptors)

Major Groove

Minor Groove
Base Pairing

(Donors matched to Acceptors)
Base Pairing
(Donors matched to Acceptors)
Base Pairing
(Donors matched to Acceptors)
Base Pairing
(Donors matched to Acceptors)
Base Pairing
(Donors matched to Acceptors)

Good base pairing
Base Pairing
(Donors matched to Acceptors)

Good base pairing
Watson-Crick facing
Base Pairing
(Donors matched to Acceptors)

Good base pairing
Watson-Crick facing
but *Anti-Watson-Crick* orientation
Base Pairing
(Donors matched to Acceptors)
Base Pairing
(Donors matched to Acceptors)
Base Pairing
(Donors matched to Acceptors)

Good base pairing
Base Pairing
(Donors matched to Acceptors)

Good base pairing
WC-Hoogsteen facing
Bad Base Pairing
(Donors *not* matched to Acceptors)
Bad Base Pairing
(Donors to Acceptors with **terrible angles**)

![Diagram of base pairing](image)
Wild (but good) Base Pairing
Wild (but good) Base Pairing
Wild (but good) Base Pairing
Wild (but good) Base Pairing

G-quartet
(Telomeres)
AT Base Pair

Ten H-Bonds
How important are H-bonds in DNA?

Table 1. Free Energies and Melting Temperatures for Dodecamer Duplexes Containing a Variable T−X, F−X, B−X, or D−X Base Pair (X = A, T, C, G)

<table>
<thead>
<tr>
<th>duplex</th>
<th>$T_m$ (°C) $^a$</th>
<th>$ΔG^{298}$ (kcal)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5'-CTTTCGAGATTC-3'</td>
<td>39.4</td>
<td>12.3</td>
</tr>
<tr>
<td>3'-GAAAAAGAGTTT-5'</td>
<td>26.4</td>
<td>8.7</td>
</tr>
<tr>
<td>5'-CTTTCGAGATTC-3'</td>
<td>30.7</td>
<td>9.3</td>
</tr>
<tr>
<td>3'-GAAAAAGAGTTT-5'</td>
<td>27.1</td>
<td>8.9</td>
</tr>
<tr>
<td>5'-CTTTCGAGATTC-3'</td>
<td>21.4</td>
<td>7.4</td>
</tr>
<tr>
<td>3'-GAAAAAGAGTTT-5'</td>
<td>25.0</td>
<td>8.2</td>
</tr>
<tr>
<td>5'-CTTTCGAGATTC-3'</td>
<td>23.0</td>
<td>8.0</td>
</tr>
<tr>
<td>3'-GAAAAAGAGTTT-5'</td>
<td>20.2</td>
<td>7.3</td>
</tr>
<tr>
<td>5'-CTTTCGAGATTC-3'</td>
<td>21.0</td>
<td>7.5</td>
</tr>
<tr>
<td>3'-GAAAAAGAGTTT-5'</td>
<td>22.9</td>
<td>7.8</td>
</tr>
<tr>
<td>5'-CTTTCGAGATTC-3'</td>
<td>20.1</td>
<td>7.6</td>
</tr>
<tr>
<td>3'-GAAAAAGAGTTT-5'</td>
<td>20.3</td>
<td>6.7</td>
</tr>
<tr>
<td>5'-CTTTCGAGATTC-3'</td>
<td>20.8</td>
<td>7.4</td>
</tr>
<tr>
<td>3'-GAAAAAGAGTTT-5'</td>
<td>22.2</td>
<td>7.6</td>
</tr>
<tr>
<td>5'-CTTTCGAGATTC-3'</td>
<td>19.7</td>
<td>7.4</td>
</tr>
<tr>
<td>3'-GAAAAAGAGTTT-5'</td>
<td>17.6</td>
<td>6.9</td>
</tr>
</tbody>
</table>

$^a$ Conditions: 100 mM NaCl, 10 mM MgCl$_2$, 10 mM Na·Pipes, pH 7.0, 1.6 μM each strand.
How important are H-bonds in DNA?

