Nucleic Acids

Nano 501

April 18, 2007

Donald E. Bergstrom

Walther Professor of Medicinal Chemistry

Department of Medicinal Chemistry and Molecular Pharmacology

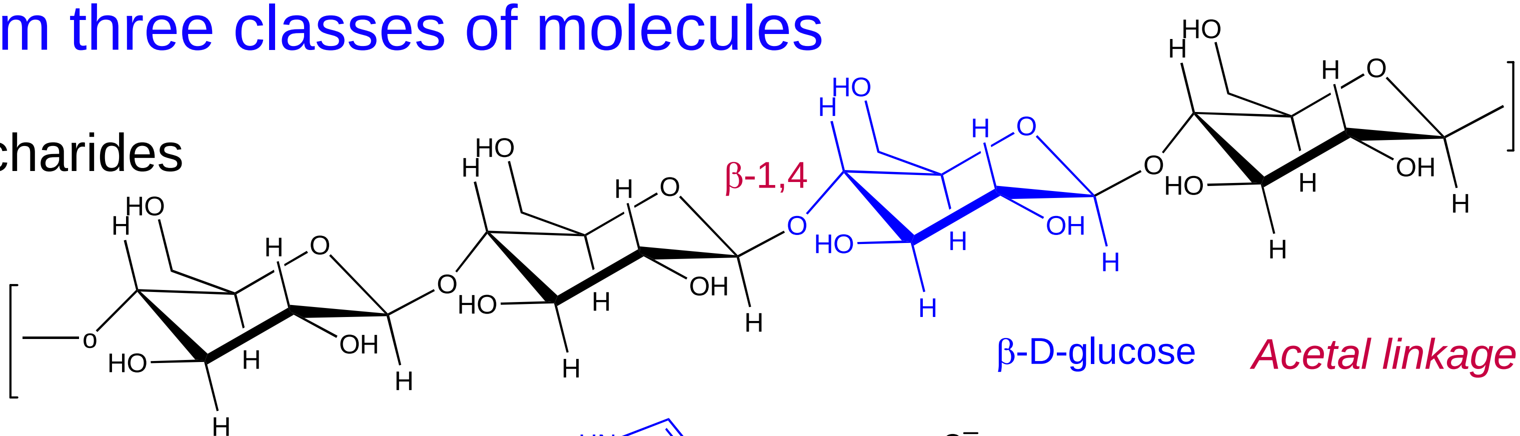
Birck Nanotechnology Center

Nucleic Acids

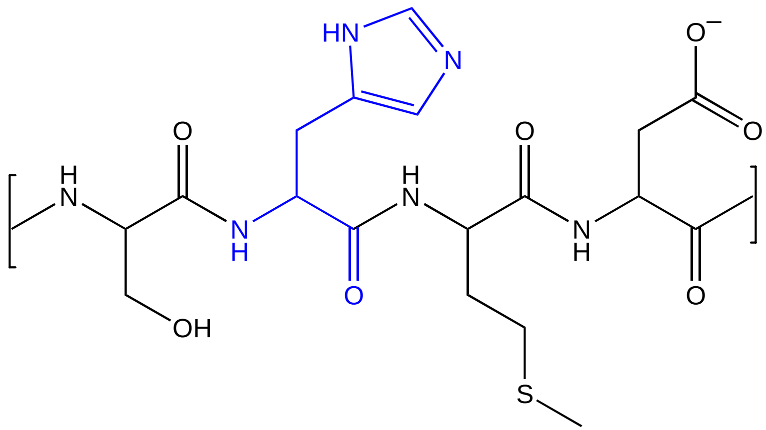
I. Structure
II. Biological Function
III. Modification of RNA
IV. Genesis of Adenine
V. Material Science

Nature builds nanoscale materials largely from three classes of molecules

Polysaccharides



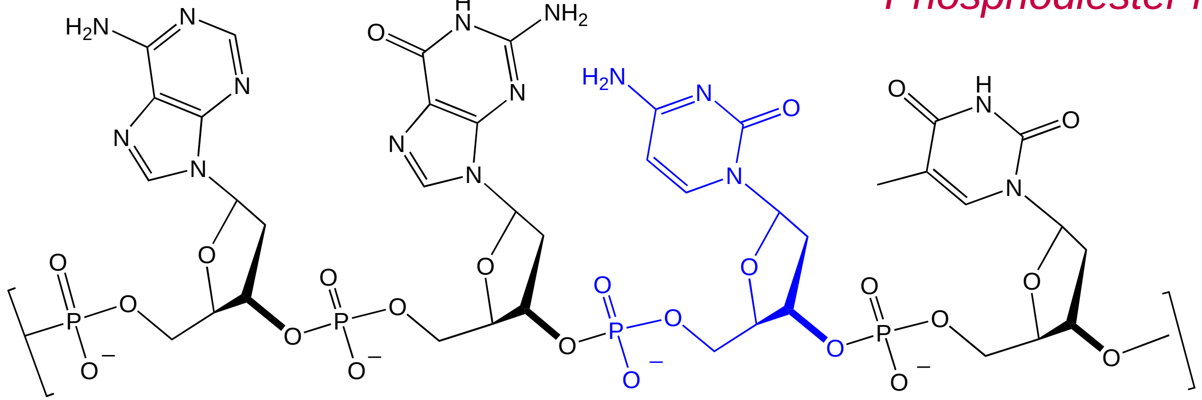
Polypeptides



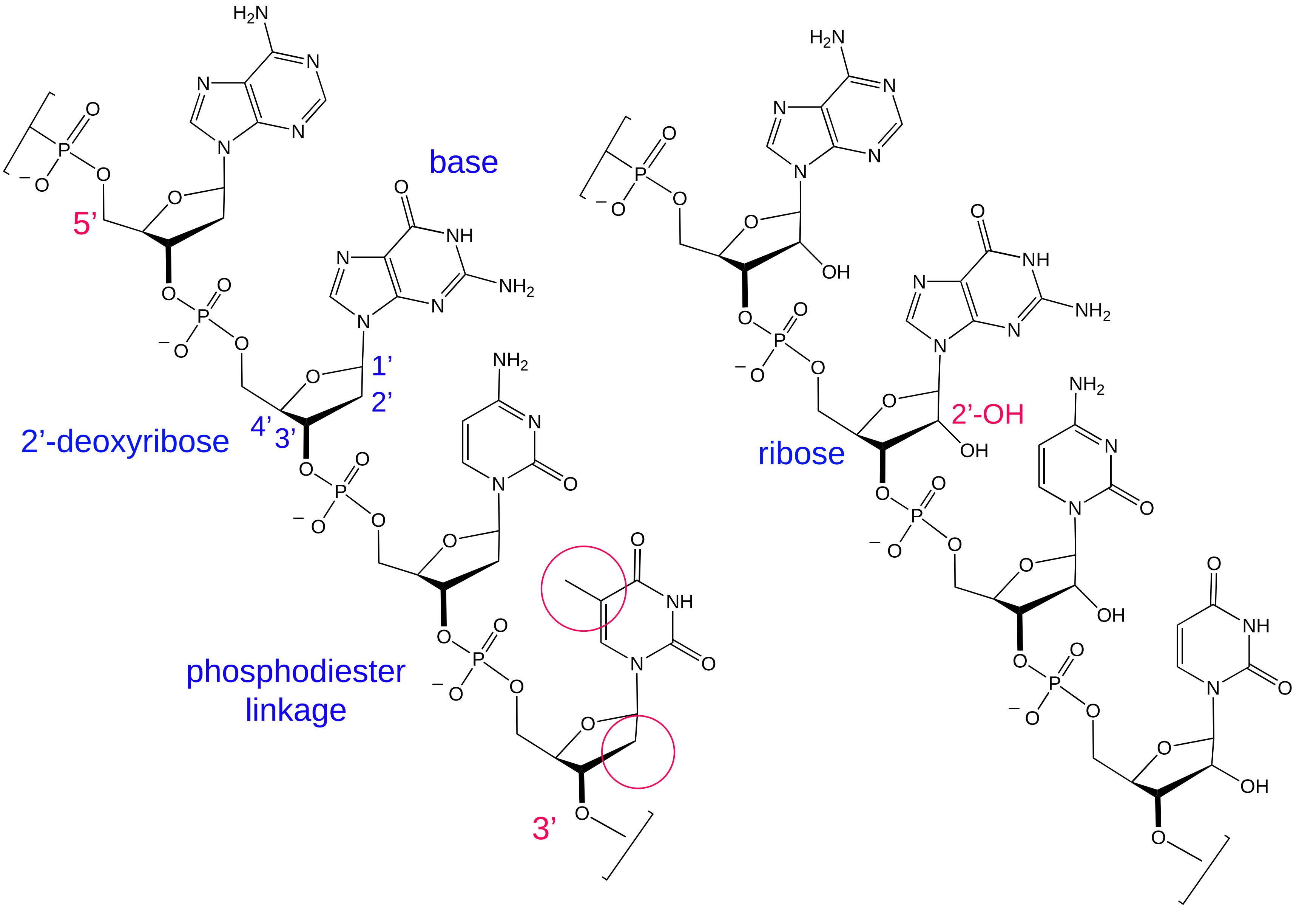
Amide linkage

Phosphodiester linkage

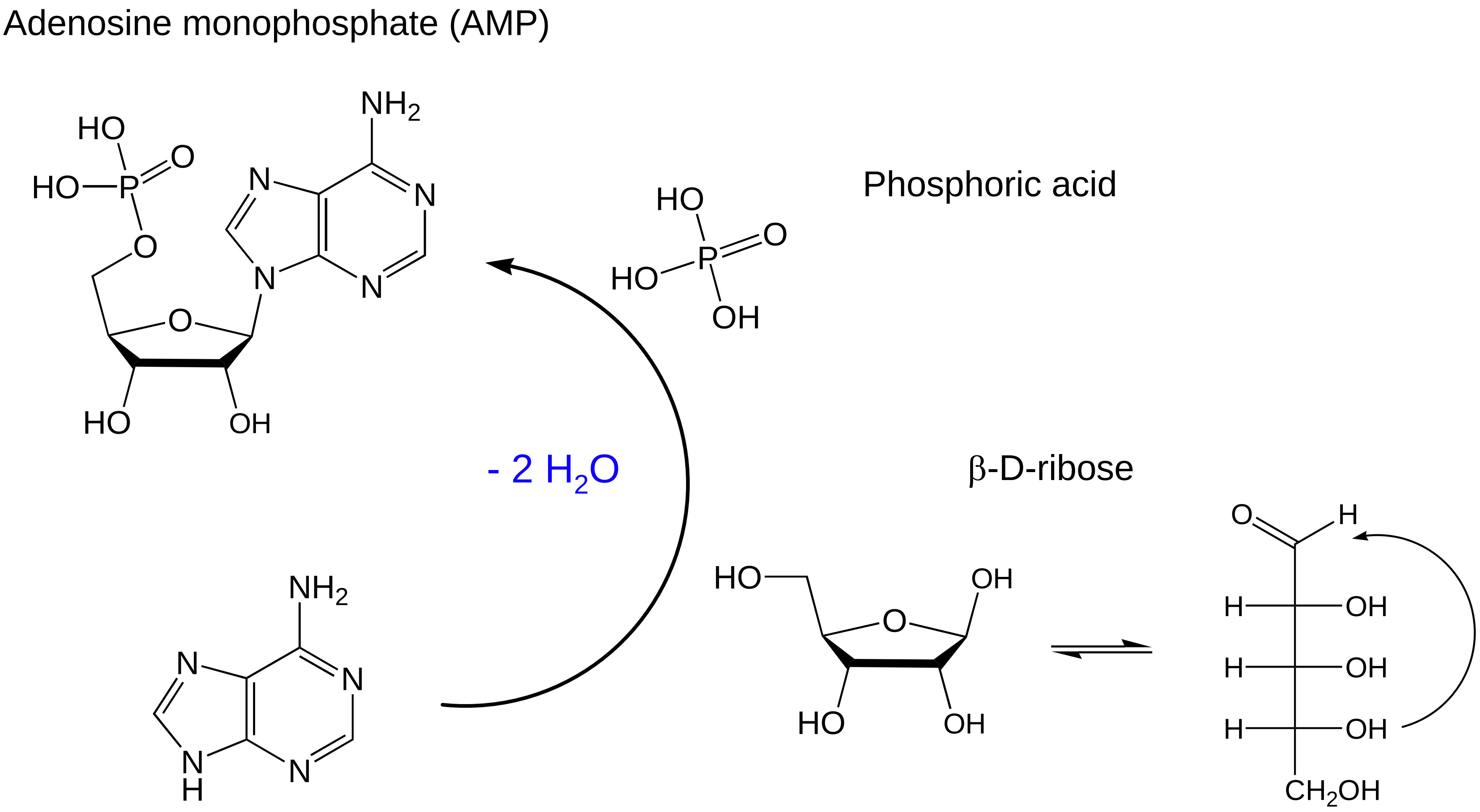
Polynucleotides



Deoxyribonucleic acid (DNA) Ribonucleic acid (RNA)

