RNA Chemistry & Structure

Roles of molecules

	RNA	DNA	proteins
genetic information	yes	yes	
catalysis	yes		yes
regulation / interactions	yes	yes	yes
structure	usually single stranded	usually duplex	lots

Catalysis and binding

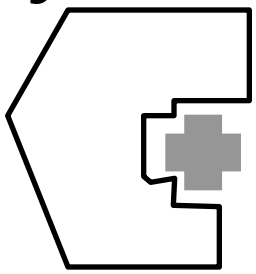
Catalysis

- proteins – classic enzymes
- RNA – less common, but well established
  (ribosome, hammerhead, ..)

Specific binding

- proteins
  - bind substrates, ligands, DNA, RNA
- DNA
  - sequence specific binding – to proteins, RNA, DNA
- RNA
  - same as DNA +
  - specific catalysis implies specific recognition
  - switches and regulators

Recognition / binding specificity



Protein view – via evolution

- protein scaffold / framework positions groups
- in binding / reactive region specific groups interact
- lots of chemical groups to choose from (20 amino acids)

DNA – not thought of in these terms

- some specificity
  - regulatory binding proteins are sequence specific

RNA

- sequence specificity for binding proteins
- RNAzymes, aptamers, selex
- binding of arbitrary small molecules

Structure

DNA

- mostly thought of as double helix

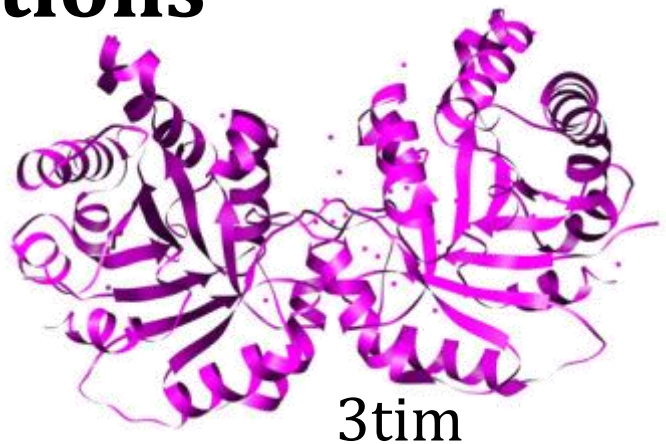
Protein (simple dogma)

- from a specific sequence to a well defined structure
- less often – floppy, unstructured, mobile, alternative folds

RNA

- does an RNA sequence fold up to a well defined structure ?
  - all possible RNA's ?
  - biological RNA's ?
  - some RNA's ?

Structure Expectations



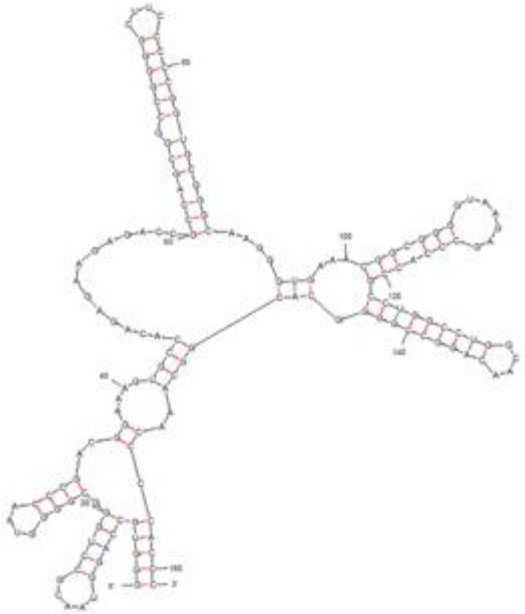
3tim

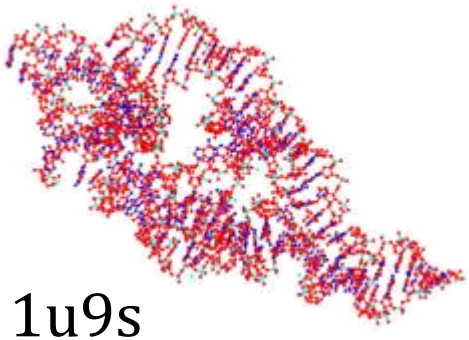
Protein

- usually 3D
- rarely secondary structure

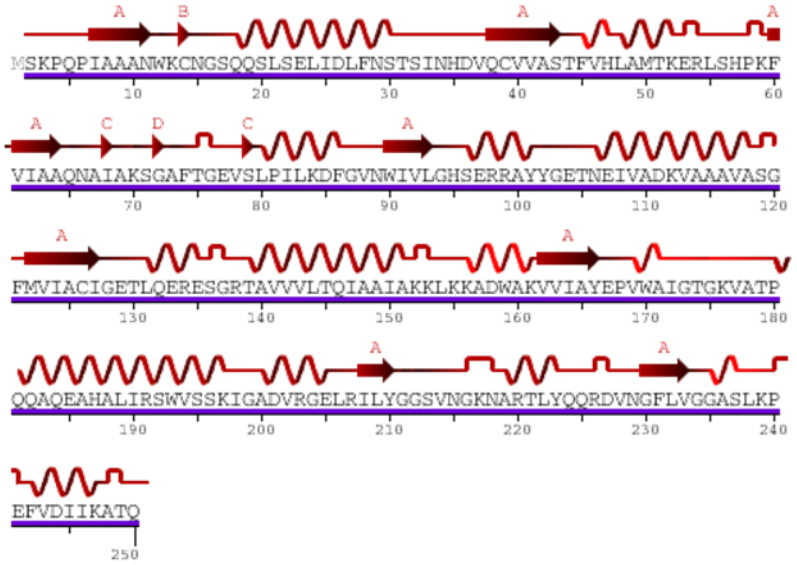
RNA

- usually secondary structure





1u9s



Structural Data

Proteins

- 10^5 or about 3×10^4 interesting ones

RNA

- 2.8×10^3 structures with some RNA
- 45 with RNA + DNA (no protein)
- 1058 with pure RNA

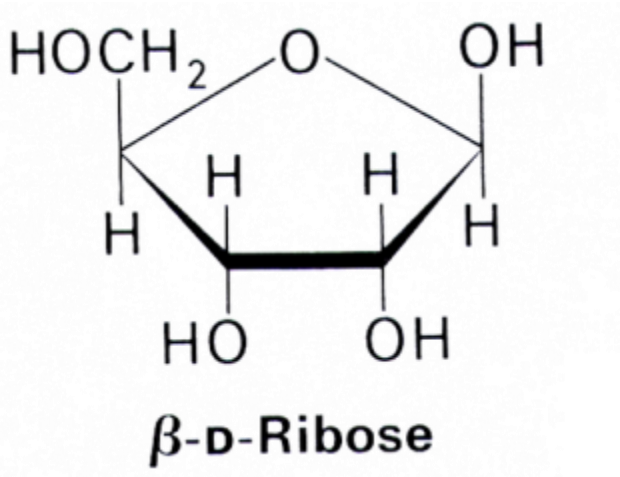
Determining structures

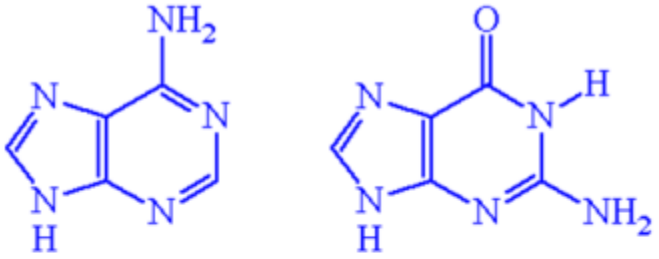
- general – RNA hard to handle (RNases)
- crystallography
- NMR
  - assignments very difficult (only 4 kinds of base)

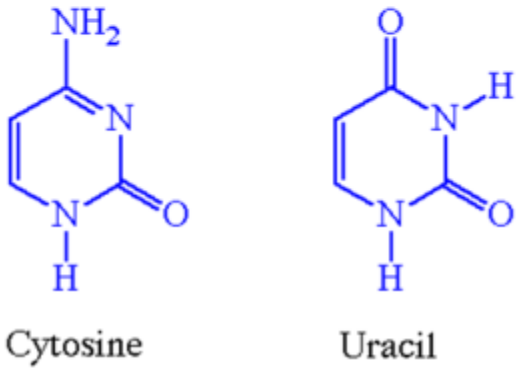
RNA structure

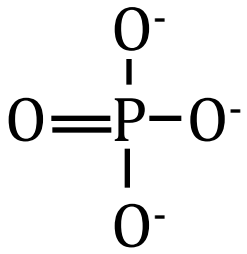
3 components

- ribose (sugar)
- phosphate (PO_4)
- base (nucleotide)





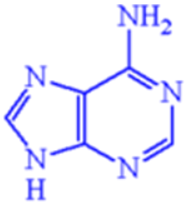




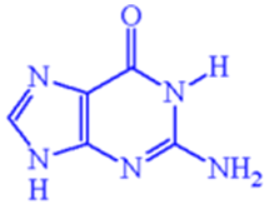
pictures from Stryer, Biochemistry, WH Freeman and Company, New York, 1981

RNA Bases

purines

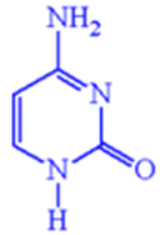


Adenine

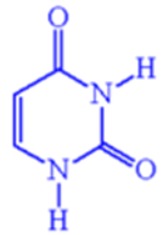


Guanine

pyrimidines



Cytosine

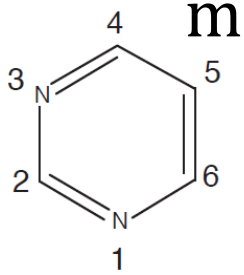


Uracil

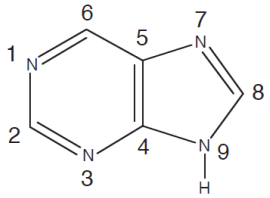
Are they like protein residues ?

- not classified by chemistry
- do they have interactions ?
  - yes

mother shapes



pyrimidine

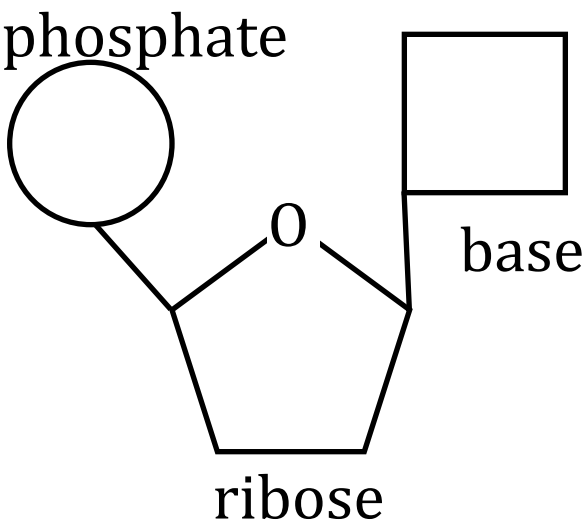


purine

- numbering not used much
- putting pieces together…

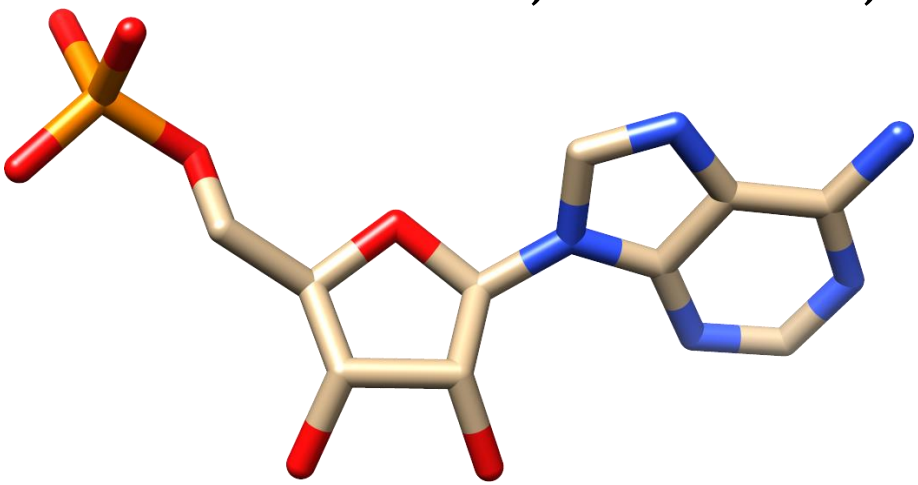
pictures from Stryer, Biochemistry, WH Freeman and Company, New York, 1981

RNA structure

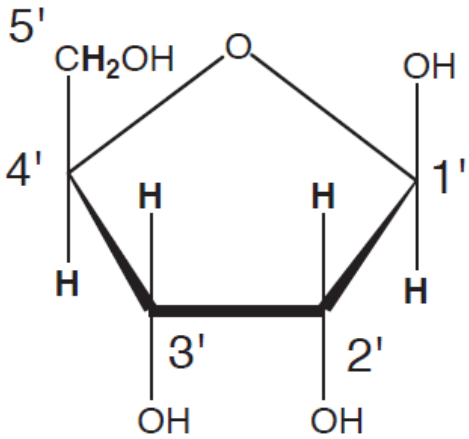


- joining the components

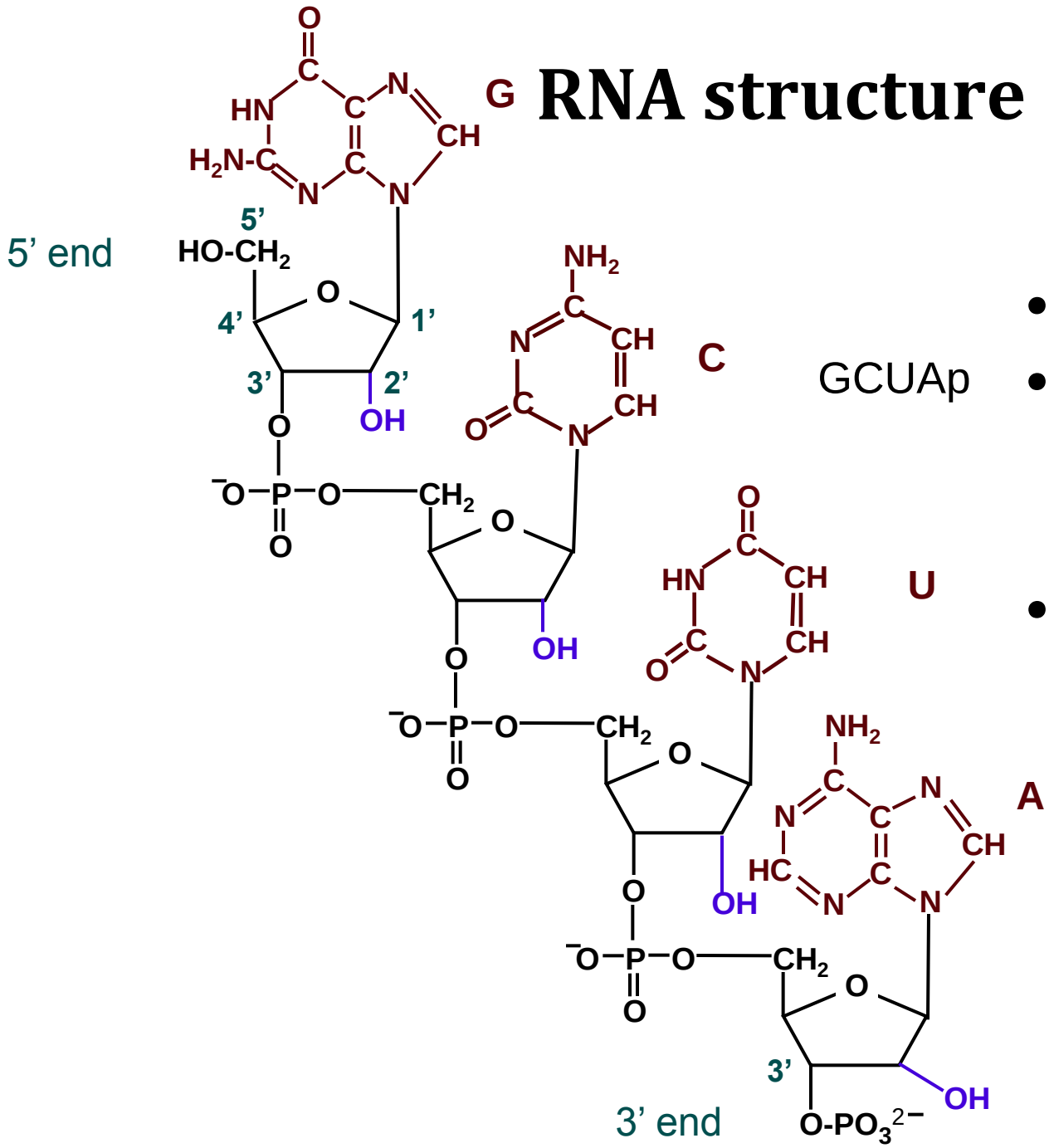
- adenosine 5'-monophosphate
  - not adenine, adenosine, …



