Lecture plan

	Termin	Dozent	Thema
1	1.4.	Andrew	RNA Chemie
2	8.4.	Andrew	Ursprung der RNA-Welt
3	15.4.	Uli	Turnover (Synthese, Abbau, Splicing, Editing)
4	22.4.	Nicolas	nicht kodierende RNAs in Prokaryoten
5	29.4.	Nicolas	nicht kodierende RNAs in Eukaryoten
6	6.5.	Nicolas	RNA Interferenz
7	20.5.	Nicolas	Riboswitches
8	27.5.	Nicolas	Evolution *in vitro*
9	3.6.	Andrew	Strukturvorhersage I
10	10.6.	Andrew	Strukturvorhersage II
11	17.6.	Andrew	Strukturvorhersage III
12	24.6.	Uli	RNasen+Ribozyme
13	1.7.	Uli	das Ribosom
14	8.7.	Uli	Telomere und RNA

Seminar Thema

- try to put first page of each paper in stine
- first come, first served

The RNA world

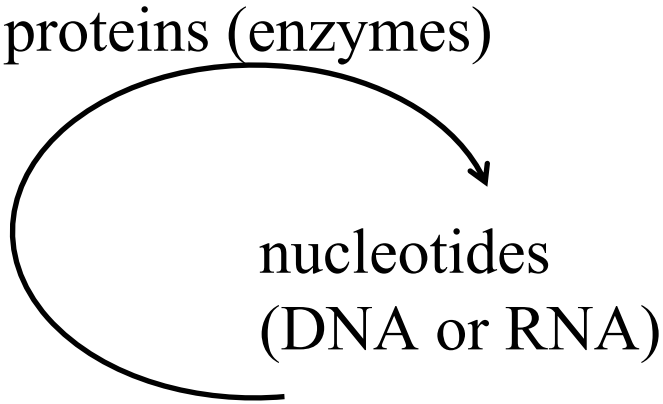
- Based on RNA World chapter 2
- Questions
  - did life start with RNA ?
  - RNA similar ?
  - something completely different ?

- What is life ?

Today versus history

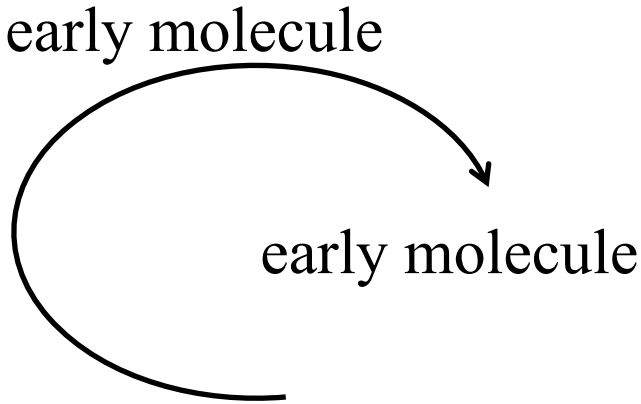
Picture today

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



Suggestion

- one molecule
  - self copying
    - possibilities
    1. protein like
    2. nucleotide like
    3. something else
- Remember: this is templated
  - later remove this requirement



