# 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

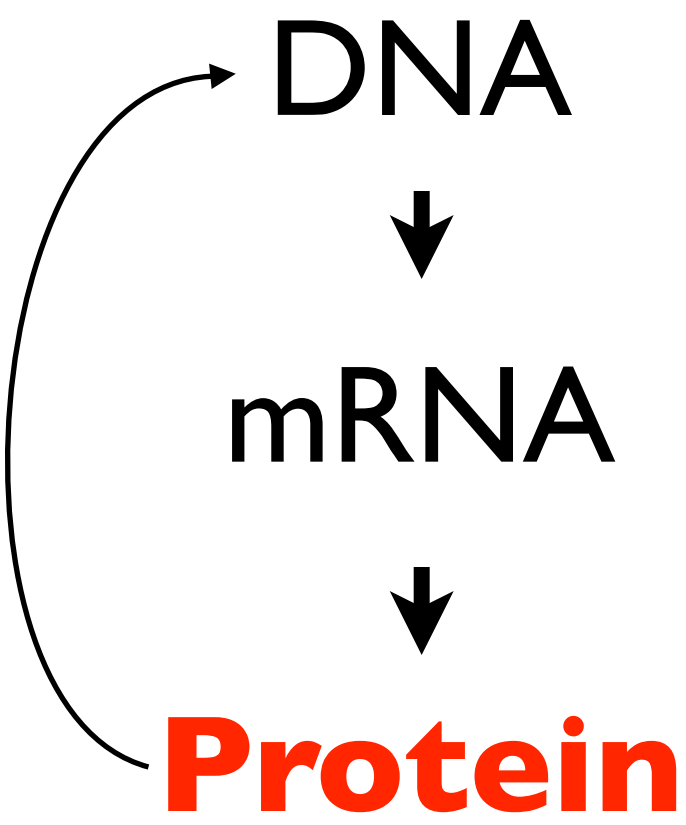
↓

mRNA

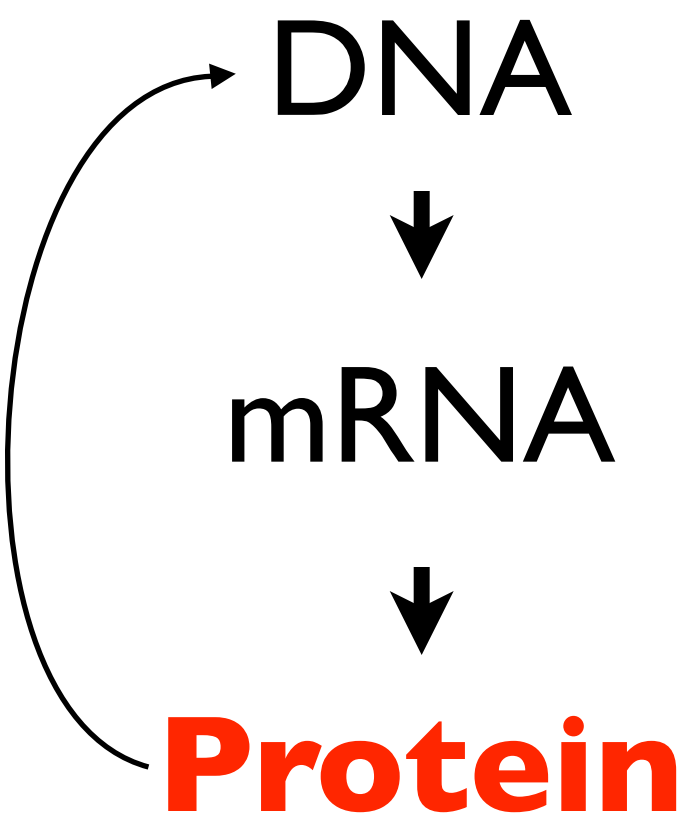
↓

**Protein**

# In the beginning…

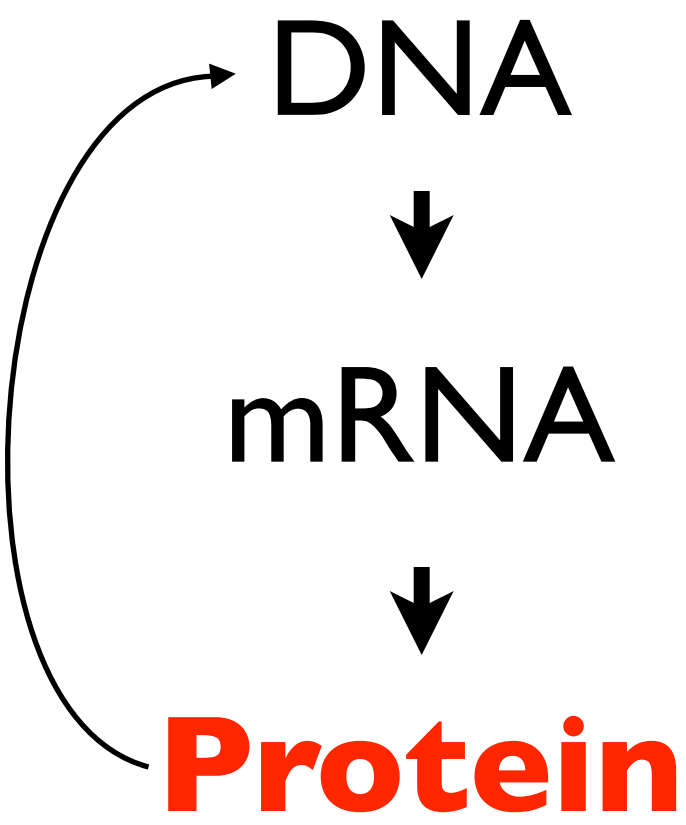


# In the beginning…




Archival information storage

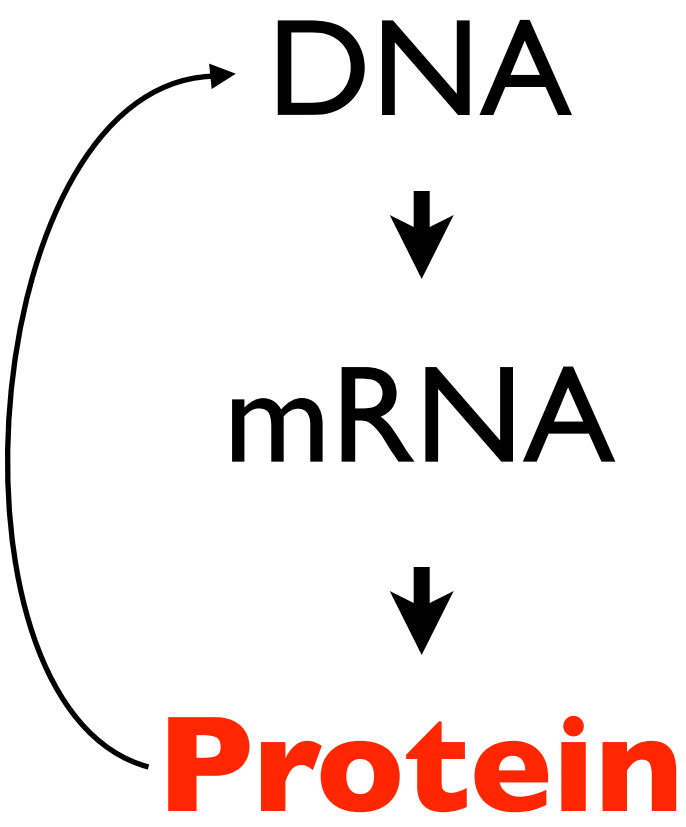
# In the beginning…




Archival information storage

Transient information storage

# In the beginning…

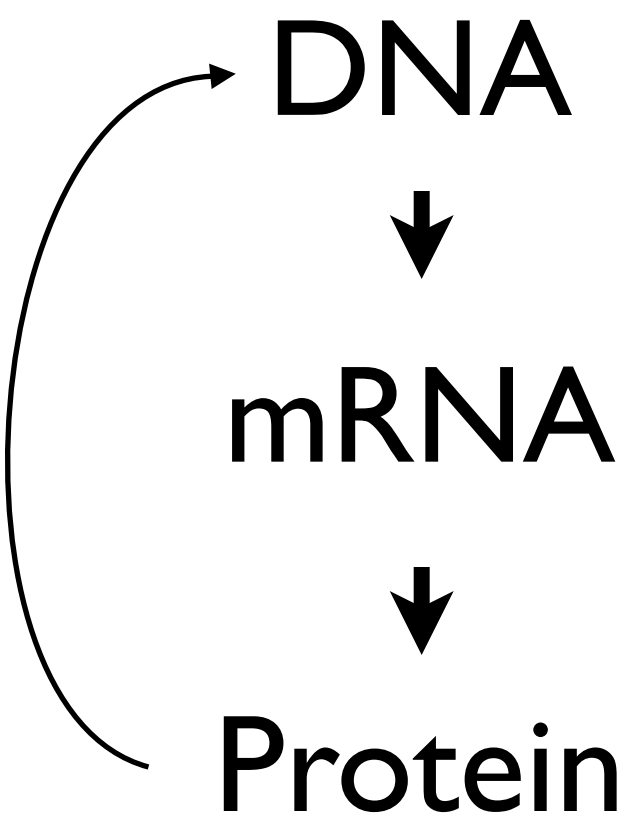



Archival information storage

Transient information storage

Catalysis, structure, regulation, et al.

# Chicken & Egg?

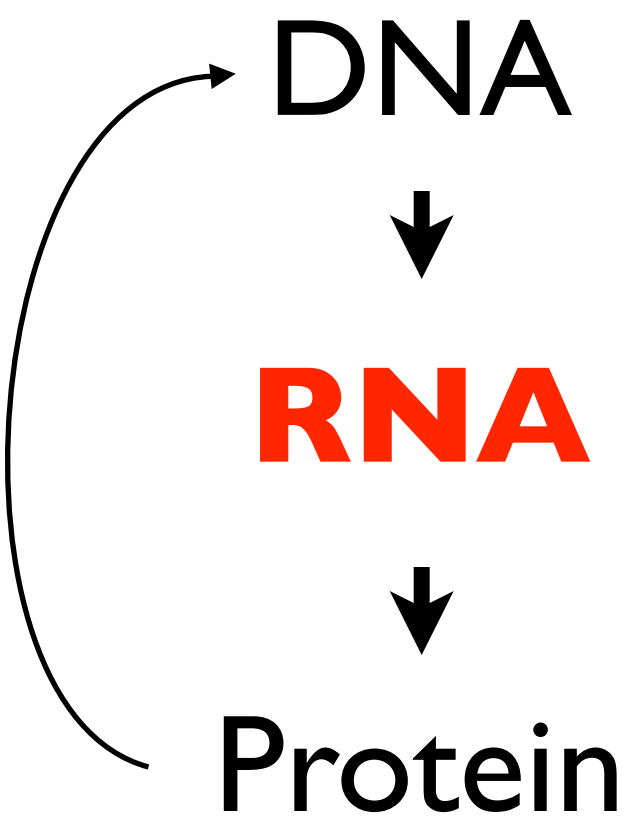



Archival information storage

Transient information storage

Catalysis, structure, regulation, et al.

# RNA can do everything



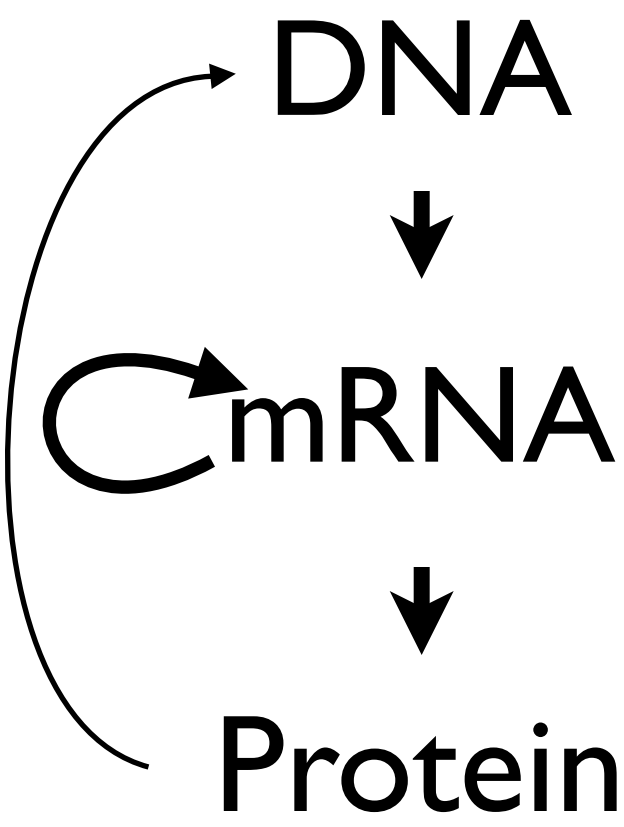

Archival information storage

Transient information storage
**Catalysis!**

1980-2000

Catalysis, structure, regulation, et al.

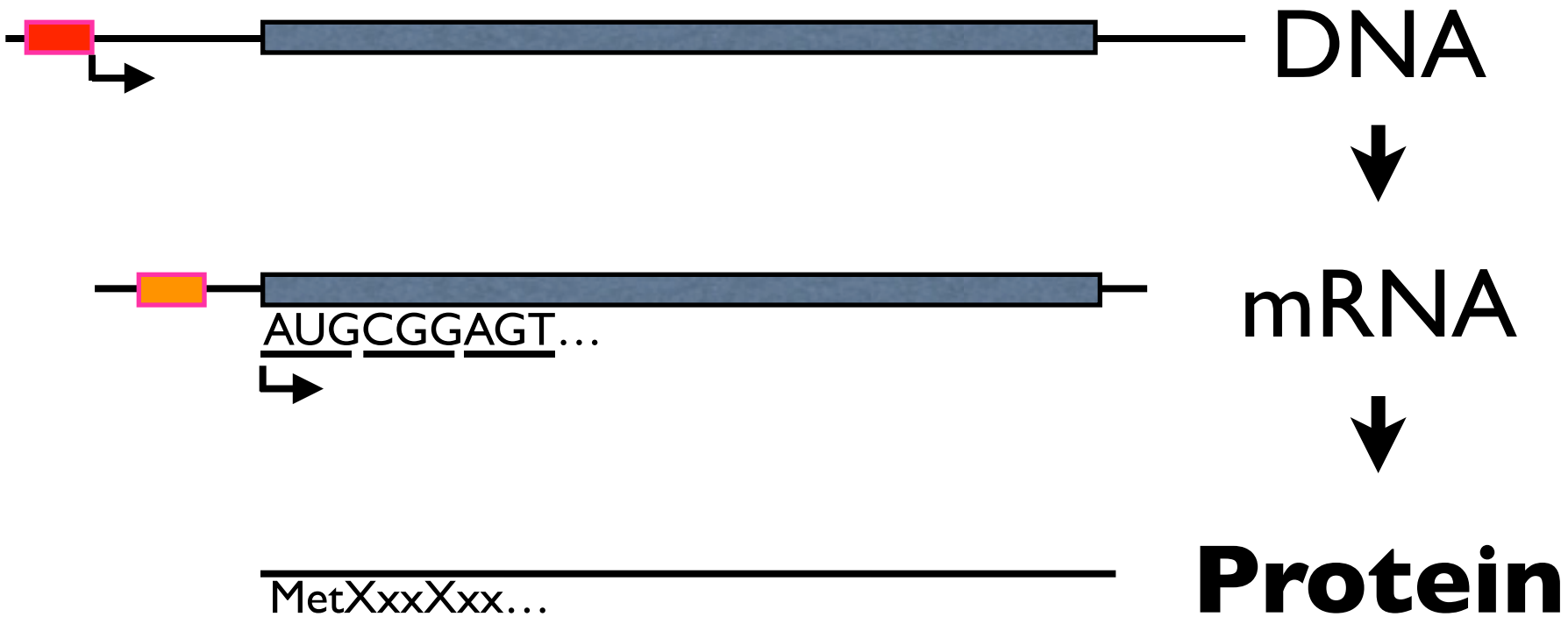
# RNA - primordial molecule




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

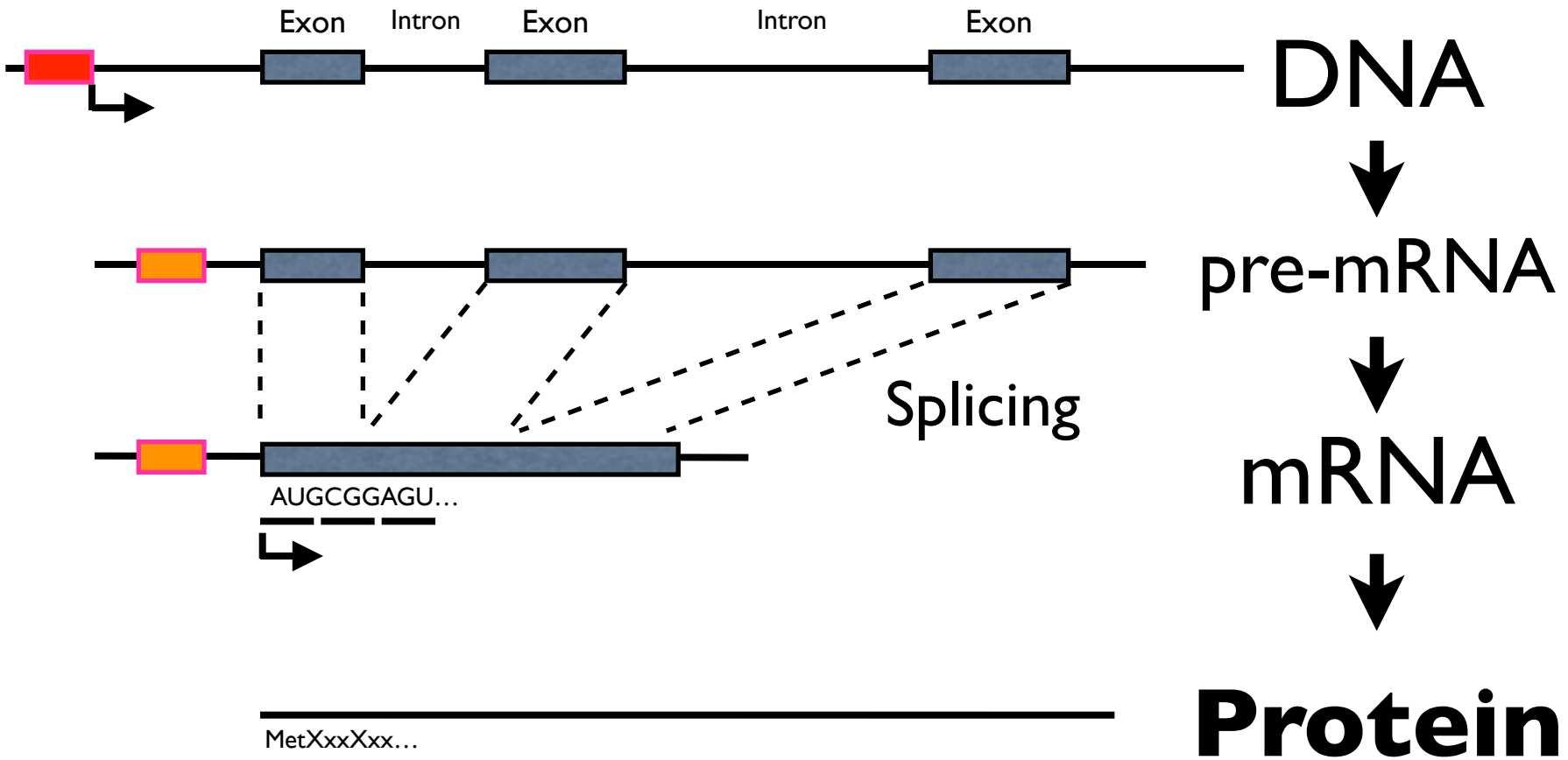
But **proteins** came along to supplant everything and make the world, evolutionarily, what it is today. All hail the protein!

# Seems simple…

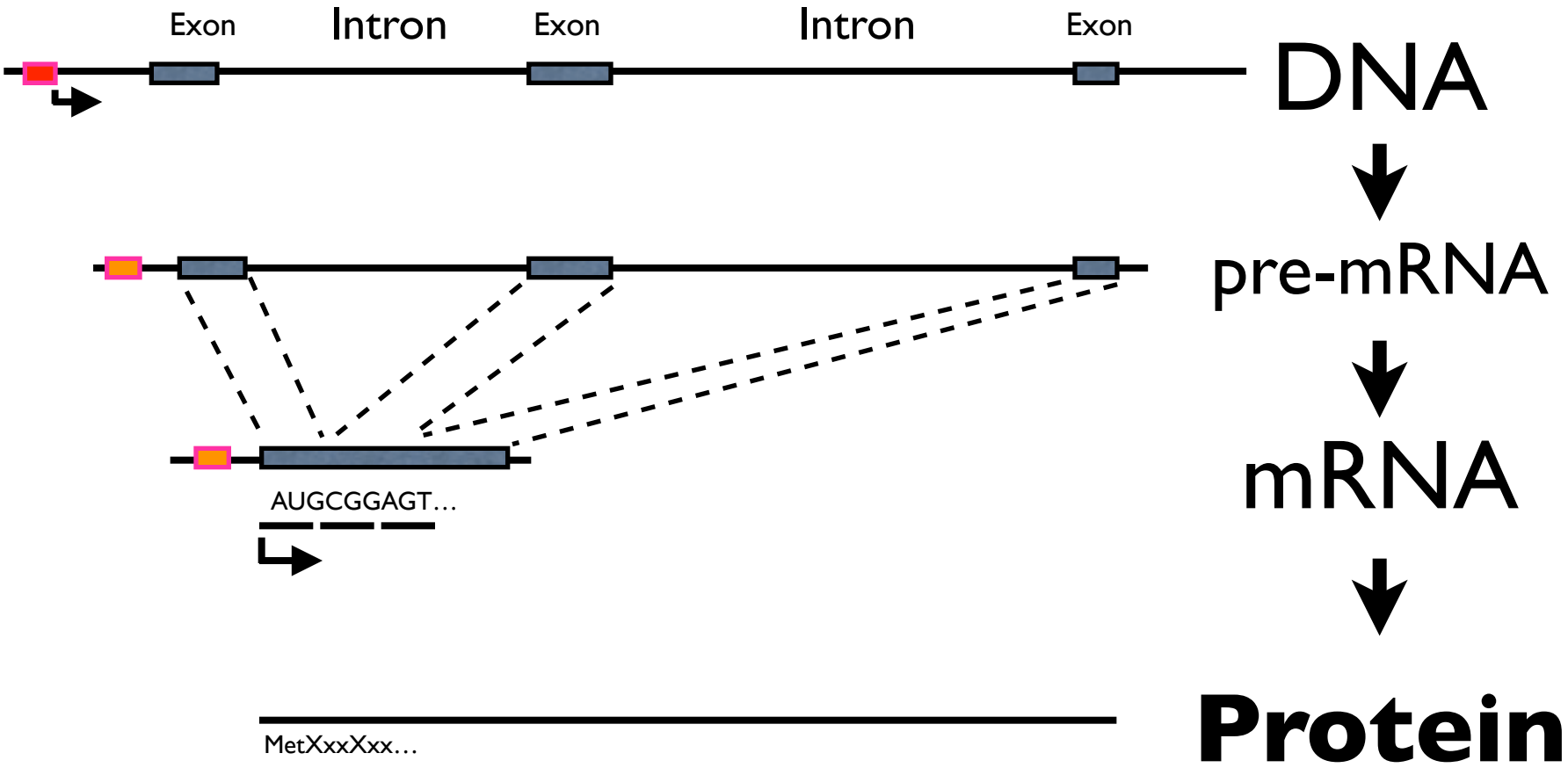


# Seems simple?

# Even more complicated…

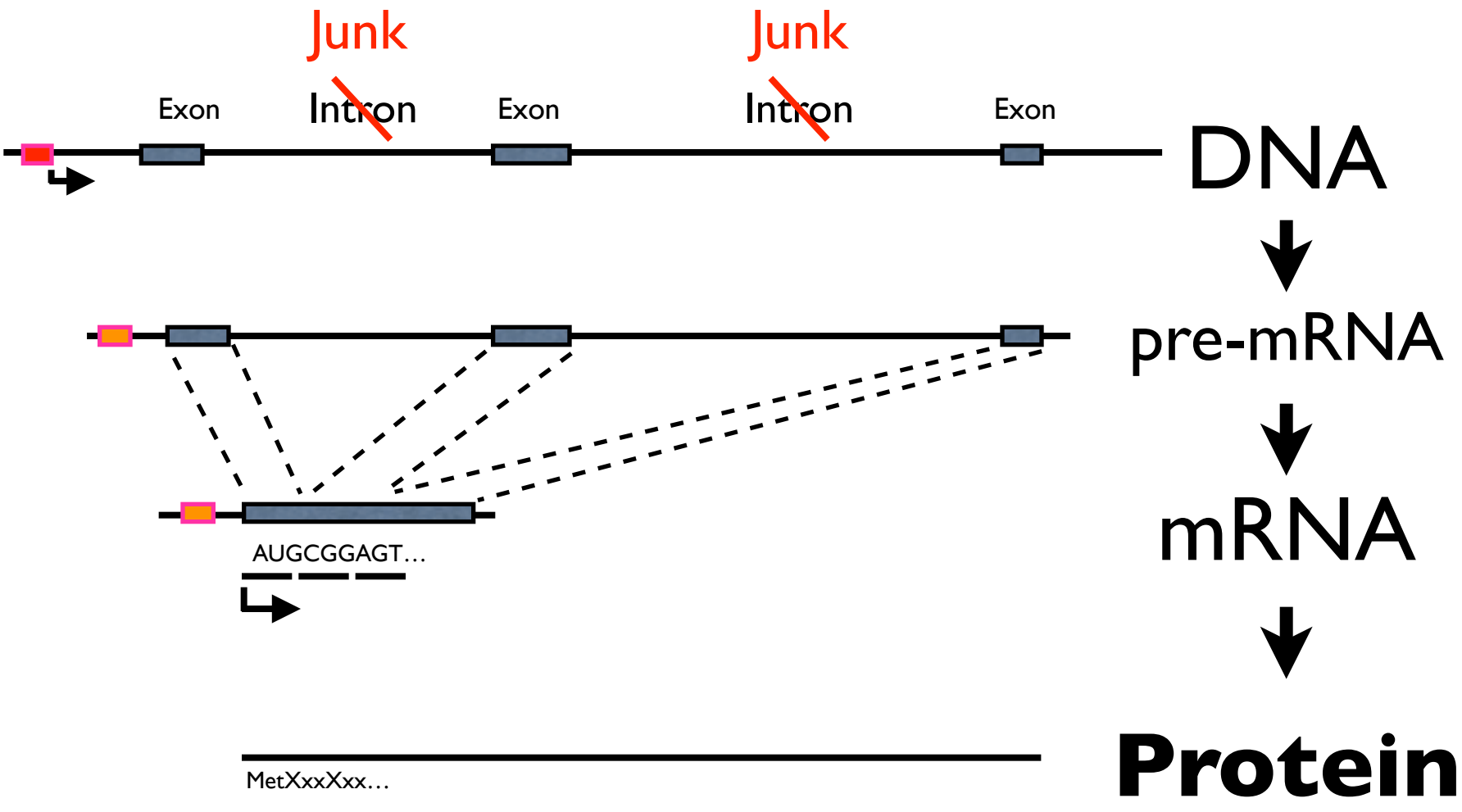


# Reality…


Exon
Intron
Exon
Intron
Exon
DNA
pre-mRNA
mRNA
AUGCGGAGT…
MetXxxXxx…
Protein

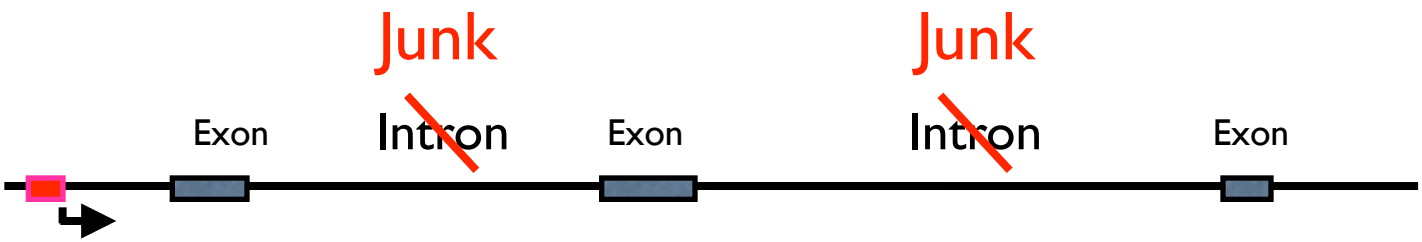
# Reality…

## and our perception of it



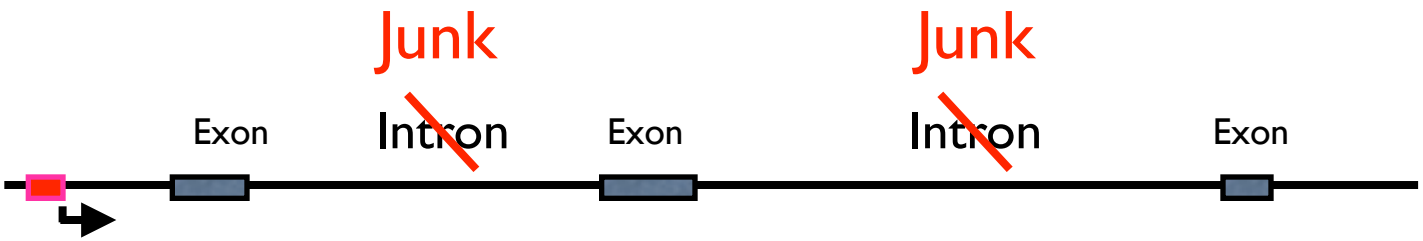
# Reality…

## 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

Polymerases

Kinases

Hydrogenases

Receptors

Oxygenases

Proteases

# 20th Century View

Polymerases

Kinases

Hydrogenases

Receptors

Oxygenases

Proteases

Nucleic acids → Trash

# 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

# 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 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

# 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

# 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

# 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

# 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 of the gene mutations that produce genetic diseases and cancers do so by effecting pre-mRNA editing

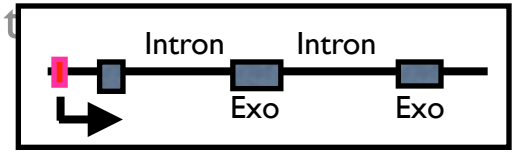
# 21st Century News

## Genome project

### How? RNA editing and alternative splicing

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

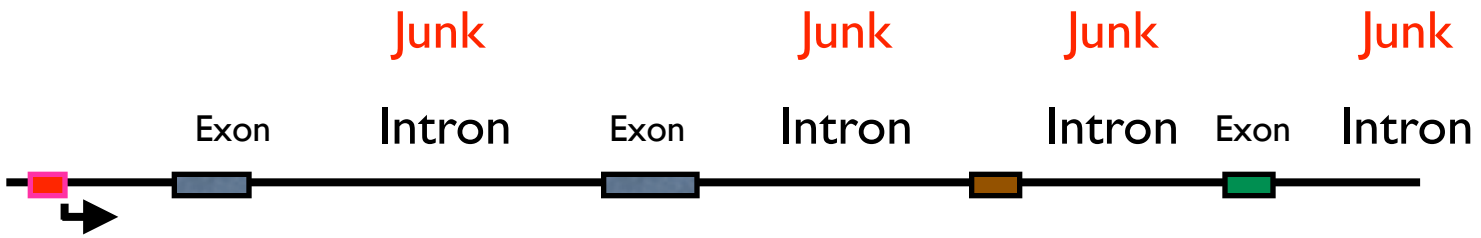
The prevalence of alternative editing appears t
with an organism's complexity



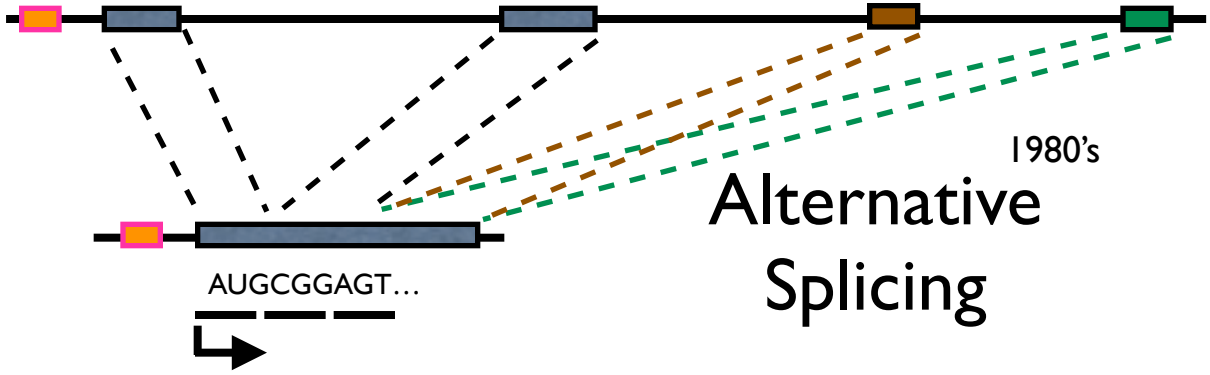

Humans have the highest number of introns per gene of any organism

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

# 21st Century News




DNA

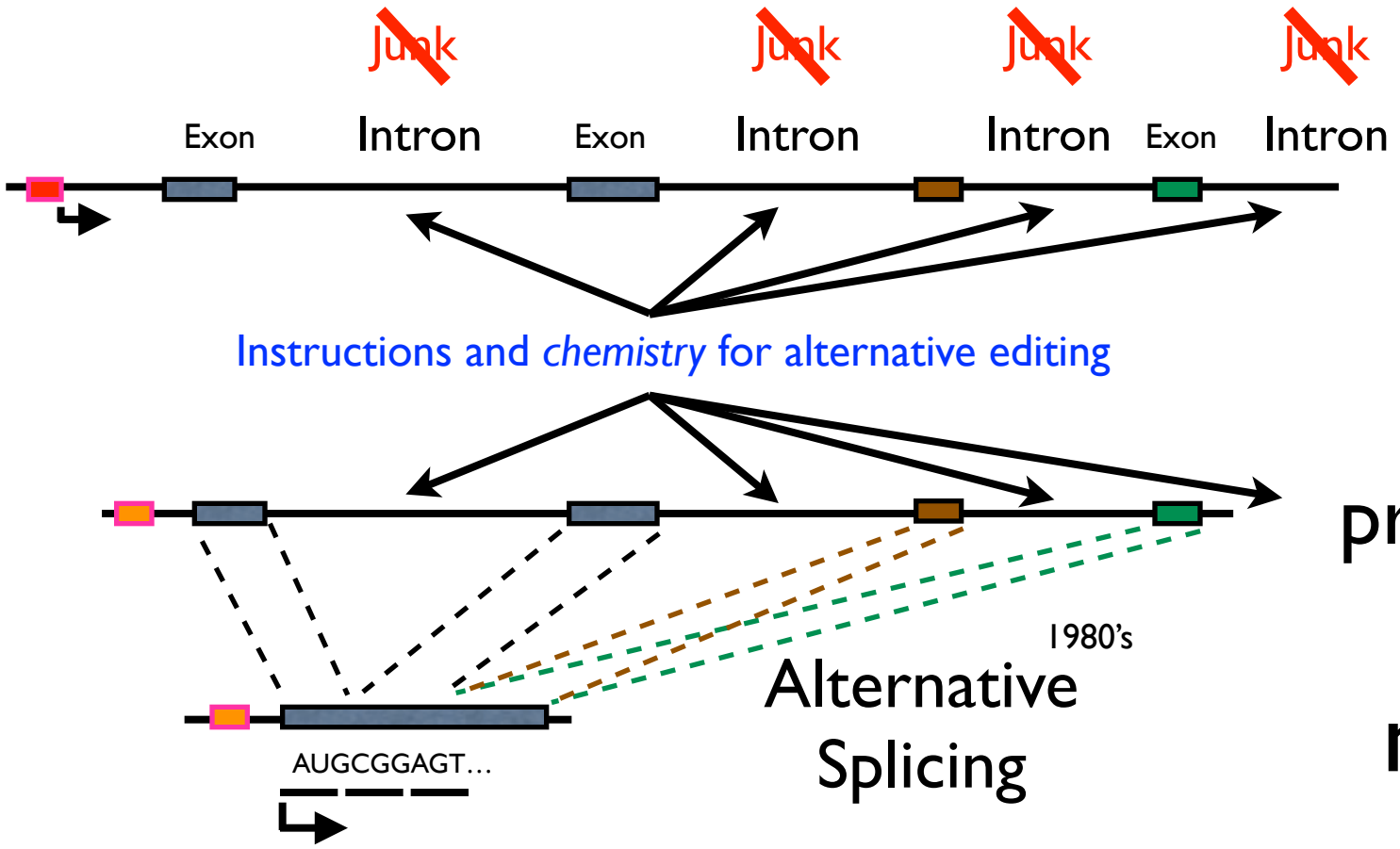



pre-mRNA

mRNA

**Protein**

# 21st Century News




DNA

pre-mRNA

mRNA

**Protein**

# 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…

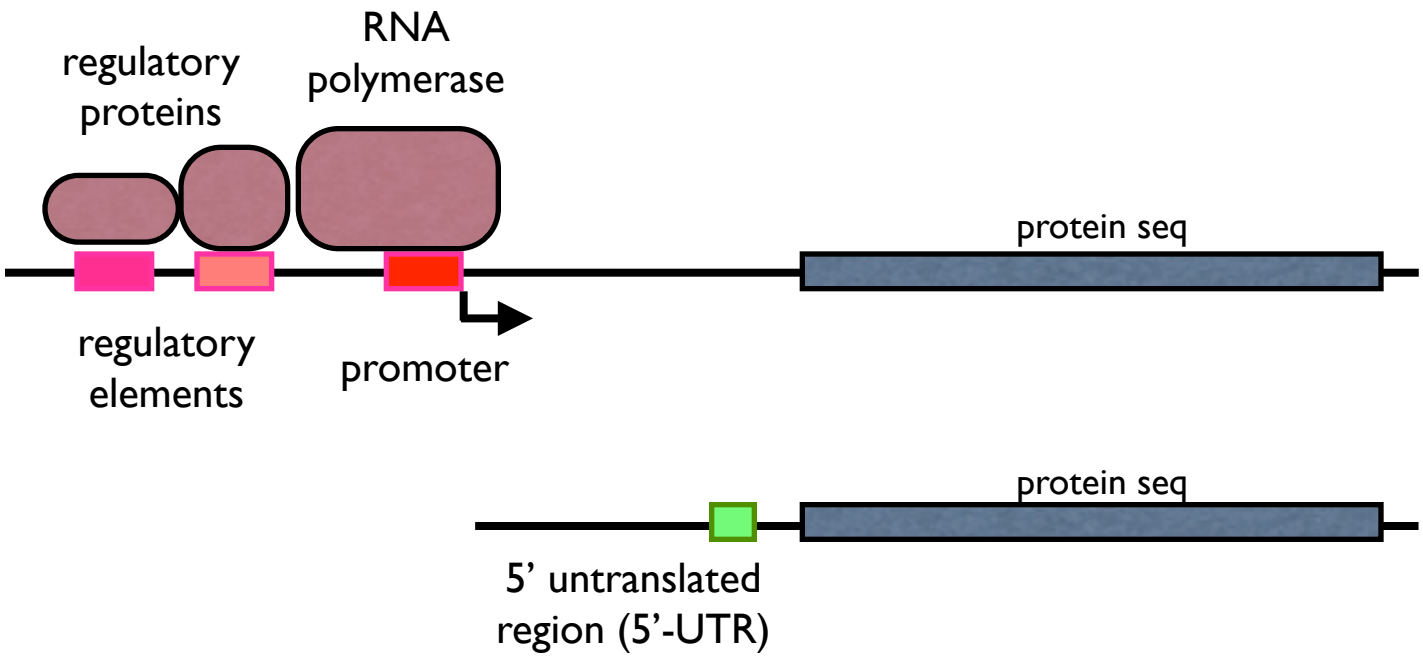
With the advent of PCR, nucleic acids have been recognized as extremely powerful combinatorial tools in the test tube

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

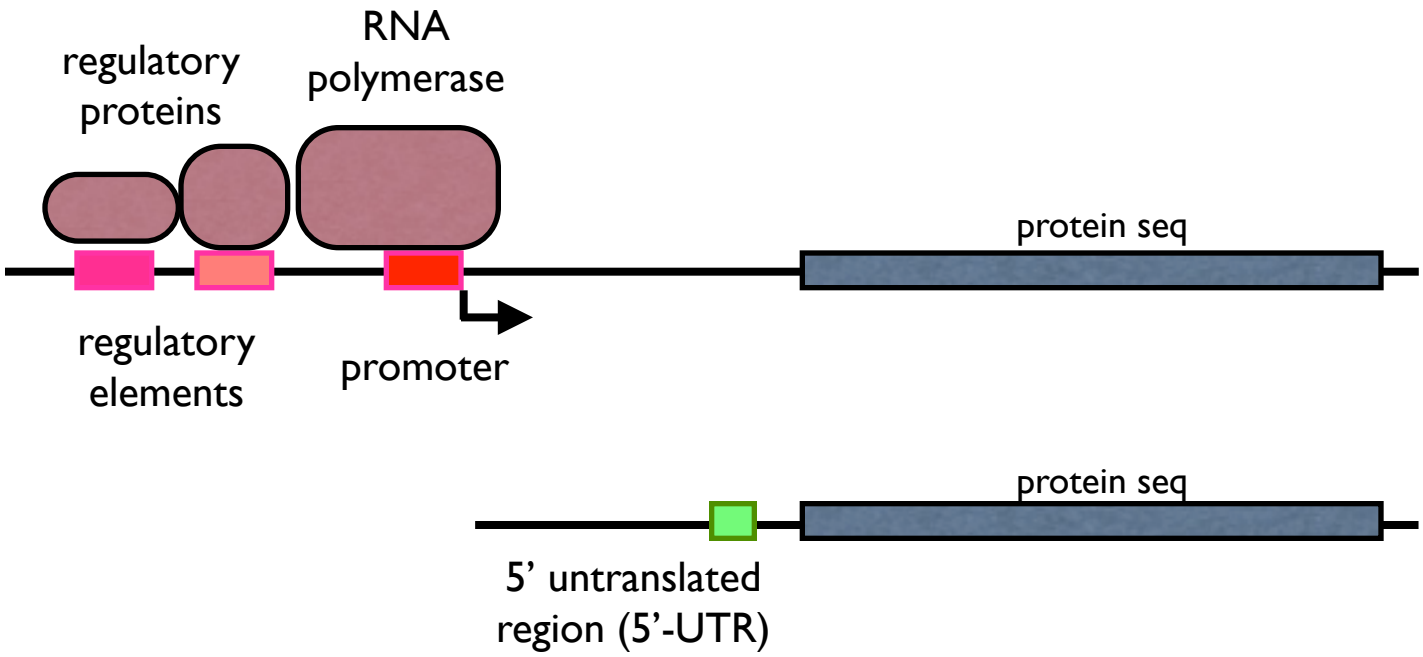
↓

mRNA

↓

Protein

# More gene regulation




DNA

↓

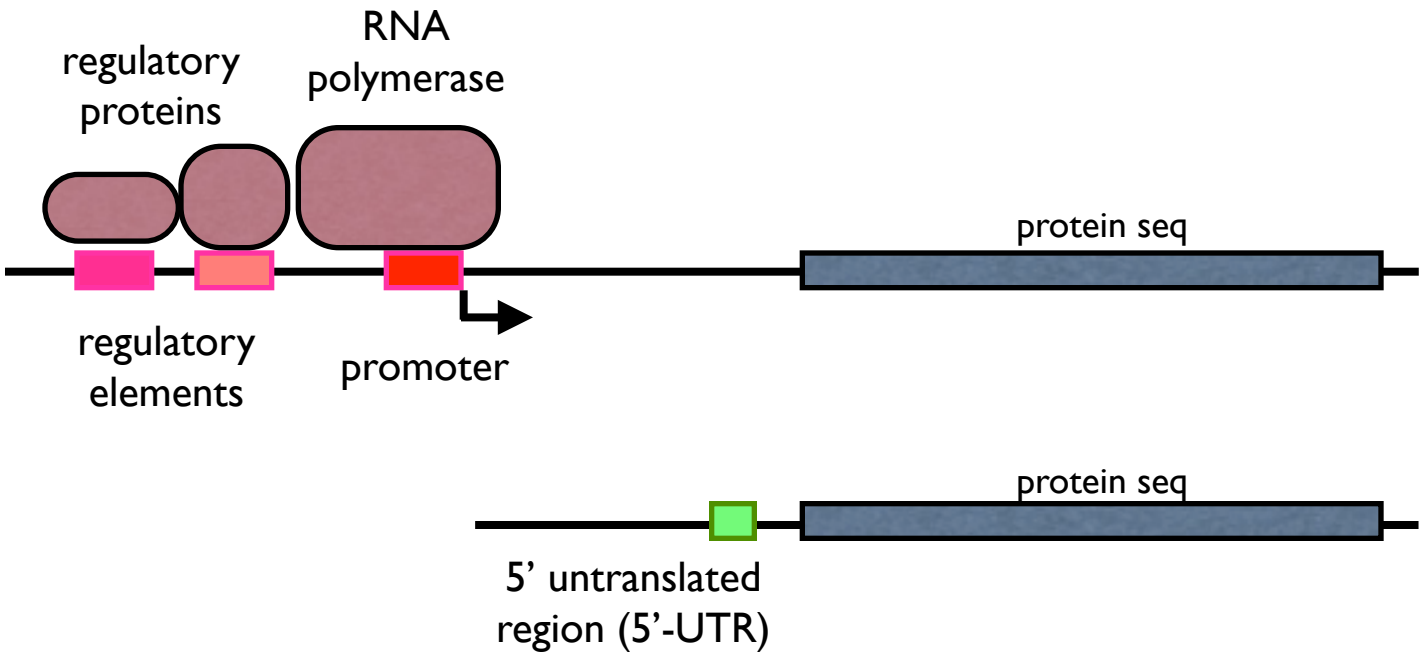
mRNA

↓

Protein

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

# More gene regulation




DNA

↓

mRNA

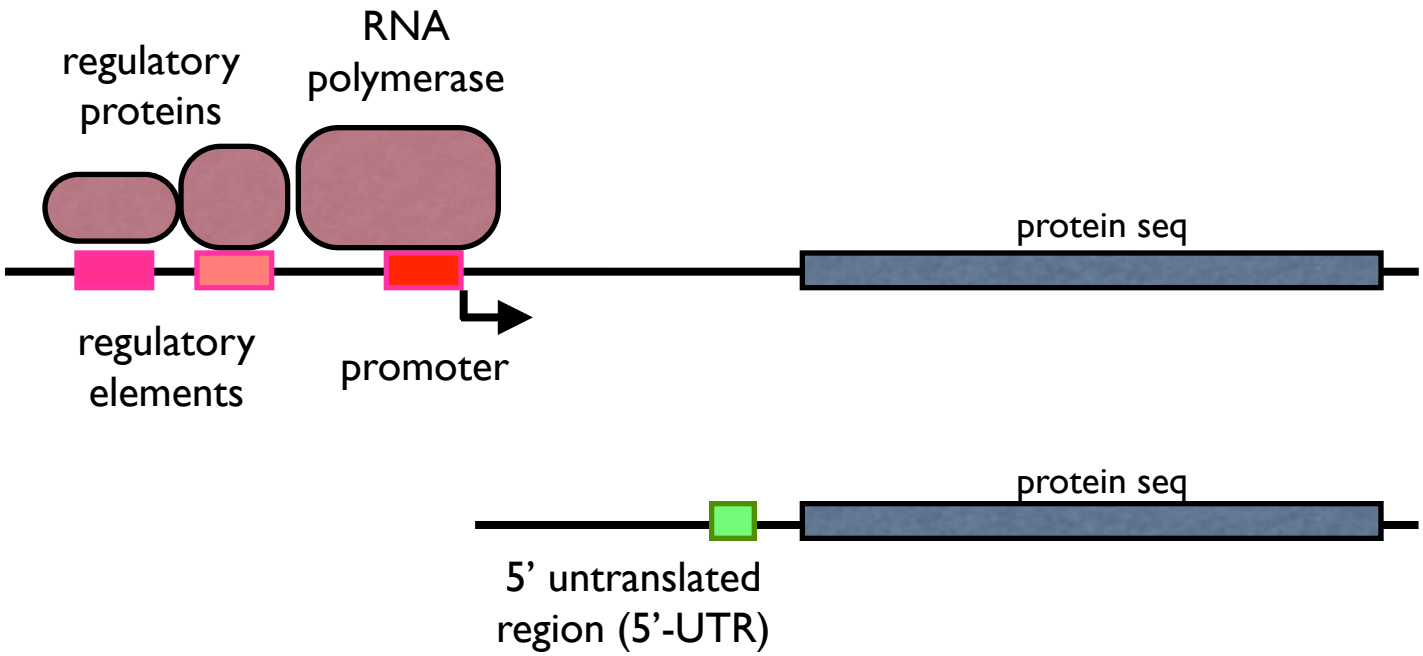
↓

Protein

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




DNA

↓

mRNA

↓

Protein

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?

**2004**

↘ One week: a Nature paper!