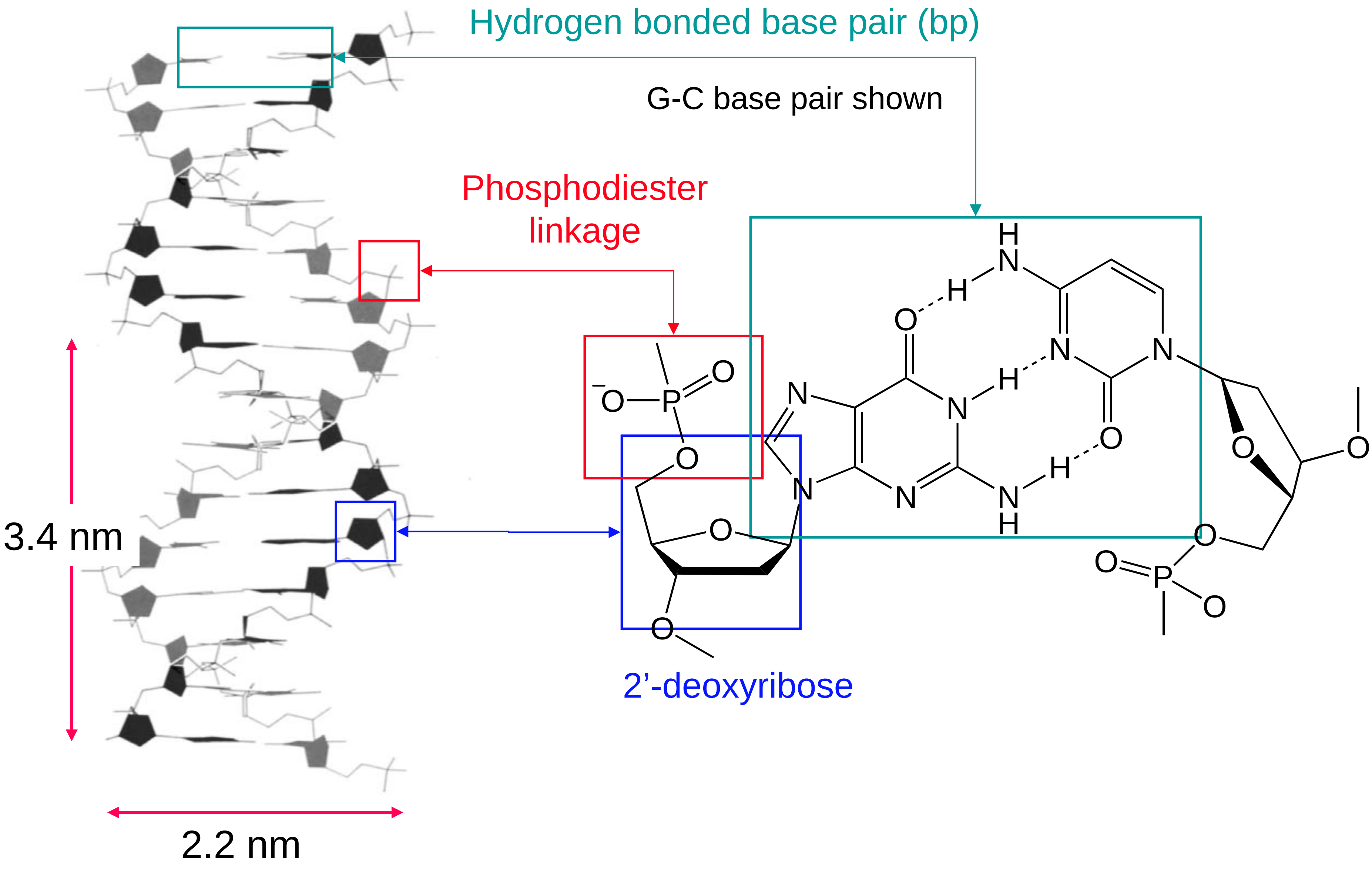


DNA monomeric unit

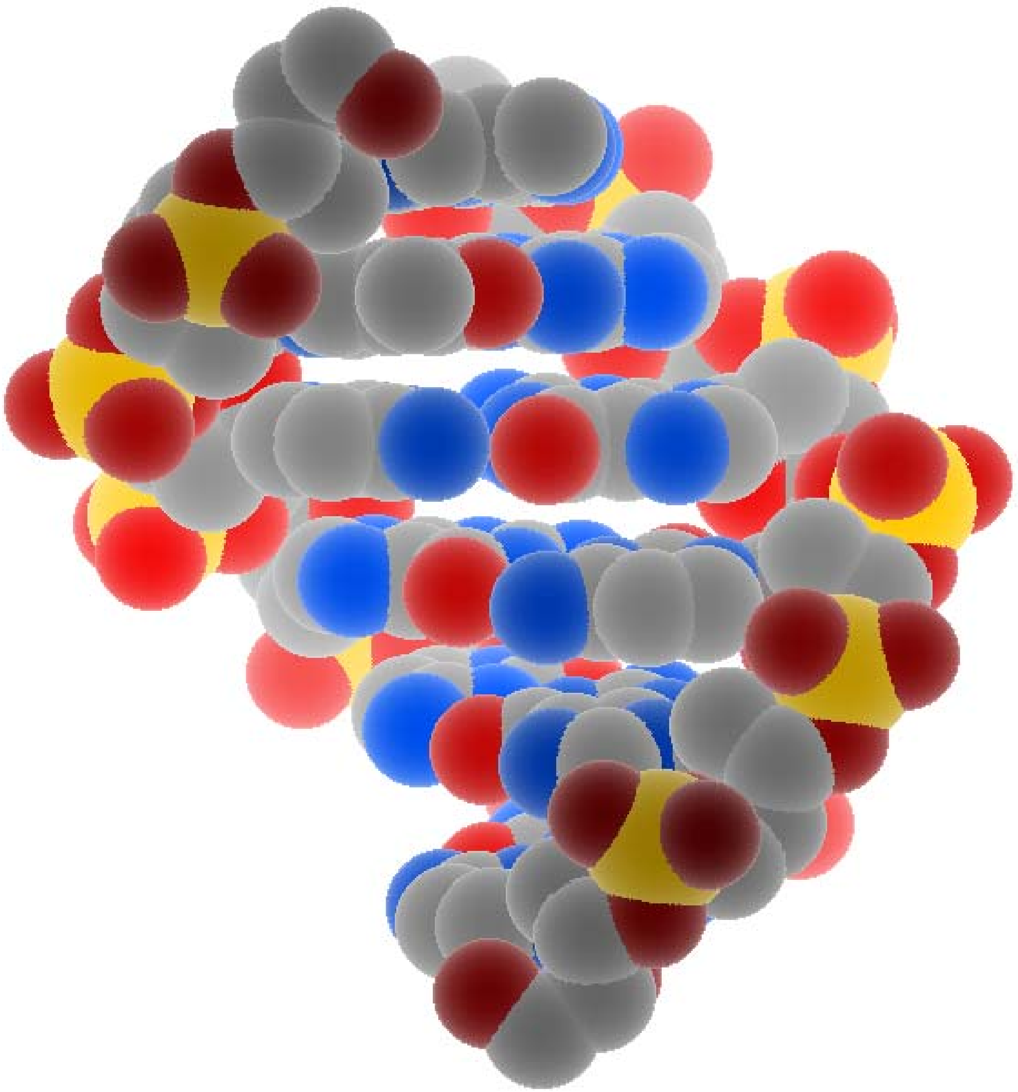


B-DNA Structure

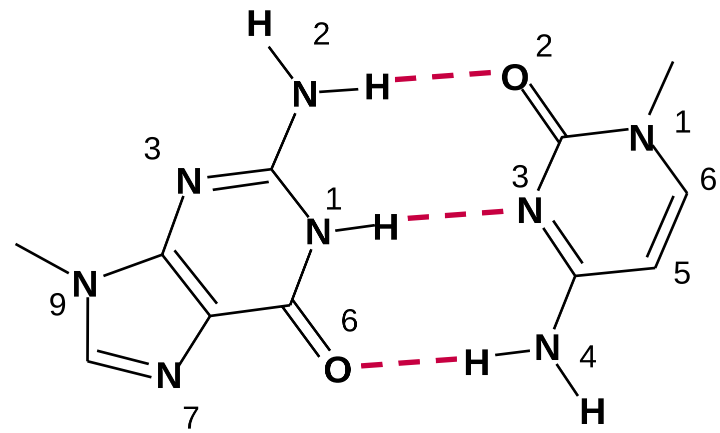
10 bp / helix turn



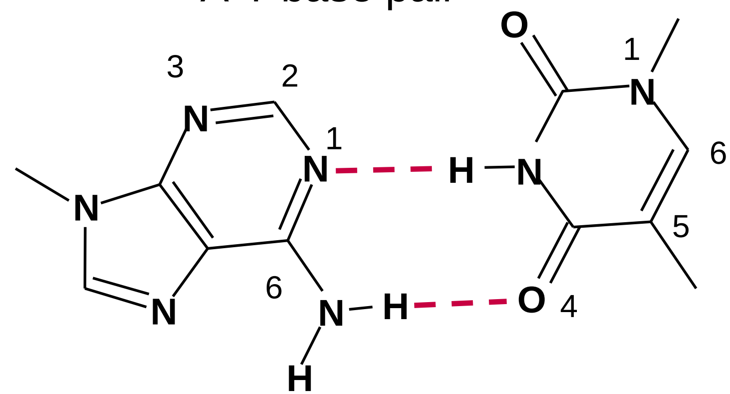
Space filling model of B-DNA structure



G-C base pair



A-T base pair



Central Dogma of Molecular Biology

DNA

Transcription

RNA

Translation

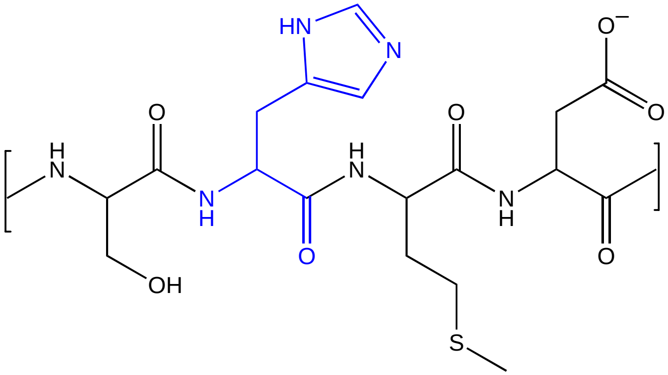
Proteins

5’….TCA CAT ATG GAT….3’

Template strand 3’….AGT GTA TAC CTA….5’

mRNA 5’….UCA-CAU-AUG-GAU….3’

…Ser-His-Met-Asp…



The Human Genome

Human nuclear DNA

3.1 billion bp

~ 24,000 genes (3% of genome)

Genes in the human code for the construction of:

Proteins ~ 23,000

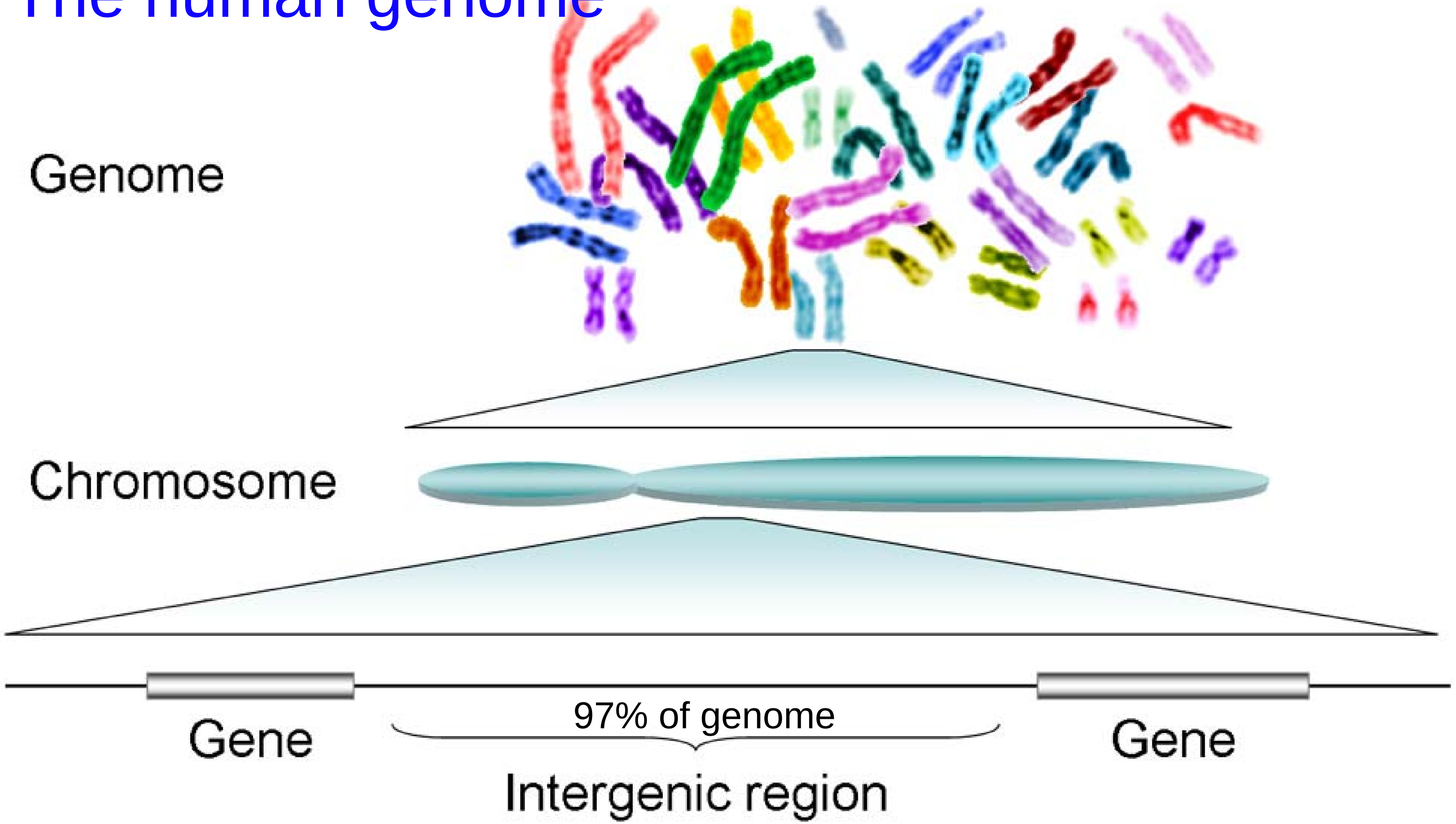
miRNA ~ 400

rRNA ~ 7

snRNA ~ 170

tRNA ~ 60

The human genome



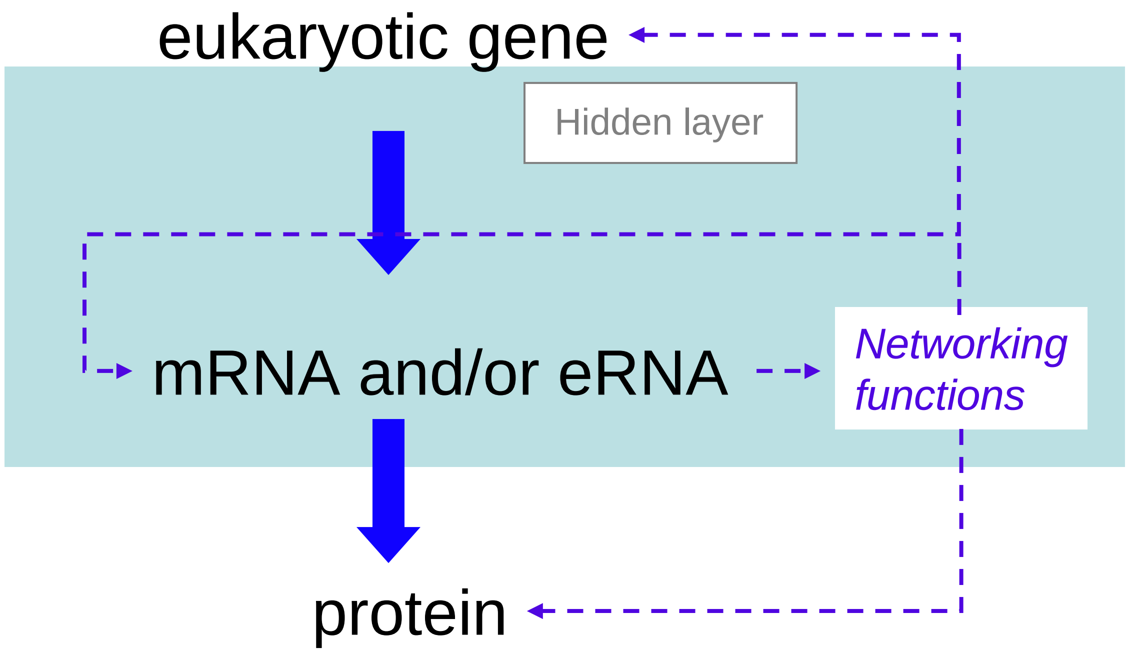
The human genome is composed of 23 pairs of chromosomes (46 *in total*), each of which contain hundreds of genes separated by *intergenic regions*. Intergenic regions may contain regulatory sequences and non-coding DNA. http://en.wikipedia.org/wiki/Human_genome

RNA

	Name	Size	Function
miRNA	microRNA	21-23 nt	regulation of gene expression
mRNA	messengerRNA	1400 nt (average)	code for protein synthesis
snRNA	small nuclear RNA	40-50 nt	pre-mRNA splicing
tRNA	transfer RNA	73-93 nt	protein synthesis
ncRNA	non-coding RNA	small	Unknown function
rRNA	ribosomal RNA	5000 nt (varies)	organization of protein synthesis machinery