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

Equilibrium

- Reaction A + B ↔ C + D
- Decay A ↔ B + C, then

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

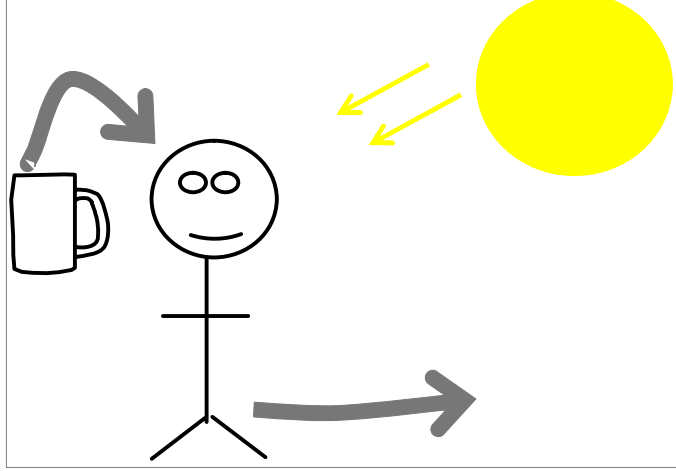
- in a closed system, if

$$\Delta G = RT \ln \frac{[B][C]}{[A]}$$

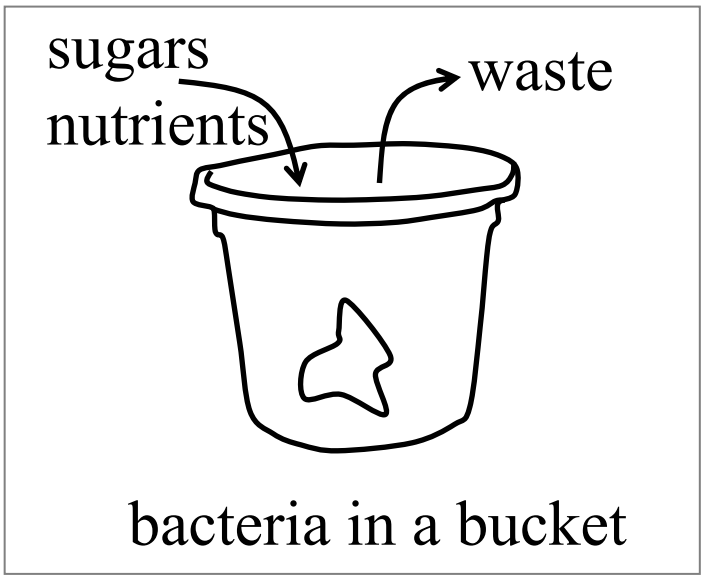
$$\ln \frac{[B][C]}{[A]} = 0$$ you are dead (or very sleepy)

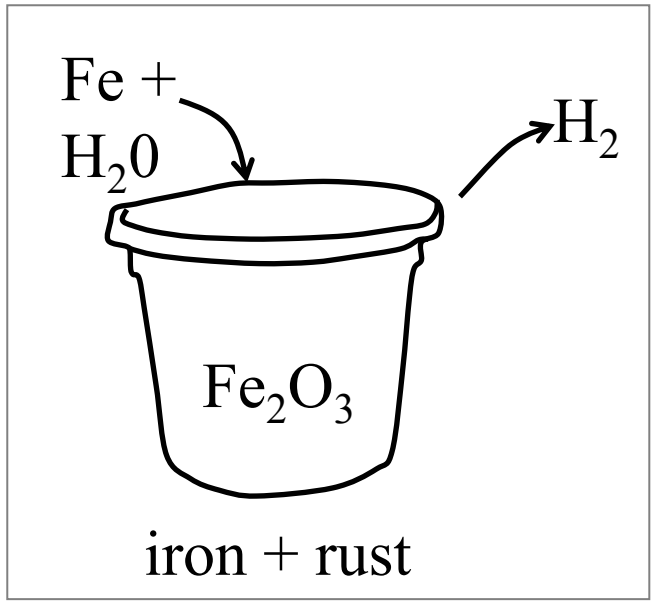
- how do we treat this ?
  - life is in "steady state"
- flux of A = - flux of BC so J_A=-J_{BC}

Steady state systems



- Input of energy
  - maintenance of order
- grows
- catalytic and specific





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

Rust

- why is rust not life (what we would like)
  - rather low information
  - no ability to change and evolve

information / entropy

- entropy is easy to define
  - N_{states} equal probability $S = -k \ln N_{states}$
  - or with different probabilities $S = -k \sum_{i=1}^{N_{states}} p_i \ln p_i$

- life has information, but what is it ?

Information

- Sometimes information = entropy
- for alphabet size λ, information per character $h = -k\sum_{i=1}^{\lambda} p_i \ln p_i$
  - note if some letters are more common, info reduced
- why is rust not living ?
  - alphabet size is 1 – no information
- pretend a genome is a digit in alphabet of possible genomes
- what about an e. coli ? ≈ 5 million base pairs (5×10^6)
- how many states could e. coli's genome have ? $4^{5\times10^6} \approx 10^{3000000}$
  - of these possibilities, very few are used
  - "information" per genome is big

- Claim
  - evolution is information increase via selection

Complexity

- Smallest genomes
  - viruses – few proteins – parasitic
- free living ?
  - a few hundred proteins
- if life came from a simple soup
  - why are there no traces ?
  - is there a minimum complexity for life ?

Life

- rust can catalyse the production of rust
- the process of "life" can copy sequence$_1$ or sequence$_2$
- this flexibility necessary for evolution