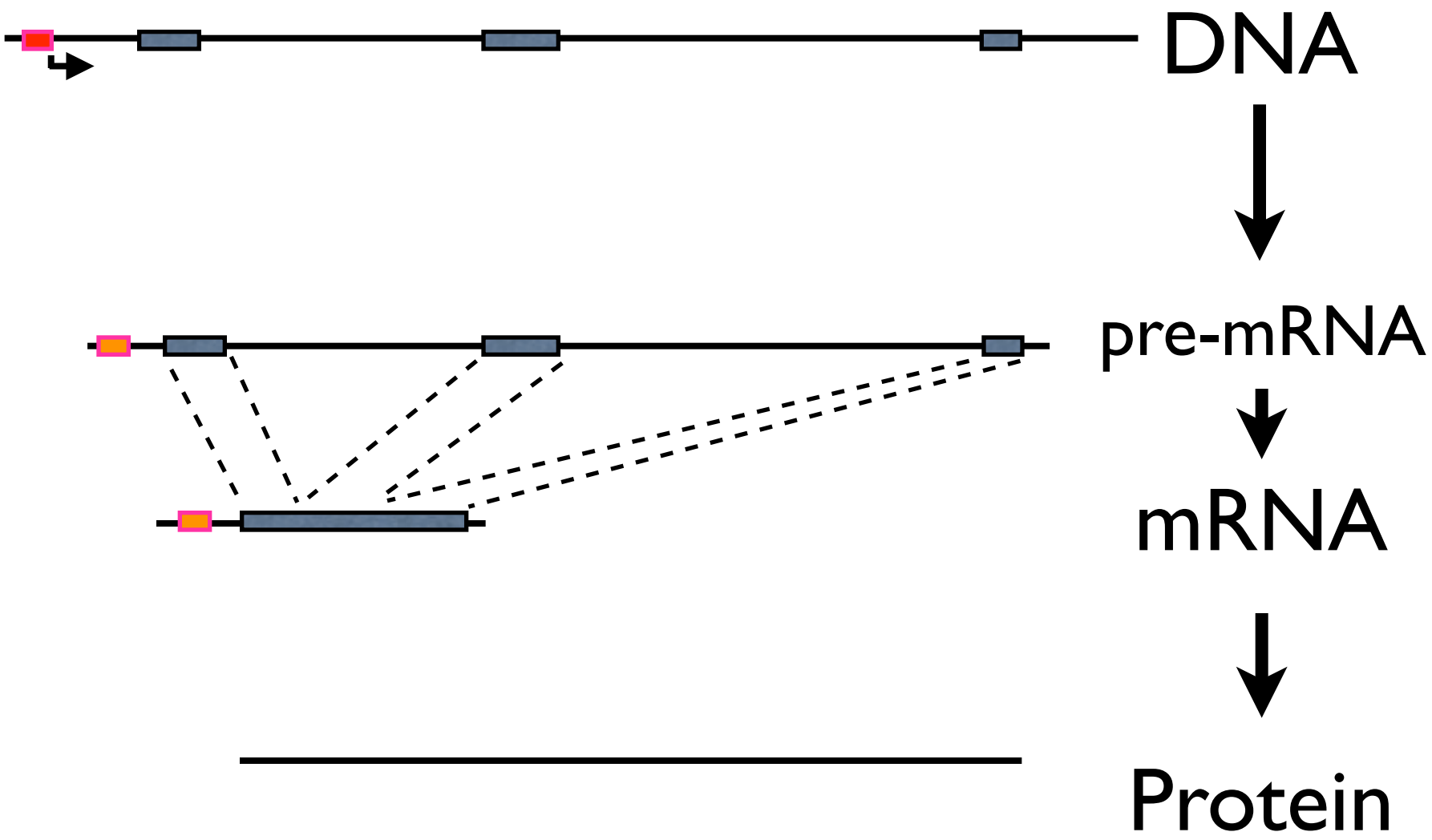
# Project Encode (2007)

(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

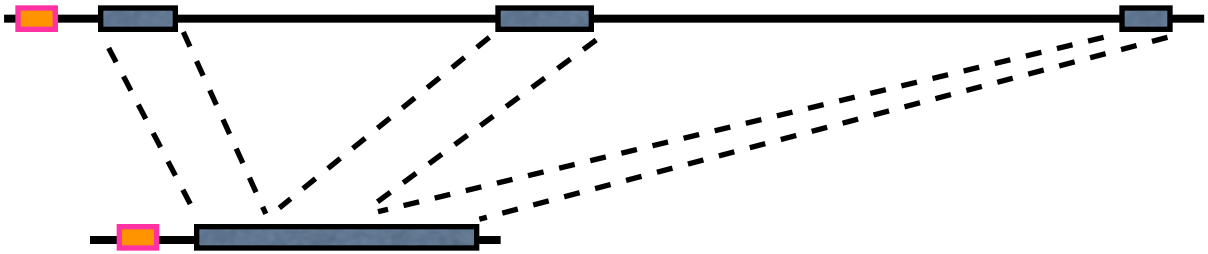


# 21st Century Opportunities

DNA

early '80s, but…

Alternative splicing

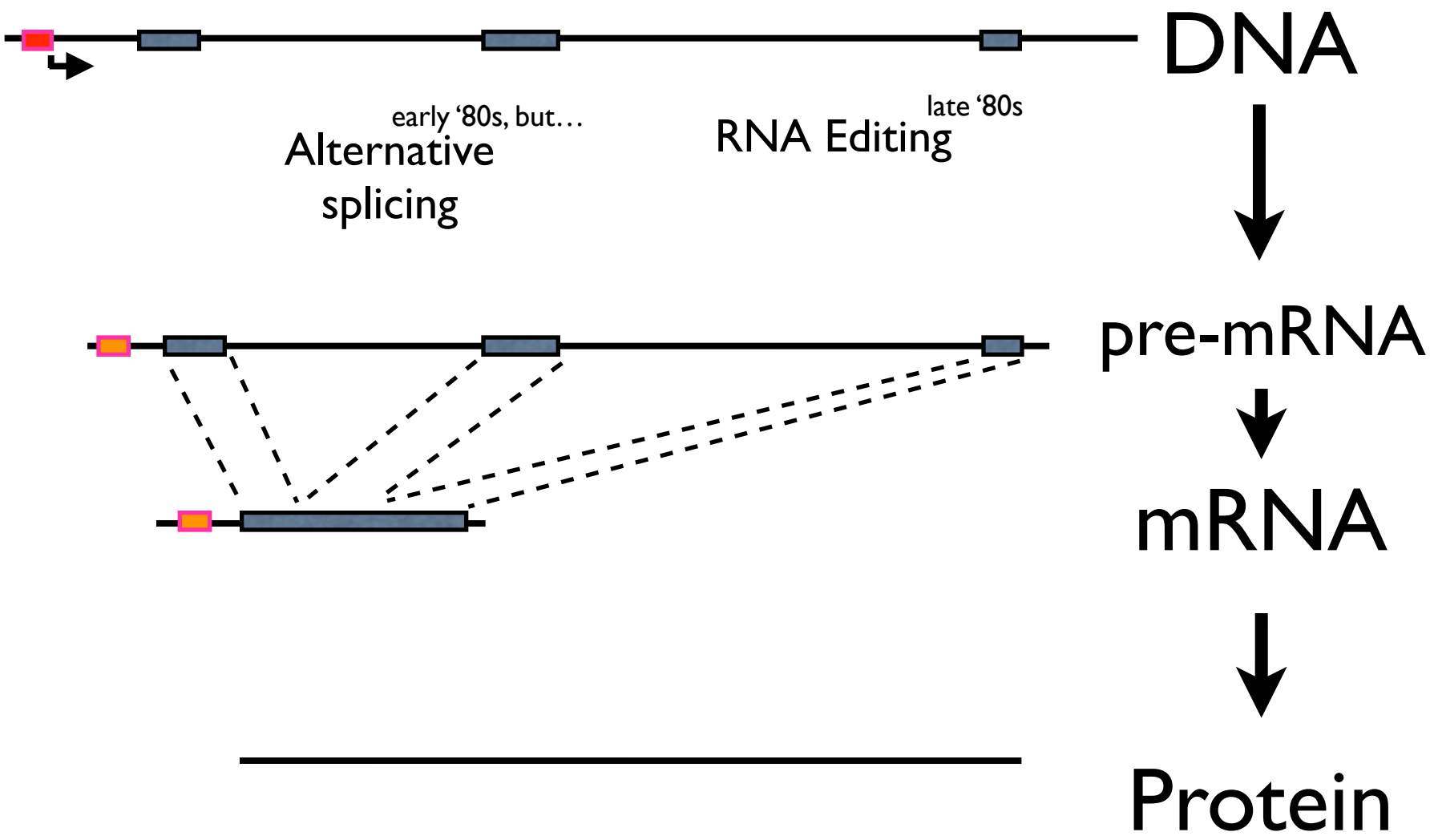


pre-mRNA

mRNA

Protein

# 21st Century Opportunities



# 21st Century Opportunities

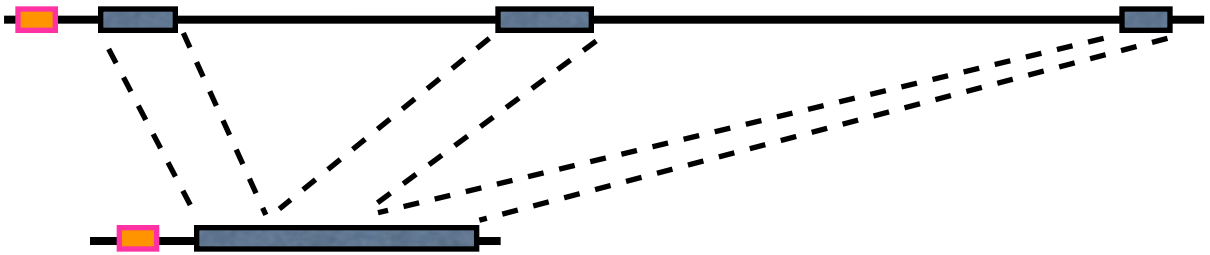
DNA

early '80s, but…

Alternative splicing

late '80s

RNA Editing



pre-mRNA

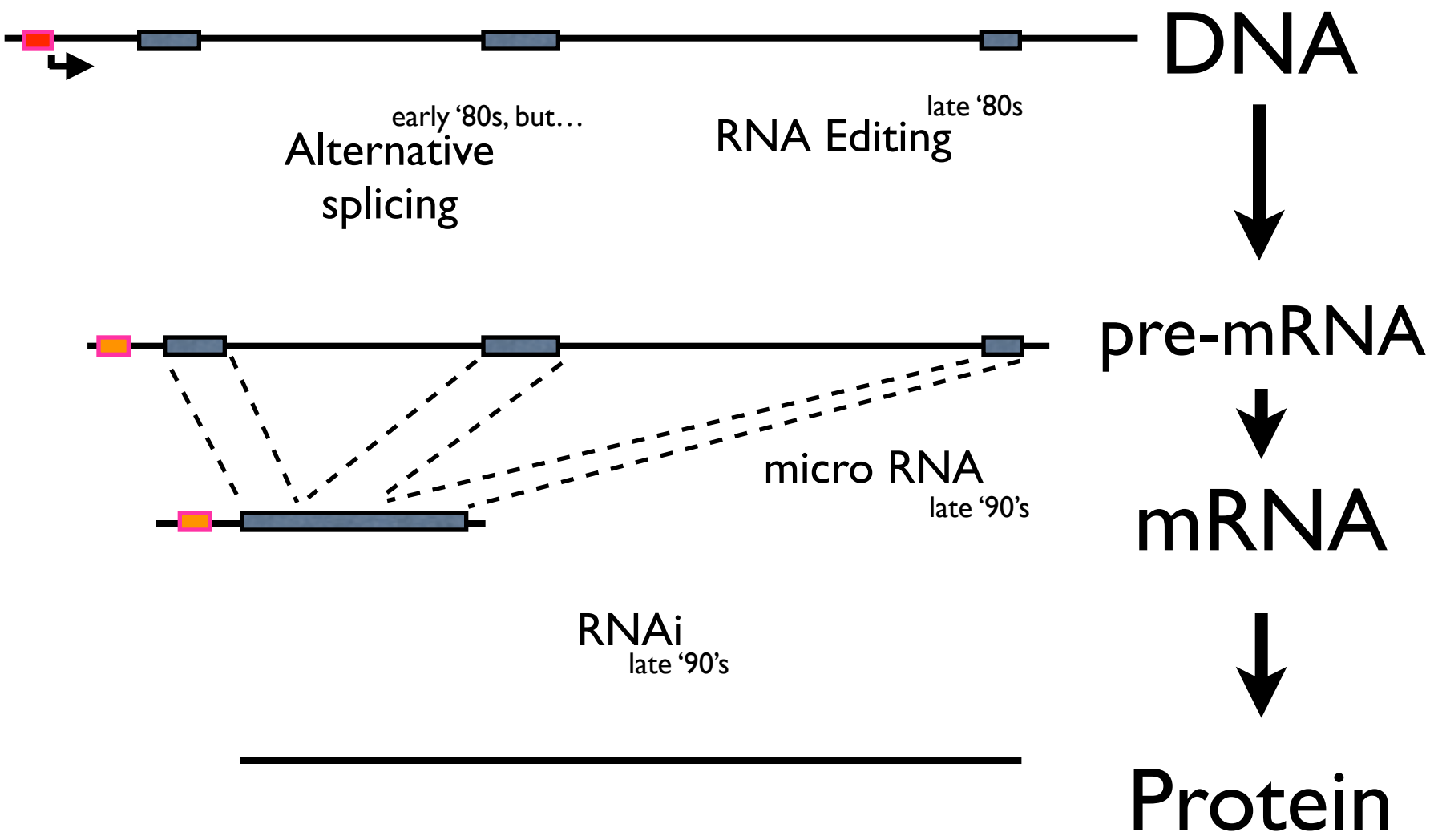
mRNA

RNAi

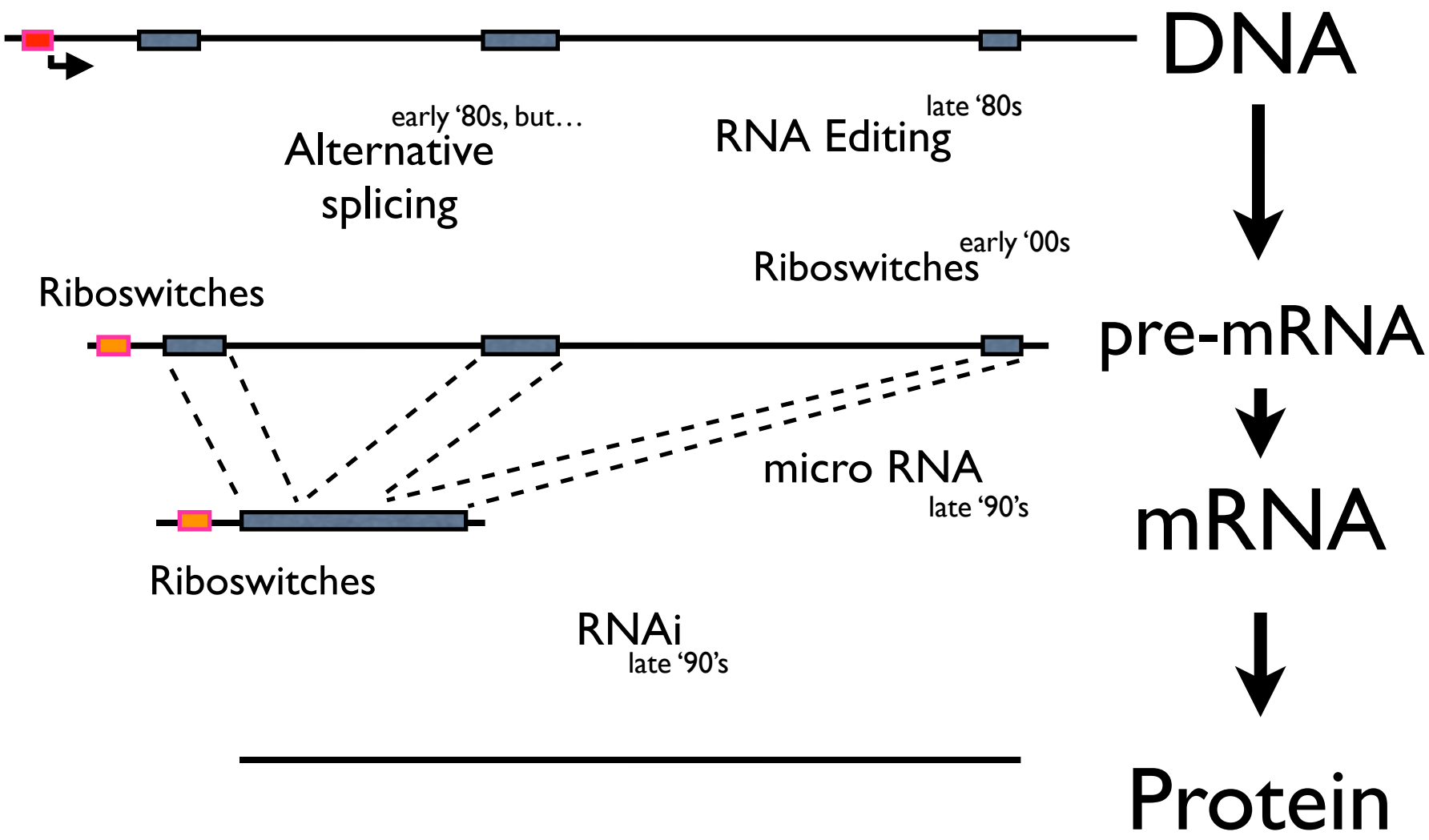
late '90's

Protein

# 21st Century Opportunities



# 21st Century Opportunities



# 21st Century Opportunities

DNA

early '80s, but…

Alternative splicing

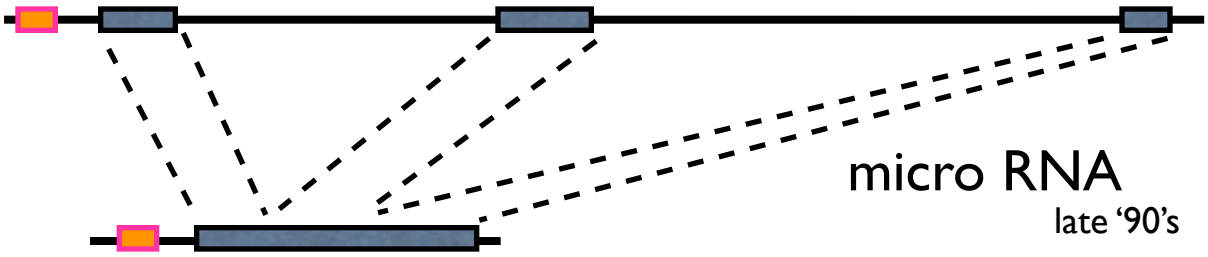
late '80s

RNA Editing

early '00s

Riboswitches

Riboswitches



pre-mRNA

micro RNA

late '90's

mRNA

Riboswitches

RNAi

late '90's

2010

To be discovered…

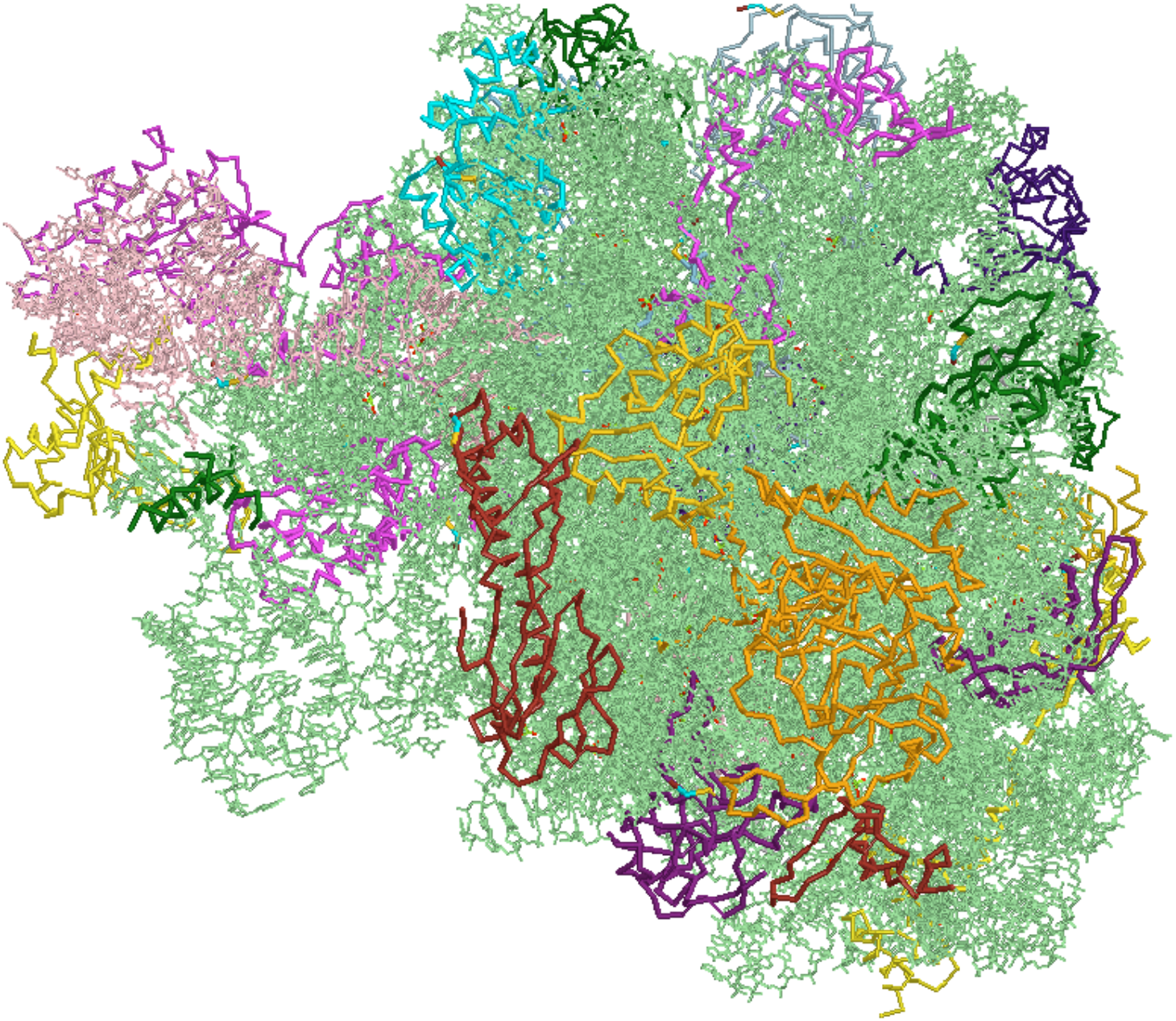
"so 20th century"

Protein

# 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?)

**ACCGCCACCGAAG**

**TGGCGGTGGCTTC**

or

**ACCGCCACCGAAG**

**TGGCGGTGGCTTA**

# Which is more stable?

(which has a higher melting temperature?)

**ACCGCCACCGAAG**
**TGGCGGTGGCTTC**

51.6° C

or

**ACCGCCACCGAAG**
**TGGCGGTGGCTTA**

# Which is more stable?

(which has a higher melting temperature?)

**ACCGCCACCGAAG**
**TGGCGGTGGCTTC**

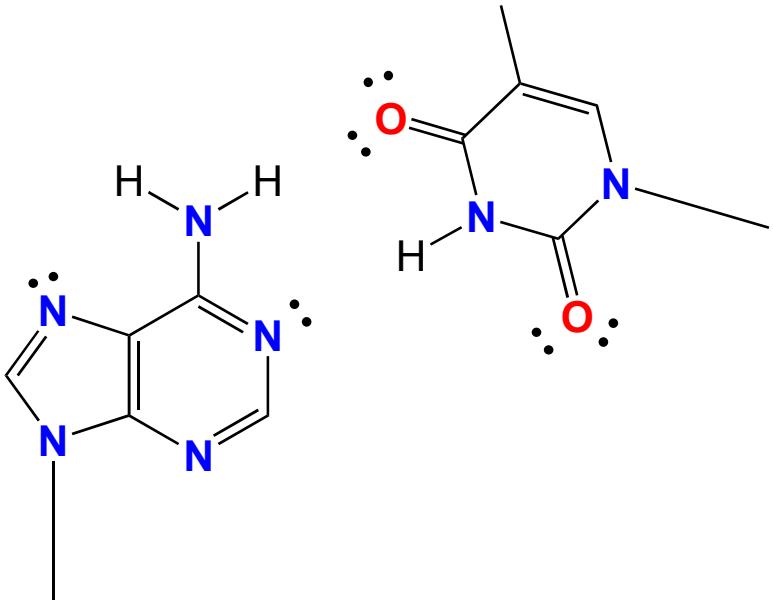
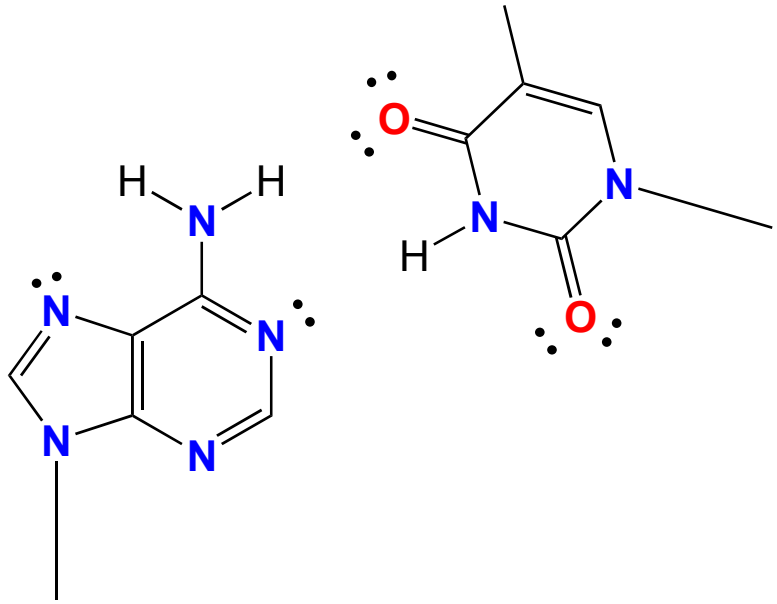
51.6° C

