Intro 1: Last week's take home lessons

Life & computers : Self-assembly

Math: be wary of approximations

Catalysis & Replication

Differential equations: dy/dt=ky(1-y)

Mutation & the single molecule: Noise is overcome

Directed graphs & pedigrees

Bell curve statistics: Binomial, Poisson, Normal

Selection & optimality

Intro 2: Today's story, logic & goals

Biological side of Computational Biology

- Elements & Purification
- Systems Biology & Applications of Models
- Life Components & Interconnections
- Continuity of Life & Central Dogma
- Qualitative Models & Evidence
- Functional Genomics & Quantitative models
- Mutations & Selection

Elements For most NA & protein backbones:

C,H,N,O,P,S

6+13 Useful for many species:

Na, K, Fe, Cl, Ca, Mg, Mo, Mn, Se, Cu, Ni, Co, **Si**

Group	1	2		3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Period																			
1	1 H																		2 He
2	3 Li	4 Be												5 B	6 C	7 N	8 O	9 F	10 Ne
3	11 Na	12 Mg												13 Al	14 Si	15 P	16 S	17 Cl	18 Ar
4	19 K	20 Ca		21 Sc	22 Ti	23 V	24 Cr	25 Mn	26 Fe	27 Co	28 Ni	29 Cu	30 Zn	31 Ga	32 Ge	33 As	34 Se	35 Br	36 Kr
5	37 Rb	38 Sr		39 Y	40 Zr	41 Nb	42 Mo	43 Tc	44 Ru	45 Rh	46 Pd	47 Ag	48 Cd	49 In	50 Sn	51 Sb	52 Te	53 I	54 Xe
6	55 Cs	56 Ba	*	71 Lu	72 Hf	73 Ta	74 W	75 Re	76 Os	77 Ir	78 Pt	79 Au	80 Hg	81 Tl	82 Pb	83 Bi	84 Po	85 At	86 Rn
7	87 Fr	88 Ra	**	103 Lr	104 Rf	105 Db	106 Sg	107 Bh	108 Hs	109 Mt	110 Uun	111 Uuu	112 Uub	113 Uut	114 Uuq	115 Uup	116 Uuh	117 Uus	118 Uuo

*Lanthanoids	*	57 La	58 Ce	59 Pr	60 Nd	61 Pm	62 Sm	63 Eu	64 Gd	65 Tb	66 Dy	67 Ho	68 Er	69 Tm	70 Yb
**Actinoids	**	89 Ac	90 Th	91 Pa	92 U	93 Np	94 Pu	95 Am	96 Cm	97 Bk	98 Cf	99 Es	100 Fm	101 Md	102 No

From atoms to (bio)molecules

H_2O	H_2 , O_2	H^+ , OH^-
CH_4	C_{60}	CO_3^-
NH_3	N_2	NO_3^-
H_2S	S_n	SO_4^{--} Mg^{++}
PH_3		$K^+PO_4^{--}$ Na^+
Gas	**Elemental**	**Salt**

Purify

Elements,
molecules,
assemblies,
organelles,
cells,
organisms

chromatography

Clonal growth

Purified history

Pre 1970s: Column/gel purification revolution

Mid-1970s: Recombinant DNA brings clonal (single-step) purity.

1984-2002: Sequencing genomes & automation aids return to whole systems.

Intro 2: Today's story, logic & goals

Biological side of Computational Biology

Elements & Purification

•Systems Biology & Applications of Models

Life Components & Interconnections

Continuity of Life & Central Dogma

Qualitative Models & Evidence

Functional Genomics & Quantitative models

Mutations & Selection

"A New Approach To Decoding Life: Systems Biology" Ideker et al 2001

(http://arep.med.harvard.edu/pdf/Ideker01b.pdf)

1. **Define all components of the system.**

2. **Systematically perturb and monitor components of the system (genetically or environmentally).**

3. **Refine the model such that its predictions most closely agree with observations.**

4. **New perturbation experiments to distinguish among model hypotheses.**

Systems biology critique

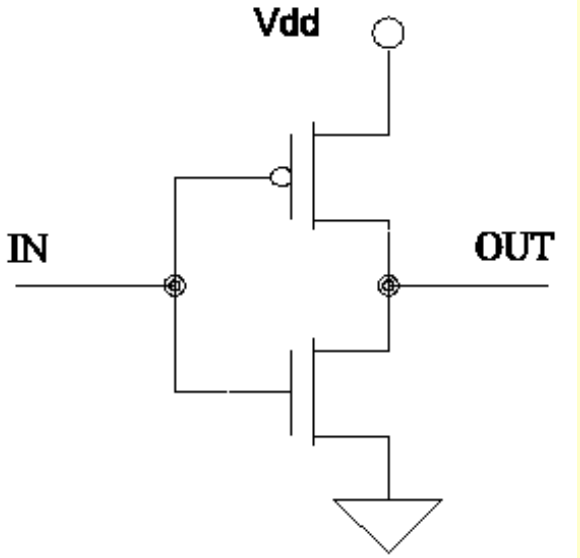
An old approach. New spins:
1. "*all* components" 2. "*Systematically* perturb"

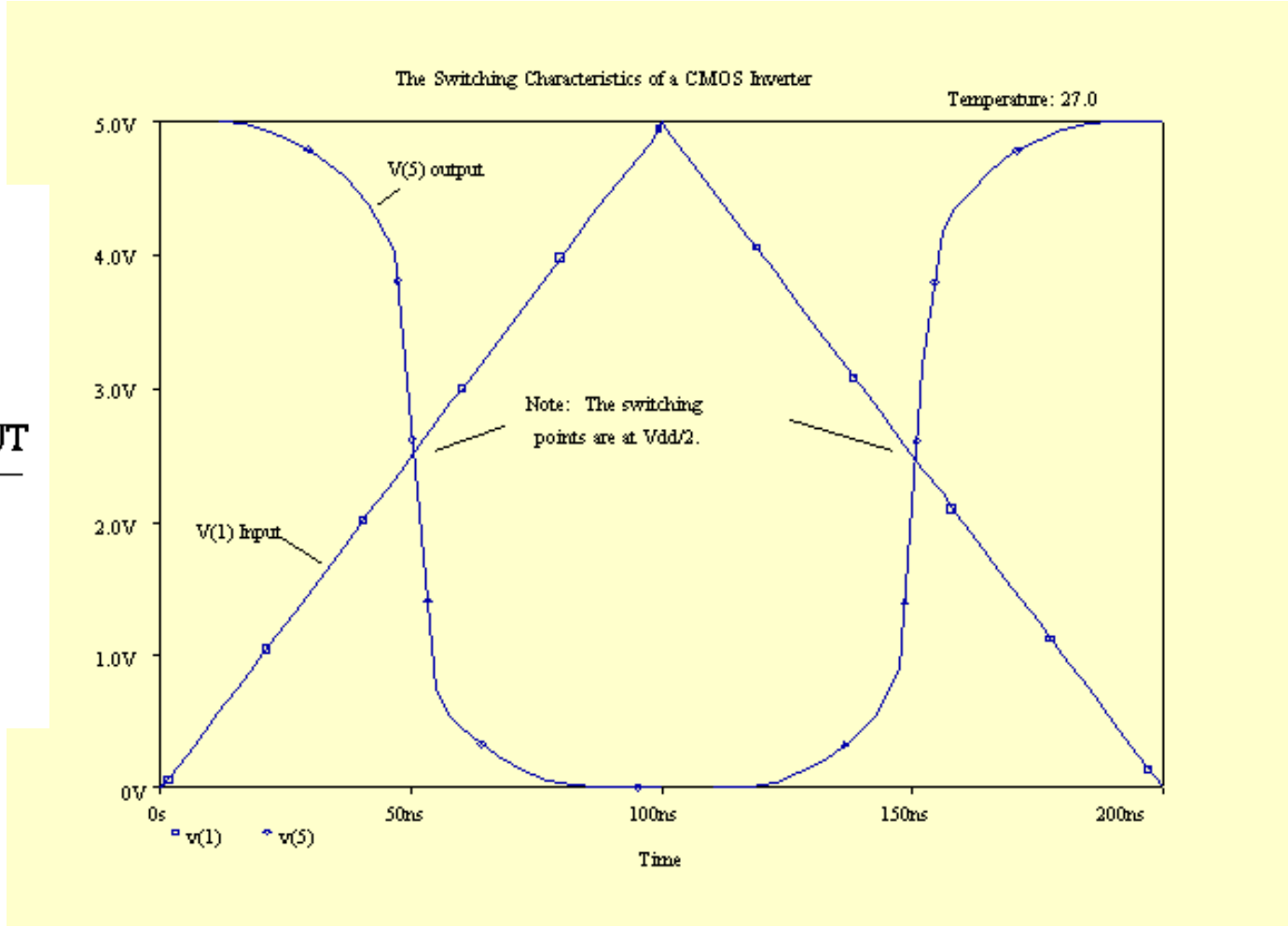
Unstated opportunities?
3. Refine the model without overfitting.
Methods to recapture unautomated data.
Explicit(automatic?) logical connections.

4. Optimization of new perturbation experiments & technologies.

Automation, *ultimate applications*, & synthetics as standards for: search, merge, check

Transistors > inverters > registers > binary adders > **compilers** > application programs





Spice simulation of a CMOS inverter(figures) (http://et.nmsu.edu/~etti/spring97/electronics/cmos/cmostran.html)

Why?

#0. Why sequence the genome(s)? To allow #1,2,3 below.

#1. Why map variation?
#2. Why obtain a complete set of human RNAs, proteins & regulatory elements?
#3. Why understand comparative genomics and how genomes evolved? To allow #4 below.

#4. Why quantitative biosystem models of molecular interactions with multiple levels (atoms to cells to organisms & populations)?
To share information. Construction is a test of understanding & to make **useful products**.