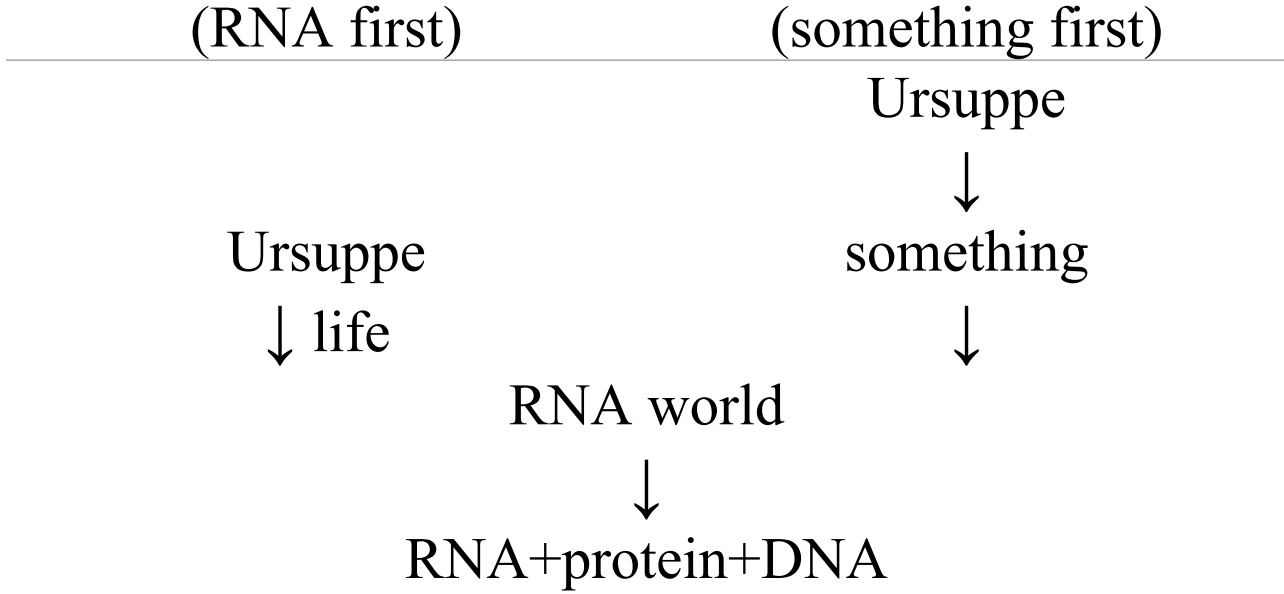
Summary of life

- not at equilibrium / consuming energy
- catalytic
- creating information
- copying with possibility of change / selection
- minimum complexity ? no evidence yet

RNA world existence

Two views

- how likely is the RNA first view ?



RNA world definition

- replication of RNA (directed)
- Watson-Crick base pairing
- no protein catalysis

  - did it exist ?

Why believe in an RNA world ?

Vague…

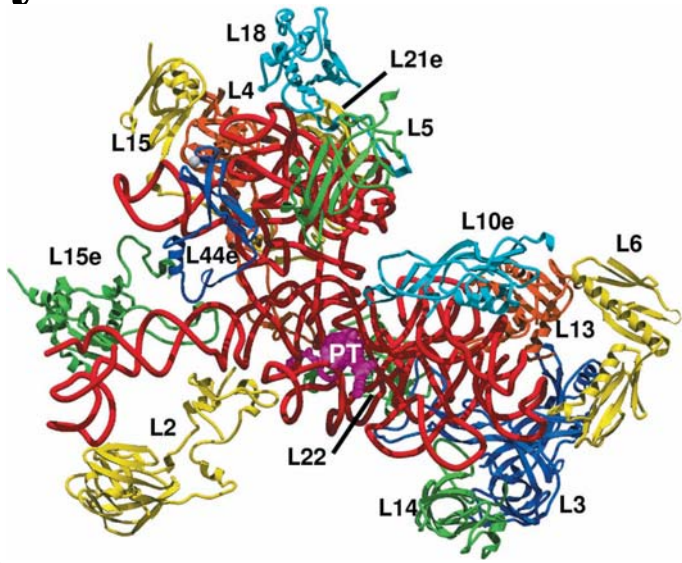
- nucleotides carry information
- many examples of catalytic activity

Strong

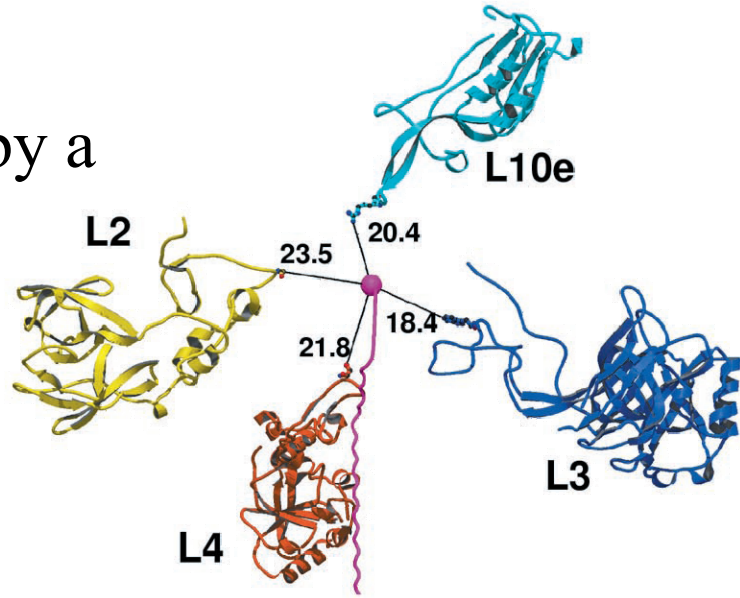
- ribosome..
- active site rather well known