Table 1. Free Energies and Melting Temperatures for Dodecamer Duplexes Containing a Variable T–X, F–X, B–X, or D–X Base Pair (X = A, T, C, G)

<table>
<thead>
<tr>
<th>duplex</th>
<th>$T_m$ (°C)</th>
<th>$\Delta G_{298}$ (kcal)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5'-CTTTTTCTTTCTT 3'-GAAAAAGAGAAA</td>
<td>39.4</td>
<td>12.3</td>
</tr>
<tr>
<td>5'-CTTTTTCTTTCTT 3'-GAAAAAGAGAAA</td>
<td>28.4</td>
<td>8.7</td>
</tr>
<tr>
<td>5'-CTTTTTCTTTCTT 3'-GAAAAAGAGAAA</td>
<td>30.7</td>
<td>9.3</td>
</tr>
<tr>
<td>5'-CTTTTTCTTTCTT 3'-GAAAAAGAGAAA</td>
<td>27.1</td>
<td>8.9</td>
</tr>
<tr>
<td>5'-CTTTTTCTTTCTT 3'-GAAAAAGAGAAA</td>
<td>21.4</td>
<td>7.4</td>
</tr>
<tr>
<td>5'-CTTTTTCTTTCTT 3'-GAAAAAGAGAAA</td>
<td>25.0</td>
<td>8.2</td>
</tr>
<tr>
<td>5'-CTTTTTCTTTCTT 3'-GAAAAAGAGAAA</td>
<td>23.0</td>
<td>8.0</td>
</tr>
<tr>
<td>5'-CTTTTTCTTTCTT 3'-GAAAAAGAGAAA</td>
<td>20.2</td>
<td>7.3</td>
</tr>
<tr>
<td>5'-CTTTTTCTTTCTT 3'-GAAAAAGAGAAA</td>
<td>21.0</td>
<td>7.5</td>
</tr>
<tr>
<td>5'-CTTTTTCTTTCTT 3'-GAAAAAGAGAAA</td>
<td>22.9</td>
<td>7.8</td>
</tr>
<tr>
<td>5'-CTTTTTCTTTCTT 3'-GAAAAAGAGAAA</td>
<td>20.1</td>
<td>7.6</td>
</tr>
<tr>
<td>5'-CTTTTTCTTTCTT 3'-GAAAAAGAGAAA</td>
<td>20.3</td>
<td>6.7</td>
</tr>
<tr>
<td>5'-CTTTTTCTTTCTT 3'-GAAAAAGAGAAA</td>
<td>20.8</td>
<td>7.4</td>
</tr>
<tr>
<td>5'-CTTTTTCTTTCTT 3'-GAAAAAGAGAAA</td>
<td>22.2</td>
<td>7.6</td>
</tr>
<tr>
<td>5'-CTTTTTCTTTCTT 3'-GAAAAAGAGAAA</td>
<td>19.7</td>
<td>7.4</td>
</tr>
<tr>
<td>5'-CTTTTTCTTTCTT 3'-GAAAAAGAGAAA</td>
<td>17.6</td>
<td>6.9</td>
</tr>
</tbody>
</table>

$^a$ Conditions: 100 mM NaCl, 10 mM MgCl₂, 10 mM Na·Pipes, pH 7.0, 1.6 μM each strand.
How important are H-bonds in DNA?

Table 1. Free Energies and Melting Temperatures for Dodecamer Duplexes Containing a Variable T–X, F–X, B–X, or D–X Base Pair (X = A, T, C, G)