Grand (& useful) Challenges

A) From atoms to evolving minigenome-cells.

- Improve *in vitro* macromolecular synthesis.
- Conceptually link atomic (mutational) changes to population evolution (via molecular & systems modeling).
- Novel polymers for smart-materials, mirror-enzymes & drug selection.

B) From cells to tissues.

- Model combinations of external signals & genome-programming on expression.
- Manipulate stem-cell fate & stability.
- Engineer reduction of mutation & cancerous proliferation.
- Programmed cells to replace or augment (low toxicity) drugs.

C) From tissues to systems

- Programming of cell and tissue morphology.
- Quantitate robustness & evolvability.
- Engineer sensor-effector feedback networks where macro-morphologies determine the functions; past (Darwinian) or future (prosthetic).

Intro 2: Today's story, logic & goals

Biological side of Computational Biology

Elements & Purification
Systems Biology & Applications of Models
•Life Components & Interconnections
Continuity of Life & Central Dogma
Qualitative Models & Evidence
Functional Genomics & Quantitative models
Mutations & Selection

Number of component types (guesses)

	Mycoplasma	Worm	Human
Bases	.58M	>97M	3000M
DNAs	1	7	25
Genes	.48k	>19k	34k-150k
RNAs	.4k	>30k	.2-3M
Proteins	.6k	>50k	.3-10M
Cells	1	959	10^{14}

(http://www.nature.com/cgi-taf/DynaPage.taf?file=/nature/journal/v409/n6822/full/409860a0_fs.html)

(http://www.ensembl.org/Genesweep/)

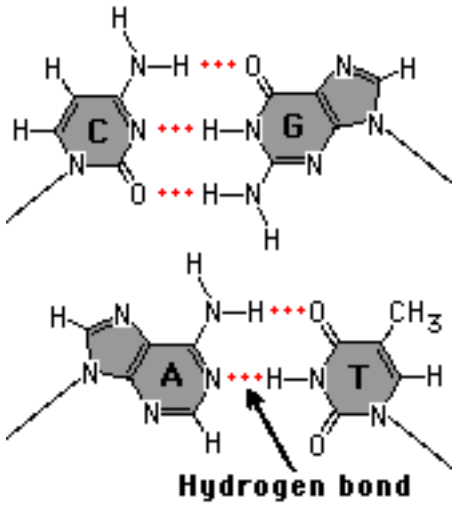
From monomers to polymers

Complementary surfaces

Watson-Crick base pair

(Nature April 25, 1953)

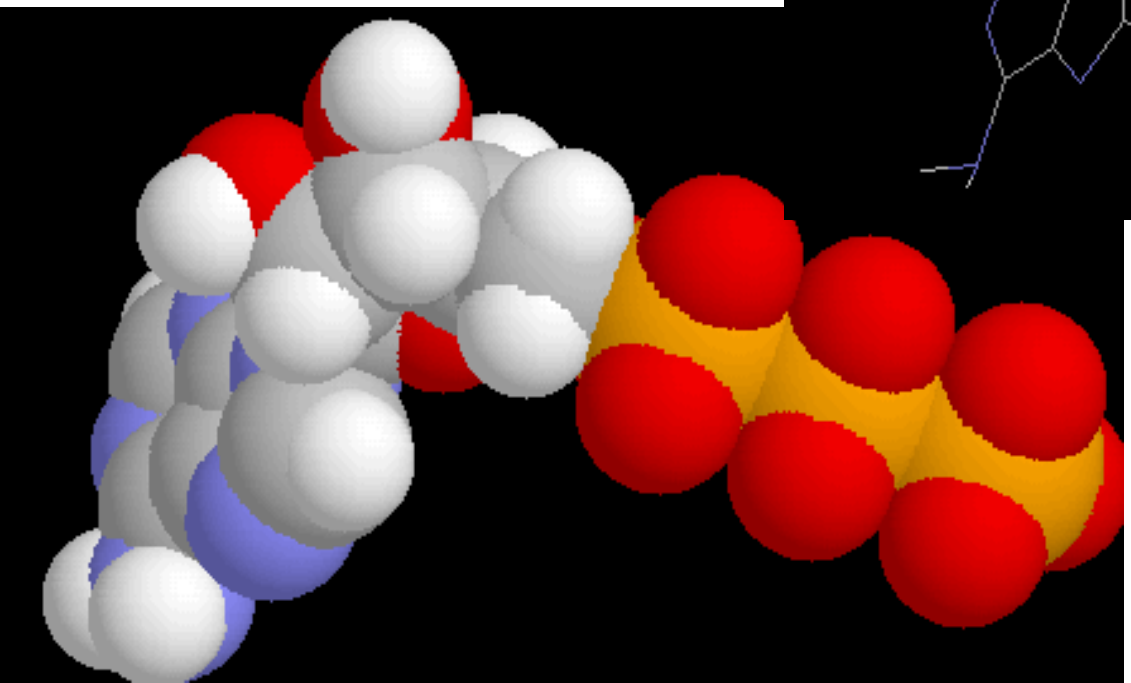
(http://www.sil.si.edu/Exhibitions/Science-and-the-Artists-Book/bioc.htm#27)



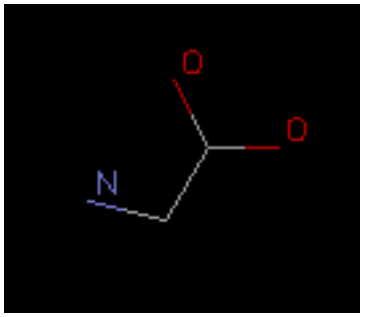
Nucleotides

dATP

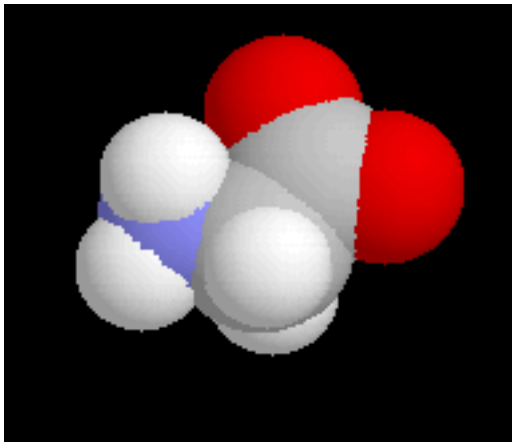
rATP



The simplest amino acid component of proteins



Glycine
Gly
G

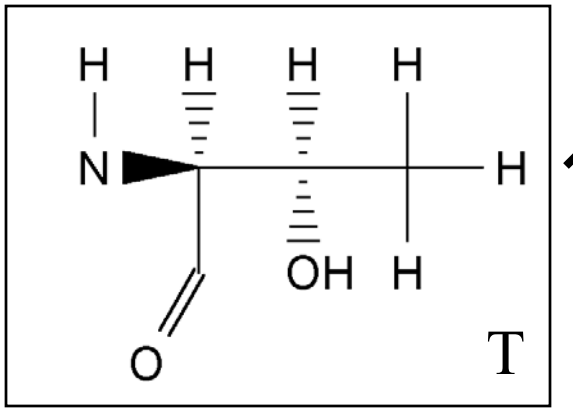


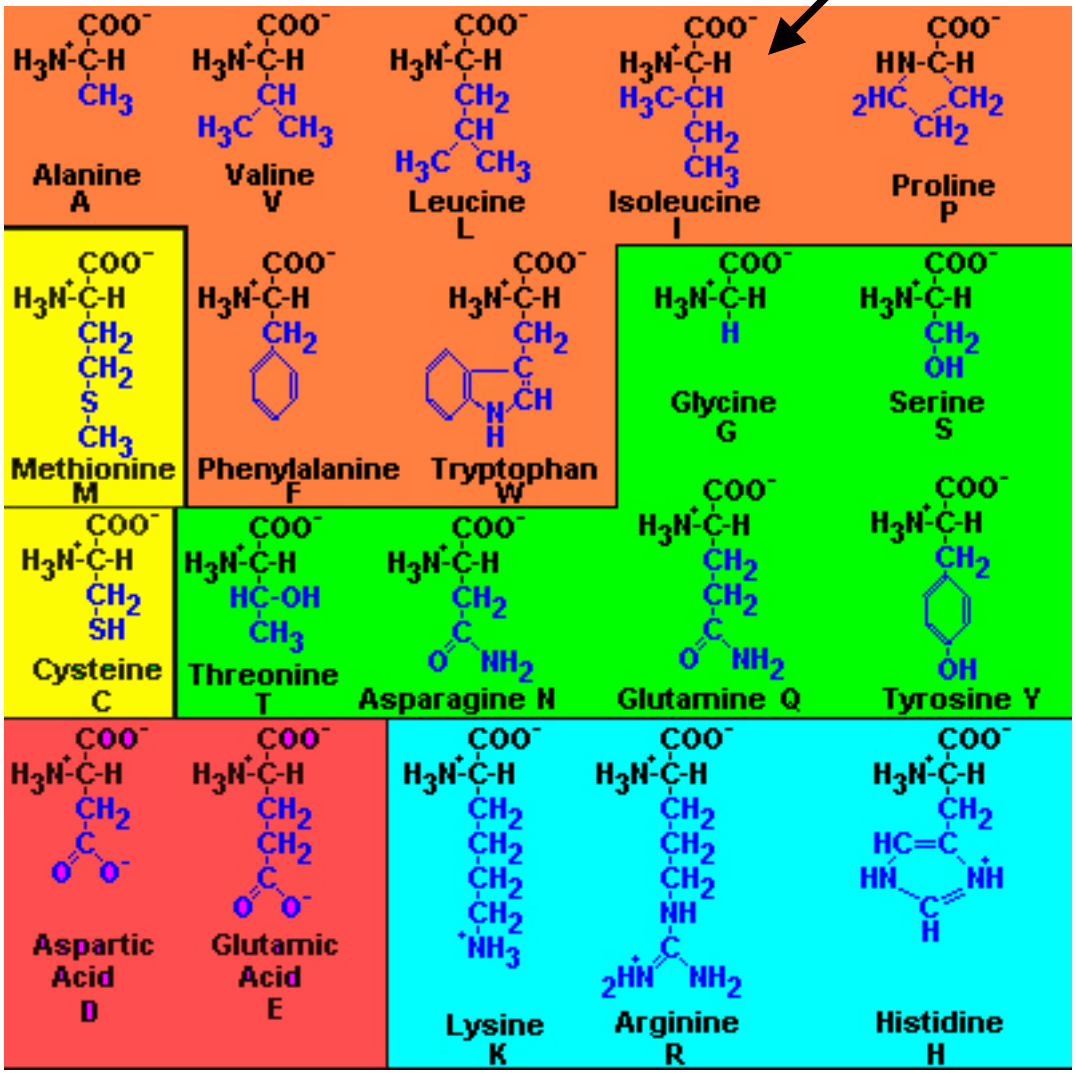
config(glycine,[
substituent(aminoacid_L_backbone),
substituent(hyd),
linkage(from(aminoacid_L_backbone,car(1)),
to(hyd,hyd(1)),
nil,single)]).