- note numbering on sugar ring



RNA structure



- negative charges
- directional
  - 5' to 3'
- notation
  - always 5' to 3'

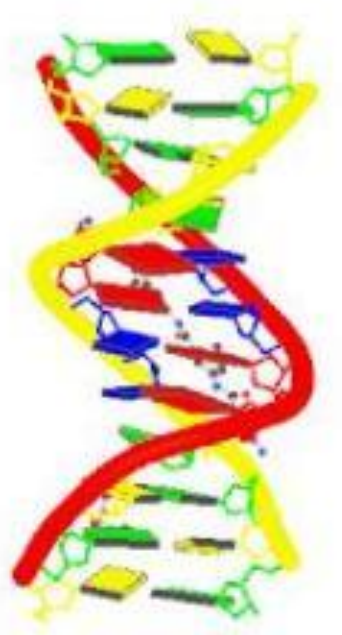
from Parson, W, http://courses.washington.edu/bioc441/lectures/ppt/13.DNA&RNA.ppt

H bonding

What holds the pairs of a helix together ? H-bonds

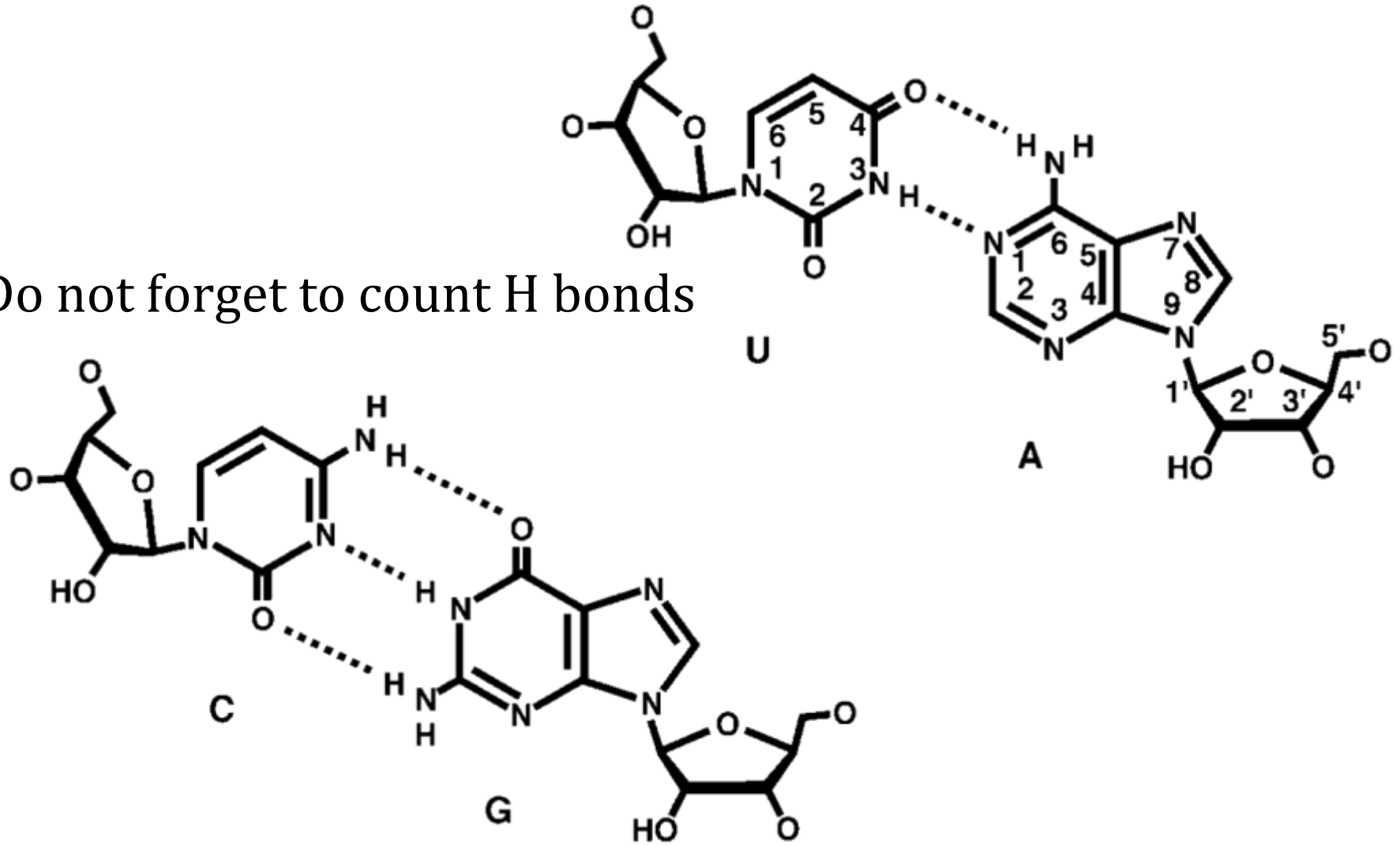
- applies to RNA
- rules from proteins
  - H-bond donors are NH, OH
  - acceptors – anything with partial –'ve

Historic H-bonding pairs…



Historic H-bonding pairs

Do not forget to count H bonds



from Batey, BT, Rambo, RP, Doudna, JA, Angew. Chem. Int. Ed. (1999), 38, 2326-2343

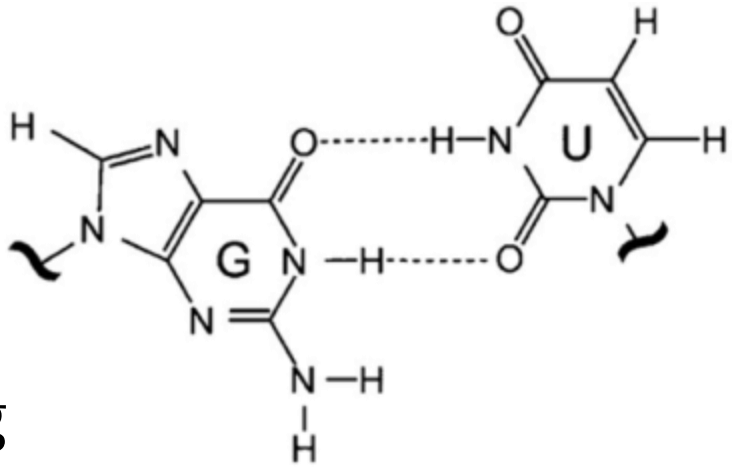
Historic viewpoint

- RNA has 4 bases + GC, AU base pairs
- H-bond pairs look flat
  - not true

Other common H-bond partner

Contrast with DNA (GC and AT)

- almost no mismatches in DNA



RNA (GC, AU) much more interesting

- third base pair GU (rather common)
- lots of weaker pairs possible

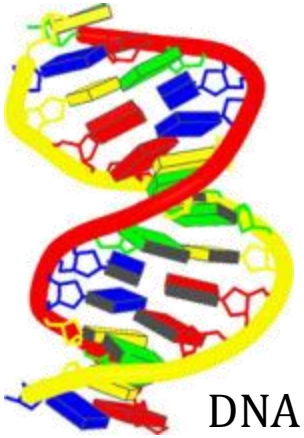
Possible RNA structures

DNA ? nearly always similar helix

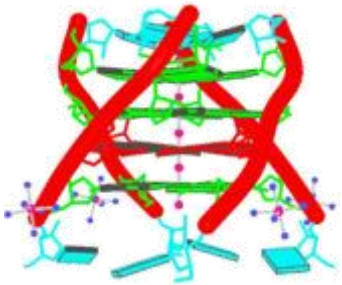
- some debate about A, B, Z, ..

RNA

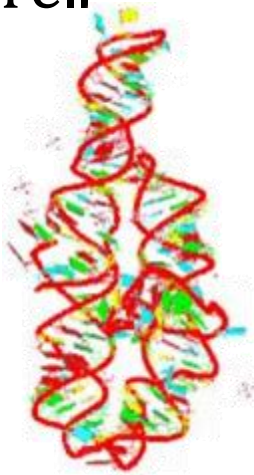
- lots of varieties known
- nomenclature..



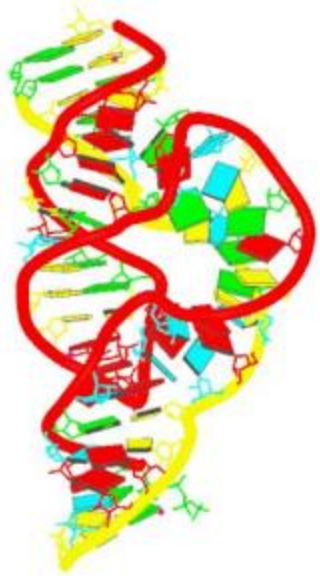
DNA
duplex
140D



tetraplex
1mdg



group I intron
1hr2



hammerhead
2oeu

tRNA
1evv

RNA coordinates / nomenclature

As for proteins: PDB format

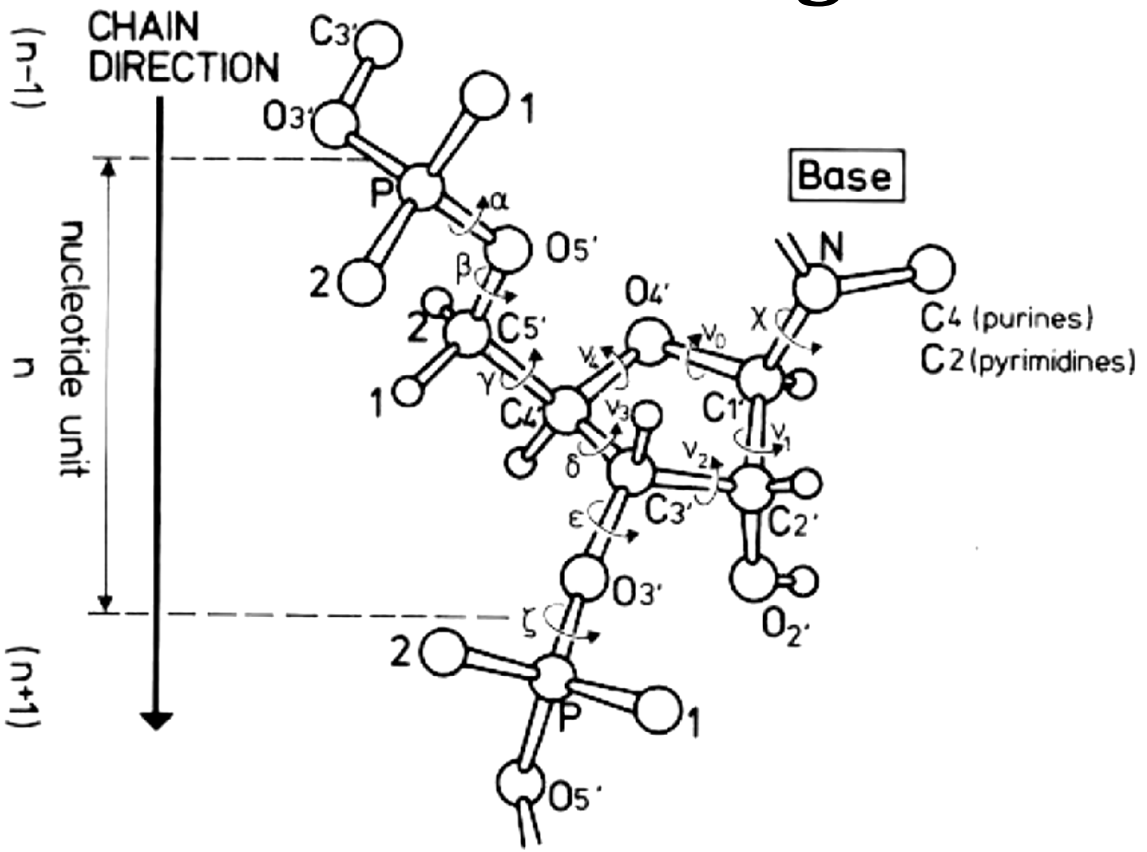
```
ATOM      1  O5*   G A 103      58.355  47.332  91.116  1.00175.32
ATOM      2  C5*   G A 103      57.373  48.210  90.636  1.00175.32
ATOM      3  C4*   G A 103      56.962  47.802  89.224  1.00175.19
ATOM      4  O4*   G A 103      58.148  47.463  88.474  1.00175.34
ATOM      5  C3*   G A 103      56.096  46.543  89.152  1.00175.03
```

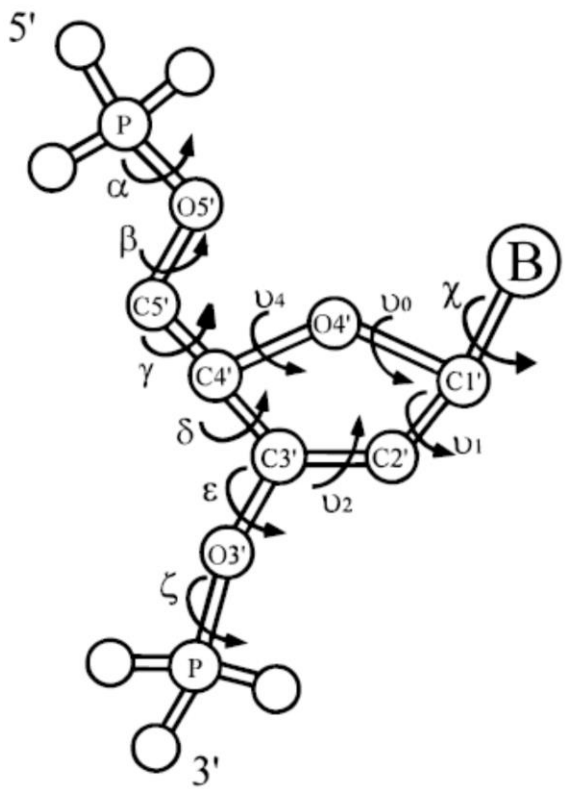
As for proteins

- dihedral angles are useful

- Unlike proteins (φ,ψ) there are 8 (α, β, γ...)

dihedral angle nomenclature





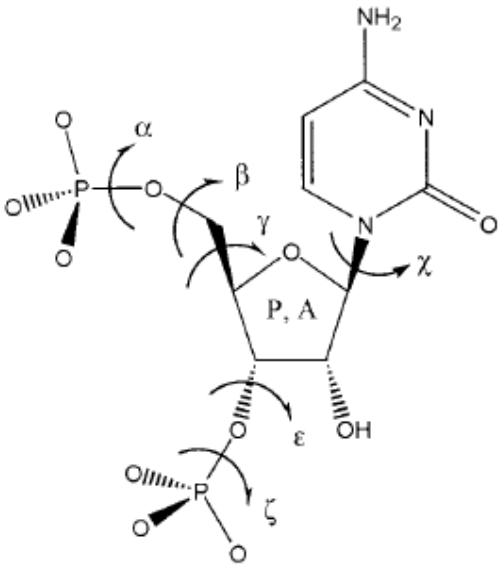
from Marino, JP, Schwalbe, H., Griesinger, C, Acc. Chem. Res. 32, 614-623 (1999)

from Saenger, W. Principles of Nucleic Acid Structure, Springer, N.Y. 1984

dihedral angle nomenclature

8 angles

- α, β, γ, ε, ζ, χ
- 2 for sugar (P, A)



- too many for me – how to simplify ?

what if two angles are highly correlated ?

