(re)constructing life: we must not forget the chassis

antoine danchin 唐善・安東

amabiotics sas / amabiotics intl ltd hk

international lectures series, universität bremen

bremen, october 24th, 2012

synthetic biology in context

synthetic biology beyond the hype

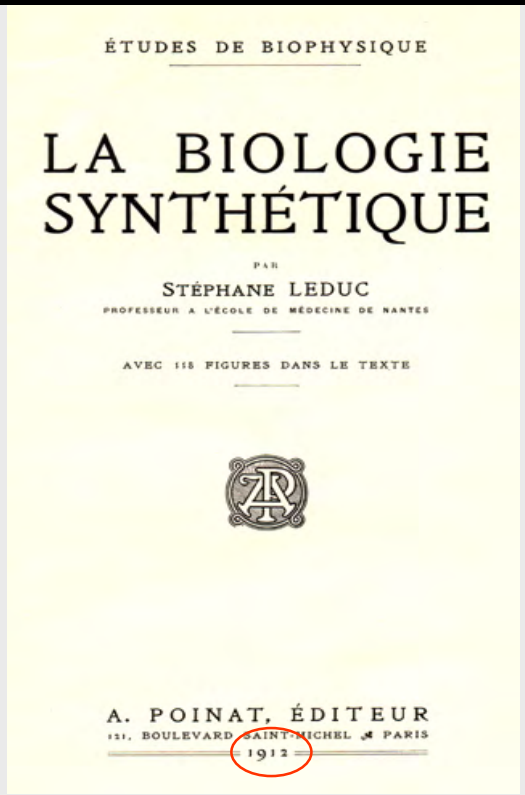
- **reconstructing and understanding**: forgetting the "black box" sb reconstructs life to explore whether we understand what life is and **uncover missing entities using engineering principles**

- **abstracting**: sb keeps the laws defining life, and applies them using objects of a different physico-chemical nature (xenobiology)

- **engineering**: sb designs and standardises « biobricks » to construct programs using a « **chassis** » with man's interests' goals

- **evolving**: sb combines design and evolution to use (poorly understood) principles that drive adaptation; **there is an in-built principle meant to trap information in living organisms**

overlooked history


ÉTUDES DE BIOPHYSIQUE

LA BIOLOGIE SYNTHÉTIQUE

PAR

STÉPHANE LEDUC

PROFESSEUR A L'ÉCOLE DE MÉDECINE DE NANTES

AVEC 118 FIGURES DANS LE TEXTE

A. POINAT, ÉDITEUR

121, BOULEVARD SAINT-MICHEL & PARIS

1912

stéphane leduc
1853 - 1939

Reassembly of Living Cells from Dissociated Components

kw jeon, ij lorch, jf danielli science **1970** 167: 1627-8

Abstract. *Combining the techniques of nuclear transplantation and cytoplasmic transfer, dissociated amoeba nuclei, cytoplasm, and membranes were reassembled to form viable amoebae. The techniques of cell reassembly appear to be sufficiently adequate so that any desired combination of cytoplasm, nucleus, and membrane can be assembled into living cells.*

IV.

Artificial Synthesis of New Life Forms

bulletin of atomic scientists
december **1972** 28: 20-24

JAMES F. DANIELLI
1911-1984

The age of synthesis is in its infancy, but is clearly discernable. In the last decade (1960-70), we have seen the first syntheses of a protein, a gene, a virus, a cell, and of allophenic mice. Nothing with such dramatic implications has ever been seen in biology before. Previously, plant and animal breeders have been able to create what are virtually new species, and have been able to do so at a rate which is of the order of 10^4 times that of average evolutionary processes. A further increase in rate is now on the horizon. We need a few additional "firsts" before this will occur: (1) to be able to synthesize a chromosome from genes and other appropriate macromolecules; (2) to be able to insert a chromosome into a cell; or, alternatively to (1) and (2), to be able (3) to insert genes into a cell in some other way; (4) we must also learn how to bring the set of genes, which is introduced into a cell, within the domain of cellular control mechanisms, so that they do not run wild in the cell. None of these problems appear to be of exceptional difficulty.

a standing enigma: babies are born very young!

but ageing is sometimes positive, and this is foreign to standard engineering knowledge

contrary to intuition, mixing a population of young bacteria with an old culture, the old one outgrows the young one ("growth advantage in stationary phase": gasp phenotype)

is this compatible with synthetic biology? with scaling up?

which processes underlie this phenomenon?

which genes allow information to accumulate?

cells as computers making computers

life requires:

- a program (a "book of recipes": **replicated**)

recursive information transfer and trapping

=> coding from one level to a second level introduces an essential asymmetry (conceptually **different from feedback**, feedforward**)**

- a machine ("chassis") allowing the program to be expressed (**reproduced**) and defining an inside and an outside

- a dynamic coupling process: metabolism (chemical interchange)

synthetic life asks that one places the program within a chassis

program and chassis

sb usually aims at creating novel programs, assuming that previously characterized chassis, preferably from gras organisms, will yield expected results