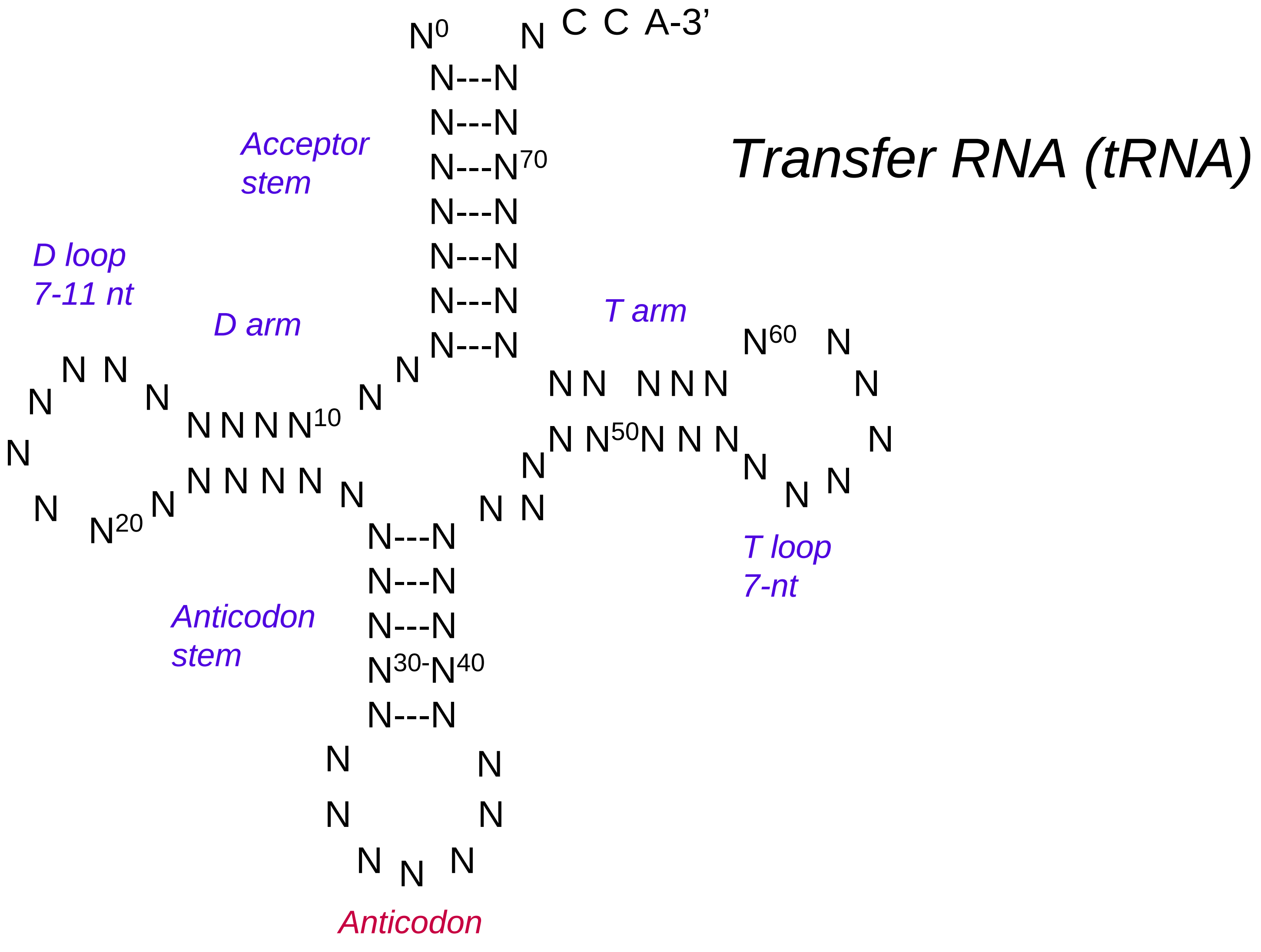
RNA Regulatory Networks

Undiscovered dark matter of the eukaryotic cell

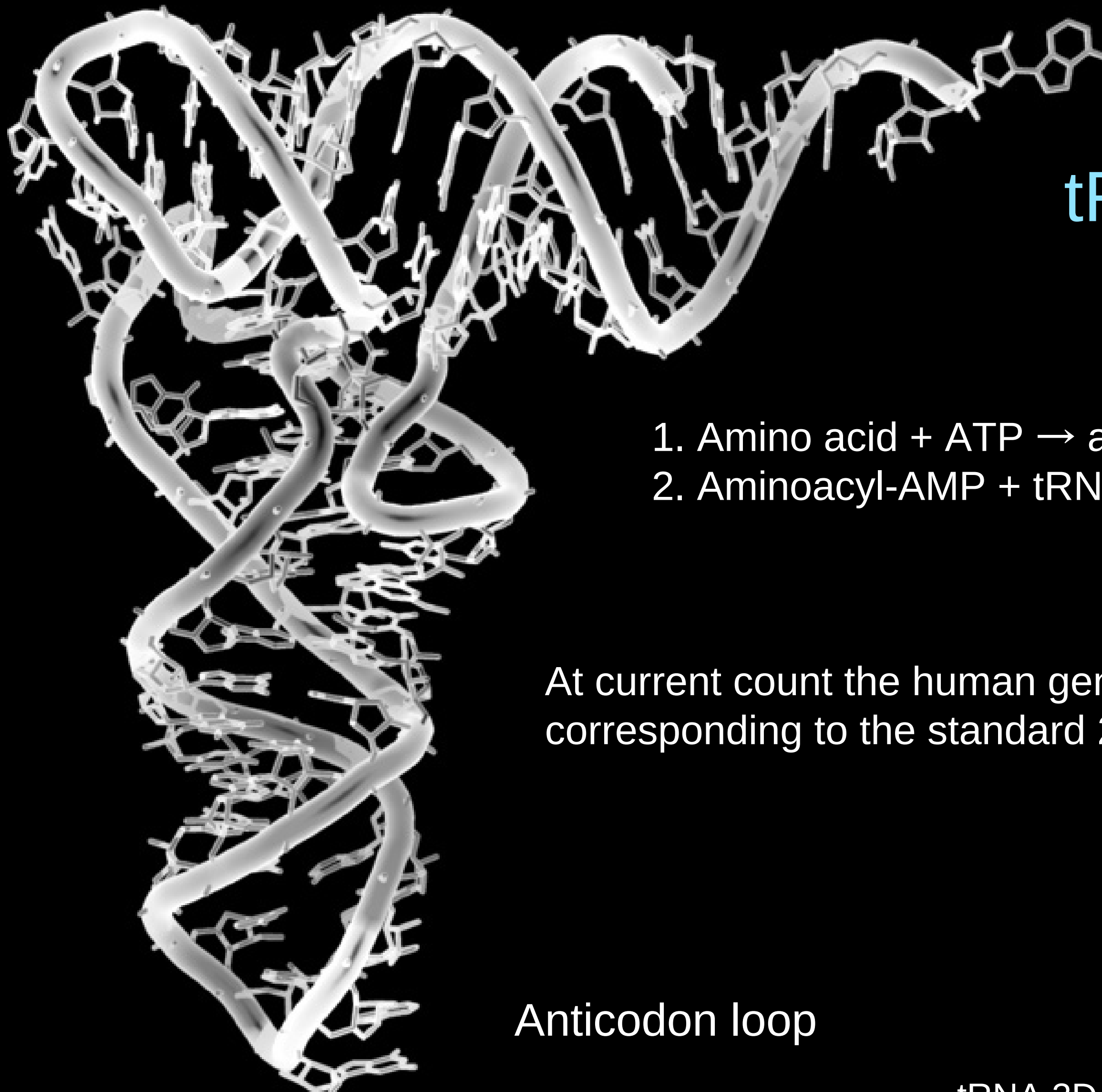


The Functional Genomics of Noncoding RNA, John S. Mattick, *Science* **2005**, *309*, 1527..

RNAs typically have extensive secondary structure



RNAs are more complex than DNA



tRNA 3D structure

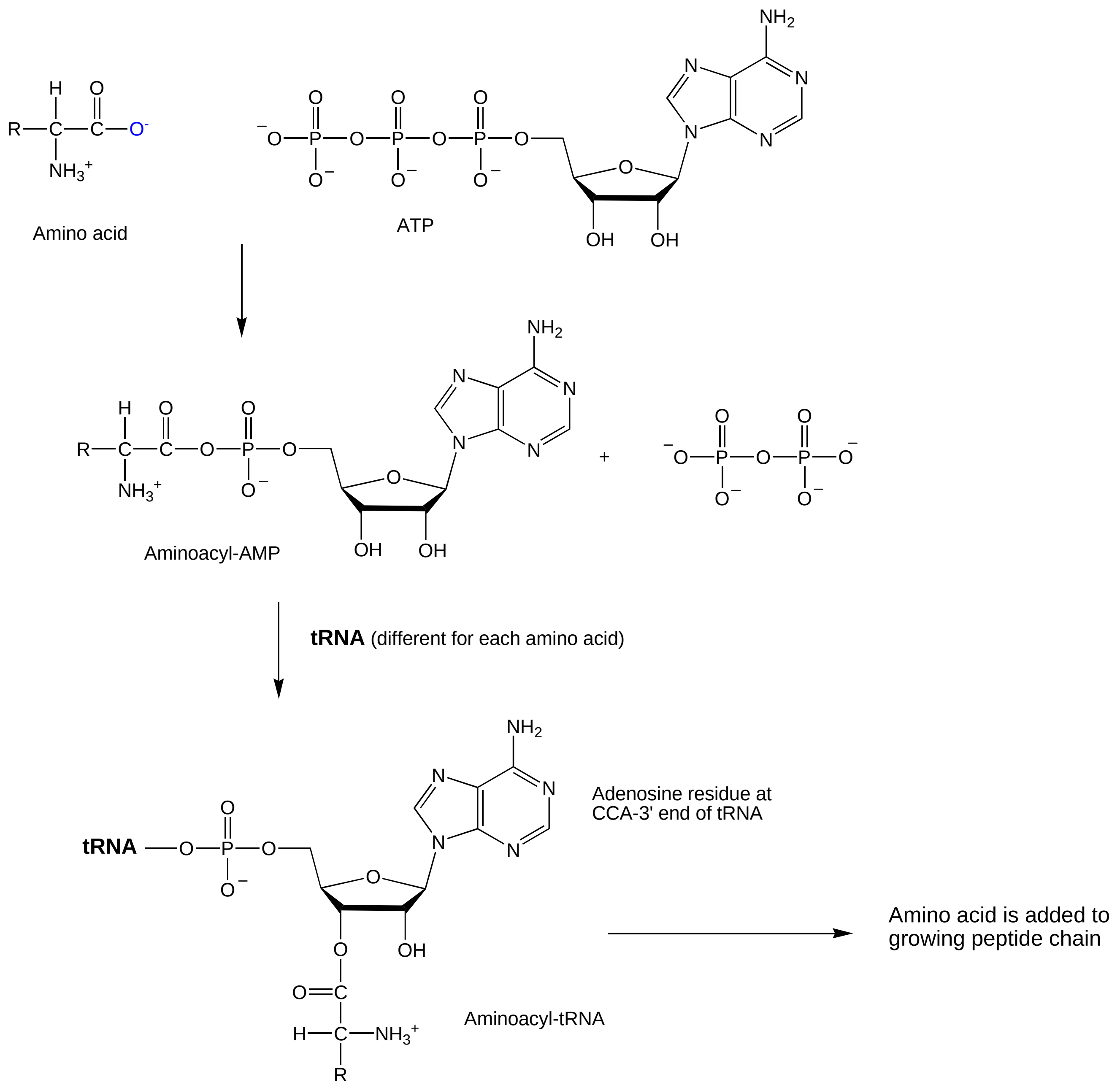
1. Amino acid + ATP → aminoacyl-AMP + Ppi
2. Aminoacyl-AMP + tRNA → aminoacyl-tRNA + AMP

At current count the human genome contains 448 genes corresponding to the standard 20-amino acid decoding tRNAs

Anticodon loop

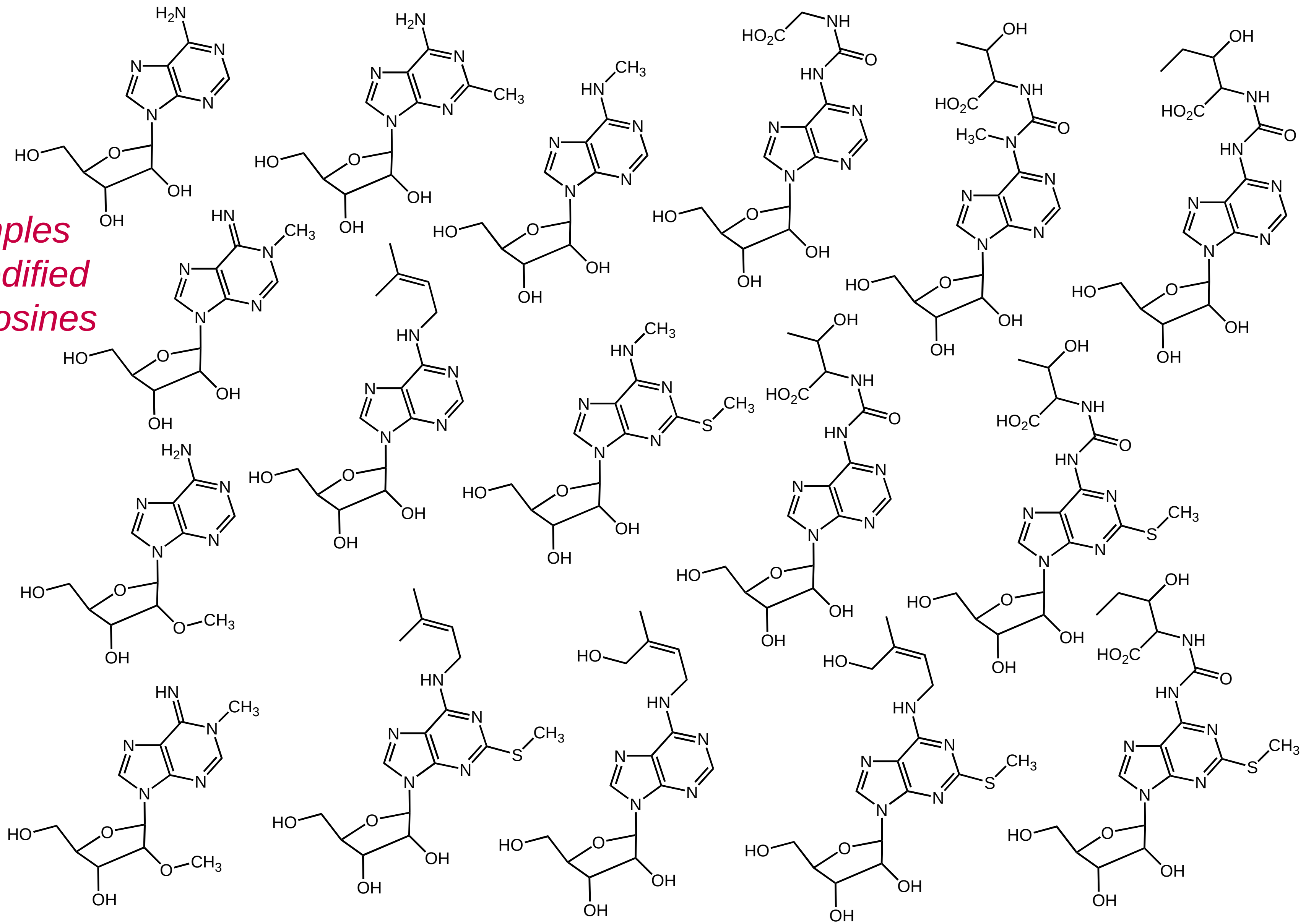
tRNA 3D structure -1974 Alexander Rich, Aaron Klug & coworkers- *Wikipedia*

tRNAs are critical intermediates in protein synthesis



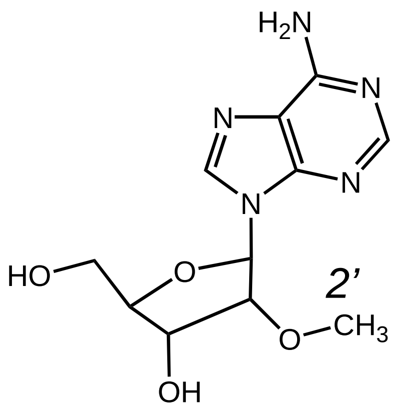
Nucleotides in tRNA are often modified

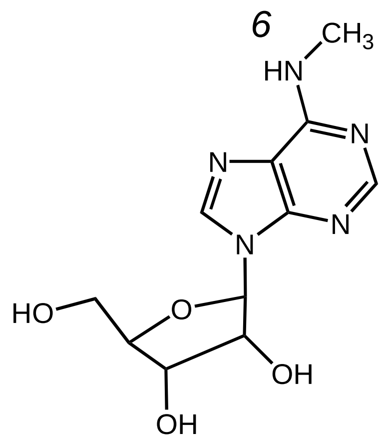
Examples of modified adenosines



Nucleotide Methylation

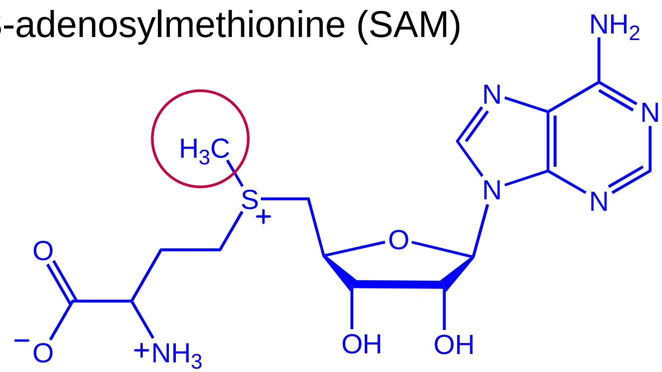
The most common modification in tRNA, mRNA, and rRNA





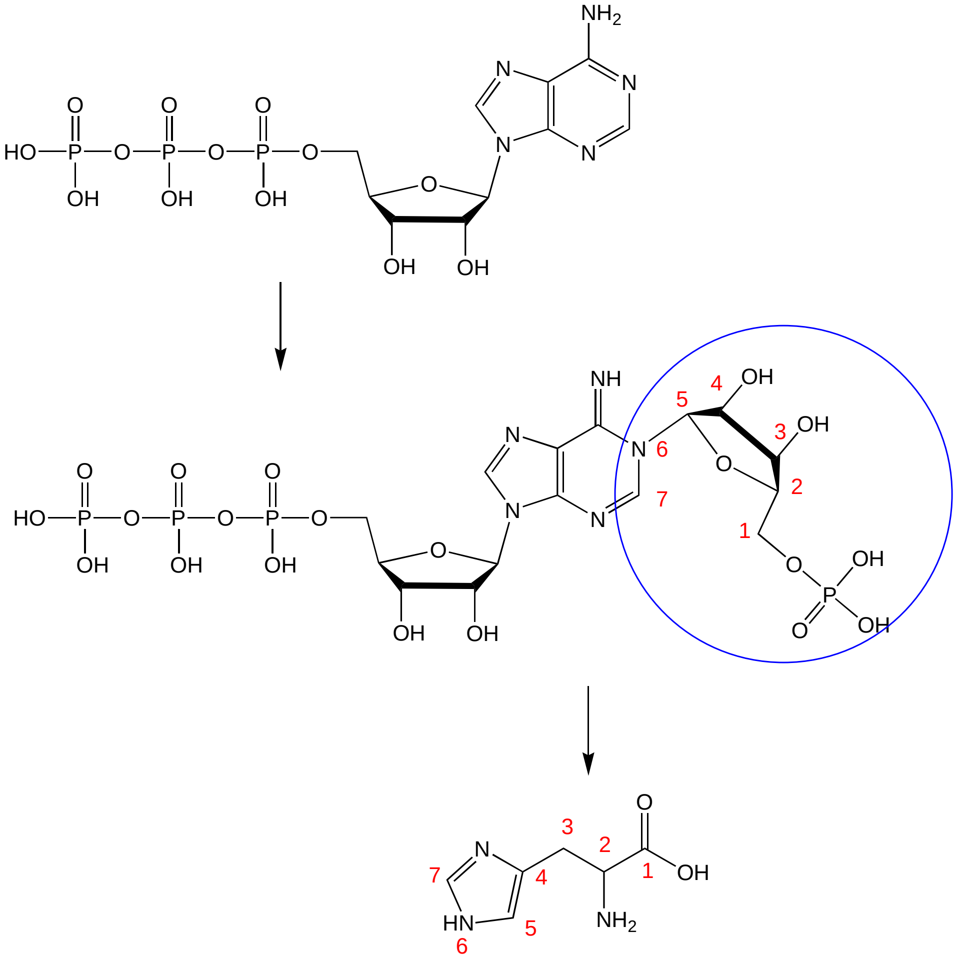
N-6 is sometimes also modified in DNA

The co-enzyme that facilitates adenosine methylation is S-adenosylmethionine (SAM)