Klotho **(http://www.ibc.wustl.edu/klotho)**

Smiles String: [CH2]([NH3+])[C](=[O])[O-]

20 Amino acids of 280





www.people.virginia.edu/~rjh9u/aminacid.html

www-nbrf.georgetown.edu/pirwww/search/textresid.html

Intro 2: Today's story, logic & goals

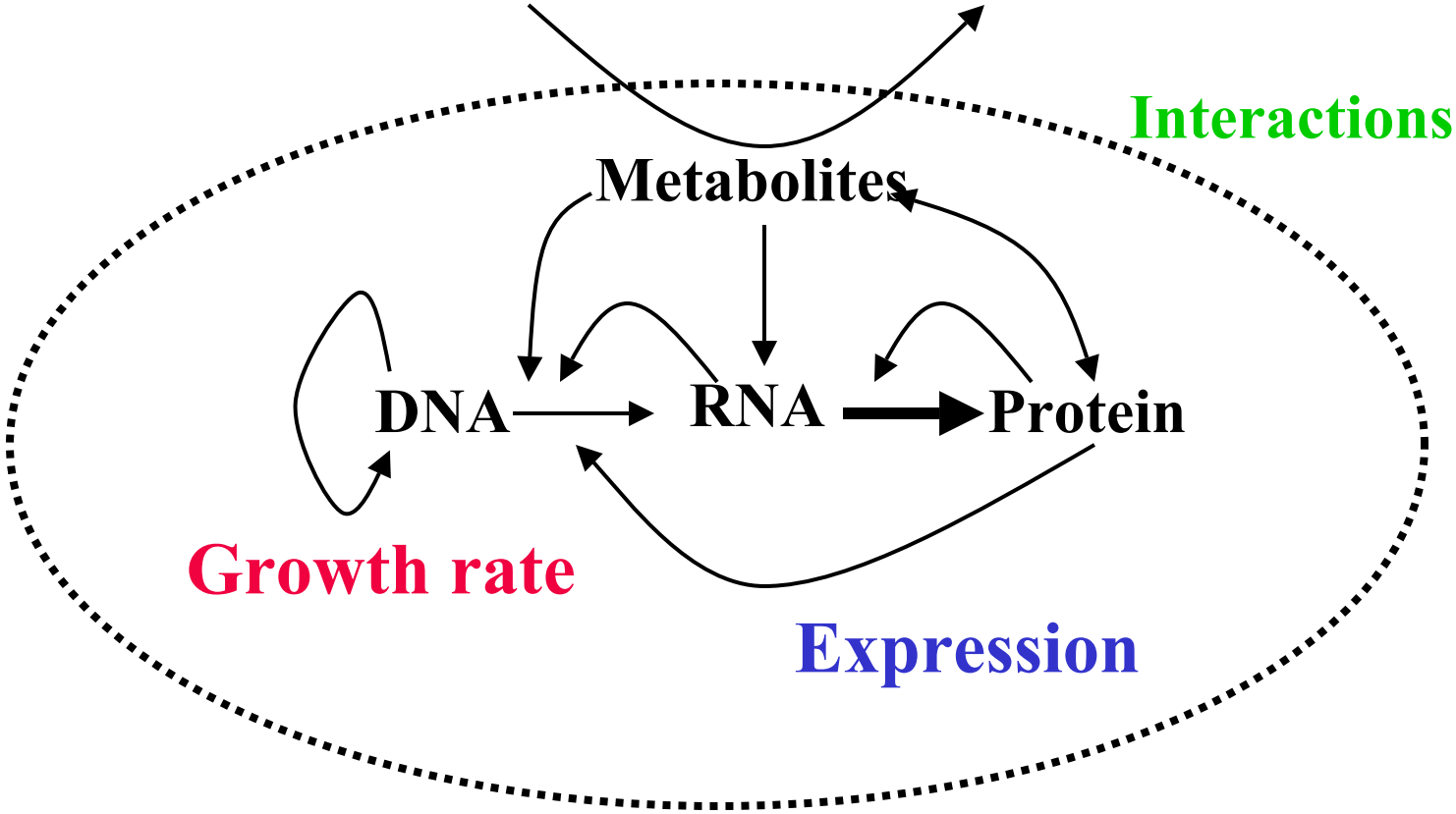
Biological side of Computational Biology

Elements & Purification
Systems Biology & Applications of Models
Life Components & Interconnections
•Continuity of Life & Central Dogma
Qualitative Models & Evidence
Functional Genomics & Quantitative models
Mutations & Selection

Continuity of Life & Central Dogma

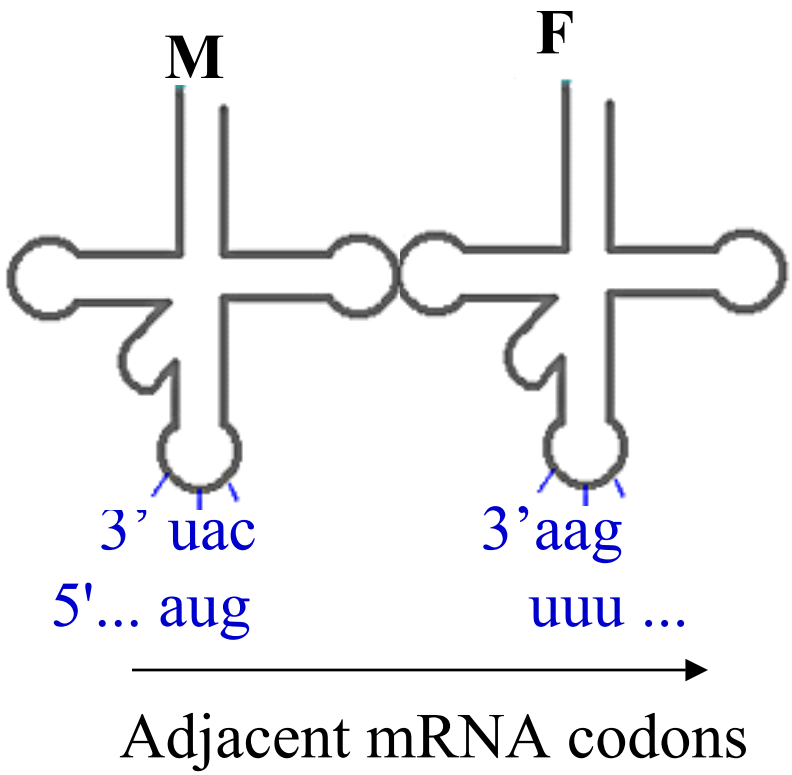
Self-assembly, Catalysis, Replication, Mutation, Selection

Regulatory & Metabolic Networks



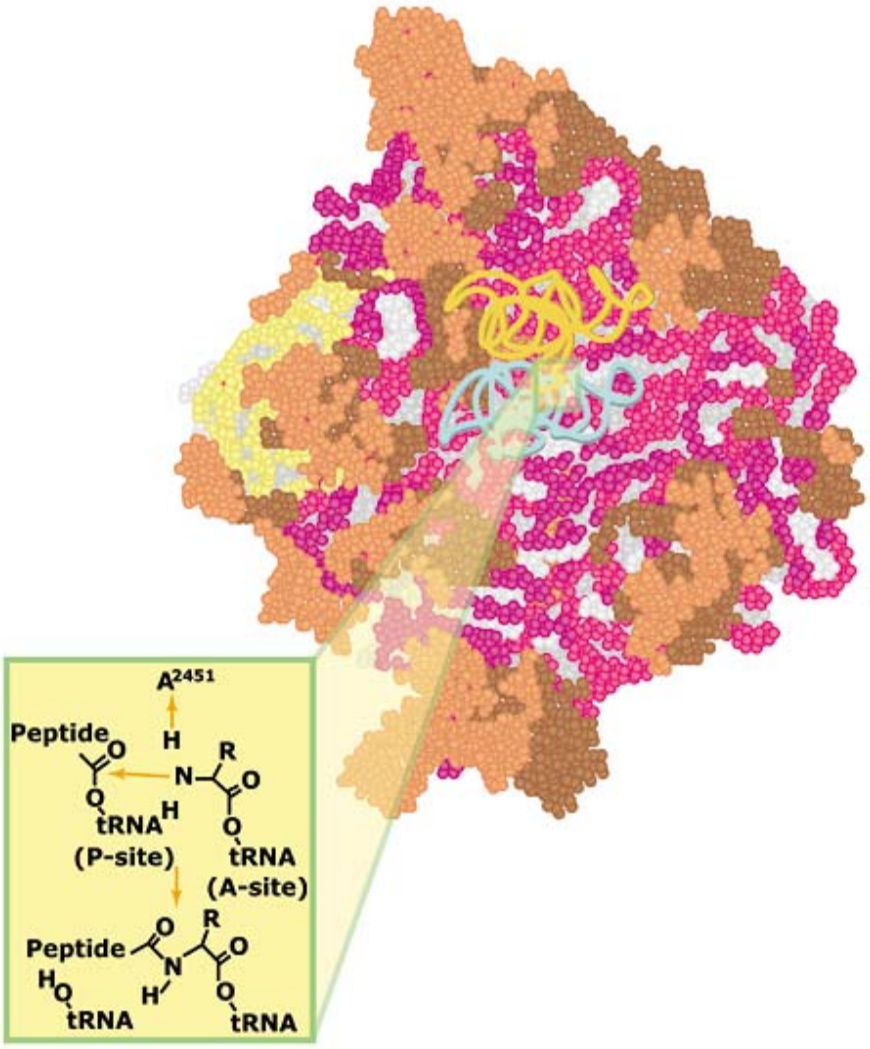
Polymers: **Initiate, Elongate, Terminate, Fold, Modify, Localize, Degrade**