- if we know *x*, then *y* is probably known

ideas for classification…

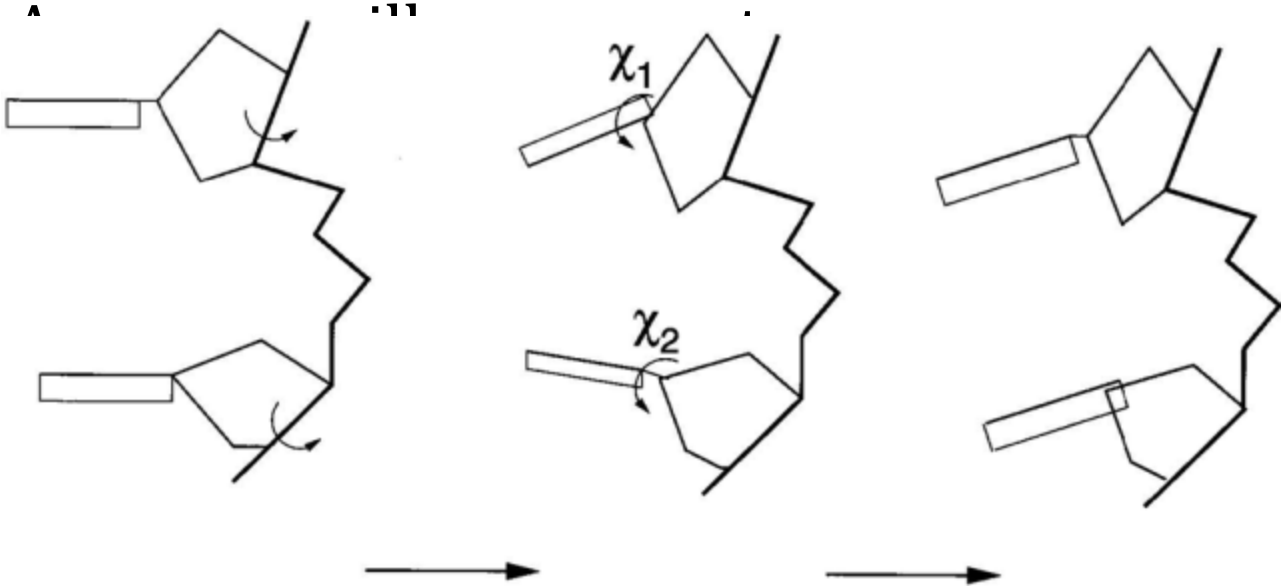
Describing RNA conformation

Example approach – look for correlations

- principle component analysis (quick detour if necessary)

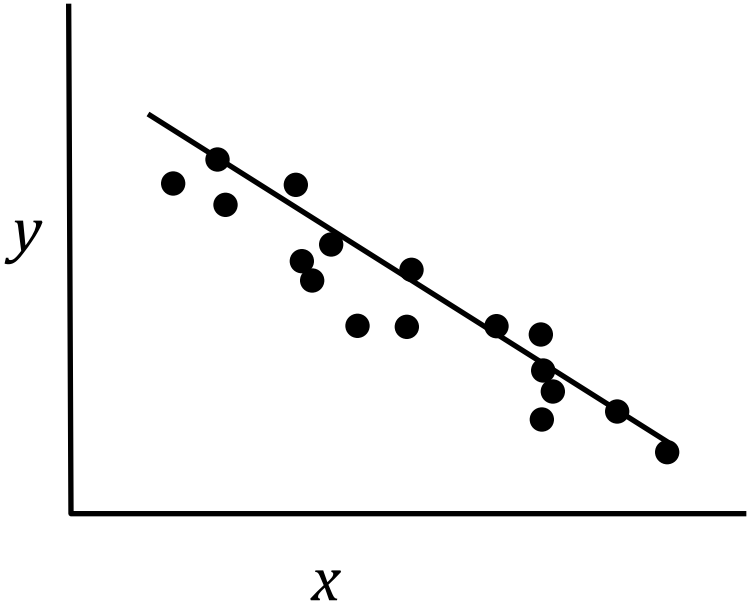
What if sugars move in two residues ?

- energetically, would like to maintain base pairing…
- P,
  -



Beckers, MLM & Buydens, MC, (1998), J. Comput. Chem. 19, 695-715.

PCA reminder



I have two dimensional data

- could well be described by a first (component) and
- maybe second component

n-dimensional data

- how much of variance is described by 1st, 2nd, … components

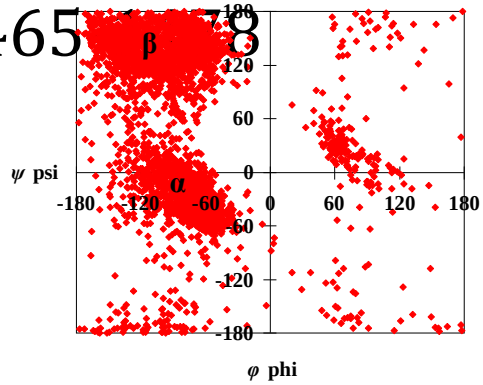
Describing RNA conformation

Claim

- most conformations are well described by 3 variables

Alternative…

- do not work in terms of real dihedral angles
- invent reference points
- example study…
  - Duarte, CM & Pyle, AM, (1998) 284, 1465



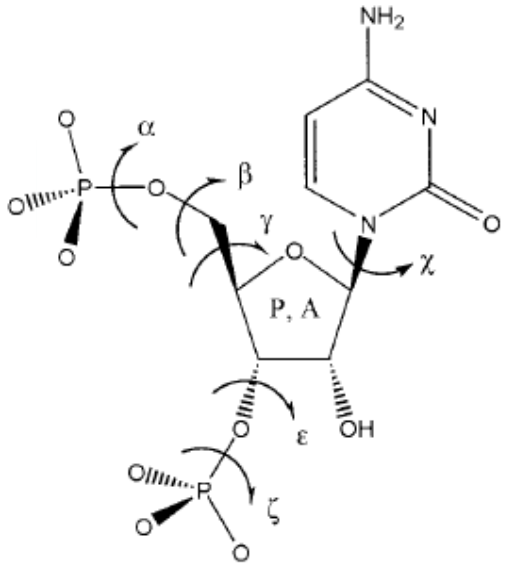
remember ramachandran plots in proteins

- can one do something similar in RNA ?

Reduced RNA conformation

Basic idea

- pick 4 atoms that are not sequential
- define a simplified backbone
  - P-C_4-P-C_4-P-C_4-…
- leads to "pseudo-torsion" angles



η

$C4_{n-1}$-P_n-$C4_n$-P_{n+1}

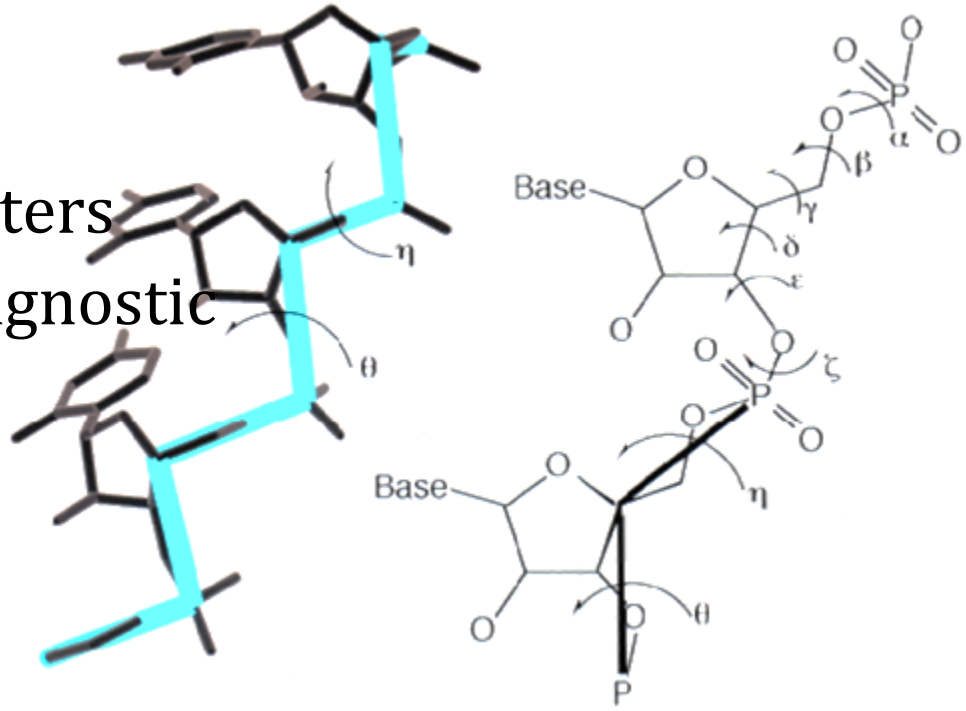
θ

P_n-$C4_n$-P_{n+1}-$C4_{n+1}$

Reduced RNA conformation

Plan of authors

- take 52 structures
  - (≈700 nucleotides)
  - collect η, θ
    - see if there are clusters
    - see if angles are diagnostic

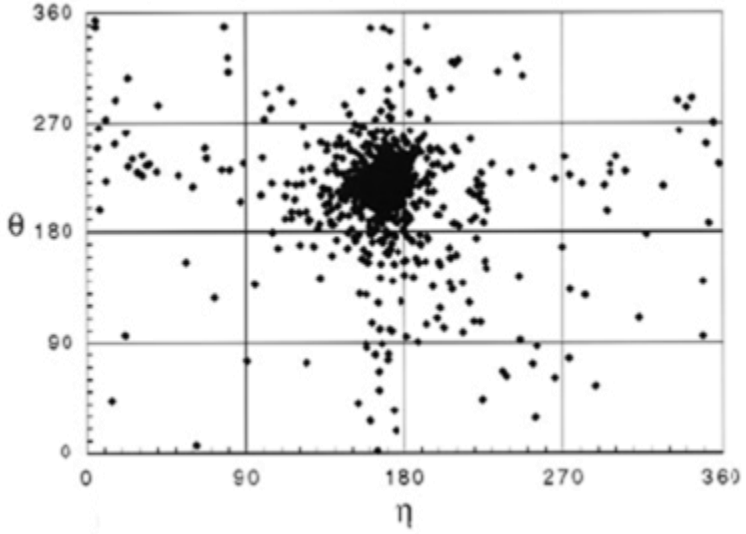


Duarte, CM & Pyle, AM, (1998) 284, 1465-1478

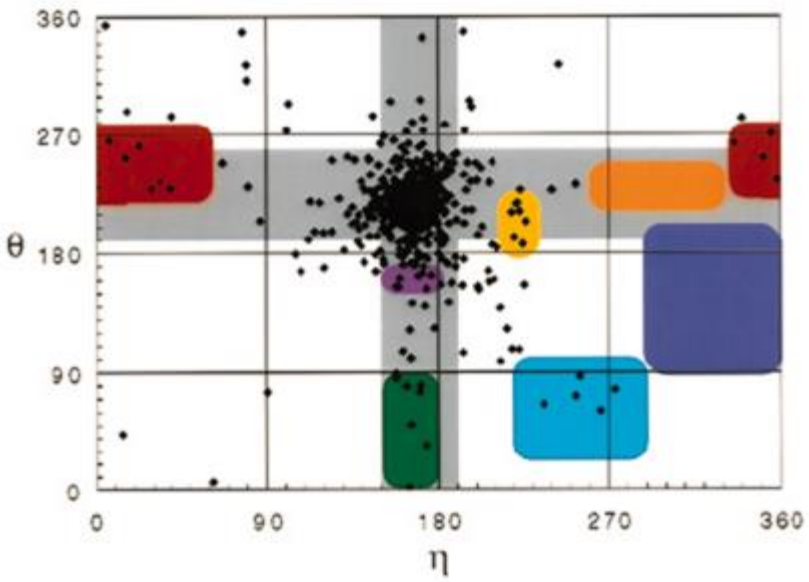
Reduced RNA conformation

Do you see clusters ?

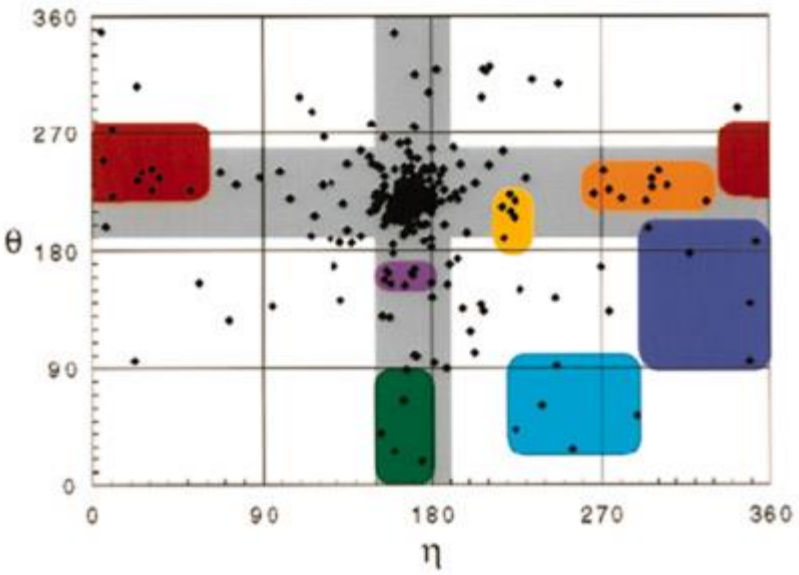
- main set of points …
- boring RNA helix



-



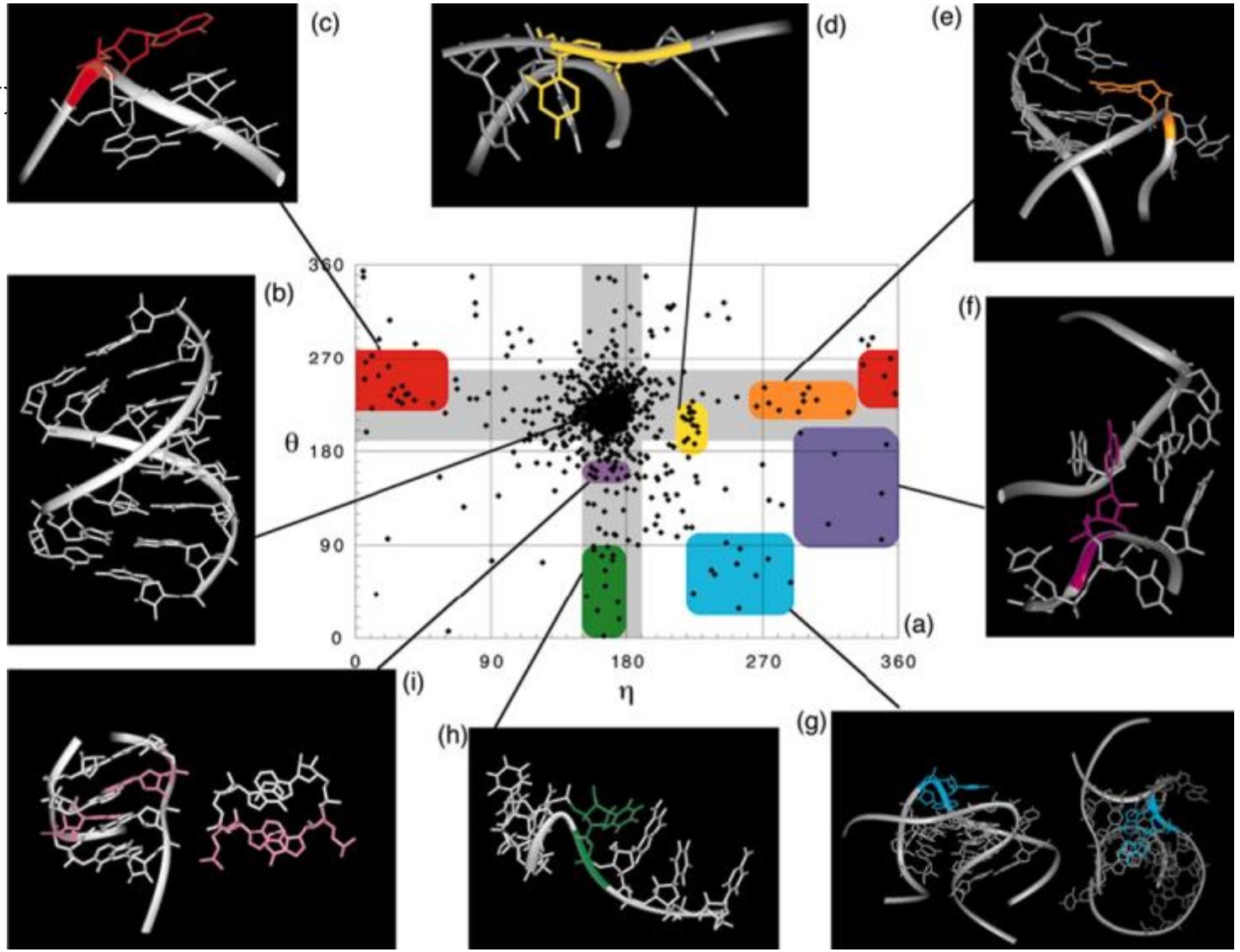
no tertiary interactions



yes tertiary interactions

Duarte, CM & Pyle, AM, (1998) 284, 1465-1478

Reduced RNA conformation



Duarte, CM & Pyle, AM, (1998) 284, 1465-1478

Reduced RNA conformation

We are interested in a critical look at ideas

How to read this…

- if you measure a pair of η, θ pseudo-angles
  - could you guess if something is wrong in structure ?
  - could you use this to categorise the conformation ?
- are there better ways to categorise structure ?