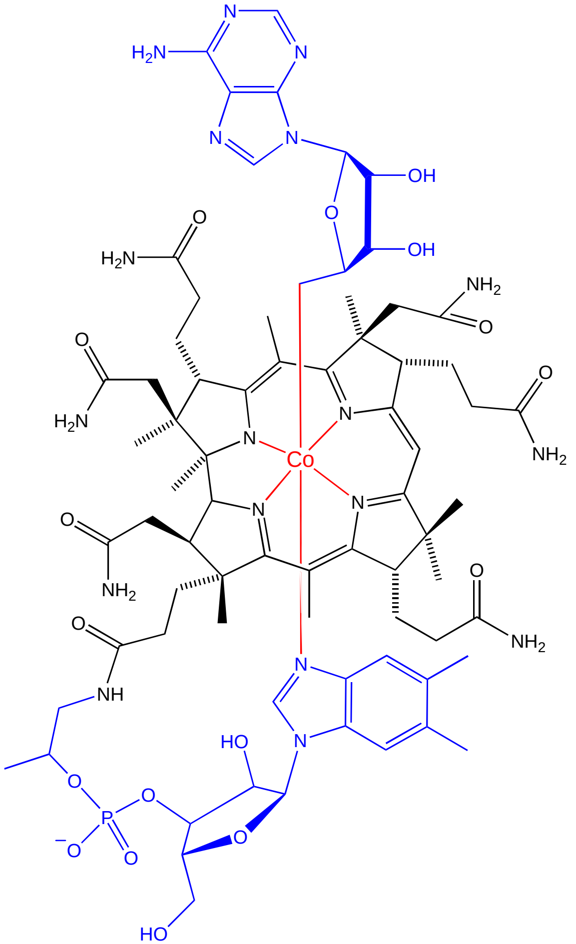


Adenosine is responsible for the biosynthesis of DNA from RNA

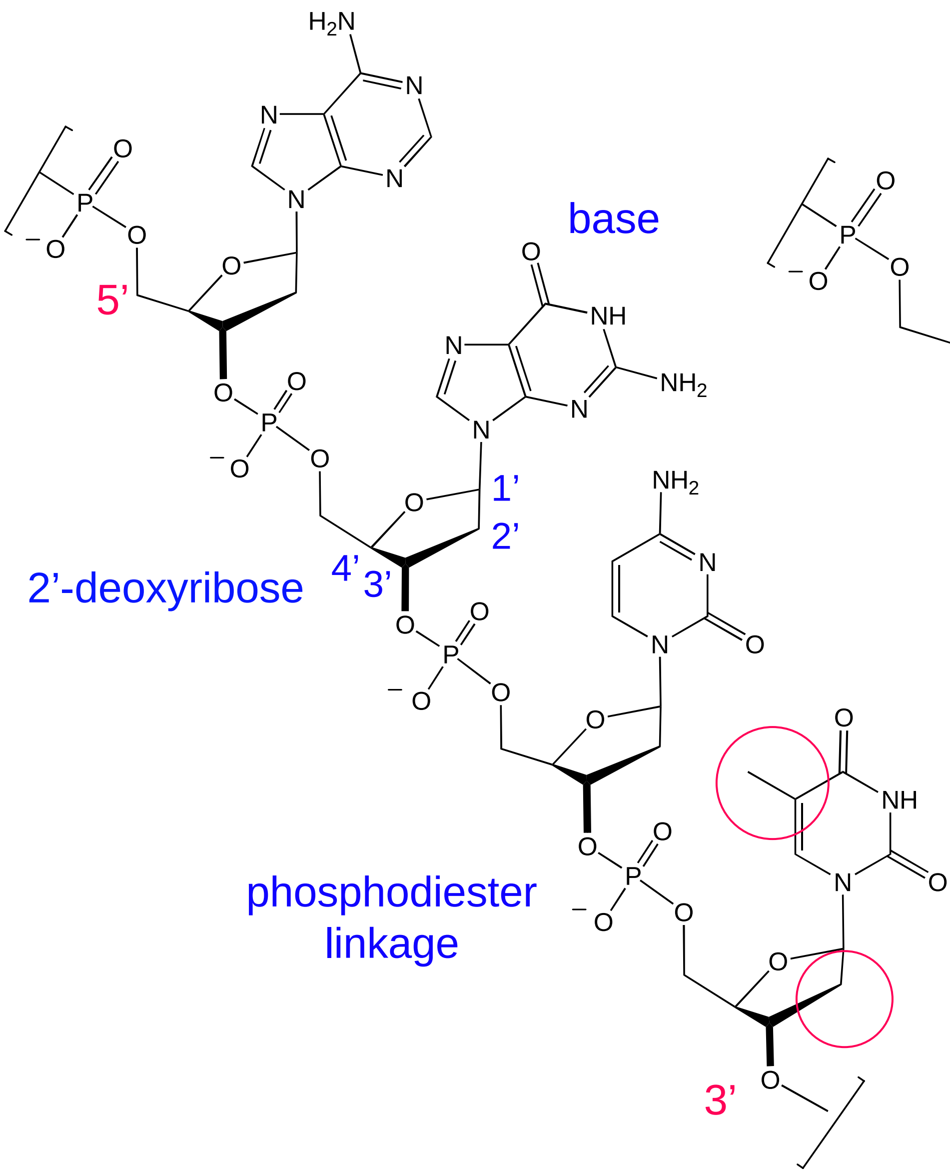
1) ATP is the biosynthetic precursor of histidine. C7 of histidine becomes the C5 methyl group of thymine



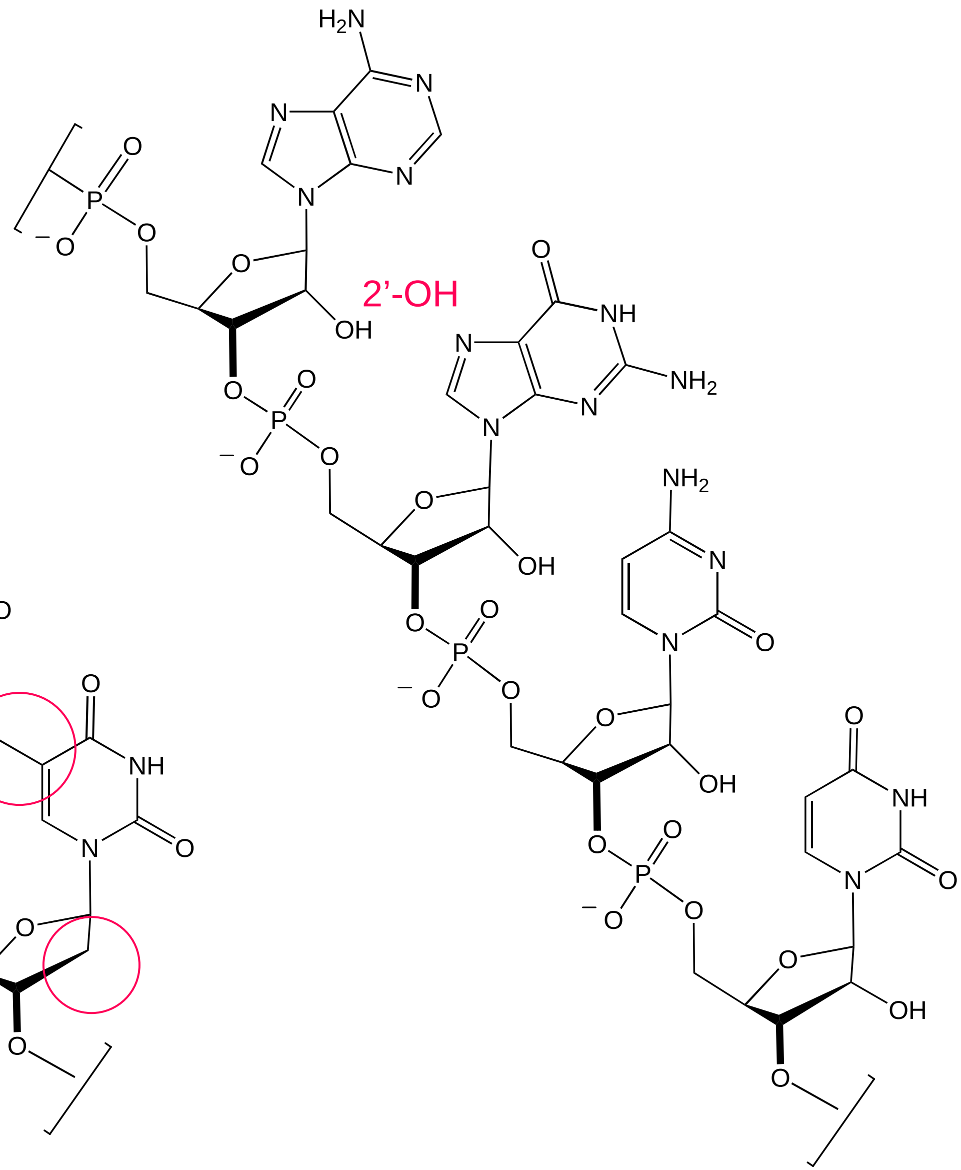
2) The coenzyme adenosylcobalamin participates in the reduction of the 2'-OH to 2'-H



Deoxyribonucleic acid (DNA)



Ribonucleic acid (RNA)



Why did DNA evolve from RNA?

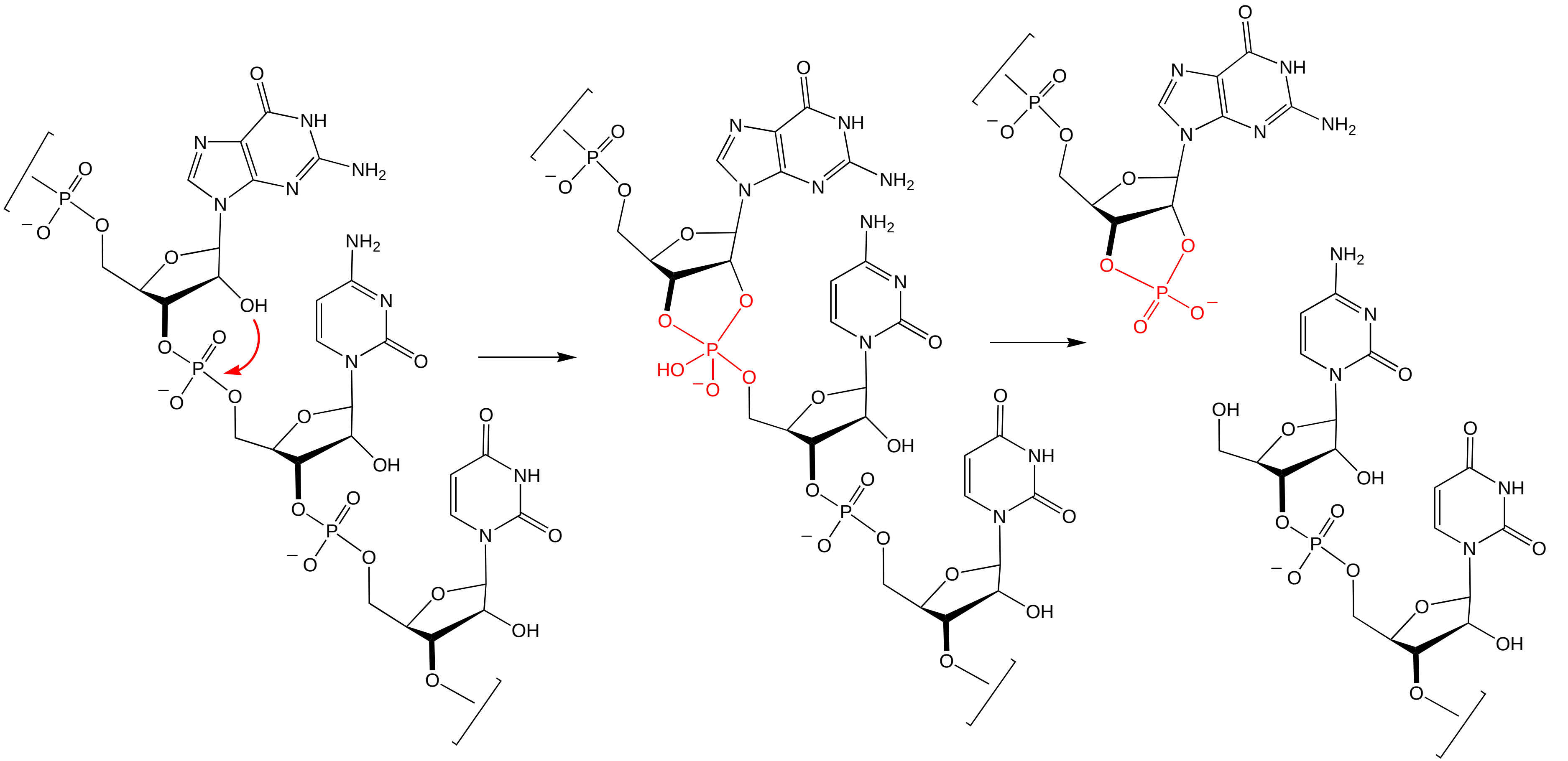
The 2’-OH in RNA participates in a relatively facile degradation pathway to monomers

Removal of the 2’-OH eliminates this degradation pathway - As a result DNA is a more stable polymer.

Cytosine is prone to spontaneous deamination in water to yield uracil

Addition of the C5 methyl provides a mechanism for efficient repair of spontaneous cytosine to uracil mutations since uracil can be distinguished from thymine.