escherichia coli
saccharomyces cerevisiae
pseudomonas putida

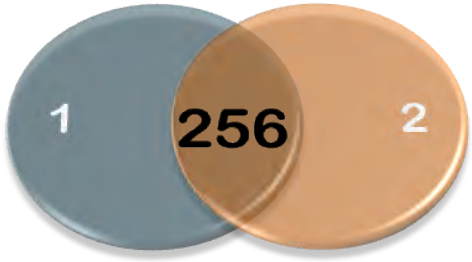
are preferred candidates

and mycoplasmas for proof-of-concept

quest for the minimal genome

genomes are not rosetta stones…

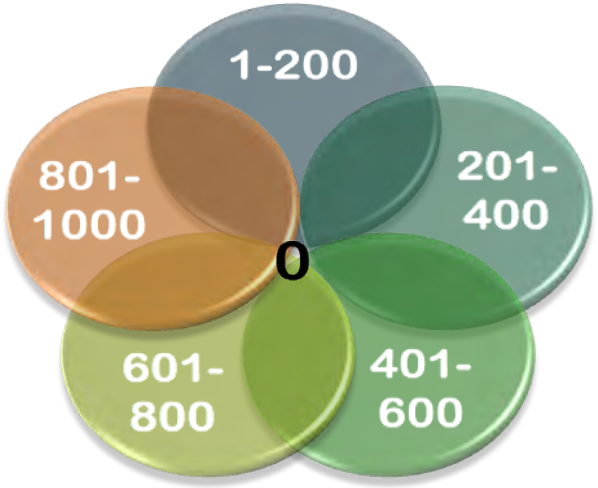
m. genitalium *h. influenzae*



Mushegian AR and Koonin EV. PNAS 1996; 93:10268-73

Koonin EV. Nat. Rev. Microbiol. 2003;1:127-36

Ciccarelli FD, et al. Science. 2006;311:1283-7



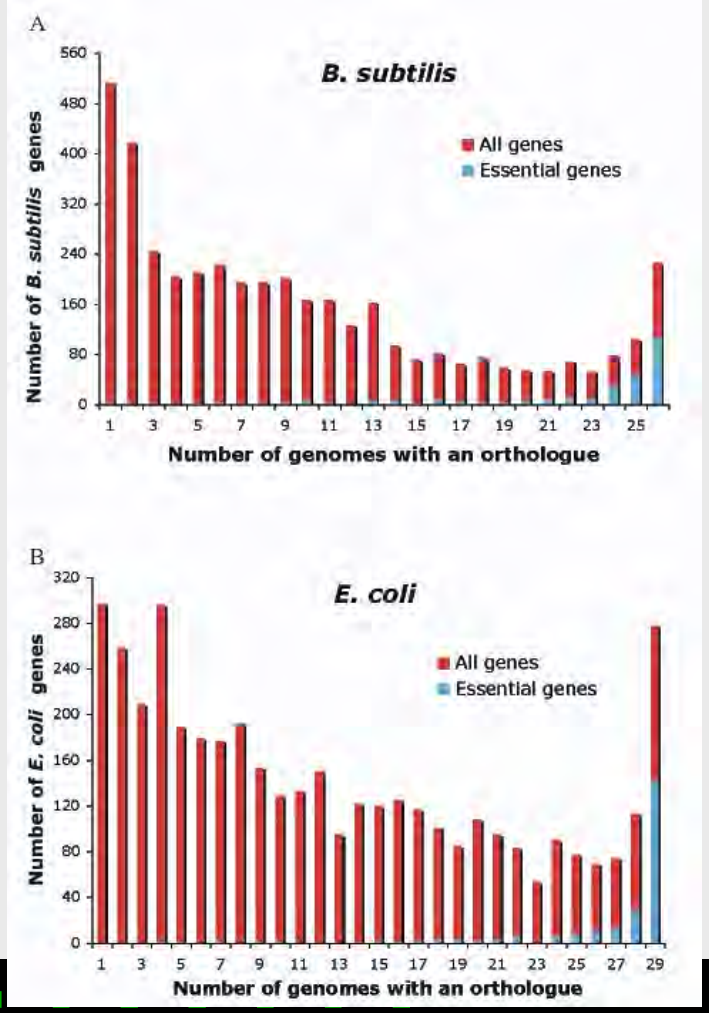
Lagesen K et al. Microbiology. 2010;156:603-8

p e r s i s t e n t g e n e s

essential genes are located in the leading strand; they are conserved in a majority of genomes; by contrast the genes that are conserved and located in the leading strand make a particular category, which doubles the number of « essential » genes

these genes make a universal category; 400-500 genes persist in a majority of bacterial genomes; they are not only involved in the three processes needed for life, but in maintenance and in adaptation to transient phenomena; a fraction manages the evolution of the organism

gene persistence

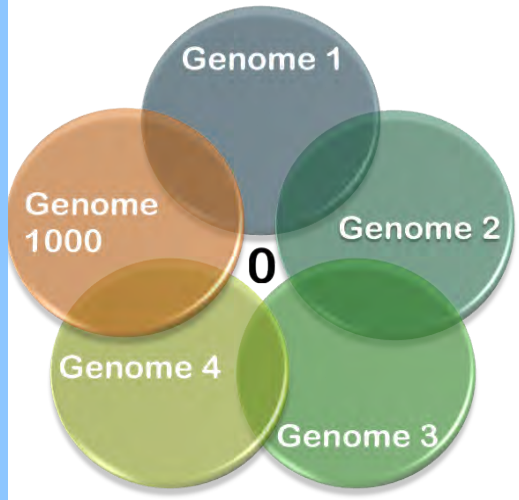


which functional category?

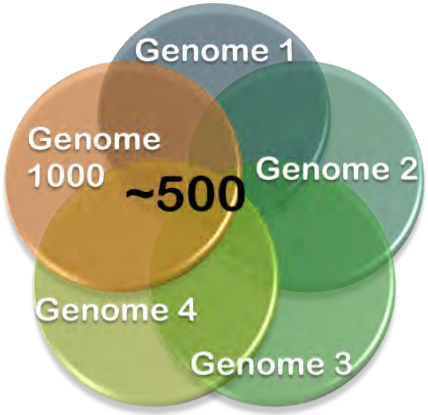
- information transfer
- compartmentalization
- intermediary metabolism
- stress, maintenance and repair

highly non random!

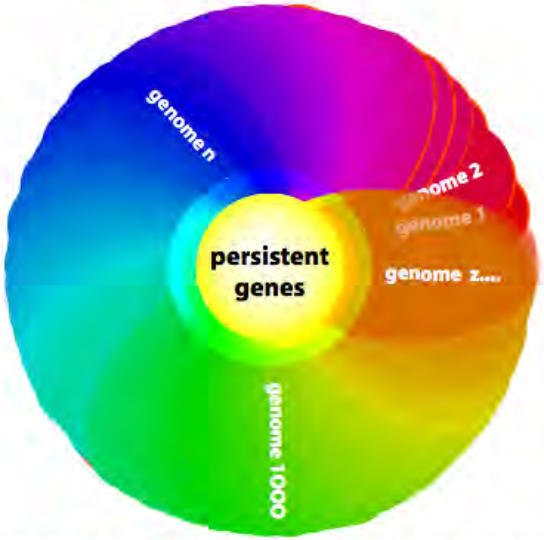
… genes persist in a quorum of genomes



Conserved orthologs



Persistent genes



genomes overlap; as more genomes are compared progressively less orthologs are shared until their number falls to zero

persistent genes are orthologs that belong to a quorum of genomes, above a threshold computed using a measure that retains frequent genes that tend to cluster together