<table>
<thead>
<tr>
<th>duplex</th>
<th>$T_m$ (°C)</th>
<th>$\Delta G^\circ_{298}$ (kcal)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5'-CTTTCCTTTTCT</td>
<td>39.4</td>
<td>12.3</td>
</tr>
<tr>
<td>3'-GAAAAAGAAAAA</td>
<td>26.4</td>
<td>8.7</td>
</tr>
<tr>
<td>5'-CTTTCCTTTTCT</td>
<td>30.7</td>
<td>9.3</td>
</tr>
<tr>
<td>3'-GAAAAAGAAAAA</td>
<td>27.1</td>
<td>8.9</td>
</tr>
<tr>
<td>5'-CTTTCCTTTTCT</td>
<td>21.4</td>
<td>7.4</td>
</tr>
<tr>
<td>3'-GAAAAAGAAAAA</td>
<td>25.0</td>
<td>8.2</td>
</tr>
<tr>
<td>5'-CTTTCCTTTTCT</td>
<td>23.0</td>
<td>8.0</td>
</tr>
<tr>
<td>3'-GAAAAAGAAAAA</td>
<td>20.2</td>
<td>7.3</td>
</tr>
<tr>
<td>5'-CTTTCCTTTTCT</td>
<td>21.0</td>
<td>7.5</td>
</tr>
<tr>
<td>3'-GAAAAAGAAAAA</td>
<td>22.9</td>
<td>7.8</td>
</tr>
<tr>
<td>5'-CTTTCCTTTTCT</td>
<td>20.1</td>
<td>7.6</td>
</tr>
<tr>
<td>3'-GAAAAAGAAAAA</td>
<td>20.3</td>
<td>6.7</td>
</tr>
<tr>
<td>5'-CTTTCCTTTTCT</td>
<td>20.8</td>
<td>7.4</td>
</tr>
<tr>
<td>3'-GAAAAAGAAAAA</td>
<td>22.2</td>
<td>7.6</td>
</tr>
<tr>
<td>5'-CTTTCCTTTTCT</td>
<td>19.7</td>
<td>7.3</td>
</tr>
<tr>
<td>3'-GAAAAAGAAAAA</td>
<td>17.6</td>
<td>6.9</td>
</tr>
</tbody>
</table>

*Conditions: 100 mM NaCl, 10 mM MgCl$_2$, 10 mM Na-PIPES, pH 7.0, 1.6 μM each strand.*
How important are H-bonds in DNA?

![Chemical structures](Image)

Table 1. Free Energies and Melting Temperatures for Dodecamer Duplexes Containing a Variable T-X, F-X, B-X, or D-X Base Pair (X = A, T, C, G)