Summary

- RNA structure as per Watson-Crick, old text books
- How are RNA structures different to DNA ?
- What are the biological roles ?
- Can we neatly summarise RNA structures ?
  - see what information (angles) are necessary
  - define alternative angles

- Next..
  - What is life ?

The RNA world

Definitions of life

Evidence for RNA world

Before RNA world

Problems with RNA world

Alternatives (maybe there was no RNA world)

History

Start of life

- proteins are catalysts –necessary to copy DNA..
  - until…

Origin of life

The RNA world

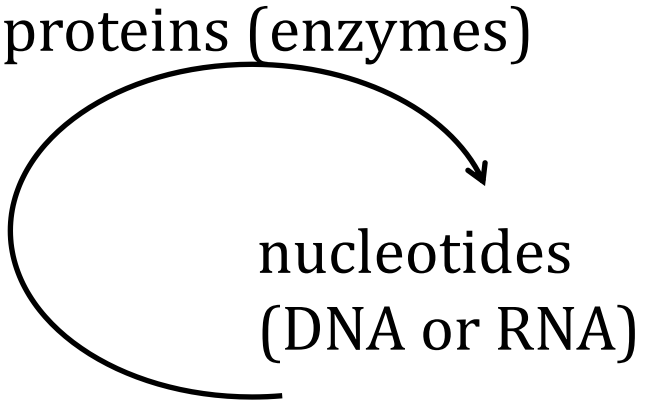
from Walter Gilbert

UNTIL recently, when one thought of the varied molecular processes at the origin of life, one imagined that the first self-useful exon to pass from one replicating structure to an unrelated one.

This picture of the RNA world is one of by arranging them according to an RNA template using other RNA molecules such as the RNA core of the ribosome. This process would make the first proteins, which would simply be better enzymes than their RNA counterparts. I suggest that protein molecules do not carry out enzymic reactions of a different nature from RNA molecules but are able to perform the same reactions more effectively

- 1986: first RNAzymes found
- start of RNA world story
  - today …

Today versus history



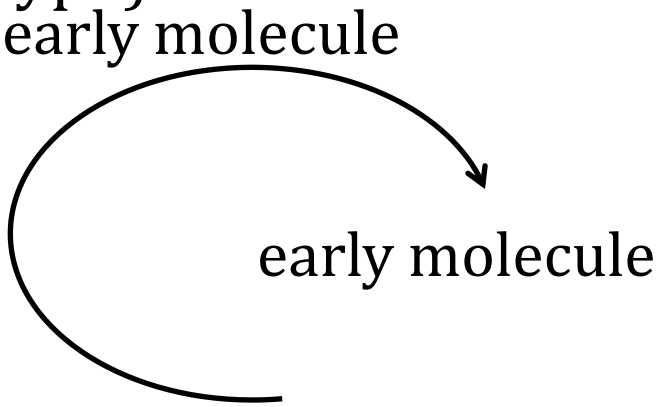
Picture today

- simultaneous development of
  - proteins (copying)
  - nucleotides (information storage)

Suggestion

- one molecule (phenotype+genotype)
  - self copying
    - possibilities

    1. protein like
    2. nucleotide like
    3. something else



This is templated

What is life ? Practical – not philosphical

Practical – not philosophical

- people, trees, …
- bacteria
- viruses ?
- infectious DNA / RNA ?

Some concepts

- life consumes energy – better formulated
- life avoids equilibrium, needs energy, consumes entropy
- evolution

Equilibrium

Reaction A + B ↔ C + D $\Delta G = RT \ln \frac{[C][D]}{[A][B]}$

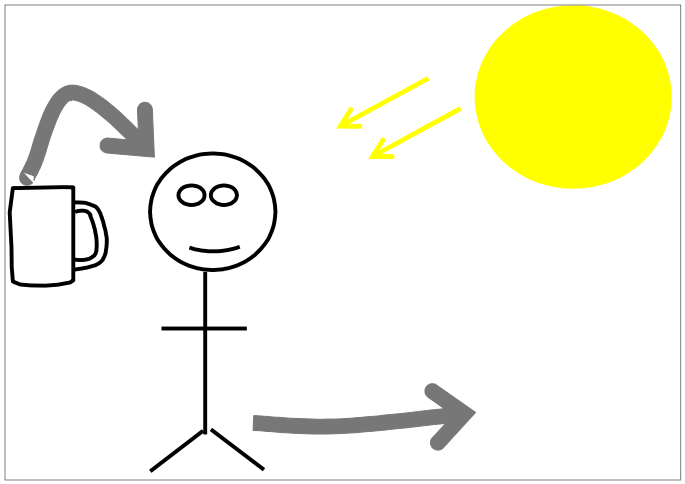
Decay A ↔ B + C, then $\Delta G = RT \ln \frac{[B][C]}{[A]}$

In a closed system, if $\ln \frac{[B][C]}{[A]} = 0$ you are dead

Consequence

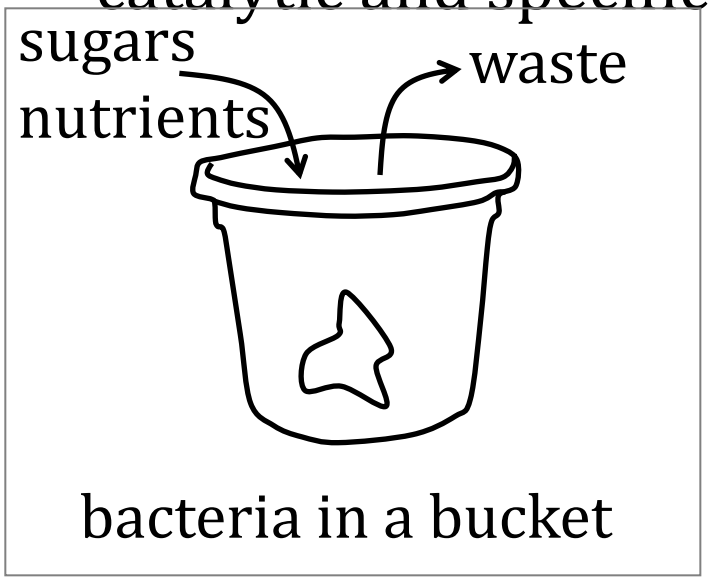
- life looks like "steady state" (stationärer Zustand)
- not Gleichgewicht

Steady state systems

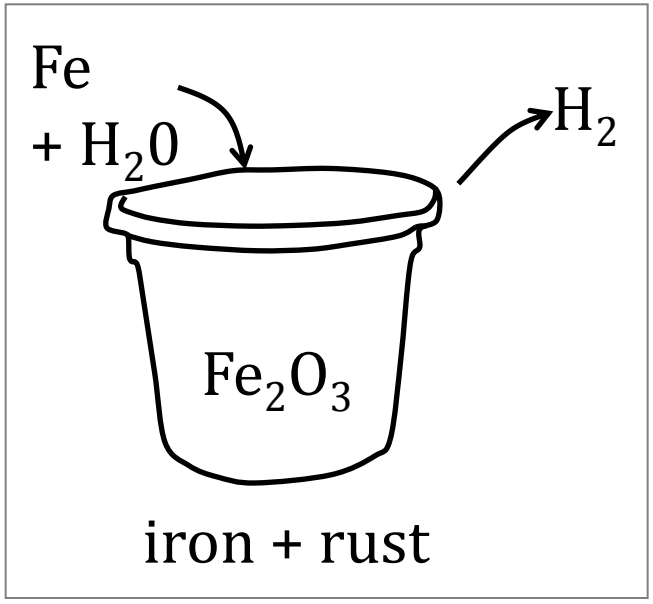


Input of energy

- maintenance of order
- grows
- catalytic and specific



bacteria in a bucket



iron + rust

- bacteria and rust
  - grow, eat nutrients, catalyse their own copying

Rust

Why is rust not life

- low information
- no ability to change and evolve

information / entropy

Entropy is easy to define

$$S = -k \sum_{i=1}^{N} p_i \ln p_i \quad \text{where we have } N \text{ states}$$

Remember properties of S

- if all states are equally populated $p_i = N^{-1}$
  $S = -k \ln N^{-1}$ (not so small)
- if only one state is populated
  $$S = -kN^{-1} \ln N^{-1}$$
- think of organisms

Information

A genome is a string amongst possible genomes

- *E. coli* ?
  ≈ 5 million base pairs (5×10^6)
  $4^{5\times10^6} \approx 10^{3000000}$ possible states (complexity ?)
- how many states are used ?
  - very few
  - "information" per genome is big

- genome of rust ? information in rust ?
  - alphabet is 1 ? length is 1 ?

Claim

- evolution is information increase via selection

Complexity

Smallest genomes

- viruses – few proteins – parasitic

Free living ?

- a few hundred proteins

Is there a minimum complexity for life ?

- no answer, but rust is very simple

Life

Rust can

- catalyse the production of rust, does not adapt

Our "life"

- has general copying machinery
- can copy sequence_1 or sequence_2
- templated copying

This flexibility necessary for evolution

How simple could evolution be ?..

Summary of life

- not at equilibrium / consuming energy
- catalytic
- creating / storing information
- copying with possibility of change / selection

Why so important ?

- Ursuppe should have these properties

RNA world properties

- genetic continuity via RNA
- Watson-Crick base pairing
- no genetically-coded proteins

  - did it exist ?

Why believe in an RNA world ?

1. both phenotype and genotype
2. roles of nucleotides
3. Selex
4. biosynthesis
5. ribosome

In turn..

Phenotype and Genotype

Proteins

- catalysts
- rarely code for other proteins

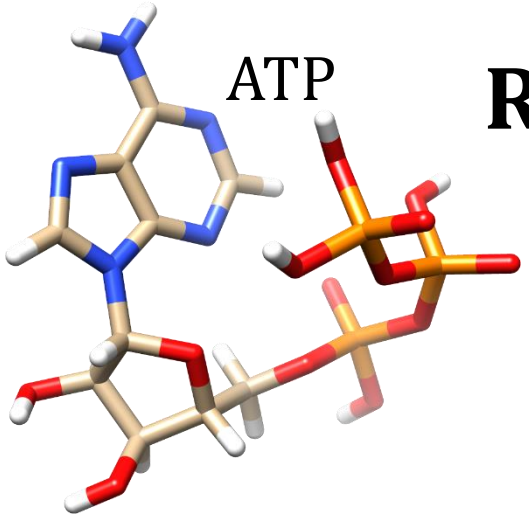
RNA

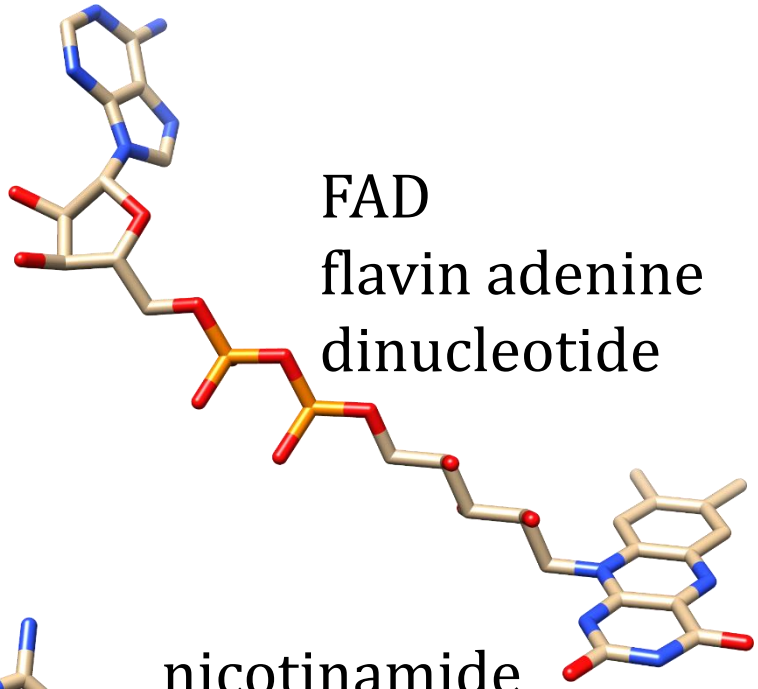
- catalysts
- does encode other DNA / RNA molecules

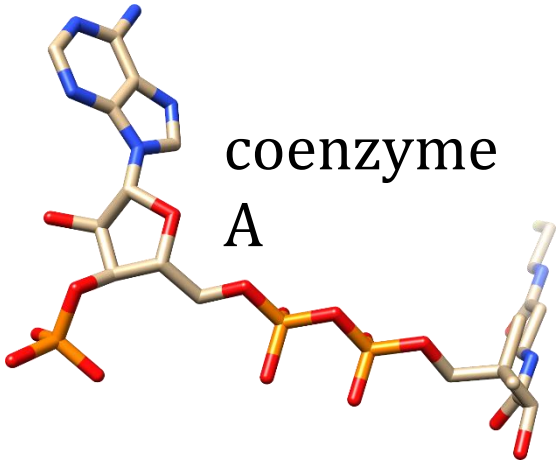
Simplicity - life started with one kind of molecule

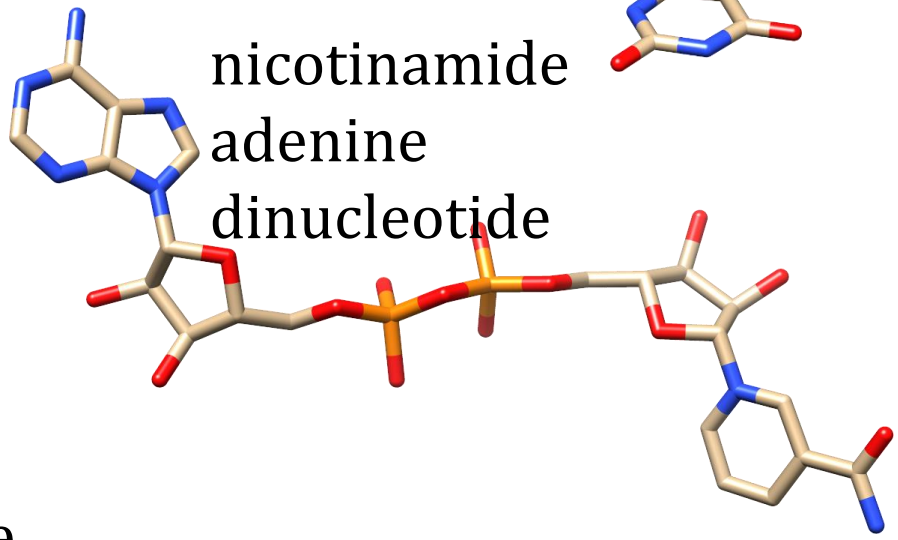
- should be RNA (RNA-like)

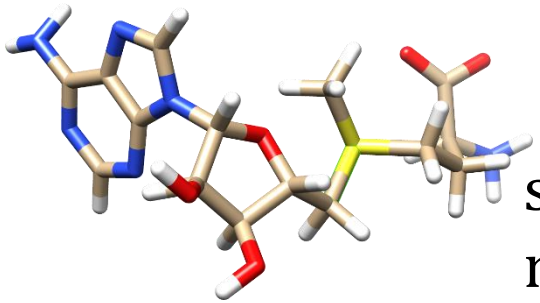
Roles of nucleotides











Why believe in an RNA world ?