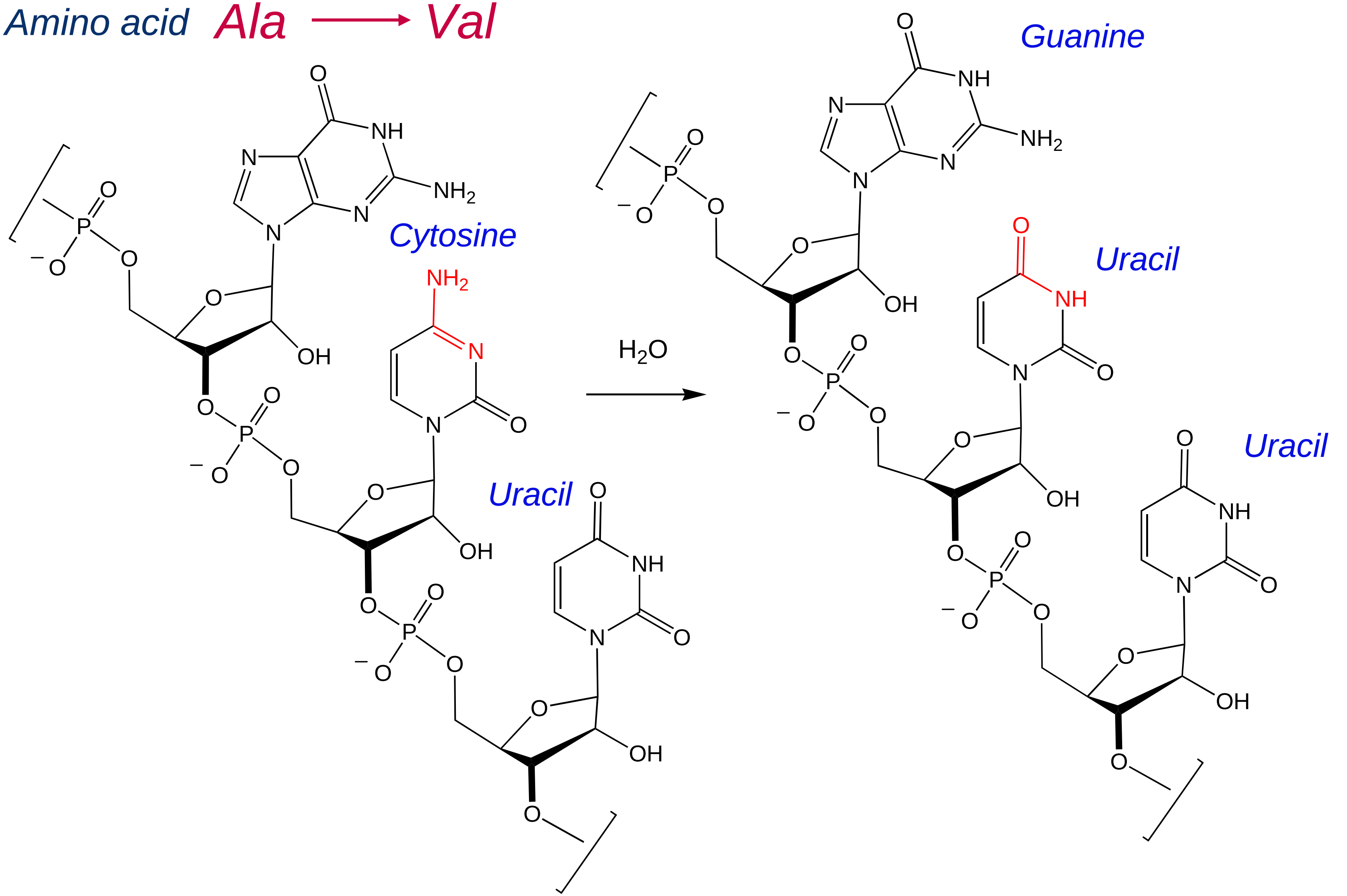
2’-OH mediated RNA degradation



Deamination of cytosine to uracil in RNA is mutagenic

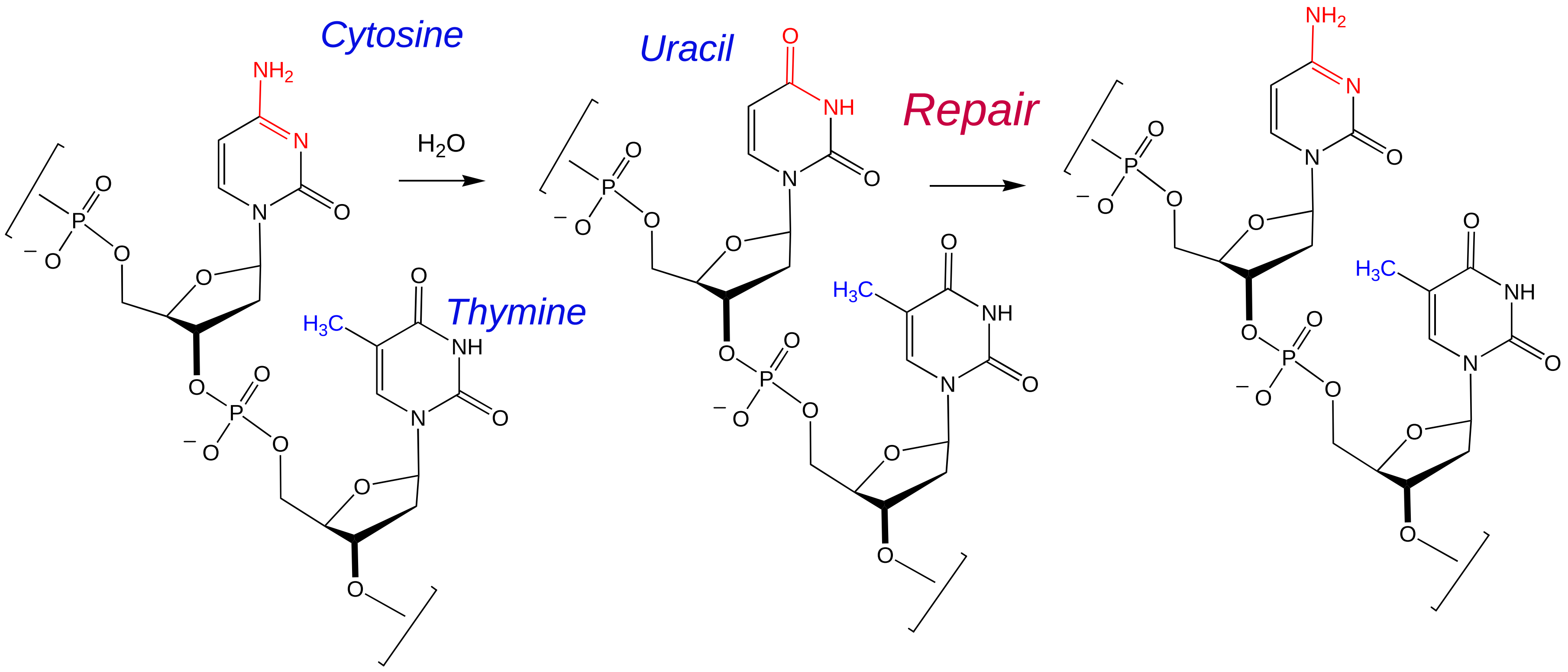
Codon ***GCU*** → ***GUU***

Amino acid ***Ala*** → ***Val***



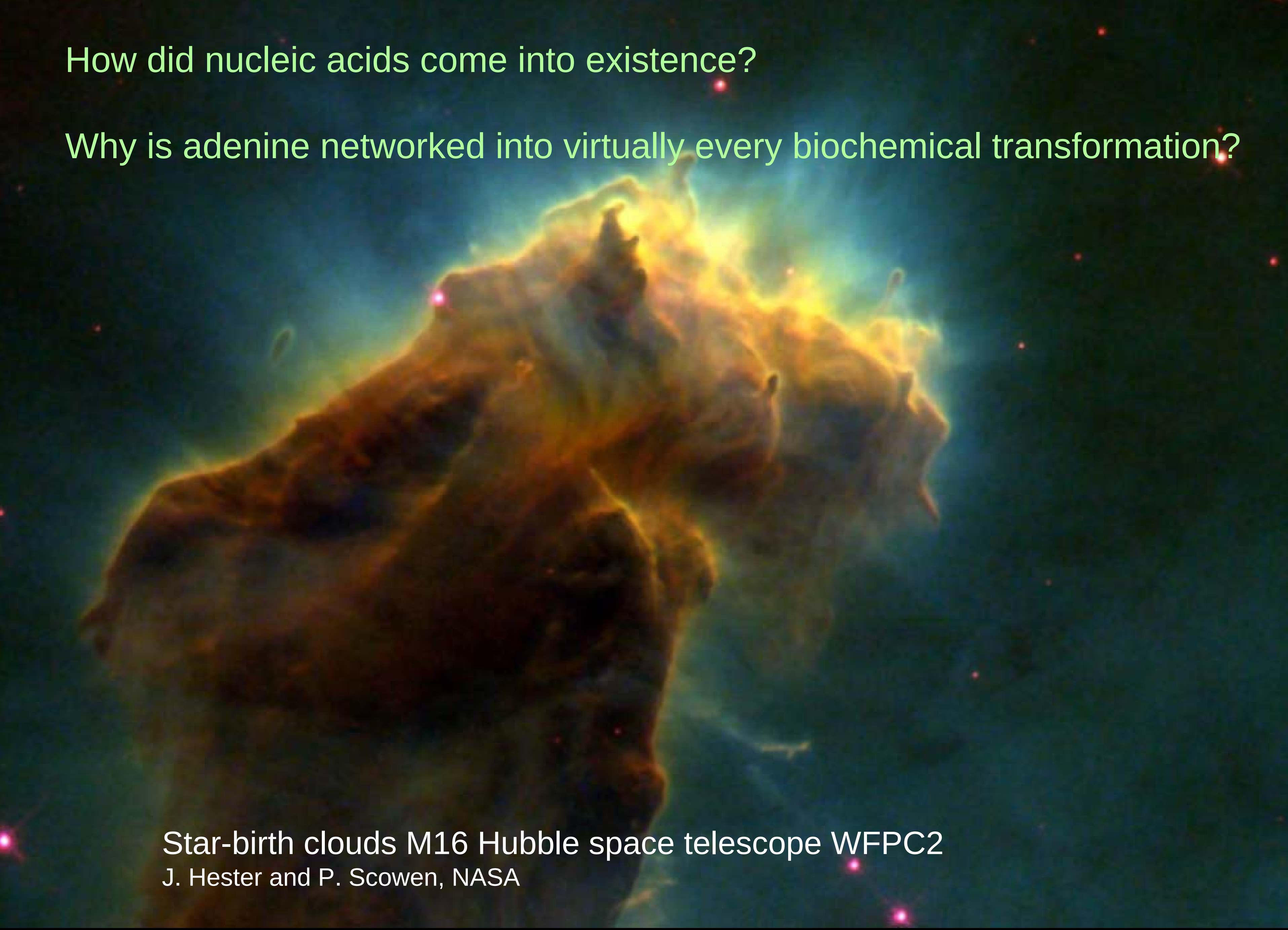
Deamination of cytosine to uracil in DNA is repaired

Repair enzymes distinguish uracil from thymine



How did nucleic acids come into existence?

Why is adenine networked into virtually every biochemical transformation?



Star-birth clouds M16 Hubble space telescope WFPC2
J. Hester and P. Scowen, NASA

Stellar generation of the elements

CNO cycle   T ~ 3 x 10^7K

$$^{12}_{6}C + {}^{1}_{1}H \longrightarrow {}^{13}_{7}N + \gamma \longrightarrow {}^{13}_{6}C + e^+$$

$$^{13}_{6}C + {}^{1}_{1}H \longrightarrow {}^{14}_{7}N + \gamma$$

$$^{14}_{7}N + {}^{1}_{1}H \longrightarrow {}^{15}_{8}O + \gamma \longrightarrow {}^{15}_{7}N + e^+$$

$$^{15}_{7}N + {}^{1}H \longrightarrow {}^{12}_{6}C + {}^{4}He$$

The reaction cycle is autocatalytic, which is also a fundamental characteristic of biological processes

Triple alpha process (Red-Giant Core)

$T \sim 10^8 K$

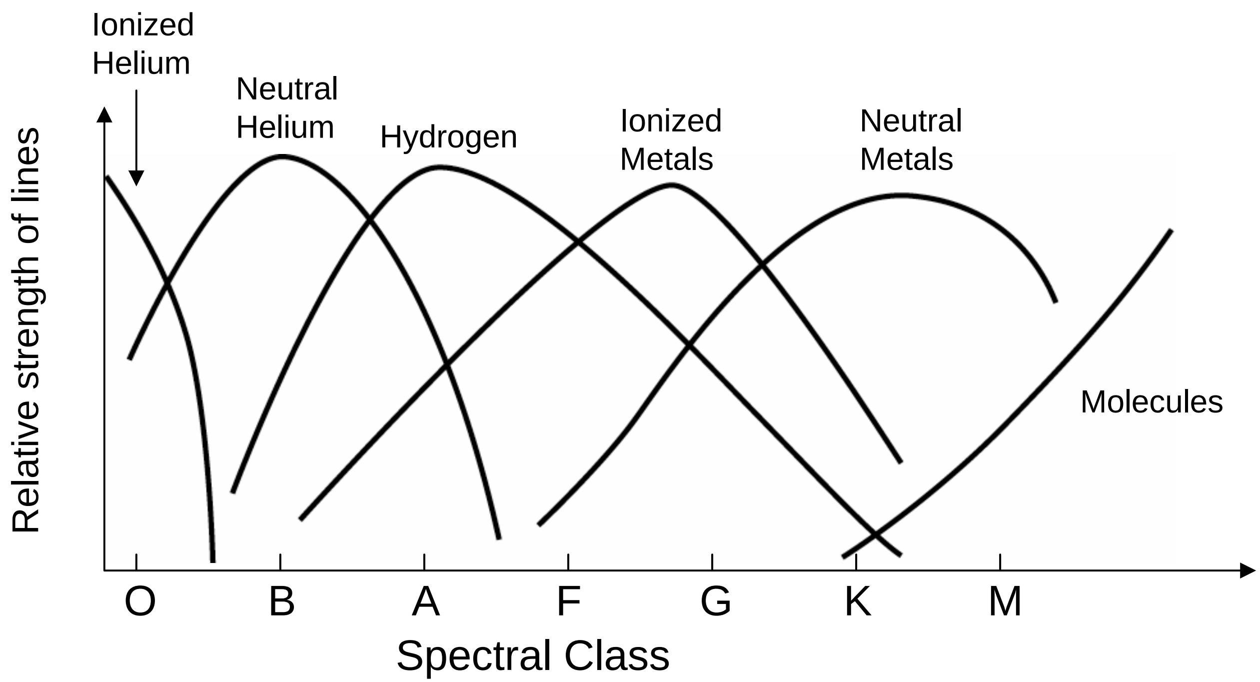
$$2\ {}^{4}_{2}He \rightleftharpoons {}^{8}_{4}Be \qquad T_{1/2} = 7 \times 10^{-17}\ s$$

$$ {}^{8}_{4}Be + {}^{4}_{2}He \rightleftharpoons {}^{12}_{6}C^{*} \longrightarrow {}^{12}_{6}C + 2\gamma$$

Harvard Spectral Classification

Overseas [Bulletin/Broadcast]: A Flash! Godzilla Kills Mothra! (Rodan Named Successor)

http://www.shef.ac.uk/physics/people/pacrowther/obafgkmrns.html



Harvard Spectral Classification created by Annie Jump Canon 1918-1924

www.shef.ac.uk/physics/people/pacrowther/spectral_classification.html

O	B	A	F	G	K	M R N
Violet	Blue	Blue	Blue-white	Yellow-white	Orange-Red	Red
> 28,000K	10,000-28,000K	7,500-10,000K	6,000-7,500K	5,000-6,000K	3,500-5,000K	< 3,500K

Low temperature processes

$T \sim 10^4$K Hydrogen exists as H^+

$T \sim 5 \times 10^3$K Hydrogen exists as H•

T ~ 5000 - 6000K CH star

G-type giant (G5 to K5) with strong molecular bands for CH

T < 3500 K Carbon star (red giant)

Spectra indicate abundance of C2 and CN as well as other carbon compounds

T > 1000 K C star (giant)(R and N in Harvard Classification)

Spectra show C2, CN, and CH, but no metallic oxides

T > 1000 K CN strong star (giant)

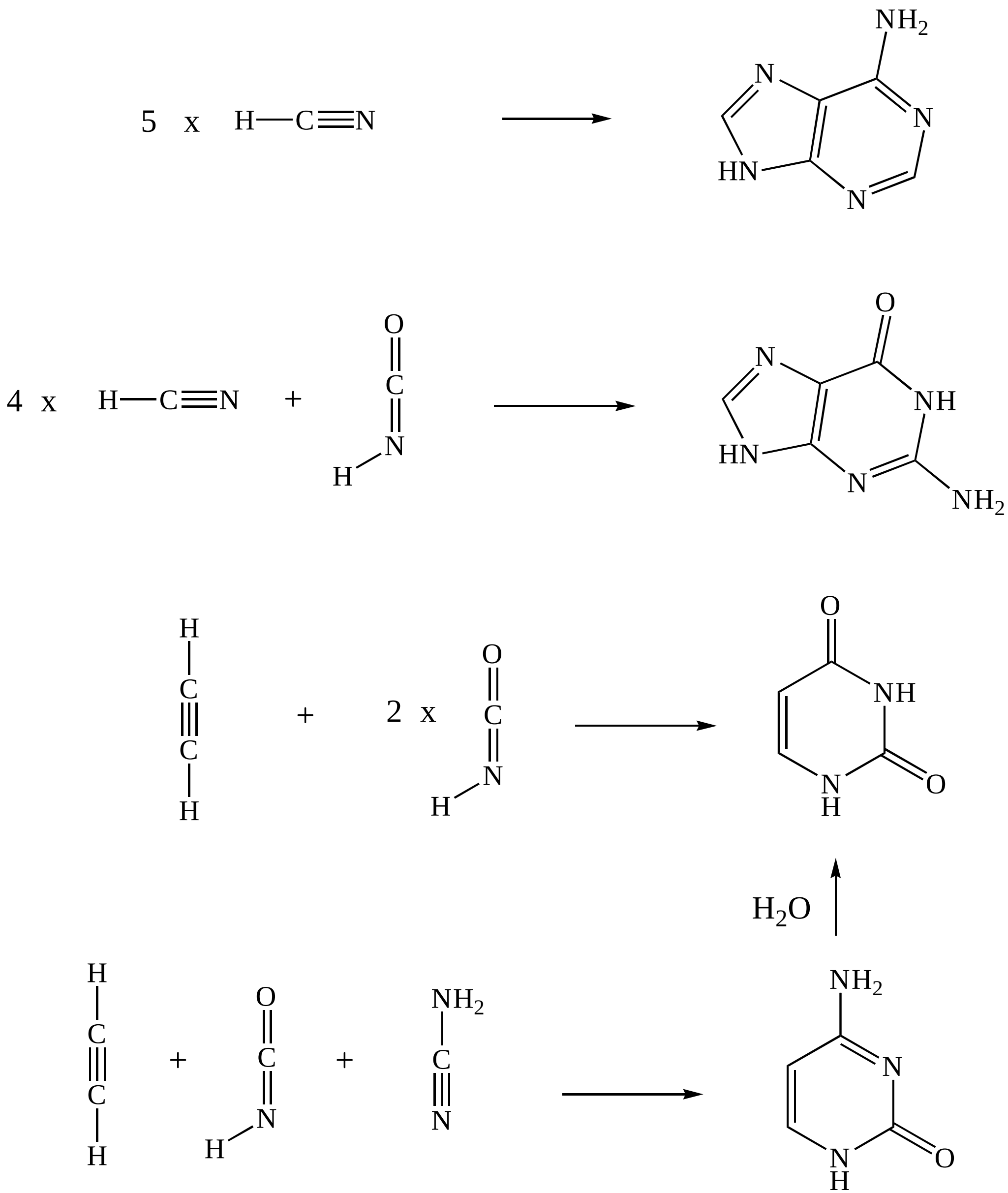
Spectra for CN as well as metallic oxides

http://www.site.uottawa.ca:4321/astronomy/index.html#latestar

Orion Nebula where new stellar systems are being created; many are surrounded by clouds of HCN or H_2CO. NASA/JPL-Caltech

Interstellar CHNO species RNA bases

The nucleic acid heterocyclic bases are formally polymers of interstellar CHNO species