"The" Genetic Code



x=	u□ □	c□	a□	g□	
Uxu□	**F**	**S**	**Y**	**C**	
uxc□					
uxa□			-	-	**TER**
uxg□			-	**W**	
Cxu□	**L**	**P**	**H**	**R**	
cxc□					
cxa□			**Q**		
cxg□					
axu□	**I**	**T**	**N**	**S**	
axc□					**C-S**
axa□			**K**	**R**	
axg□	**M**				**NH+**
gxu□	**V**	**A**	**D**	**G**	
gxc□					**O-**
gxa□			**E**		
gxg□					**H:D/A**

Translation
t-,m-,r-RNA



Large macromolecular complexes:
Ribosome: 3 RNAs (over 3 kbp plus over 50 different proteins)

Science (2000) 289: 878, 905, 920, 3D coordinates.
The ribosome is a ribozyme.
(http://www.sciencemag.org/cgi/content/full/289/5481/905),

(http://www.rcsb.org/pdb/cgi/explore.cgi?pid=8478969223009&pdbId=1FFK)

Perl Dogma (EditPlus) (http://www.editplus.com/)

```
 4  #### The Central Dogma ####
 5  ############################
 6
 7  ### Genome
 8  $DNA_seq = "ATGACCCTACTAGATCATCTATGATAGCTCAT";
 9
10  ### Transcription
11  $RNA_seq = $DNA_seq;
12  $RNA_seq =~ s/T/U/gi;
13  print "$RNA_seq\n";
14
15  ### Translation
16  $position = 0;
17  while (substr $RNA_seq,$position,3) {
18        $codon = substr $RNA_seq,$position,3;
19        print translate_codon($codon);
20        $position = $position + 3;
21  }
22  sub translate_codon {
23        if ($_[0] =~ /GC./i) {return Ala;}
24        if ($_[0] =~ /UGC|UGU/i) {return Cys;}
```

Continuity & Diversity of life

Genomes 0.5 to 7 Mbp **10 Mbp to 1000 Gbp**

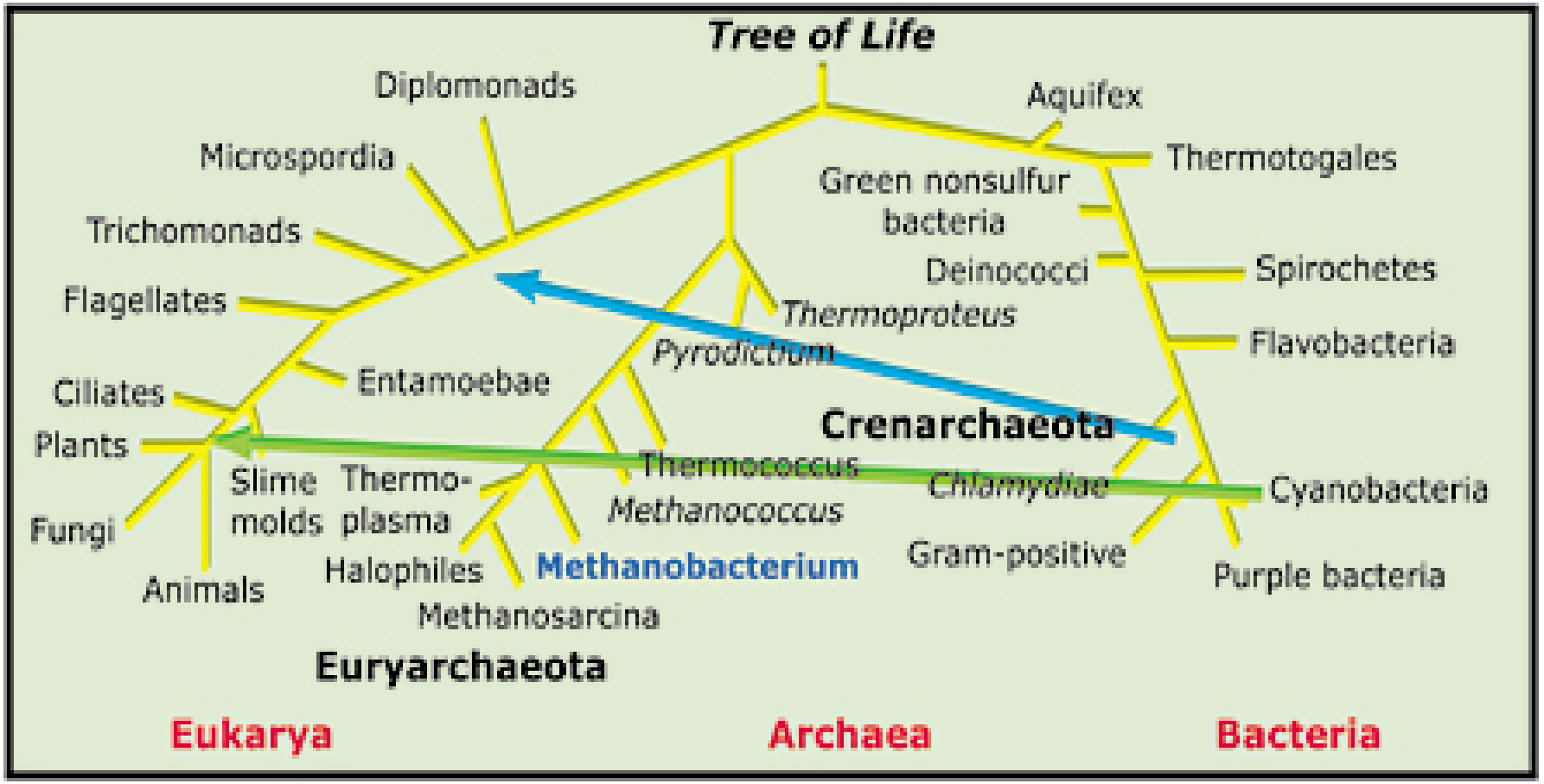


Figure (http://216.190.101.28/GOLD/)

How many living species?

5000 bacterial species per gram of soil (<70% DNA bp identity)
Millions of non-microbial species (& dropping)
Whole genomes: 45 done since 1995, 322 in the pipeline!(ref)
(http://216.190.101.28/GOLD/).
Sequence bits: 16234 (in 1995) to **79961** species (in 2000) NCBI

& Why study more than one species?
Comparisons allow discrimination of
subtle functional constraints.

Genetic codes (ncbi)

(http://www.ncbi.nlm.nih.gov/htbin-post/Taxonomy/wprintgc?mode=c)

```
1. "Standard Code"
Base1  = TTTTTTTTTTTTTTTTCCCCCCCCCCCCCCCCAAAAAAAAAAAAAAAAGGGGGGGGGGGGGGGG
Base2  = TTTTCCCCAAAAGGGGTTTTCCCCAAAAGGGGTTTTCCCCAAAAGGGGTTTTCCCCAAAAGGGG
Base3  = TCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAG
  AAs  = FFLLSSSSYY**CC*WLLLLPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGG
Starts = ---M---------------M---------------M---------------------------
2. The Vertebrate Mitochondrial Code
  AAs  = FFLLSSSSYY**CCWWLLLLPPPPHHQQRRRRIIMMTTTTNNKKSS**VVVVAAAADDEEGGGG
Starts = --------------------------------MMMM---------------M-----------
3. The Yeast Mitochondrial Code
  AAs  = FFLLSSSSYY**CCWWTTTTPPPPHHQQRRRRIIMMTTTTNNKKSSRRVVVVAAAADDEEGGGG
Starts = ----------------------------------MM---------------------------
11. The Bacterial "Code"
  AAs  = FFLLSSSSYY**CC*WLLLLPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGG
Starts = ---M---------------M------------MMMM---------------M-----------
14. The Flatworm Mitochondrial Code
  AAs  = FFLLSSSSYYY*CCWWLLLLPPPPHHQQRRRRIIIMTTTTNNNKSSSSVVVVAAAADDEEGGGG
Starts = --------------------------------M---------------------------
22. Scenedesmus obliquus mitochondrial Code
  AAs  = FFLLSS*SYY*LCC*WLLLLPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGG
Starts = --------------------------------M---------------------------
```

Translational reprogramming

Gesteland, R. F. and J. F. Atkins. 1996. Recoding - Dynamic reprogramming of translation (1996). Ann. Rev.Biochem 65:741-768

Herbst KL, et al. 1994
PNAS 91:12525-9
A mutation in ribosomal protein
L9 affects ribosomal hopping
during translation of gene 60
from bacteriophage T4.
"Ribosomes hop over a 50-nt coding gap during translation..."