or

**ACCGCCACCGAAG**
**TGGCGGTGGCTTA**

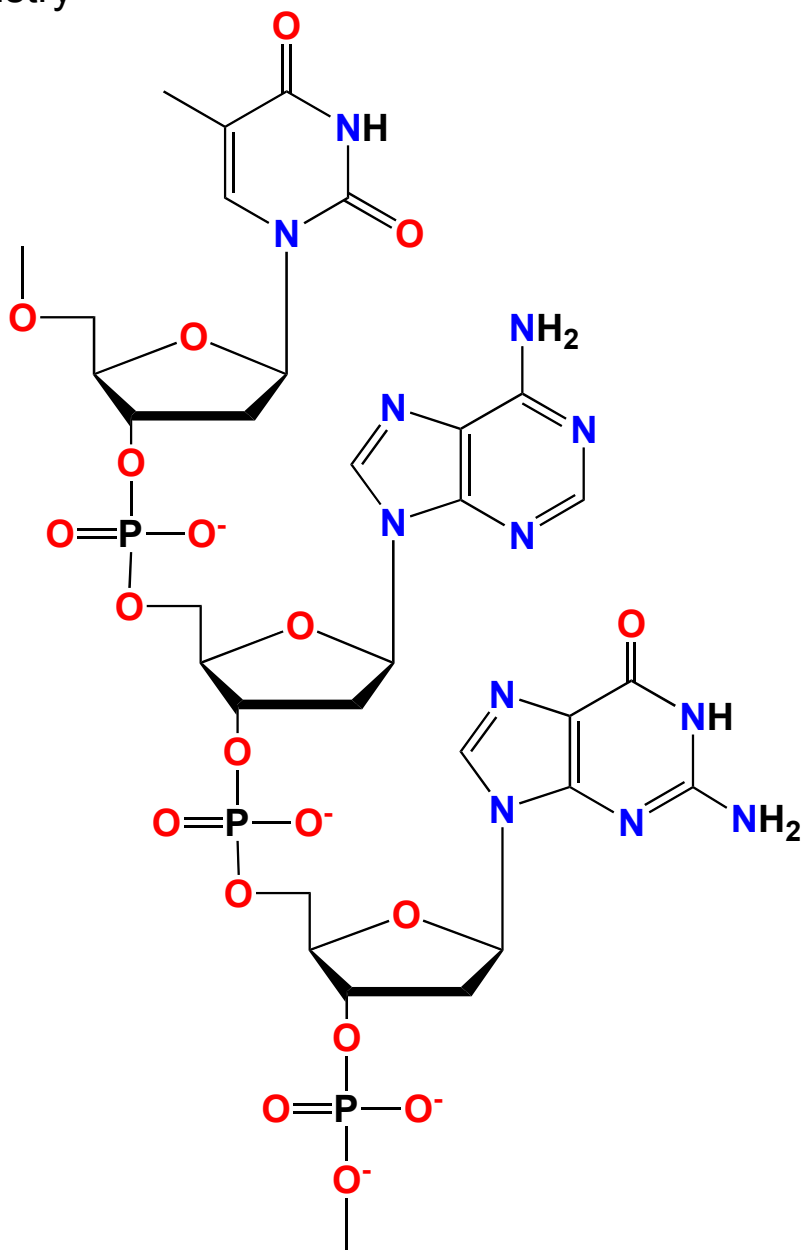
52.5° C

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


H
H
N
N
N
N
N
O
N
H
O

H
H
N
N
N
N
N
O
N
H
O

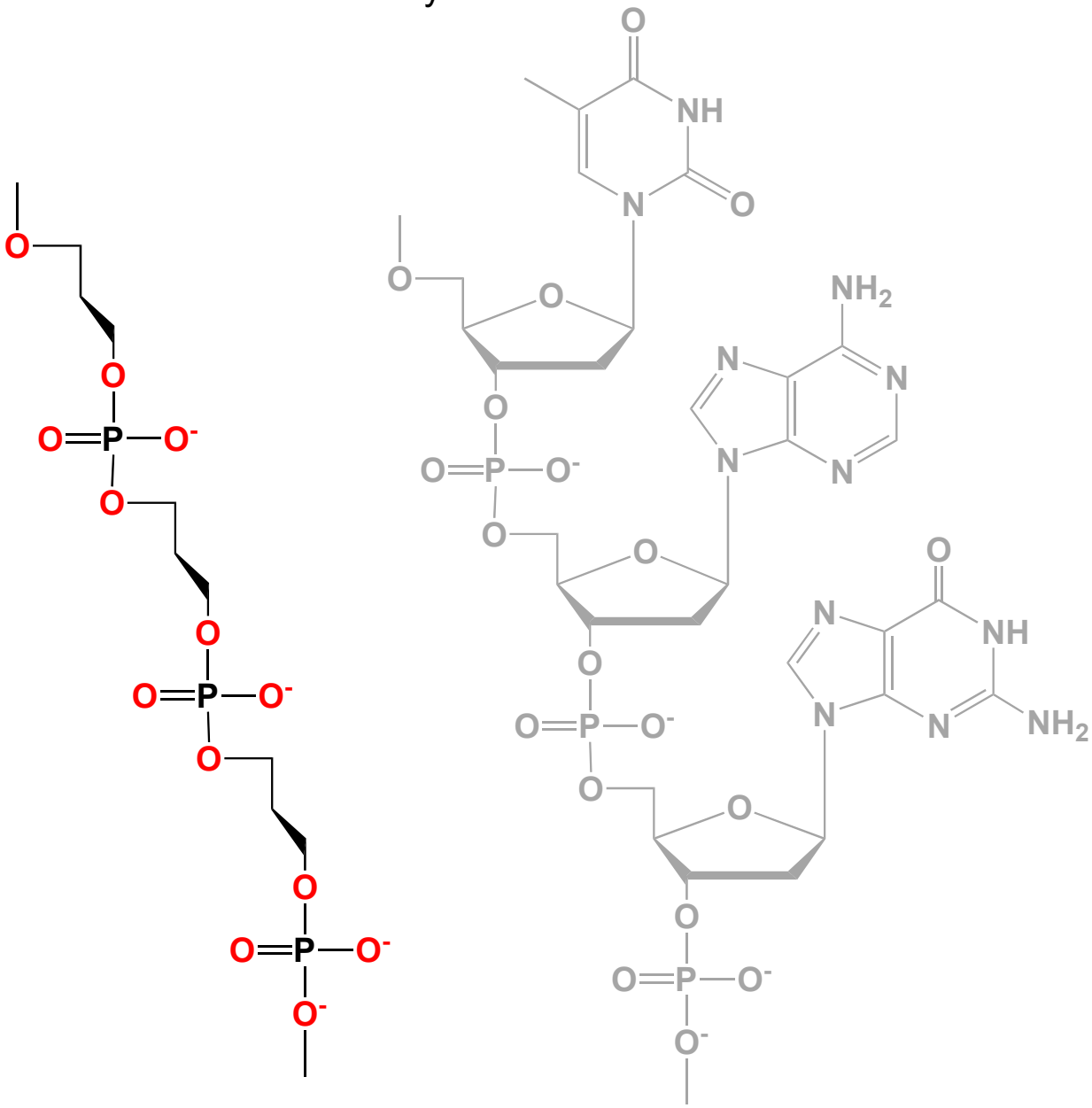
# DNA

A look at the Chemistry



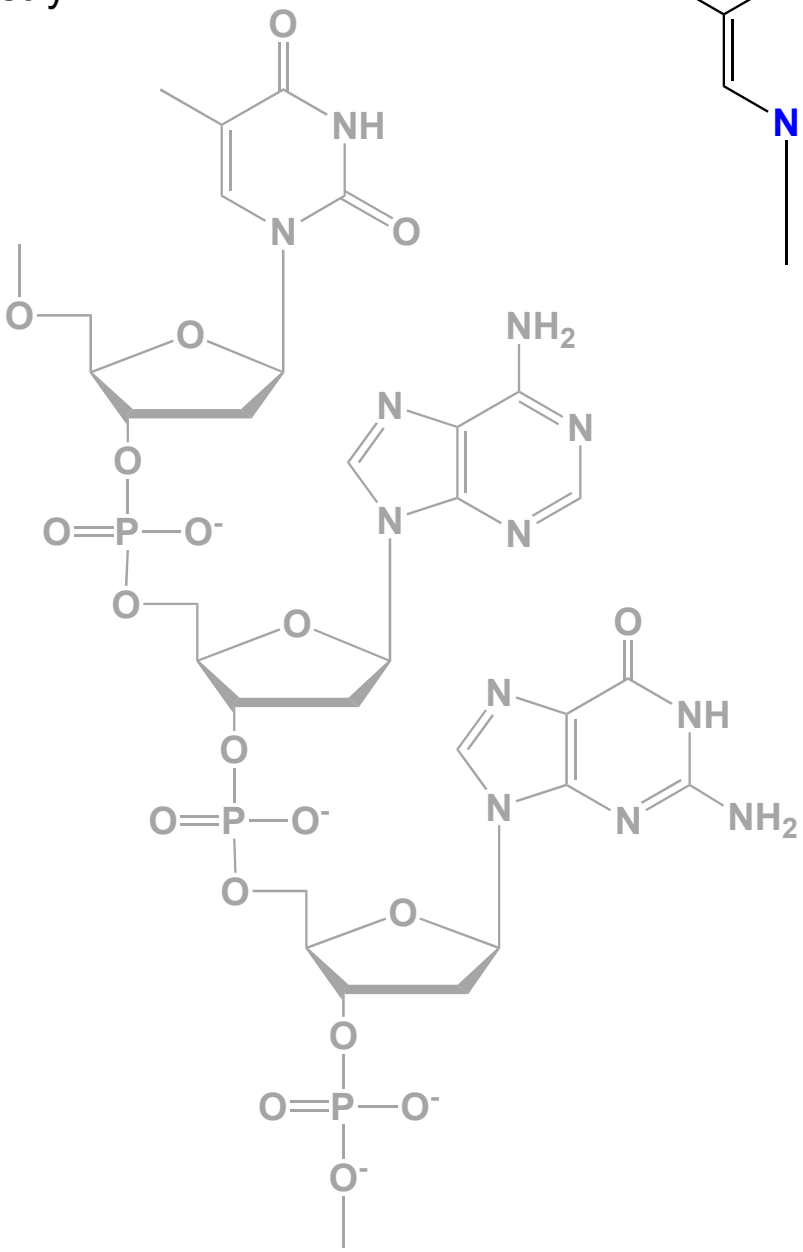
# DNA

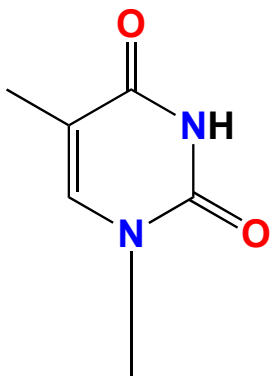
A look at the Chemistry

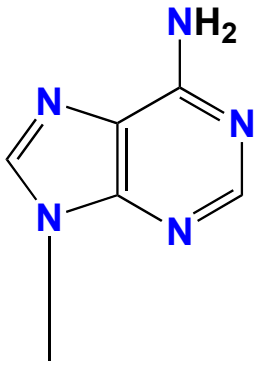


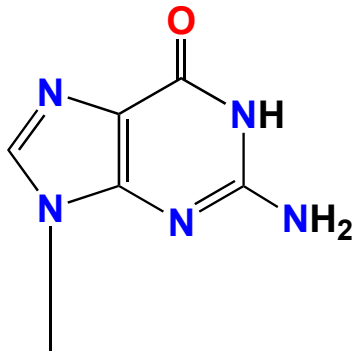
# DNA

A look at the Chemistry



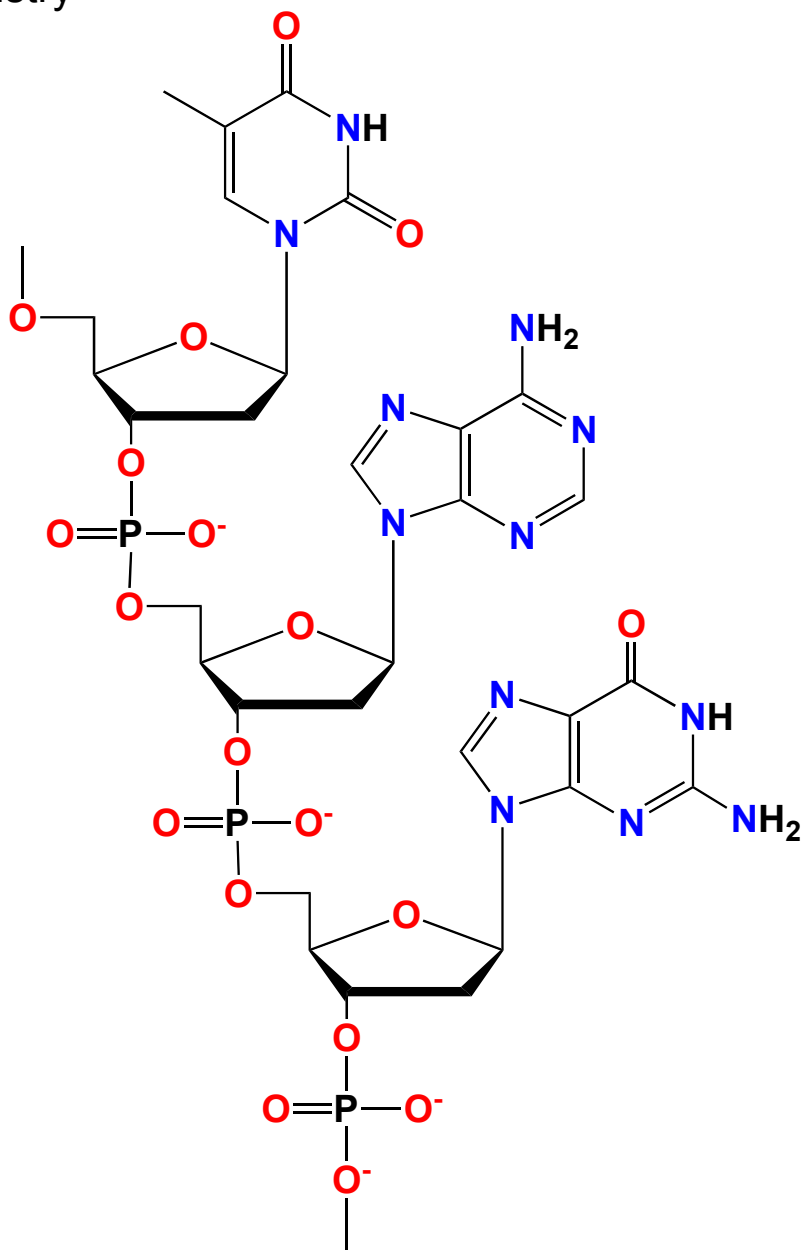






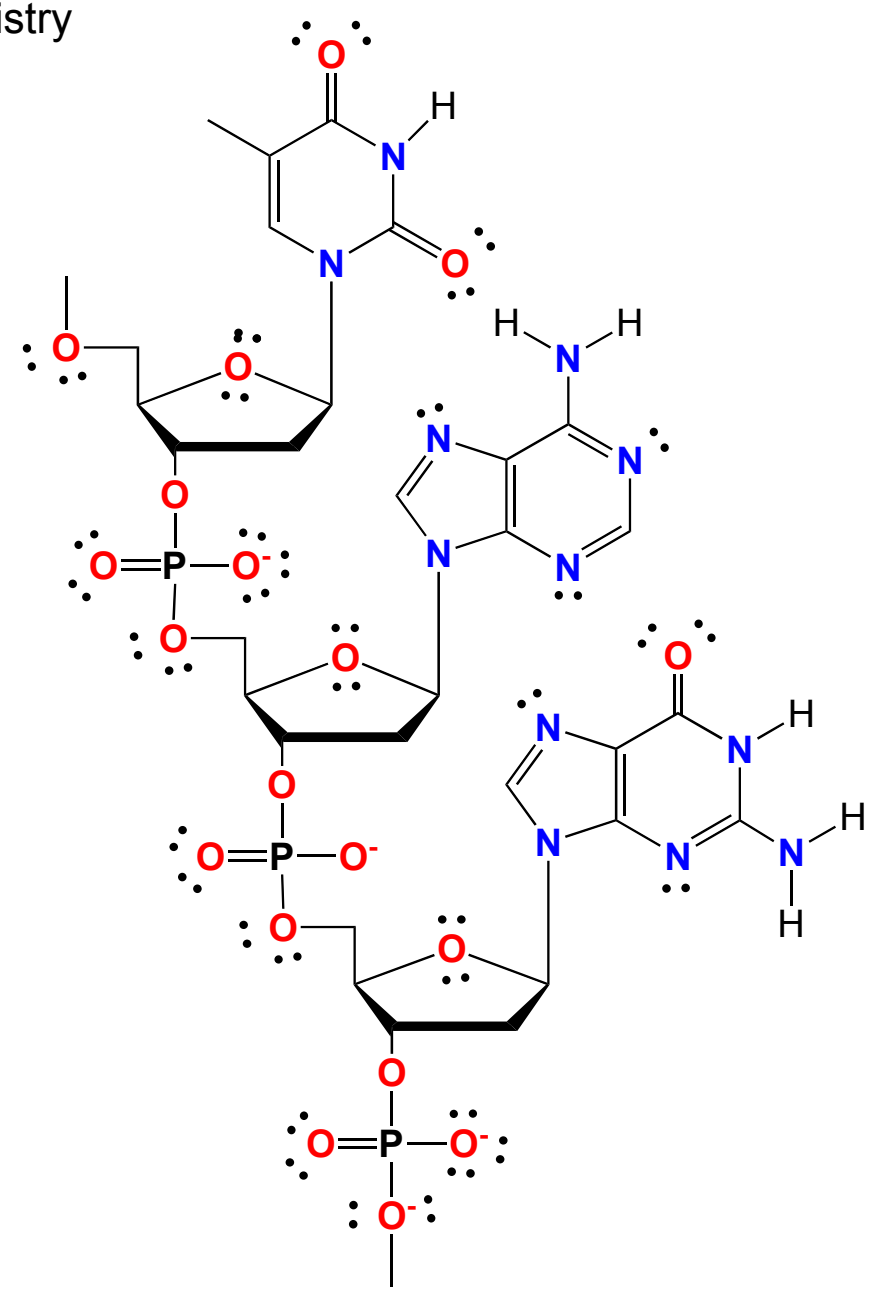
# DNA

A look at the Chemistry



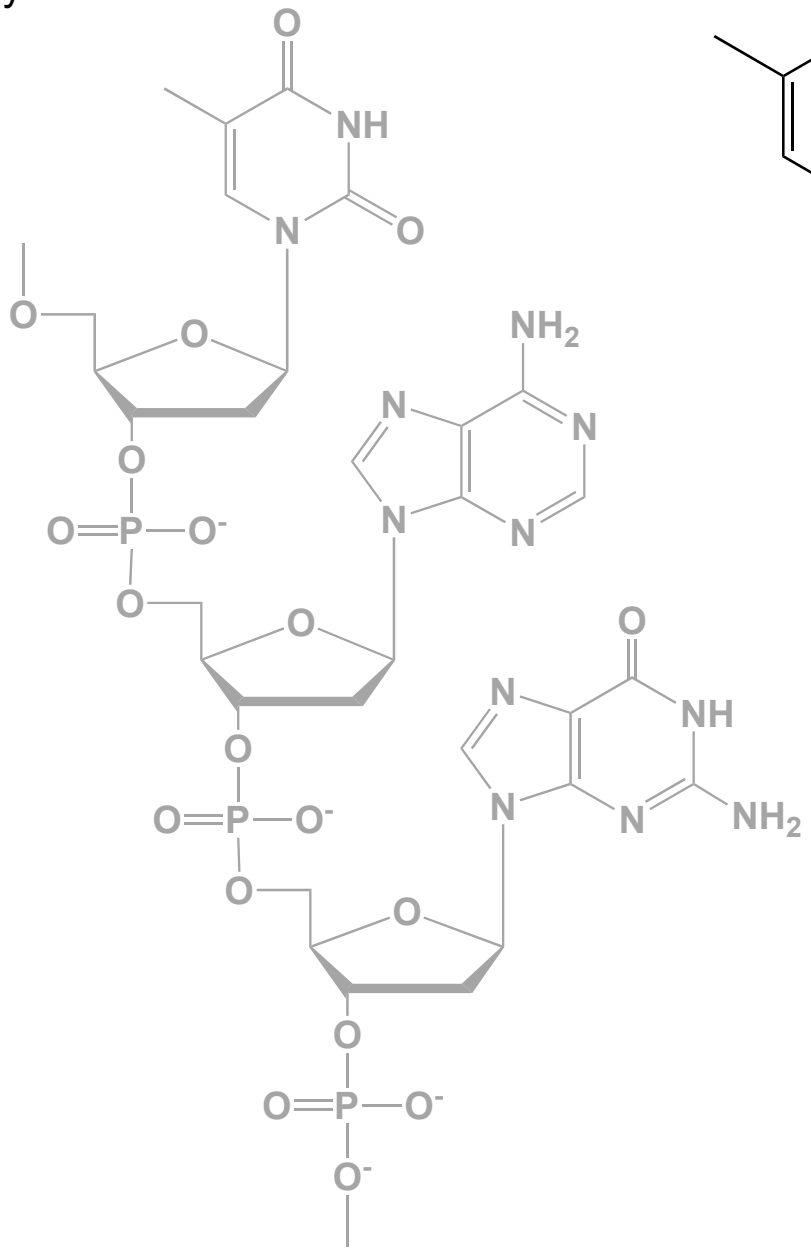
# DNA

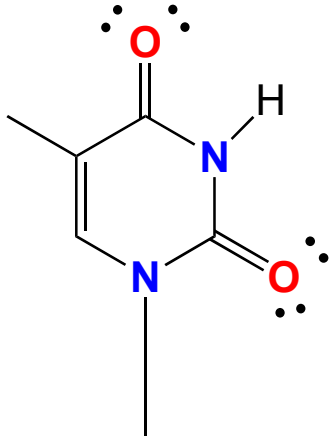
A look at the Chemistry



# DNA

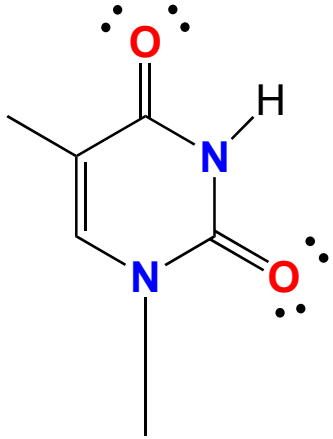
A look at the Chemistry

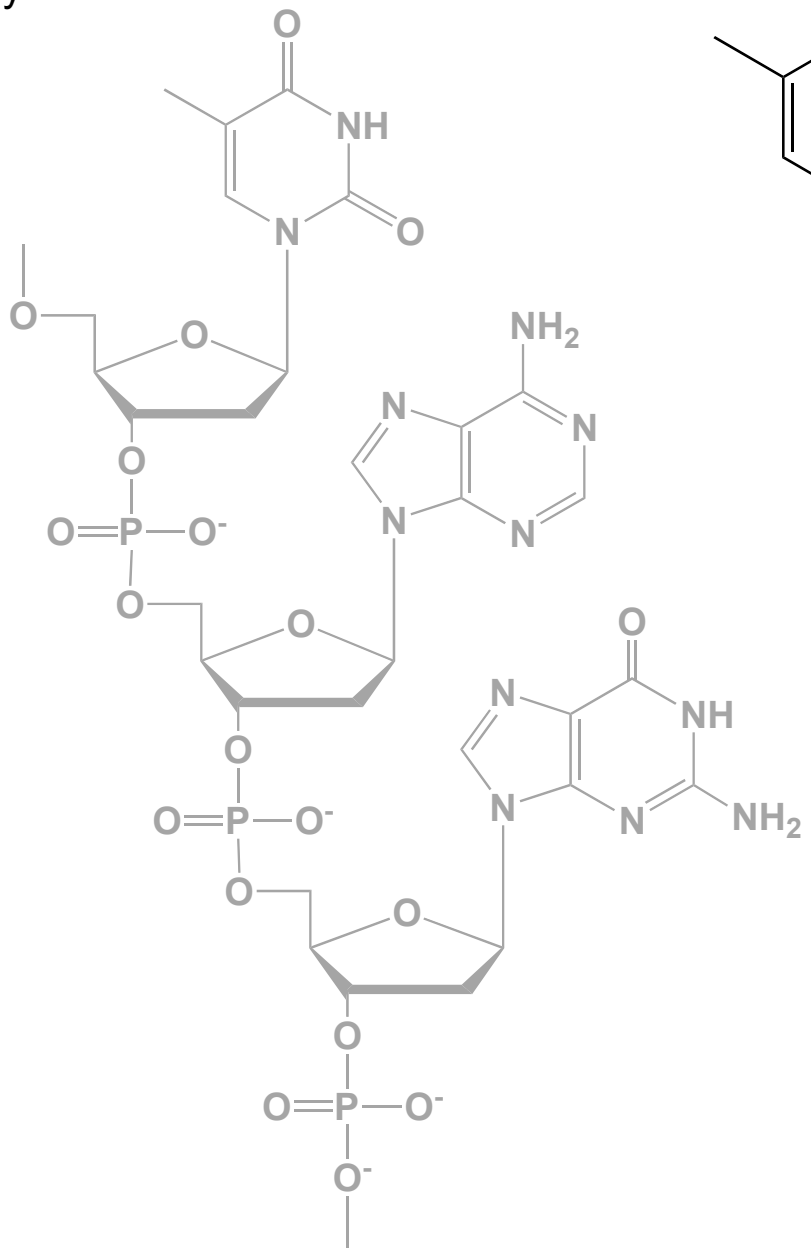


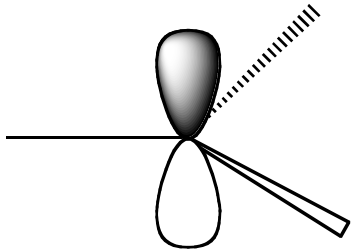



# DNA

A look at the Chemistry

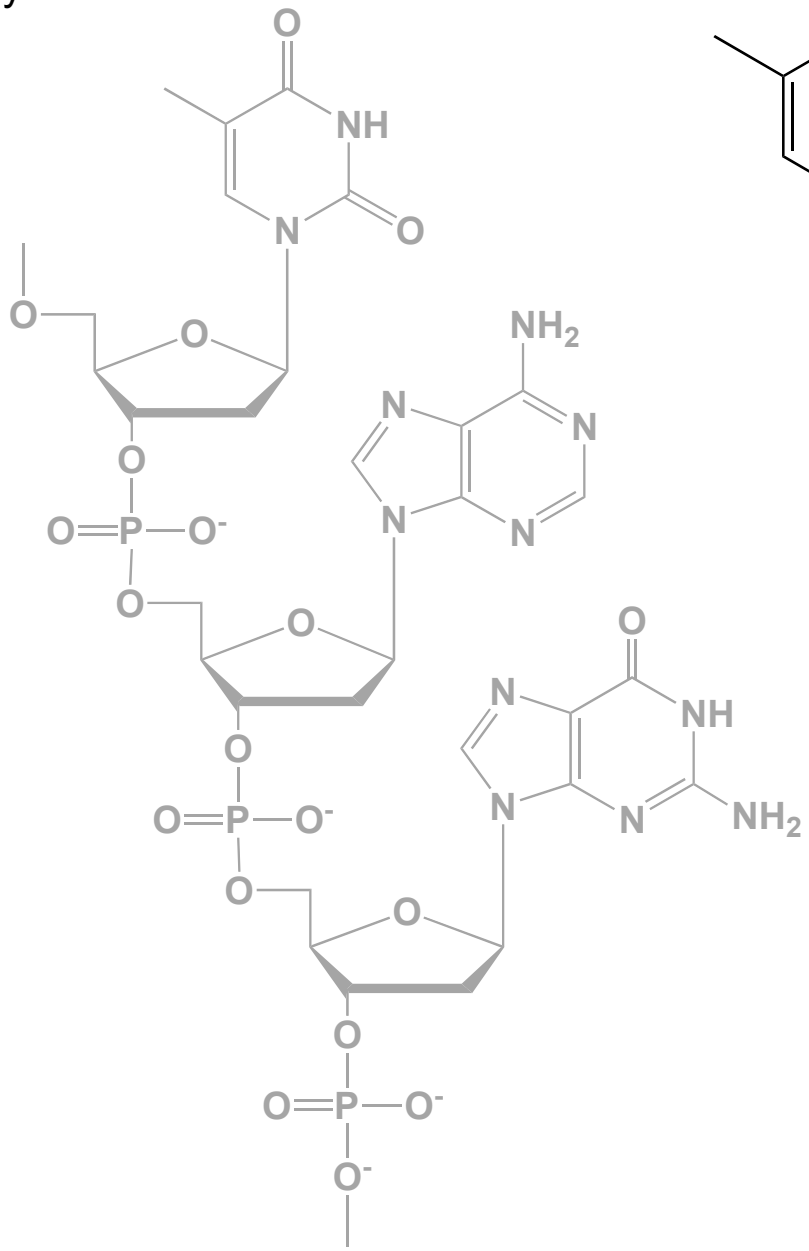


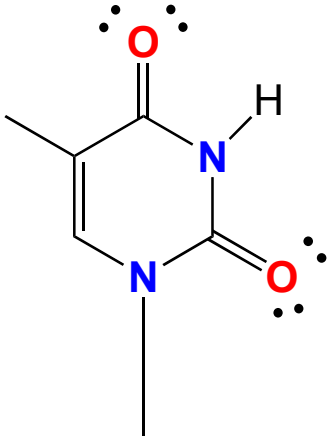


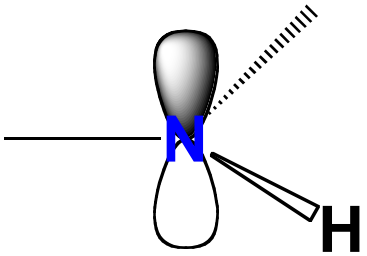


# DNA

A look at the Chemistry

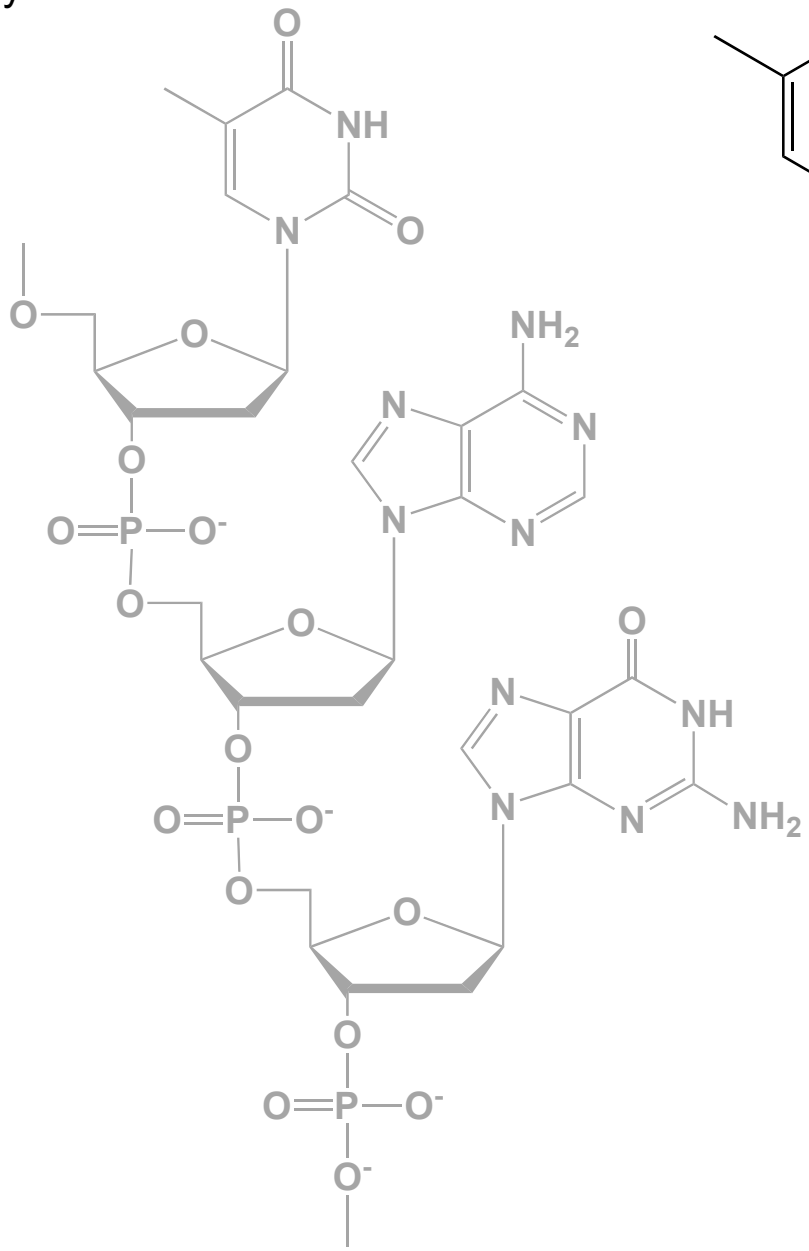


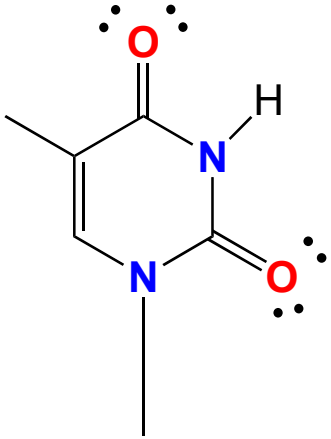



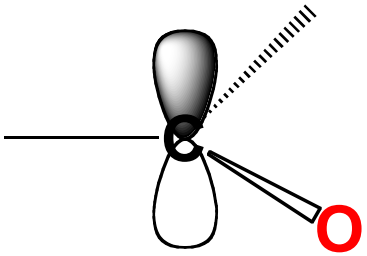



# DNA

A look at the Chemistry

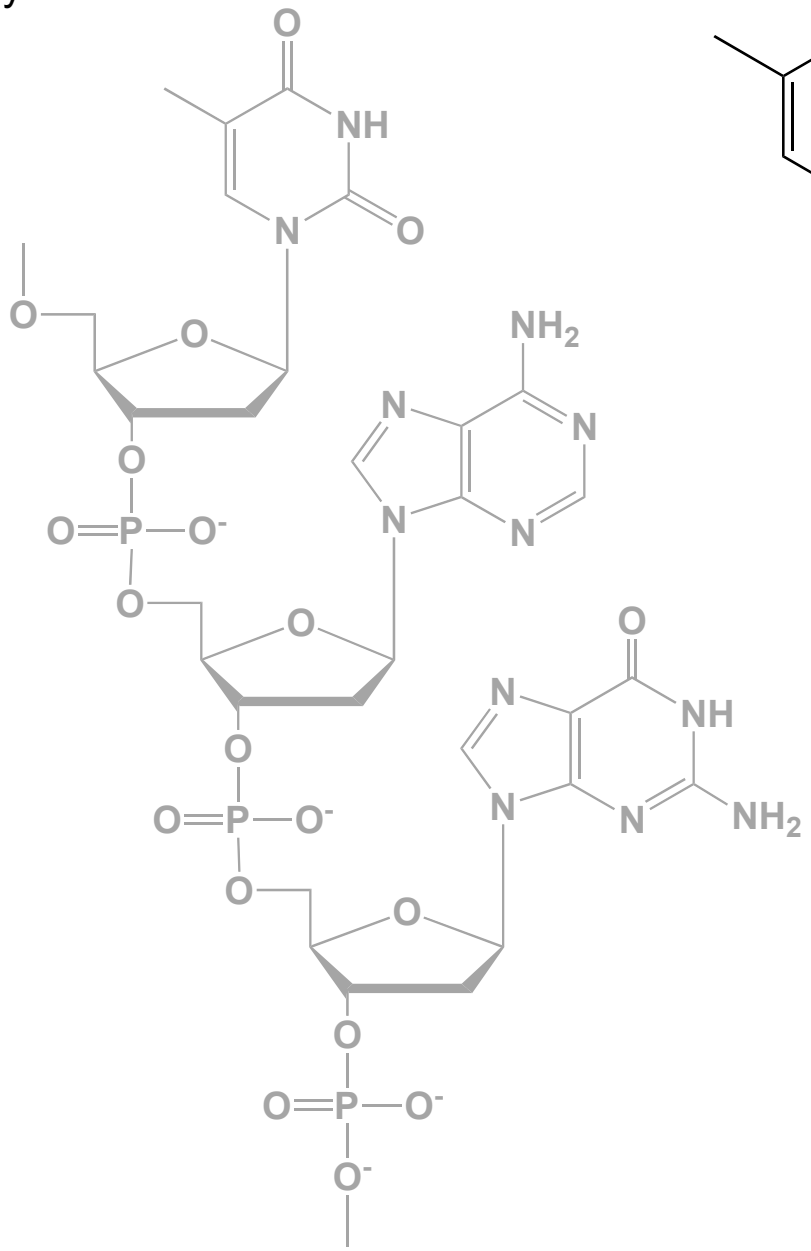


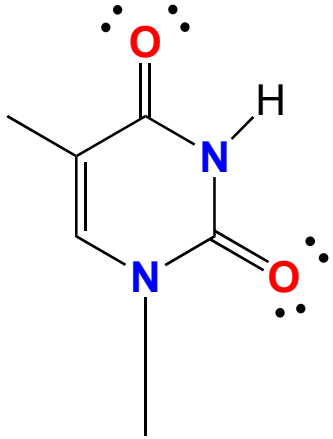



# DNA

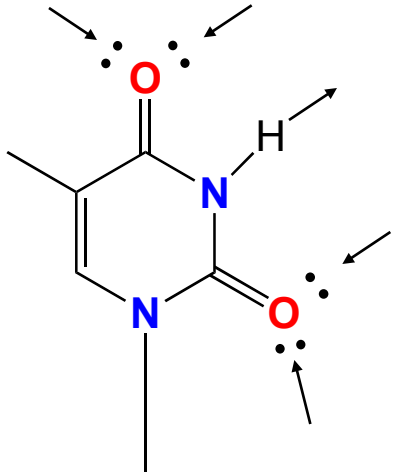
A look at the Chemistry

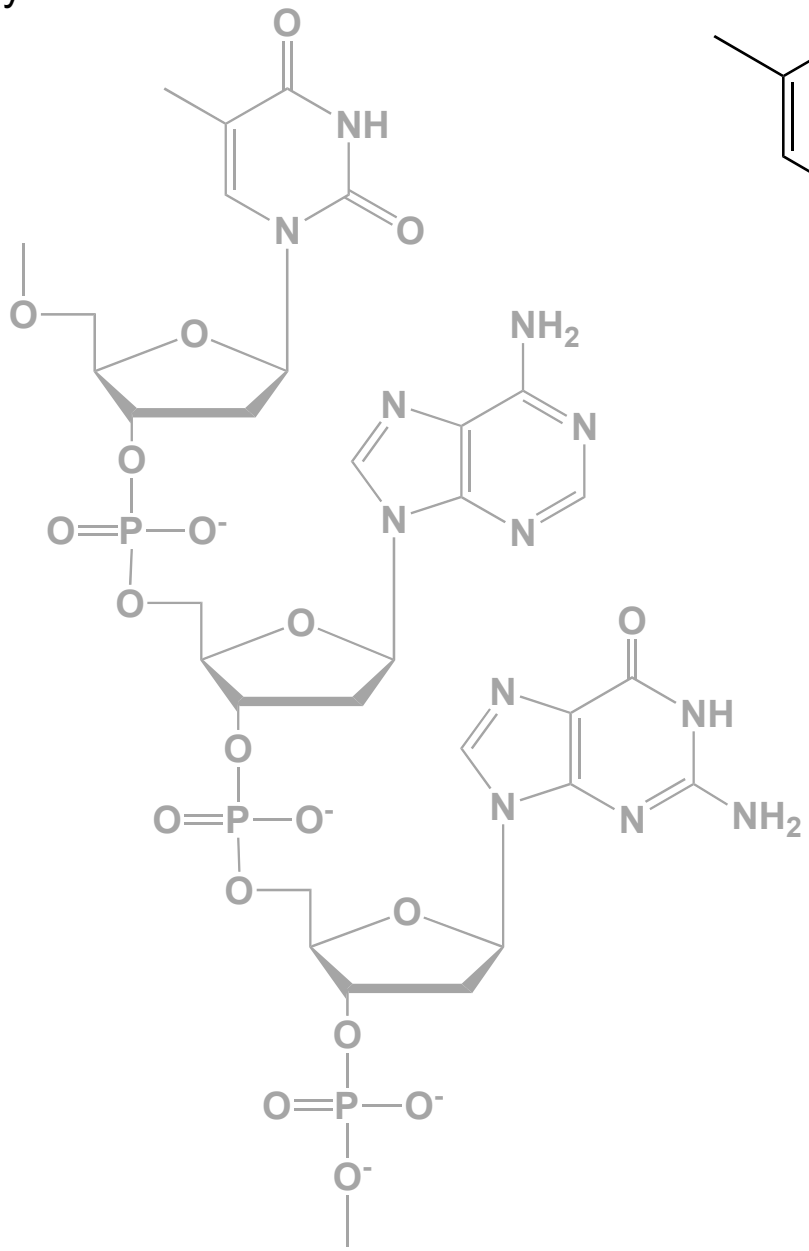


# DNA

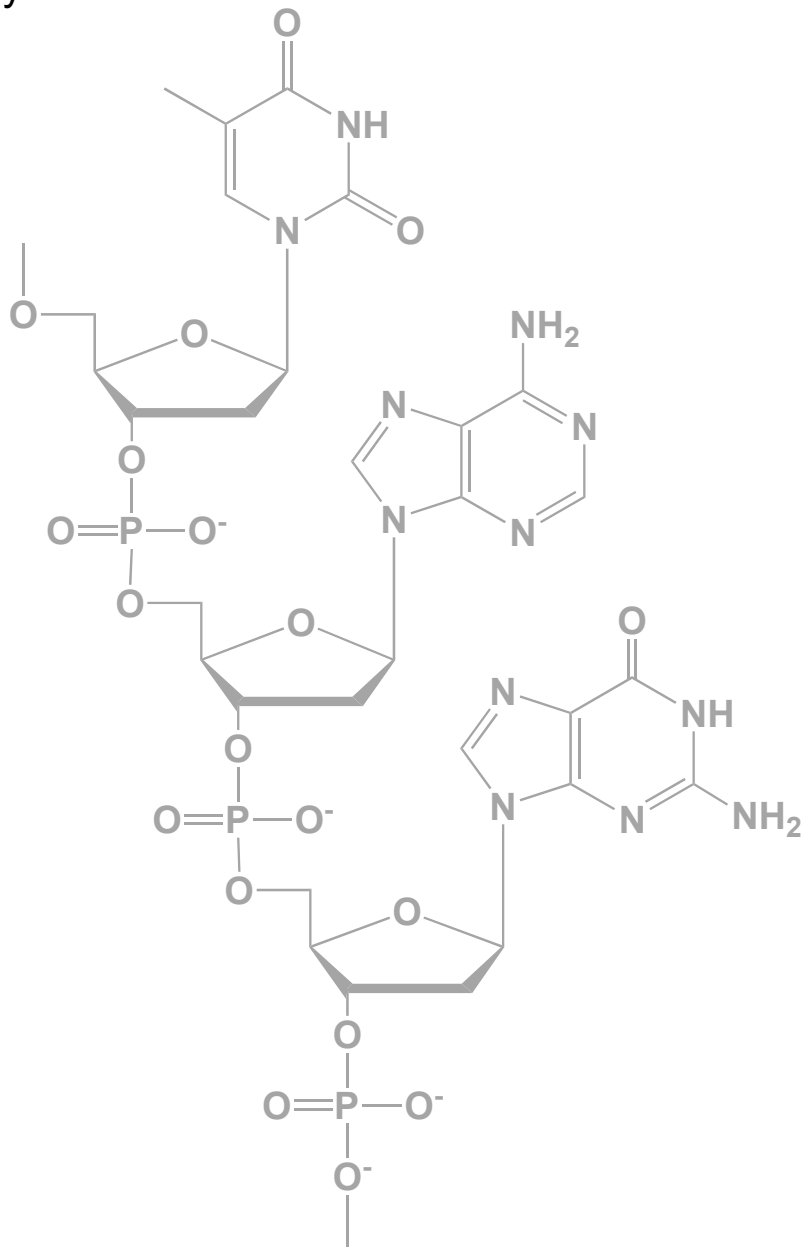
A look at the Chemistry

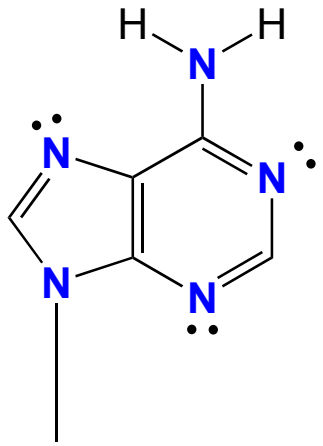


# DNA

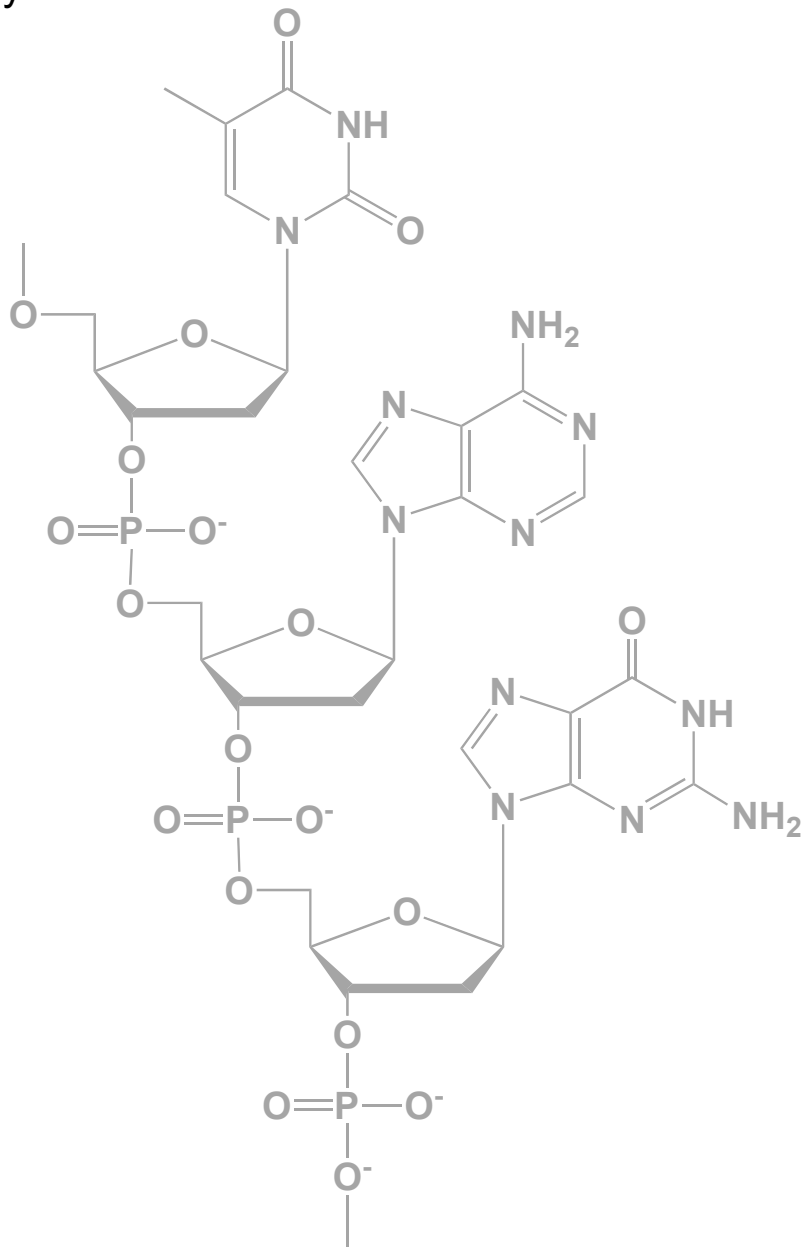
A look at the Chemistry

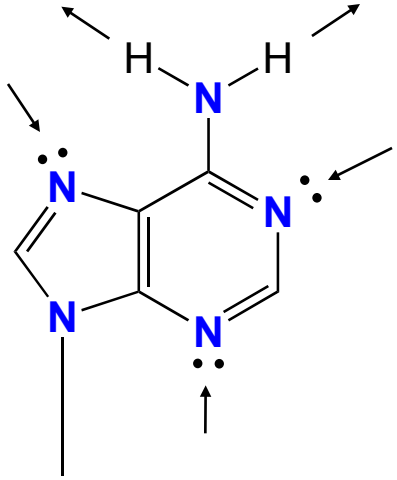




# DNA

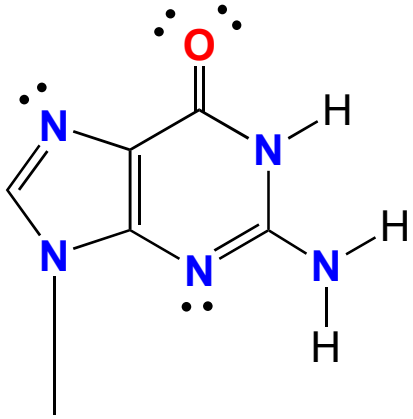
A look at the Chemistry

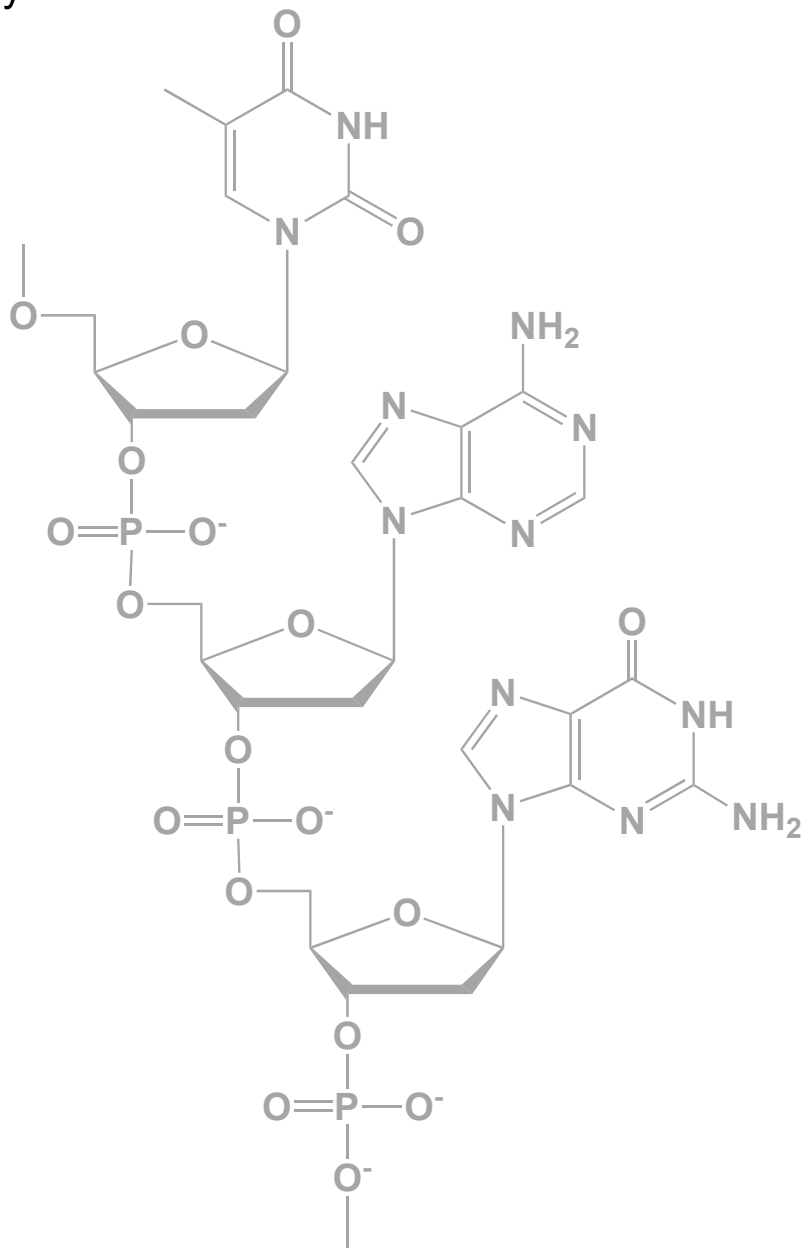




# DNA

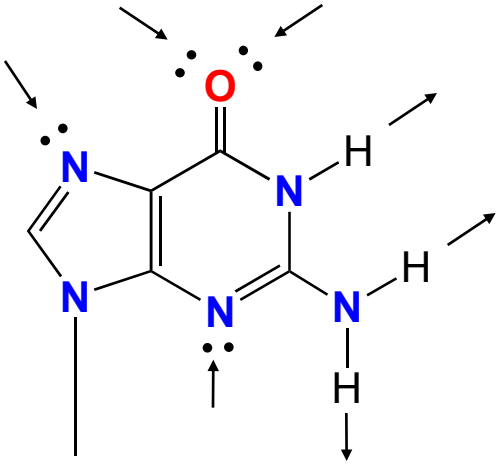
A look at the Chemistry

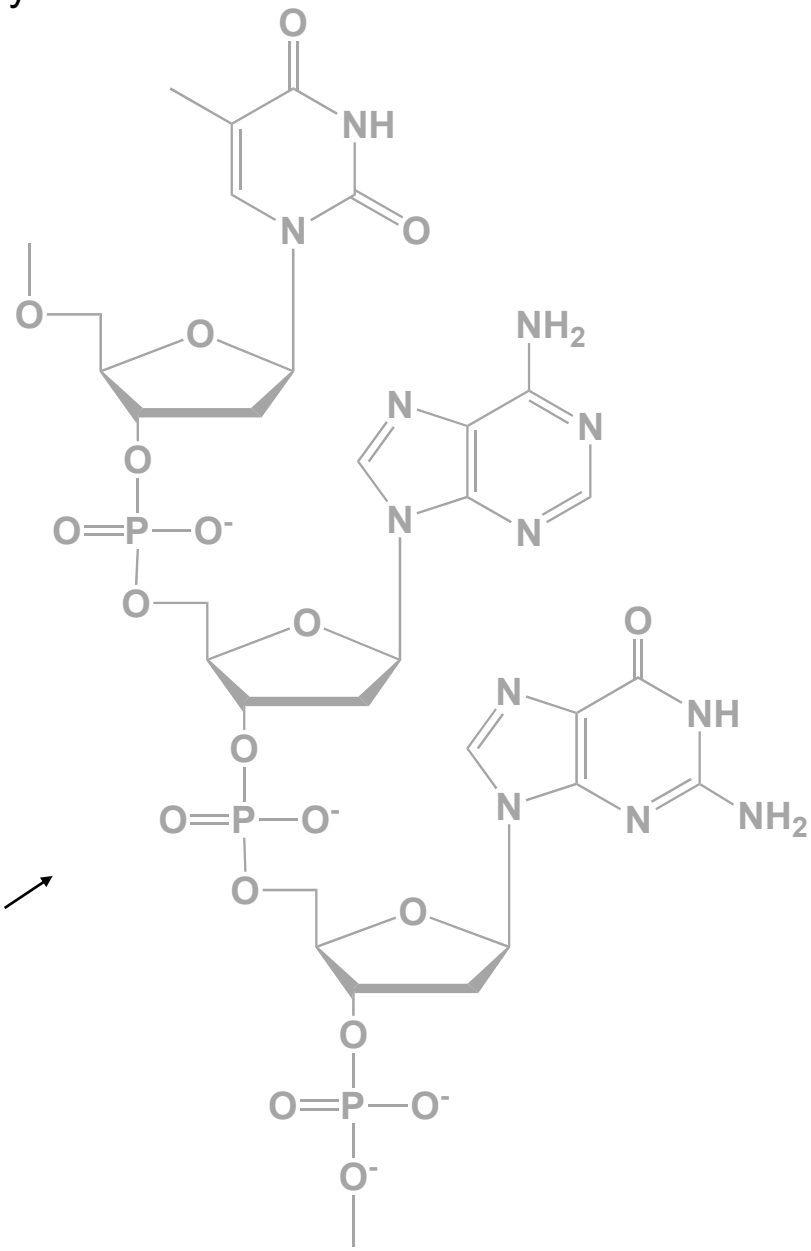




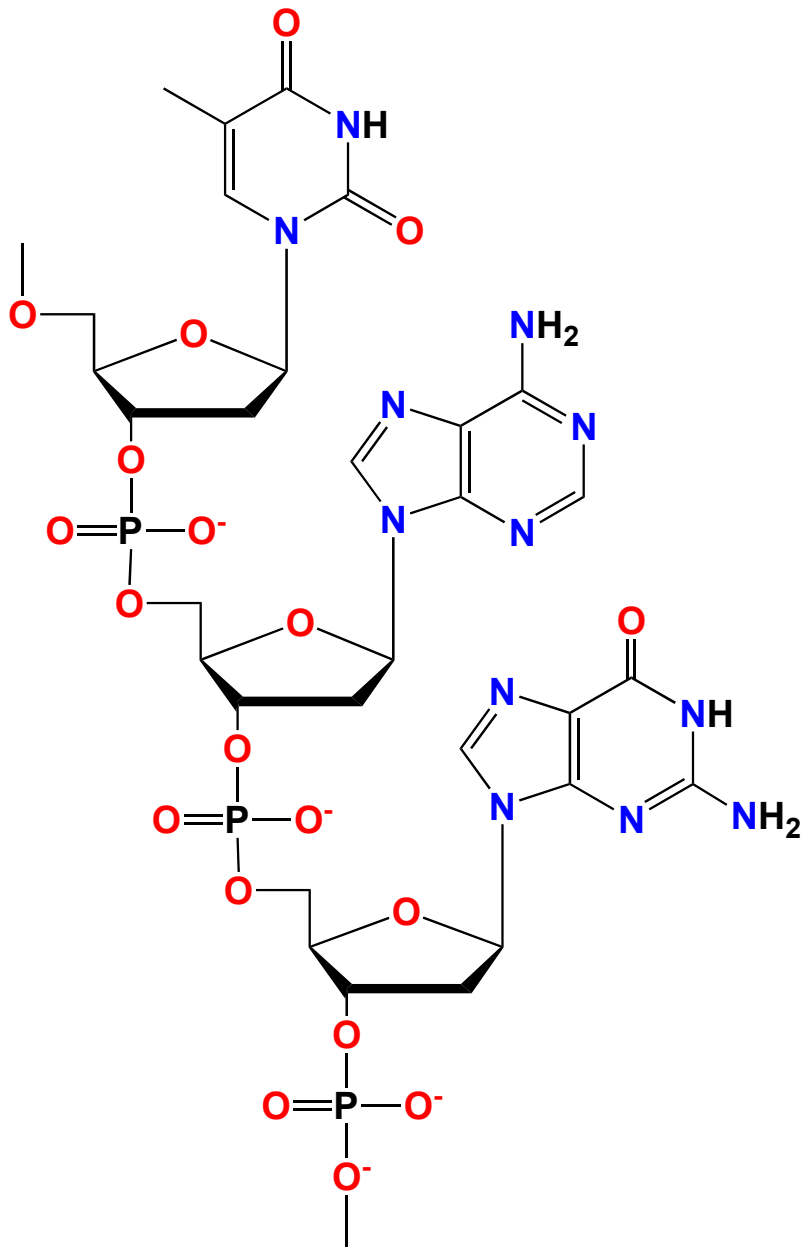
# DNA

A look at the Chemistry

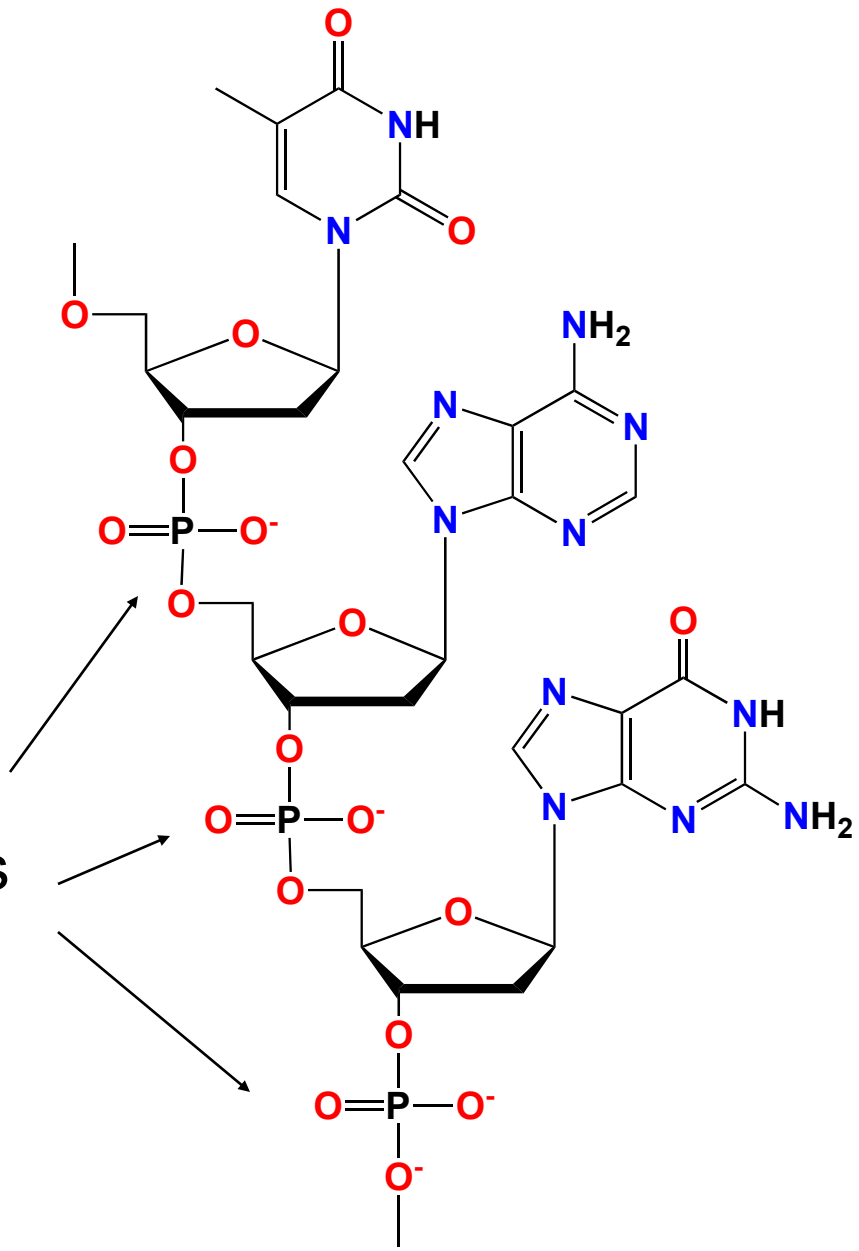




# What forces are important?

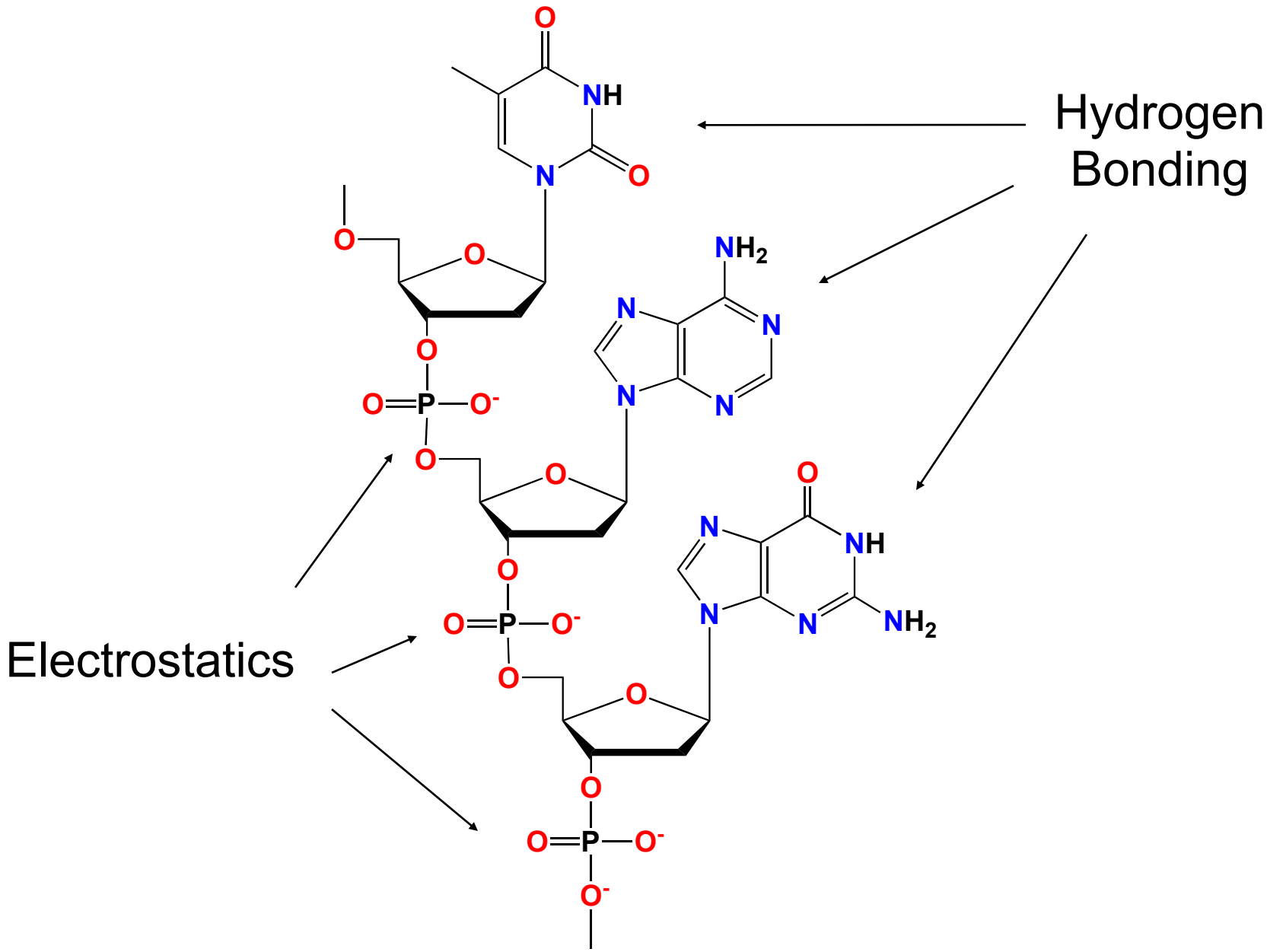


# What forces are important?

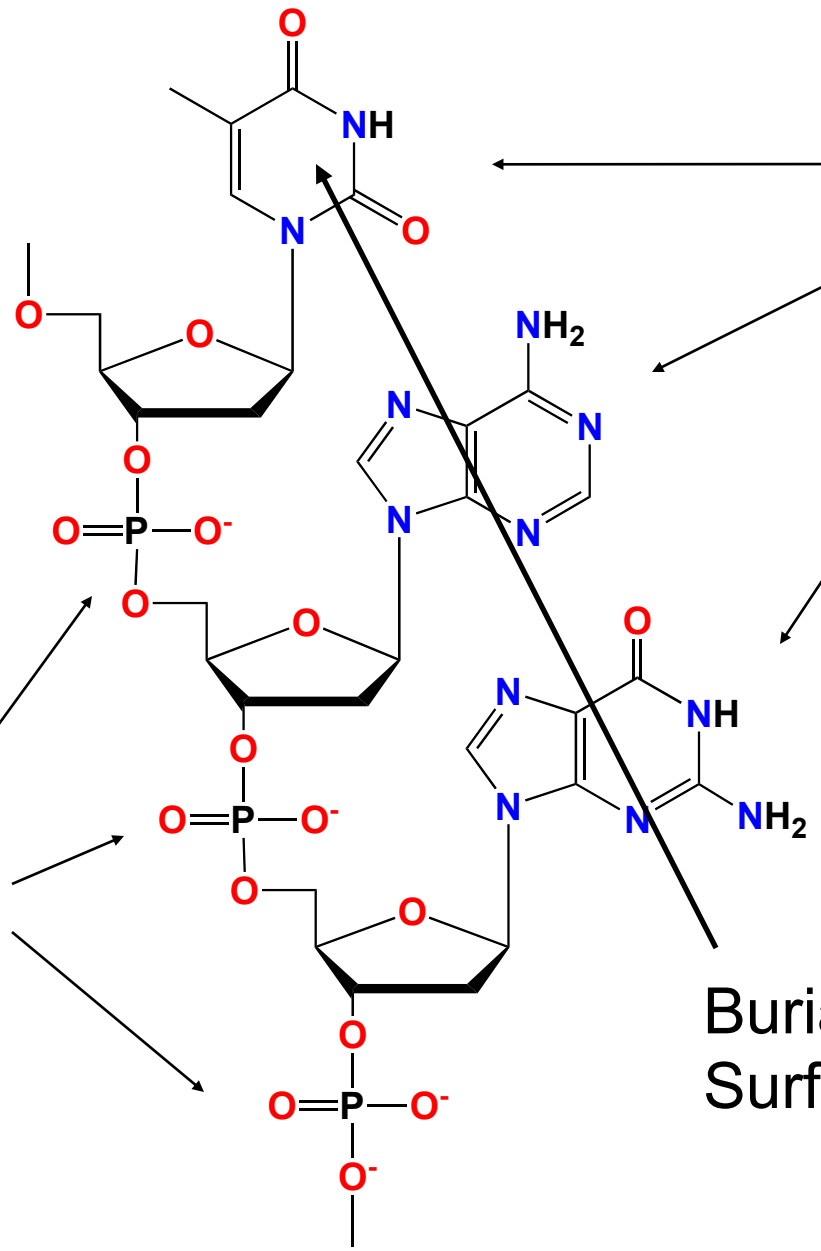



Electrostatics

# What forces are important?



# What forces are important?




Hydrogen Bonding

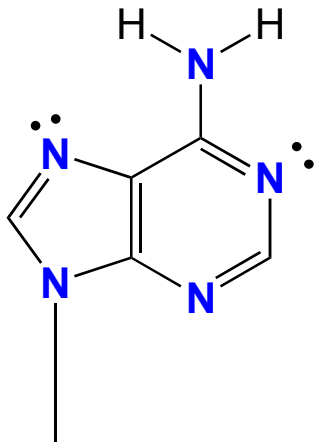
Electrostatics

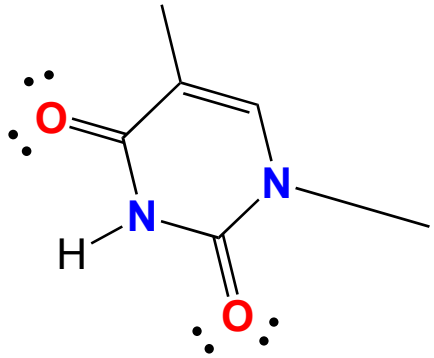
Burial of Hydrophobic Surface & Stacking

# Base Pairing

(Donors matched to Acceptors)

A



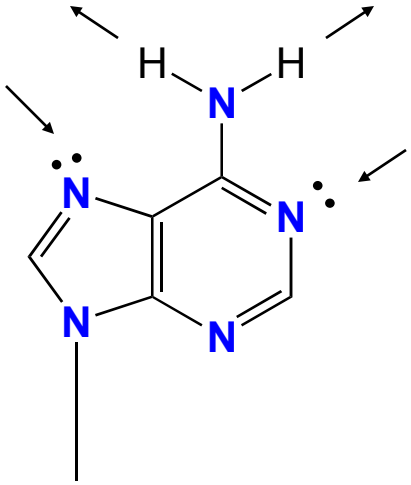



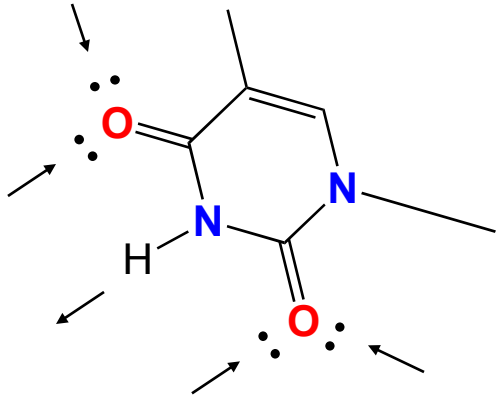

T

# Base Pairing

(Donors matched to Acceptors)

A

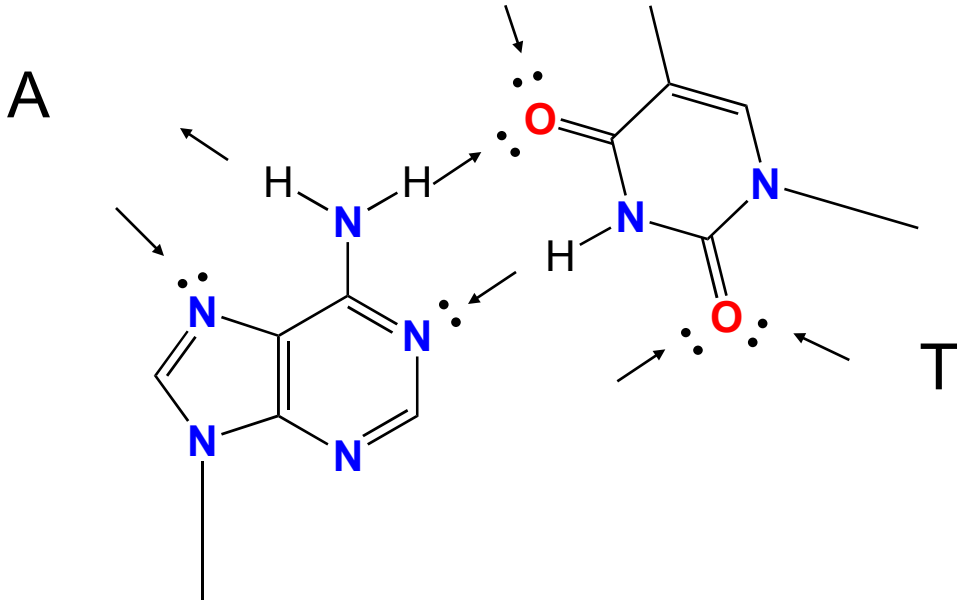



T

# Base Pairing

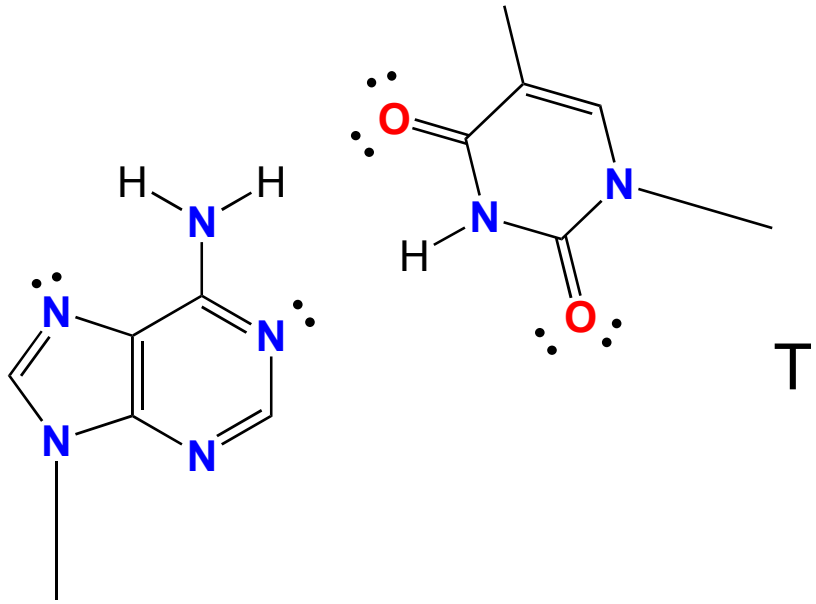
(Donors matched to Acceptors)



# Base Pairing

(Donors matched to Acceptors)

A



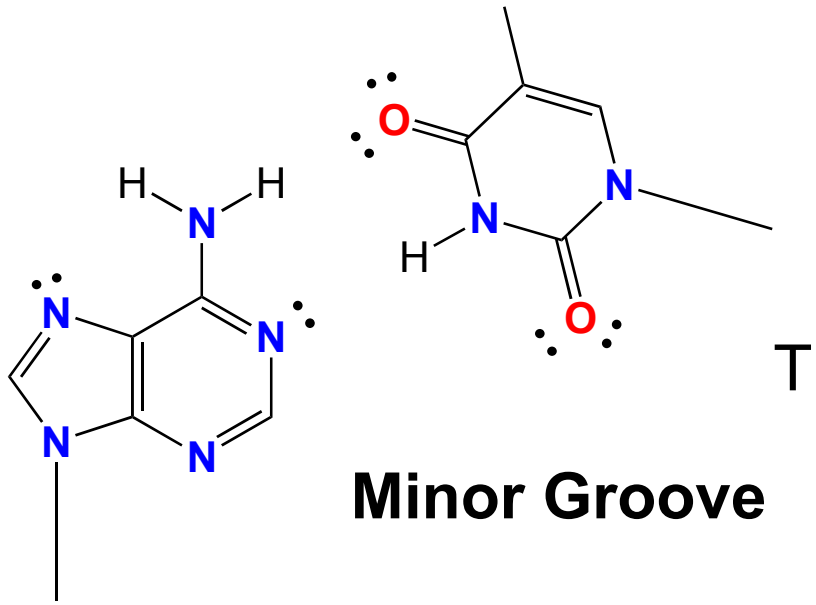

T

# Base Pairing

(Donors matched to Acceptors)

**Major Groove**

A

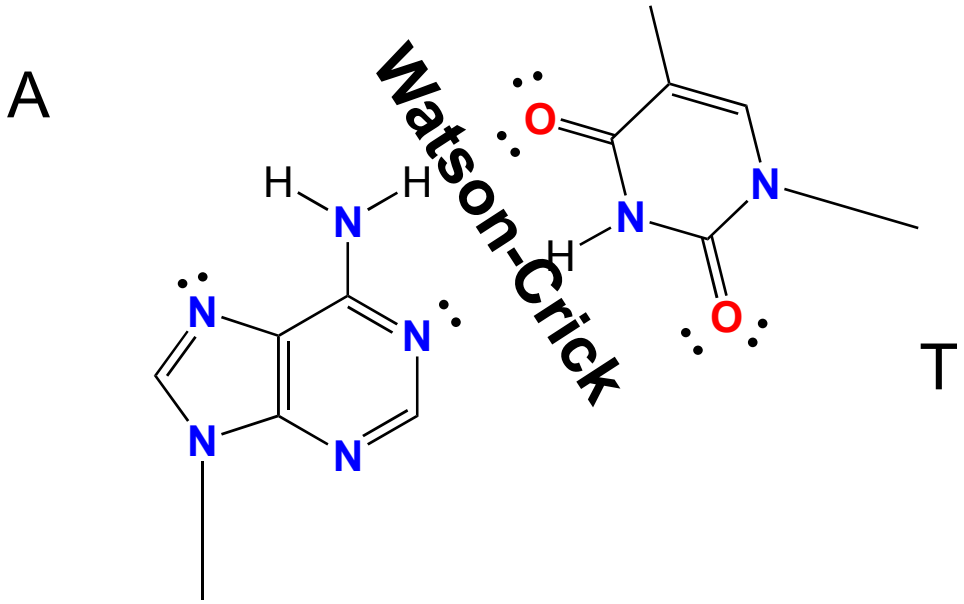



T

**Minor Groove**

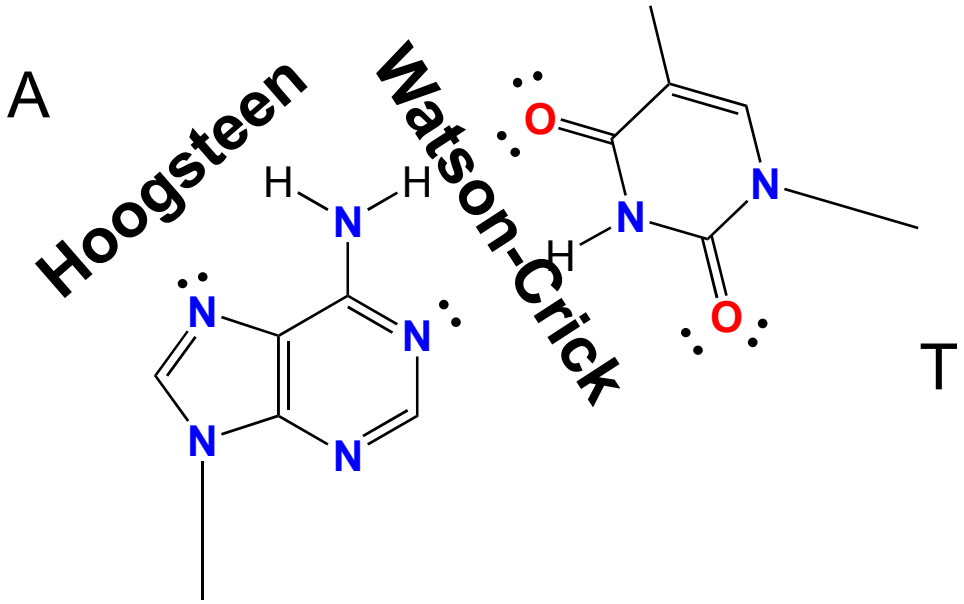
# Base Pairing

(Donors matched to Acceptors)



# Base Pairing

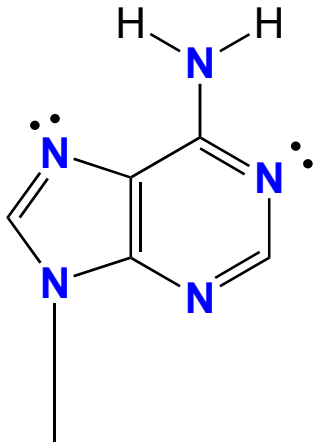
(Donors matched to Acceptors)

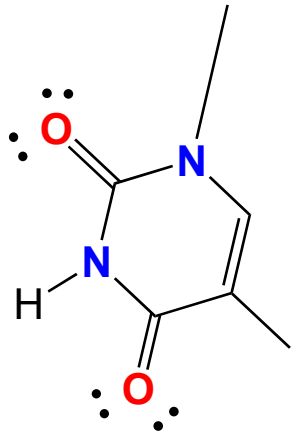


# Base Pairing

(Donors matched to Acceptors)

A



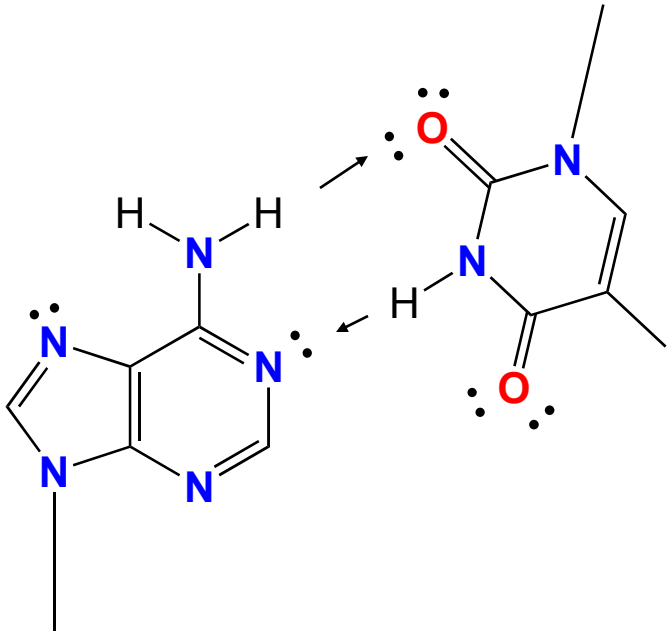




T

# Base Pairing

(Donors matched to Acceptors)

A

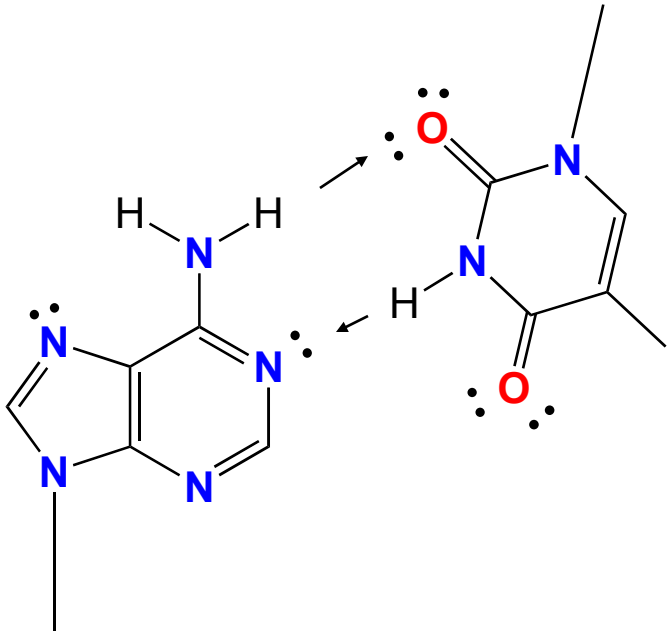



T

# Base Pairing

(Donors matched to Acceptors)

A



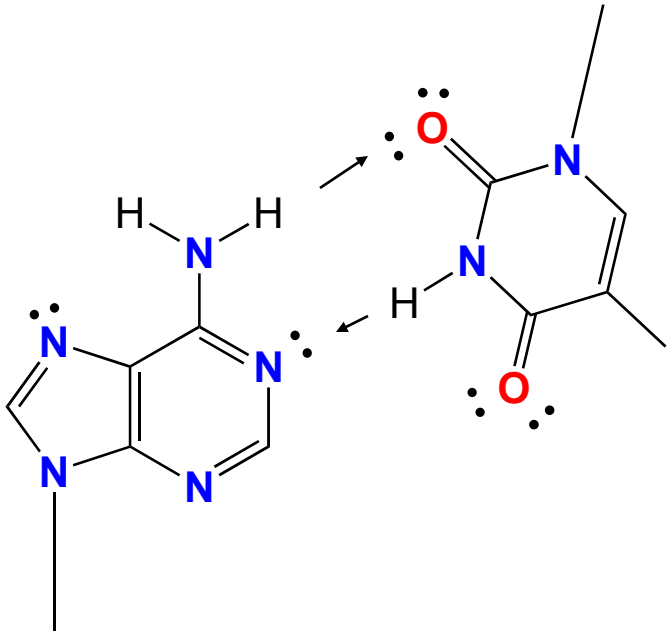

T

Good base pairing

# Base Pairing

(Donors matched to Acceptors)