<table>
<thead>
<tr>
<th>duplex</th>
<th>$T_m$ (°C)</th>
<th>$\Delta G^\circ_{298}$ (kcal)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5-CTTTTCTTTCTTT</td>
<td>39.4</td>
<td>12.3</td>
</tr>
<tr>
<td>3-GAAAAAGAGAAA</td>
<td>39.4</td>
<td>12.3</td>
</tr>
<tr>
<td>5-CTTTTCTTTCTTT</td>
<td>28.4</td>
<td>8.7</td>
</tr>
<tr>
<td>3-GAAAAAGAGAAA</td>
<td>28.4</td>
<td>8.7</td>
</tr>
<tr>
<td>5-CTTTTCTTTCTTT</td>
<td>30.7</td>
<td>9.3</td>
</tr>
<tr>
<td>3-GAAAAAGAGAAA</td>
<td>30.7</td>
<td>9.3</td>
</tr>
<tr>
<td>5-CTTTTCTTTCTTT</td>
<td>30.7</td>
<td>9.3</td>
</tr>
<tr>
<td>3-GAAAAAGAGAAA</td>
<td>30.7</td>
<td>9.3</td>
</tr>
<tr>
<td>5-CTTTTCTTTCTTT</td>
<td>27.1</td>
<td>8.9</td>
</tr>
<tr>
<td>3-GAAAAAGAGAAA</td>
<td>27.1</td>
<td>8.9</td>
</tr>
<tr>
<td>5-CTTTTCTTTCTTT</td>
<td>39.4</td>
<td>12.3</td>
</tr>
<tr>
<td>3-GAAAAAGAGAAA</td>
<td>39.4</td>
<td>12.3</td>
</tr>
<tr>
<td>5-CTTTTCTTTCTTT</td>
<td>21.4</td>
<td>7.4</td>
</tr>
<tr>
<td>3-GAAAAAGAGAAA</td>
<td>21.4</td>
<td>7.4</td>
</tr>
<tr>
<td>5-CTTTTCTTTCTTT</td>
<td>25.0</td>
<td>8.2</td>
</tr>
<tr>
<td>3-GAAAAAGAGAAA</td>
<td>25.0</td>
<td>8.2</td>
</tr>
<tr>
<td>5-CTTTTCTTTCTTT</td>
<td>23.0</td>
<td>8.0</td>
</tr>
<tr>
<td>3-GAAAAAGAGAAA</td>
<td>23.0</td>
<td>8.0</td>
</tr>
<tr>
<td>5-CTTTTCTTTCTTT</td>
<td>20.2</td>
<td>7.3</td>
</tr>
<tr>
<td>3-GAAAAAGAGAAA</td>
<td>20.2</td>
<td>7.3</td>
</tr>
<tr>
<td>5-CTTTTCTTTCTTT</td>
<td>21.0</td>
<td>7.5</td>
</tr>
<tr>
<td>3-GAAAAAGAGAAA</td>
<td>21.0</td>
<td>7.5</td>
</tr>
<tr>
<td>5-CTTTTCTTTCTTT</td>
<td>22.9</td>
<td>7.8</td>
</tr>
<tr>
<td>3-GAAAAAGAGAAA</td>
<td>22.9</td>
<td>7.8</td>
</tr>
<tr>
<td>5-CTTTTCTTTCTTT</td>
<td>20.1</td>
<td>7.6</td>
</tr>
<tr>
<td>3-GAAAAAGAGAAA</td>
<td>20.1</td>
<td>7.6</td>
</tr>
<tr>
<td>5-CTTTTCTTTCTTT</td>
<td>20.3</td>
<td>6.7</td>
</tr>
<tr>
<td>3-GAAAAAGAGAAA</td>
<td>20.3</td>
<td>6.7</td>
</tr>
<tr>
<td>5-CTTTTCTTTCTTT</td>
<td>20.3</td>
<td>6.7</td>
</tr>
<tr>
<td>3-GAAAAAGAGAAA</td>
<td>20.3</td>
<td>6.7</td>
</tr>
<tr>
<td>5-CTTTTCTTTCTTT</td>
<td>20.3</td>
<td>6.7</td>
</tr>
<tr>
<td>3-GAAAAAGAGAAA</td>
<td>20.3</td>
<td>6.7</td>
</tr>
<tr>
<td>5-CTTTTCTTTCTTT</td>
<td>20.3</td>
<td>6.7</td>
</tr>
<tr>
<td>3-GAAAAAGAGAAA</td>
<td>20.3</td>
<td>6.7</td>
</tr>
<tr>
<td>5-CTTTTCTTTCTTT</td>
<td>22.2</td>
<td>7.6</td>
</tr>
<tr>
<td>3-GAAAAAGAGAAA</td>
<td>22.2</td>
<td>7.6</td>
</tr>
<tr>
<td>5-CTTTTCTTTCTTT</td>
<td>19.7</td>
<td>7.4</td>
</tr>
<tr>
<td>3-GAAAAAGAGAAA</td>
<td>19.7</td>
<td>7.4</td>
</tr>
<tr>
<td>5-CTTTTCTTTCTTT</td>
<td>17.6</td>
<td>6.9</td>
</tr>
<tr>
<td>3-GAAAAAGAGAAA</td>
<td>17.6</td>
<td>6.9</td>
</tr>
</tbody>
</table>

$^a$ Conditions: 100 mM NaCl, 10 mM MgCl$_2$, 10 mM Na·Pipes, pH 7.0, 1.6 μM each strand.
How important are H-bonds in DNA?

Table 1. Free Energies and Melting Temperatures for Dodecamer Duplexes Containing a Variable T–X, F–X, B–X, or D–X Base Pair (X = A, T, C, G)