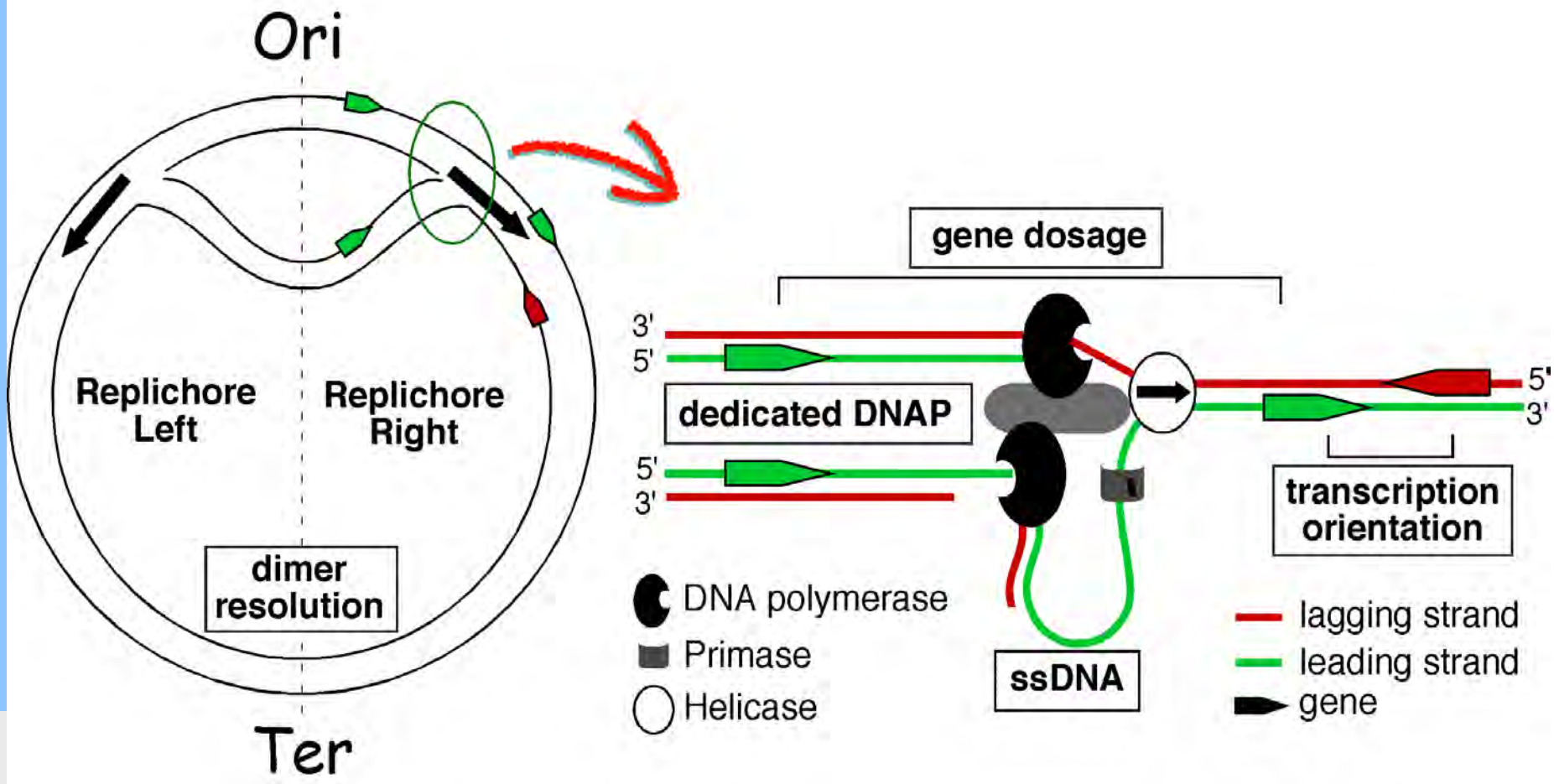
functional analysis bottom - up

comparative genomics expects that genes shared by multiple genomes (persistent genes) are likely to be essential

they share common features, in particular they are expressed from the leading dna strand, reducing transcription / replication conflicts

this constraint is important for engineering

transcription / replication conflicts



the program has a material support!

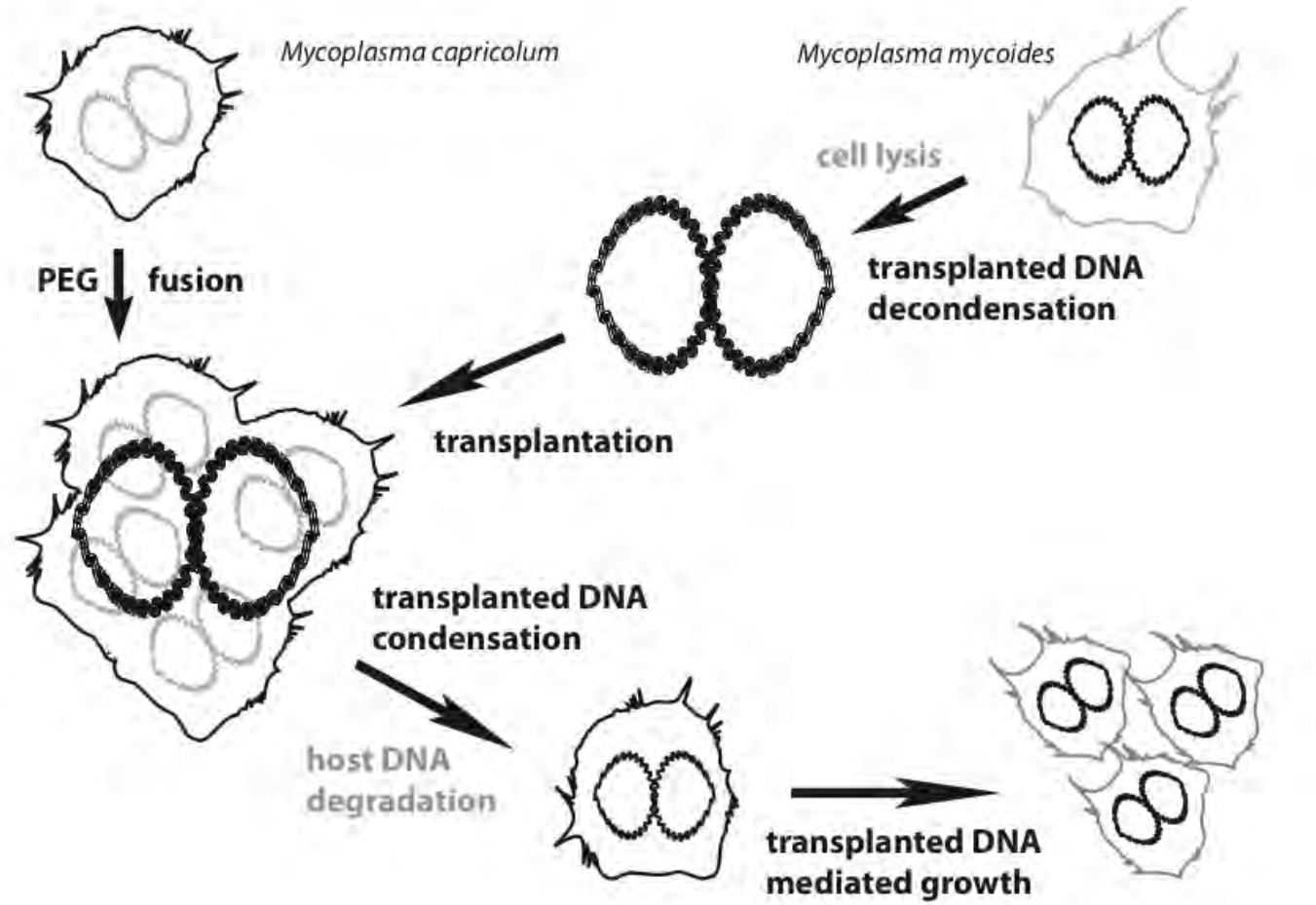
it is not enough to have a dna molecule with the right sequence, it needs to be correctly folded