A

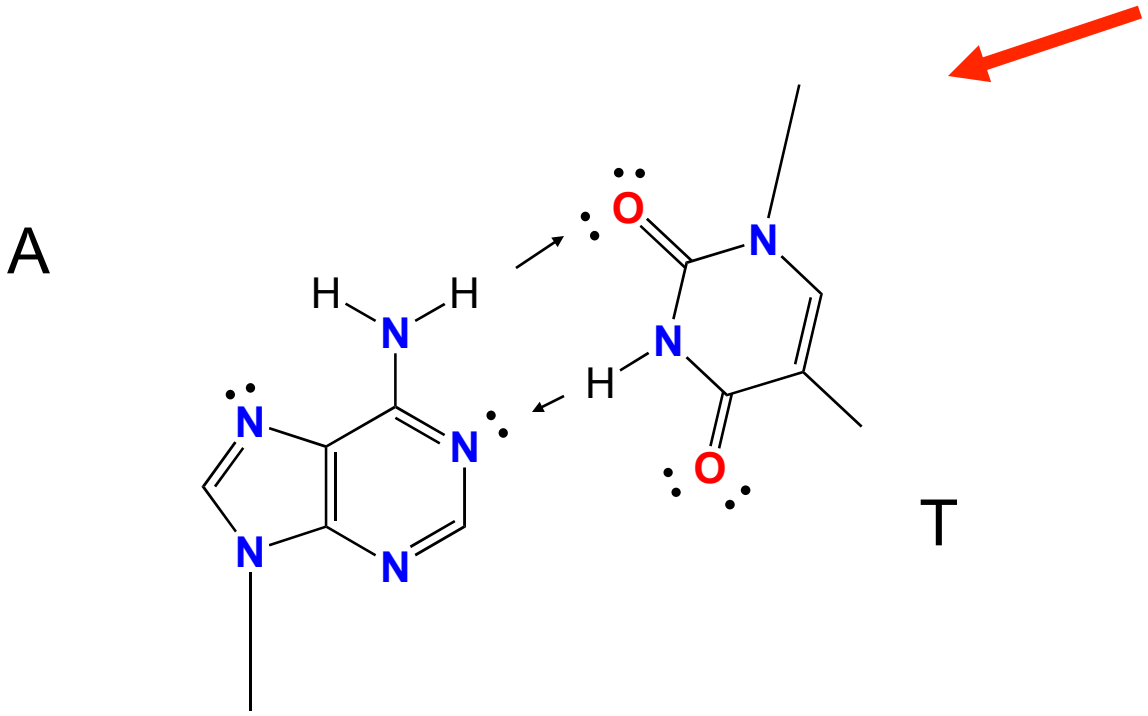



T

Good base pairing
Watson-Crick facing

# Base Pairing

(Donors matched to Acceptors)

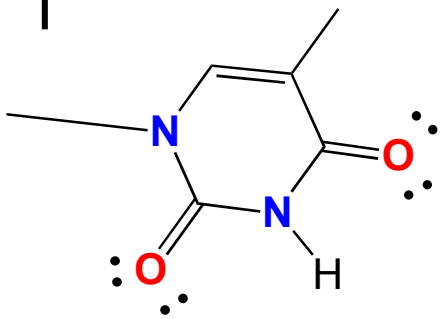


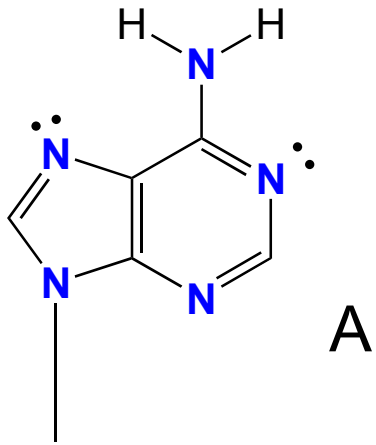

Good base pairing
Watson-Crick facing
but *Anti-Watson-Crick* orientation

# Base Pairing

(Donors matched to Acceptors)

T



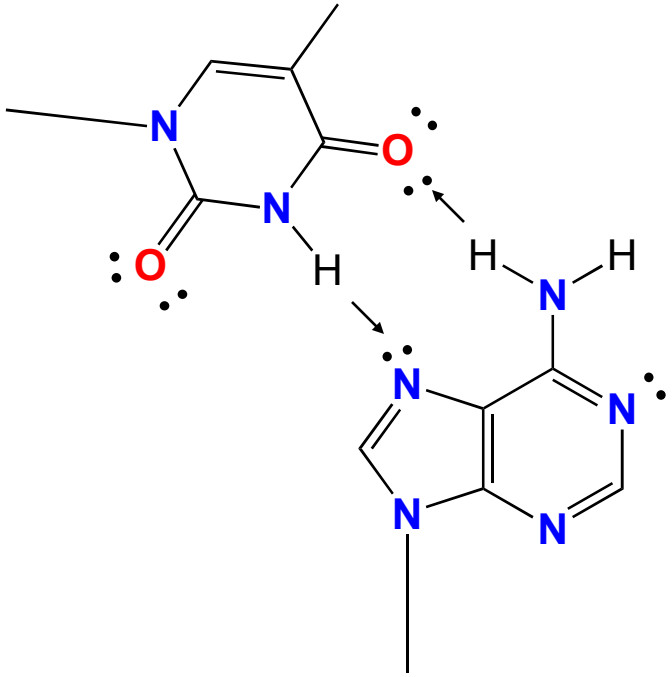




A

# Base Pairing

(Donors matched to Acceptors)

T

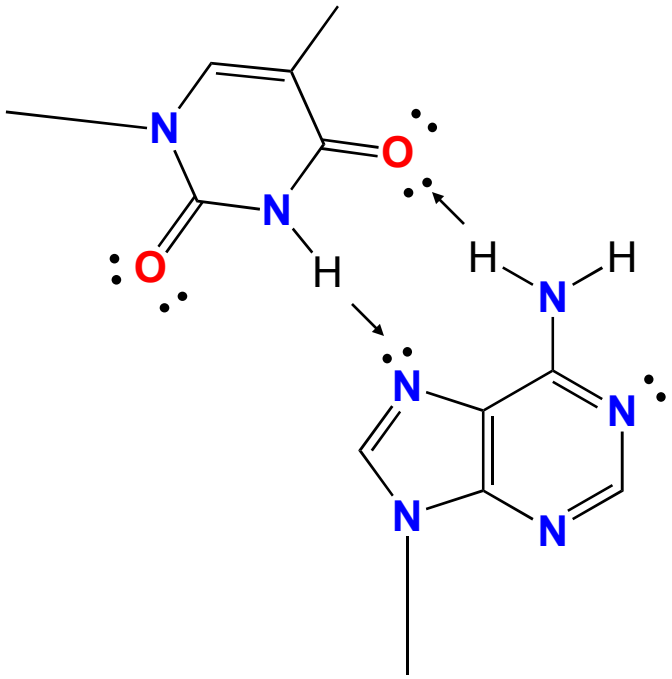



A

# Base Pairing

(Donors matched to Acceptors)

T



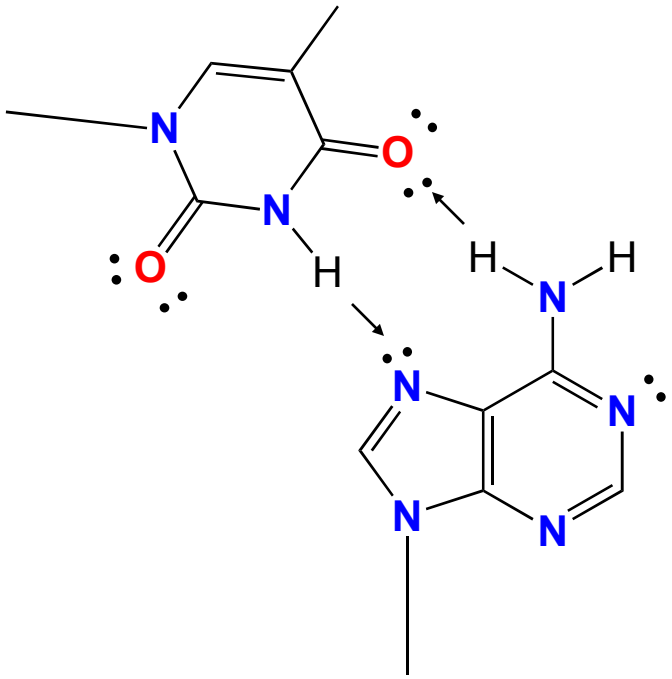

A

Good base pairing

# Base Pairing

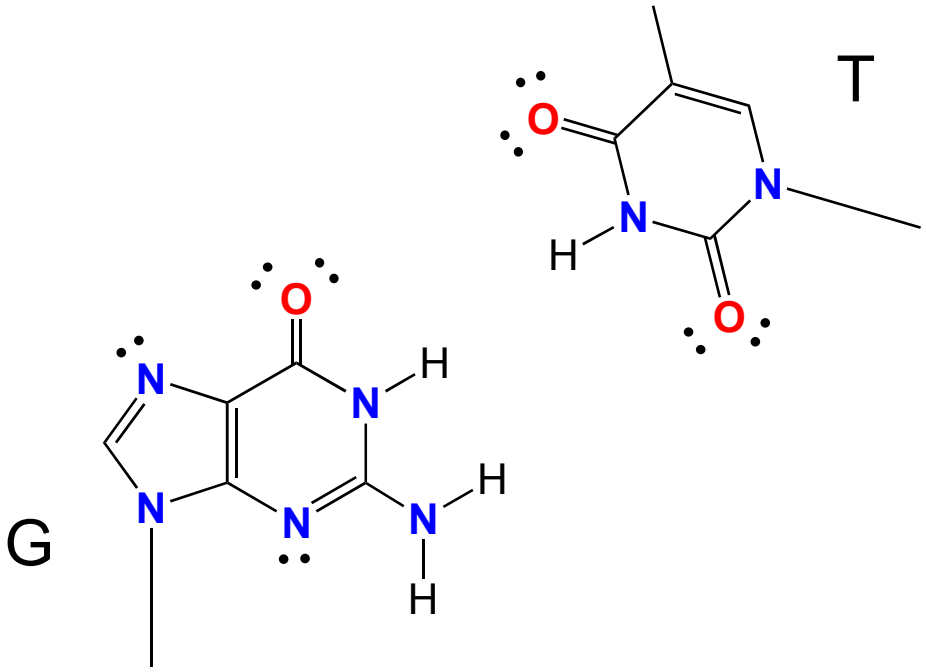
(Donors matched to Acceptors)

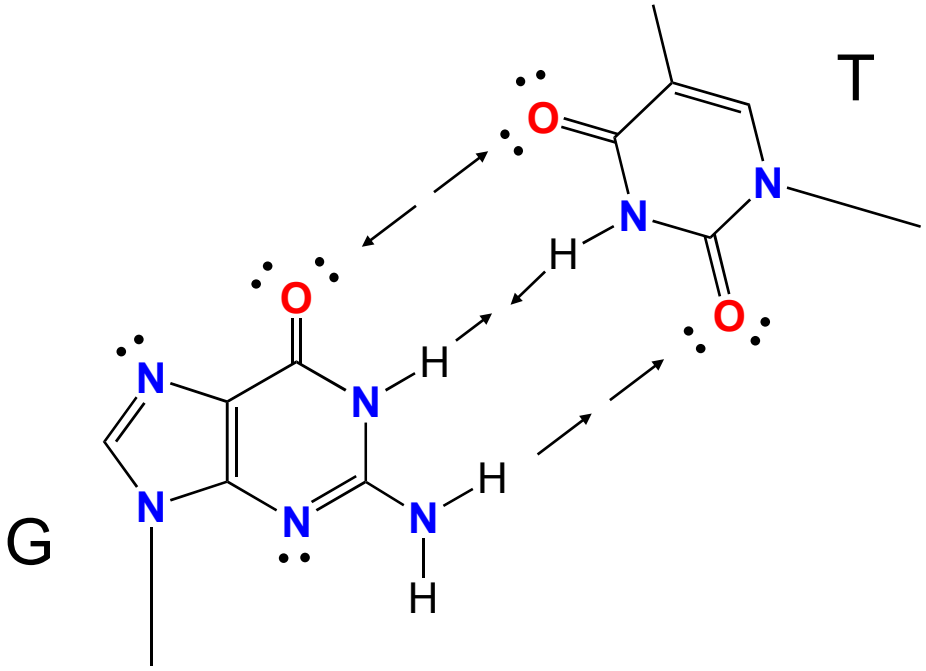
T

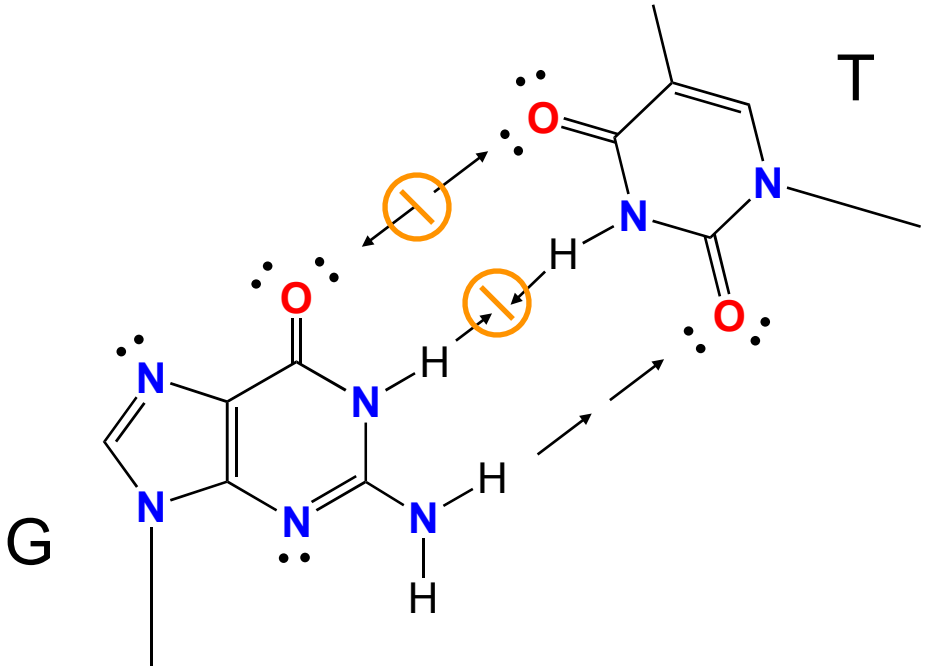



A

Good base pairing
WC-Hoogsteen facing

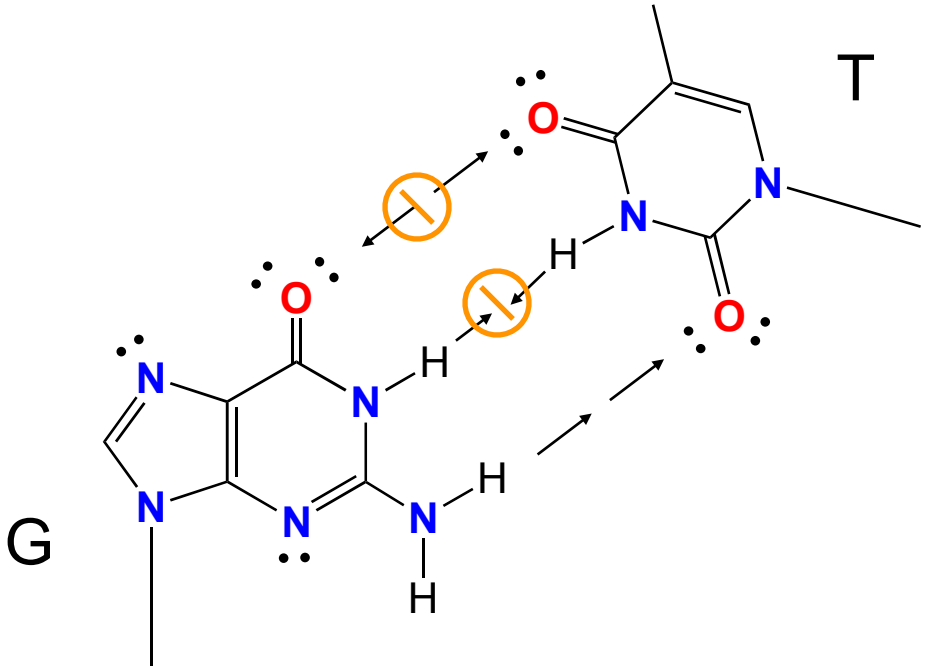

T
O
N
H
N
O
G
O
N
H
N
N
N
H
N
H

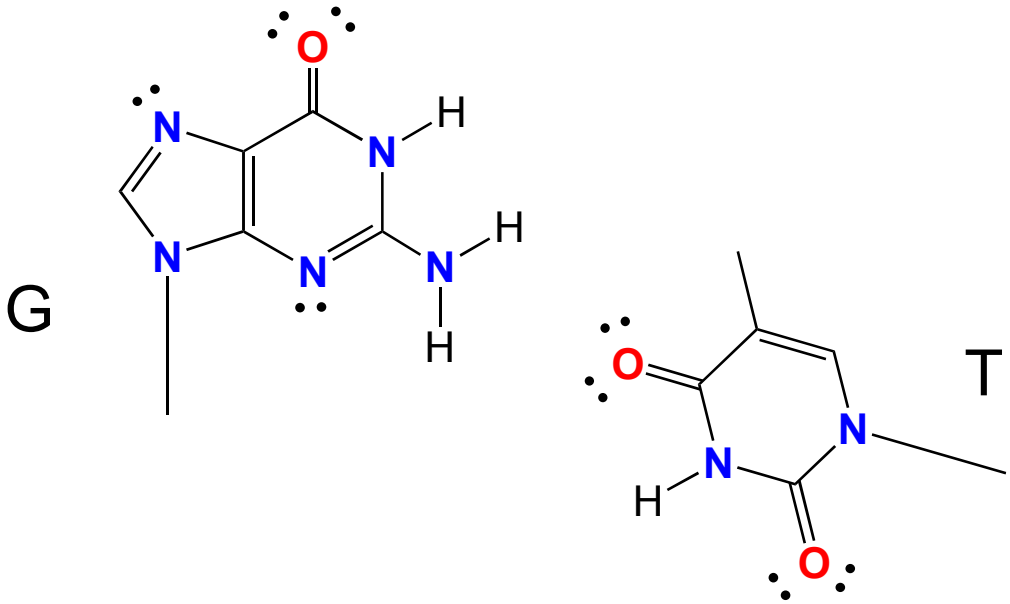

T
G
O
N
H
O
N
N
H
O
N
N
N
H
H

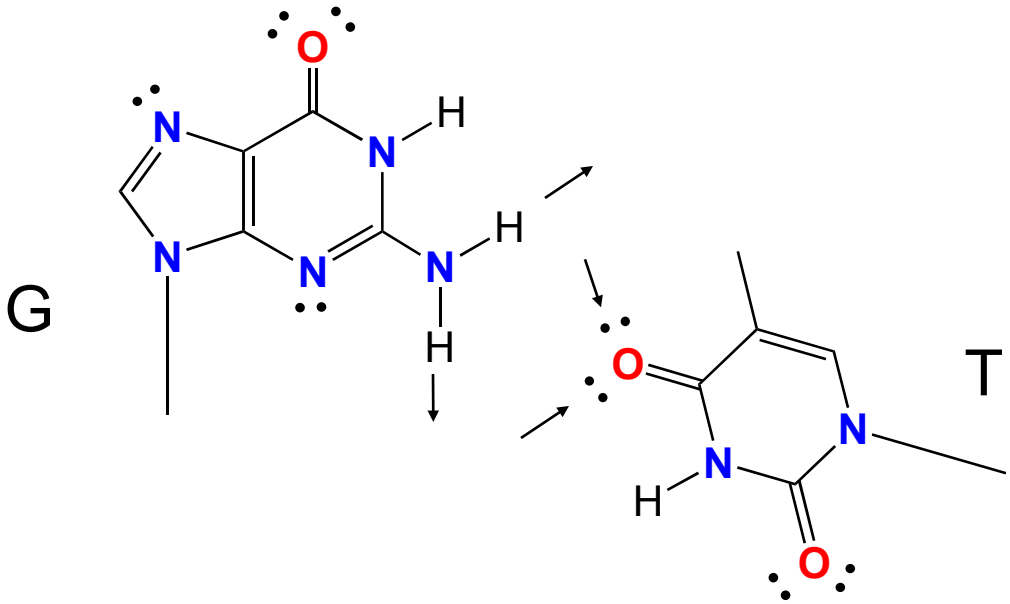

T
G
O
N
H
O
N
N
H
O
N
N
N
H
H

# Bad Base Pairing

(Donors ***not*** matched to Acceptors)

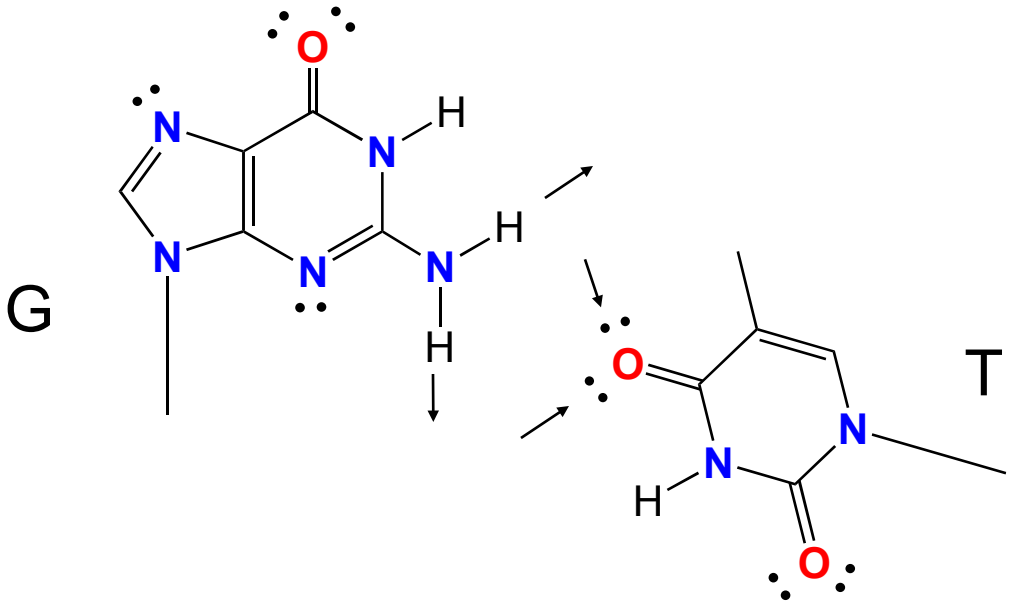



G
O
N
H
N
N
H
N
N
H
T
O
N
N
H
O

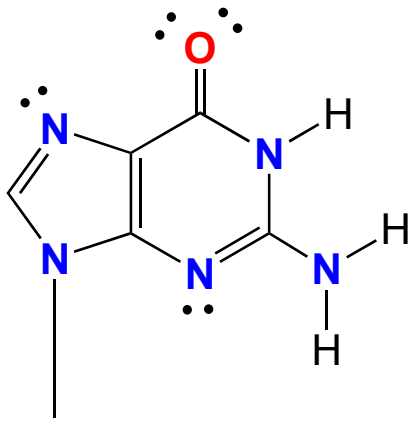

G
O
N
H
N
N
N
H
N
H
O
T
N
N
H
O

# Bad Base Pairing

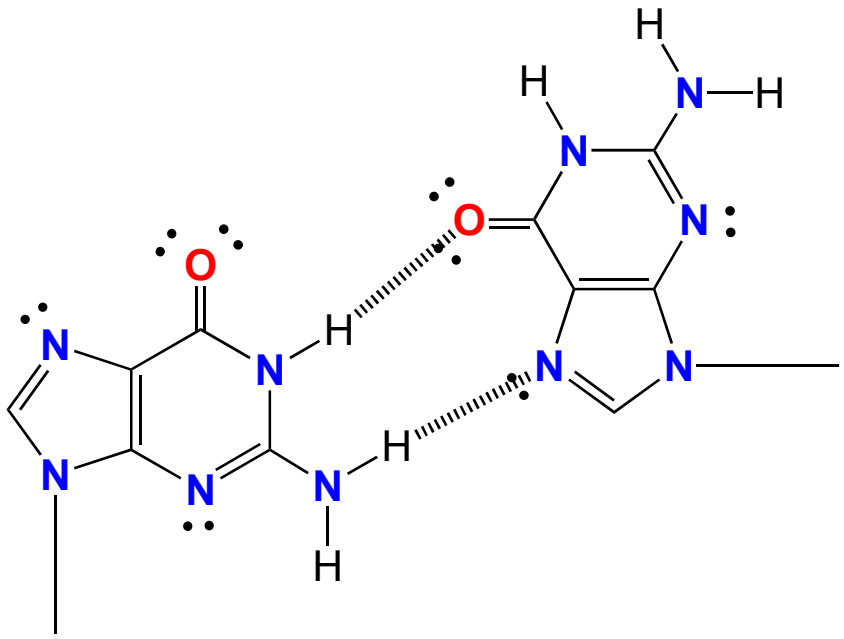
(Donors to Acceptors with ***terrible angles***)



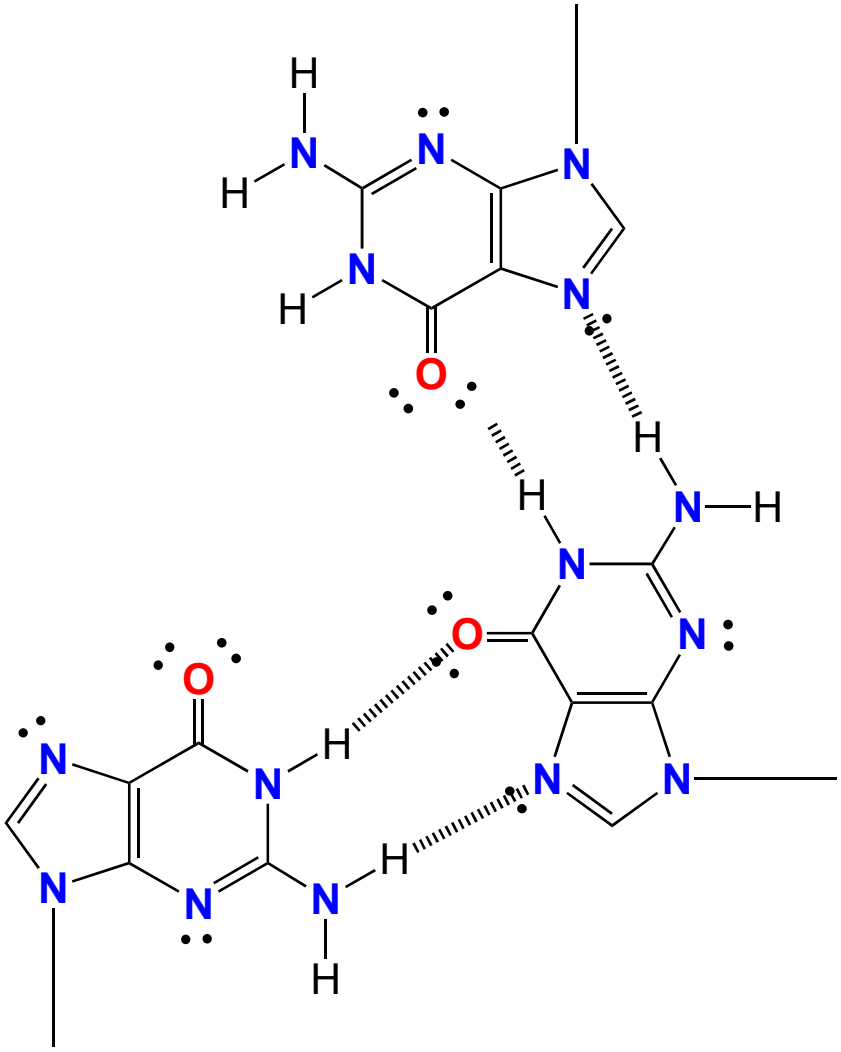
# Wild (but good) Base Pairing



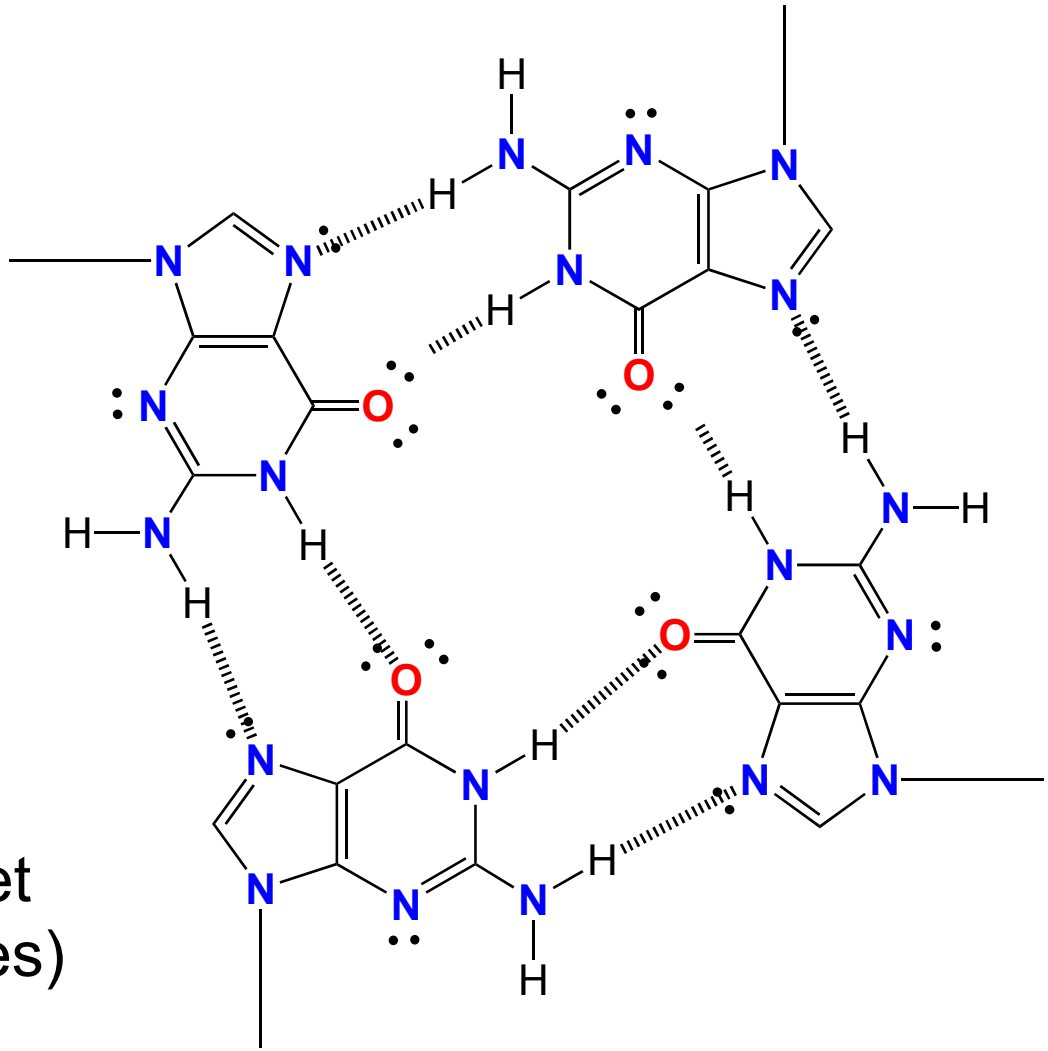
# Wild (but good) Base Pairing



# Wild (but good) Base Pairing

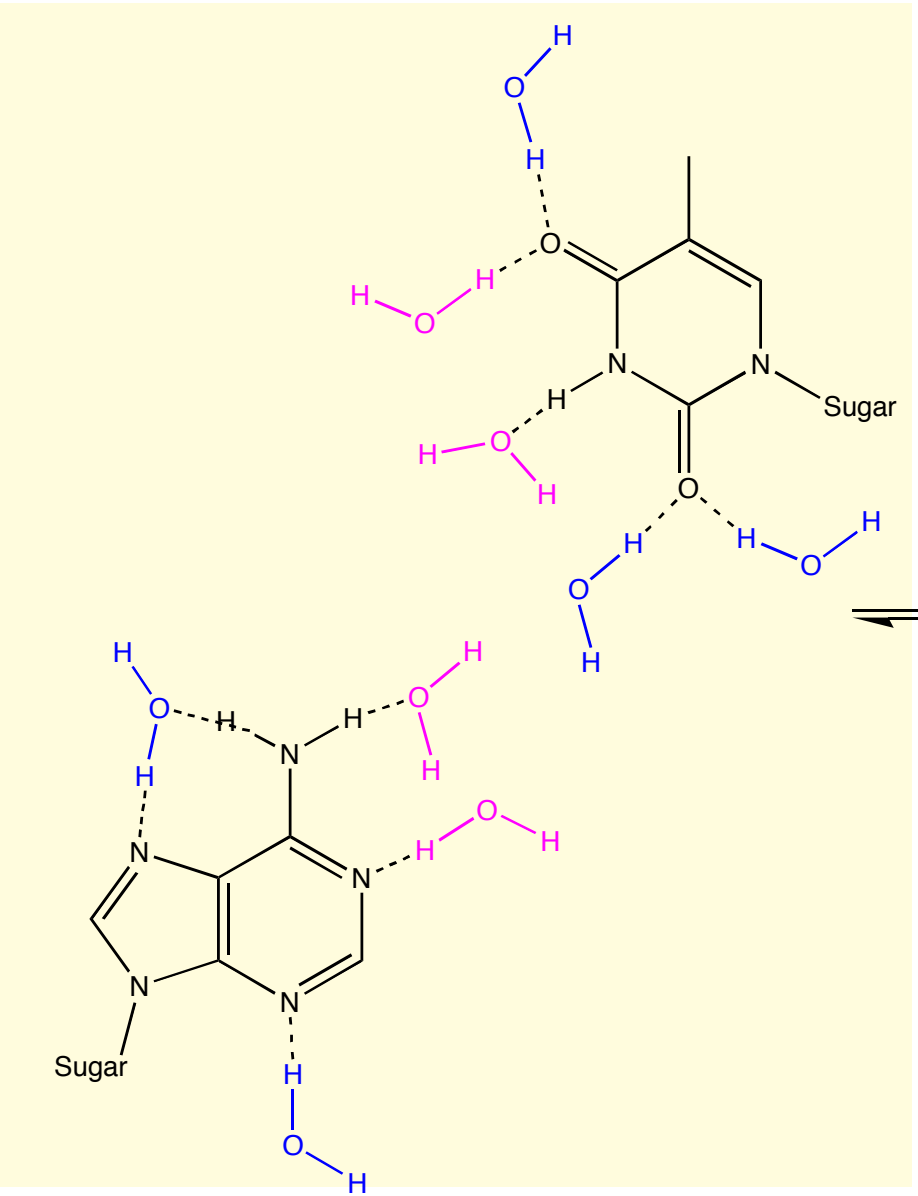


# Wild (but good) Base Pairing



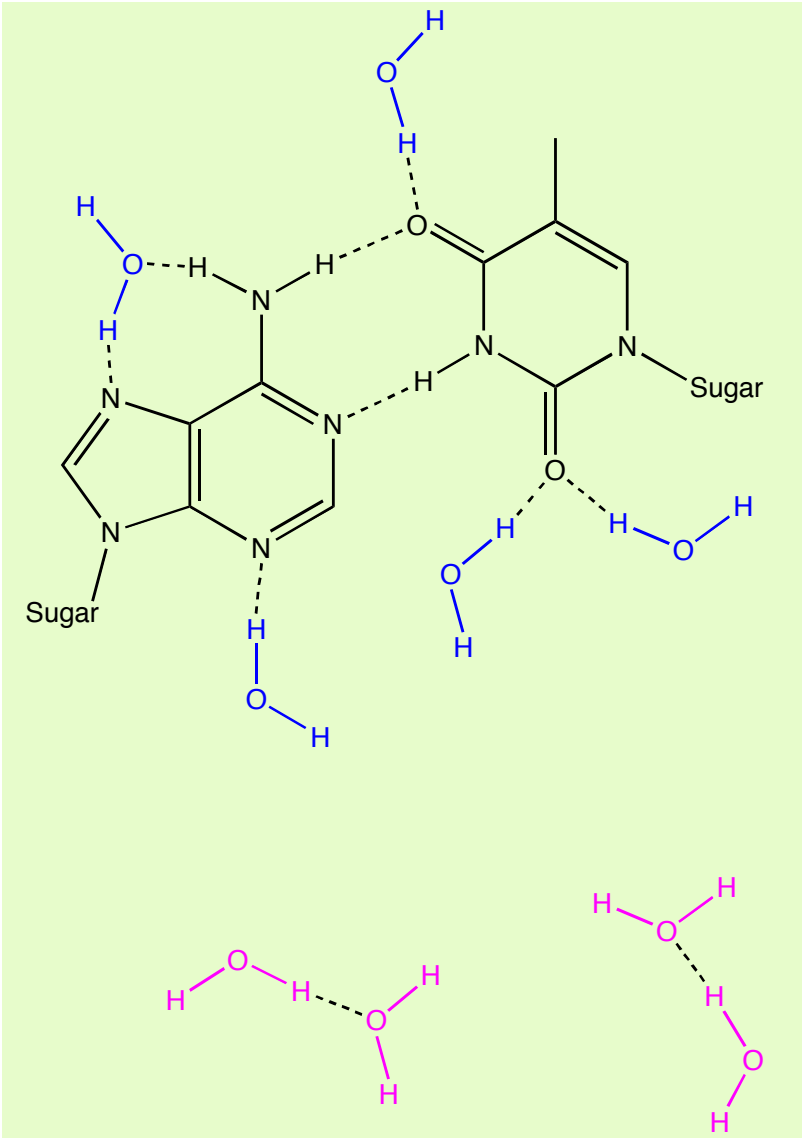
G-quartet
(Telomeres)

# AT Base Pair

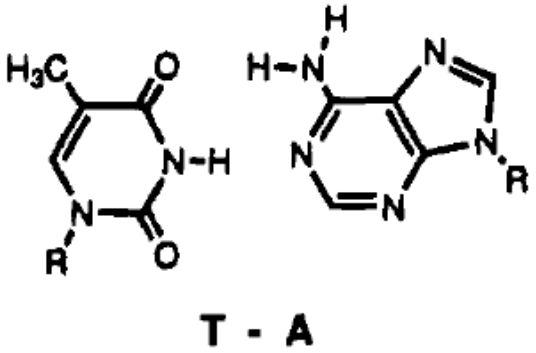



Ten H-Bonds

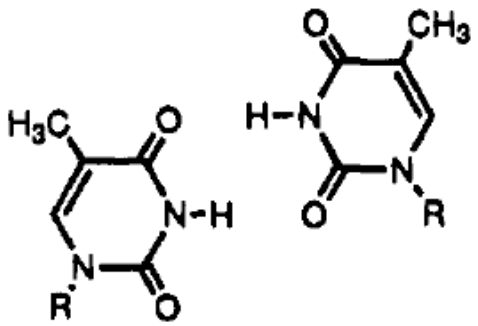
Ten H-Bonds



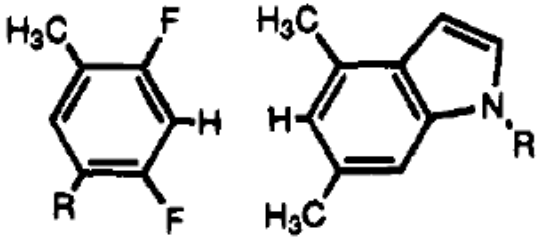
# How important are H-bonds in DNA?



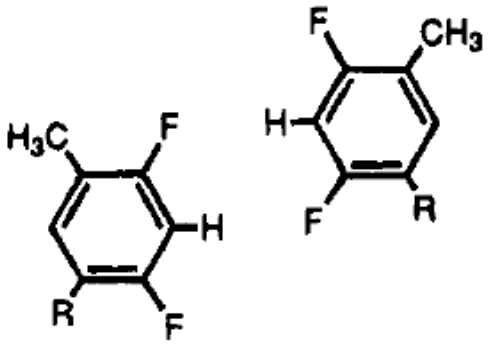
T - A



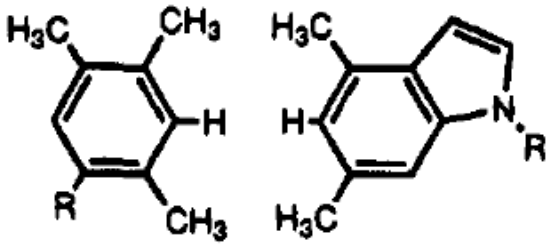
T - T



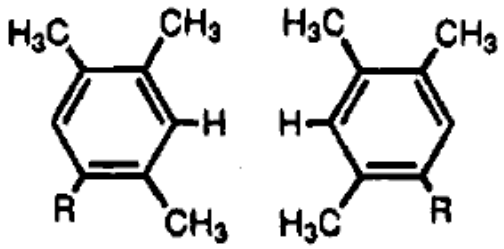
F - D



F - F



B - D



B - B

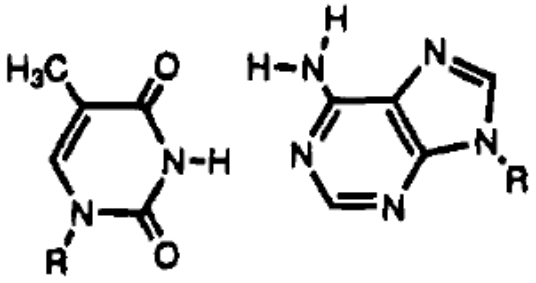


**Table 1.** Free Energies and Melting Temperatures for Dodecamer Duplexes Containing a Variable T−X, F−X, B−X, or D−X Base Pair (X = A, T, C, G)

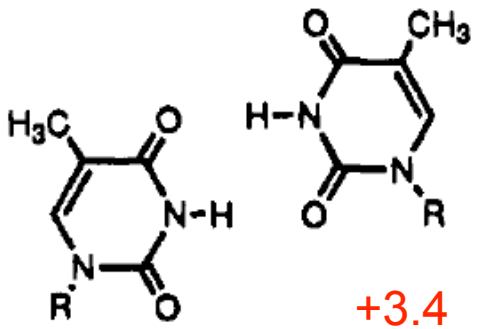
| duplex | $T_m$ (°C) [a] | $-\Delta G°_{25}$ (kcal) |
|---|---|---|
| 5′-CTTTTCTTTCTT<br>3′-GAAAAGAAAGAA | 39.4 | 12.3 |
| 5′-CTTTTCTTTCTT<br>3′-GAAAAGCAAGAA | 26.4 | 8.7 |
| 5′-CTTTTCTTTCTT<br>3′-GAAAAGGAAGAA | 30.7 | 9.3 |
| 5′-CTTTTCTTTCTT<br>3′-GAAAAGTAAGAA | 27.1 | 8.9 |
| 5′-CTTTTCFTTCTT<br>3′-GAAAAGAAAGAA | 21.4 | 7.4 |
| 5′-CTTTTCFTTCTT<br>3′-GAAAAGCAAGAA | 25.0 | 8.2 |
| 5′-CTTTTCFTTCTT<br>3′-GAAAAGGAAGAA | 23.0 | 8.0 |
| 5′-CTTTTCFTTCTT<br>3′-GAAAAGTAAGAA | 20.2 | 7.3 |
| 5′-CTTTTCBTTCTT<br>3′-GAAAAGAAAGAA | 21.0 | 7.5 |
| 5′-CTTTTCBTTCTT<br>3′-GAAAAGCAAGAA | 22.9 | 7.8 |
| 5′-CTTTTCBTTCTT<br>3′-GAAAAGGAAGAA | 20.1 | 7.6 |
| 5′-CTTTTCBTTCTT<br>3′-GAAAAGTAAGAA | 20.3 | 6.7 |
| 5′-CTTTTCDTTCTT<br>3′-GAAAAGAAAGAA | 20.8 | 7.4 |
| 5′-CTTTTCDTTCTT<br>3′-GAAAAGCAAGAA | 22.2 | 7.6 |
| 5′-CTTTTCDTTCTT<br>3′-GAAAAGGAAGAA | 19.7 | 7.4 |
| 5′-CTTTTCDTTCTT<br>3′-GAAAAGTAAGAA | 17.6 | 6.9 |

[a] Conditions: 100 mM NaCl, 10 mM $MgCl_2$, 10 mM Na · PIPES, pH 7.0, 1.6 $\mu$M each strand.

# How important are H-bonds in DNA?

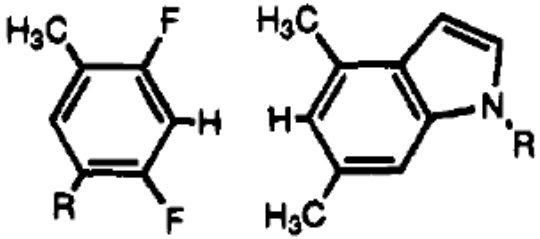


T - A

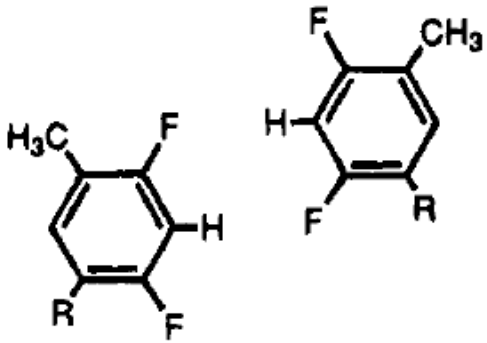


T - T

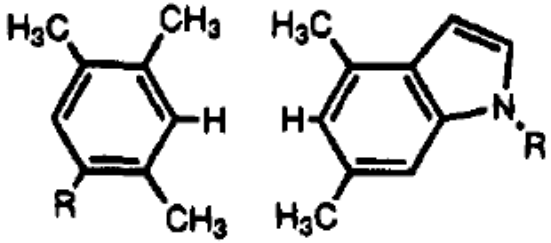
+3.4



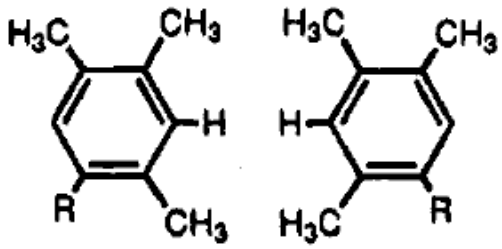
F - D



F - F



B - D



B - B

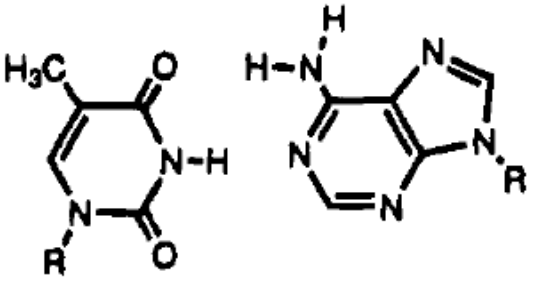
*J. Am. Chem. Soc., Vol. 117, No. 7, 1995* 1867

**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)

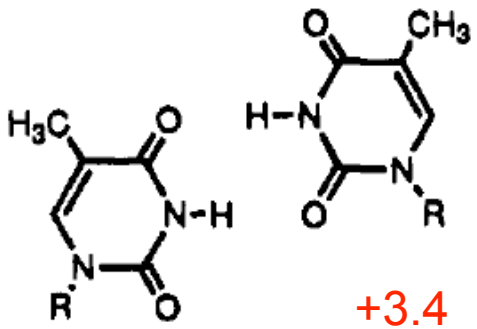
| duplex | $T_m$ (°C) [a] | $-\Delta G°_{25}$ (kcal) |
|---|---|---|
| 5′-CTTTTCTTTCTT<br>3′-GAAAAGAAAGAA | 39.4 | 12.3 |
| 5′-CTTTTCTTTCTT<br>3′-GAAAAGCAAGAA | 26.4 | 8.7 |
| 5′-CTTTTCTTTCTT<br>3′-GAAAAGGAAGAA | 30.7 | 9.3 |
| 5′-CTTTTCTTTCTT<br>3′-GAAAAGTAAGAA | 27.1 | 8.9 |
| 5′-CTTTTCFTTCTT<br>3′-GAAAAGAAAGAA | 21.4 | 7.4 |
| 5′-CTTTTCFTTCTT<br>3′-GAAAAGCAAGAA | 25.0 | 8.2 |
| 5′-CTTTTCFTTCTT<br>3′-GAAAAGGAAGAA | 23.0 | 8.0 |
| 5′-CTTTTCFTTCTT<br>3′-GAAAAGTAAGAA | 20.2 | 7.3 |
| 5′-CTTTTCBTTCTT<br>3′-GAAAAGAAAGAA | 21.0 | 7.5 |
| 5′-CTTTTCBTTCTT<br>3′-GAAAAGCAAGAA | 22.9 | 7.8 |
| 5′-CTTTTCBTTCTT<br>3′-GAAAAGGAAGAA | 20.1 | 7.6 |
| 5′-CTTTTCBTTCTT<br>3′-GAAAAGTAAGAA | 20.3 | 6.7 |
| 5′-CTTTTCDTTCTT<br>3′-GAAAAGAAAGAA | 20.8 | 7.4 |
| 5′-CTTTTCDTTCTT<br>3′-GAAAAGCAAGAA | 22.2 | 7.6 |
| 5′-CTTTTCDTTCTT<br>3′-GAAAAGGAAGAA | 19.7 | 7.4 |
| 5′-CTTTTCDTTCTT<br>3′-GAAAAGTAAGAA | 17.6 | 6.9 |

[a] Conditions: 100 mM NaCl, 10 mM $MgCl_2$, 10 mM Na · PIPES, pH 7.0, 1.6 $\mu$M each strand.

# How important are H-bonds in DNA?

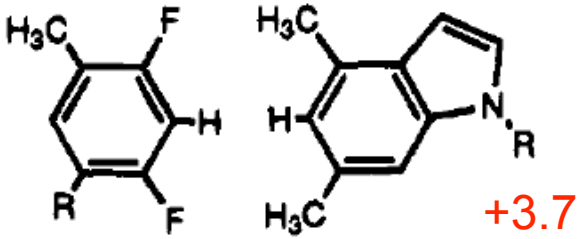


T - A



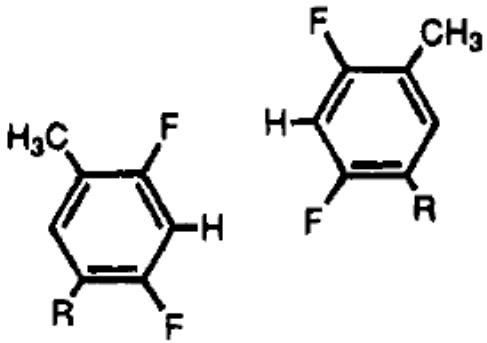
T - T

+3.4

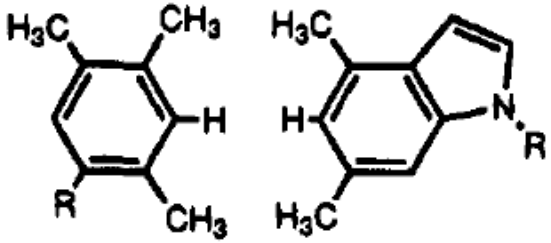


F - D

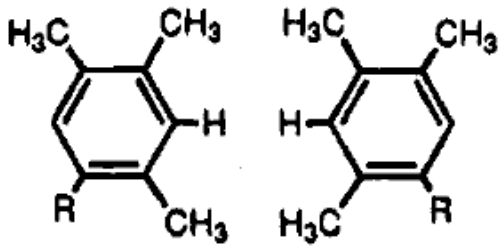
+3.7



F - F



B - D



B - B

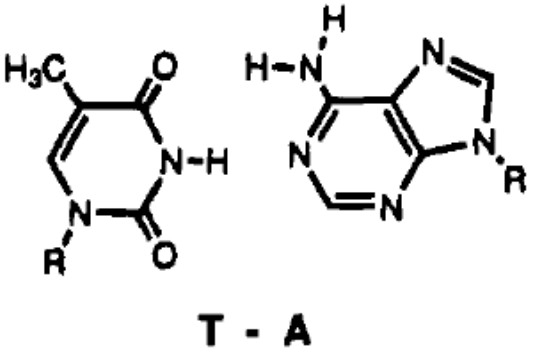
J. Am. Chem. Soc., Vol. 117, No. 7, 1995 1867

**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)

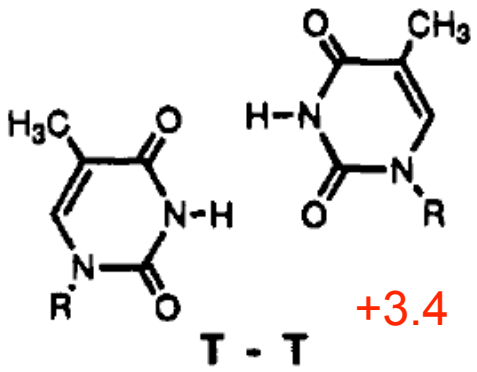
| duplex | $T_m$ (°C) [a] | $-\Delta G°_{25}$ (kcal) |
|---|---|---|
| 5′-CTTTTCTTTCTT<br>3′-GAAAAGAAAGAA | 39.4 | 12.3 |
| 5′-CTTTTCTTTCTT<br>3′-GAAAAGCAAGAA | 26.4 | 8.7 |
| 5′-CTTTTCTTTCTT<br>3′-GAAAAGGAAGAA | 30.7 | 9.3 |
| 5′-CTTTTCTTTCTT<br>3′-GAAAAGTAAGAA | 27.1 | 8.9 |
| 5′-CTTTTCFTTCTT<br>3′-GAAAAGAAAGAA | 21.4 | 7.4 |
| 5′-CTTTTCFTTCTT<br>3′-GAAAAGCAAGAA | 25.0 | 8.2 |
| 5′-CTTTTCFTTCTT<br>3′-GAAAAGGAAGAA | 23.0 | 8.0 |
| 5′-CTTTTCFTTCTT<br>3′-GAAAAGTAAGAA | 20.2 | 7.3 |
| 5′-CTTTTCBTTCTT<br>3′-GAAAAGAAAGAA | 21.0 | 7.5 |
| 5′-CTTTTCBTTCTT<br>3′-GAAAAGCAAGAA | 22.9 | 7.8 |
| 5′-CTTTTCBTTCTT<br>3′-GAAAAGGAAGAA | 20.1 | 7.6 |
| 5′-CTTTTCBTTCTT<br>3′-GAAAAGTAAGAA | 20.3 | 6.7 |
| 5′-CTTTTCDTTCTT<br>3′-GAAAAGAAAGAA | 20.8 | 7.4 |
| 5′-CTTTTCDTTCTT<br>3′-GAAAAGCAAGAA | 22.2 | 7.6 |
| 5′-CTTTTCDTTCTT<br>3′-GAAAAGGAAGAA | 19.7 | 7.4 |
| 5′-CTTTTCDTTCTT<br>3′-GAAAAGTAAGAA | 17.6 | 6.9 |

[a] Conditions: 100 mM NaCl, 10 mM $MgCl_2$, 10 mM Na · PIPES, pH 7.0, 1.6 $\mu$M each strand.