Roles of nucleotides

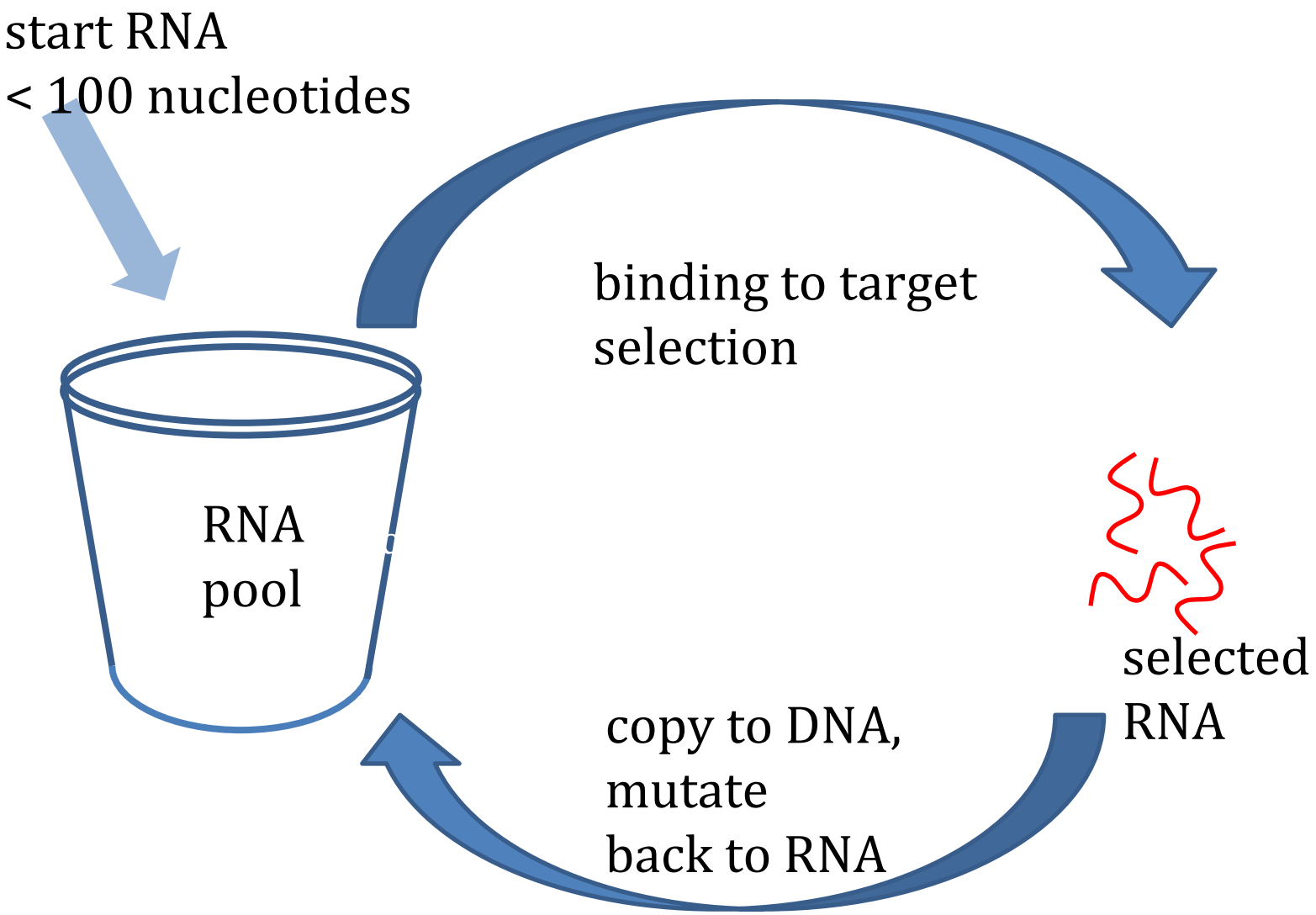
Cofactors, nucleotides, energy

- basically nucleotides
  - ATP, FAD, NAD, TPP, …

SELEX

- topic in other lectures (Prof Hahn, 4. Woche)
- Basic Idea

SELEX (2 minute version)



SELEX (2 minute version)

Empirical

- fishing in an RNA soup, one can find all kinds of activities / binding abilities
- can one find binding / stabilization of transition states ?

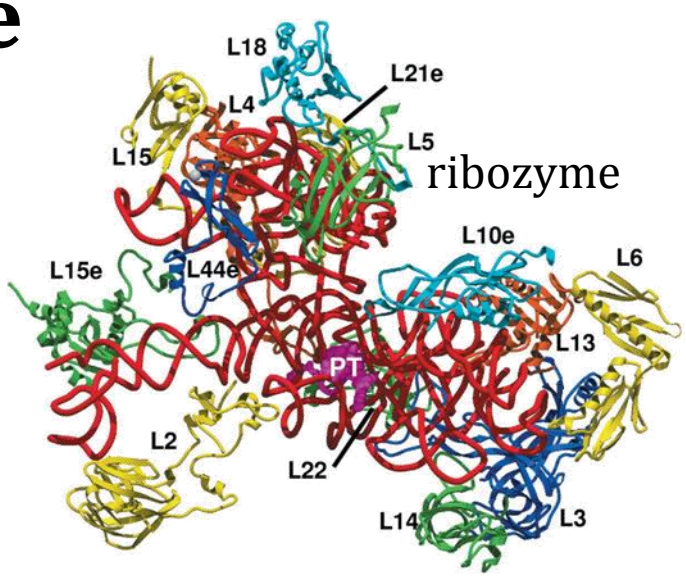
Interpretation

- activities are present in random soup waiting to be found
- start of life was just a big selection experiment

Biosynthesis

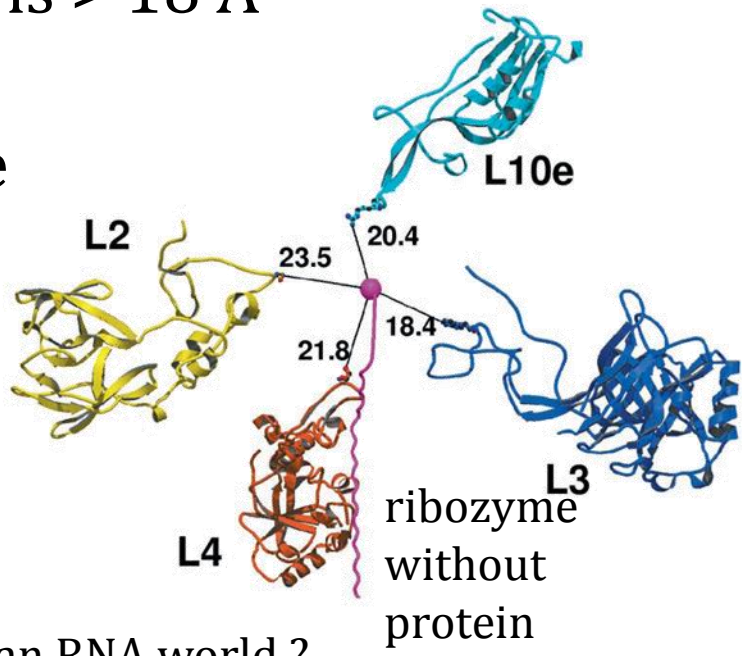
- much machinery devoted to RNA biosynthesis
  many enzymatic steps
- DNA is just a modification afterwards
- looks as if RNA is the older molecule

The ribosome



- incredibly conserved
- part of ribosome near active site

- remove all the RNA
- the nearest protein to active site is > 18 Å
- the fundamental operation of making proteins from a template
  - carried out by a ribozyme



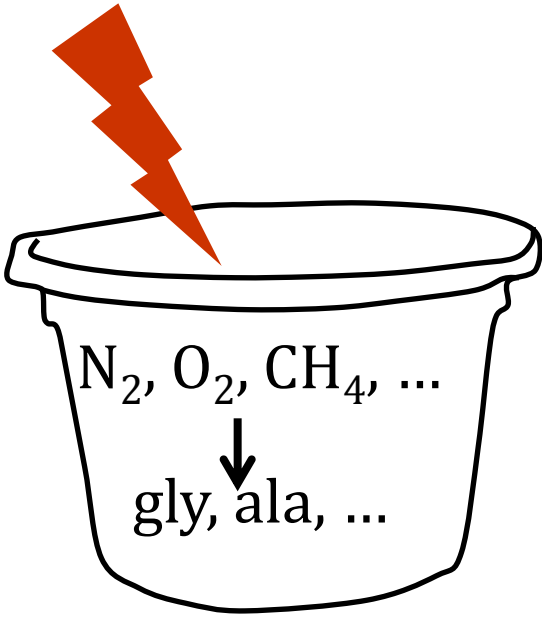
Nissen P, Hansen J, Ban N, Moore PB, Steitz TA., Science 289, 920-930 (2000)

Why believe in an RNA world ?

RNA World – requirements

Source of basic requirements

- ribose
- bases (A, C, G, U + more T, I, X, …)



Vague source

- Miller experiments from 1950's

Can one make nucleosides ? nucleotides ?

- polynucleotides ?

Lots of problems… Later

How to make nucleotides ?

N-sugar + xxx-PO_4 →N-sugar-PO_4 (+H_20 + xxx)

ribozymes have been made for related reactions

- quite plausible
  - no really good candidates yet

Abiotic ?

- many examples of catalysis exist
  - Pb^{2+}, BO_3^{3-}, …

RNA replicase

One model – we have one replicase

- basic requirement – replicase should
  - act on itself (or similar copies)
  - should produce
    - itself or
    - complementary copies

Length constraints

- define fidelity q = probability that one residue is correctly added
- probability of copying chain length n correctly = q^n
- no mistakes – no evolution

q	n	perfect copies
0.9	4	0.66
0.9	10	0.35
0.95	10	0.65
0.95	20	0.36

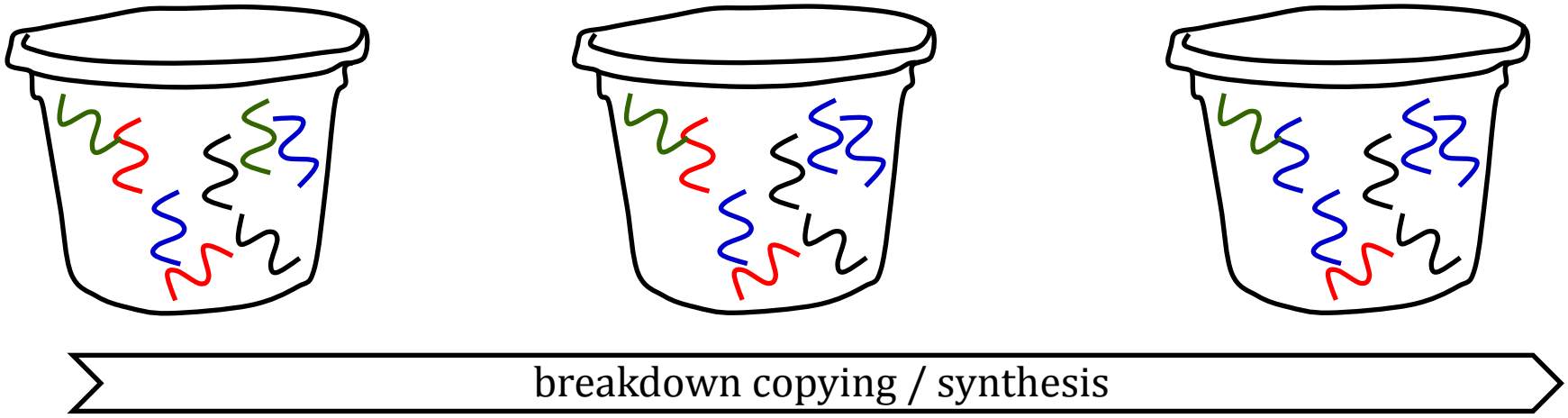
Replicase Quality

- Is there are magic q ?
- Must we wait for some chemicals with correct q ?

- No ! Evolution helps

evolution without cells (primordial slime)

What do we need for evolution ? Not much



If the blue molecule and related variants

- copies itself better
- is copied by other molecules
- resistant to breakdown

It will eventually dominate

First replicase

How likely are we to take a random soup of nucleotides

- ribozyme of 40 bases
- $q = 0.9$
    - not very likely, but if
- a replicase starts
    - copies related molecules better than unrelated

If it copies better / faster it will be selected for and evolve

- could this happen ?
    - copying by other catalysts using RNA as template

Alternative Genetic Systems

Must we start with RNA ?

If not, bias is towards a system

- can pair specifically with RNA sequences
  - XYZW pairs to ACGU so we can have template directed RNA synthesis
  - explains transition to RNA world
- should form an open (helical) structure

Why do we like this ?

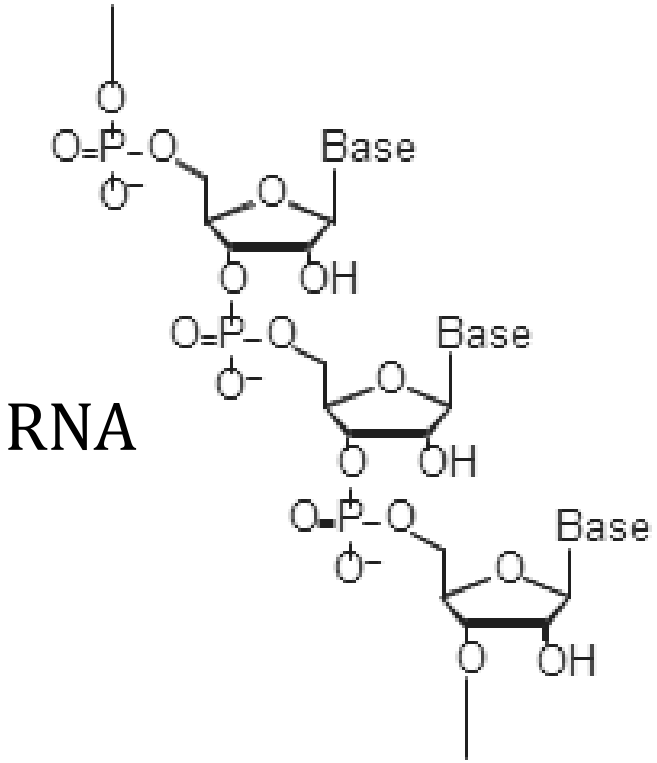
- Maybe we do not have to explain the abiotic synthesis of RNA building blocks