d n a t r a n s p l a n t a t i o n

upon lysis dna is prone to expand as unavoidable nicks cut strands randomly

it cannot enter a single host cell

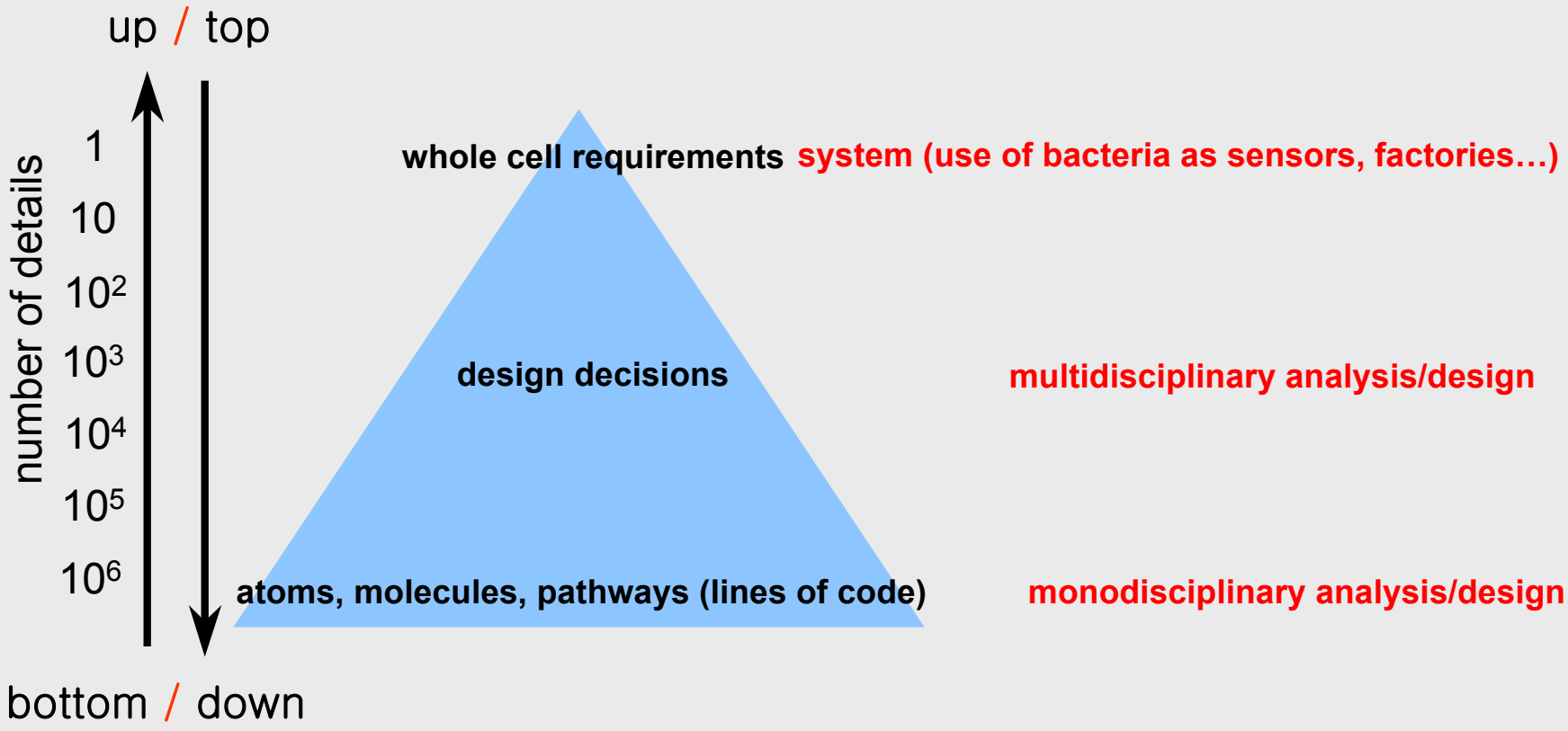
PEG makes a macro cell that can accomodate it



lartigue c, glass ji, alperovich n, pieper r, parmar pp, hutchison ca, 3rd, smith ho, venter jc: genome transplantation in bacteria: changing one species to another. *science* 2007, **317**(5838):632-638.

the functional genome

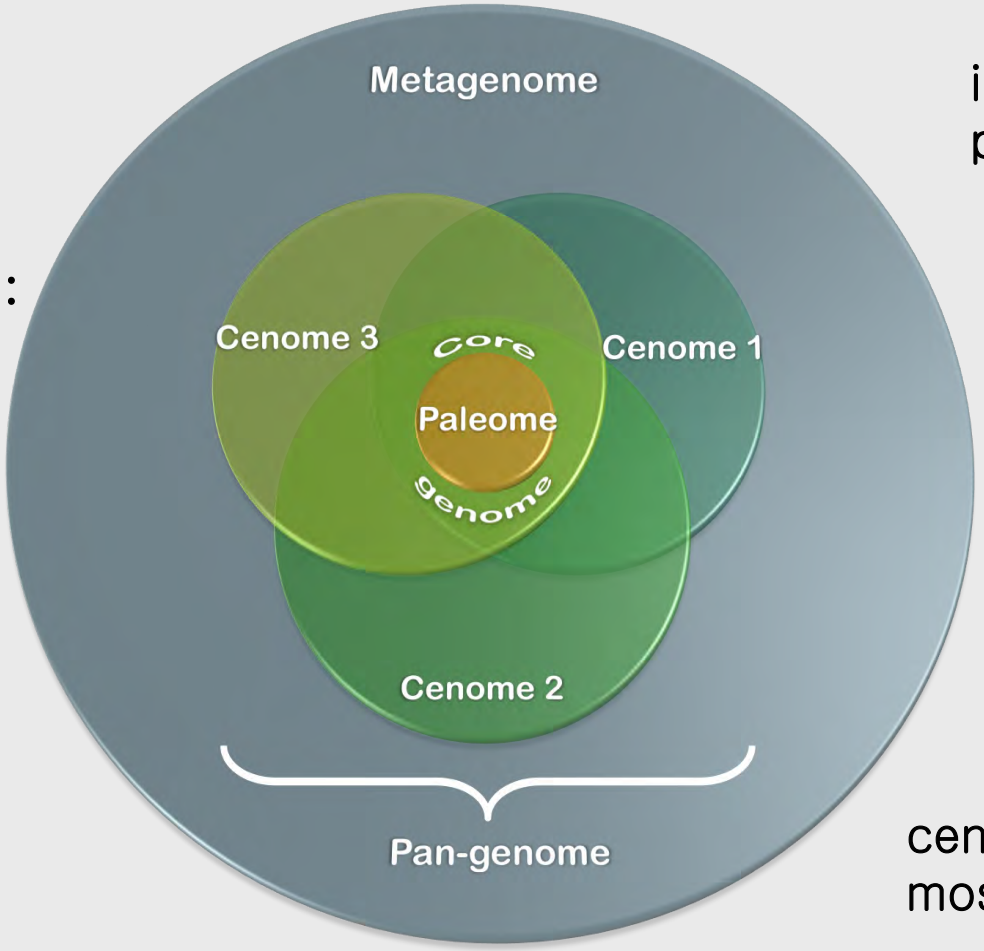
functional analysis



two histories; two functions

pan genome:

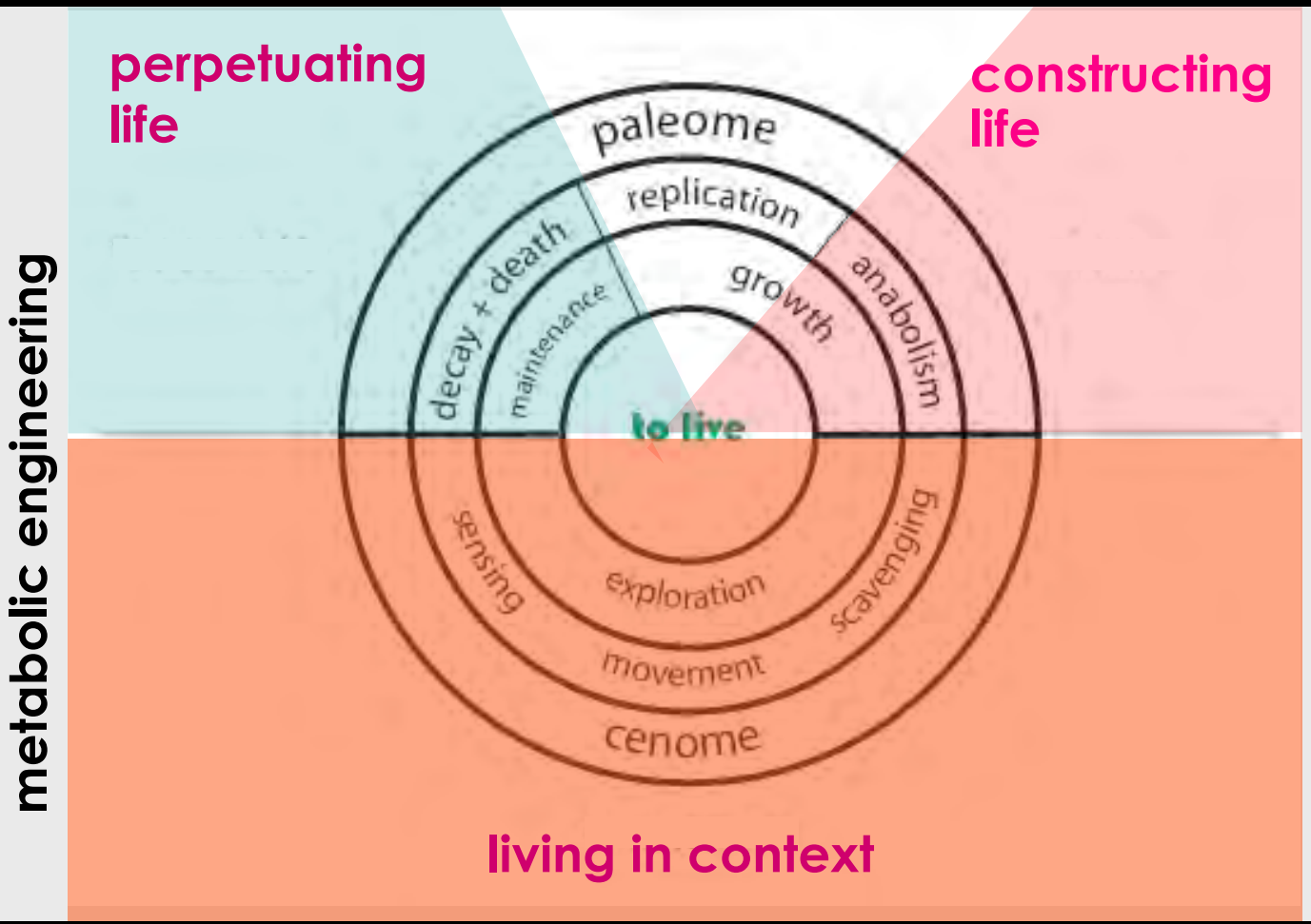
paleome
+
cenome



in e. coli
paleome: 1,900 genes

cenome > 40,000 genes
mostly hgt

a tale of two genomes



the genome is functionally organized

some genes tend to stay close to one another:

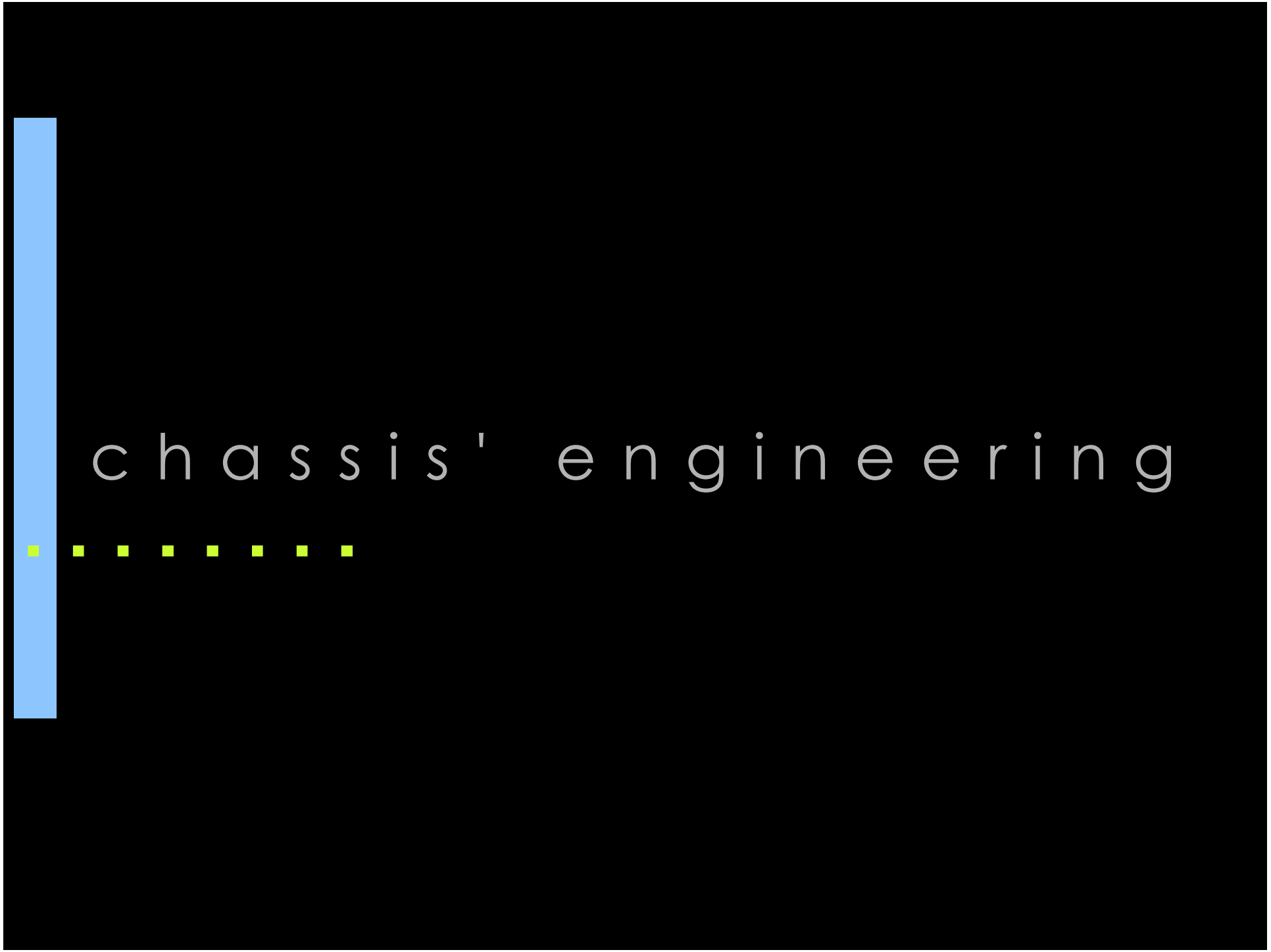
- persistent genes (present in a large quorum of genomes: **no** ubiquitous genes)
- rare genes (present in specific strains of a given species)

the latter are easily accounted for, as they come from horizontal gene transfer; what about the former?