Ribosome is an ribozyme

- 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

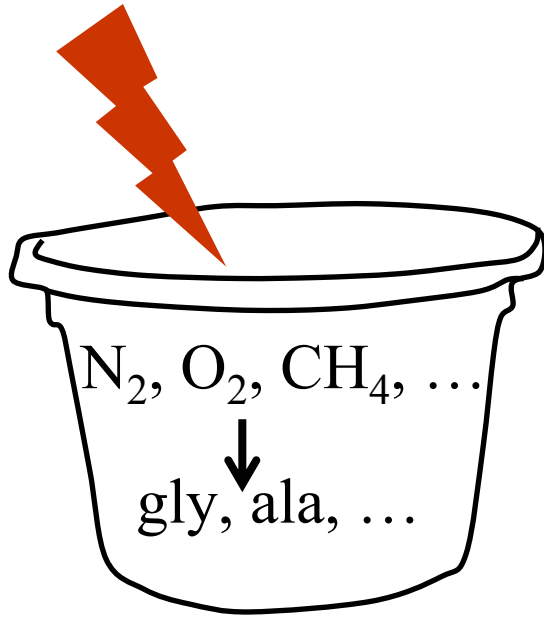


Nissen P, Hansen J, Ban N, Moore PB, Steitz TA., Science 289, 920-930 (2000)

RNA World – requirements

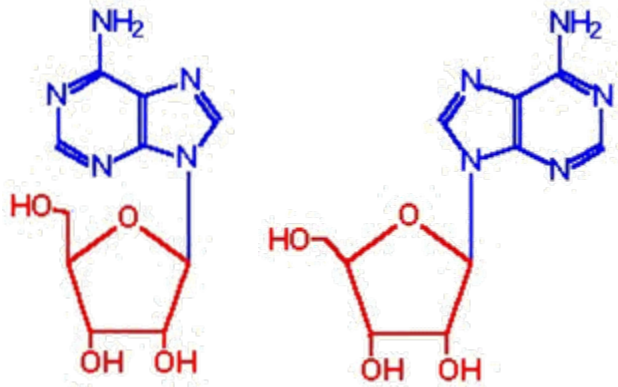
- source of basic requirements
  - ribose
  - bases (A, C, G, U + more T, I, X, …)
- vague source
  - Miller experiments from 1950's
- can one make nucleosides ? nucleotides ?
  - polynucleotides ?

- lots of problems



Specificity

- 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
- a list of problems..



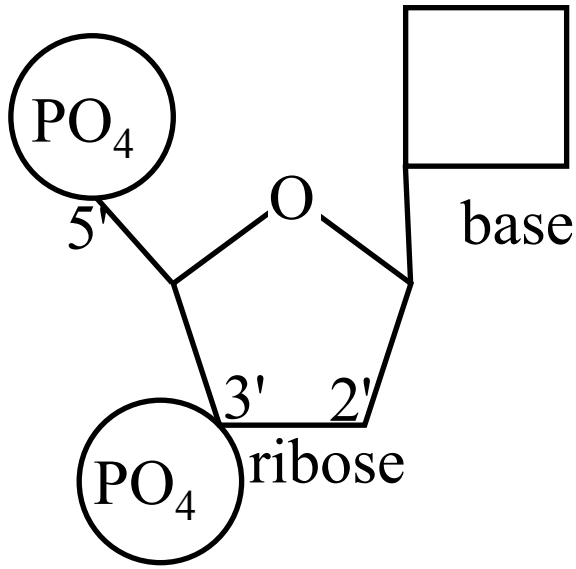
syn vs *anti*

polynucleotide synthesis (problems)

- nucleotide (base+sugar+PO_4) → polyphosphate (base+sugar+$(PO_4)_n$)
  - plausible without enzymes
- nucleoside +PO_4→NTP more difficult without enzymes
  - necessary in modern life chemistry
  - enzyme requirement
    - good to distinguish life from random chemistry
    - hard to see in random soup (urschleim)

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

5' to 3' who cares ?

Mental picture / preconception

- one RNA acts as a template for itself or inverse
- regular geometry

Possibility – catalysis of one "regioselective" product by

- metal ions
- adsorption on surfaces, minerals
- specially folded RNA

Replication without enzymes

- possibility – soup with 4 activated bases (N- 5'$PO_4$2-Me-imidazole) where N is A, G, C, U
- add poly C – directed copying to mostly poly G
- problem
  - poly-G like to form tetramers
- poly (G_mC_n) copies to poly (G_nC_m) but
  - when [G] > [C]
    - tends to form self interactions – not good as a template

- is this a property of
  - their model system ?
  - the specific sequences ? (good ones may exist)

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

- without errors – no evolution
- interesting proposal later..

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

Joyce / Orgel – 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
  - physical separation of templates rather than true selectivity

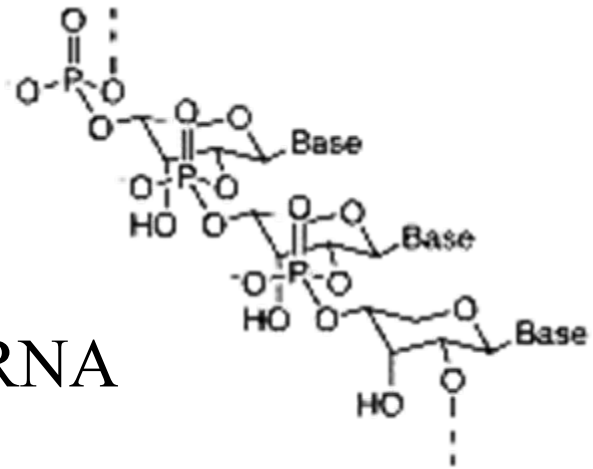
How to make nucleotides ?