# How important are H-bonds in DNA?

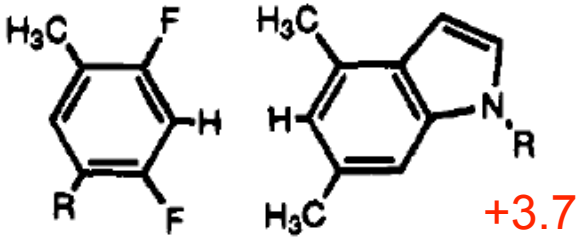


T - A



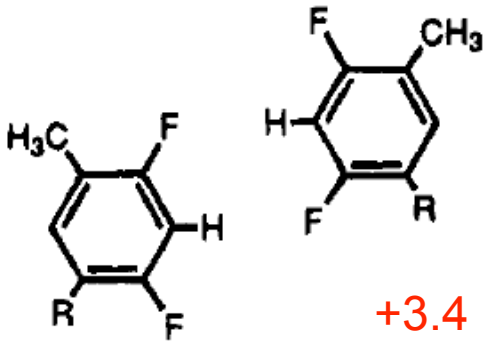
+3.4

T - T



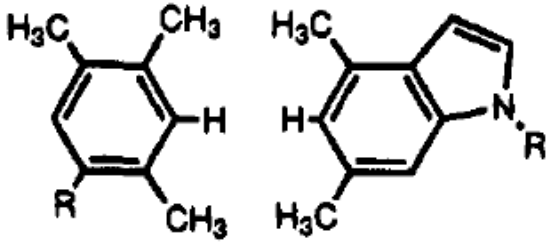
+3.7

F - D

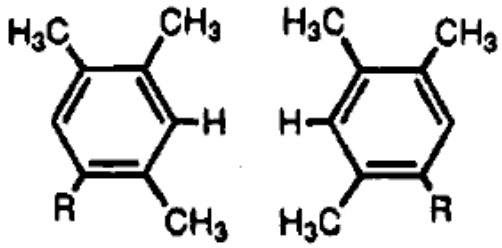


+3.4

F - F



B - D



B - B

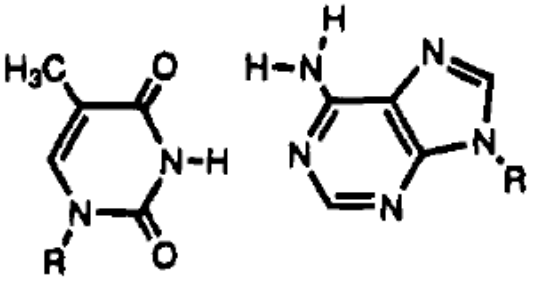
*J. Am. Chem. Soc., Vol. 117, No. 7, 1995* 1867

**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)

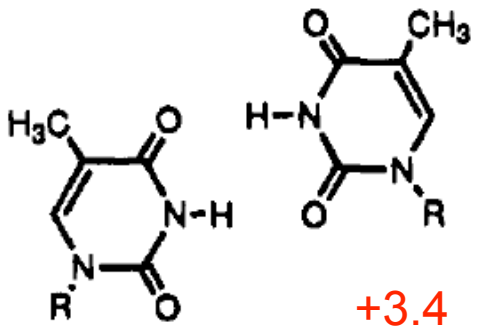
| duplex | $T_m$ (°C) [a] | $-\Delta G°_{25}$ (kcal) |
|---|---|---|
| 5′-CTTTTC**T**TTCTT<br>3′-GAAAAG**A**AAGAA | 39.4 | 12.3 |
| 5′-CTTTTC**T**TTCTT<br>3′-GAAAAG**C**AAGAA | 26.4 | 8.7 |
| 5′-CTTTTC**T**TTCTT<br>3′-GAAAAG**G**AAGAA | 30.7 | 9.3 |
| 5′-CTTTTC**T**TTCTT<br>3′-GAAAAG**T**AAGAA | 27.1 | 8.9 |
| 5′-CTTTTC**F**TTCTT<br>3′-GAAAAG**A**AAGAA | 21.4 | 7.4 |
| 5′-CTTTTC**F**TTCTT<br>3′-GAAAAG**C**AAGAA | 25.0 | 8.2 |
| 5′-CTTTTC**F**TTCTT<br>3′-GAAAAG**G**AAGAA | 23.0 | 8.0 |
| 5′-CTTTTC**F**TTCTT<br>3′-GAAAAG**T**AAGAA | 20.2 | 7.3 |
| 5′-CTTTTC**B**TTCTT<br>3′-GAAAAG**A**AAGAA | 21.0 | 7.5 |
| 5′-CTTTTC**B**TTCTT<br>3′-GAAAAG**C**AAGAA | 22.9 | 7.8 |
| 5′-CTTTTC**B**TTCTT<br>3′-GAAAAG**G**AAGAA | 20.1 | 7.6 |
| 5′-CTTTTC**B**TTCTT<br>3′-GAAAAG**T**AAGAA | 20.3 | 6.7 |
| 5′-CTTTTC**D**TTCTT<br>3′-GAAAAG**A**AAGAA | 20.8 | 7.4 |
| 5′-CTTTTC**D**TTCTT<br>3′-GAAAAG**C**AAGAA | 22.2 | 7.6 |
| 5′-CTTTTC**D**TTCTT<br>3′-GAAAAG**G**AAGAA | 19.7 | 7.4 |
| 5′-CTTTTC**D**TTCTT<br>3′-GAAAAG**T**AAGAA | 17.6 | 6.9 |

[a] Conditions: 100 mM NaCl, 10 mM $MgCl_2$, 10 mM Na · PIPES, pH 7.0, 1.6 $\mu$M each strand.

# How important are H-bonds in DNA?

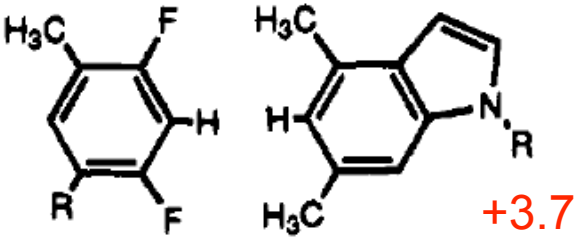


T - A



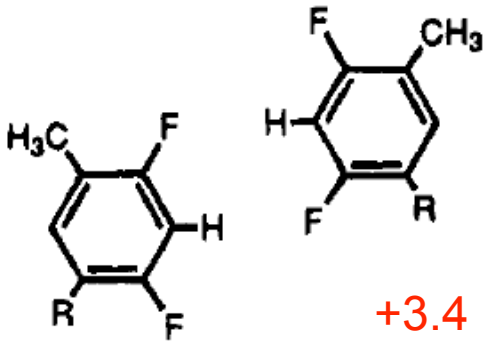
+3.4

T - T



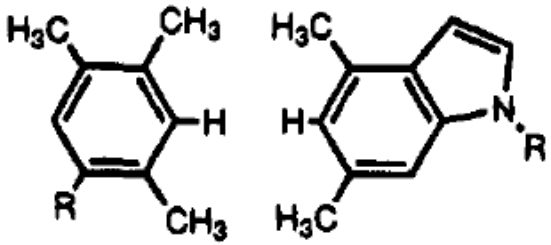
+3.7

F - D



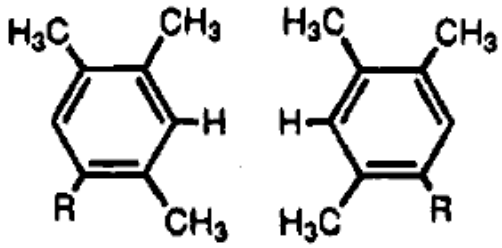
+3.4

F - F



+3.5

B - D



B - B

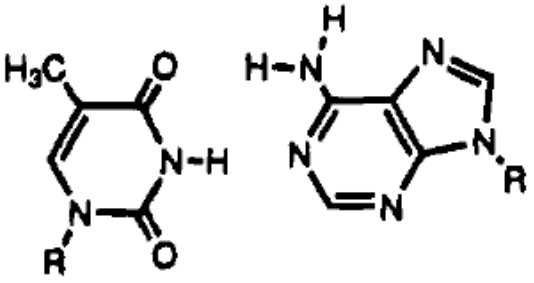
*J. Am. Chem. Soc., Vol. 117, No. 7, 1995* 1867

**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)

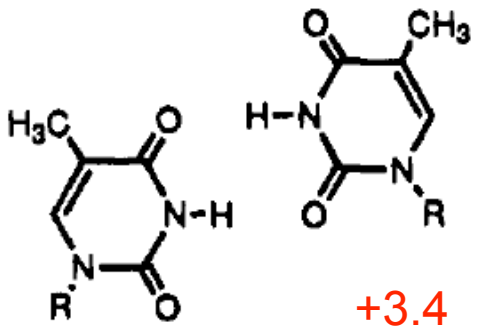
| duplex | $T_m$ (°C) [a] | $-\Delta G°_{25}$ (kcal) |
|---|---|---|
| 5′-CTTTTCTTTCTT<br>3′-GAAAAGAAAGAA | 39.4 | 12.3 |
| 5′-CTTTTCTTTCTT<br>3′-GAAAAGCAAGAA | 26.4 | 8.7 |
| 5′-CTTTTCTTTCTT<br>3′-GAAAAGGAAGAA | 30.7 | 9.3 |
| 5′-CTTTTCTTTCTT<br>3′-GAAAAGTAAGAA | 27.1 | 8.9 |
| 5′-CTTTTCFTTCTT<br>3′-GAAAAGAAAGAA | 21.4 | 7.4 |
| 5′-CTTTTCFTTCTT<br>3′-GAAAAGCAAGAA | 25.0 | 8.2 |
| 5′-CTTTTCFTTCTT<br>3′-GAAAAGGAAGAA | 23.0 | 8.0 |
| 5′-CTTTTCFTTCTT<br>3′-GAAAAGTAAGAA | 20.2 | 7.3 |
| 5′-CTTTTCBTTCTT<br>3′-GAAAAGAAAGAA | 21.0 | 7.5 |
| 5′-CTTTTCBTTCTT<br>3′-GAAAAGCAAGAA | 22.9 | 7.8 |
| 5′-CTTTTCBTTCTT<br>3′-GAAAAGGAAGAA | 20.1 | 7.6 |
| 5′-CTTTTCBTTCTT<br>3′-GAAAAGTAAGAA | 20.3 | 6.7 |
| 5′-CTTTTCDTTCTT<br>3′-GAAAAGAAAGAA | 20.8 | 7.4 |
| 5′-CTTTTCDTTCTT<br>3′-GAAAAGCAAGAA | 22.2 | 7.6 |
| 5′-CTTTTCDTTCTT<br>3′-GAAAAGGAAGAA | 19.7 | 7.4 |
| 5′-CTTTTCDTTCTT<br>3′-GAAAAGTAAGAA | 17.6 | 6.9 |

[a] Conditions: 100 mM NaCl, 10 mM $MgCl_2$, 10 mM Na · PIPES, pH 7.0, 1.6 $\mu$M each strand.

# How important are H-bonds in DNA?

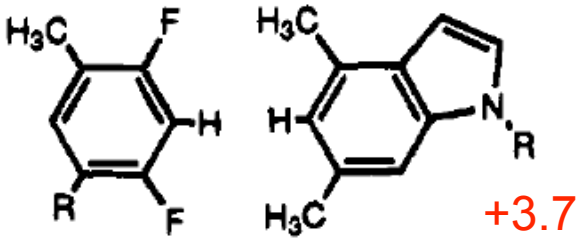


T - A



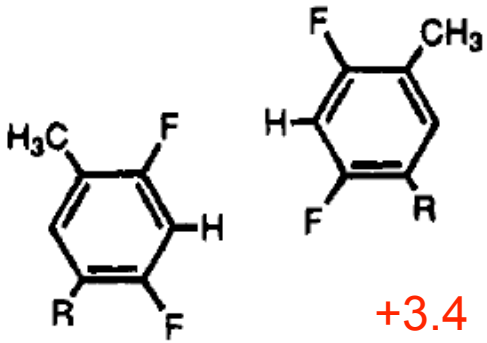
+3.4

T - T



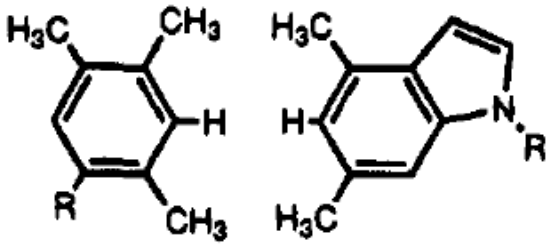
+3.7

F - D



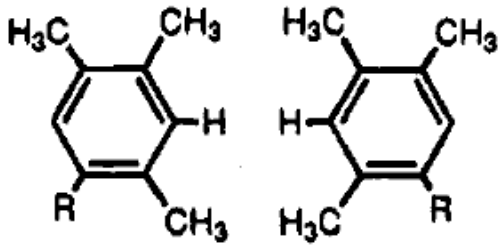
+3.4

F - F



+3.5

B - D



+3.0

B - B

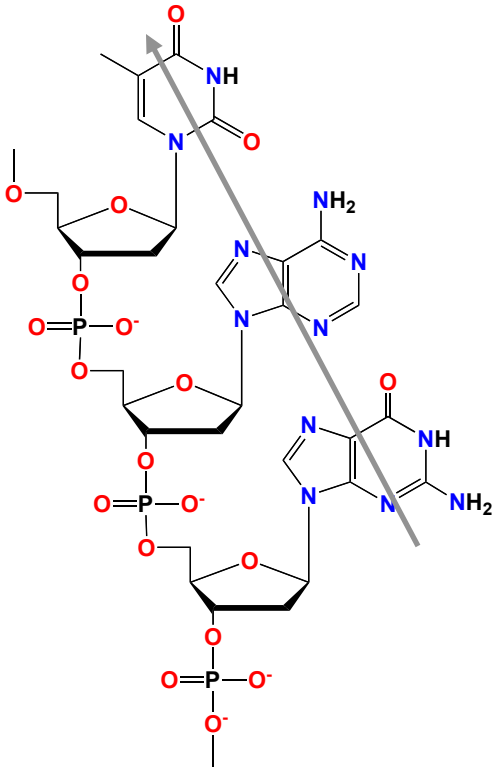
*J. Am. Chem. Soc., Vol. 117, No. 7, 1995* 1867

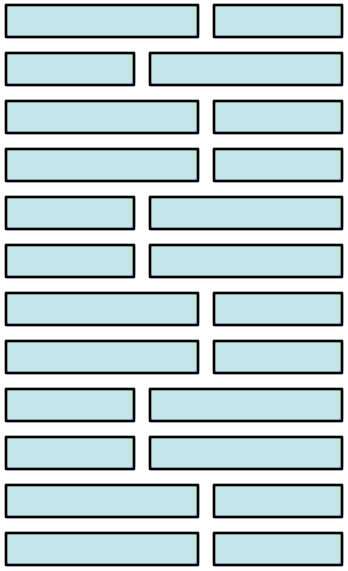
**Table 1.** Free Energies and Melting Temperatures for Dodecamer Duplexes Containing a Variable T−X, F−X, B−X, or D−X Base Pair (X = A, T, C, G)

| duplex | $T_m$ (°C) [a] | $-\Delta G°_{25}$ (kcal) |
|---|---|---|
| 5′-CTTTTCTTTCTT<br>3′-GAAAAGAAAGAA | 39.4 | 12.3 |
| 5′-CTTTTCTTTCTT<br>3′-GAAAAGCAAGAA | 26.4 | 8.7 |
| 5′-CTTTTCTTTCTT<br>3′-GAAAAGGAAGAA | 30.7 | 9.3 |
| 5′-CTTTTCTTTCTT<br>3′-GAAAAGTAAGAA | 27.1 | 8.9 |
| 5′-CTTTTCFTTCTT<br>3′-GAAAAGAAAGAA | 21.4 | 7.4 |
| 5′-CTTTTCFTTCTT<br>3′-GAAAAGCAAGAA | 25.0 | 8.2 |
| 5′-CTTTTCFTTCTT<br>3′-GAAAAGGAAGAA | 23.0 | 8.0 |
| 5′-CTTTTCFTTCTT<br>3′-GAAAAGTAAGAA | 20.2 | 7.3 |
| 5′-CTTTTCBTTCTT<br>3′-GAAAAGAAAGAA | 21.0 | 7.5 |
| 5′-CTTTTCBTTCTT<br>3′-GAAAAGCAAGAA | 22.9 | 7.8 |
| 5′-CTTTTCBTTCTT<br>3′-GAAAAGGAAGAA | 20.1 | 7.6 |
| 5′-CTTTTCBTTCTT<br>3′-GAAAAGTAAGAA | 20.3 | 6.7 |
| 5′-CTTTTCDTTCTT<br>3′-GAAAAGAAAGAA | 20.8 | 7.4 |
| 5′-CTTTTCDTTCTT<br>3′-GAAAAGCAAGAA | 22.2 | 7.6 |
| 5′-CTTTTCDTTCTT<br>3′-GAAAAGGAAGAA | 19.7 | 7.4 |
| 5′-CTTTTCDTTCTT<br>3′-GAAAAGTAAGAA | 17.6 | 6.9 |

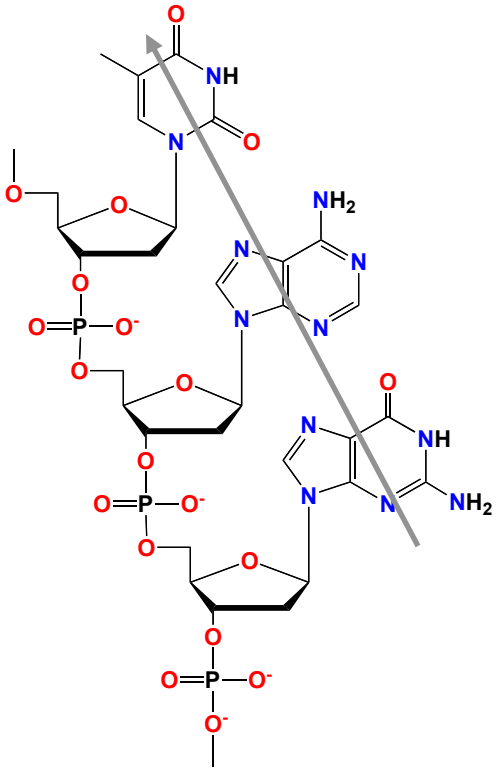
[a] Conditions: 100 mM NaCl, 10 mM $MgCl_2$, 10 mM Na · PIPES, pH 7.0, 1.6 $\mu$M each strand.

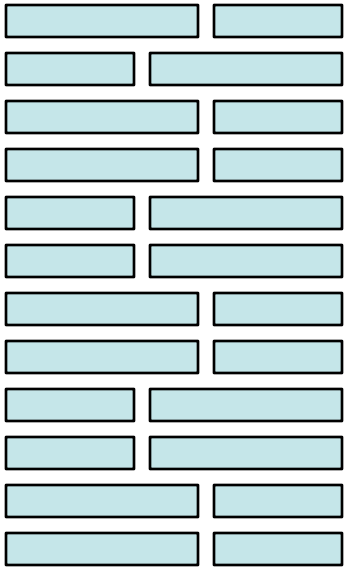
# Burial of hydrophobic surface drives helix formation (hydrophobic core / stacking interactions)



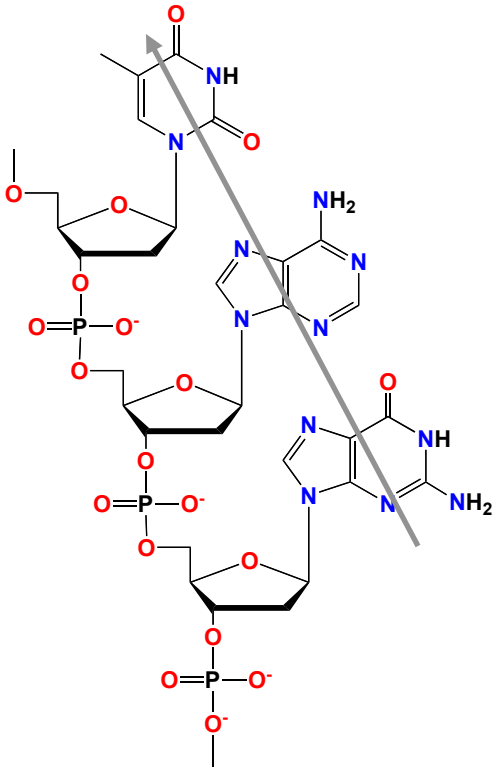
# Burial of hydrophobic surface drives helix formation (hydrophobic core / stacking interactions)

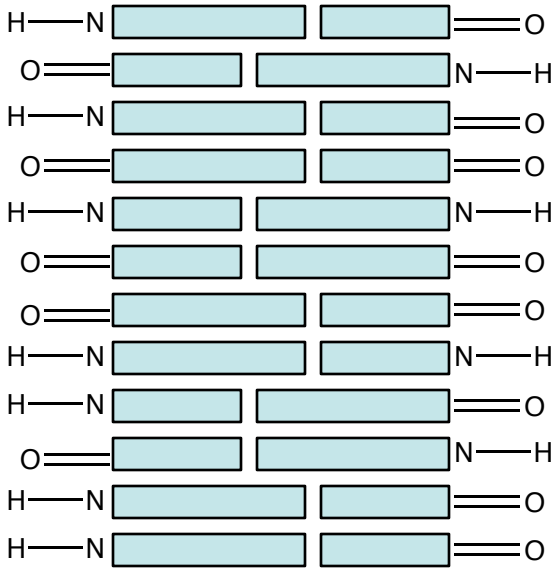




Flat faces are nonpolar

# Burial of hydrophobic surface drives helix formation (hydrophobic core / stacking interactions)

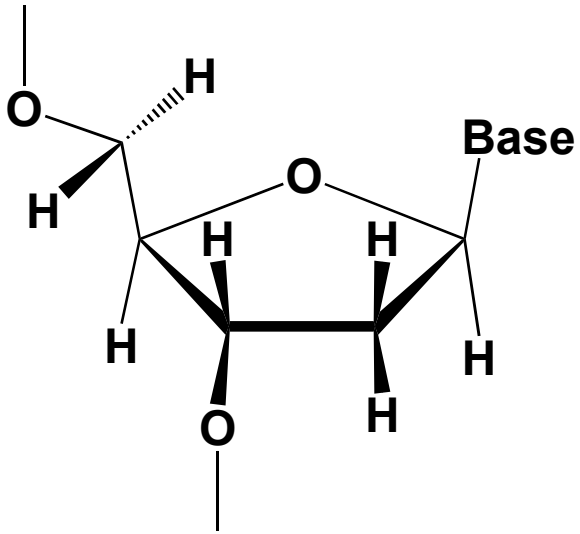


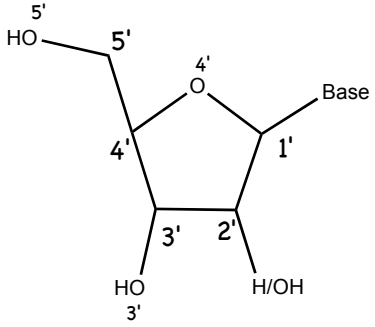


Flat faces are nonpolar
Edges are very polar (can H-bond)

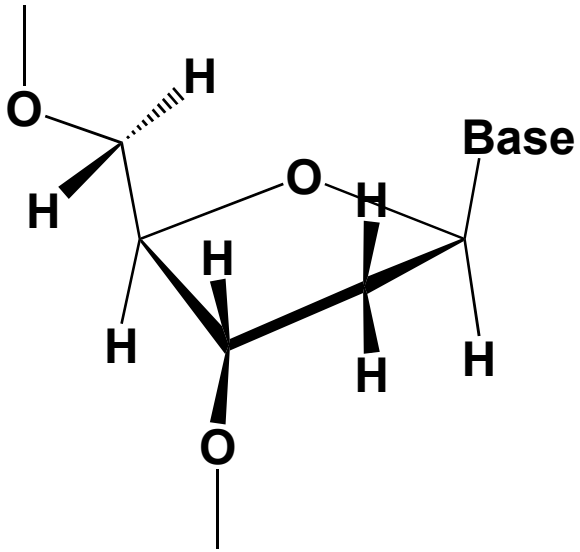
# Other chemical constraints

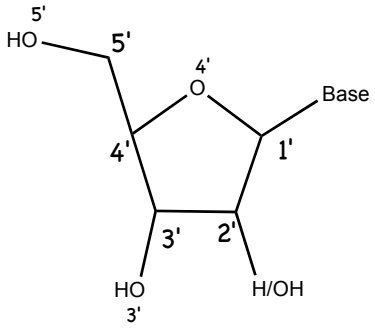
# Furanose Sugar Ring



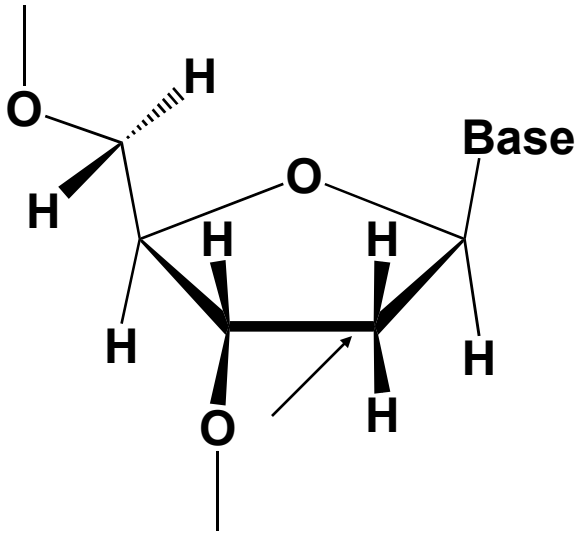



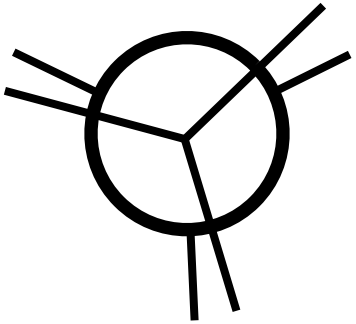
# Furanose Sugar Ring

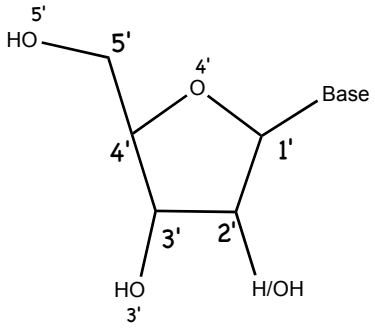


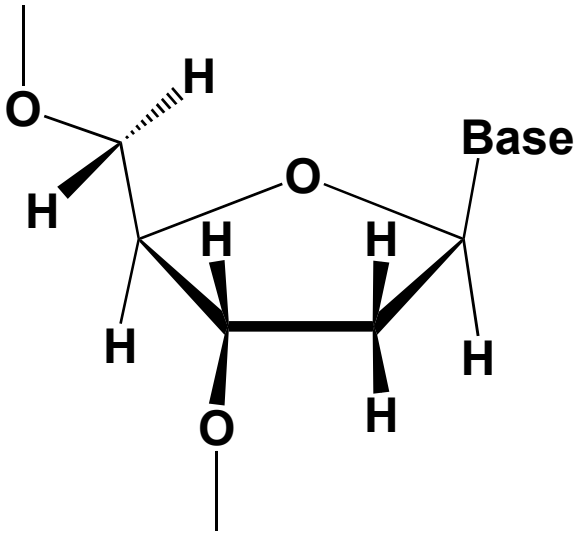
# Furanose Sugar Ring

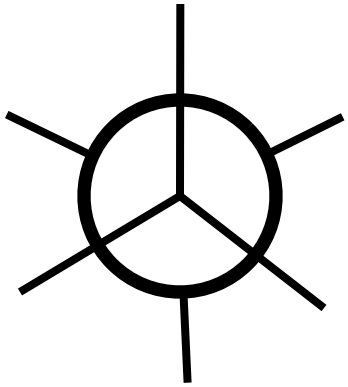


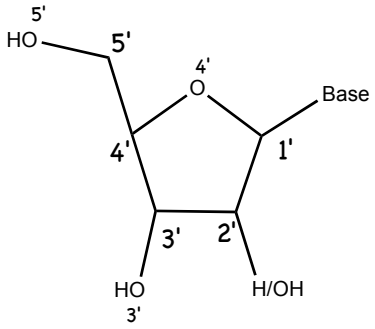





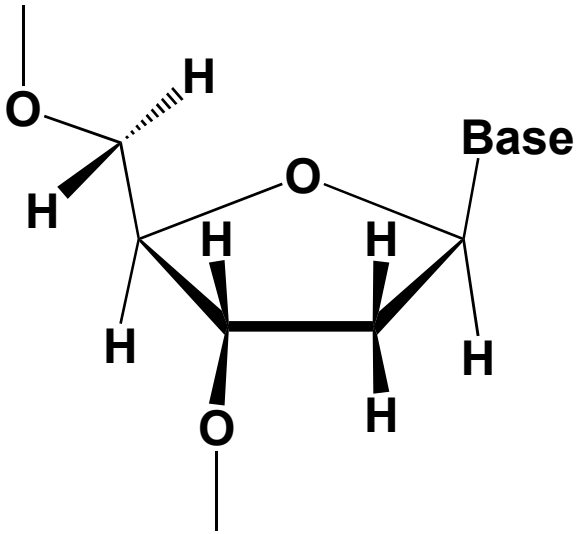
# Furanose Sugar Ring

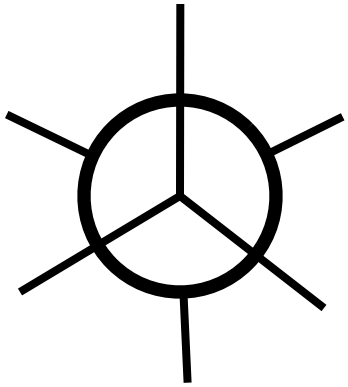


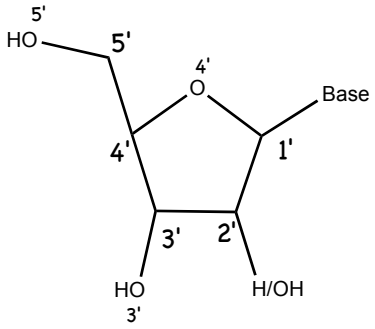



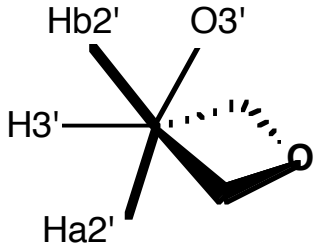


# Furanose Sugar Ring

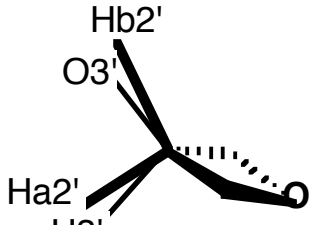




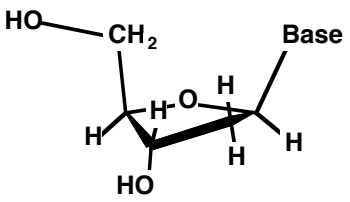




Staggered
Puckered

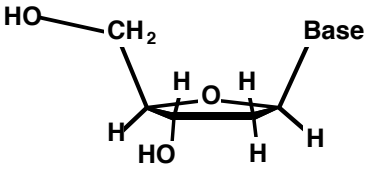



Eclipsed
Planar

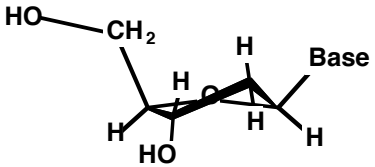
# Furanose Sugar Ring




C3' exo



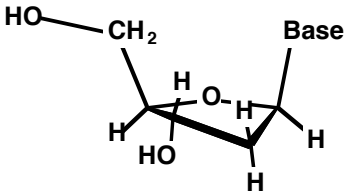

Unpuckered



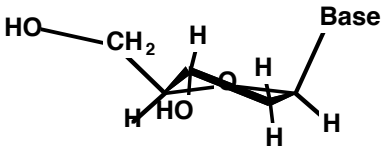

C2' endo

endo

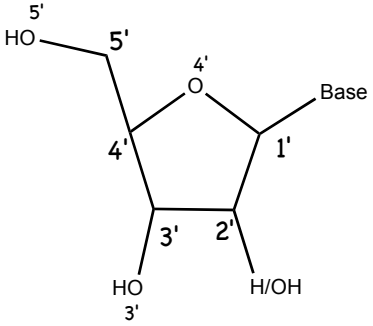
exo

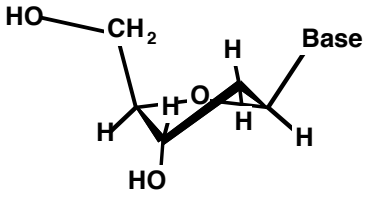



C2' exo

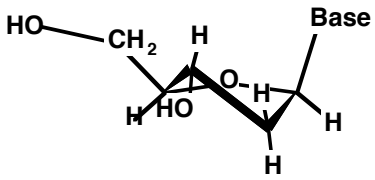



C3' endo

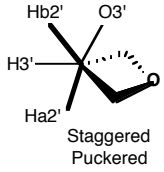


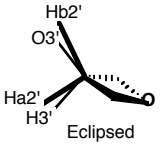

C2' endo - C3' exo




C2' exo - C3' endo




Staggered
Puckered




Eclipsed
Planar

# Hoogsteen vindicated

**Letters to Nature**

*Nature* **430**, 377-380 (15 July 2004) | doi: 10.1038/nature02692

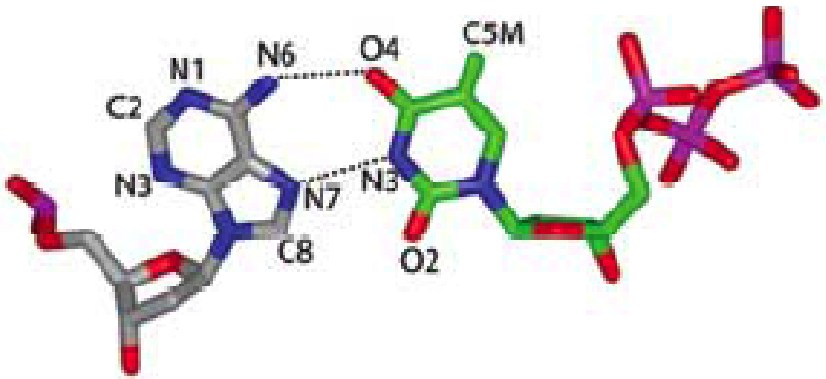
## Replication by human DNA polymerase-ι occurs by Hoogsteen base-pairing

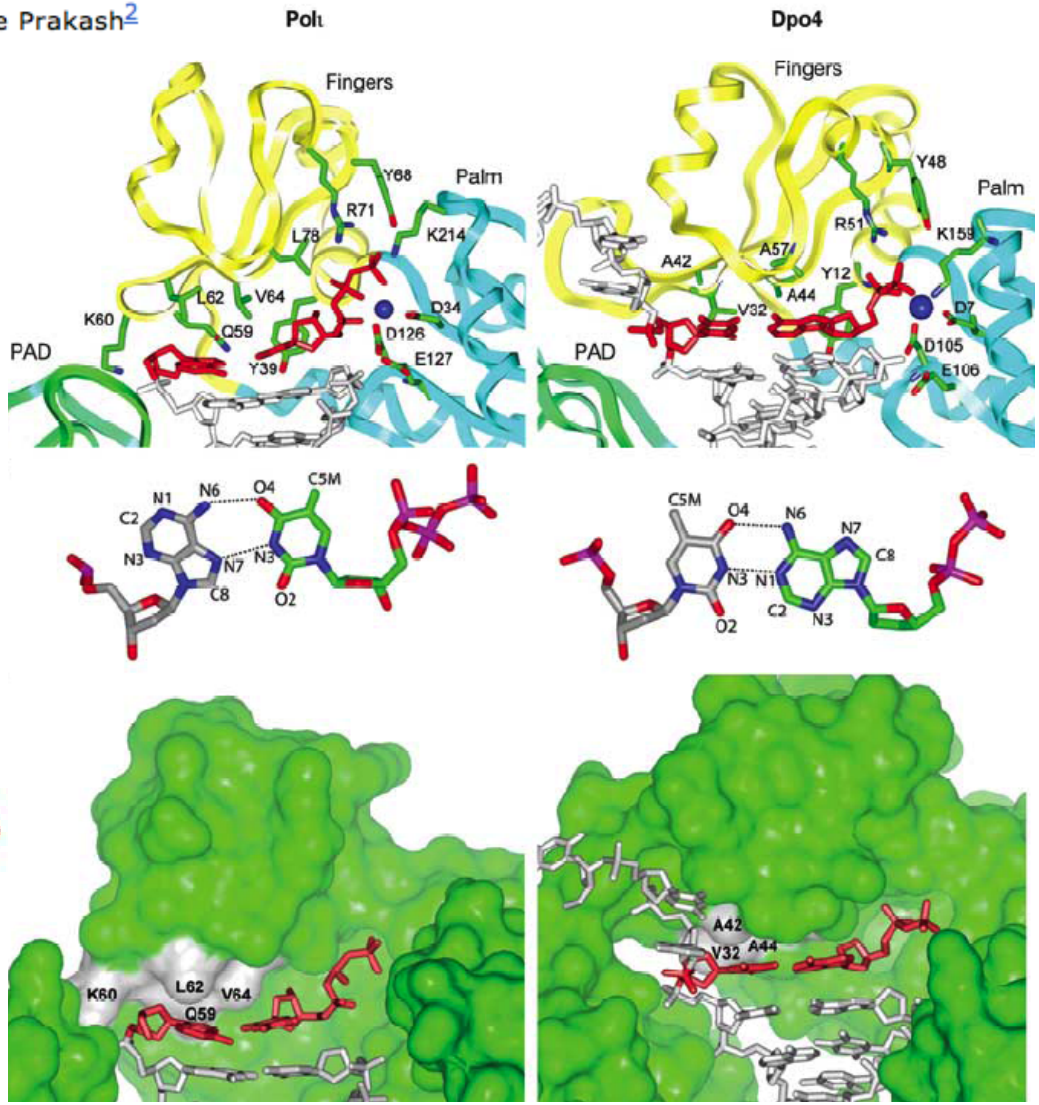
Deepak T. Nair[1], Robert E. Johnson[2], Satya Prakash[2], Louise Prakash[2] and Aneel K. Aggarwal[1]

A crystal structure from 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).



# Hoogsteen vindicated

**But… perhaps not…**

**Letters to Nature**

*Nature* **430**, 377-380 (15 July 2004) | doi: 10.1038/nature02692

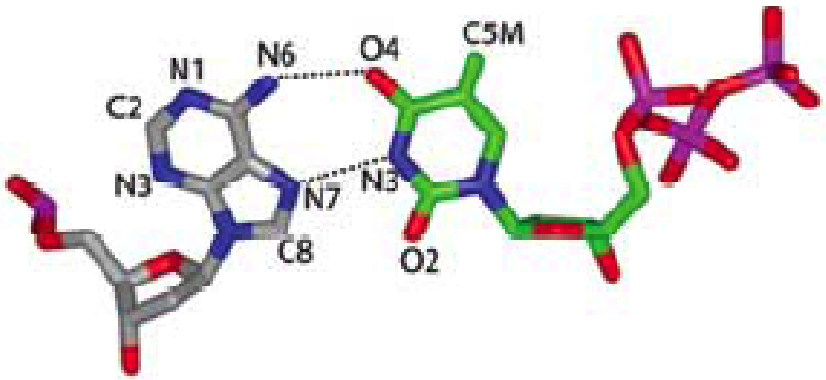
## Replication by human DNA polymerase-ι occurs by Hoogsteen base-pairing

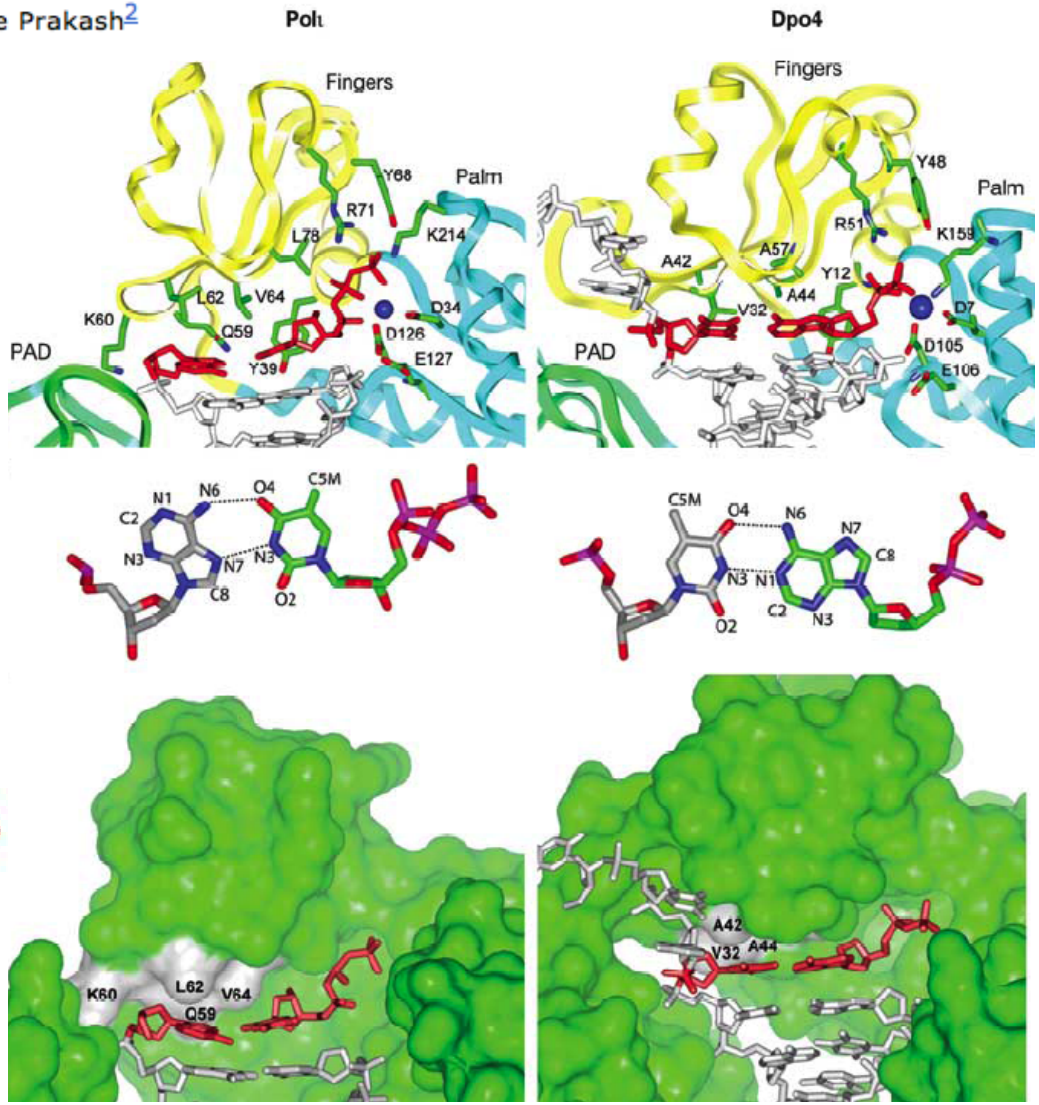
Deepak T. Nair[1], Robert E. Johnson[2], Satya Prakash[2], Louise Prakash[2] and Aneel K. Aggarwal[1]

A crystal structure from 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.

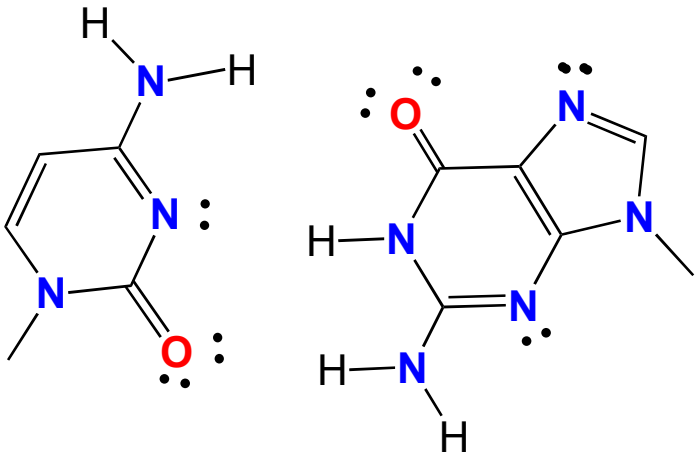
This polymerase is good at bypassing lesions in the DNA (eg, damaged Watson-Crick face of G).