a short (partial) list follows

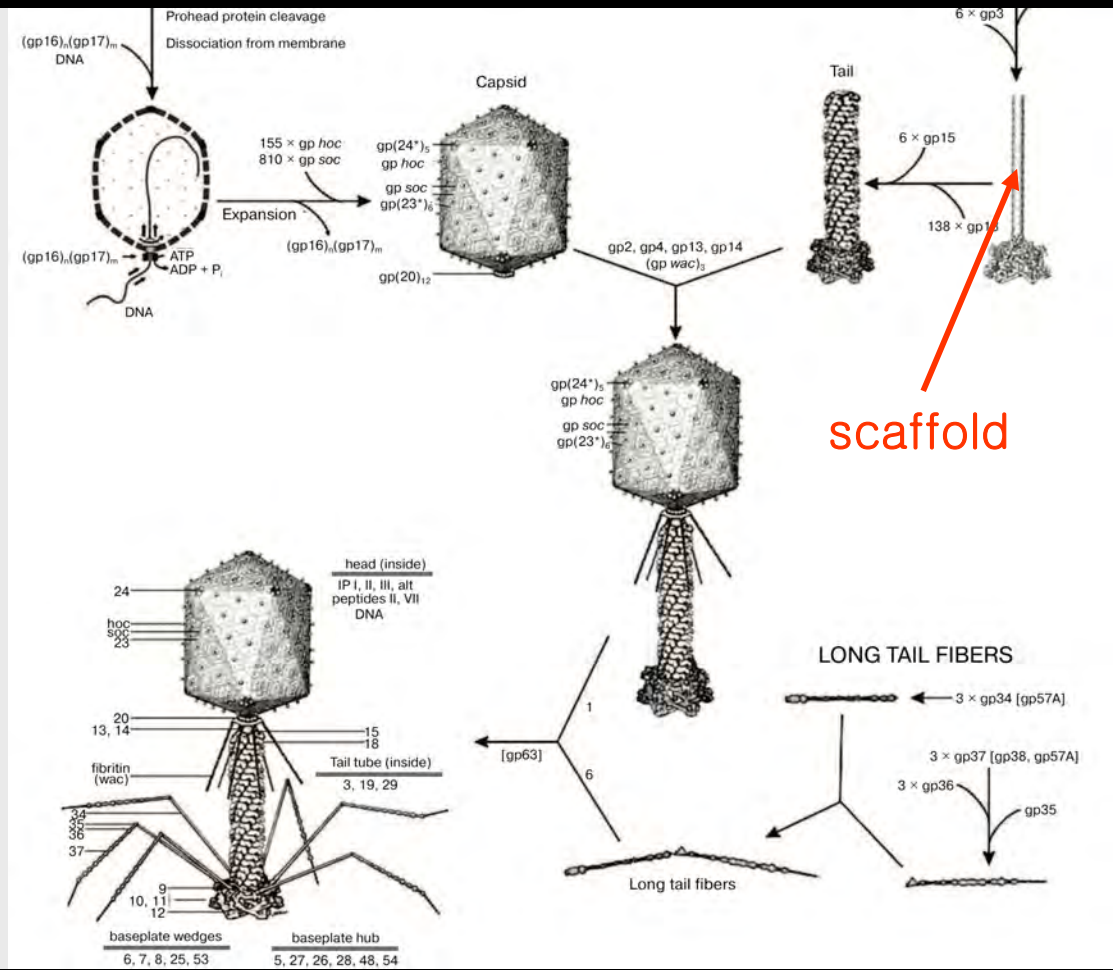
the minimal functions 1989

process	structure	length
replication	dna wielding	40 kb
transcription	transcription + coupling with translation	30 kb
translation	ribosome: ribosomal rna + 50-60 ribosomal proteins	60 kb
	trnas + trna loading + polypeptide synthesis	80 kb
core metabolism	building blocks and coenzymes	200 kb
transport	import and export	
energy management	atp synthesis and electron transfers	
specific casings	creation of an envelope	100 kb

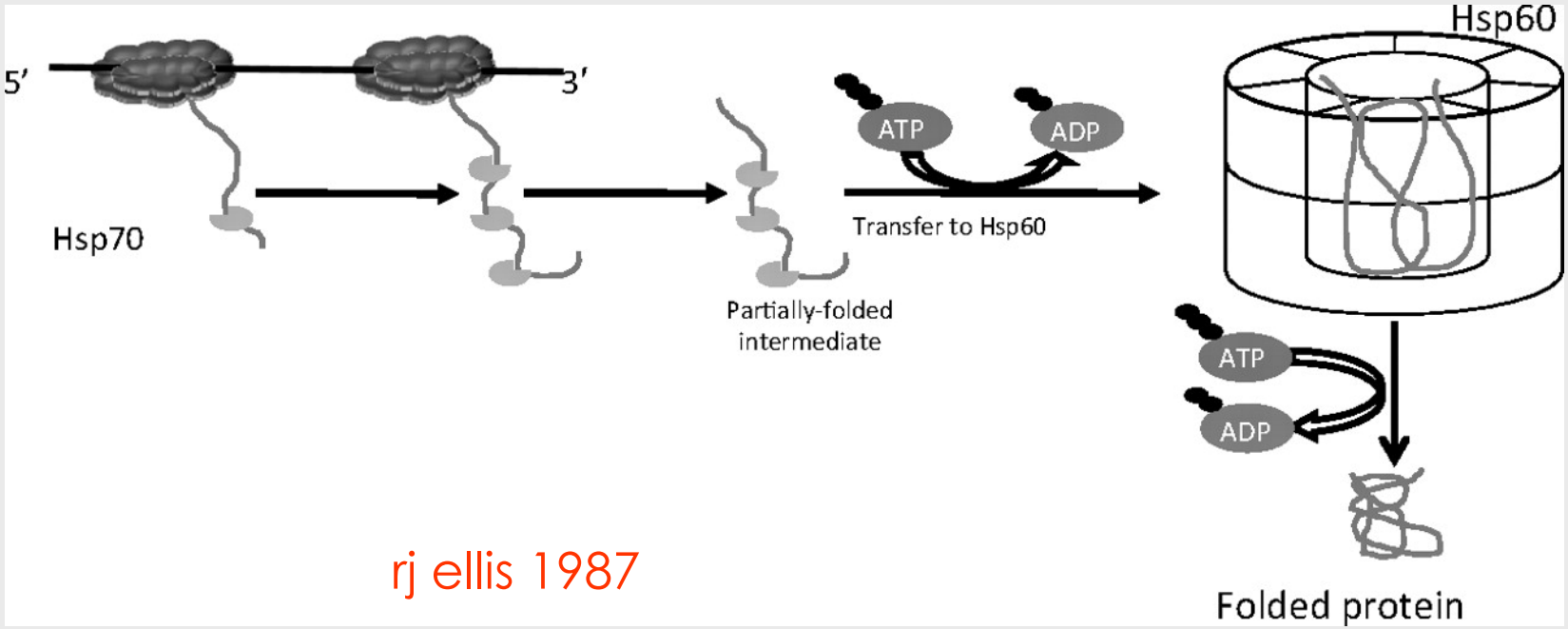

chassis' engineering

s c a f f o l d s

to build up a bacteriophage such as phage t4, a scaffold is constructed and used as a vernier to make a tail of fixed length, with the proteins of the tail making an helix structure around the scaffold that is later disposed of