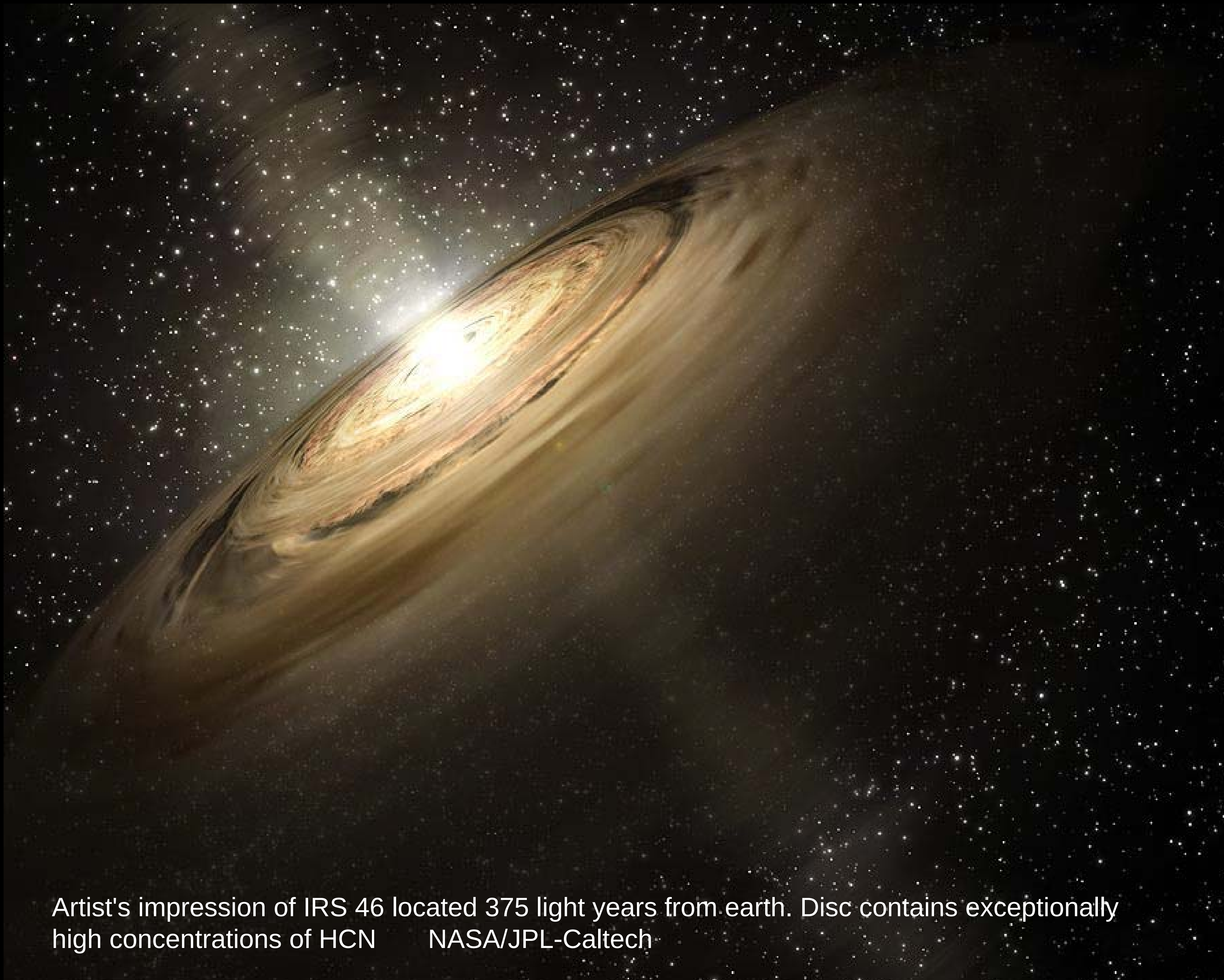


Adenine

Guanine

Uracil

Cytosine



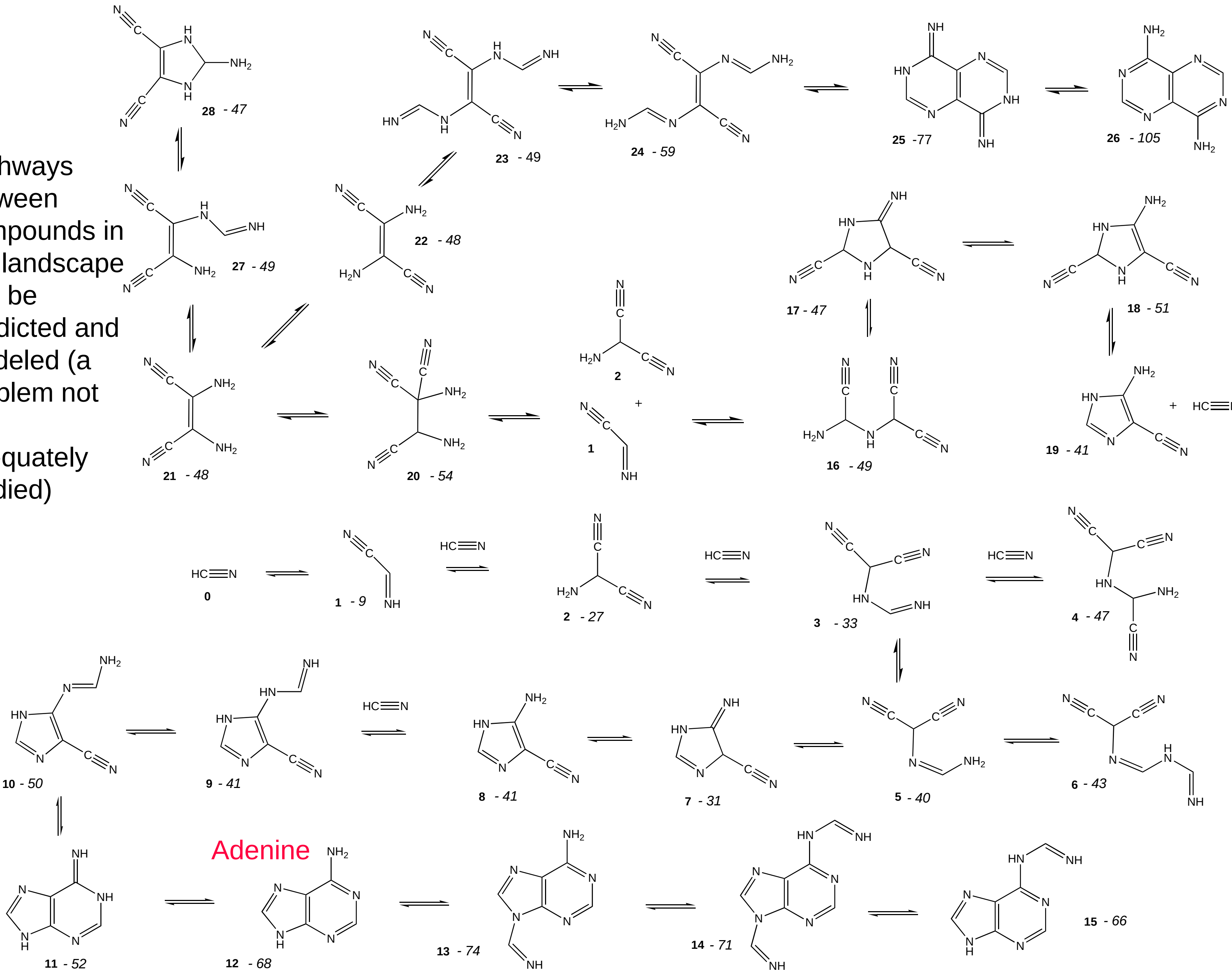
Artist's impression of IRS 46 located 375 light years from earth. Disc contains exceptionally high concentrations of HCN NASA/JPL-Caltech

HCN polymerization

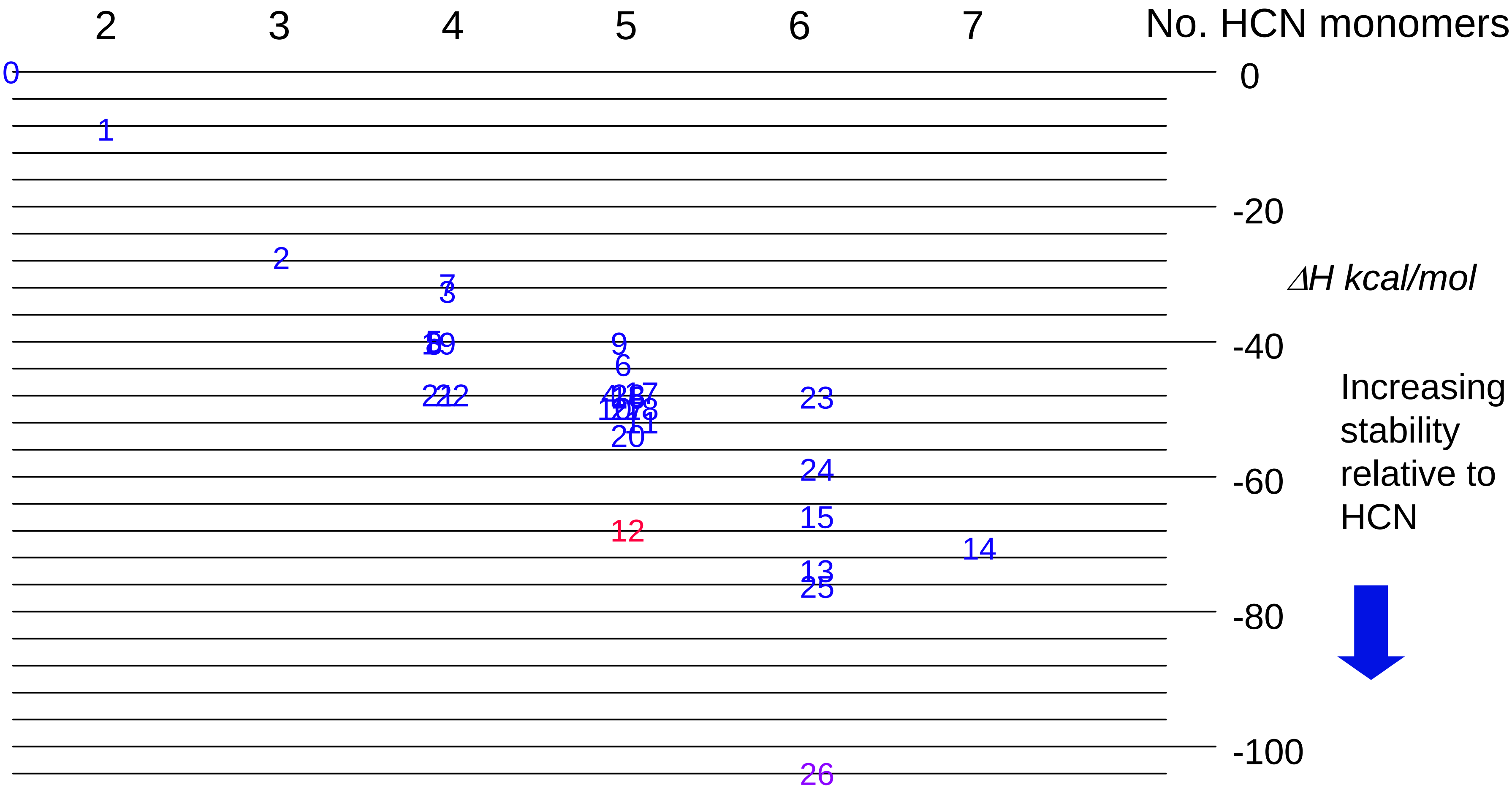
At sufficiently high concentrations HCN will polymerize to a complex mixture of products. The products and pathways will be influenced by the presence of other elements and catalysts as well as temperature and pressure.

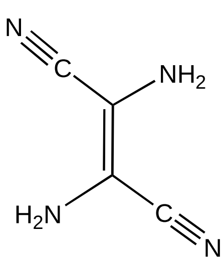
Portion of hypothetical HCN polymerization landscape

Pathways between compounds in the landscape can be predicted and modeled (a problem not yet adequately studied)

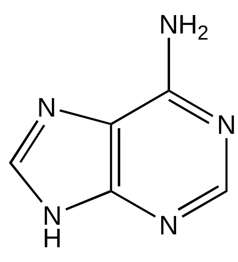


HCN polymerization landscape

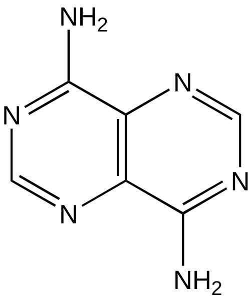




22



12 adenine



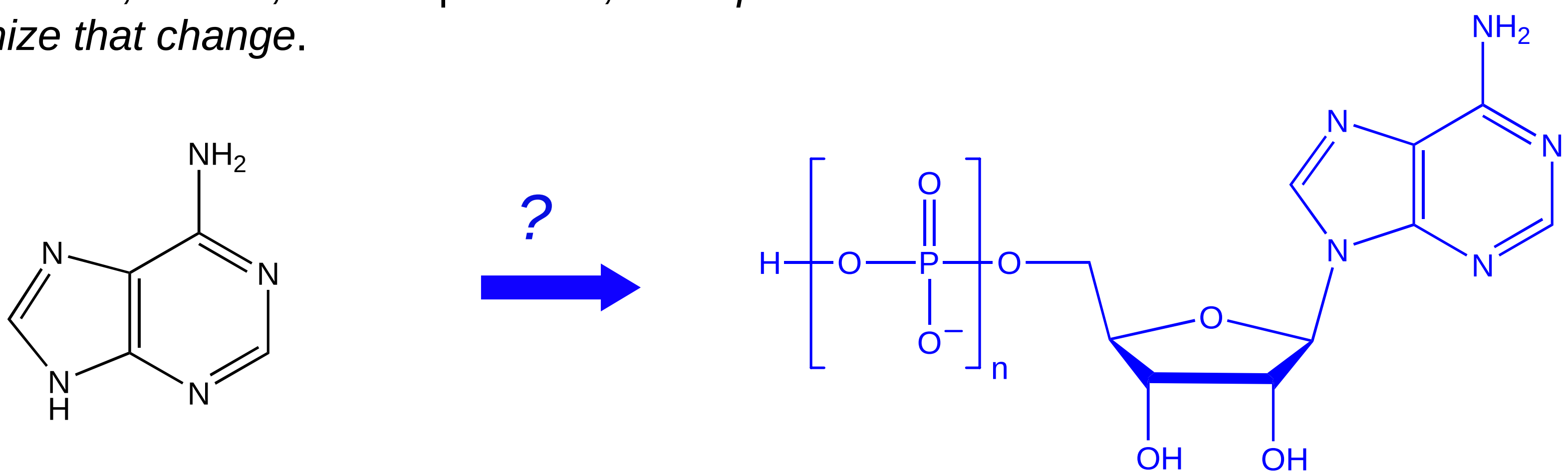
26

The path from adenine to adenosine phosphates

Adenine will be pulled from the HCN domain by interactions with other molecular domains (HCO and polyphosphate). According to *Le Chatelier's* principle other molecules in equilibrium with adenine will be pulled from the landscape as a result. In essence they become fuel for adenine.

Le Chatelier's principle

If a chemical system at equilibrium experiences a change in concentration*,* temperature*,volume, or total* pressure*; the equilibrium will shift in order to minimize that change*.



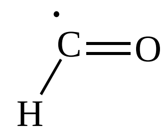
Adenine

AMP, ADP, ATP, Ap_nA

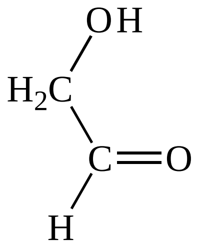
Interstellar CHO species

COH	formyl radical
HCO^+	formyl ion
HOC^+	hydroxymethylidynium
HOCO	hydroxyformyl radical
$HOCO^+$	protonated carbon dioxide
H_2CO	formaldehyde
HCOOH	formic acid
CH_3O	methoxy radical
H_2COH^+	protonated formaldehyde
CH_3OH	methanol
H_2CCO	ketene
CH_3CHO	acetaldehyde
CH_2CHOH	vinyl alcohol
c-C_2H_4O	ethylene oxide
$HCOOCH_3$	methyl formate
CH_3COOH	acetic acid
CH_2OHCHO	glycolaldehyde
CH_3CH_2OH	ethanol
CH_3OCH_3	dimethyl ether
HC_2CHO	propynal
$(H_2CO)_3$	trioxane
CH_3COCH_3	acetone