Intro 2: Today's story, logic & goals

Biological side of Computational Biology

Elements & Purification
Systems Biology & Applications of Models
Life Components & Interconnections
Continuity of Life & Central Dogma
•Qualitative Models & Evidence
Functional Genomics & Quantitative models
Mutations & Selection

Qualitative biological statements (beliefs) and evidence

metabolism
cryptic genes
information transfer
regulation
    type of regulation
    genetic unit regulated
    trigger
    trigger modulation
transport
cell processes
cell structure
location of gene products
extrachromosomal
DNA sites

Riley, GeneProtEC
(http://zeus.mbl.edu/dbadmin/genprotec/index.php3?func=list&find=0)

MIPS functions
(http://www.mips.biochem.mpg.de/proj/yeast/catalogues/funcat/)

Gene Ontology (nature of being)

The objective of GO is to provide controlled vocabularies for the description of the molecular function, biological process and cellular component of gene products. **(http://www.geneontology.org/GO.doc.html)**

...

Many aspects of biology are not included (domain structure, 3D structure, evolution, expression, etc.)... small molecules (Klotho or LIGAND)

(http://www.ibc.wustl.edu/klotho) , (http://www.genome.ad.jp/kegg/catalog/compounds.html)

Gene Ontology

GO **(http://www.geneontology.org/GO.doc.html)**

- **Molecular function**

What a gene product can do without specifying where or when. (e.g. broad "enzyme" ; narrower "adenylate cyclase“)

- **Biological process**

>1 distinct steps, time, transformation (broad: "signal transduction." narrower: "cAMP biosynthesis.")

- **Cellular component**

part of some larger object, (e.g. ribosome)_

Evidence for facts

GO (http://www.geneontology.org/GO.doc.html)

IMP	inferred from mutant phenotype
IGI	genetic interaction
IPI	physical interaction
ISS	sequence similarity
IDA	direct assay
IEP	expression pattern
IEA	electronic annotation
TAS	traceable author statement
NAS	non-traceable author statement

Direct observation

See C.elegans cell lineage & neural connections

(http://paradise.caltech.edu/~gibson/research/lineage.jpg)

Sources of Data for BioSystems Modeling:

Capillary electrophoresis
(DNA Sequencing) :
0.4Mb/day

Chromatography-Mass Spectrometry
(eg. peptide LC-ESI-MS) :
20Mb/day

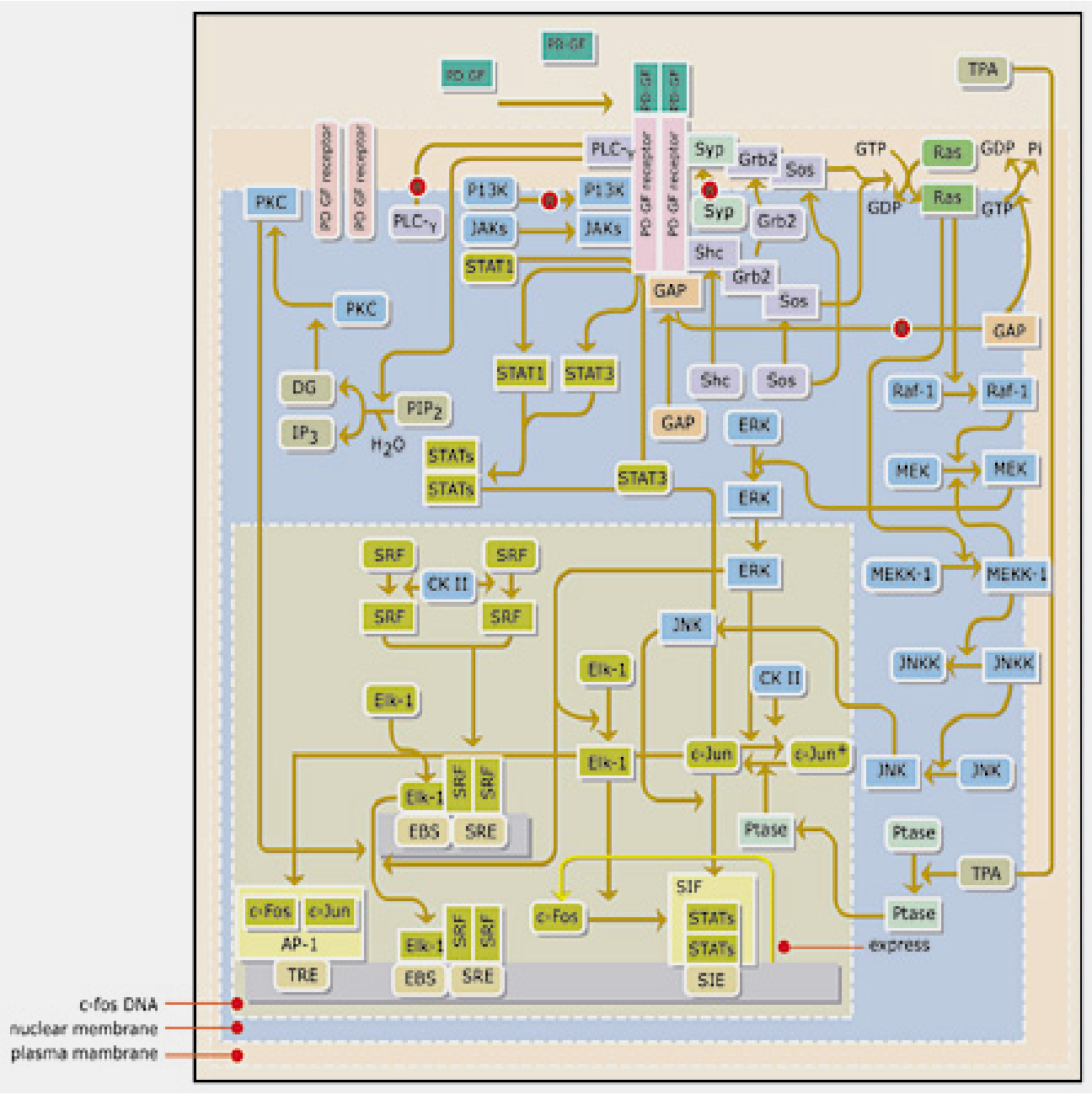
Microarray scanners (eg. RNA):
300 Mb/day **mpg**

Other microscopy (e.g. subcell, cell, tissue networks)

Signaling PAthway Database

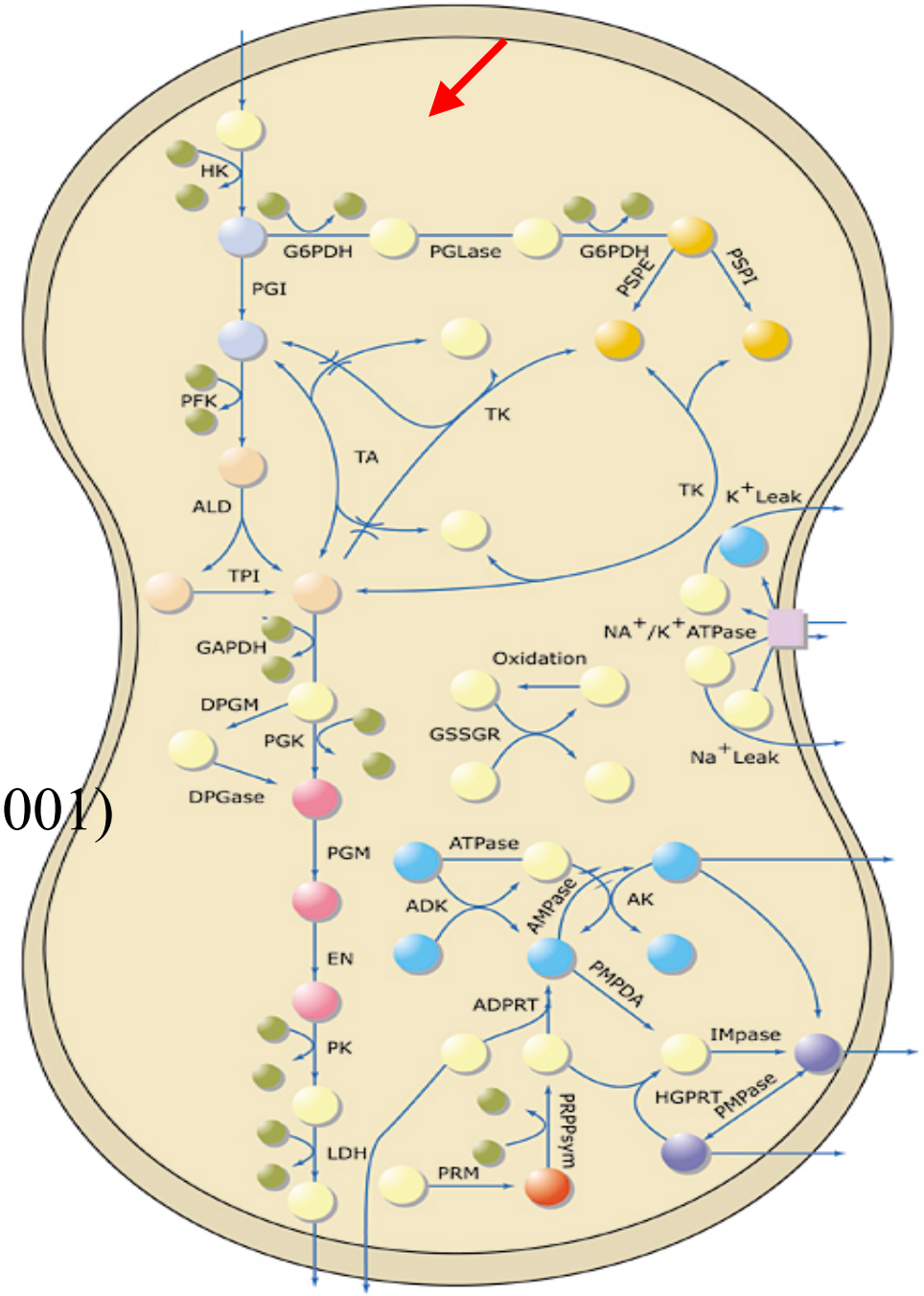
SPAD

(http://www.grt.kyushu-u.ac.jp/spad/menu.html)



Dynamic simulation of the human red blood cell metabolic network.