molecular chaperones



rj ellis 1987

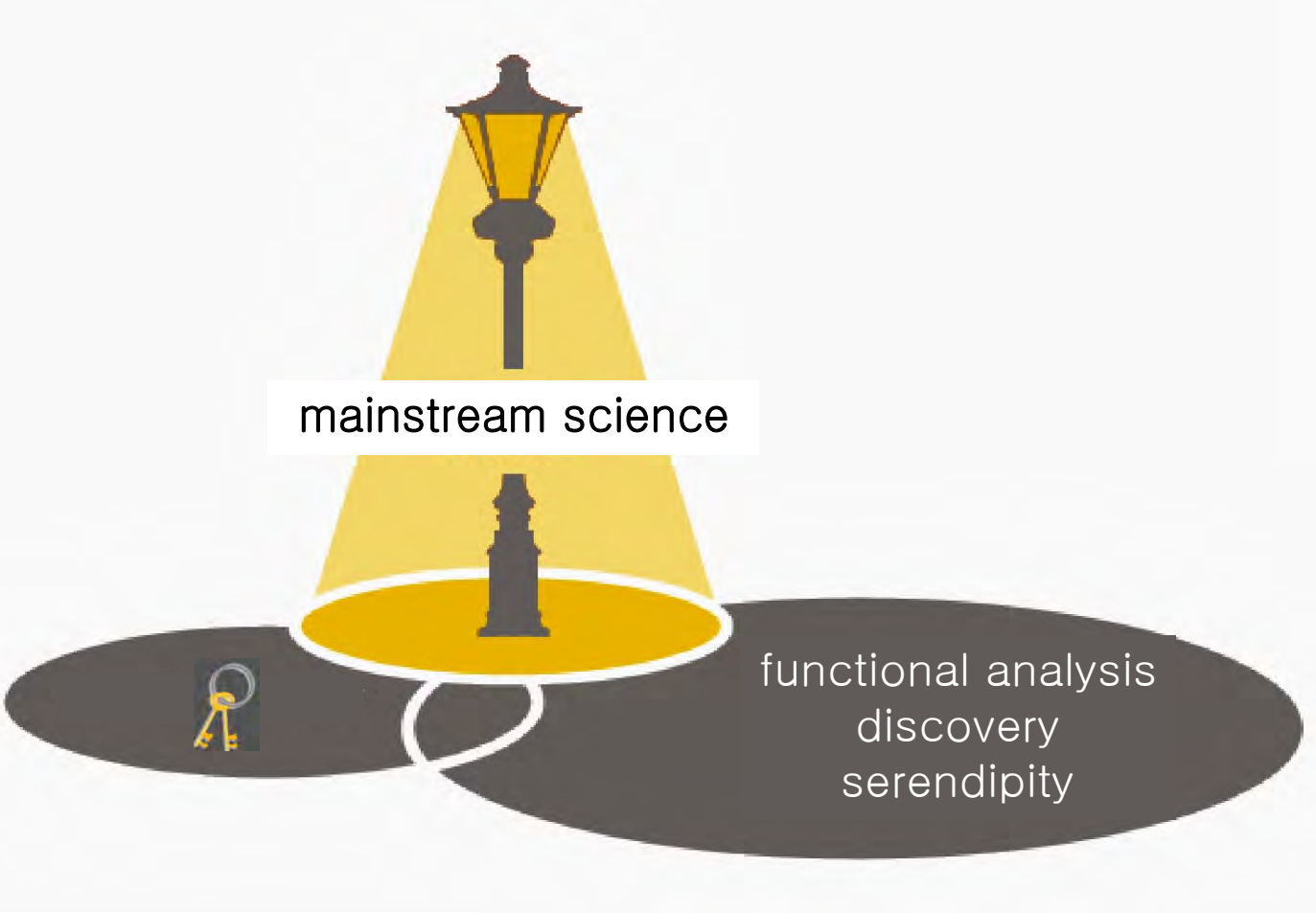
JLB JOURNAL OF LEUKOCYTE BIOLOGY

managing waste

shedder

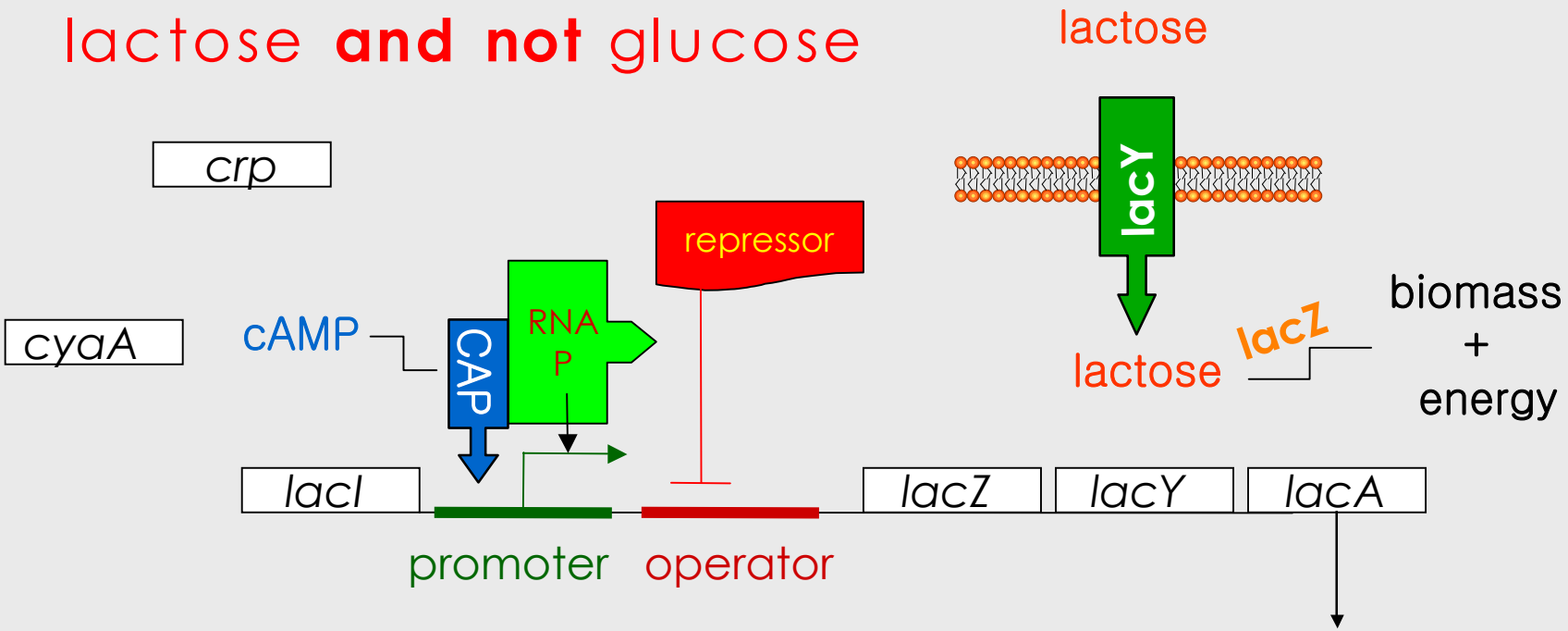
garbage chute

the lamppost effect



a chassis' engineering need

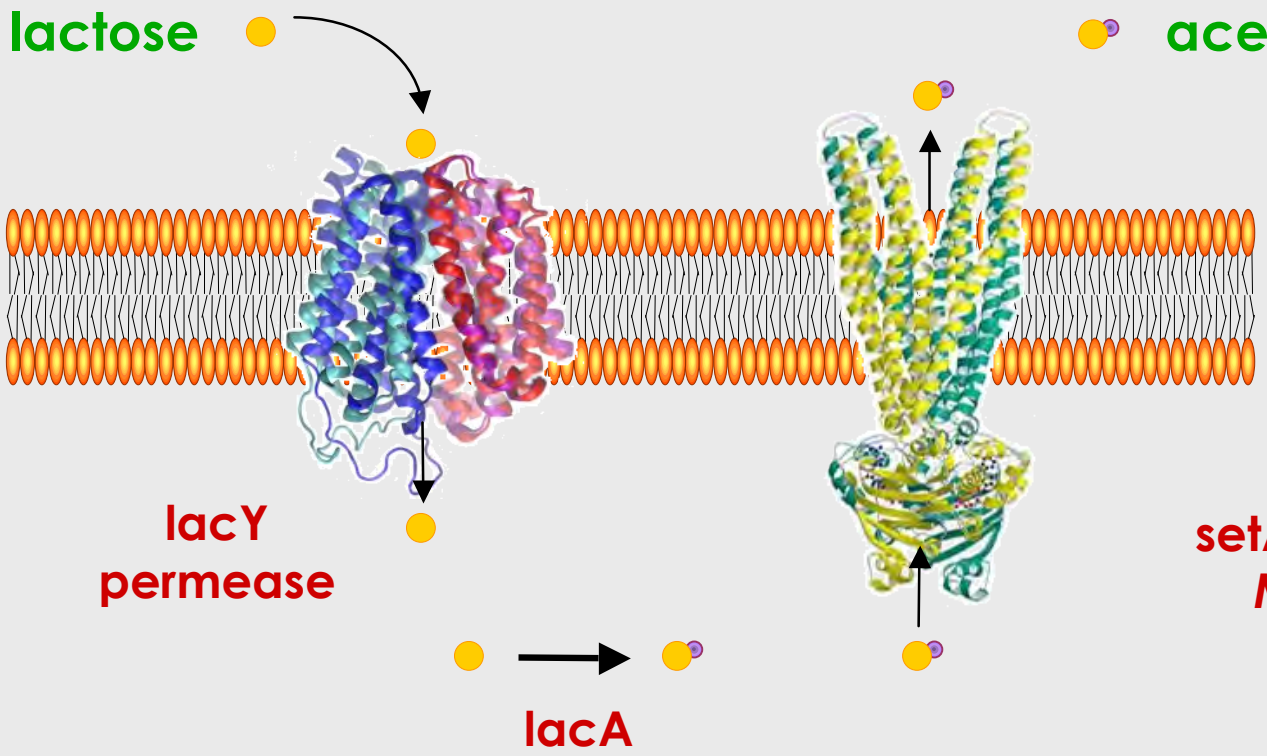
lactose **and not** glucose



what is the function of lacA?

why did we need 50 years to ask the question?

cells need safety valves, not leaks



lactose acetyl-transferase (very inefficient K_M: 1 M)