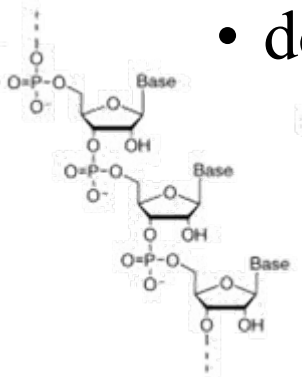
- N-sugar + xxx-PO4 $\rightarrow$N-sugar-PO_4 (+H_20)
- ribozymes have been made for related reactions
  - quite plausible
- abiotic ?
  - many examples of catalysis exist
    - Pb^{2+}, BO_3^{3-}, …
  - full catalysis requires a series of steps
    - each with specific catalysts

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
  - should form an open (helical) structure
- examples
  - replace ribose with pyranose (p-RNA)
    - stable, helical
    - does NOT form paired dimers with RNA

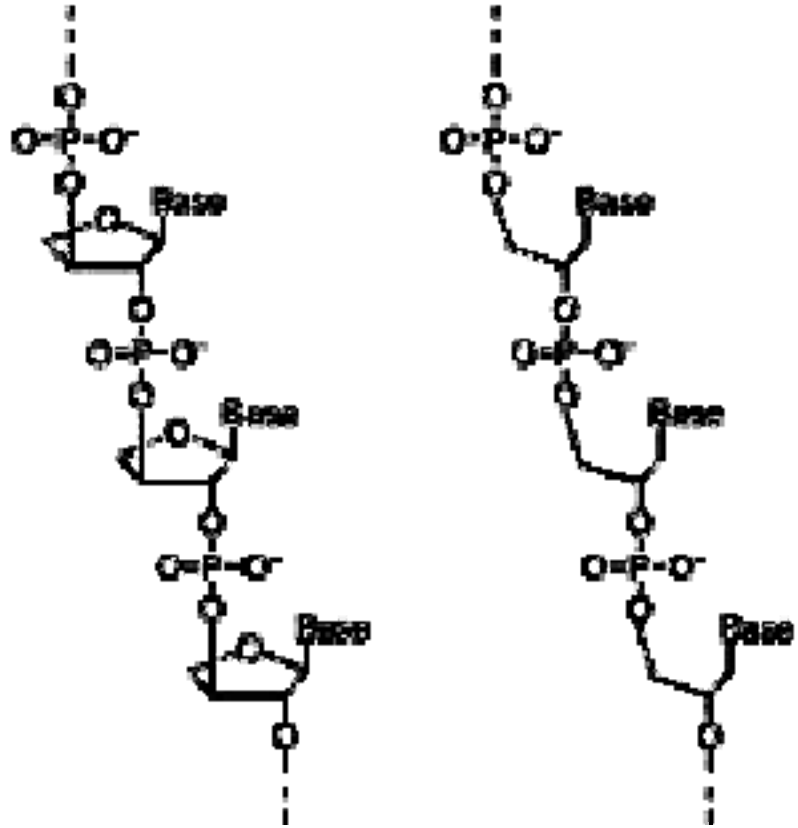




from Joyce, GF & Orgel, LE in The RNA World, (eds Gesteland, RF, Cech, TR, Atkins GF) Cold Spring Harbor Lab Press 2006

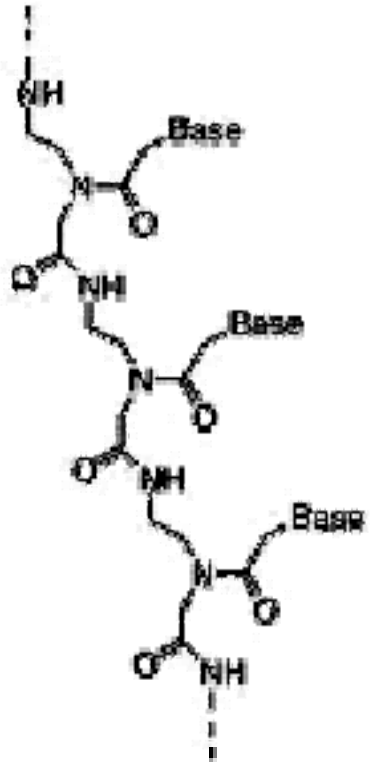
different sugars RNA

- use threose (left) TNA
  - forms stable double helix
  - threose may be easier to make
- use glycol (right) GNA
  - also forms double helices



peptide RNA

- a peptide backbone (PNA)
  - forms stable helices with DNA or RNA
    - suggests templating of modern nucleotides
- even more
  - alternating L and D-alanine
- summary…
  - pyranose (p-RNA)
  - tetrose (RNA)
  - ethylene glycol (GNA)
  - peptide NA (PNA)
  - alanyl NA (ANA)
  - …



Do we need this general templating ?

So far – search for general replicase, polymerase

- Can one build a living system from less general components ?