Jamshidi, et al(2001) Bioinformatics 17: 286-287.



Dominant alleles affecting variety of RBC proteins, malaria, drug-hemolysis, etc. Rare individually, common as a group.

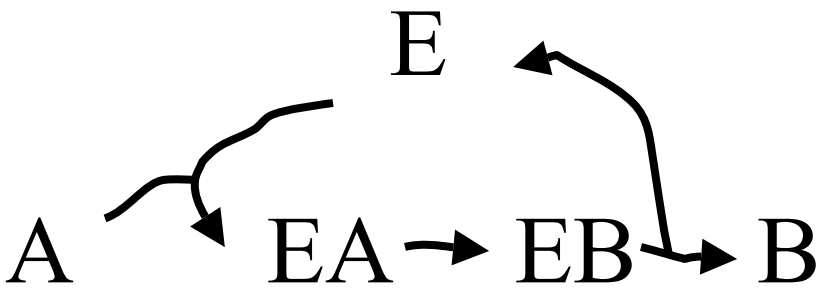
Enzyme Kinetic Expressions

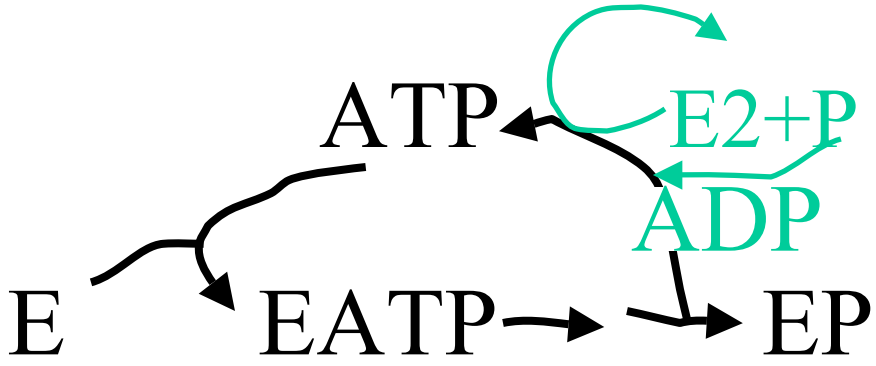
Phosphofructokinase

$$v_{PFK} = \frac{v_{mx}^{PFK}}{N_{PFK}} \left(\frac{F6P \Big/ K_{F6P}^{PFK}}{1 + F6P \Big/ K_{F6P}^{PFK}} \right) \left(\frac{Mg \bullet ATP \Big/ K_{Mg \bullet ATP}^{PFK}}{1 + Mg \bullet ATP \Big/ K_{Mg \bullet ATP}^{PFK}} \right)$$

$$N_{PFK} = 1 + L_0^{PFK} \frac{\left(1 + ATP_{free} \Big/ K_{ATP}^{PFK} \right)^4 \left(1 + Mg \Big/ K_{Mg}^{PFK} \right)^4}{\left(1 + AMP \Big/ K_{AMP}^{PFK} \right)^4 \left(1 + F6P \Big/ K_{F6P}^{PFK} \right)^4}$$

How do enzymes & substrates formally differ?



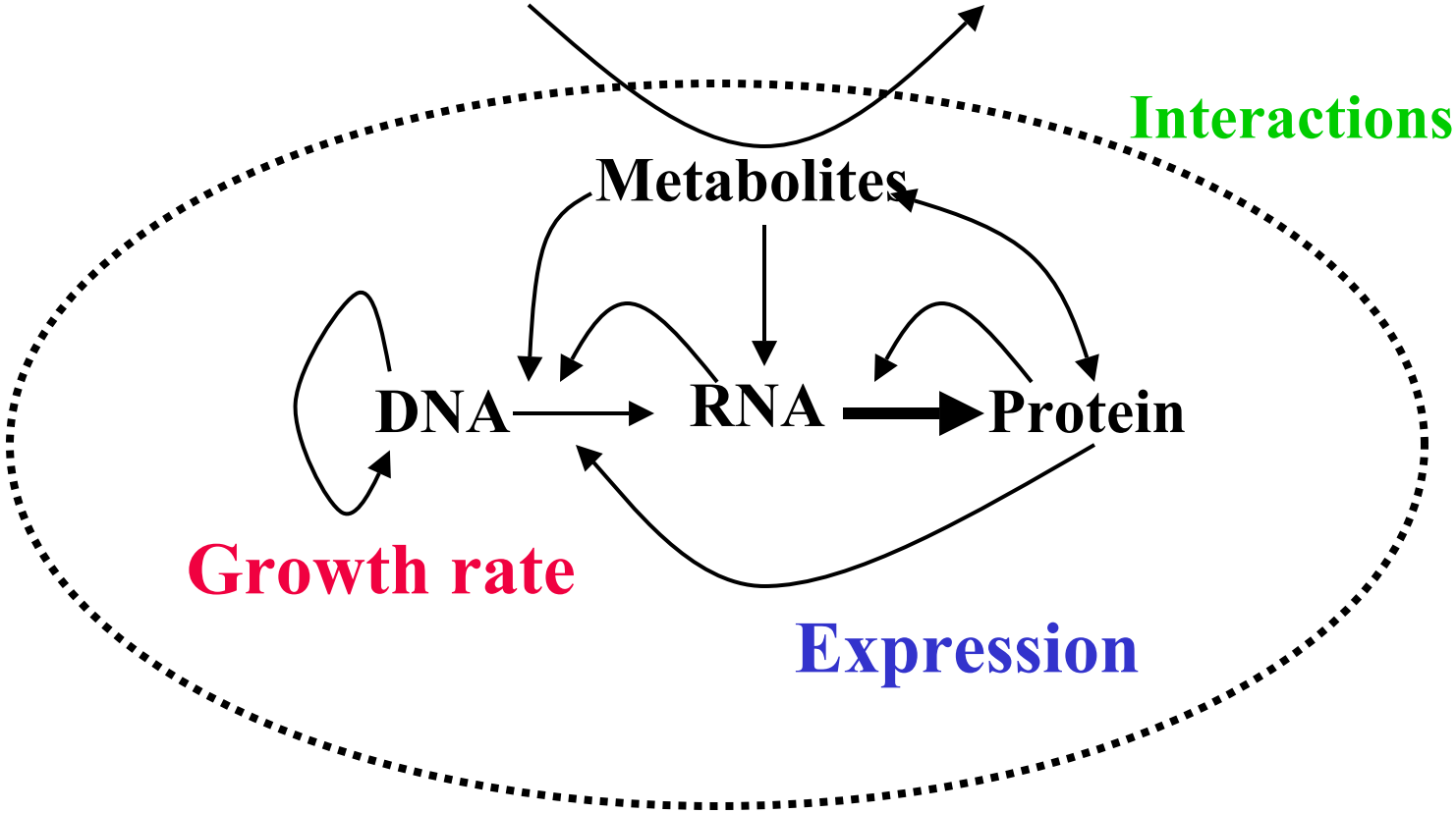


Catalysts increase the rate (&specificity) without being consumed.

Continuity of Life & Central Dogma

Self-assembly, Catalysis, Replication, Mutation, Selection

Regulatory & Metabolic Networks



Polymers: **Initiate, Elongate, Terminate, Fold, Modify, Localize, Degrade**

Intro 2: Today's story, logic & goals

Biological side of Computational Biology

Elements & Purification
Systems Biology & Applications of Models
Life Components & Interconnections
Continuity of Life & Central Dogma
Qualitative Models & Evidence
•Functional Genomics & Quantitative models
Mutations & Selection

Sources of Data for BioSystems Modeling:

Capillary electrophoresis
(DNA Sequencing) :
0.4Mb/day

Chromatography-Mass Spectrometry
(eg. peptide LC-ESI-MS) :
20Mb/day

Microarray scanners (eg. RNA):
300 Mb/day **mpg**

Other microscopy (e.g. subcell, cell, tissue networks)

Structural Genomics

(the challenge of distant homologs)

?↓?