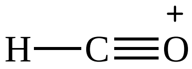
Formyl radical



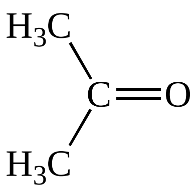
Glycolaldehyde



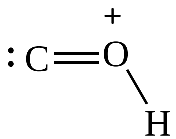
Formyl ion



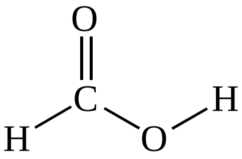
Acetone



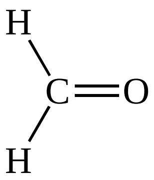
Hydroxymethylidynium



Formic acid

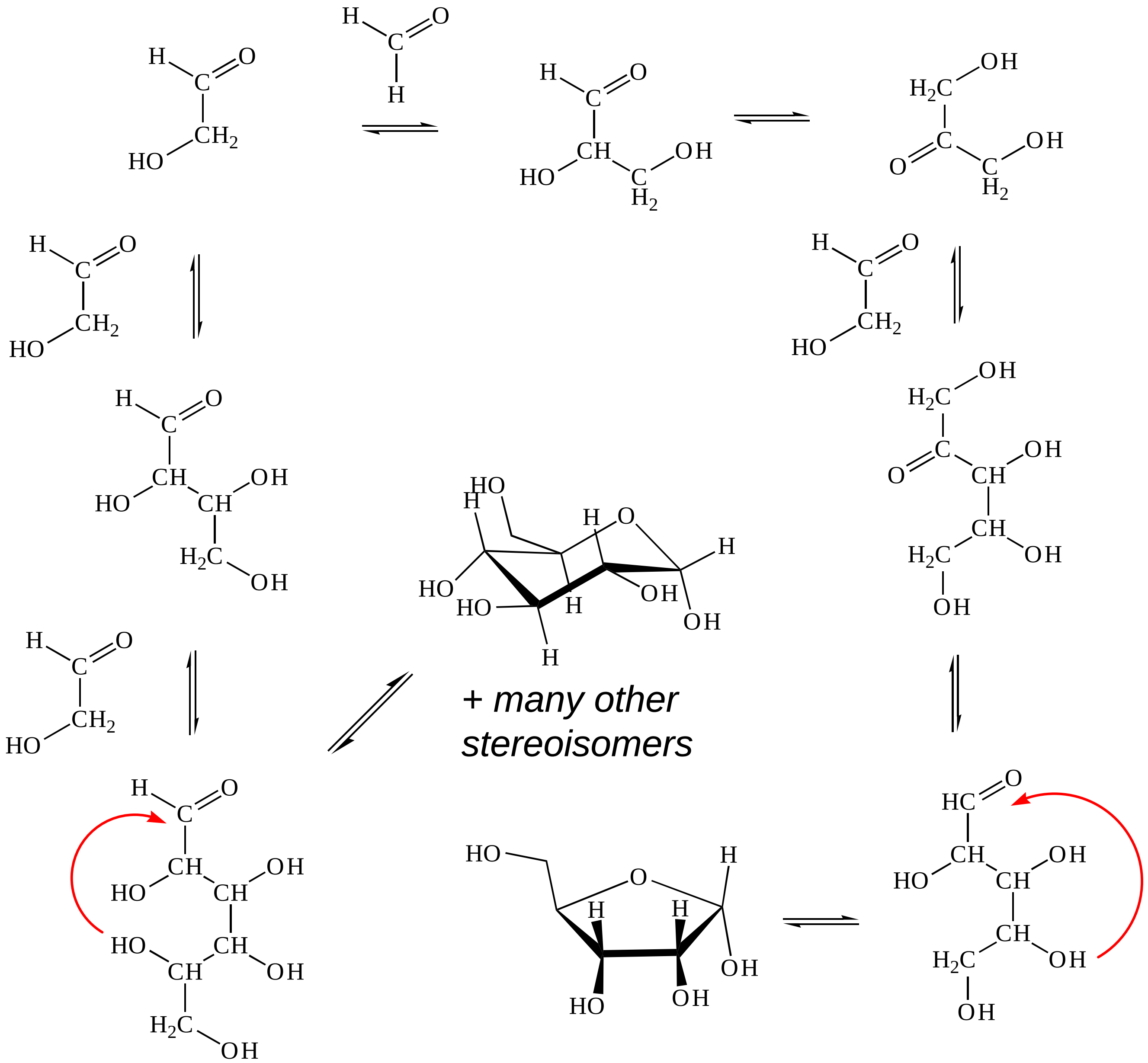


Formaldehyde

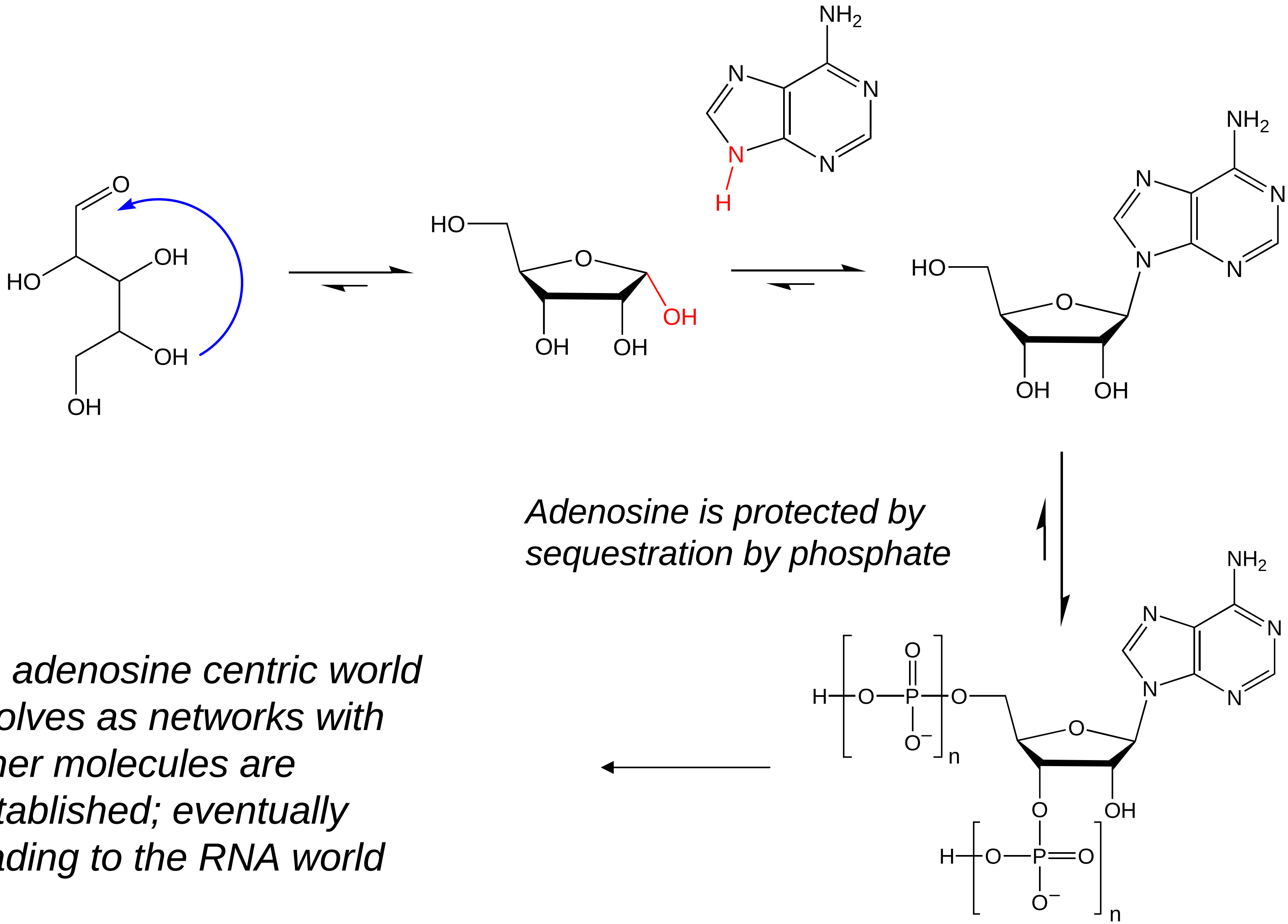


From **Constants for molecules of astrophysical interest in the gas phase: photodissociation, microwave and infrared spectra.** J. Crovisier Observatoire de Paris, Section de Meudon F-92195 Meudon, France http://www.lesia.obspm.fr/~crovisier/basemole/

Reaction pathways to monosaccharides



Path to an adenosine centric world



An adenosine centric world evolves as networks with other molecules are established; eventually leading to the RNA world

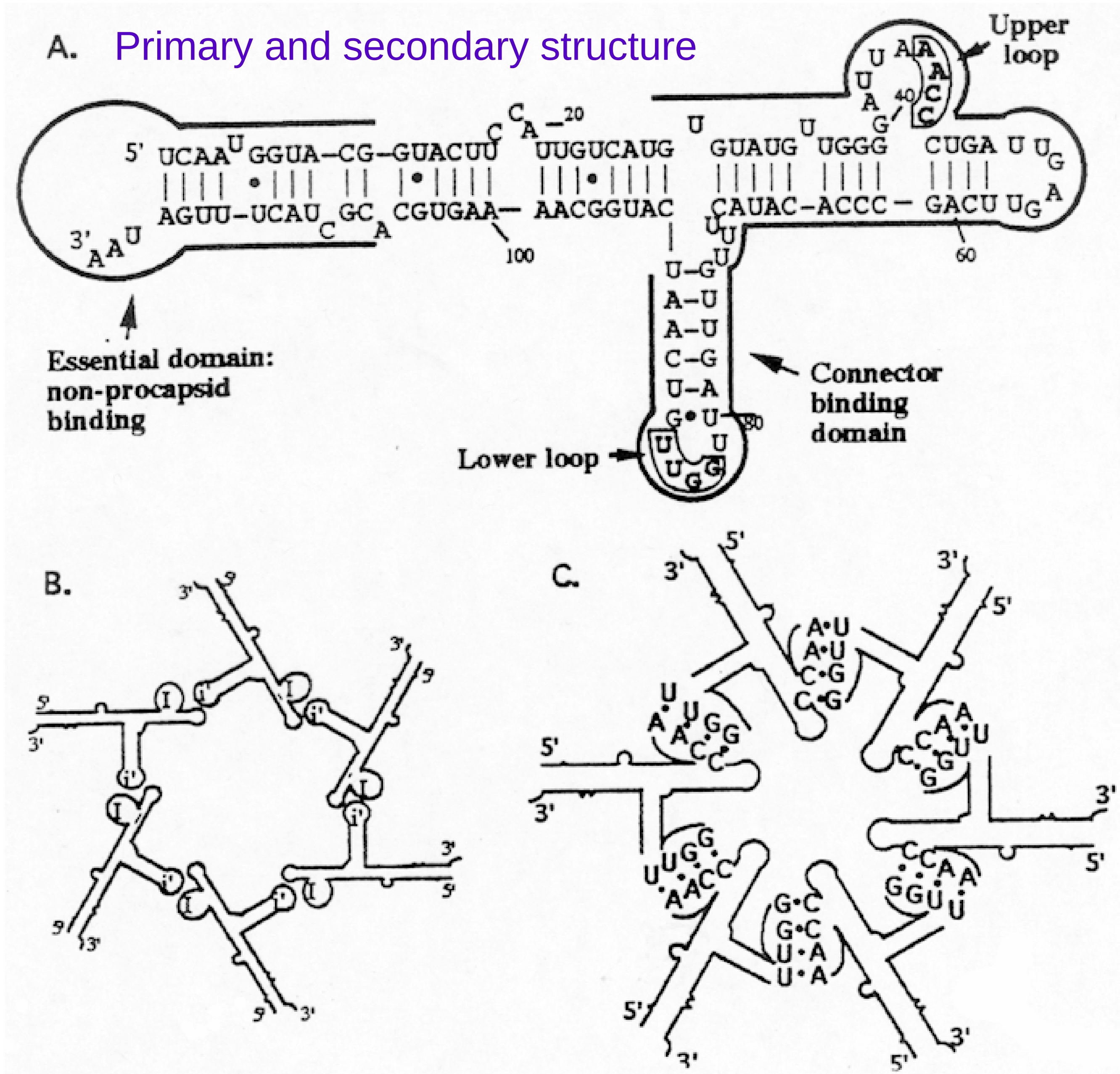
The path from adenine continues to unfold

1. Adenine captures monosaccharides
2. Adenosine recruits phosphate and other small molecules in the environment
3. Coenzymes evolve to aid adenosine survival
4. The RNA world is established
5. RNA and peptides co-evolve
6. The translation system evolves
7. DNA is established and co-evolves with RNA and proteins
8. Higher order organisms evolve

RNA can assemble into higher order structures

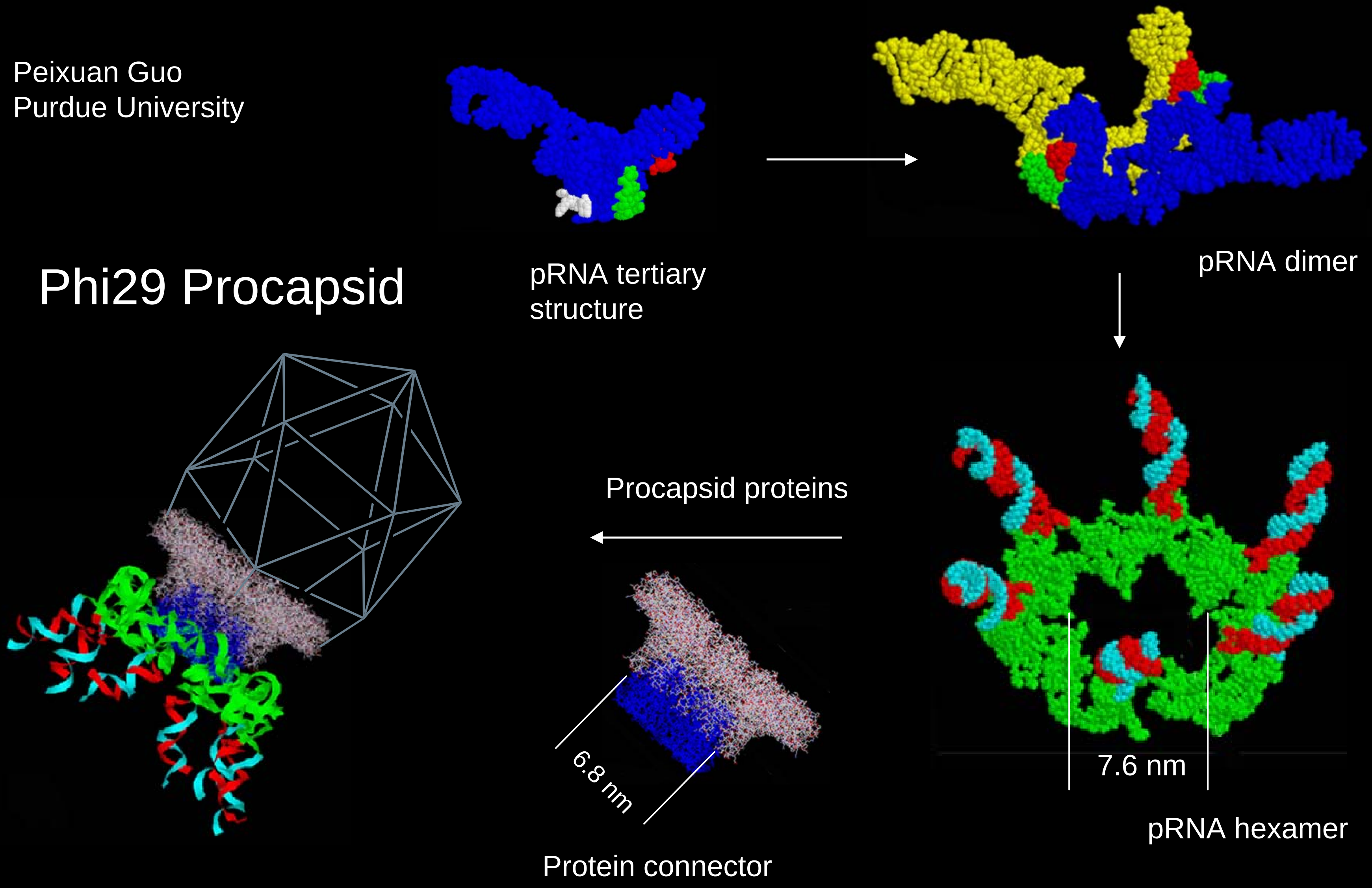
pRNA hexamer from the bacteriophage phi29

Peixuan Guo
Purdue University



pRNA hexamer assembles into a molecular motor for packaging DNA into the phi29 procapsid

Peixuan Guo
Purdue University



The path from adenine continues to unfold - 2

1. Adenine captures monosaccharides
2. Adenosine recruits phosphate and other small molecules in the environment
3. Coenzymes evolve to aid adenosine survival
4. The RNA world is established
5. RNA and peptides co-evolve
6. The translation system evolves
7. DNA is established and co-evolves with RNA and proteins
8. Higher order organisms evolve
9. Humans are recruited to develop adenine derived materials

Why nucleic acids are important materials for nanotechnology?

1. High sensitivity identification of any organism by nucleic acid sequence will be most efficiently accomplished with nanoscale devices.

2. High-throughput sequencing of the human genome for genetic diseases will only be economical if done with nanoscale technologies.