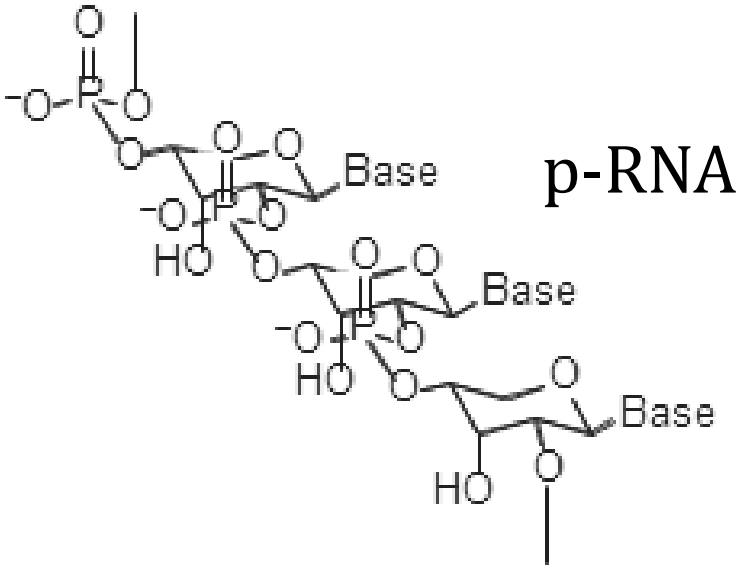
Examples…

Alternative – different sugar

pyranose form

- forms helices
- has base-pairing



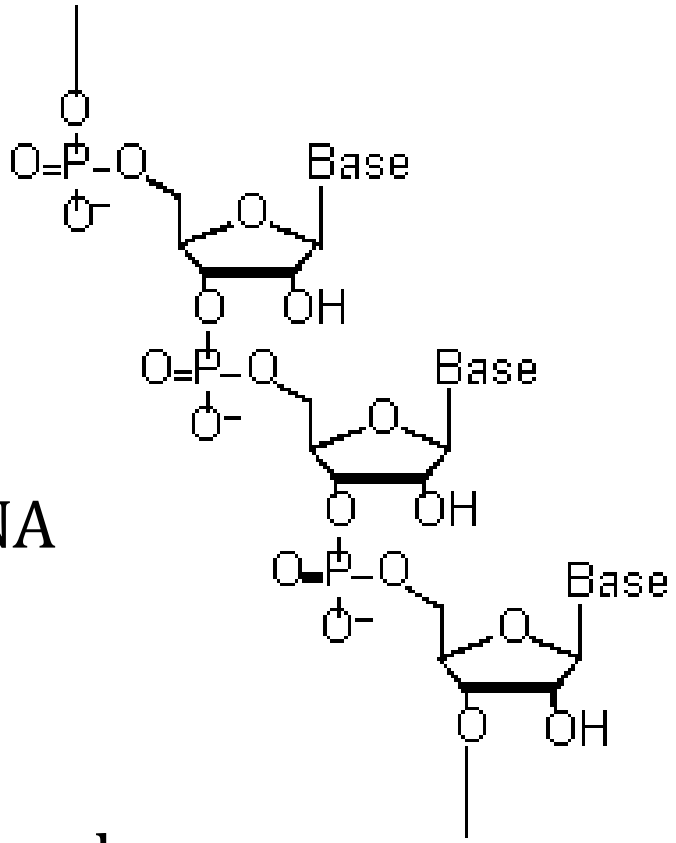


Robertson, MP & Joyce, GF, Cold Spring Harb Perspect Biol 2012;4:a003608

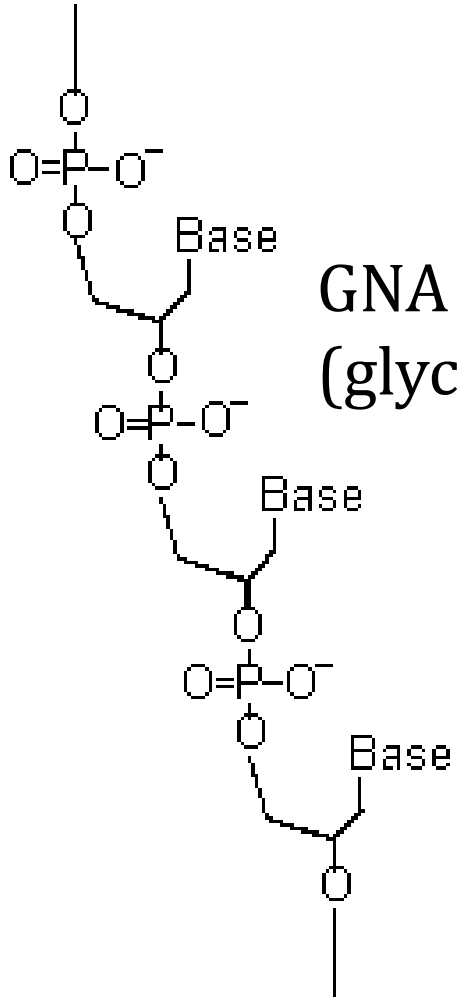
Alternative – glycol backbone

Could be called open sugar

- forms stable double helices



RNA



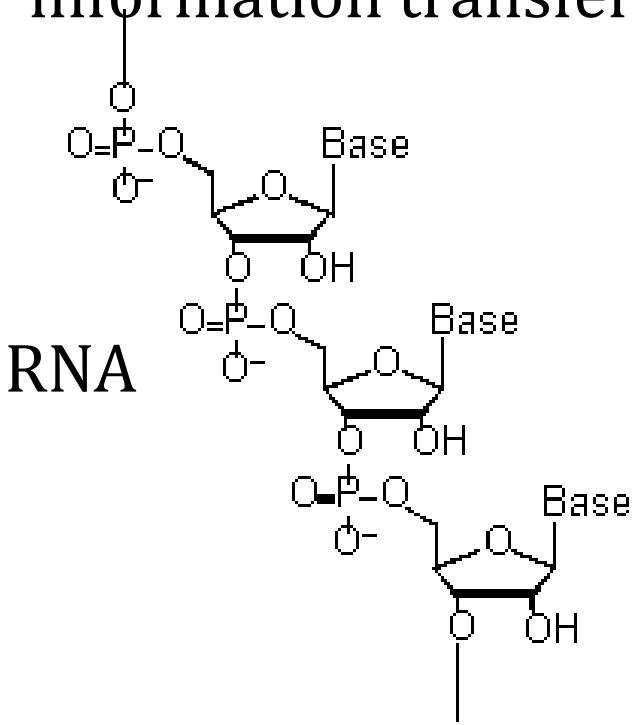
GNA (glycol)

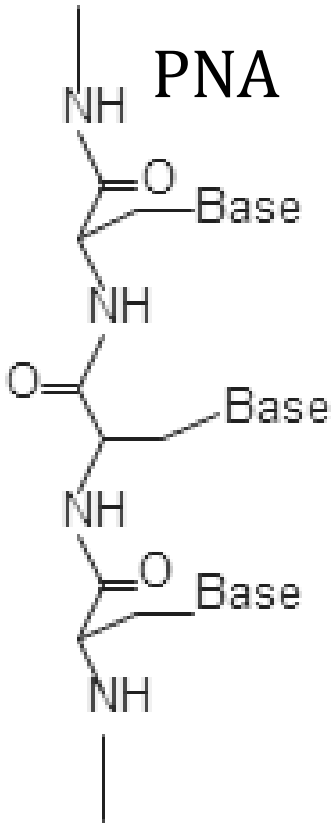
Last example..

Robertson, MP & Joyce, GF, Cold Spring Harb Perspect Biol 2012;4:a003608

Alternative - PNA

- peptide bonds on backbone
- forms helices
- binds to RNA
  - information transfer to RNA





Robertson, MP & Joyce, GF, Cold Spring Harb Perspect Biol 2012;4:a003608

Alternatives

Still more alternatives

Main point

- RNA world may have evolved from something related

Some problems…

ribosome (problems)

Usually believed to be a ribozyme.. Is it ?

Now many ribosome structures

- better resolution
- with substrates bound

Strong evidence of L27 + L16 interacting with tRNA

The point

- not everybody believes that the ribosome is a ribozyme

Other RNAzymes may not be.. (problems)

Rnase P

- maturation of mRNA
- recent RNA-free variant found (seminar topic ?)

- Are there more RNAzymes which are not RNAzymes ?

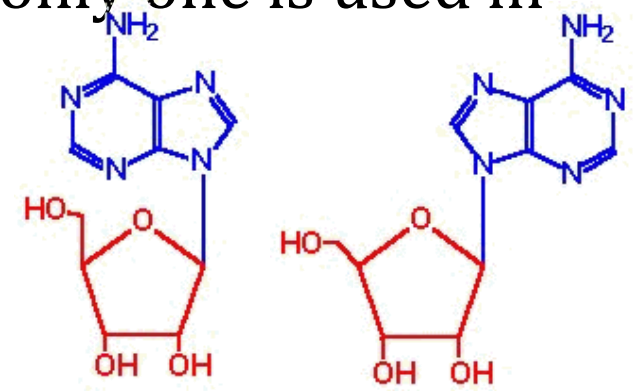
Specificity – sugars (problems)

Make sugar in lab

- condensation from smaller molecules
- result ?
  - mixture of 5 member sugars (ribose, pyranose, …)
  - ribose is not dominant

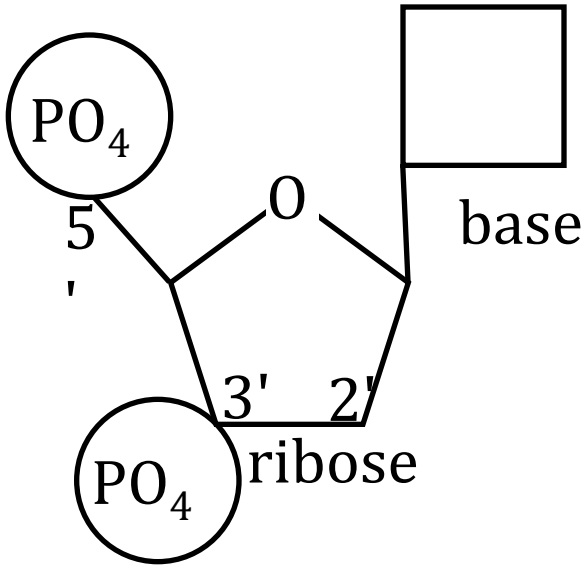
Enantiomers, isomers, ..

- details of linkages different, but only one is used in modern world
  - syn / anti, L / D



syn / anti

Joining monomers (problems)



Modern chemistry always 5' to 3'

Nucleotides (NMP)

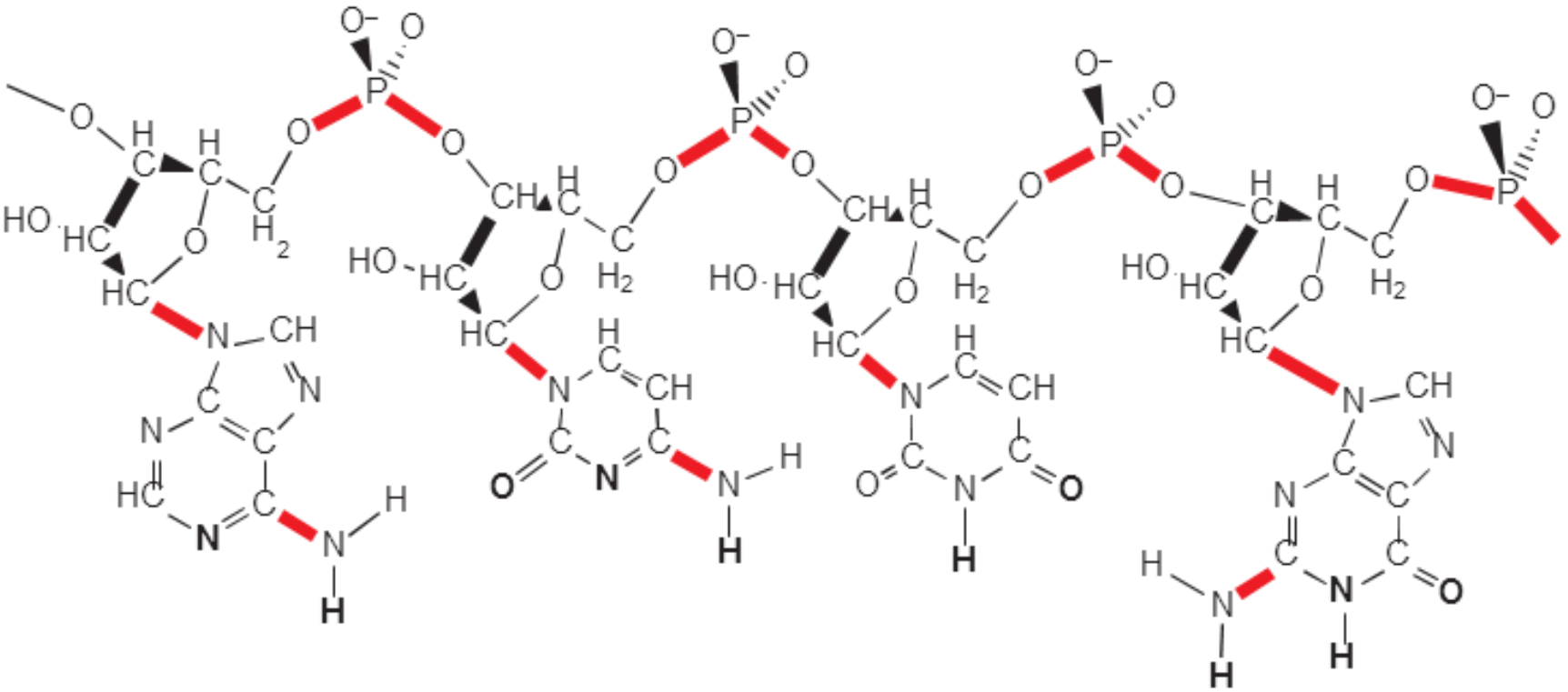
- 3 reactive groups
  - 5' PO_4, 3' OH, 2' OH

Soup of 5' NMPs and condense

- mixture of
  - 5', 5' pyrophosphate
  - 2', 5' PO_4 diester
  - 3', 5' desired diester

RNA is not very stable (problems)

All of the red bonds are subject to hydrolysis



Benner, SA, Kim, HJ, Yang, Z (2012) Cold Spring Harb Perspect Biol, 4, a003541

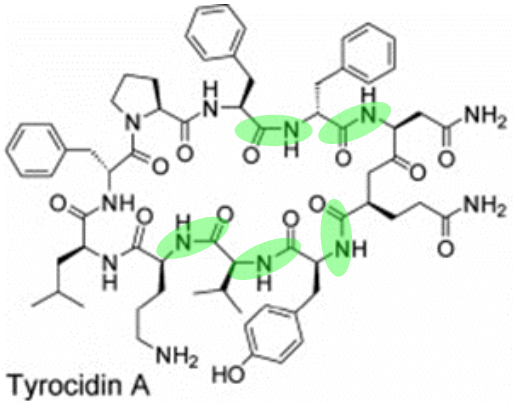
Do we need RNA for proteins ? (problems)

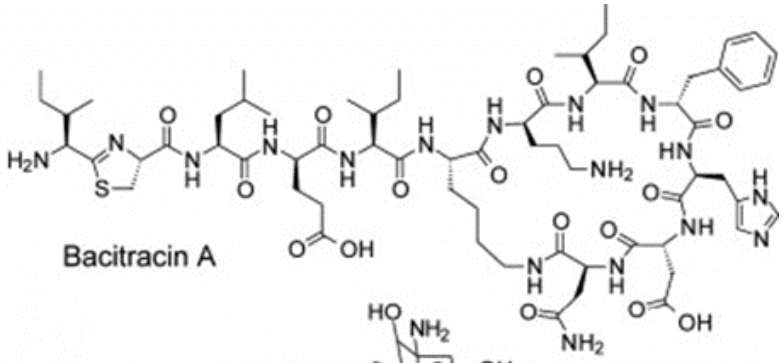
You think we need nucleotides to code for proteins, but..

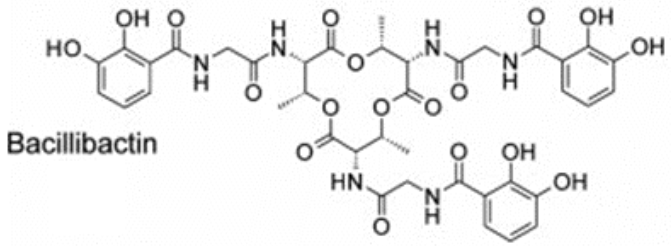
- there are many peptidyl transferases
- antamanide, glutathione ..