<table>
<thead>
<tr>
<th>duplex</th>
<th>$T_m$ (°C)</th>
<th>Δ$G^{m-29}$ (kcal)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5'-CTTTTTTTTCTTT</td>
<td>39.4</td>
<td>12.3</td>
</tr>
<tr>
<td>3'-AAAAAGAAGAAA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5'-CTTTTTTTTCTTT</td>
<td>26.4</td>
<td>8.7</td>
</tr>
<tr>
<td>3'-AAAAAGAAGAAA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5'-CTTTTTTTTCTTT</td>
<td>30.7</td>
<td>9.3</td>
</tr>
<tr>
<td>3'-AAAAAGAAGAAA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5'-CTTTTTTTTCTTT</td>
<td>27.1</td>
<td>8.9</td>
</tr>
<tr>
<td>3'-AAAAAGAAGAAA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5'-CTTTTTTTTCTTT</td>
<td>21.4</td>
<td>7.4</td>
</tr>
<tr>
<td>3'-AAAAAGAAGAAA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5'-CTTTTTTTTCTTT</td>
<td>25.0</td>
<td>8.2</td>
</tr>
<tr>
<td>3'-AAAAAGAAGAAA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5'-CTTTTTTTTCTTT</td>
<td>23.0</td>
<td>8.0</td>
</tr>
<tr>
<td>3'-AAAAAGAAGAAA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5'-CTTTTTTTTCTTT</td>
<td>20.2</td>
<td>7.3</td>
</tr>
<tr>
<td>3'-AAAAAGAAGAAA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5'-CTTTTTTTTCTTT</td>
<td>21.0</td>
<td>7.5</td>
</tr>
<tr>
<td>3'-AAAAAGAAGAAA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5'-CTTTTTTTTCTTT</td>
<td>22.9</td>
<td>7.8</td>
</tr>
<tr>
<td>3'-AAAAAGAAGAAA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5'-CTTTTTTTTCTTT</td>
<td>20.1</td>
<td>7.6</td>
</tr>
<tr>
<td>3'-AAAAAGAAGAAA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5'-CTTTTTTTTCTTT</td>
<td>20.3</td>
<td>6.7</td>
</tr>
<tr>
<td>3'-AAAAAGAAGAAA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5'-CTTTTTTTTCTTT</td>
<td>20.8</td>
<td>7.4</td>
</tr>
<tr>
<td>3'-AAAAAGAAGAAA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5'-CTTTTTTTTCTTT</td>
<td>22.2</td>
<td>7.6</td>
</tr>
<tr>
<td>3'-AAAAAGAAGAAA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5'-CTTTTTTTTCTTT</td>
<td>19.7</td>
<td>7.4</td>
</tr>
<tr>
<td>3'-AAAAAGAAGAAA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5'-CTTTTTTTTCTTT</td>
<td>17.6</td>
<td>6.9</td>
</tr>
<tr>
<td>3'-AAAAAGAAGAAA</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Conditions: 100 mM NaCl, 10 mM MgCl₂, 10 mM Na·PIPES, pH 7.0, 1.6 μM each strand.*
How important are H-bonds in DNA?

Table 1. Free Energies and Melting Temperatures for Dodecamer Duplexes Containing a Variable T–X, F–X, B–X, or D–X Base Pair (X = A, T, C, G)