3. Nanoscale devices carrying nucleic acid components will be needed for nucleic acid based drugs and gene therapies.

4. Synthetic modified nucleic acids constitute unique materials for self-assembly of multi-functional nanoscale devices.

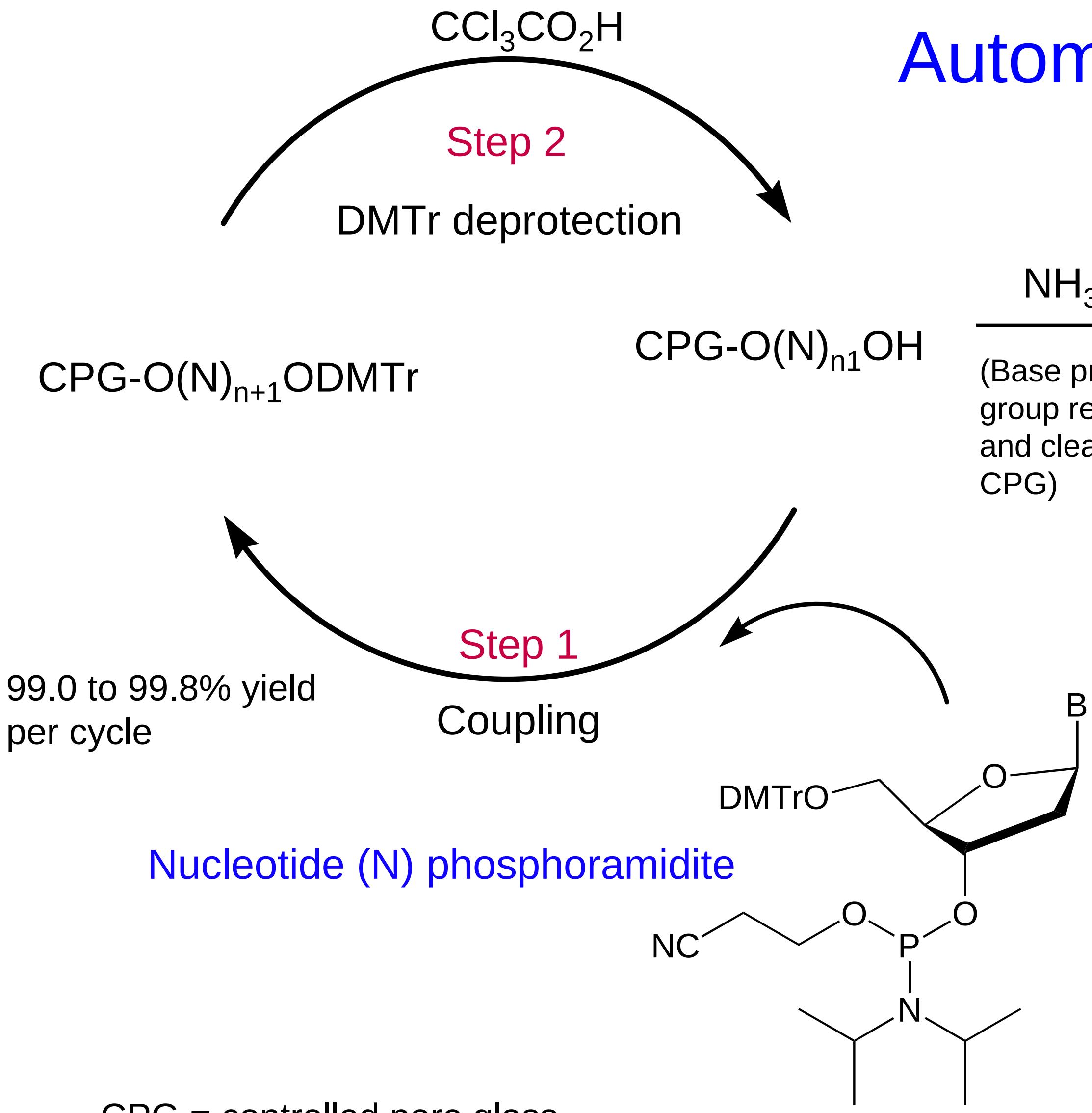
5. Nucleic acids can be used as components in nano-scale computing strategies.

A dominate theme in 21st century technology will be the merging of biology and material science.

Advantages of Synthetic Nucleic Acids as Engineering Materials

- Automated high-throughput synthesis is fast and inexpensive
- Chemically stable and biologically compatible
- Sequence-specific self-assembly into stable duplex and higher order structures
- Probes and functional groups can be attached at many different sites
- Structure can be extensively modified yet retain self-assembly properties
- Sequences can be attached to virtually any material
- Electrical and optical properties can be tuned

Automated DNA synthesis

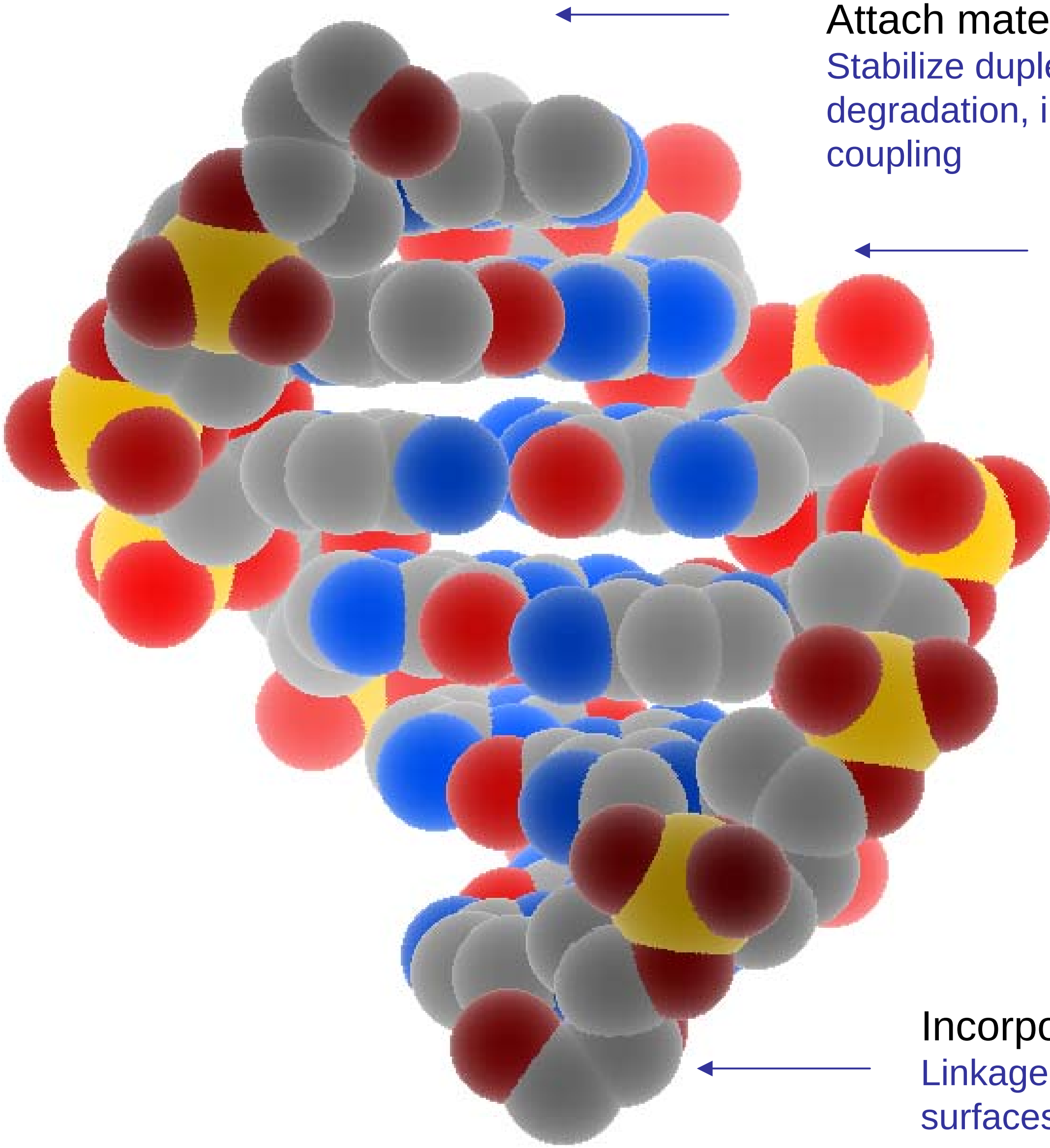


CPG = controlled pore glass

DMTr = dimethoxytrityl

B = protected adenine, cytosine, guanine,
thymine, or modified base

384 oligonucleotide parallel synthesizer



Attach materials at 3’ and 5’ ends

Stabilize duplex, protect against nuclease degradation, interface for electronic coupling

Incorporate modified bases

Modulation of biochemical, and/or electronic properties

Modify backbone

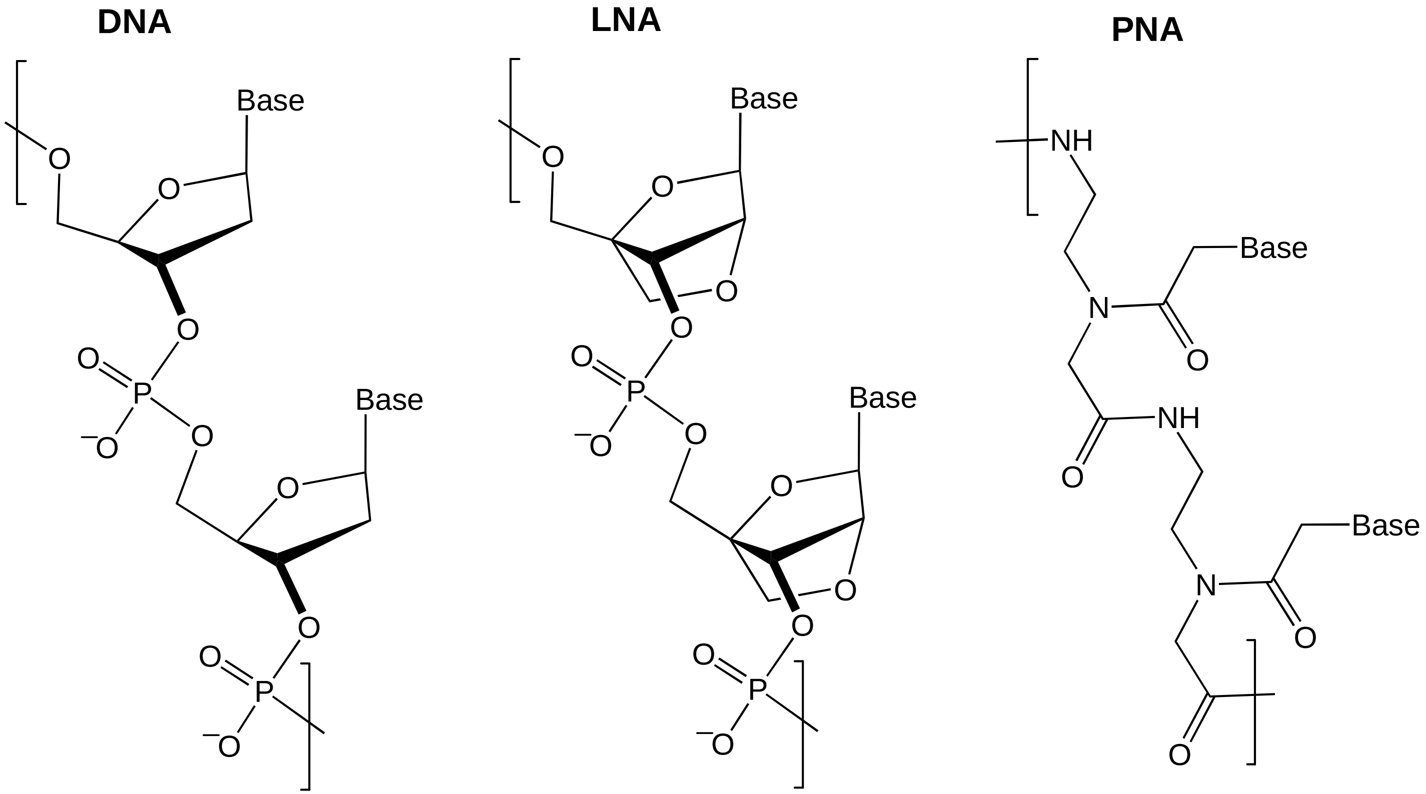
PS; PNA; LNA

Increased thermo-stability, resistance to biological degradation

Incorporate 3’ or 5’ tethers

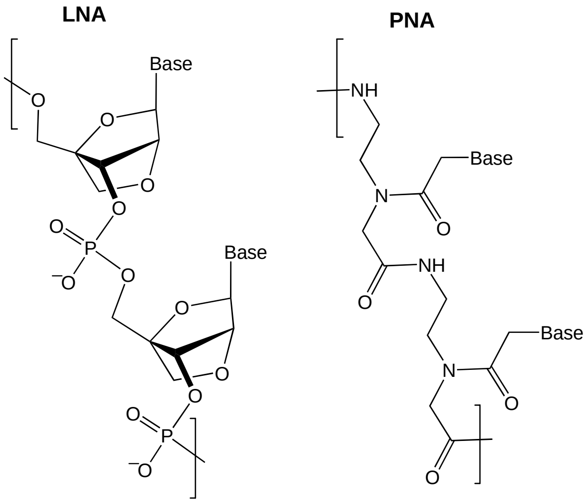
Linkage for attachment to surfaces or reporter groups

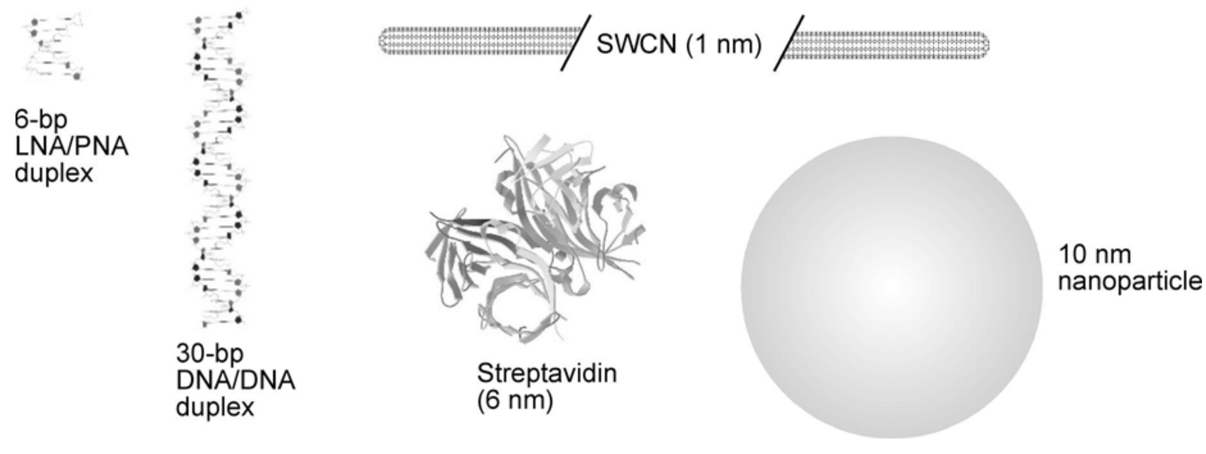
Chemists have created many other nucleic acid-like polymers that self-assemble

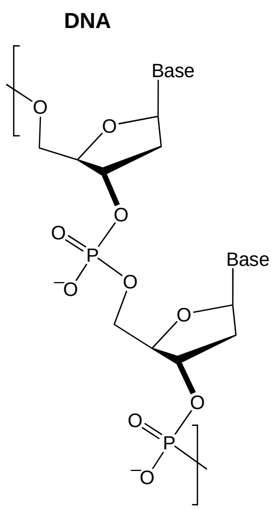


Locked nucleic acids (LNA) and peptide nucleic acids (PNA) form duplex structures that are considerably more stable than DNA and RNA

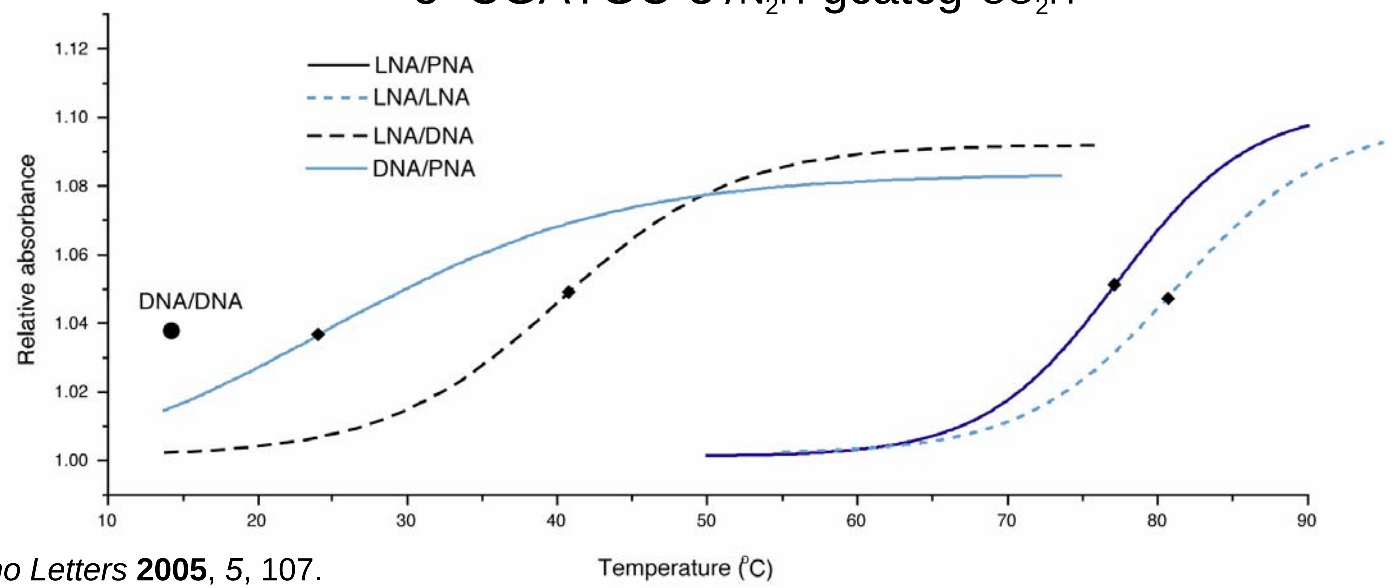
Alternative nucleic acid analogs for programmable assembly: Hybridization of LNA to PNA







5’-CGATGC-3’/N_2H-gcatcg-CO_2H



P. Ng and D. E. Bergstrom, *Nano Letters* **2005**, *5*, 107.

Re-engineering nucleic acids and nucleic acid components: future opportunities for scientists and engineers

Biologicals

- Tools for investigations of enzyme mechanism
- Tools for structural studies on protein-DNA complexes
- Components of nucleic acid based assay systems
- Diagnostic probes
- Tools for drug target validation
- Therapeutics
  - Antisense oligonucleotides
  - Decoy oligonucleotides
  - Modified oligonucleotides for gene correction
  - Sequence specific prodrug activators

Materials

- Self-assembling scaffolds for nanoscale devices
- Assemblers for reaction discovery
- Computing

Questions & Answers