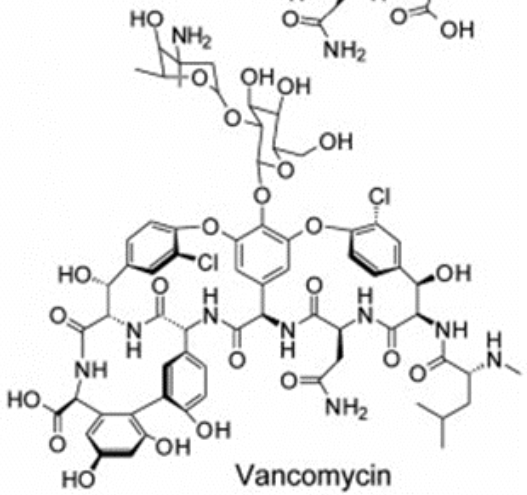
  lots of products, examples …

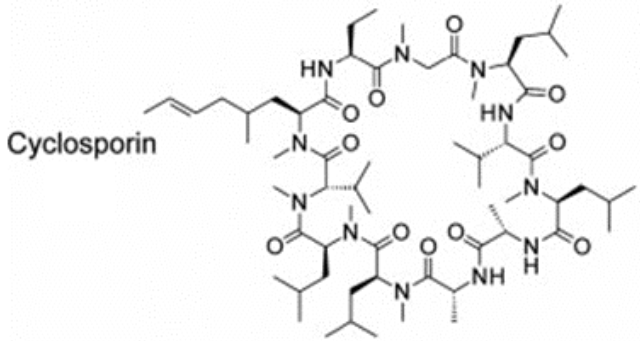
peptides – not genetically encoded

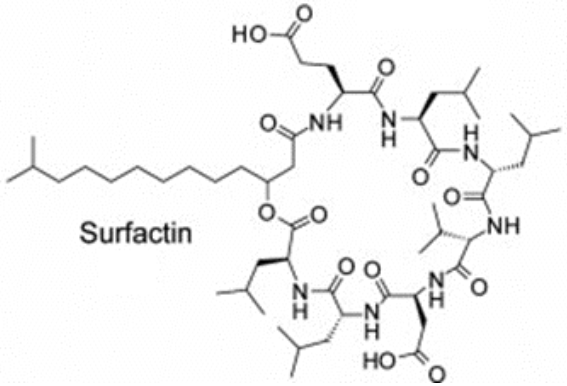












Marahiel, M.A. (2009), J. Pept. Sci. 15, 799-807

peptides – not genetically encoded

Peptides from bacteria to people

- biochemically ancient

Important point

- DNA is not the only way to store information
- Information / structure can be stored in other molecules (the synthetic enymes / peptidyl transferases)

RNA first ? Protein first ? (problems)

If the world began with RNA

- the oldest proteins should be involved in nucleotide synthesis / copying

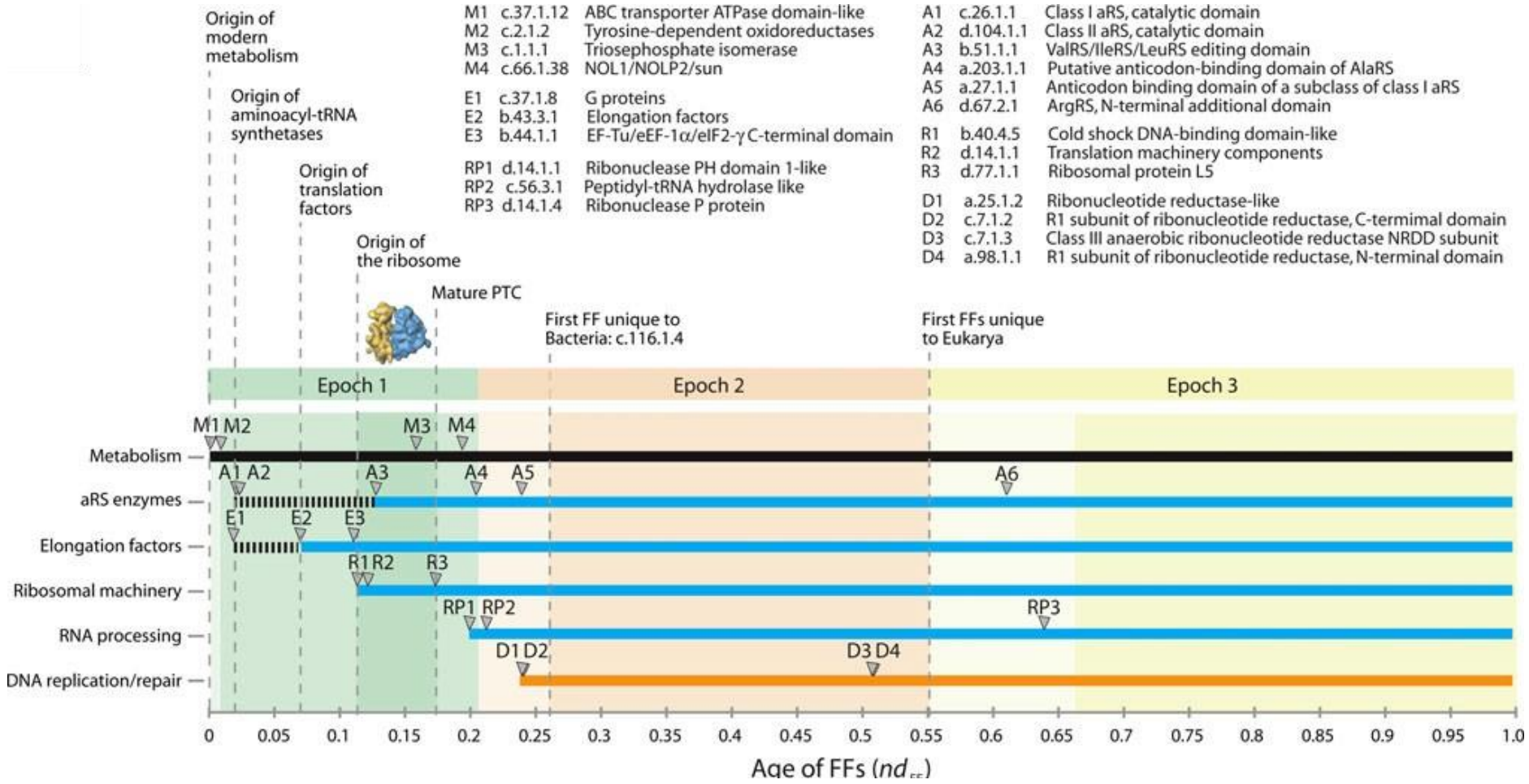
  Are they ?

Take lots of genomes

- Phylogenies (Baum des Lebens)
- trace history of proteins
- attempt to find the age of each protein

  (how far back in tree)
- …

RNA first ? Protein first ? (problems)



Caetano-Anollés, D, Kim, MK, Mittenthal, JE, Caetano-Anollés, G; J Mol Evol (2011) 72, 14-13

RNA first ? Protein first ? (problems)

Strong claim

- conventional metabolism precedes
  - RNA synthesis
  - amino-acyl tRNA synthesis

  - really all nucleotide biochemistry

General worries (problems)

Take

- several decades
- good organic chemistry labs
- lots of PhDs
- modern simulations
- modern laboratory equipment

Try to create

- a self replicating system out of abiotic components

Never really successful

Complete change of philosophy

maybe we do not need an RNA world

Do we need this general templating ?

So far – search for general replicase, polymerase

- Can one build a living system from less general components ? (nucleotides are very general)

- Examples earlier (antamanide, tyrocidin, many more)
  - what if tyrocidin catalysed the formation of antamanide which catalysed .. tyrocidin ?

What might we need for a self copying system ?

Some prerequisites are easy

First 4 points (previous slide)

- chemistry details
  - cycles of hot/cold/ dry/wet/concentration in drops, minerals..

last point

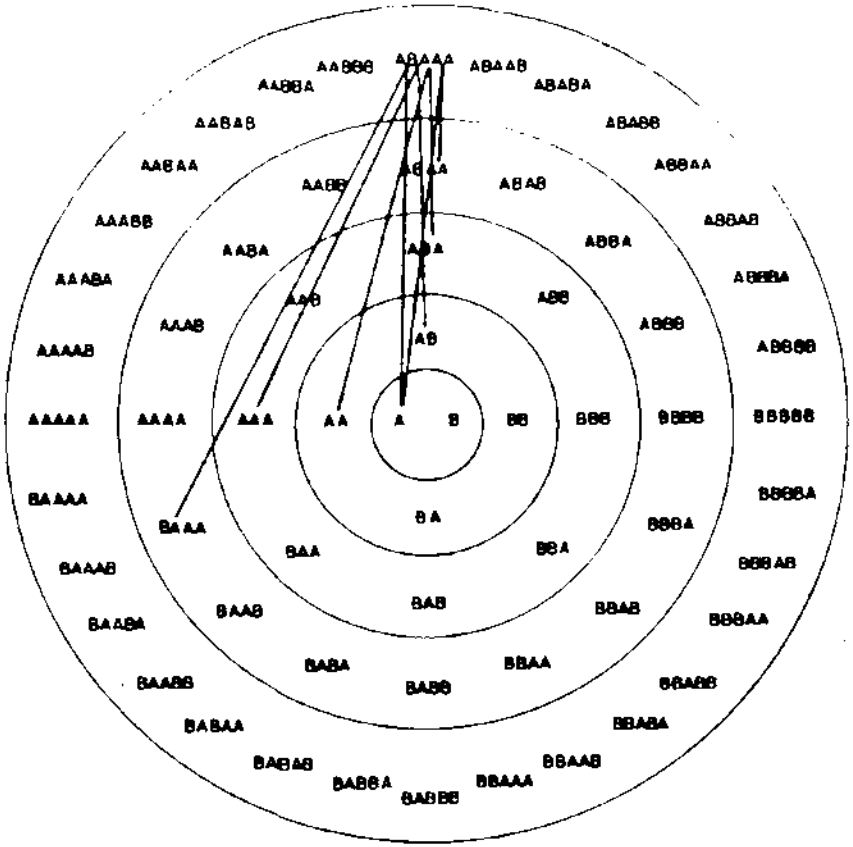
- catalytic closure

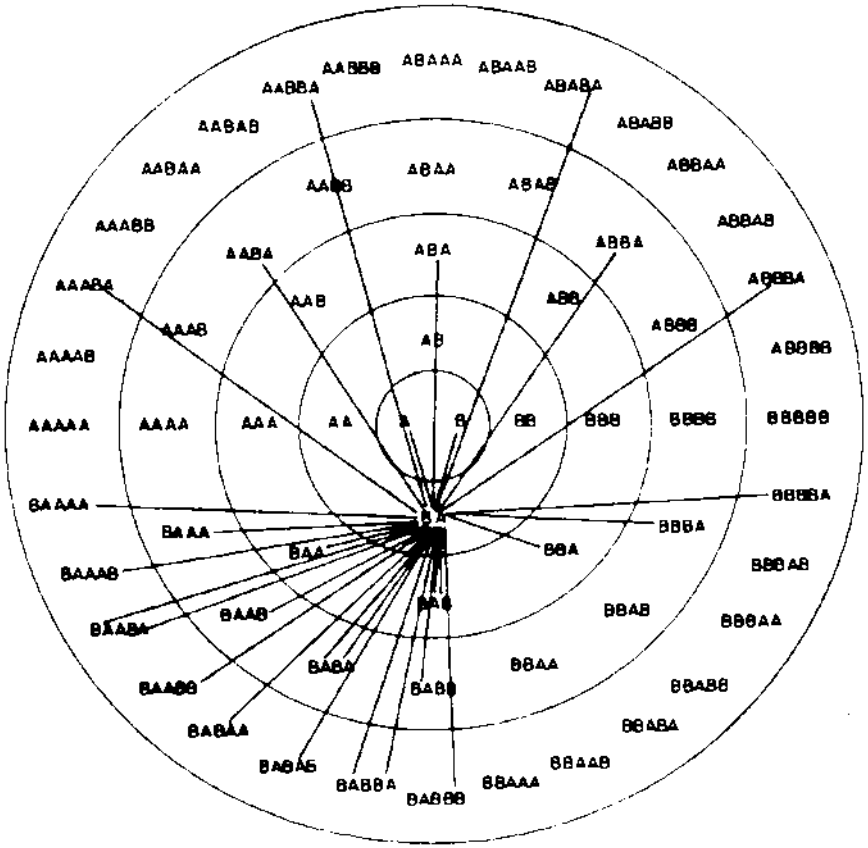
Catalytic closure

Imagine a soup of polymers with conversions

- cleavage or ligation ABCDE↔ ABC + DE

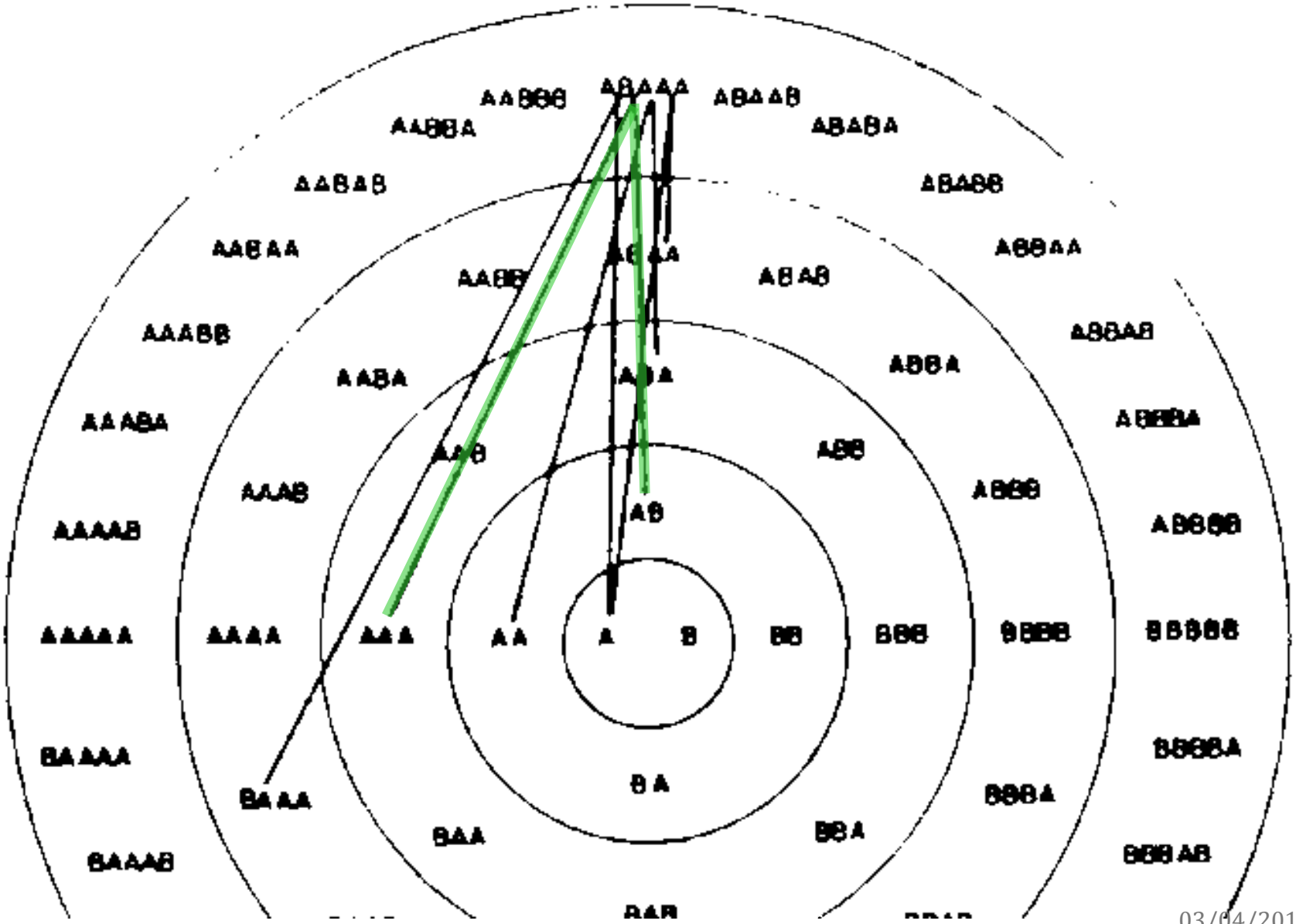
How many ways can we form a 5-mer ? or 2-mer ?