<table>
<thead>
<tr>
<th>duplex</th>
<th>$T_m$ (°C)</th>
<th>$\Delta G^\circ_{DSS}$ (kcal)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5'-CTTTGCTTCTCTT 3'-GAAGAAGAAGAAG</td>
<td>39.4</td>
<td>12.3</td>
</tr>
<tr>
<td>5'-CTTTGCTTCTCTT 3'-GAAGAAGAAGAAG</td>
<td>28.4</td>
<td>8.7</td>
</tr>
<tr>
<td>5'-CTTTGCTTCTCTT 3'-GAAGAAGAAGAAG</td>
<td>30.7</td>
<td>9.3</td>
</tr>
<tr>
<td>5'-CTTTGCTTCTCTT 3'-GAAGAAGAAGAAG</td>
<td>27.1</td>
<td>8.9</td>
</tr>
<tr>
<td>5'-CTTTGCTTCTCTT 3'-GAAGAAGAAGAAG</td>
<td>21.4</td>
<td>7.4</td>
</tr>
<tr>
<td>5'-CTTTGCTTCTCTT 3'-GAAGAAGAAGAAG</td>
<td>25.0</td>
<td>8.2</td>
</tr>
<tr>
<td>5'-CTTTGCTTCTCTT 3'-GAAGAAGAAGAAG</td>
<td>23.0</td>
<td>8.0</td>
</tr>
<tr>
<td>5'-CTTTGCTTCTCTT 3'-GAAGAAGAAGAAG</td>
<td>20.2</td>
<td>7.3</td>
</tr>
<tr>
<td>5'-CTTTGCTTCTCTT 3'-GAAGAAGAAGAAG</td>
<td>21.0</td>
<td>7.5</td>
</tr>
<tr>
<td>5'-CTTTGCTTCTCTT 3'-GAAGAAGAAGAAG</td>
<td>22.9</td>
<td>7.8</td>
</tr>
<tr>
<td>5'-CTTTGCTTCTCTT 3'-GAAGAAGAAGAAG</td>
<td>20.1</td>
<td>7.6</td>
</tr>
<tr>
<td>5'-CTTTGCTTCTCTT 3'-GAAGAAGAAGAAG</td>
<td>20.3</td>
<td>6.7</td>
</tr>
<tr>
<td>5'-CTTTGCTTCTCTT 3'-GAAGAAGAAGAAG</td>
<td>20.8</td>
<td>7.4</td>
</tr>
<tr>
<td>5'-CTTTGCTTCTCTT 3'-GAAGAAGAAGAAG</td>
<td>22.2</td>
<td>7.6</td>
</tr>
<tr>
<td>5'-CTTTGCTTCTCTT 3'-GAAGAAGAAGAAG</td>
<td>19.7</td>
<td>7.4</td>
</tr>
<tr>
<td>5'-CTTTGCTTCTCTT 3'-GAAGAAGAAGAAG</td>
<td>17.6</td>
<td>6.9</td>
</tr>
</tbody>
</table>

$^a$ Conditions: 100 mM NaCl, 10 mM MgCl$_2$, 10 mM Na-PIPES, pH 7.0, 1.6 μM each strand.
Burial of hydrophobic surface drives helix formation (hydrophobic core / stacking interactions)
Burial of hydrophobic surface drives helix formation (hydrophobic core / stacking interactions)

Flat faces are nonpolar
Burial of hydrophobic surface drives helix formation
(hydrophobic core / stacking interactions)

Flat faces are nonpolar
Edges are very polar (can H-bond)
Other chemical constraints
Furanose Sugar Ring
Furanose Sugar Ring
Furanose Sugar Ring
Furanose Sugar Ring

Eclipsed
Planar
Staggered
Puckered
Furanose Sugar Ring

C3' exo

C2' exo

Unpuckered

C2' endo

C3' endo

C2' endo - C3' exo

C2' exo - C3' endo

endo

exo

endos

exos

Staggered Puckered

Eclipsed Planar
A crystal structure from July 2004 shows that an error correcting (and error-prone) DNA polymerase uses Hoogsteen-WC base pairing to recognize the incoming substrate dNTP.

This supports biochemical studies that had been suggesting such a pairing.

This polymerase is good at bypassing lesions in the DNA (eg, damaged Watson-Crick face of G).
A crystal structure from July 2004 shows that an error correcting (and error-prone) DNA polymerase uses Hoogsteen-WC base pairing to recognize the incoming substrate dNTP.

This supports biochemical studies that had been suggesting such a pairing.

This polymerase is good at bypassing lesions in the DNA (e.g., damaged Watson-Crick face of G).
Why is Watson-Crick so good?

All four WC base pairs are isosteric.
Why is Watson-Crick so good?

All four WC base pairs are isosteric.
Why is Watson-Crick so good?

All four WC base pairs are isosteric.
Why is Watson-Crick so good?

All four WC base pairs are isosteric.
Why is Watson-Crick so good?

All four WC base pairs are *isosteric*. 
Bases that polymerases like

FIGURE 7. Examples of DNA base replacements designed to form stable pairs and/or to be replicated selectively by DNA polymerase enzymes. 17,23,24,49,54,55
Why *a helix*?
Why a helix?
Why a helix?
Why a helix?
Why *major* and *minor* grooves?
Why *major* and *minor* grooves?
Why *major* and *minor* grooves?

(a) groove
(b) groove

©1999 GARLAND PUBLISHING INC.  
A member of the Taylor & Francis Group
Why *major* and *minor* grooves?
Why is the major groove so good?