Functional Genomics

(quantitative ligand interactions)

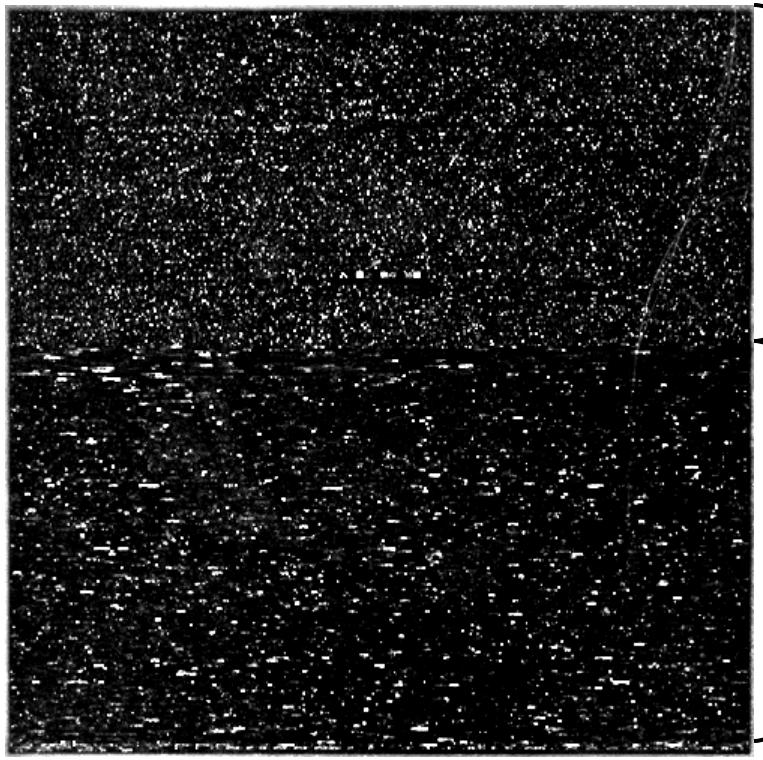
100% Sequence Identity:

1. Enolase Enzyme
2. Major Eye Lens Protein

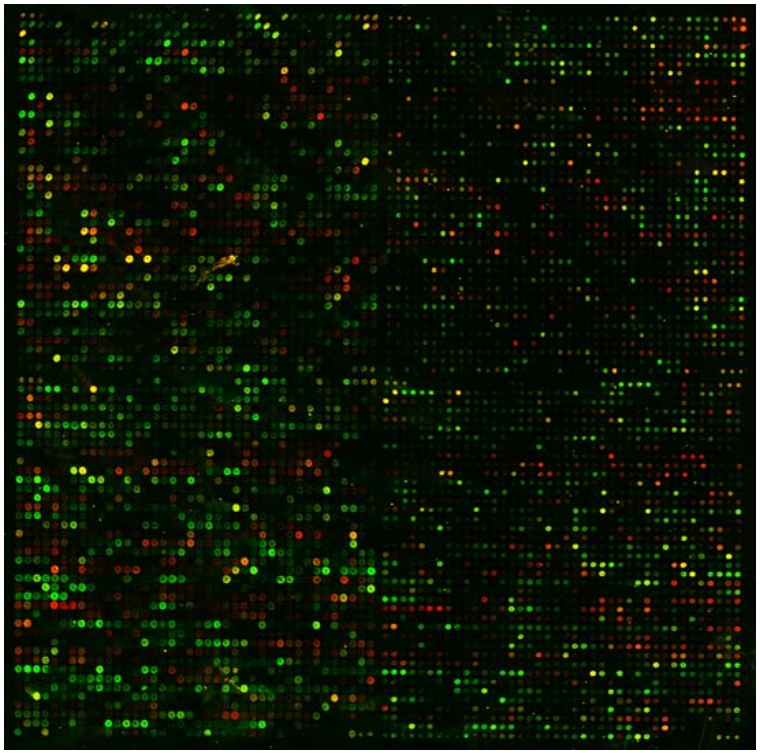
100% Sequence Identity:

1. Thioredoxin Redox
2. DNA Polymerase Processivity

mRNA expression data



Affymetrix *E. coli* oligonucleotide array



Spotted microarray mpg

What is functional genomics?

Function (1): Effects of a mutation on fitness (reproduction) summed over typical environments.
Function (2): Kinetic/structural mechanisms.
Function (3): Utility for engineering relative to a non-reproductive objective function.

Proof : Given the assumptions, the odds are that the hypothesis is wrong less than 5% of the time, keeping in mind (often hidden) multiple hypotheses.

Is the hypothesis suggested by one large dataset already answered in another dataset?

Genomics Attitude

Whole systems:
Less individual gene- or hypothesis-driven experiments;
Automation from cells to data to **model** as a proof of protocol.

Quality of data: DNA sequencing raw error: 0.01% to 10%.
Consensus of 5 to 10 error: 0.01% (1e-4)

Completion: No holes, i.e. regions with data of quality less than a goal (typically set by cost or needs of subsequent projects).

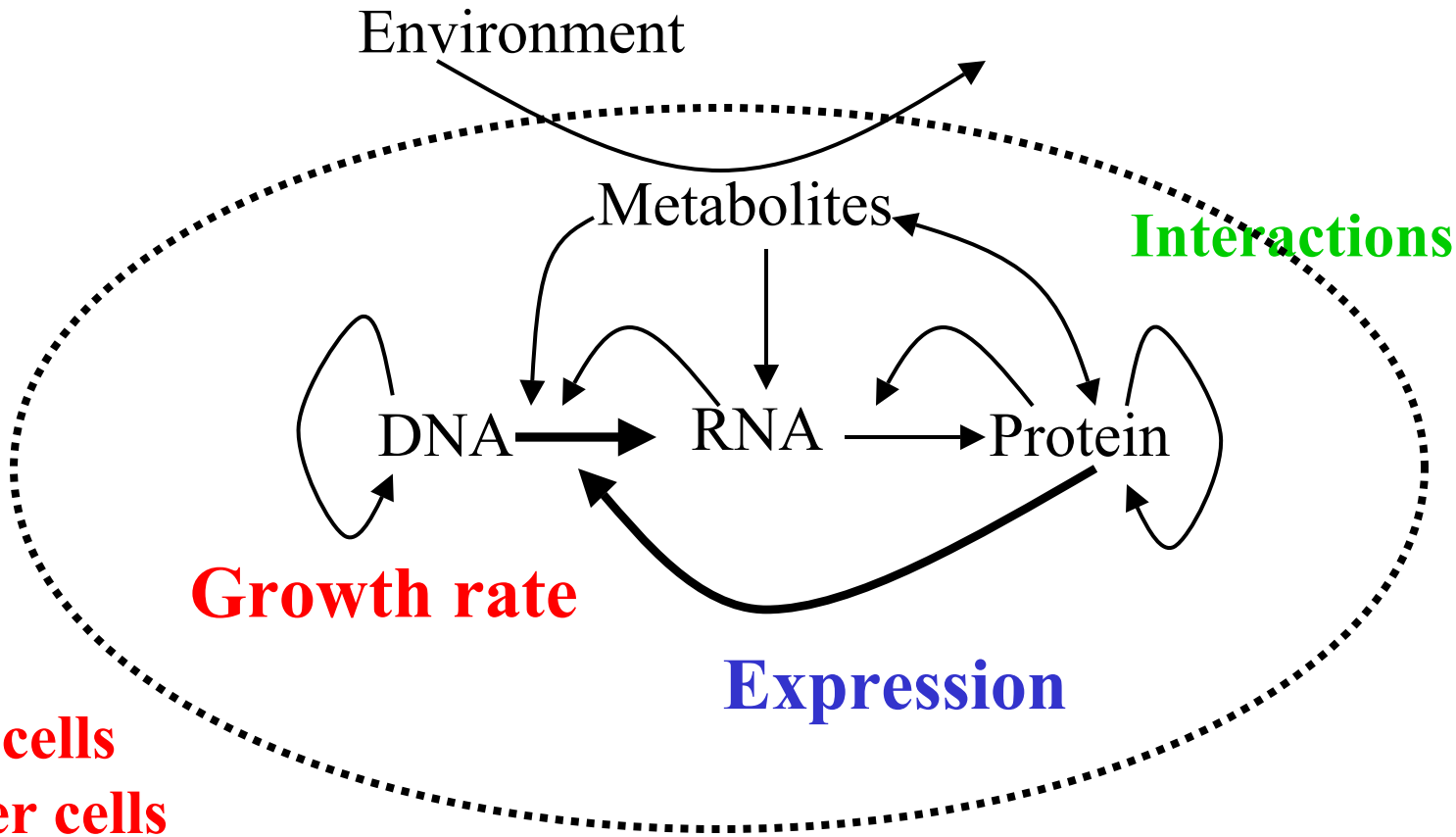
Impossible: The cost is higher than reasonable for a given a time-frame and quality assuming no technology breakthroughs.
Cost of computing vs. experimental "wet-computers".

Intro 2: Today's story, logic & goals

Biological side of Computational Biology

Elements & Purification
Systems Biology & Applications of Models
Life Components & Interconnections
Continuity of Life & Central Dogma
Qualitative Models & Evidence
Functional Genomics & Quantitative models
•Mutations & Selection