# Why is Watson-Crick so good?

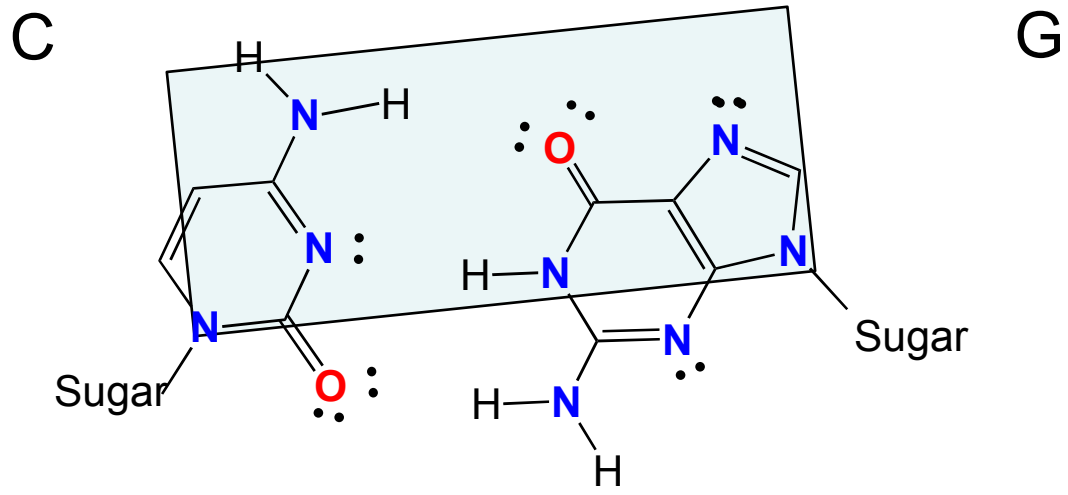
C




G

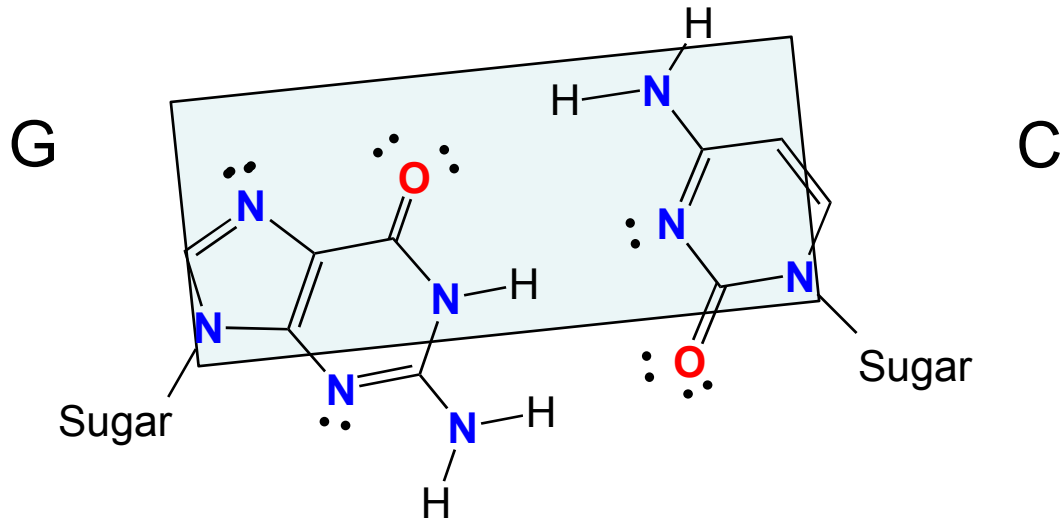
All four WC base pairs
are
***isosteric***

# Why is Watson-Crick so good?



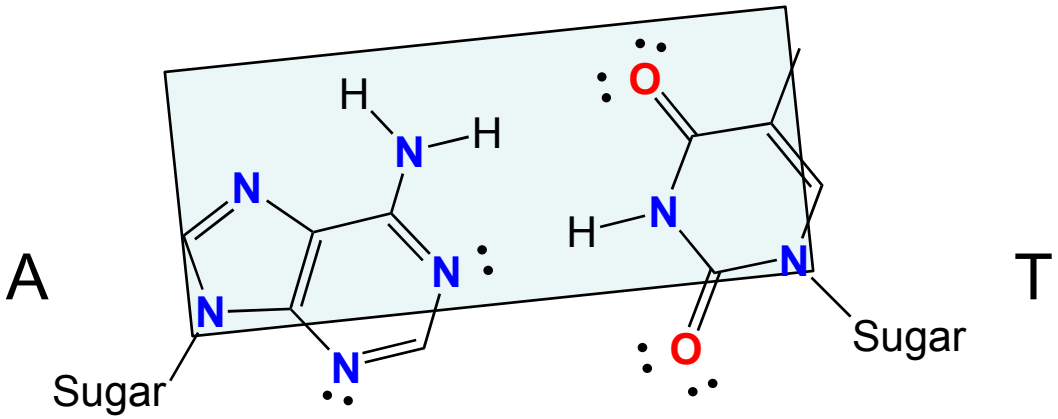

All four WC base pairs
are
***isosteric***

# Why is Watson-Crick so good?



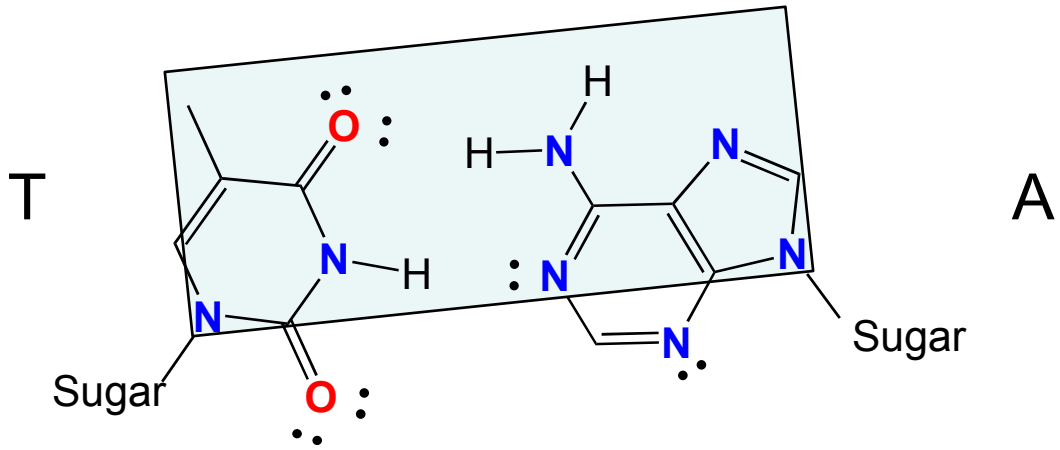

All four WC base pairs
are
***isosteric***

# 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***

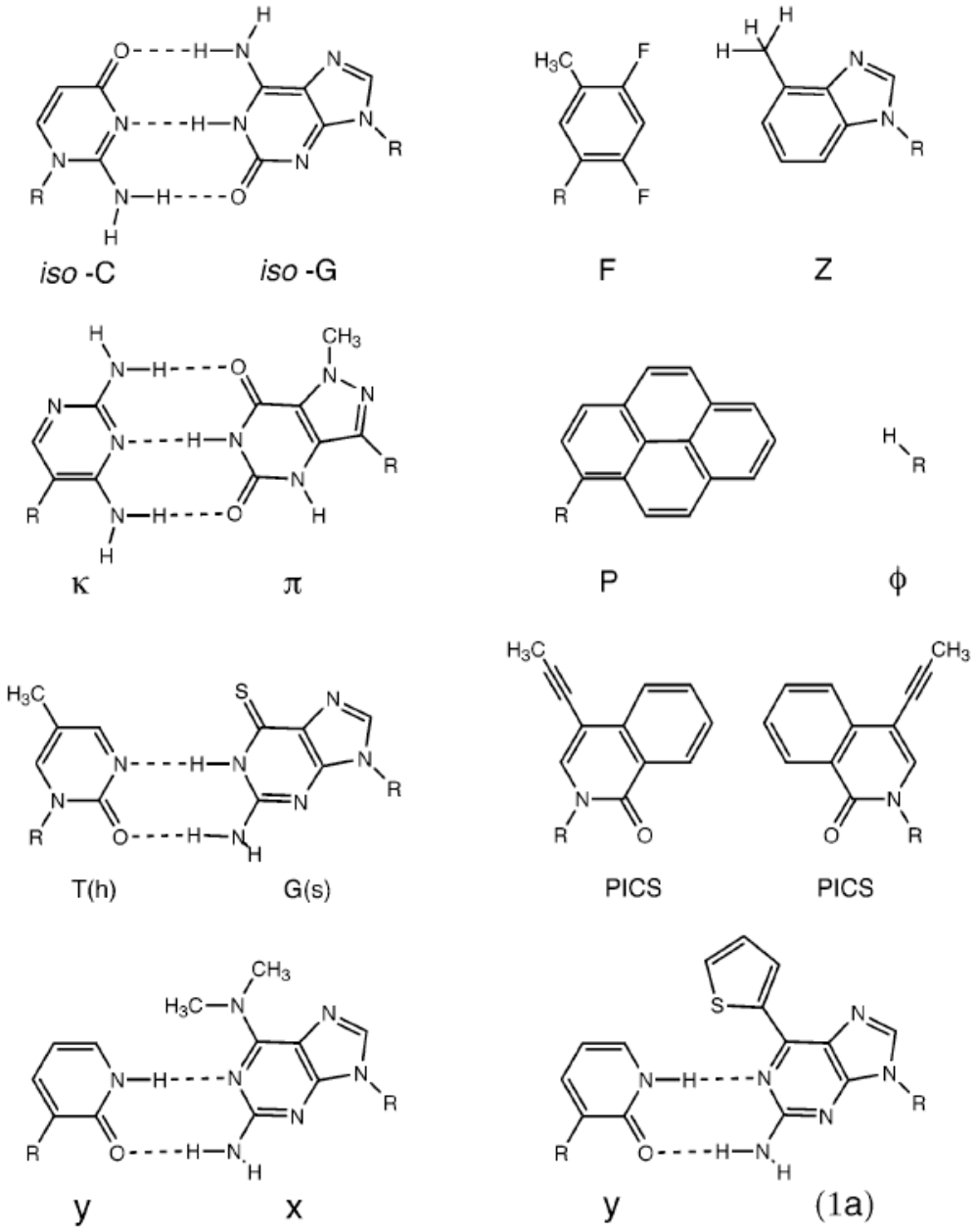
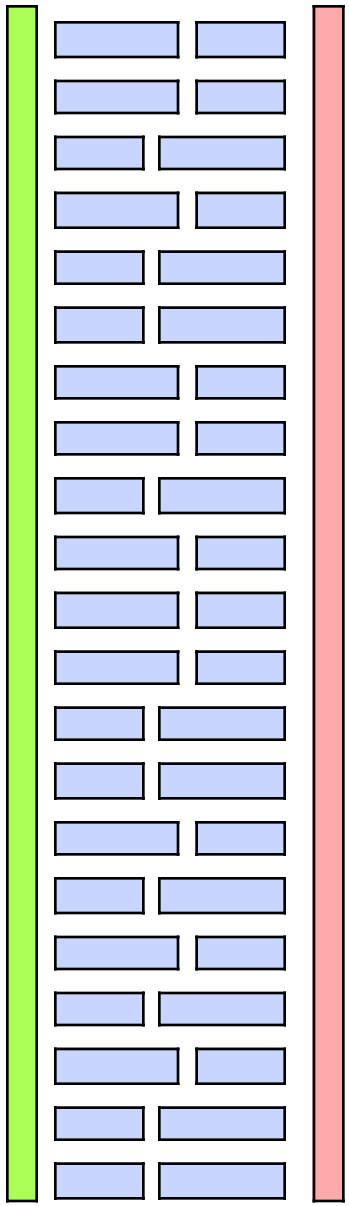
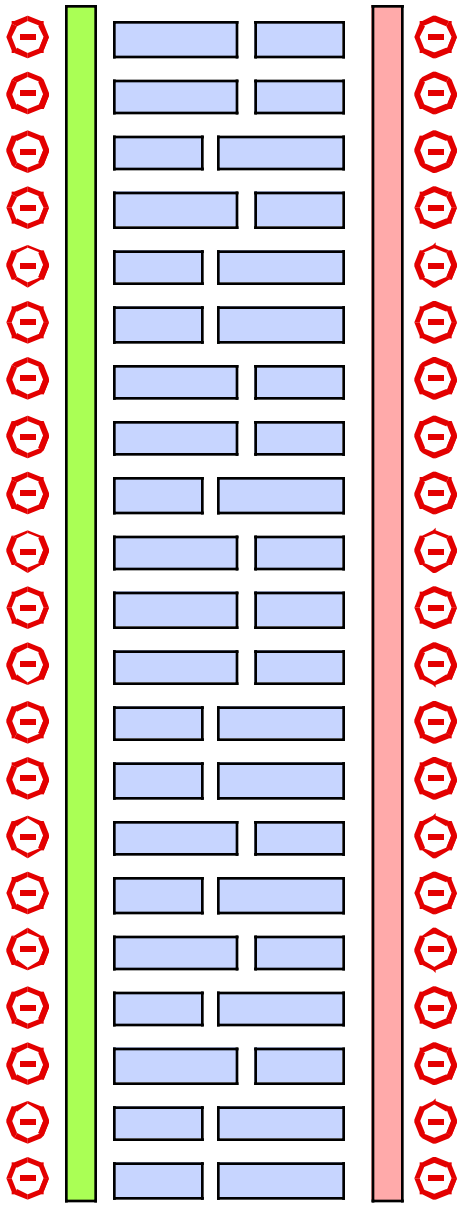



**FIGURE 7.** Examples of DNA base replacements designed to form stable pairs and/or to be replicated selectively by DNA polymerase enzymes.[17,23,24,49,54,55]

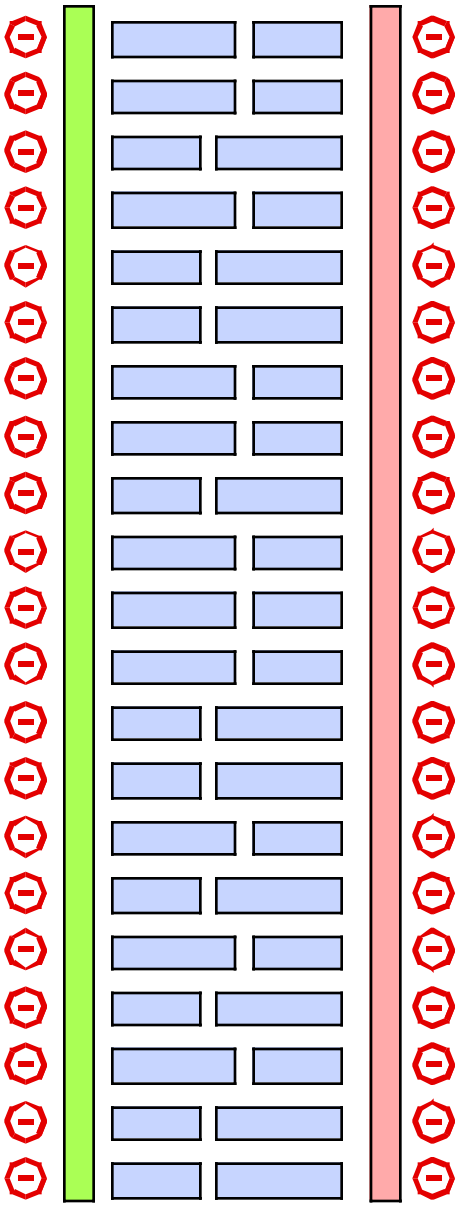
# Bases that polymerases like

# Why ***a helix***?

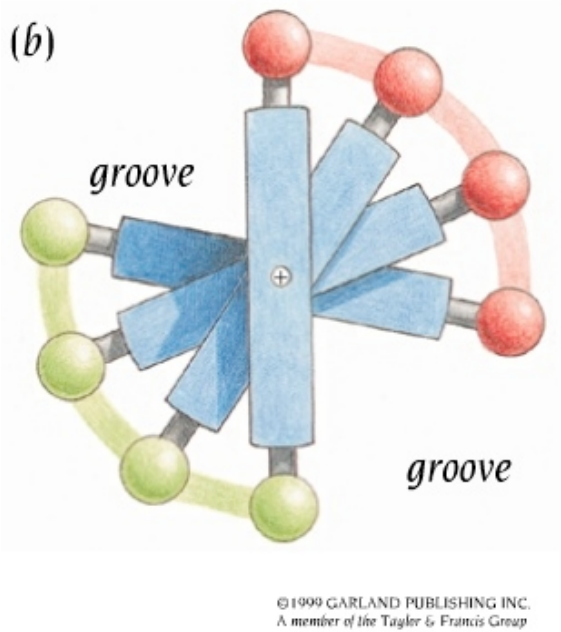




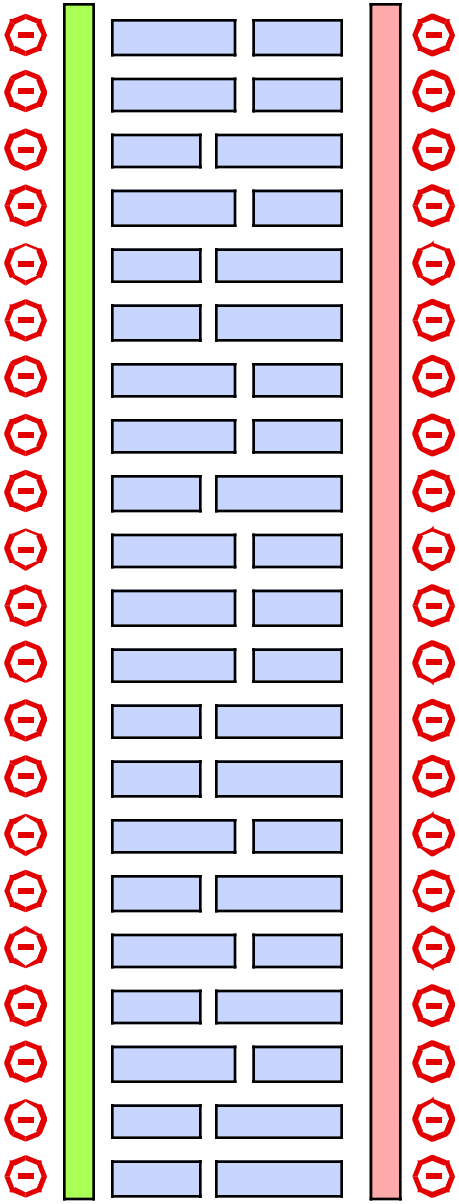
# Why ***a helix***?

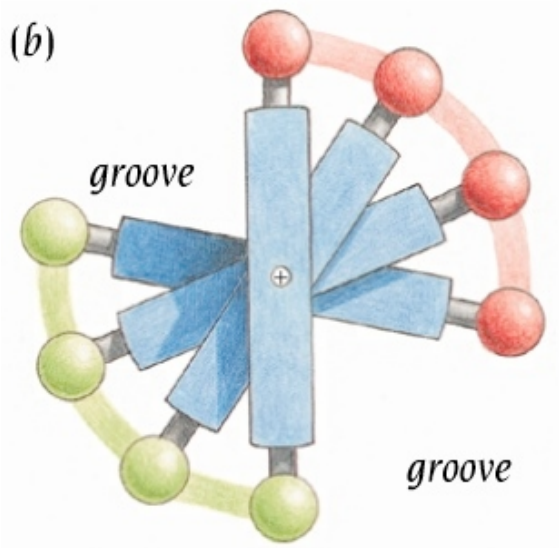


# Why ***a helix***?

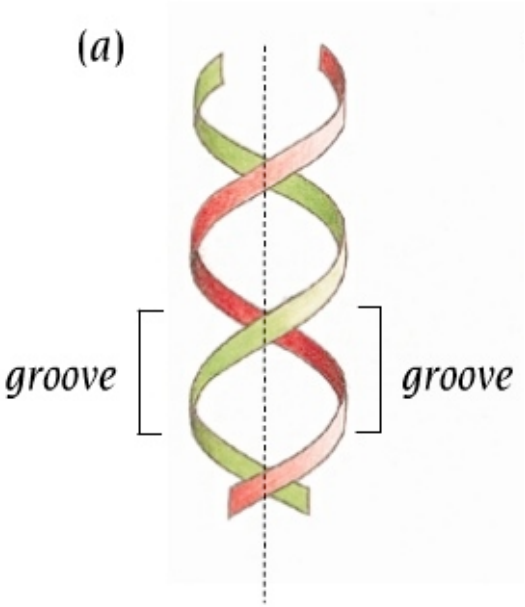


# Why ***a helix***?

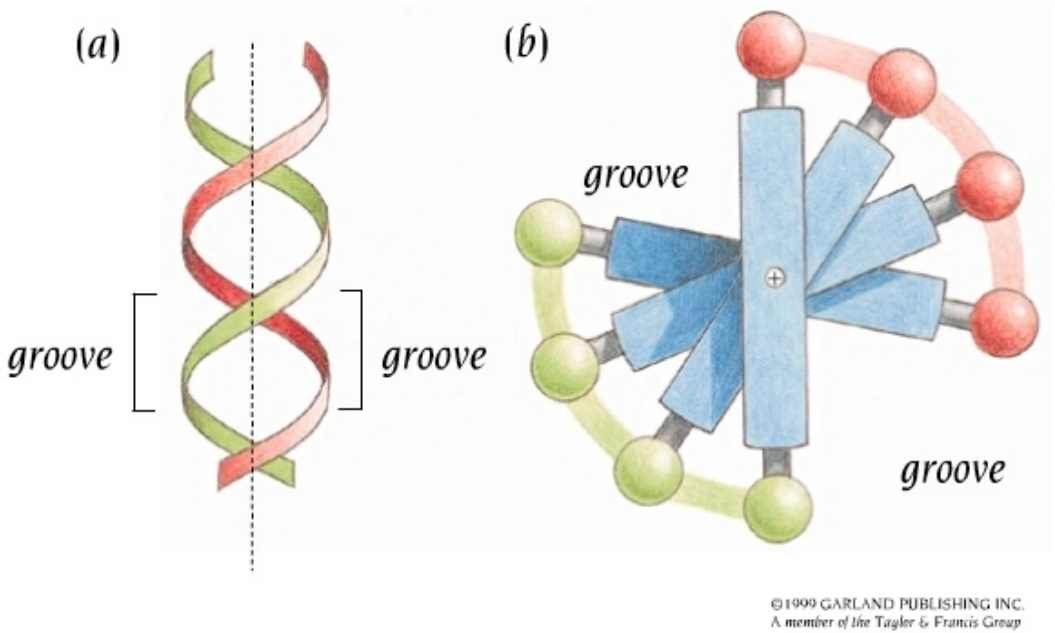




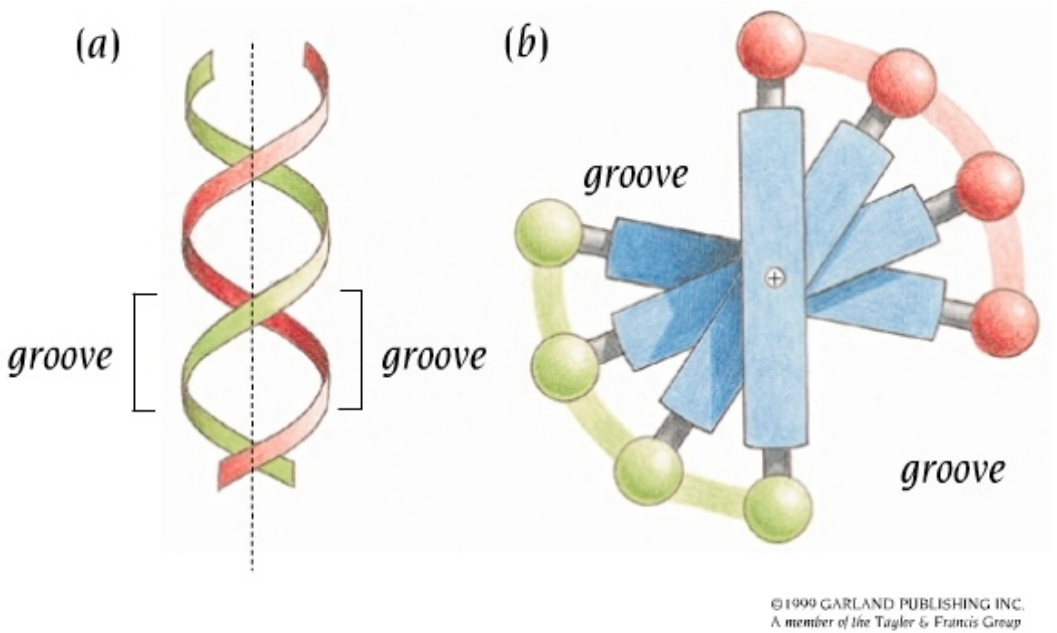


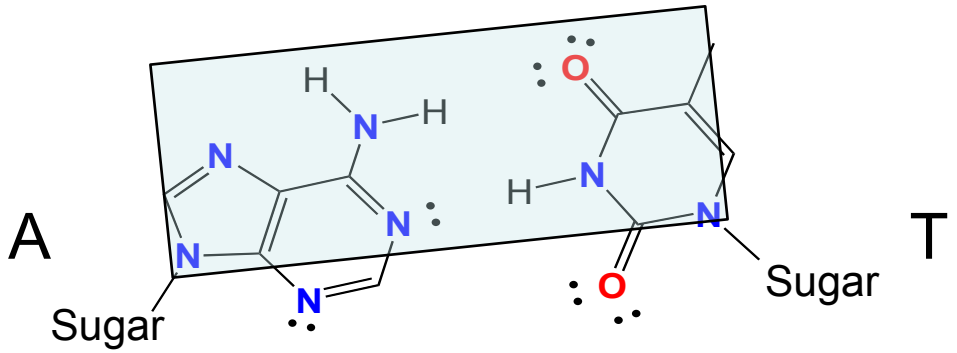
# Why ***major*** and ***minor*** grooves?



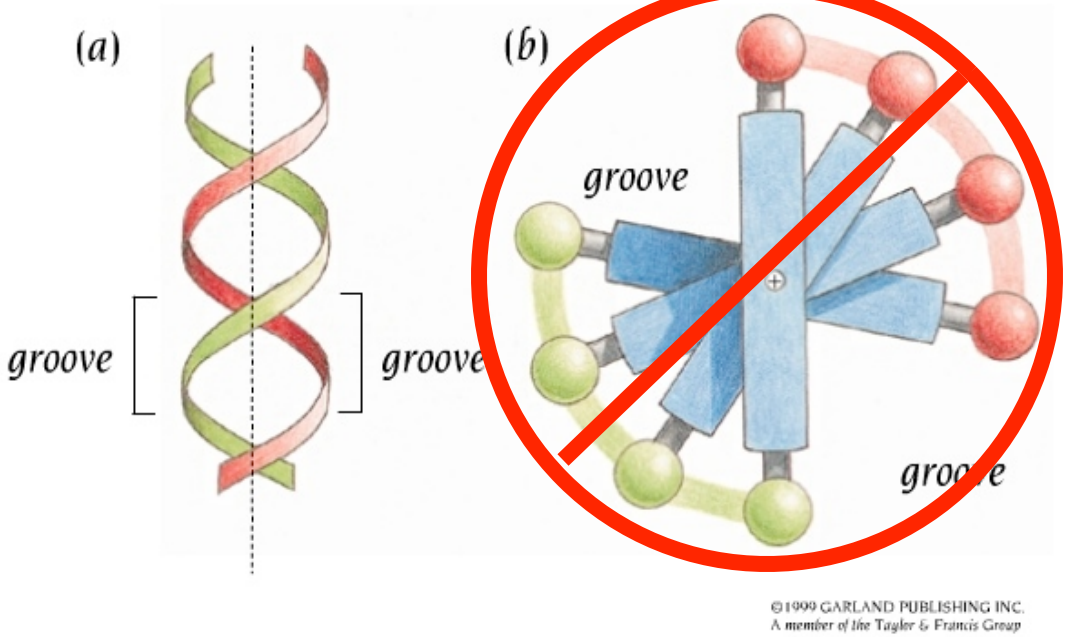
# Why ***major*** and ***minor*** grooves?

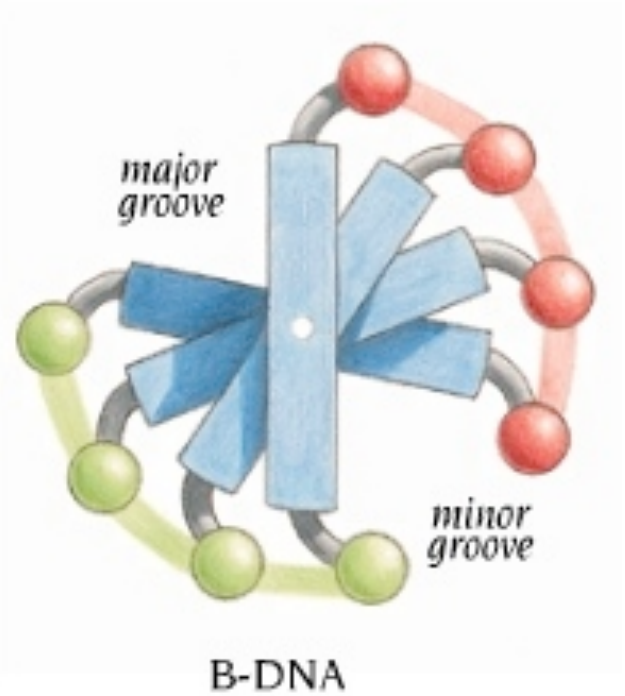


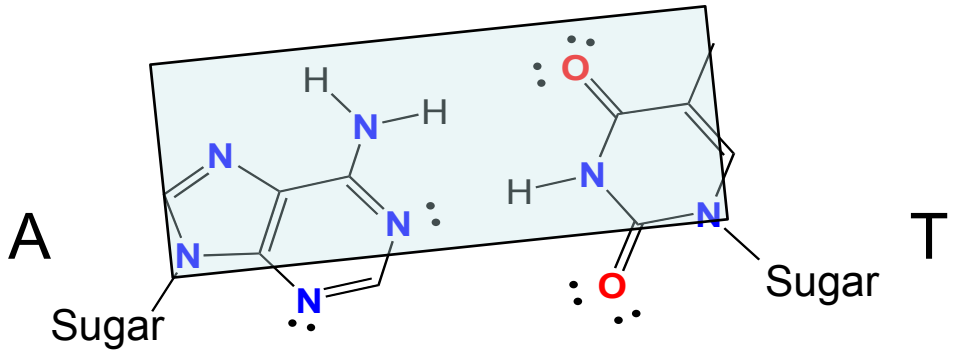


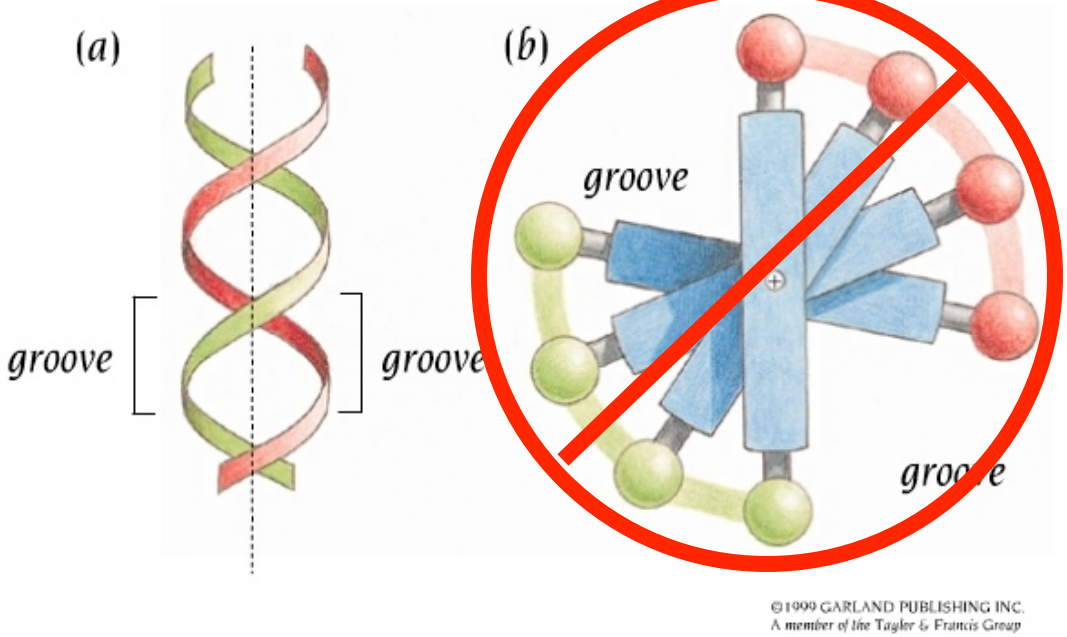
# Why ***major*** and ***minor*** grooves?

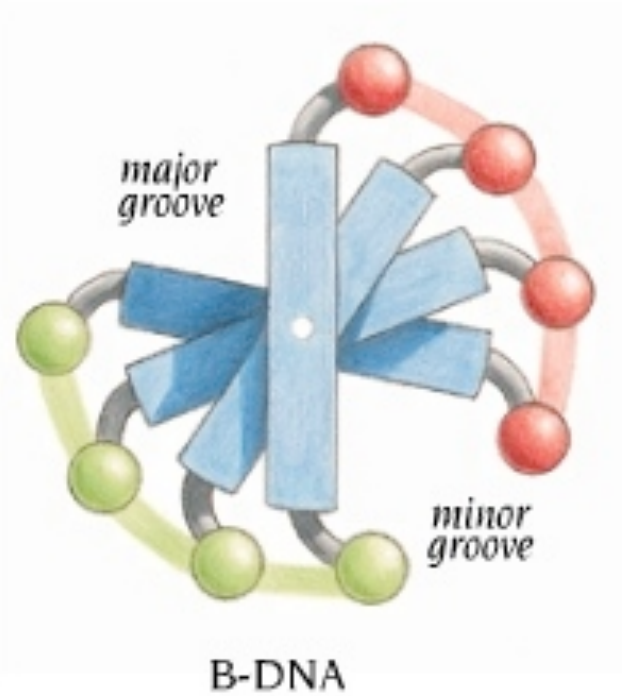


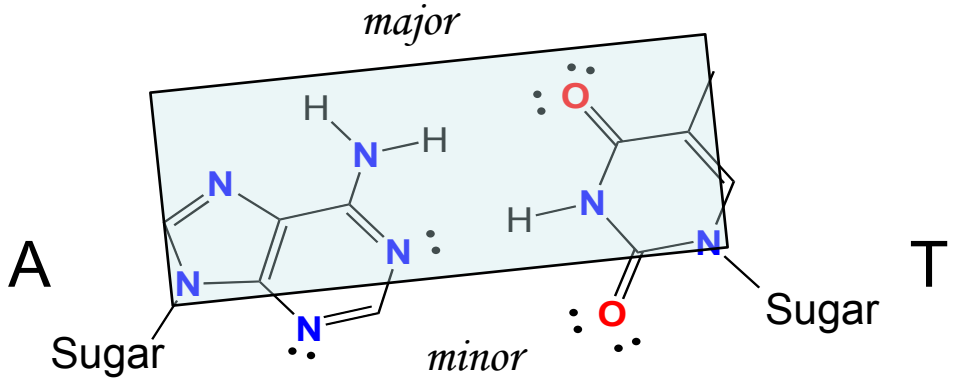



# Why ***major*** and ***minor*** grooves?



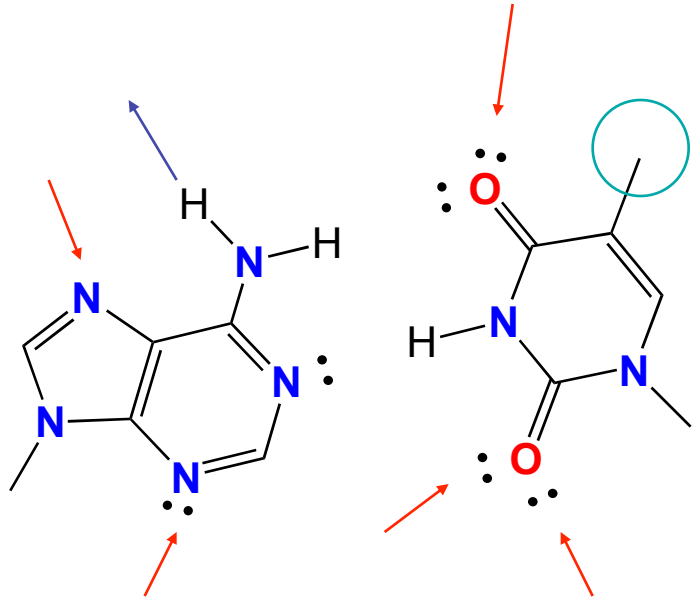



# Why is the major groove so good?

*Major Groove*

A



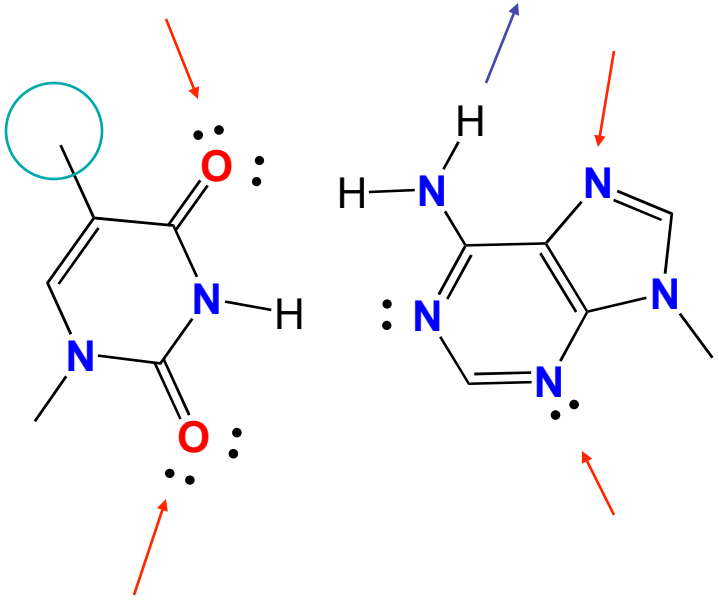

T

*Minor Groove*

# Why is the major groove so good?

*Major Groove*

T



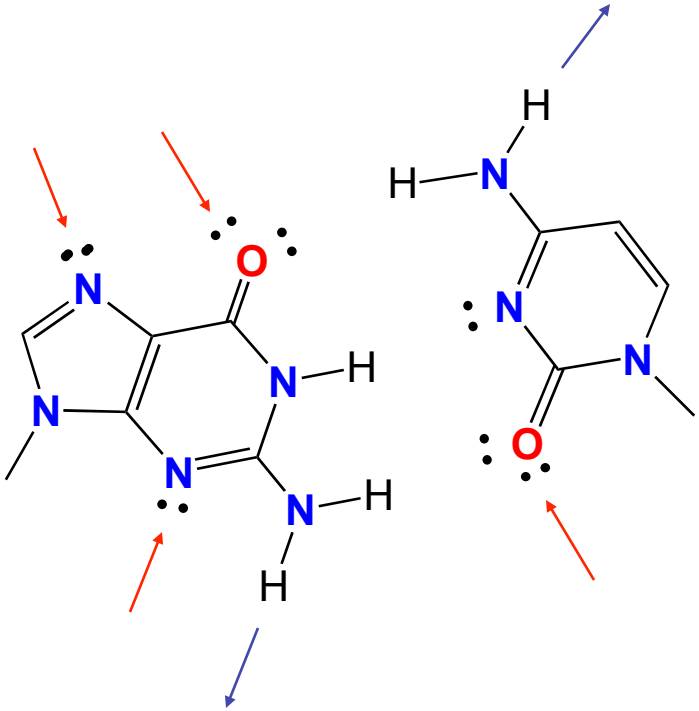

A

*Minor Groove*

# Why is the major groove so good?

*Major Groove*

G

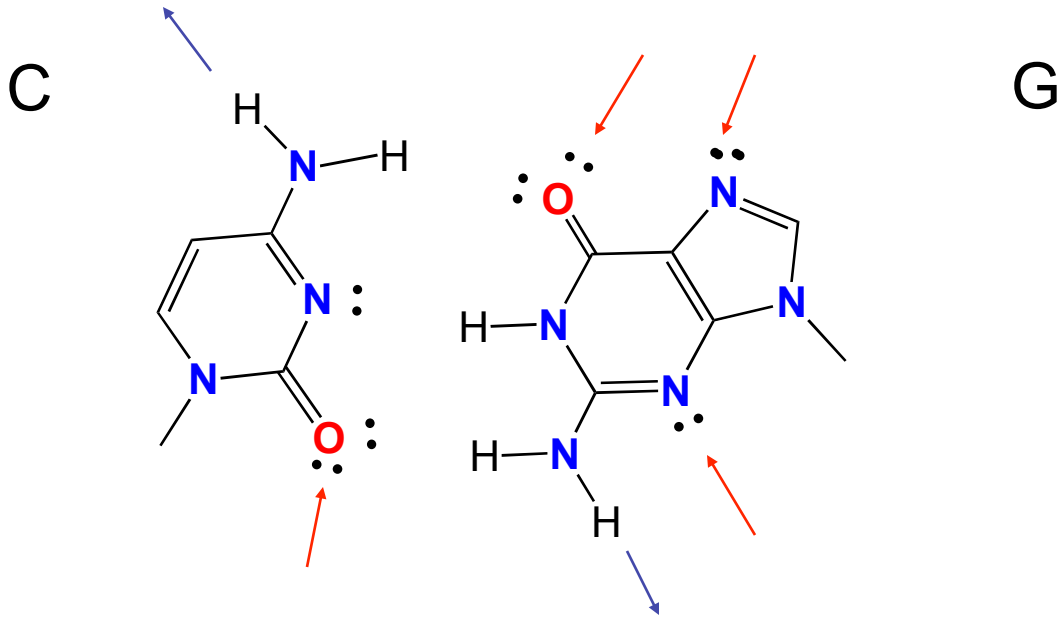



C

*Minor Groove*

# Why is the major groove so good?

*Major Groove*




*Minor Groove*

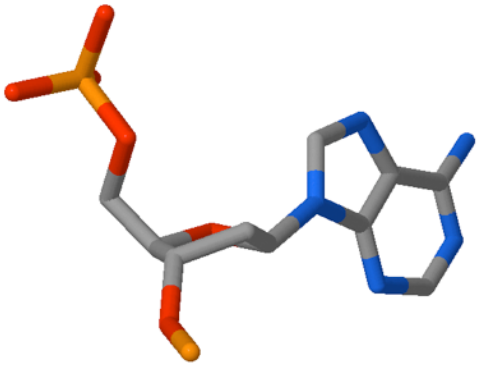
# B-form DNA

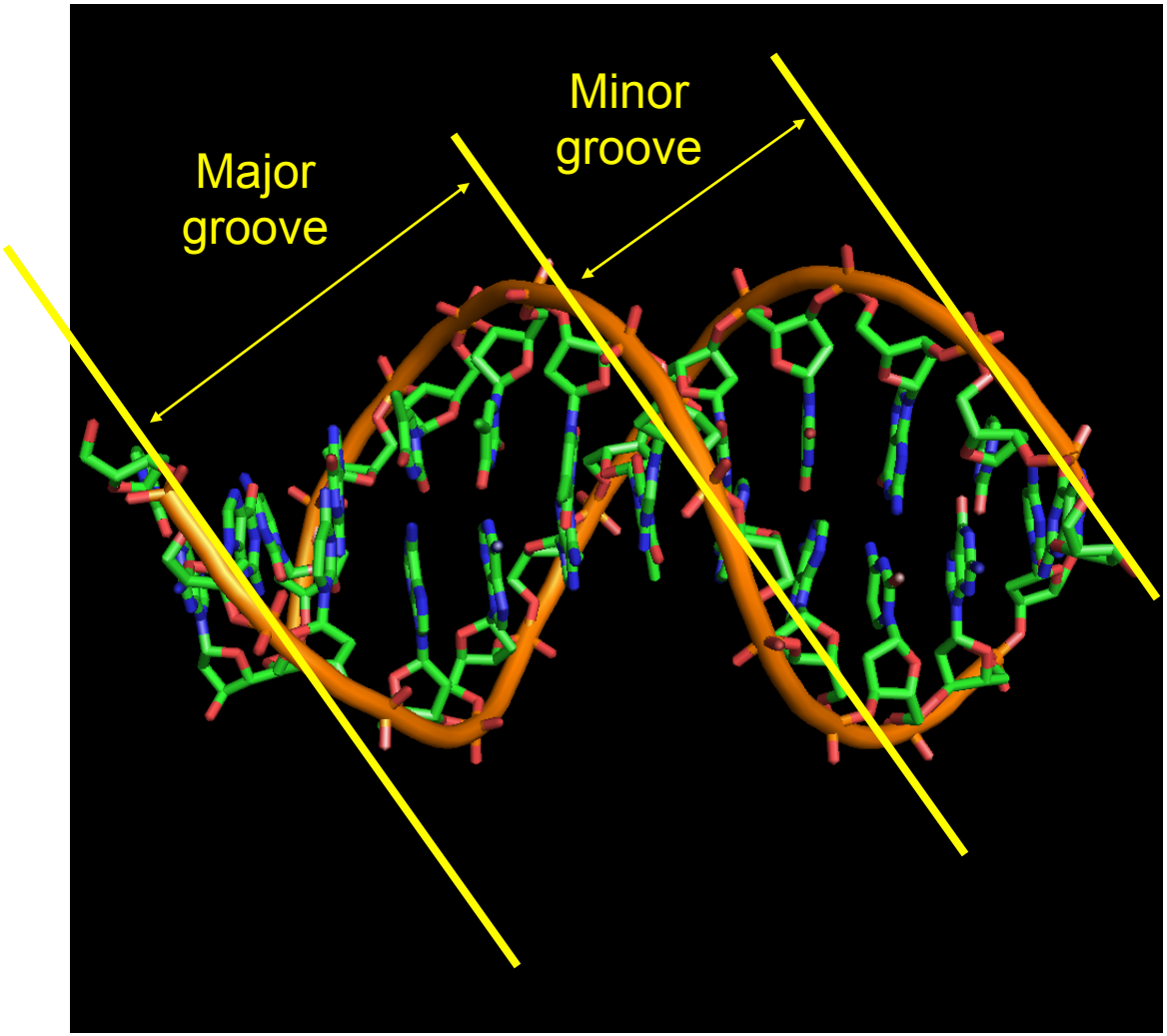
B-form

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

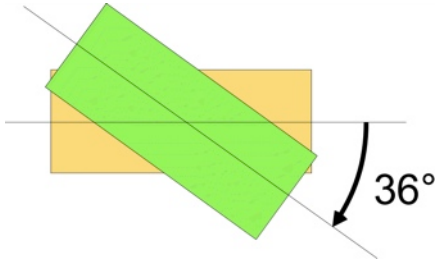
Minor groove depth = 7.5Å
Major groove depth = 8.8Å





# B-form DNA

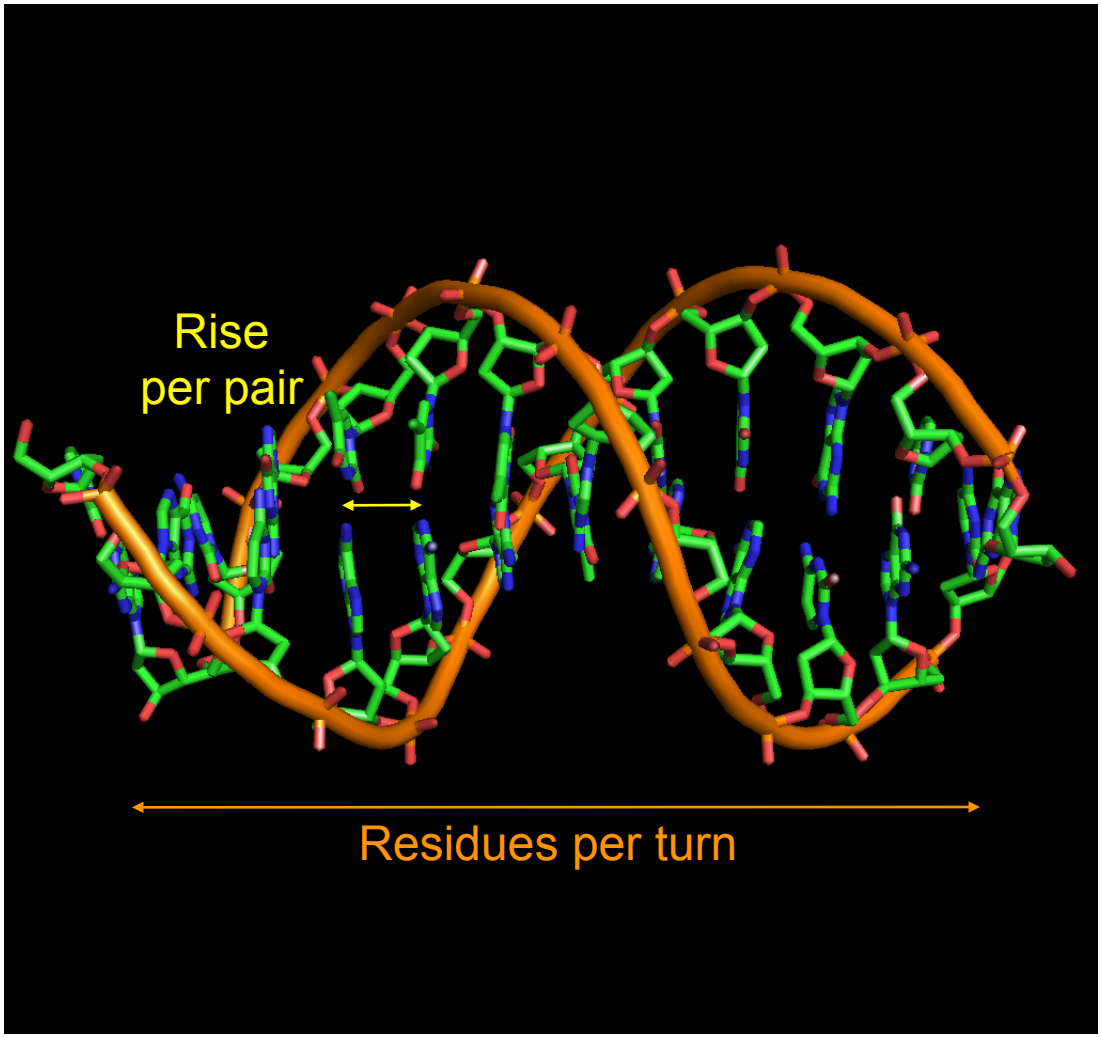
Residues per turn =10
Twist per base pair = 36°




Rise per pair = 3.4Å
c2'-endo

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

Minor groove depth = 7.5Å
Major groove depth = 8.8Å

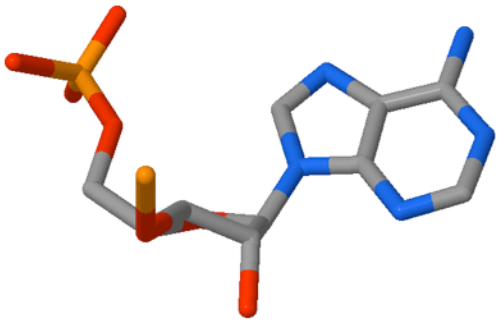


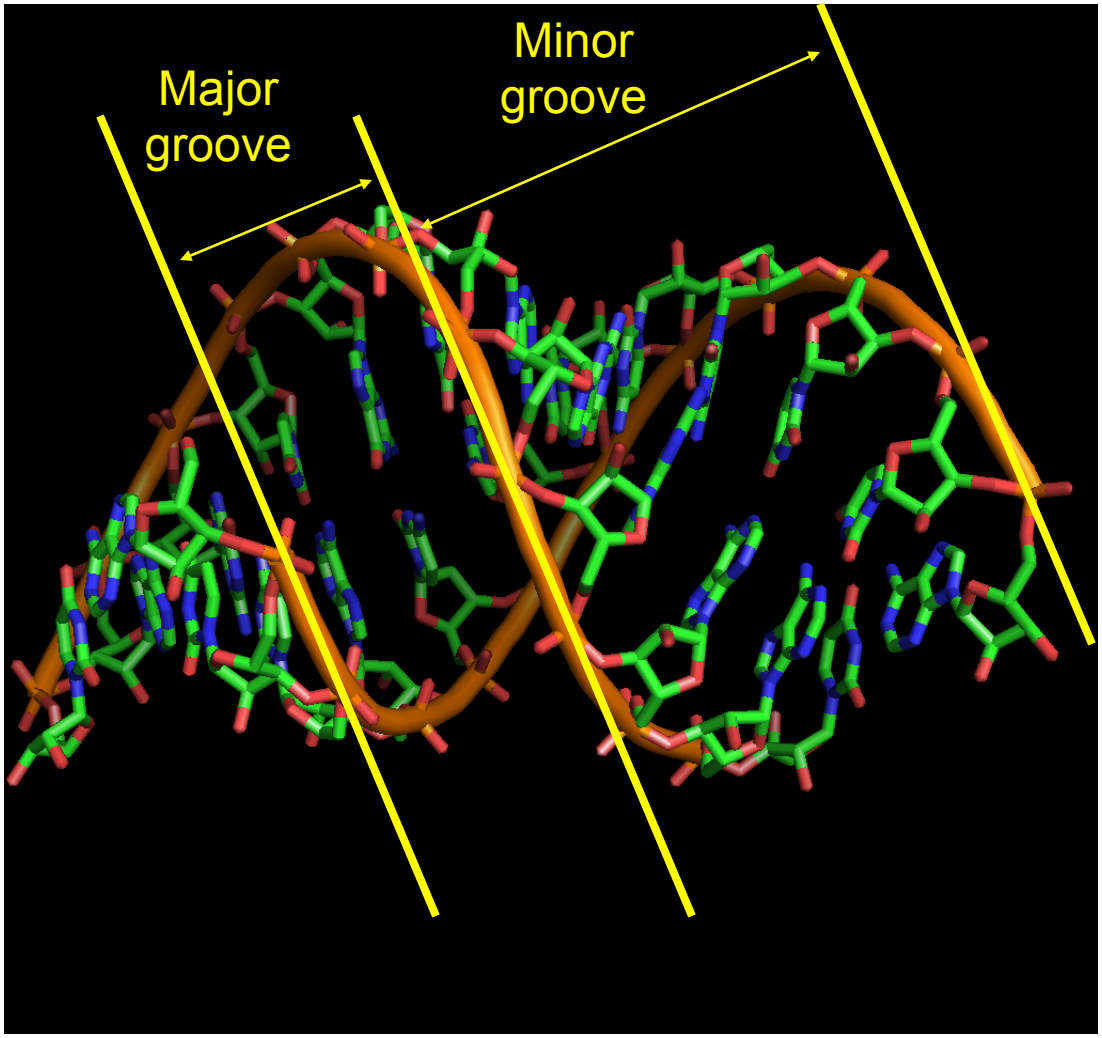
# A-form RNA

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

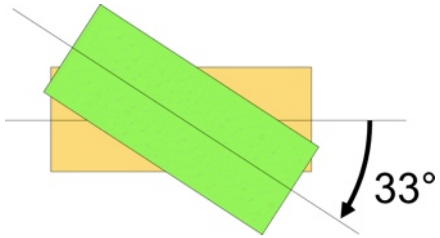
Minor groove depth = 2.8Å
Major groove depth = 13.5Å





# A-form RNA

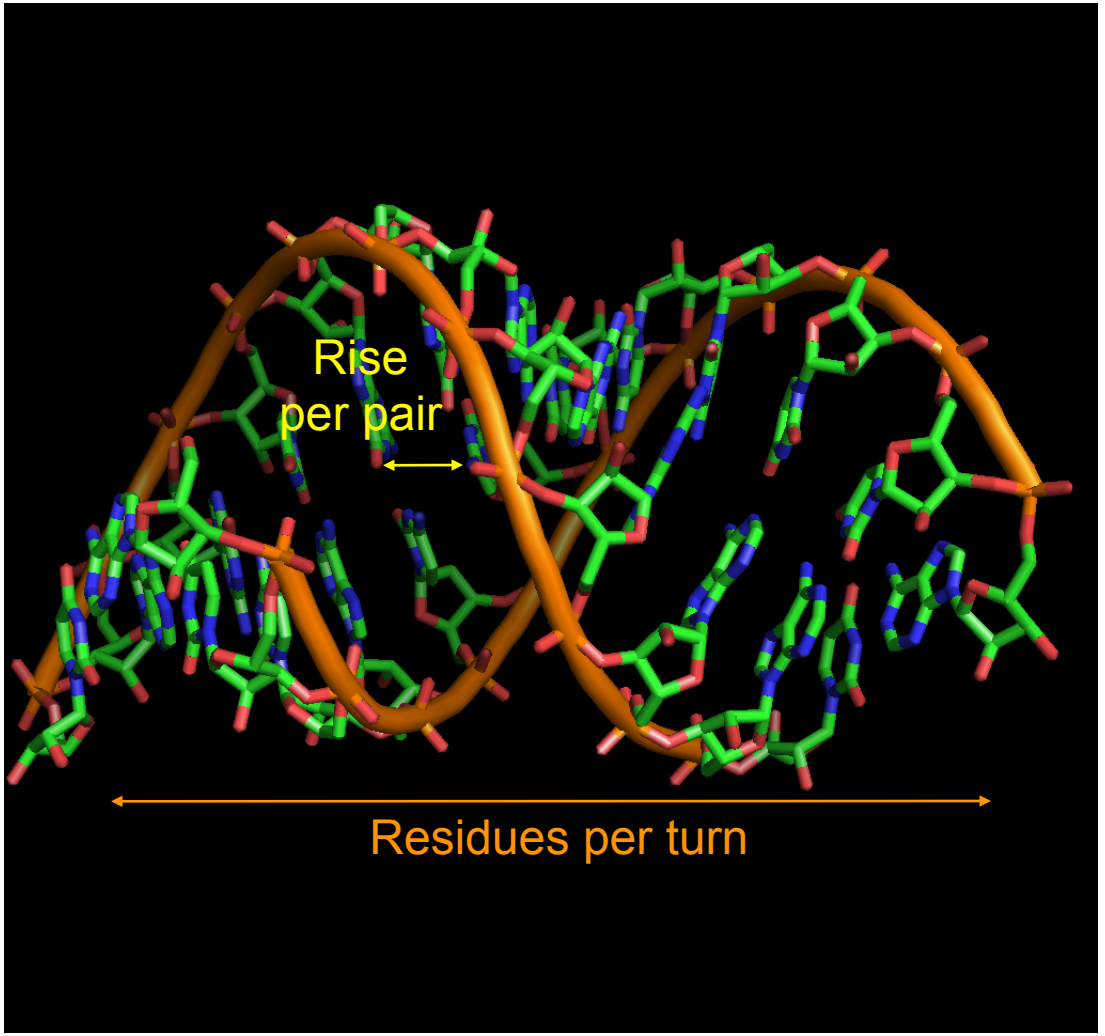
Residues per turn =11
Twist per base pair = 33°



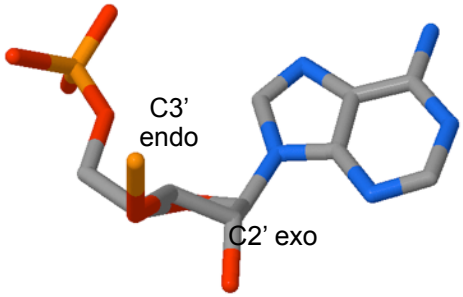

Rise per pair = 2.9Å
c3'-endo

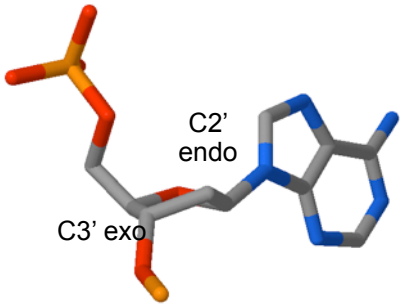
Minor groove width = 11Å
Major groove width = 2.7Å

Minor groove depth = 2.8Å
Major groove depth = 13.5Å



# Compare




## A-form (RNA)

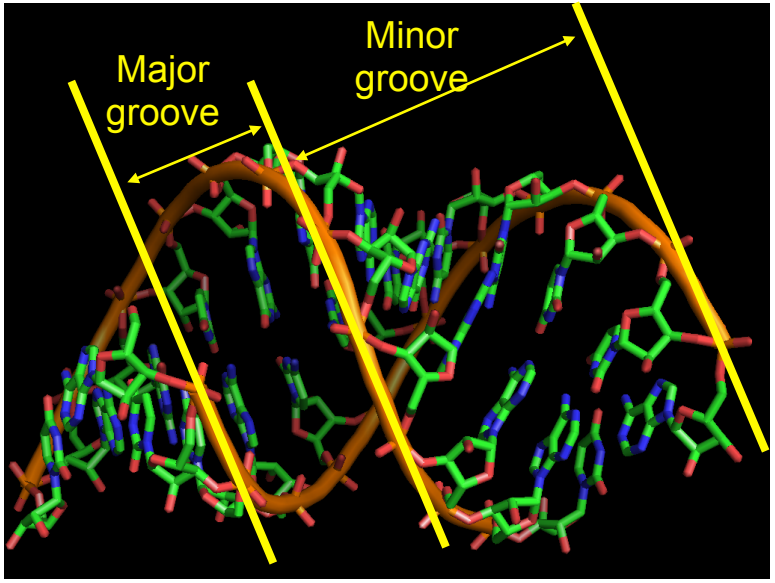
Minor groove width = 11Å
Major groove width = 2.7Å