Kauffman, SA, The Origins of Order, Oxford University Press, NY 1993

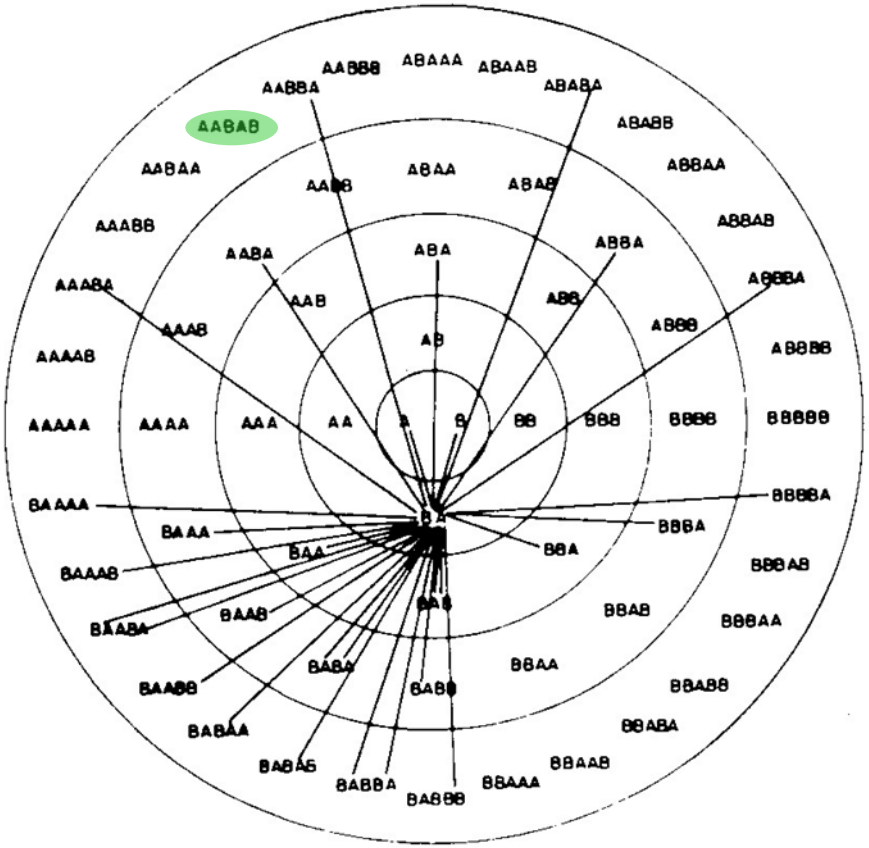
AB + AAA form ABAAA



Catalytic subset

Within set of polymers some are enzymatic for joining / breaking units

- for RNA 4×4=16 X-Y types
- pick a polymer
- with probability *p* pick a reac it catalyses
- imagine green sequence catalyses all AB bonds
  - leads to huge number of edges
- go to next sequence, maybe assign a reaction

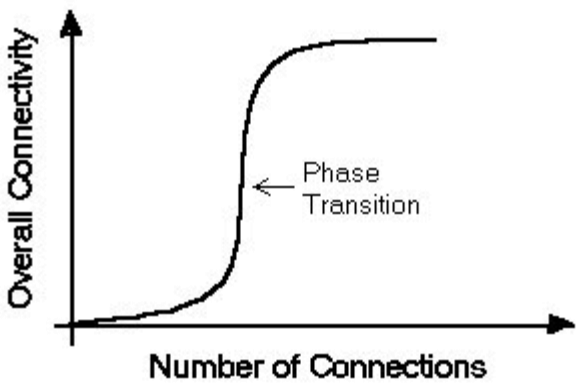


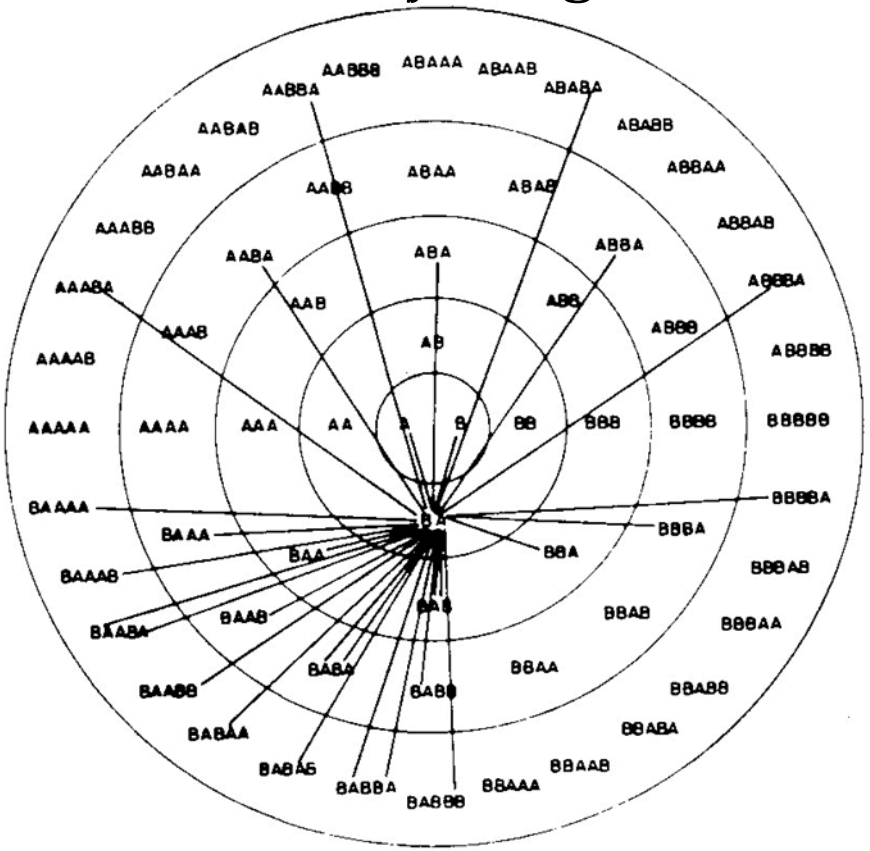
Catalytic subset

How many real enzymes and edges do we need ?

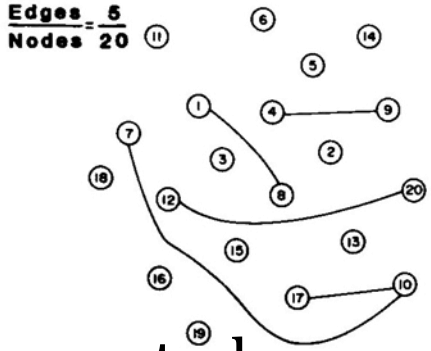
- I do not have to be able to synthesise everything

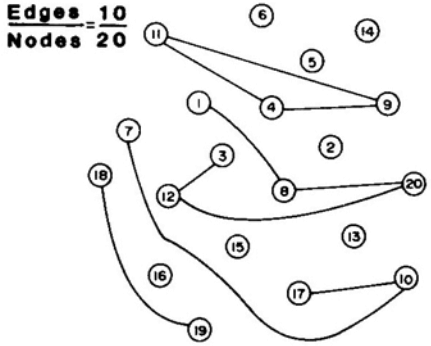
Behaviour with random graphs ?





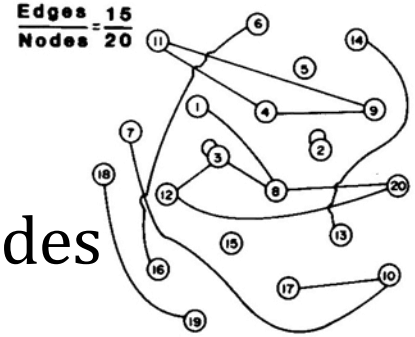
edges and connectivity

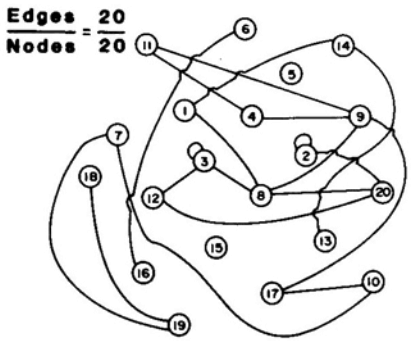


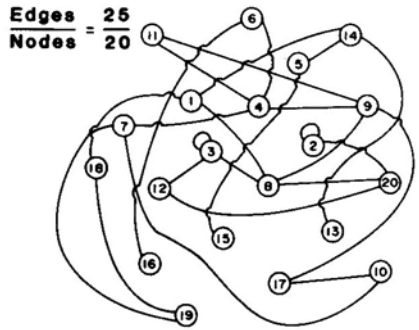


Standard results

- as edges ≈ nodes/2
  - most components are connected







When edges≈nodes

- cycles appear

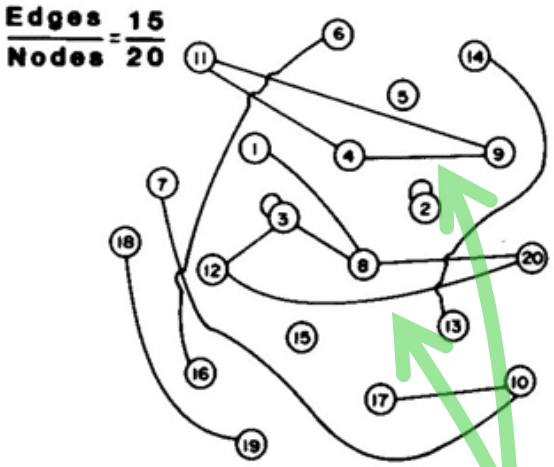
Those nodes in cycles

- can be synthesised using only other components in the cycle
- probability of cycles is near 1

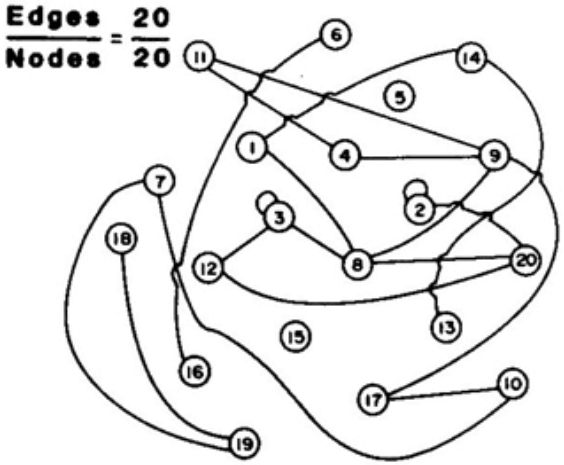
Connectivity

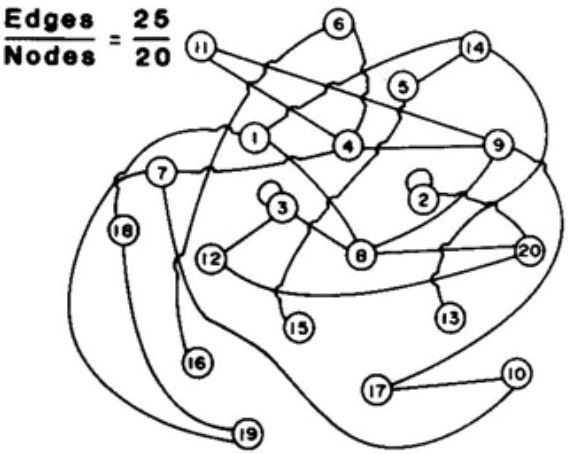
As soon as I have a cycle..

- Self-reproducing system… Life ?



here ?





Catalytic cycles

Gross simplifications

- no specificity
- one enzyme does all XY bonds regardless of context
- all rates the same…

Reasoning valid for 4 bases (RNA) or 20 residues (protein)

Auto-catalytic model

Without real "information" system

- self reproducing
- may have errors, tolerance of errors = evolution

- life may emerge suddenly
- order appears suddenly (Entropy disappears ..OK ?)

Autocatalytic model consequence

Anti-evolution

- what are the chances of molecules coming together to form a 200 residue protein ?
- what are the chances of a hurricane blowing bricks and building a house ?

This model

- the hurricane does not have to re-assemble a house
- any self-sustaining network will do
- our world is just one outcome
- whatever chemistry is most successful…

Experimental evidence

Not like ribosomes (difficult to explain without an RNA world)

Artificial systems.. example

RNA example of cooperating cycles

- ribozyme with four regions, ABCD
- three autocatalytic reactions
  - A+BCD →ABCD
  - AB+CD →ABCD
  - ABC+D →ABCD

- ABCD is a better catalyst than the parts
- recognition / pairing site can be varied
- possibility of cooperation
  - $A_1 + B_1\ C_1D_1 \xrightarrow[A_2B_2C_2D_2]{} A_1B_1C_1D_1$

Vaidya,N Manapat, ML. Chen, IA Xulvi-Brunet,IR, Hayden, EJ, Lehman, N, Nature, 2012, 491, 72

RNA example of cooperating cycles

Throw all ingredients into bucket

- A, AB, ABC, BCD, CD, D × sequence variants

48 possible products (comes from joining and sequence)

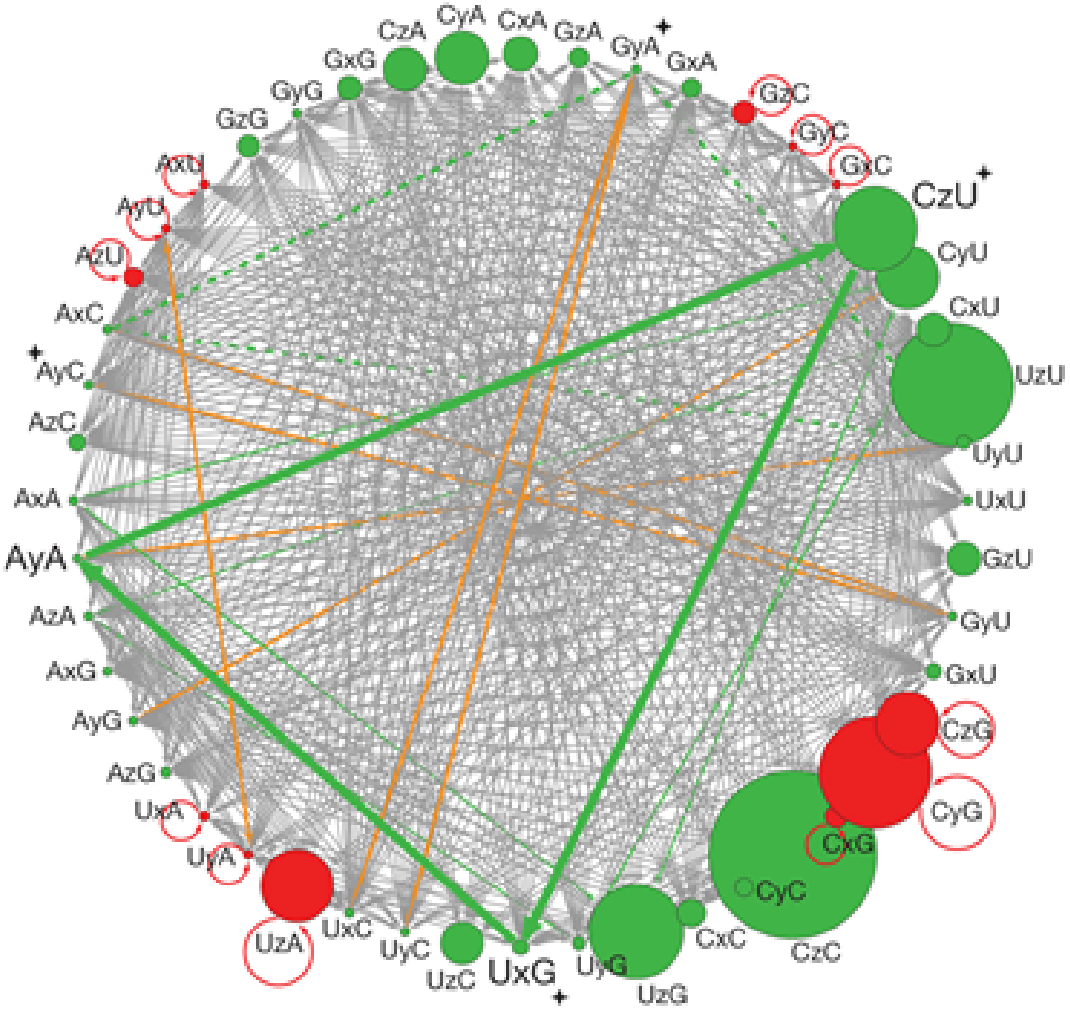
Results ?

48 products

size: how much of product after 8 hr

red: autocatalysts

green: cooperators



Claim:

cooperators are winners

Proof ? No – nice example of feasibility

Vaidya,N Manapat, ML. Chen, IA, Xulvi-Brunet, IR, Hayden, EJ, Lehman, N, Nature, 2012, 491, 72

For an Exam

- characteristics of life
- evidence for RNA world
- problems with RNA world
- auto-catalytic models

Summary

- life, entropy, information
- evolution, errors and tolerance of errors
- RNA world
  - ribosome – strong evidence
  - search for (possibly indirect) template directed replication
  - difficult to specify exact reactions producing
    - activated monomers
    - polymers
- search for simple template-directed replication may not be necessary
- self reproducing system may spontaneously form