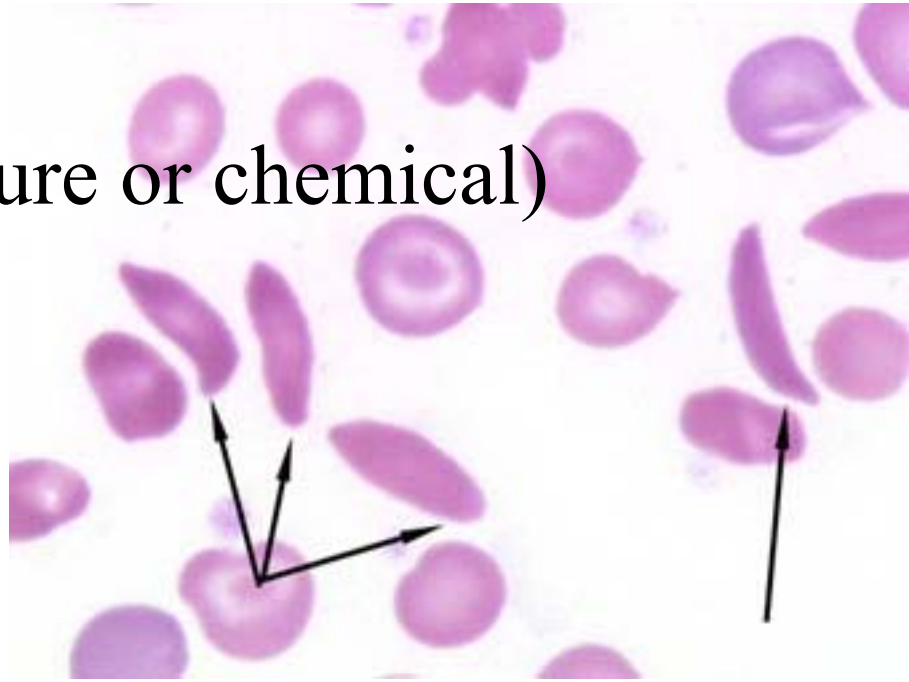
Mutations and selection



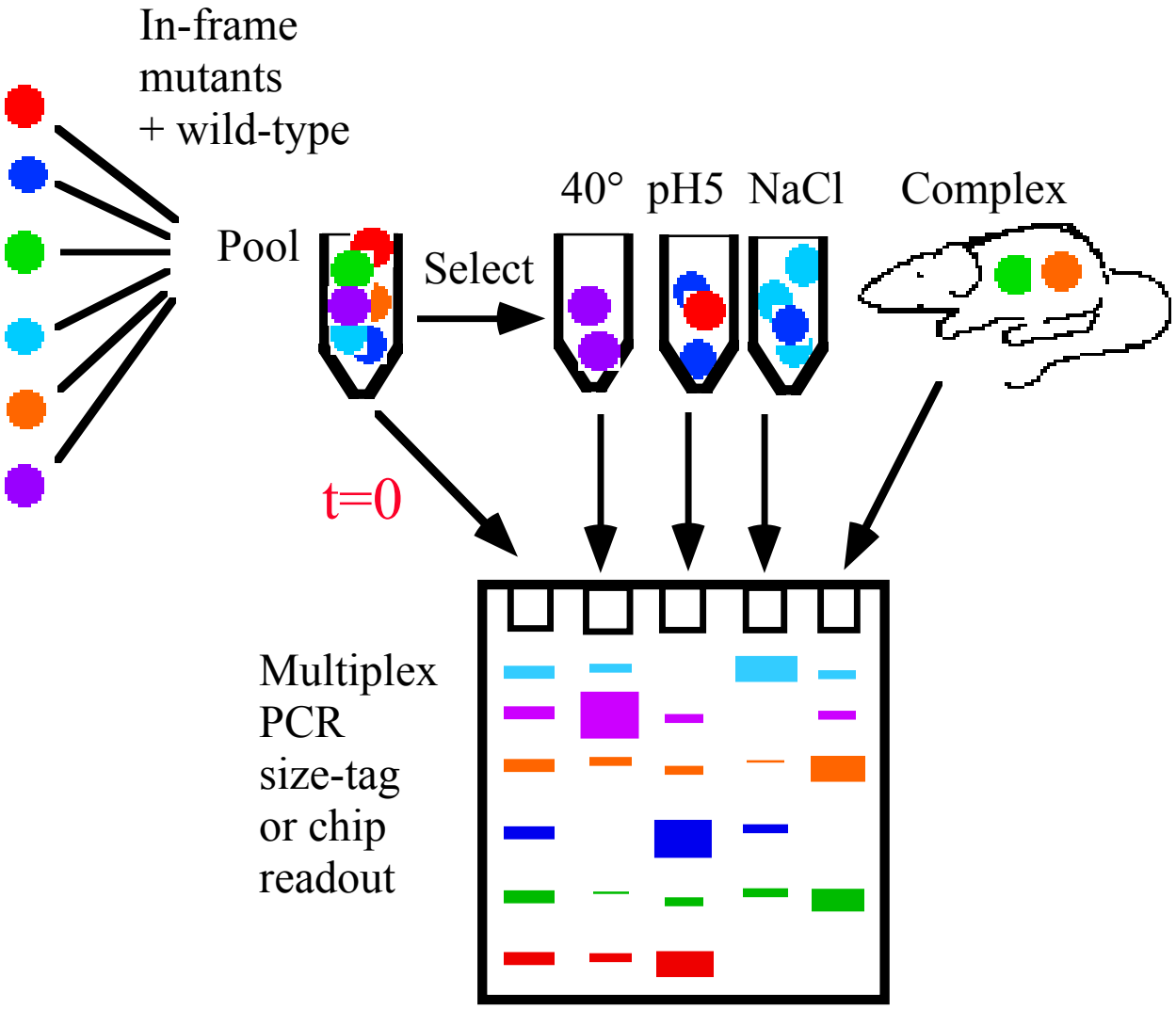
stem cells
cancer cells
viruses
organisms

Types of Mutants

Null: PKU
Dosage: Trisomy 21
Conditional (e.g. temperature or chemical)
Gain of function: HbS
Altered ligand specificity



Multiplex Competitive Growth Experiments

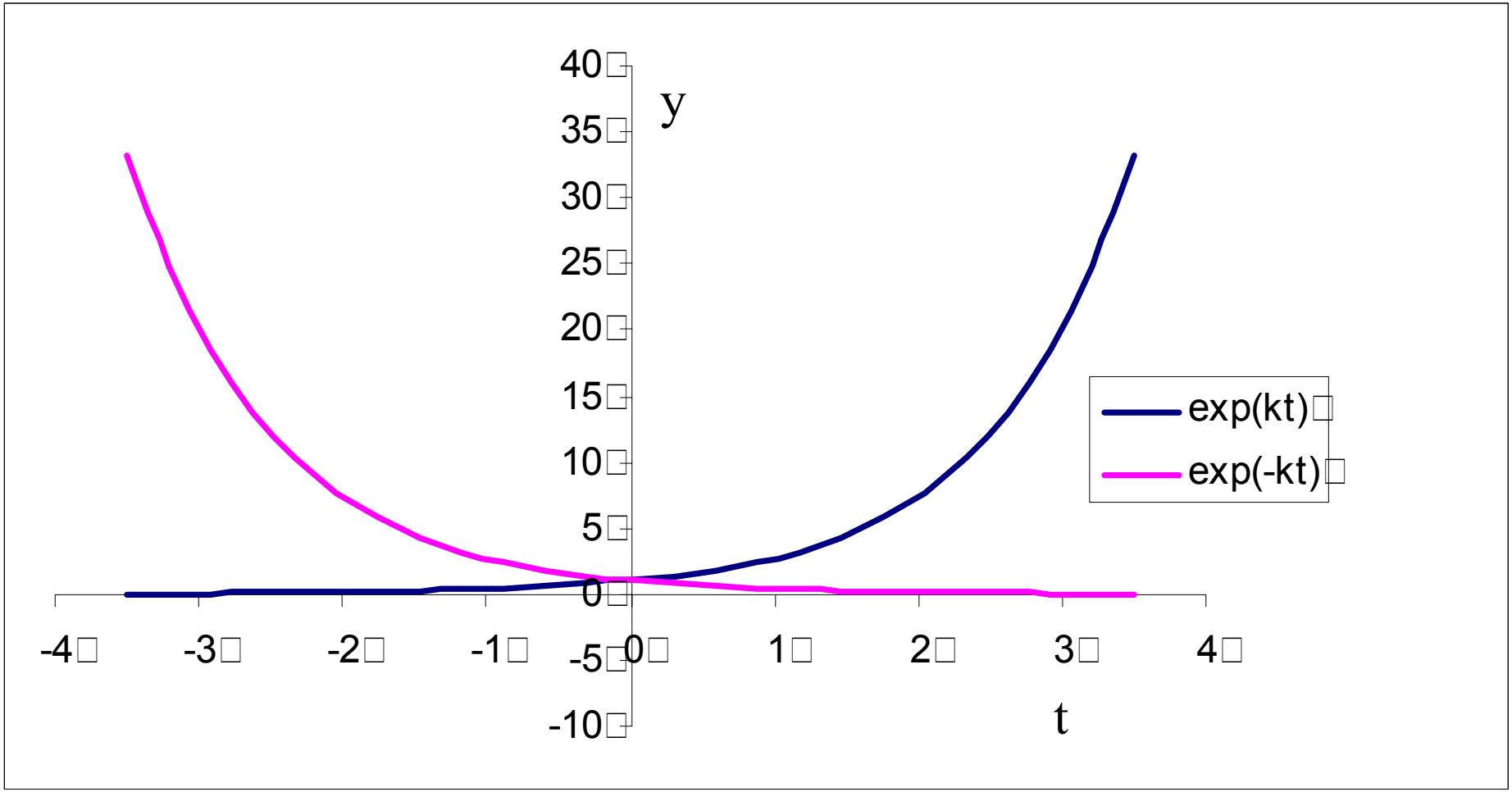


Growth & decay

dy/dt = ky

$y = Ae^{kt}$; e = 2.71828...

k=rate constant; half-life=$\log_e(2)/k$



Ratio of strains over environments, **e** , times, $\mathbf{t_e}$, selection coefficients, $\mathbf{s_e}$,

$R = R_o \exp[-\Sigma s_e t_e]$

80% of 34 random yeast insertions have s<-0.3% or s>0.3%
t=160 generations, e=1 (rich media); ~50% for t=15, e=7.
Should allow comparisons with population allele models.

Multiplex competitive growth experiments:

Thatcher, et al. (1998) PNAS 95:253.
Link AJ (1994) thesis; (1997) J Bacteriol 179:6228.
Smith V, et al. (1995) PNAS 92:6479.
Shoemaker D, et al. (1996) Nat Genet 14:450.

Intro 2: Today's story, logic & goals

Biological side of Computational Biology

Elements & Purification
Systems Biology & Applications of Models
Life Components & Interconnections
Continuity of Life & Central Dogma
Qualitative Models & Evidence
Functional Genomics & Quantitative models
Mutations & Selection