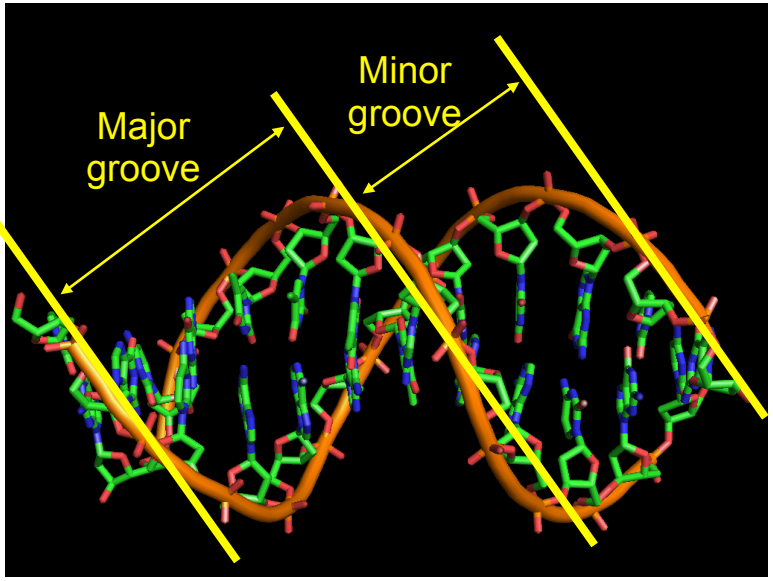
Minor groove depth = 2.8Å
Major groove depth = 13.5Å

## B-form (DNA)

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

Minor groove depth = 7.5Å
Major groove depth = 8.8Å



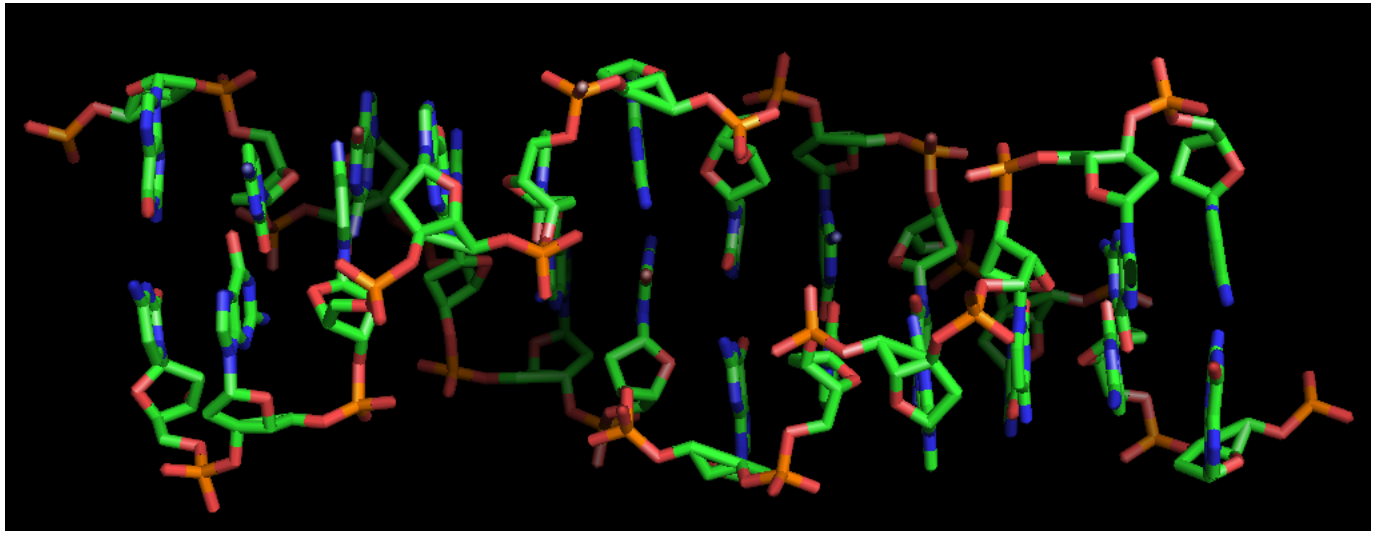
# Z-DNA

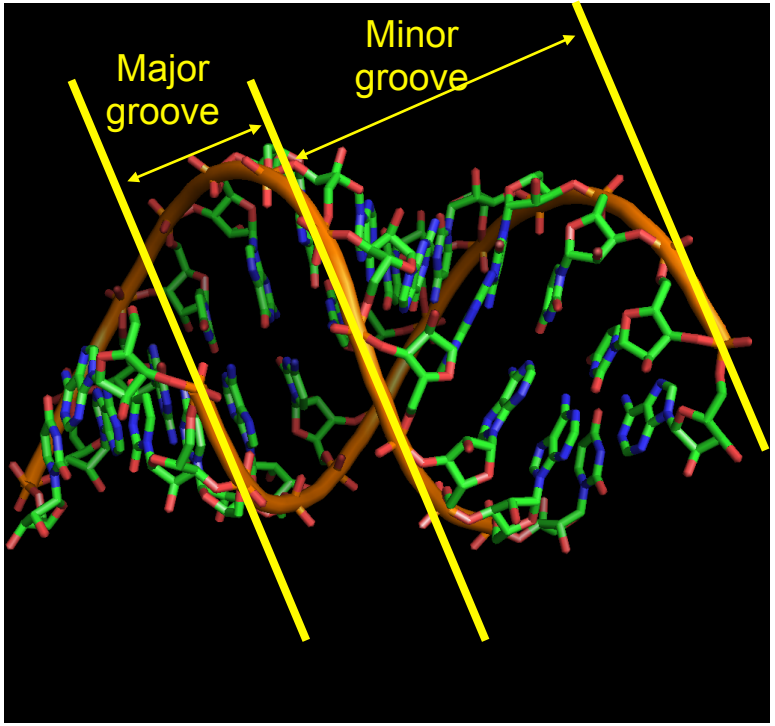
Residues per turn =12
Twist per base pair = -9 / -51°

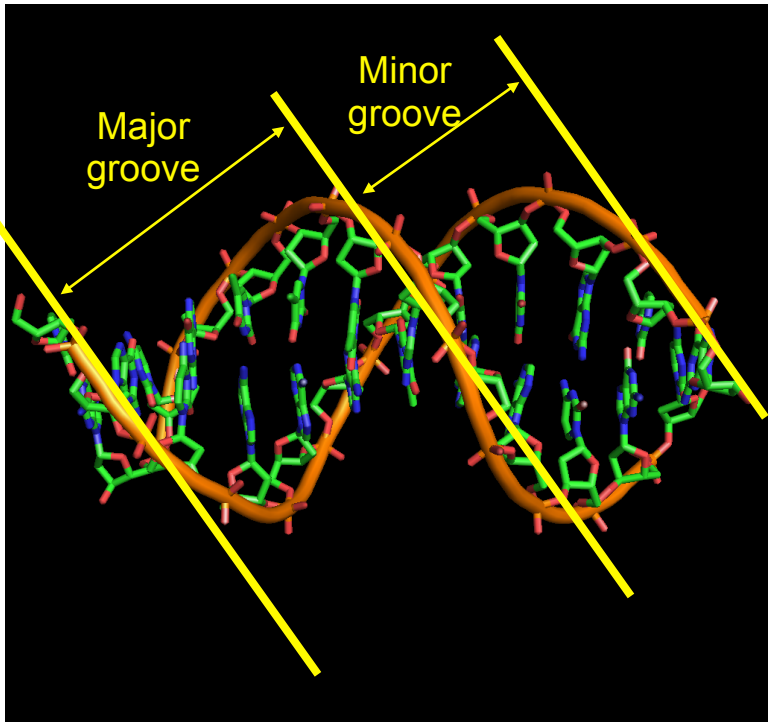
Rise per pair = 3.7Å
c3'-endo(syn) / c2'-endo