Major Groove

Minor Groove
Why is the major groove so good?

Major Groove

Minor Groove
Why is the major groove so good?
Why is the major groove so good?

*Major Groove*

*Minor Groove*
B-form DNA

B-form

Residues per turn = 10
Twist per base pair = 36°
Rise per pair = 3.4Å
c2'-endo

Minor groove width = 5.7Å
Major groove width = 11.7Å

Minor groove depth = 7.5Å
Major groove depth = 8.8Å
B-form DNA

Residues per turn = 10
Twist per base pair = 36°
Rise per pair = 3.4 Å
C2'-endo
Minor groove width = 5.7 Å
Major groove width = 11.7 Å
Minor groove depth = 7.5 Å
Major groove depth = 8.8 Å
A-form RNA

- Residues per turn = 11
- Twist per base pair = 33°
- Rise per pair = 2.9Å
- c3’-endo

- Minor groove width = 11Å
- Major groove width = 2.7Å

- Minor groove depth = 2.8Å
- Major groove depth = 13.5Å
A-form RNA

- Residues per turn = 11
- Twist per base pair = 33°
- Rise per pair = 2.9Å
- c3'-endo

- Minor groove width = 11Å
- Major groove width = 2.7Å
- Minor groove depth = 2.8Å
- Major groove depth = 13.5Å
Compare

A-form (RNA)

- Minor groove width = 11 Å
- Major groove width = 2.7 Å
- Minor groove depth = 2.8 Å
- Major groove depth = 13.5 Å

B-form (DNA)

- Minor groove width = 5.7 Å
- Major groove width = 11.7 Å
- Minor groove depth = 7.5 Å
- Major groove depth = 8.8 Å
Z-DNA

- Residues per turn = 12
- Twist per base pair = -9 / -51°
- Rise per pair = 3.7 Å
- c3'-endo(syn) / c2'-endo
- Minor groove width = 2.0 Å
  - Major groove width = 8.8 Å
  - Minor groove depth = 13.8 Å
  - Major groove depth = 3.7 Å
Ends of DNA duplexes

“Blunt” ends
Ends of DNA duplexes

“Blunt” ends
Protein - Nucleic Acid Interactions

Major Groove Interactions
Protein - Nucleic Acid Interactions

Gln

Asn

Arg

Major Groove Interactions
Nucleic Acid “Triples / Platforms”

Major Groove Interactions
Winged Helix DNA Binding Domain
Classic helix-turn-helix
Winged Helix DNA Binding Domain
Classic helix-turn-helix
Winged Helix DNA Binding Domain
Classic helix-turn-helix

Hrfx1 bound to its X-box binding site

Arg58
Simple Structure - Hairpin
Classic Structure - Pseudoknot
tRNA
AMP Aptamer
RNA
Ribosome
An RNA machine with protein cofactors
Ribosome - Secondary Structure
Science 289(5481), 905 - 920, 2000
The Complete Atomic Structure of the Large Ribosomal Subunit at 2.4 Å Resolution
Ban, Nissen, Hansen, Moore, & Steitz
The Ribosome: a wealth of RNA structure
Structural Motifs
Nucleic Acids Research, 2009, Vol. 37, No. 4 e29

Tetraloop

π-turn

Ω-turn

S2-motif

E-loop

kink-turn
Xu and Shi, “Composite RNA aptamers as functional mimics of proteins”
Nucleic Acids Research 2009 37(9):e71
Which is more stable?
(which has a higher melting temperature?)

ACCGCCACCGAAG
TGGCGGTGGCTTC

or

ACCGCCACCGAAG
TGGCGGTGGCTTA
Which is more stable?
(which has a higher melting temperature?)

ACCGCCACCGAAG
TGGCGGTGGCTTC

or

ACCGCCACCGAAG
TGGCGGTGGCTTA

51.6° C
Which is more stable?
(which has a higher melting temperature?)

ACCGCCACCGAAG
TGGCGGTGGCTTC  

or

ACCGCCACCGAAG
TGGCGGTGGCTTA  

51.6° C

52.5° C

Calculations from http://www.idtdna.com/analyze/Applications/OligoAnalyzer/