Examples

- peptides made without ribosomes
  - antamanide
  - glutathione
- the "information" is stored in enzyme structures

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

Requirements for DNA or Protein world

1. RNA can catalyse formation and cleavage of internucleotide bonds
2. abiotic formation of the monomers
3. solutions must be concentrated (small volume)
4. anabolic flux (making larger polymers)
5. catalytic closure
   formation of each member of set is catalysed by some other member

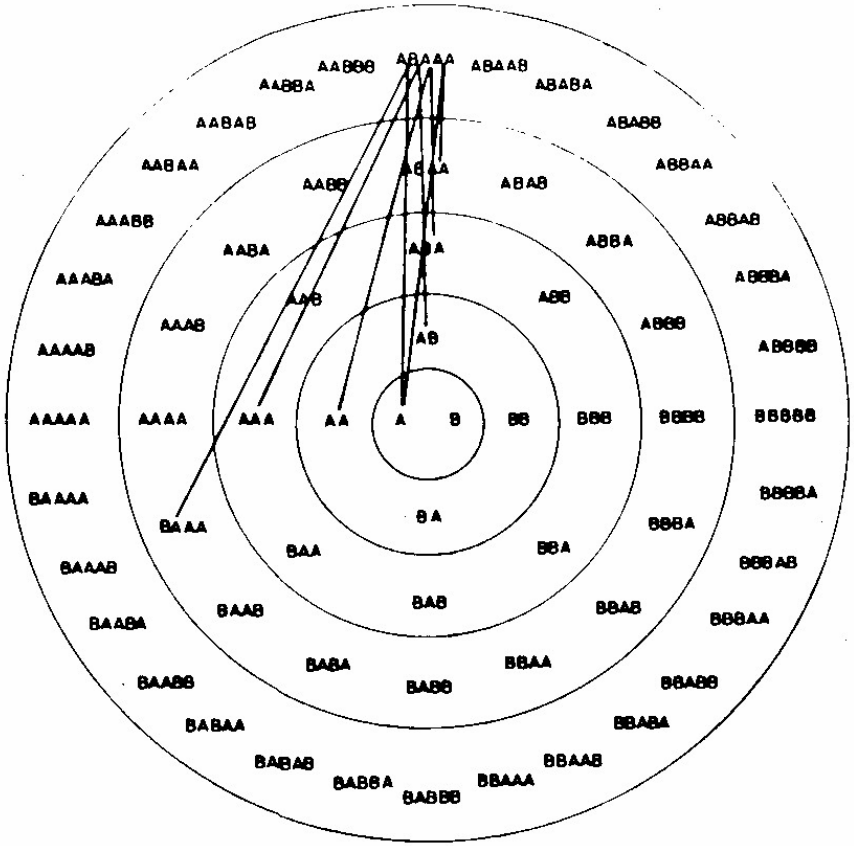
- we could apply these rules to proteins or nucleotides
  - change nature of monomer
- consider the first four problems