Minor groove width = 2.0Å
Major groove width = 8.8Å

Minor groove depth = 13.8Å
Major groove depth = 3.7Å

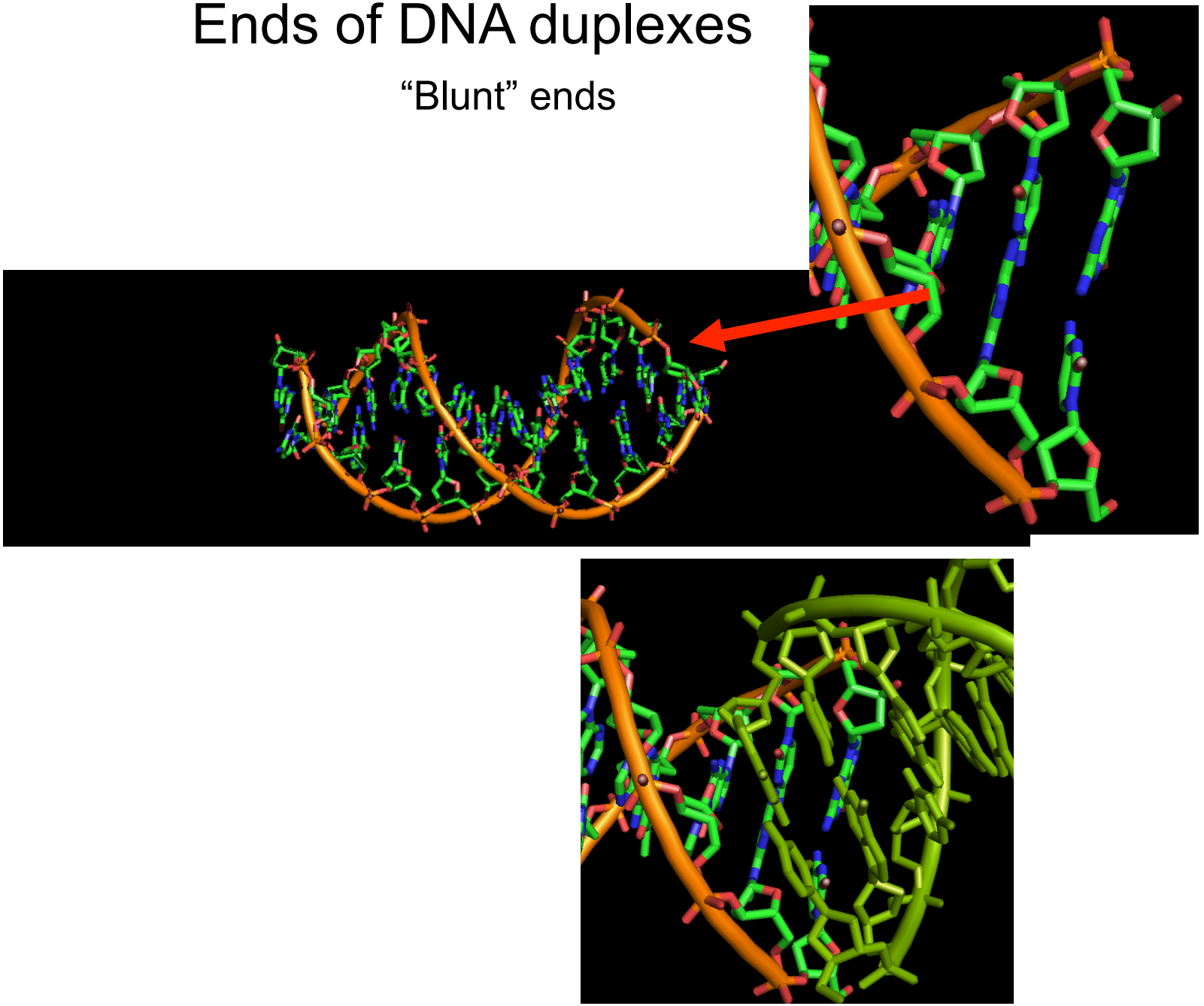




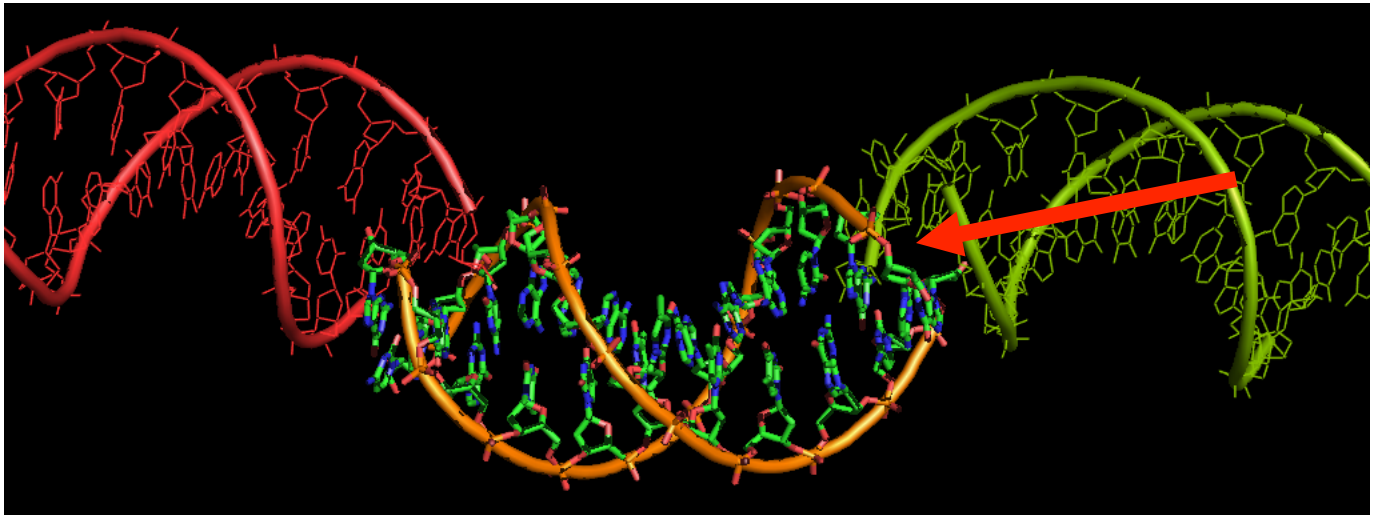
# Ends of DNA duplexes

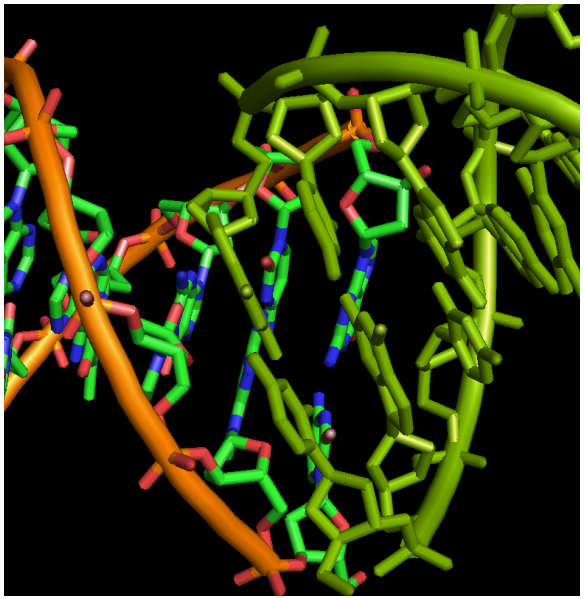
"Blunt" ends



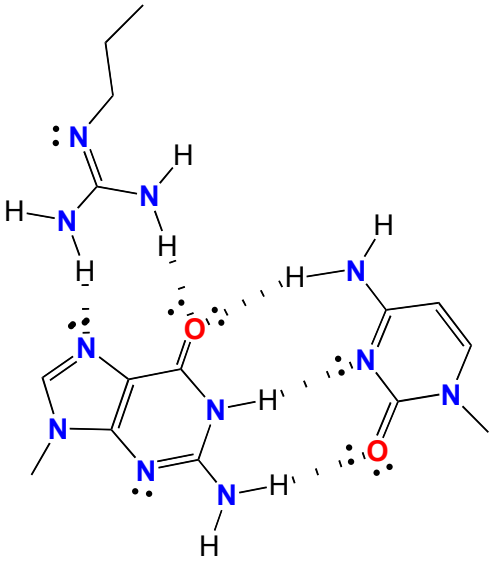
# Ends of DNA duplexes

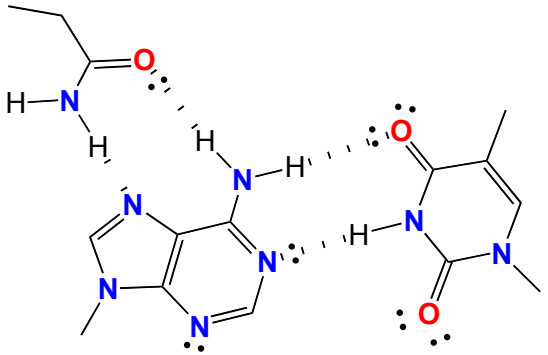
"Blunt" ends



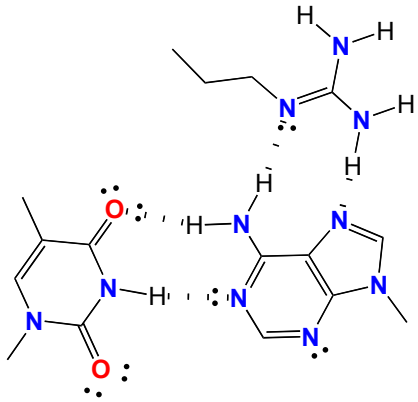


# Protein - Nucleic Acid Interactions





Major Groove Interactions

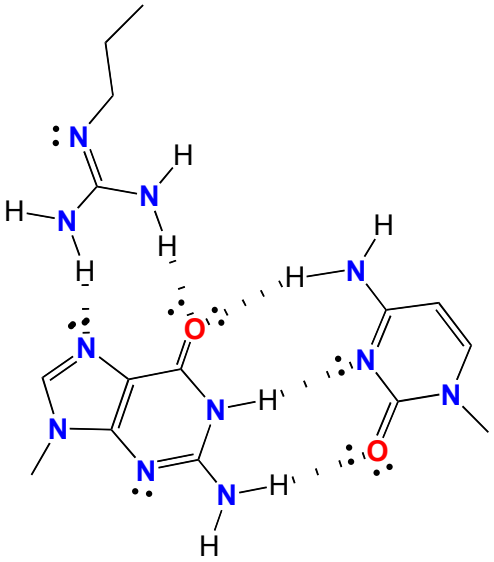


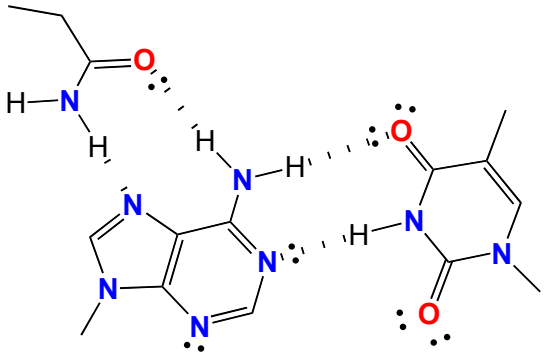
# Protein - Nucleic Acid Interactions

Gln

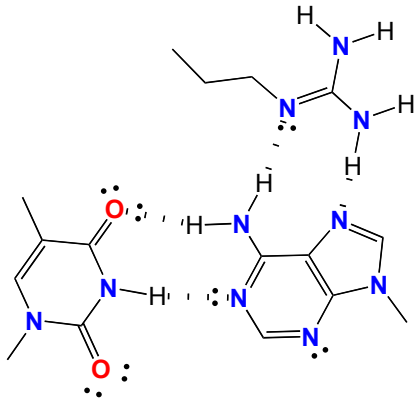
Asn

Arg

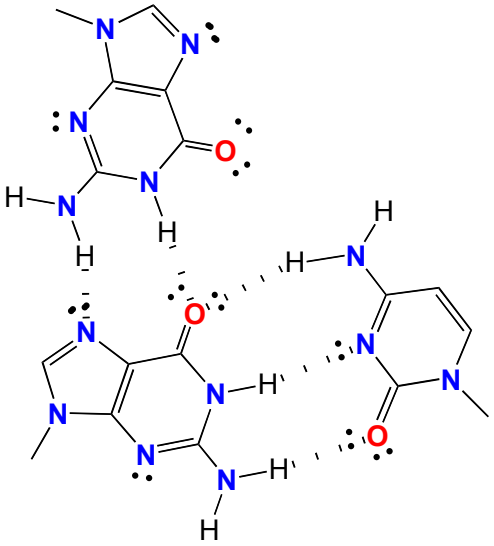


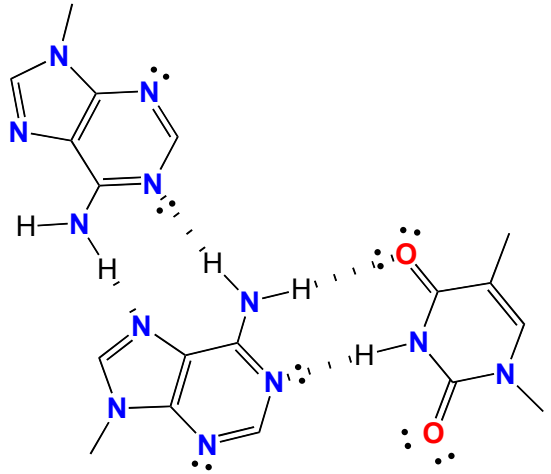


Major Groove Interactions

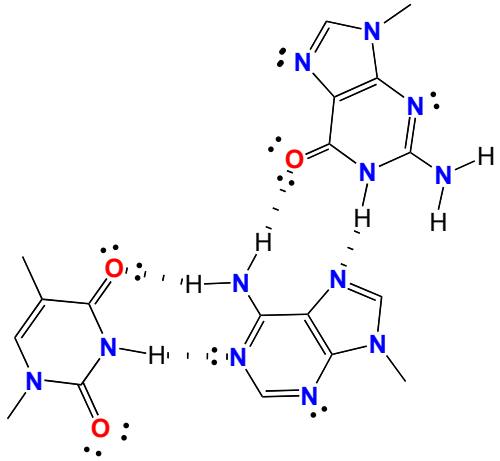


# Nucleic Acid “Triples / Platforms”



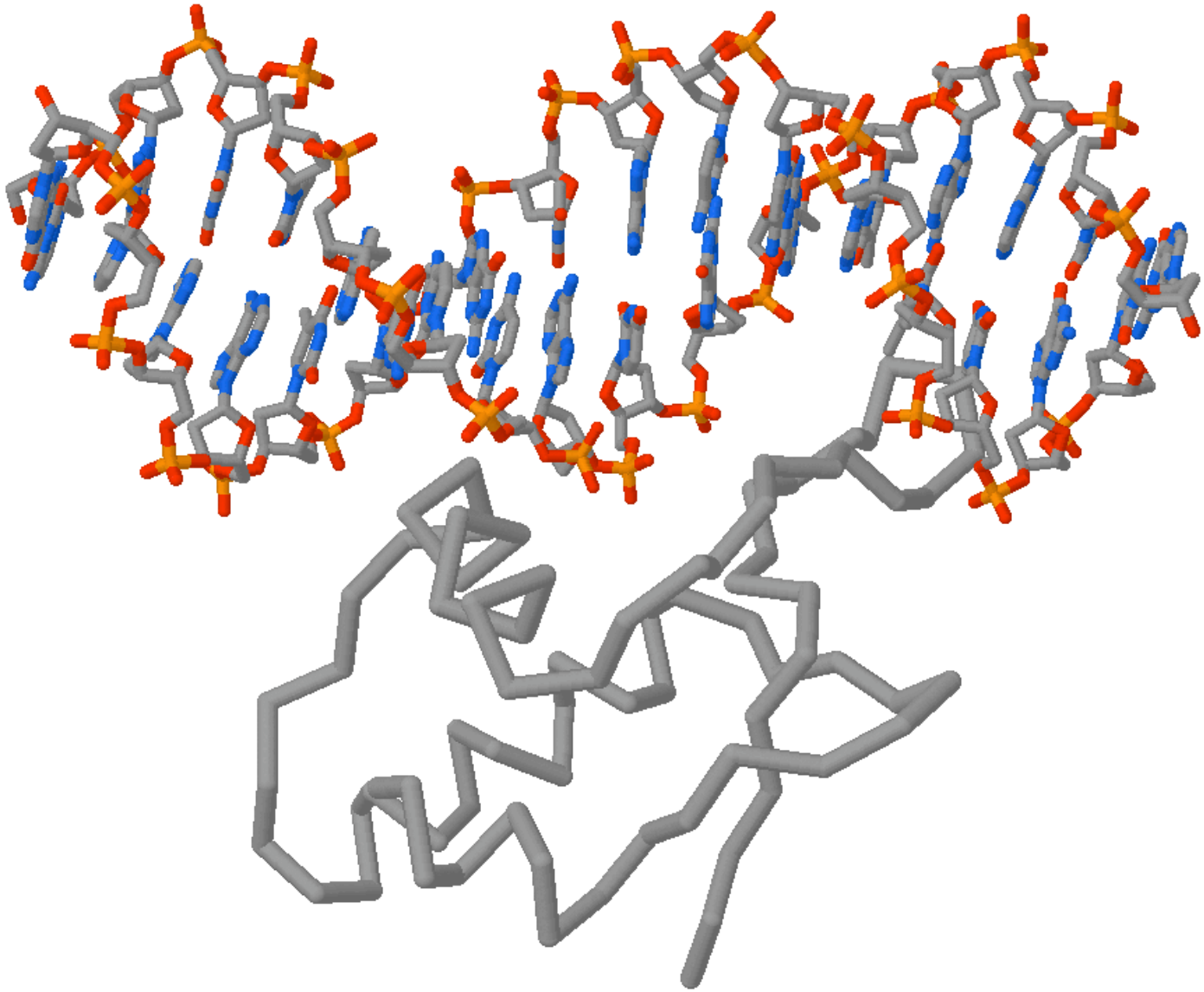


Major Groove Interactions



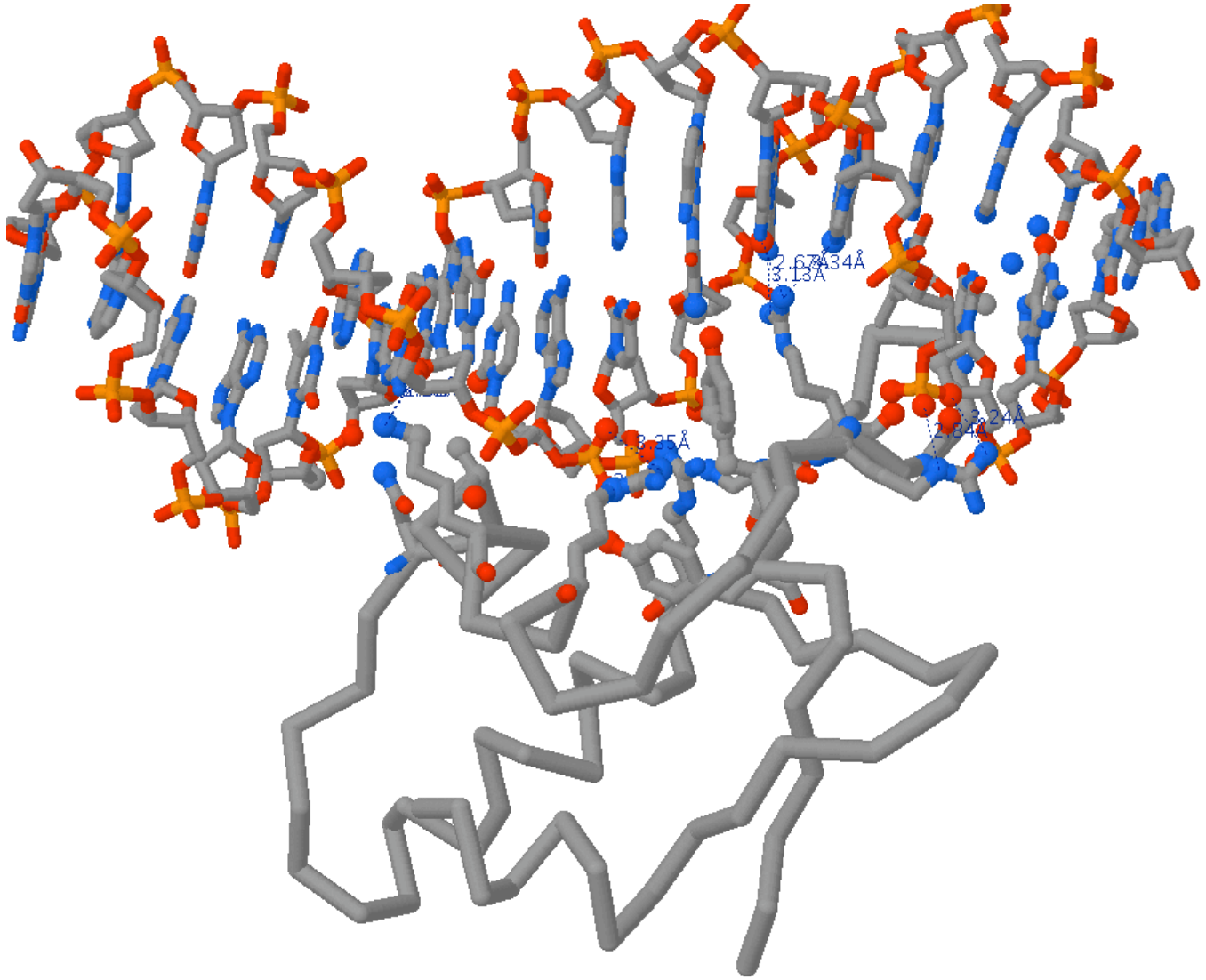
# Winged Helix DNA Binding Domain

Classic helix-turn-helix



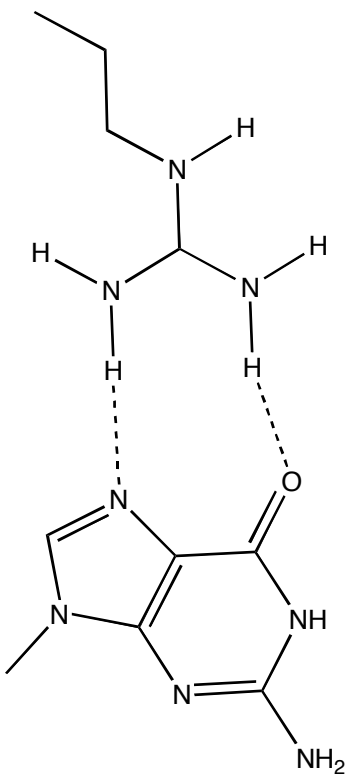
# Winged Helix DNA Binding Domain

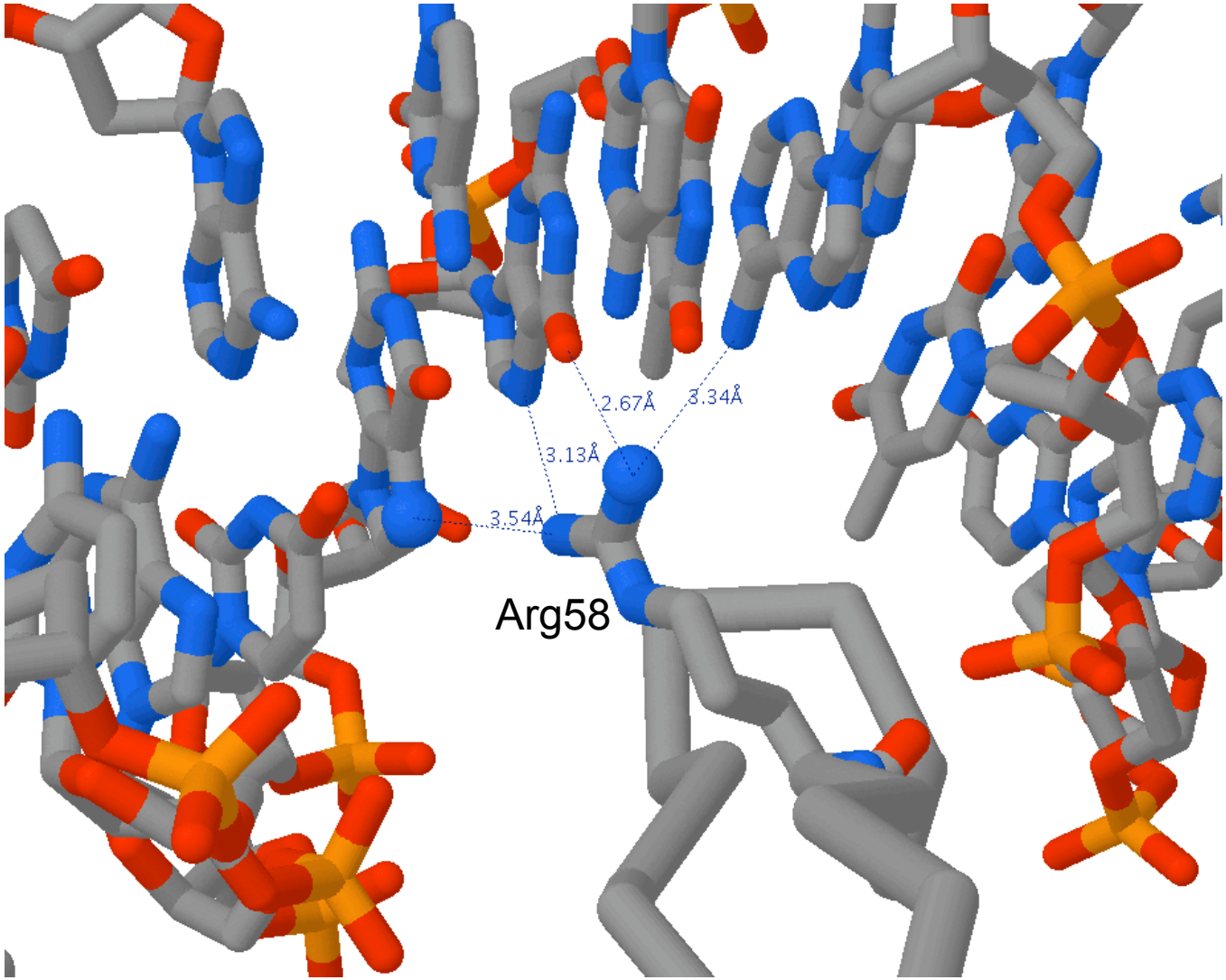
Classic helix-turn-helix



# Winged Helix DNA Binding Domain

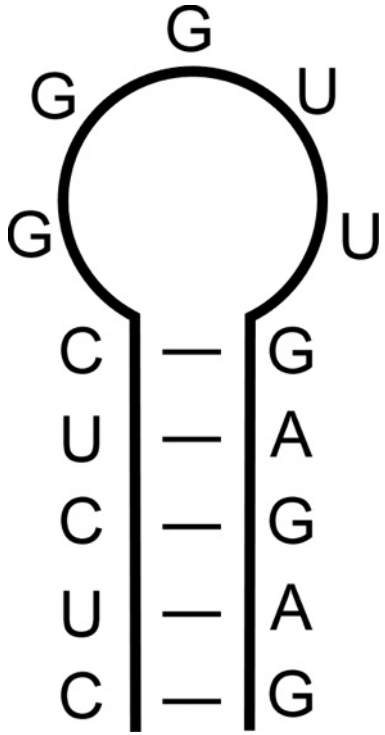
## Classic helix-turn-helix

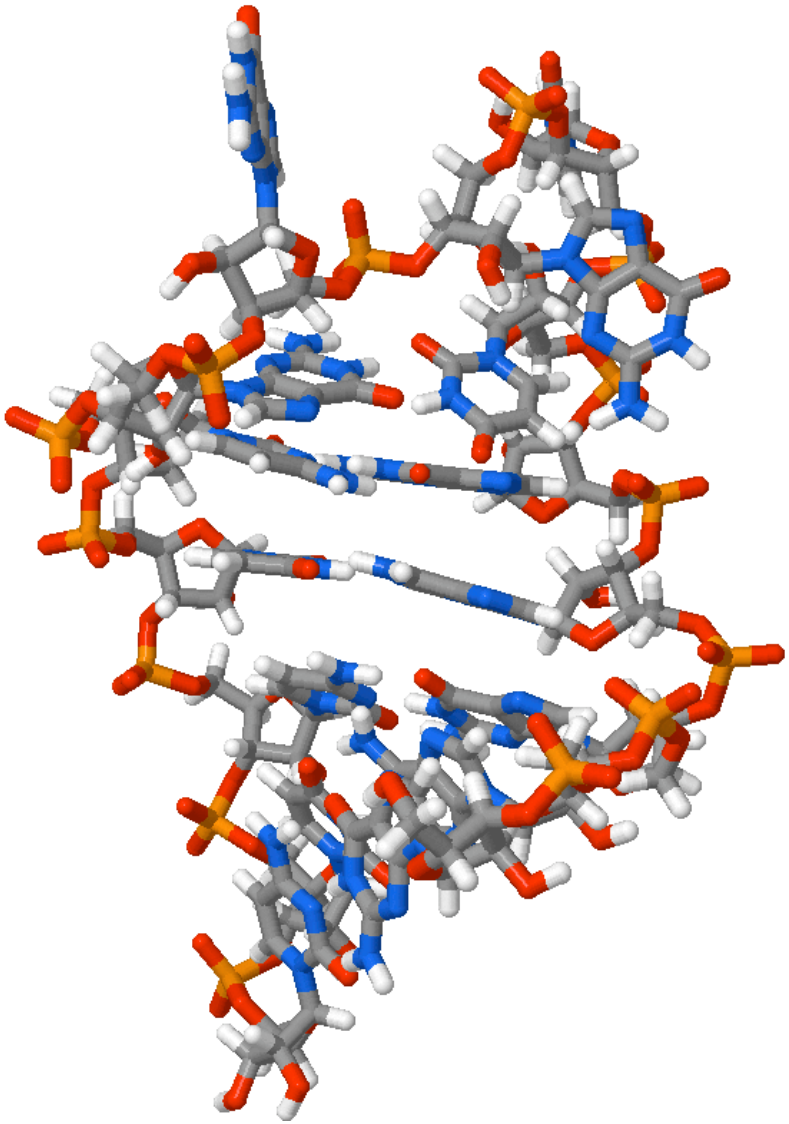


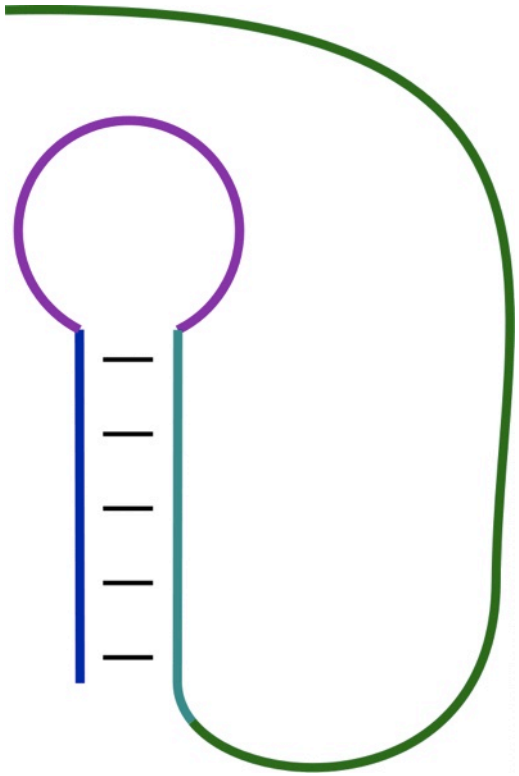

Hrfx1 bound to its X-box binding site

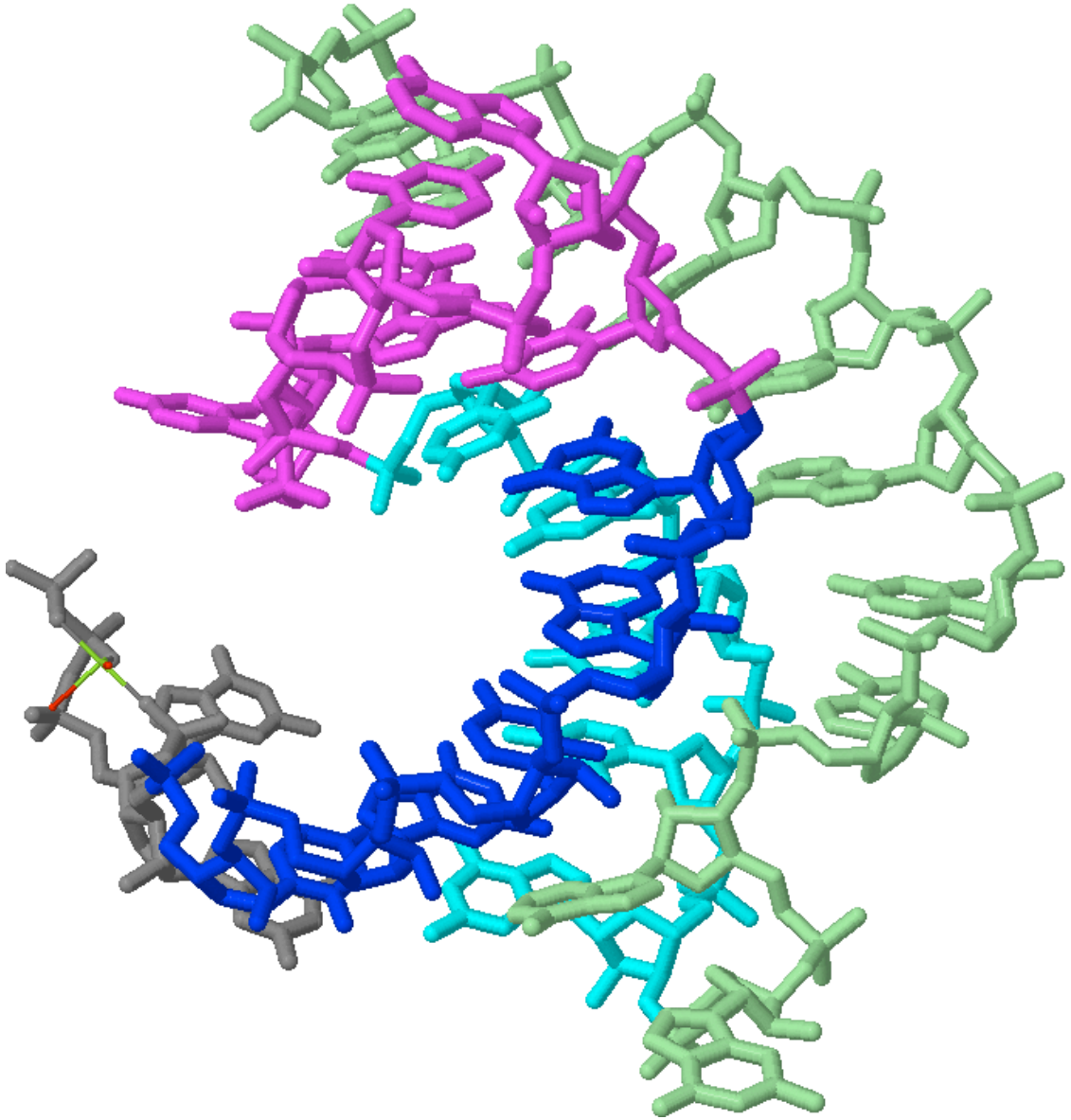
# Simple Structure - Hairpin



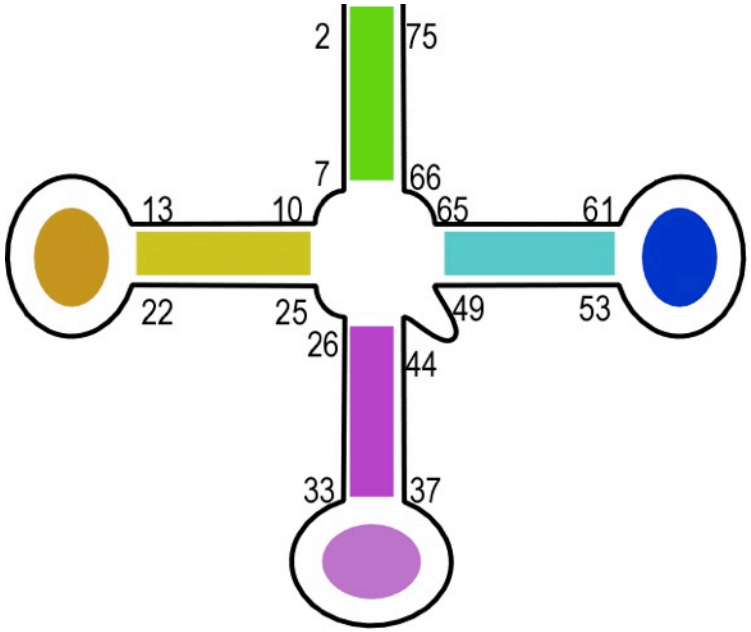



# Classic Structure - Pseudoknot

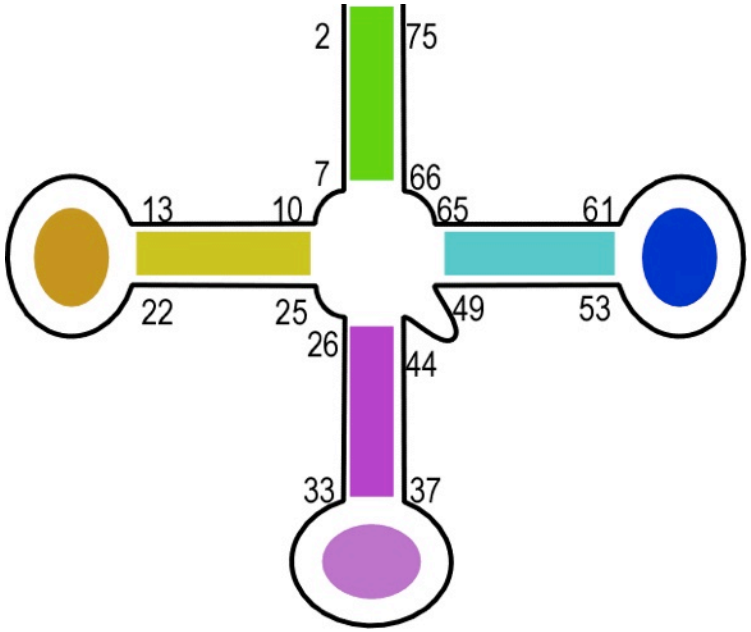


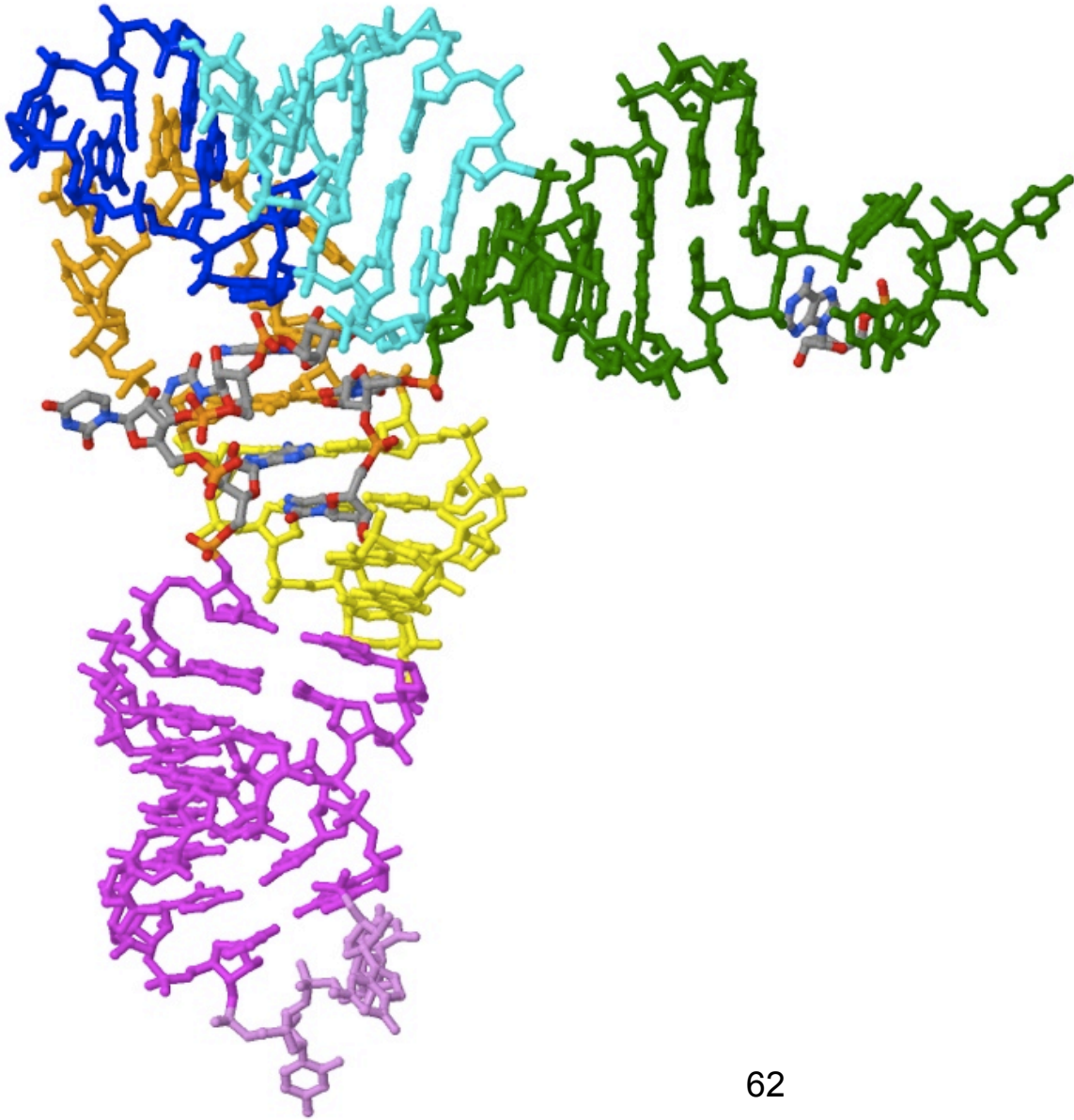


# tRNA

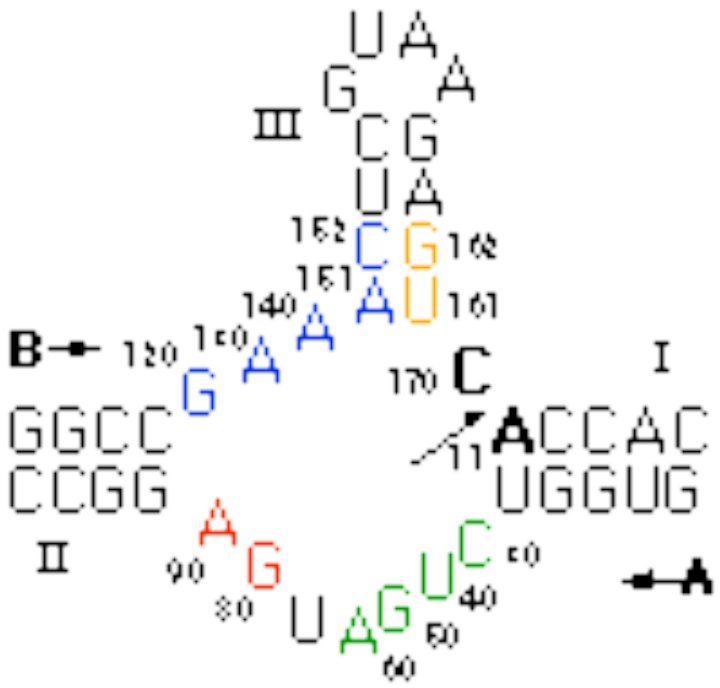


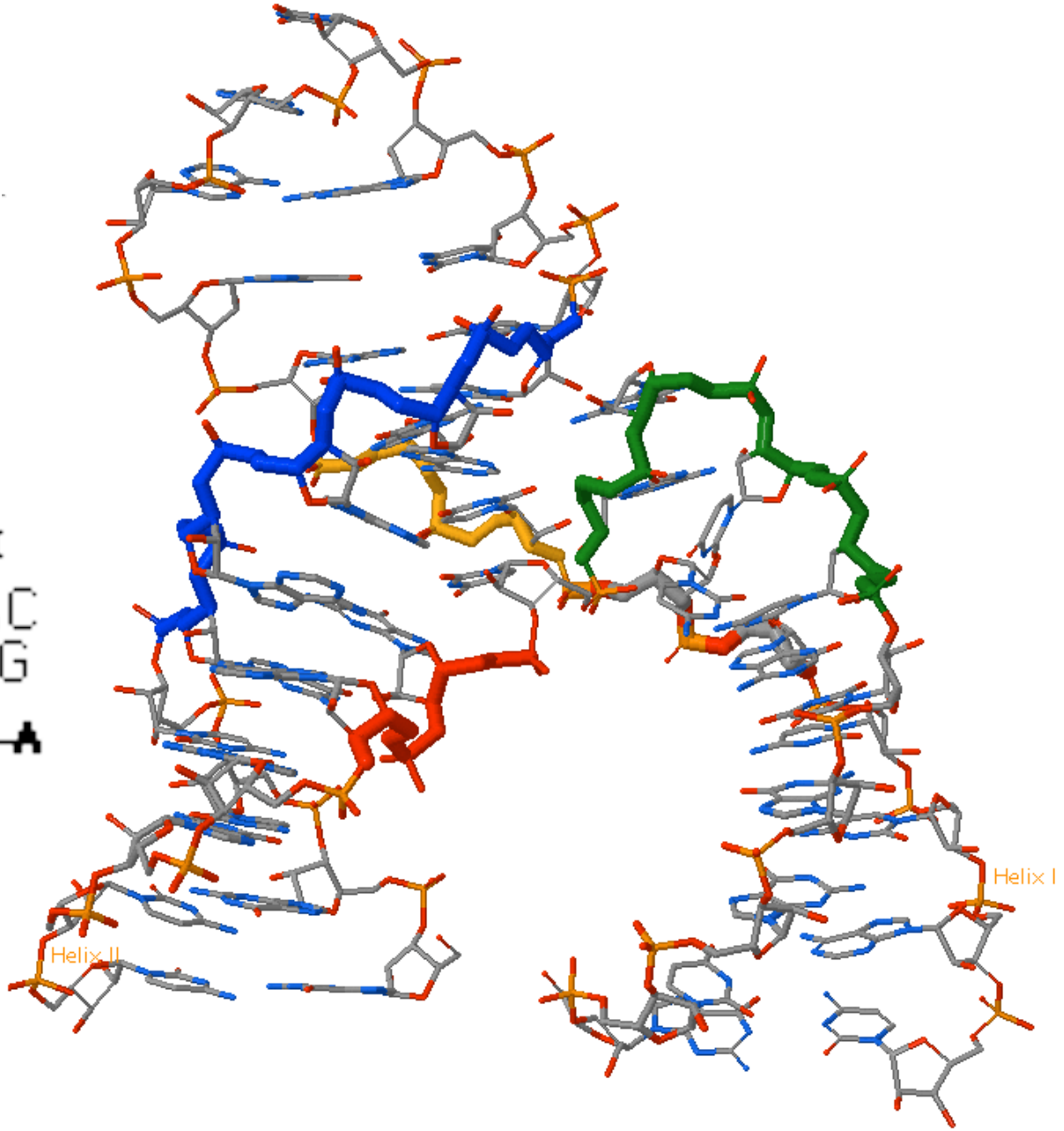
# tRNA



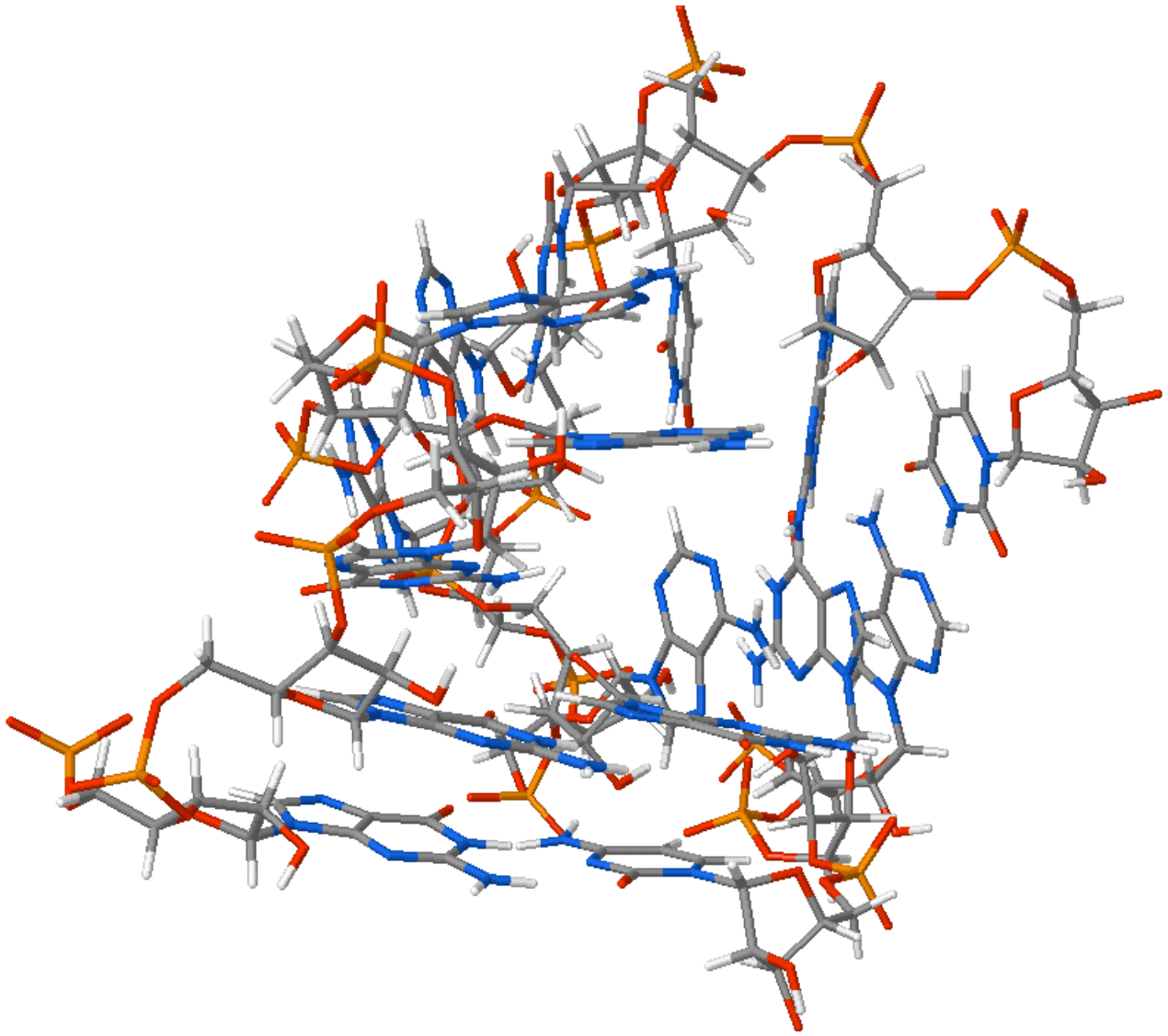



# Hammerhead Ribozyme

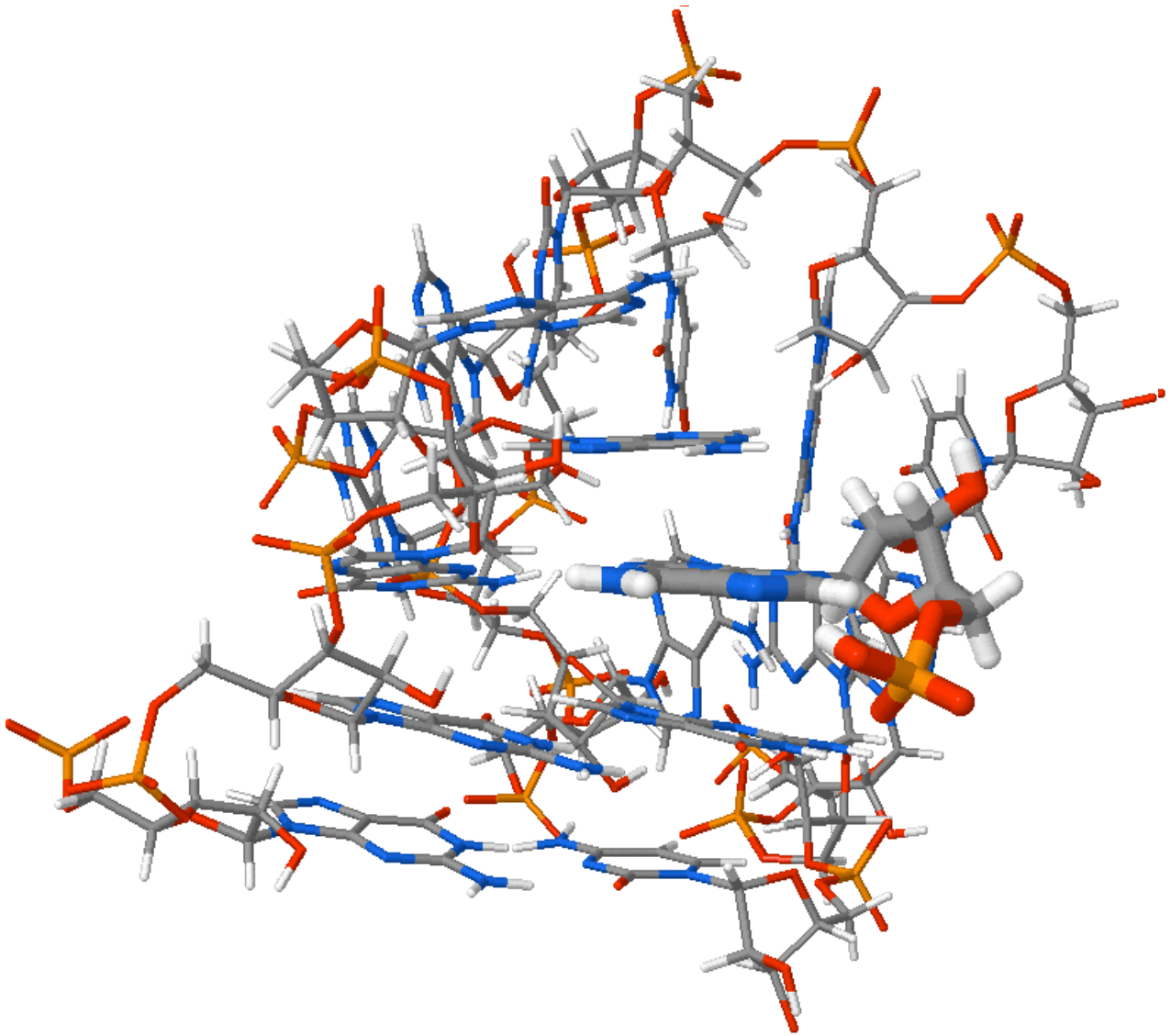


# AMP Aptamer



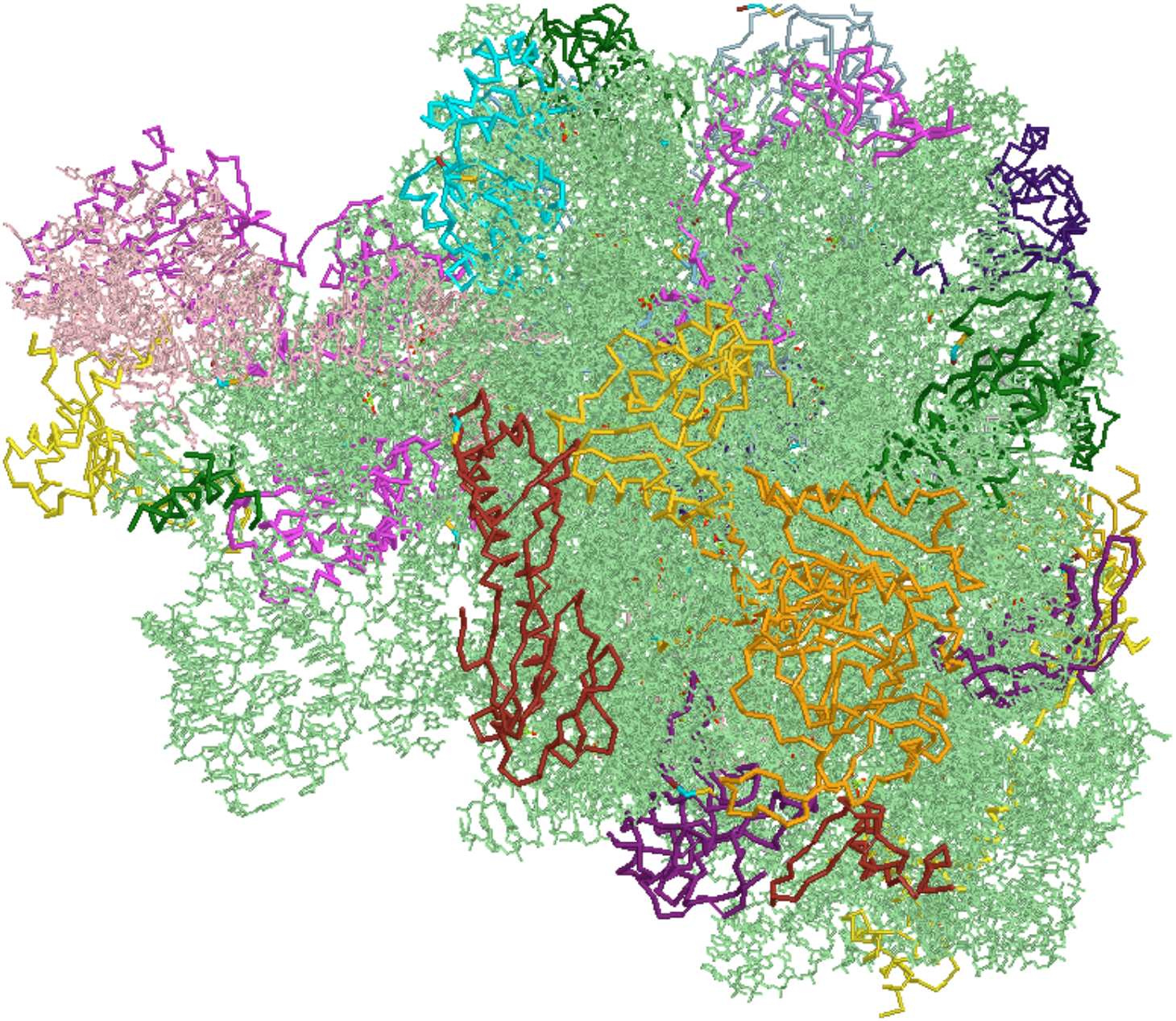
# AMP Aptamer



# RNA

# Ribosome

An RNA machine with protein cofactors

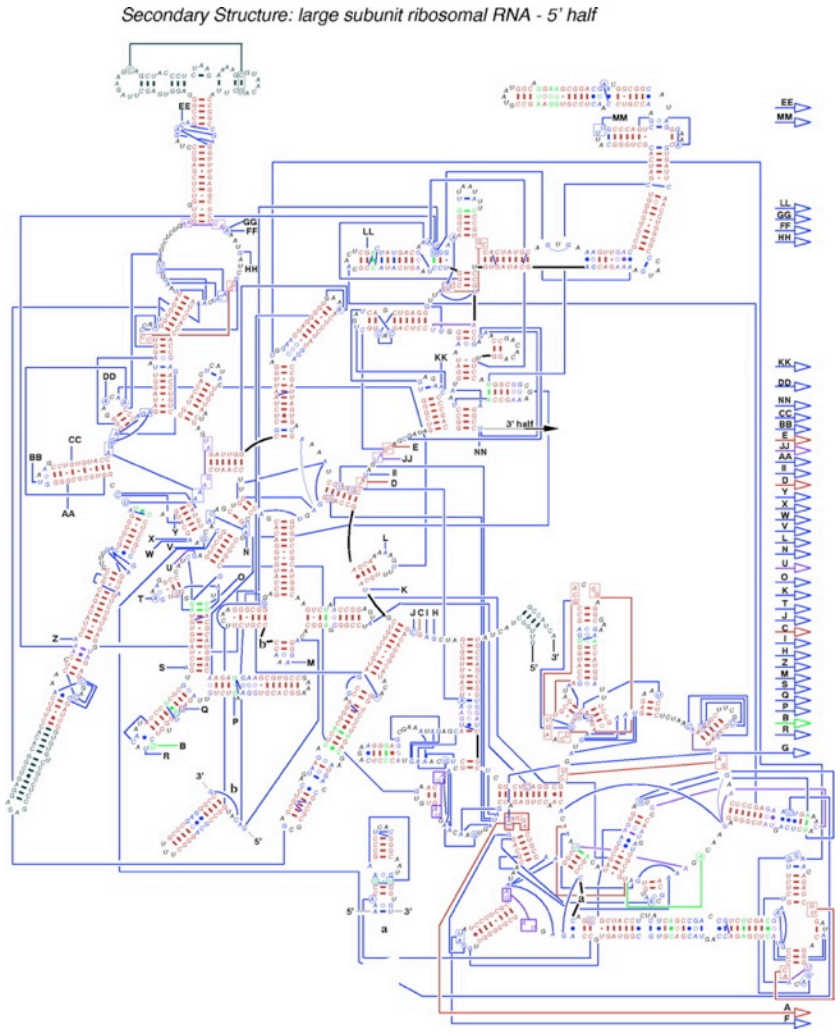


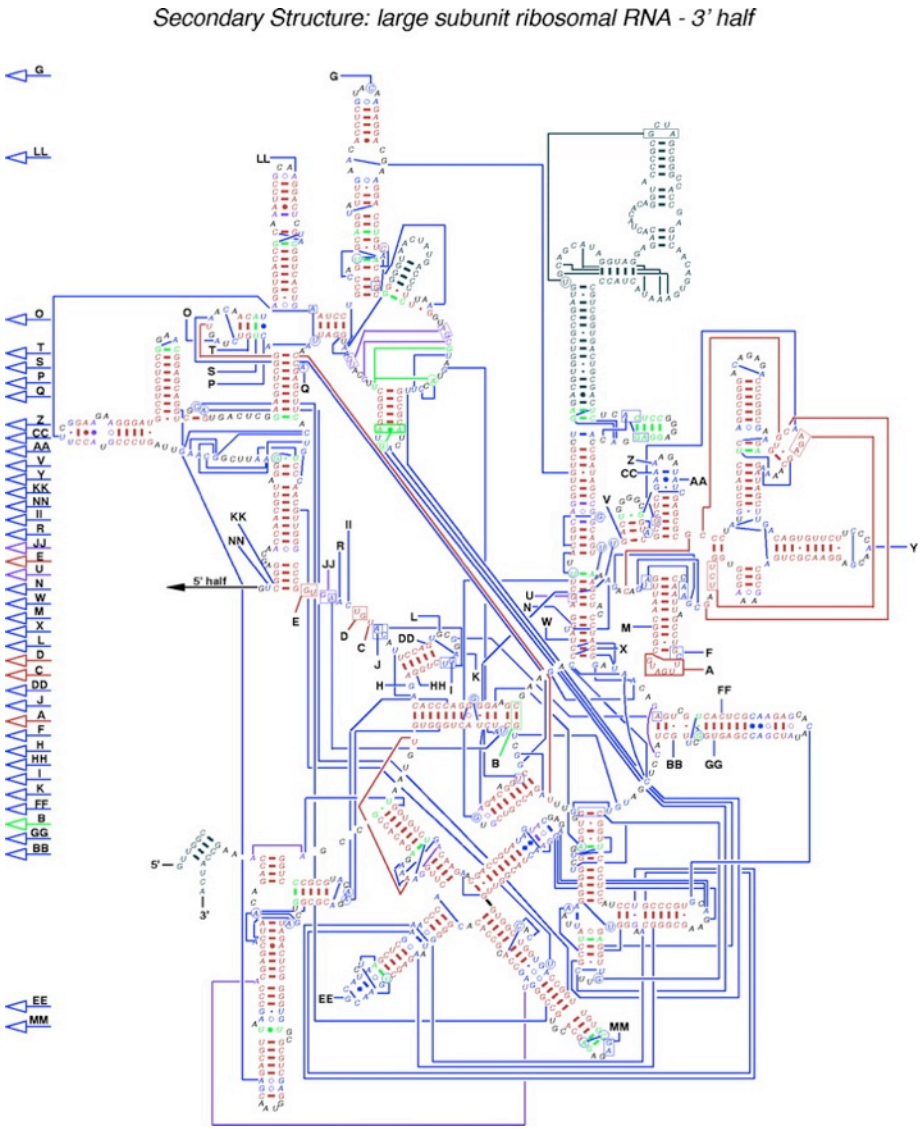
# Ribosome - Secondary Structure

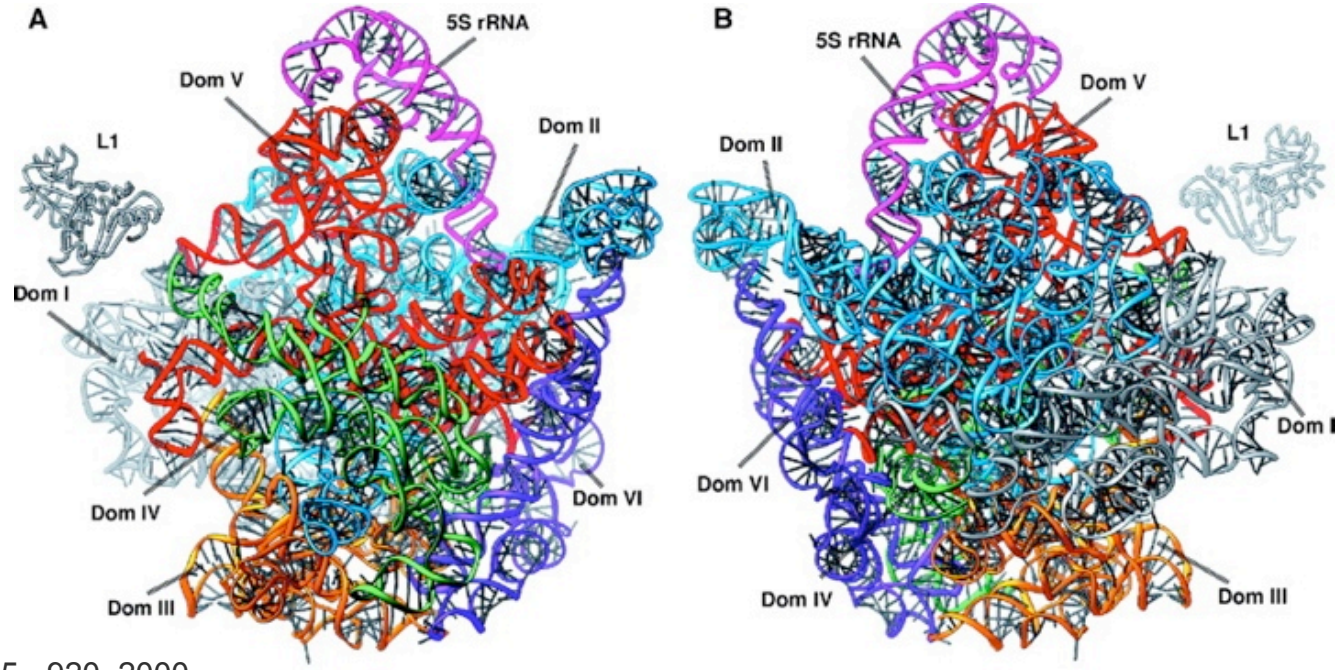
Science 289(5481), 905 - 920, 2000

**The Complete Atomic Structure of the Large Ribosomal Subunit at 2.4 Å Resolution**

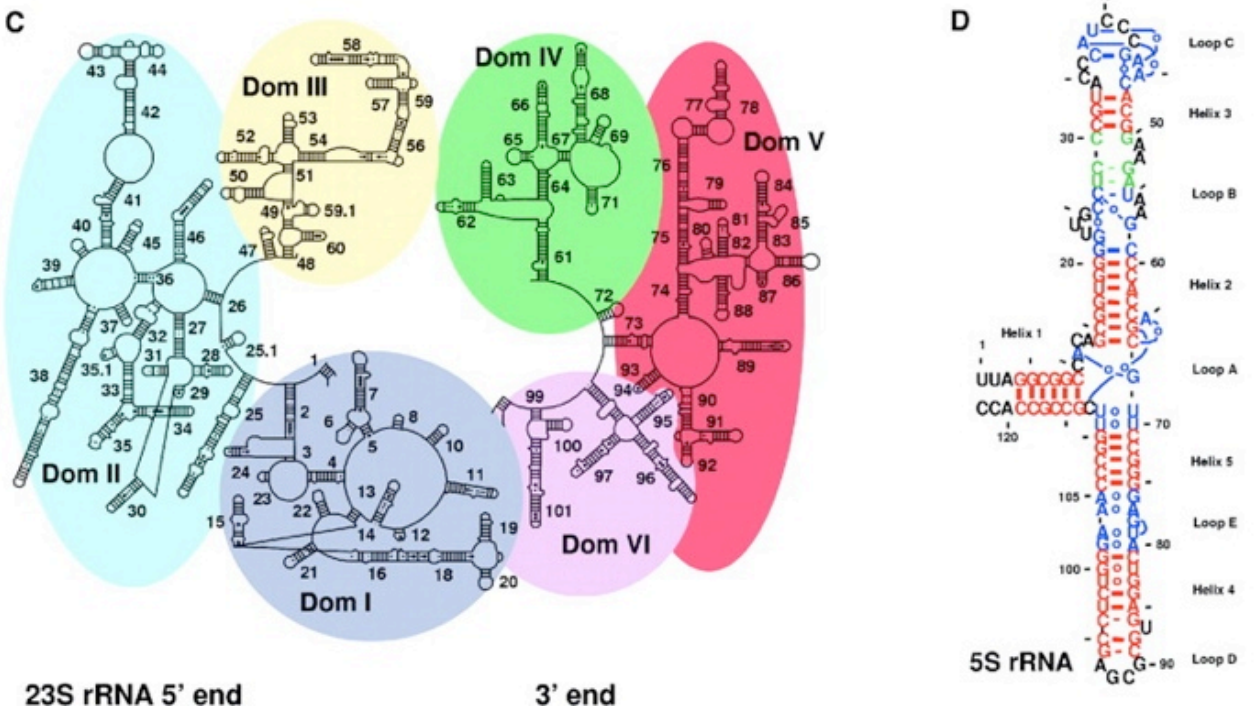
Ban, Nissen, Hansen, Moore, & Steitz






Science 289(5481), 905 - 920, 2000

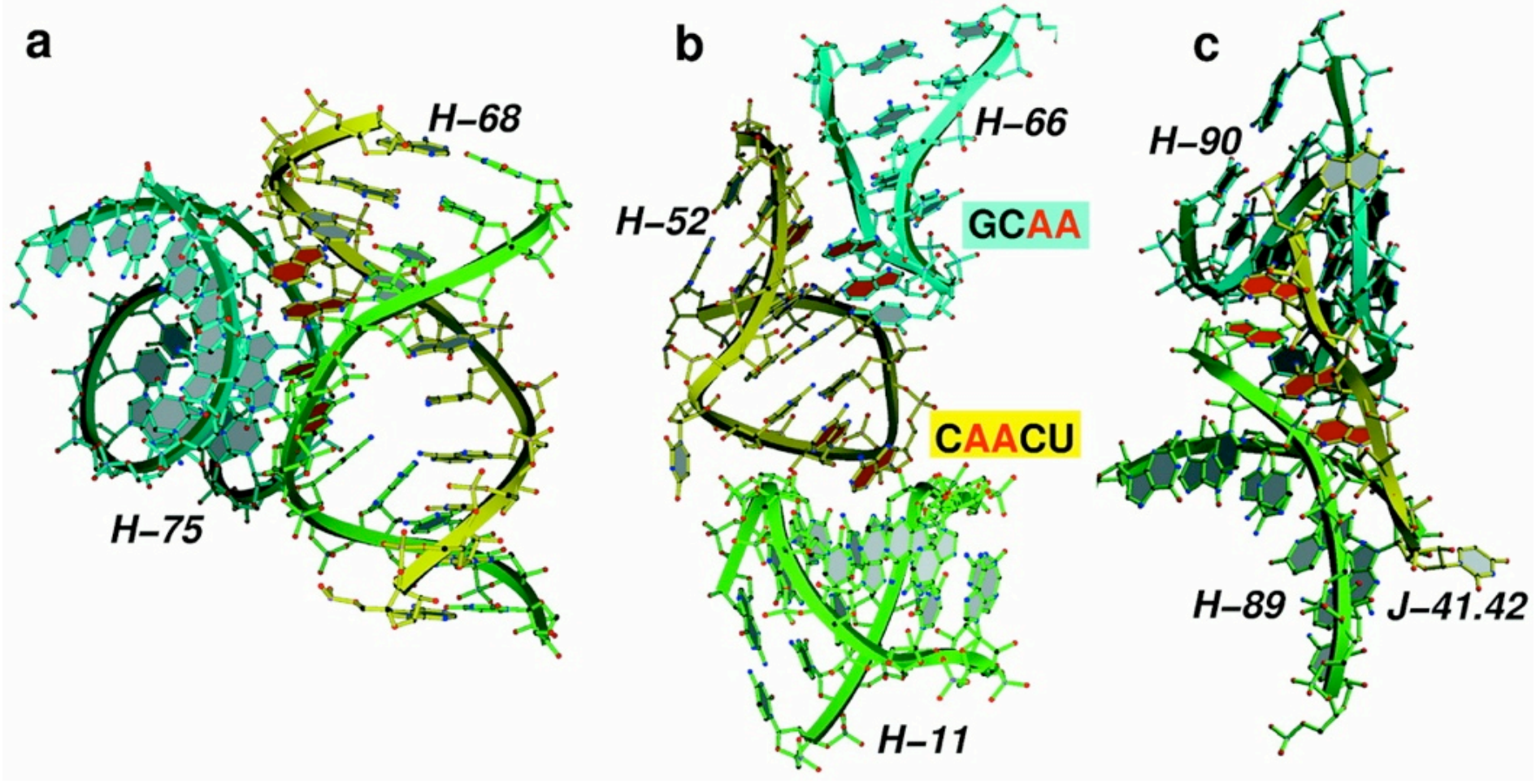


PNAS 98(9), **4899-4903**, 2001

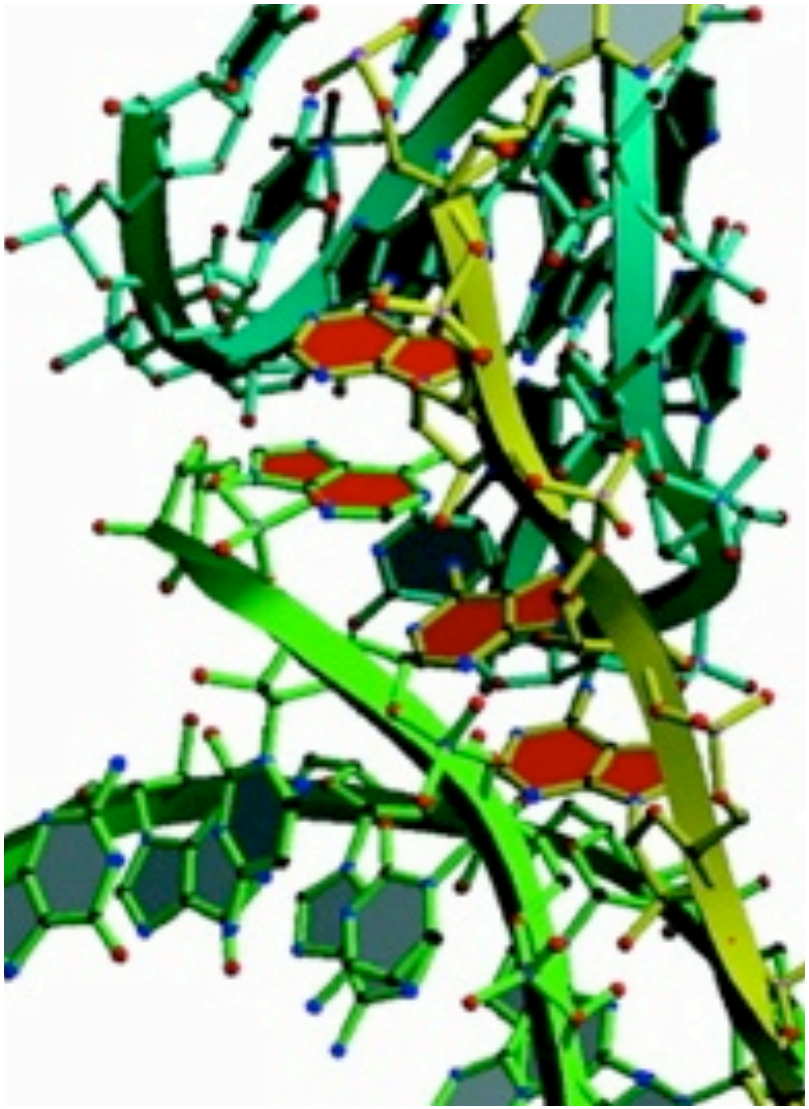
**RNA tertiary interactions in the large ribosomal subunit: The A-minor motif**

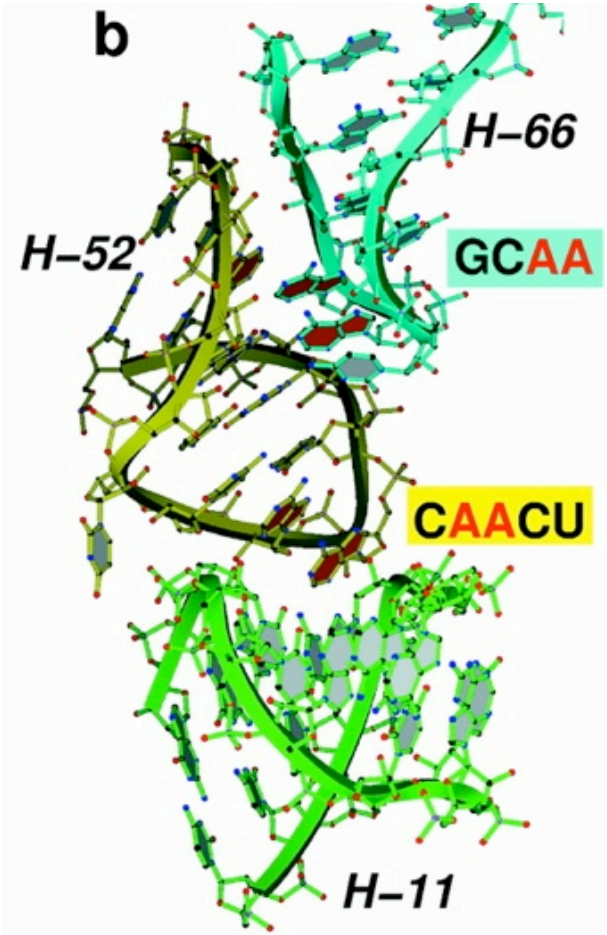
Nissen, Ippolito, Ban, Moore, & Steitz

# The Ribosome: a wealth of RNA structure

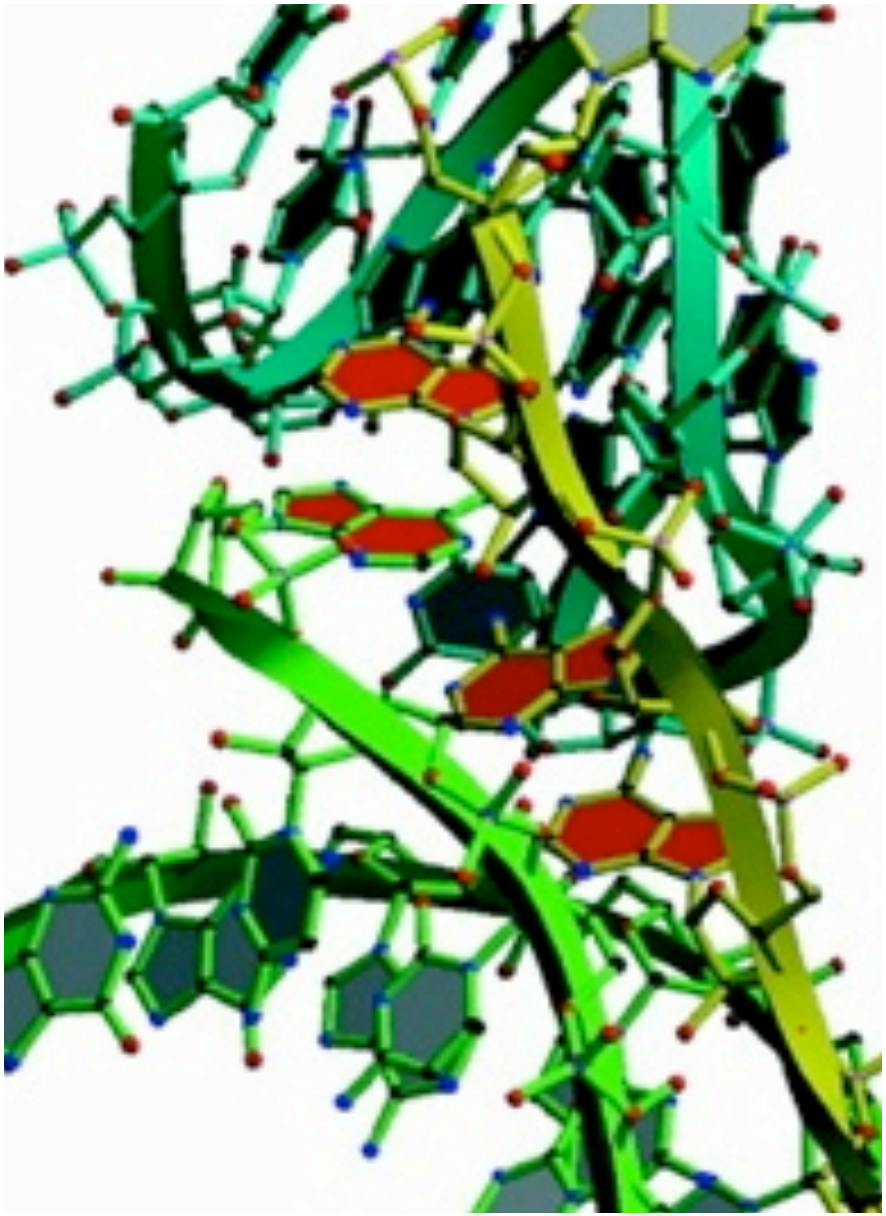


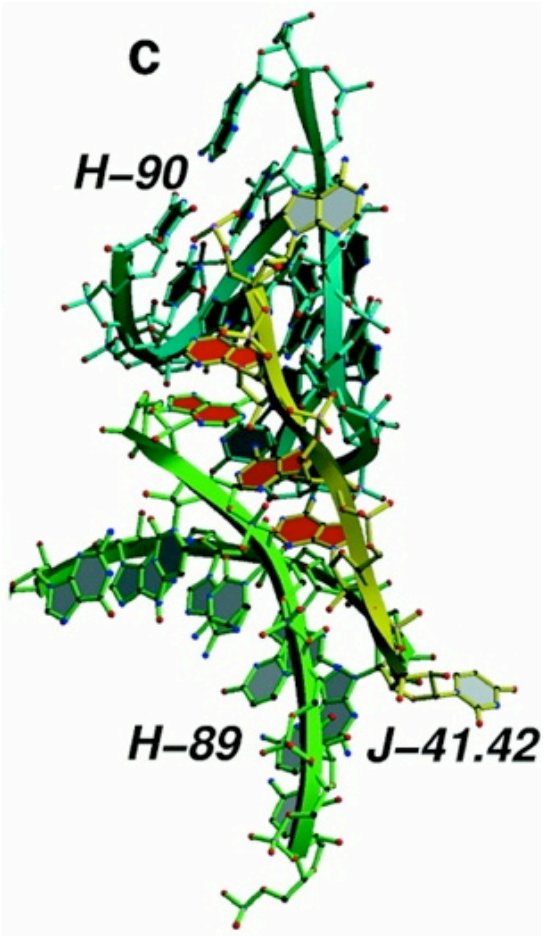
PNAS 98(9), **4899-4903**, 2001



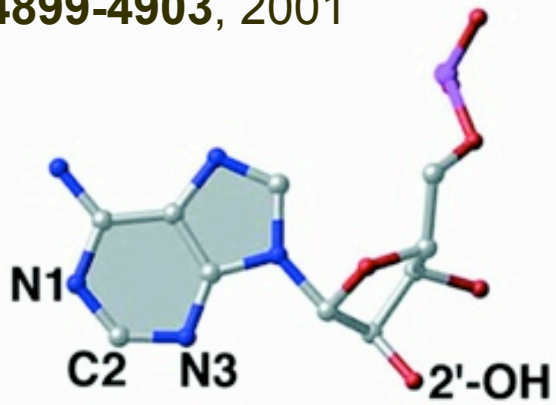


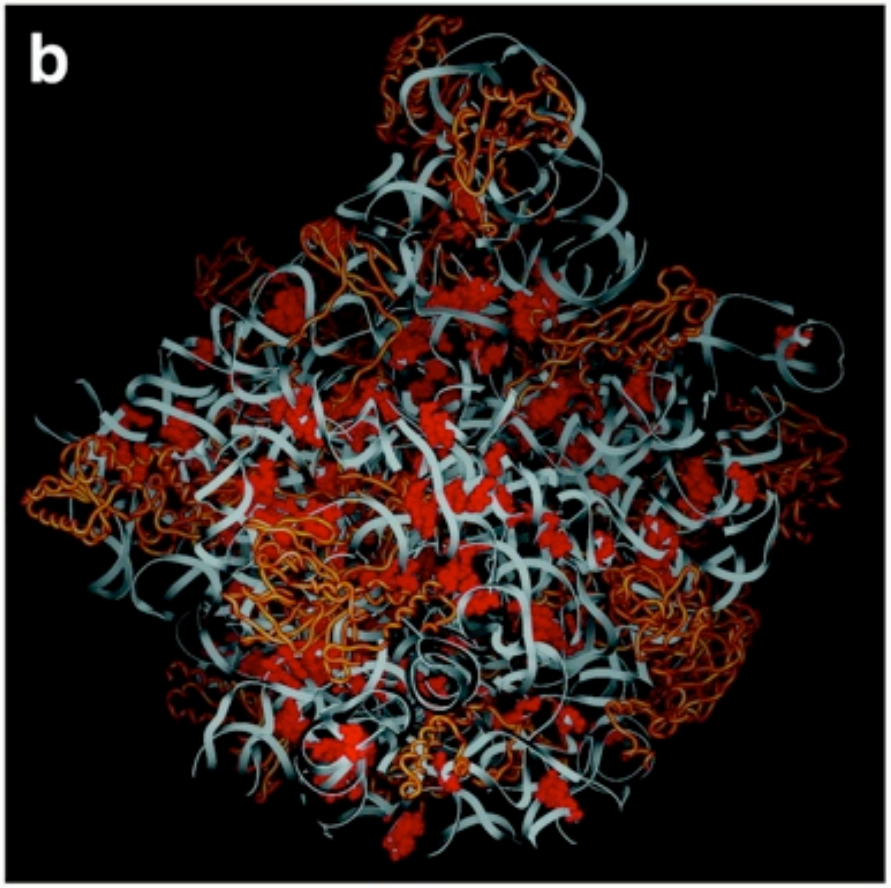
PNAS 98(9), **4899-4903**, 2001

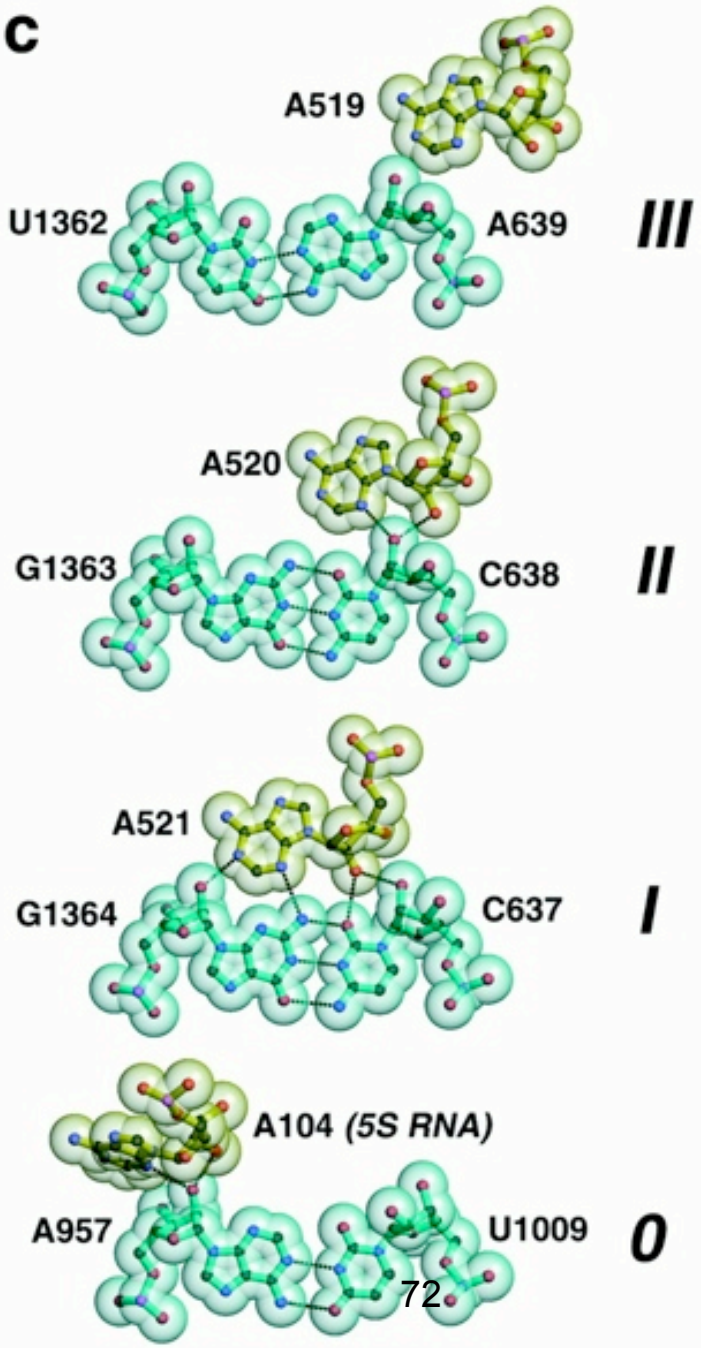




PNAS 98(9), **4899-4903**, 2001

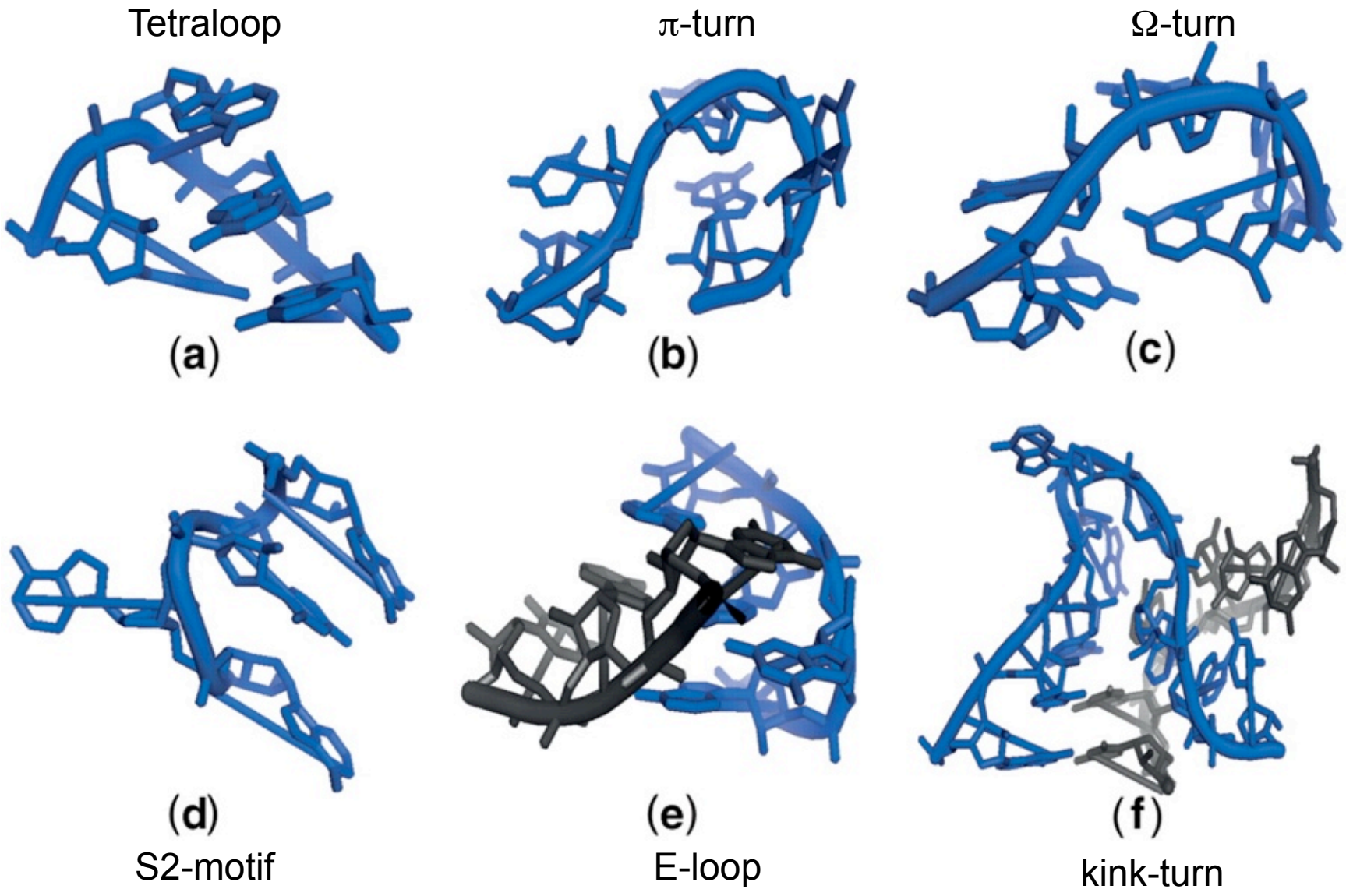


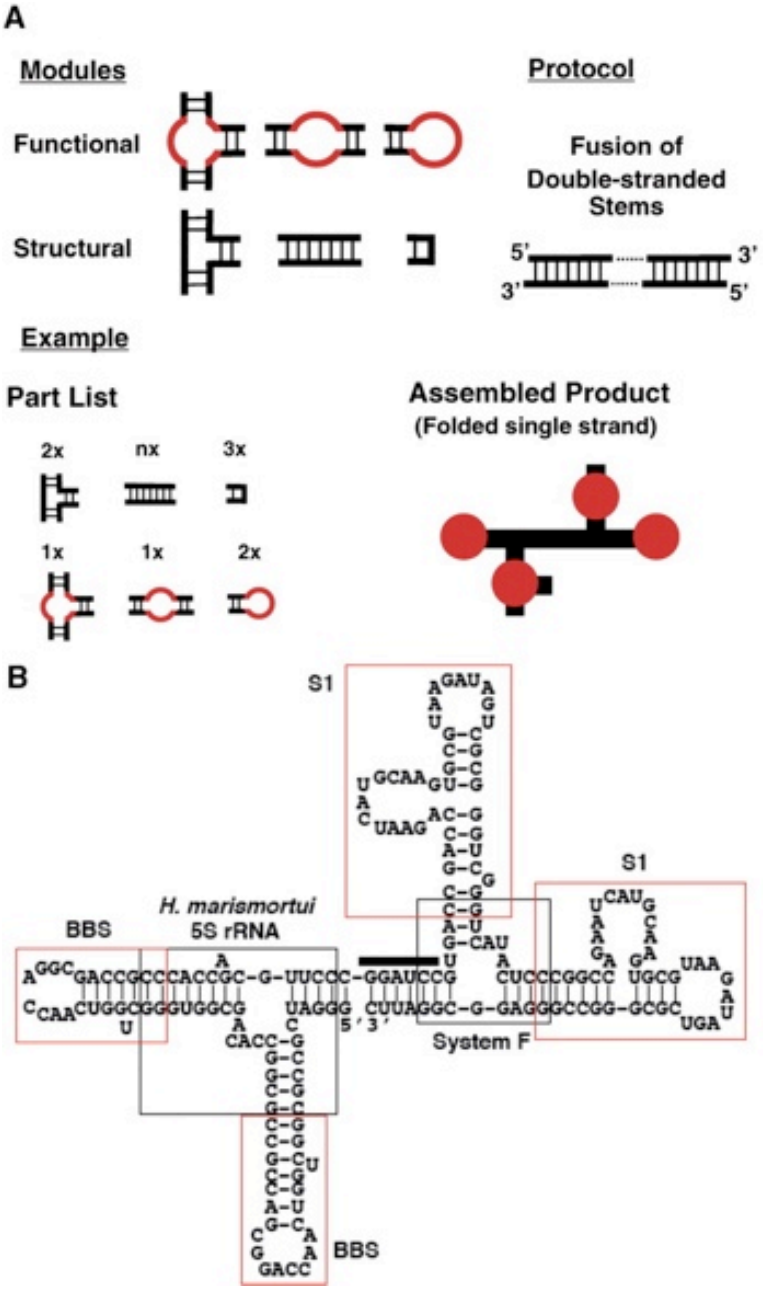



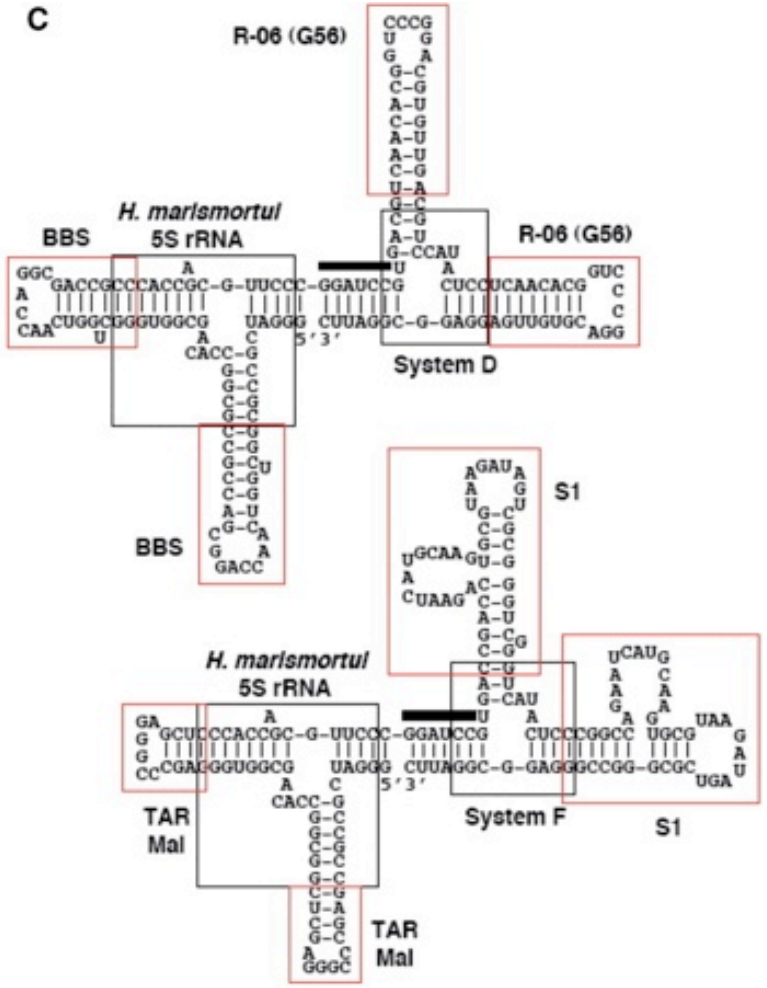
# Structural Motifs

Nucleic Acids Research, 2009, Vol. 37, No. 4 **e29**



# Where to from here? Modular Construction




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**

51.6° C

or

**ACCGCCACCGAAG**
**TGGCGGTGGCTTA**

# Which is more stable?

(which has a higher melting temperature?)

**ACCGCCACCGAAG**
**TGGCGGTGGCTTC**

51.6° C

or

**ACCGCCACCGAAG**
**TGGCGGTGGCTTA**

52.5° C

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