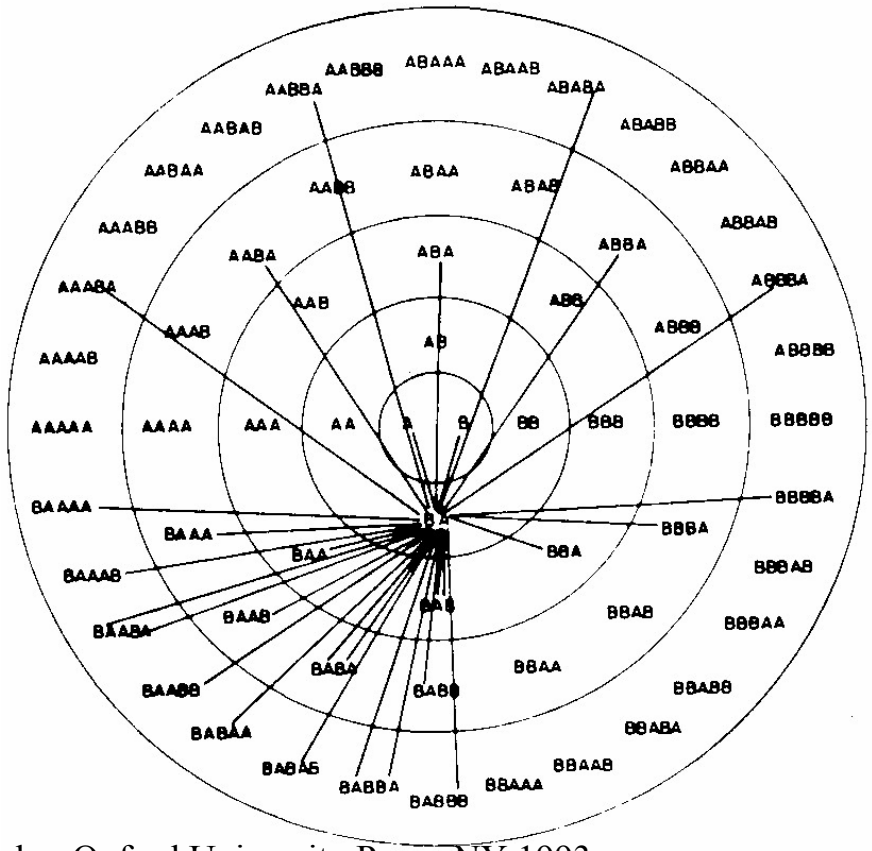
Some prerequisites are easy

1. proteolytic enzymes or ribozymes
2. tolerate a very imperfect soup of molecules, complex peptides or mixed 3',5' + 2', 5' nucleotides
3. confinement – drops, minerals, agglomerations
4. most reactions are $A+B \leftrightarrow AB + H_20$
   - removing water drives equilibrium to right
5. catalytic closure – not by simple templating

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



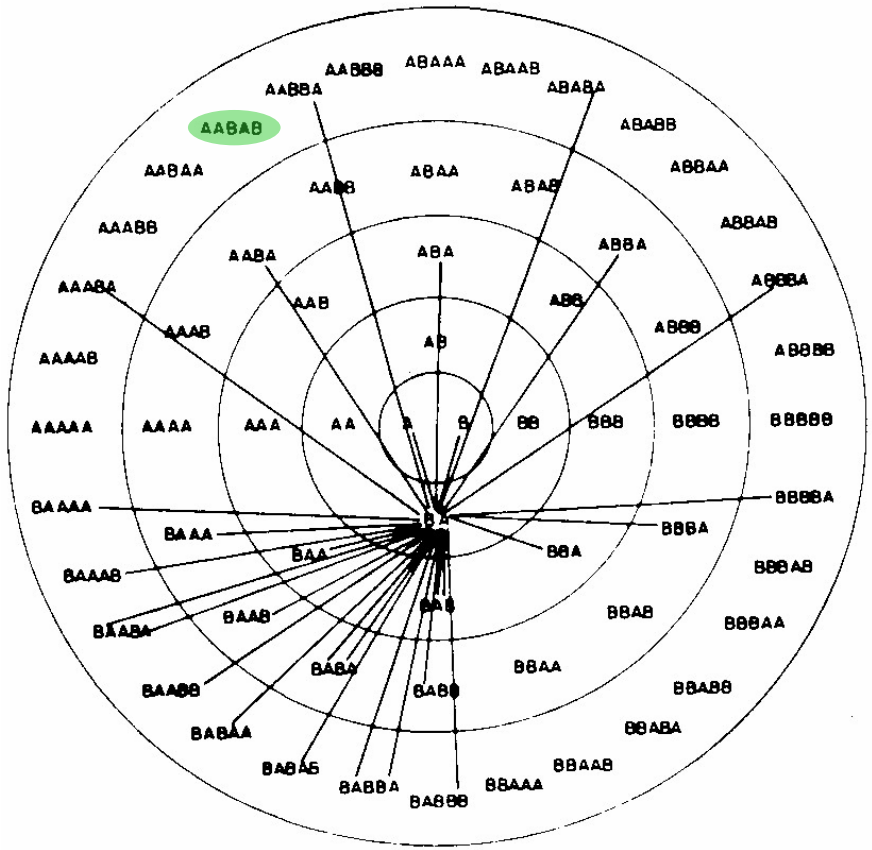


Kauffman, SA, The Origins of Order, Oxford University Press, NY 1993

Catalytic subset

- within set of polymers some are enzymatic for joining / breaking units
  - for RNA 4×4=16 X-Y types

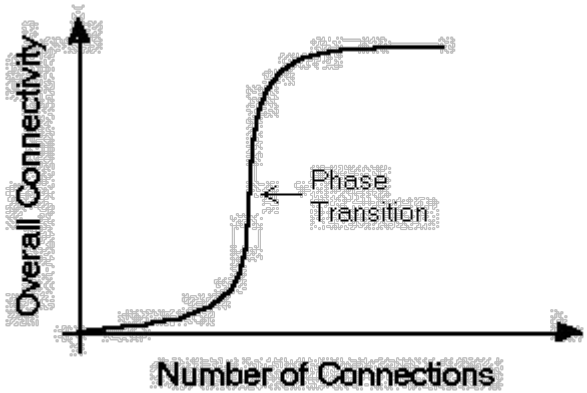
- pick a polymer
- with probability P pick a reaction it catalyses
- imagine green sequence does all AB bonds
  - leads to huge number of edges
- go to next sequence, maybe assign a reaction

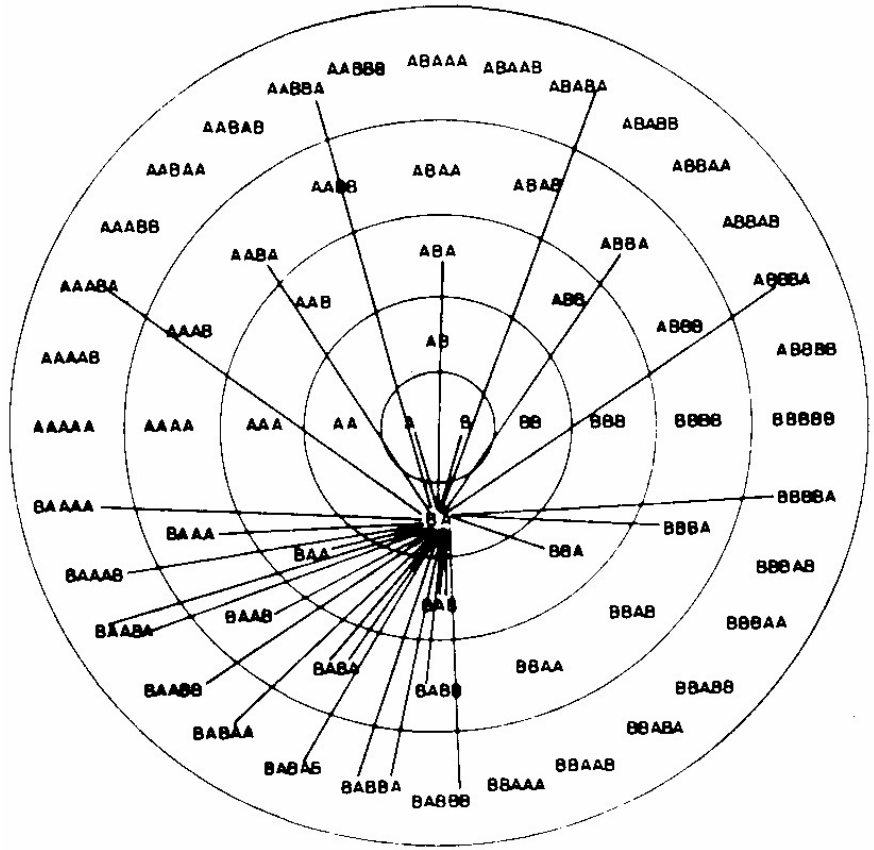


Catalytic subset

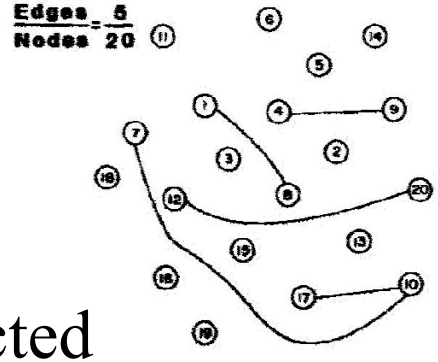
- how many real enzymes and edges do we need ?
- I do not have to be able to synthesise everything

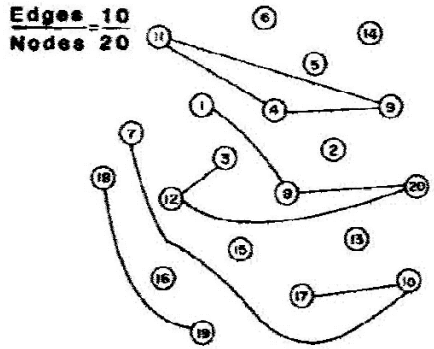
- what is the behaviour with random graphs ?



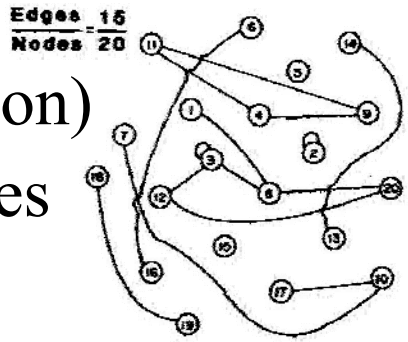


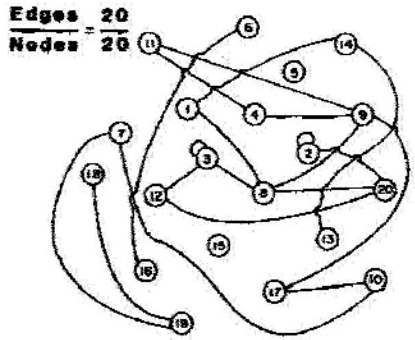
edges and connectivity

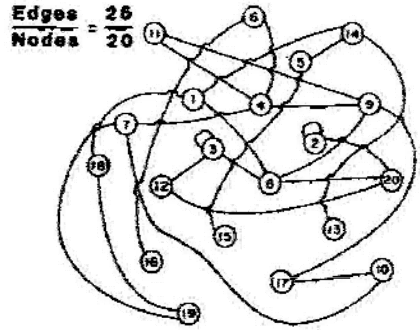




- standard results
- as edges ≈ nodes/2
  - most components are connected
  - very sudden
  - (phase transition)
- when edges≈nodes
  - cycles appear







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

Catalytic cycles

- Gross simplification here
  - random graph edges (plausible ?)
  - 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)
- main idea
  - without real "information" system
    - self reproducing
    - minimum complexity (mentioned earlier)
    - appears to self-replicate (very indirectly)
    - may have errors, tolerance of errors = evolution

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 easily spontaneously form