leftovers in sulfur metabolism

escherichia coli cysq mutants require sulfite or cysteine for growth;

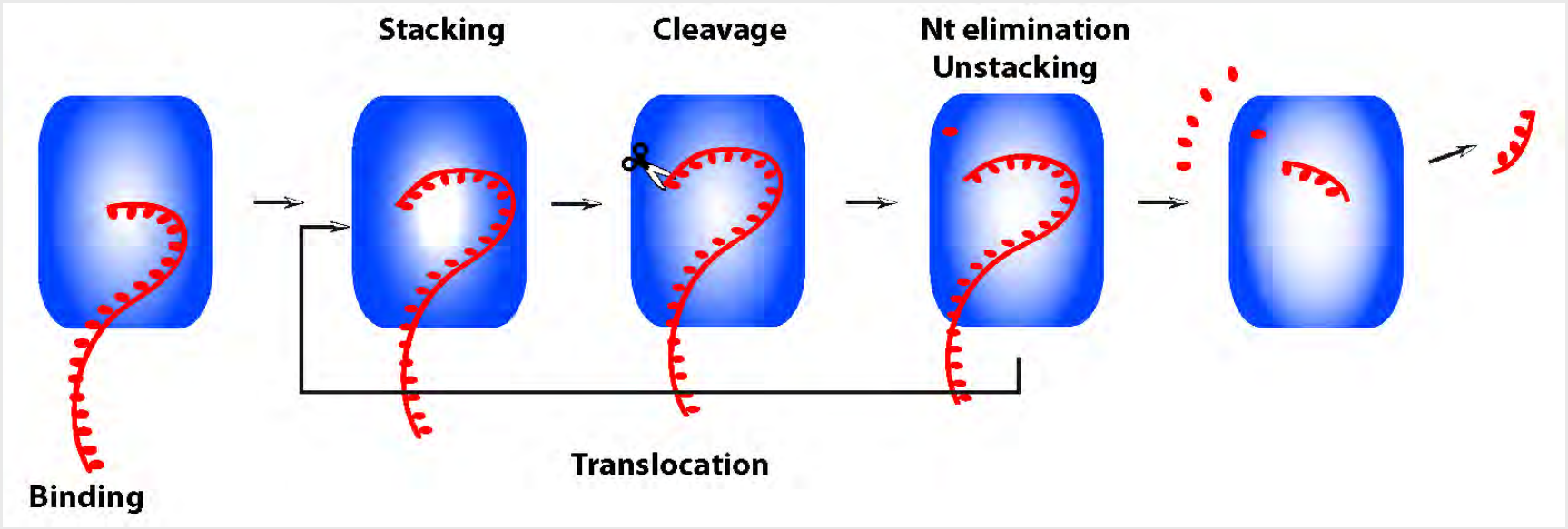
purification of pap binding proteins identified *cysq* and also protein orn that hydrolyzes very small rna molecules (nanornas);

complementation of an e. *coli orn* defect by libraries from a variety of bacterial genomes revealed proteins from several structural descents, some of which also hydrolyzing pap, i.e. playing the role that *cysq* plays

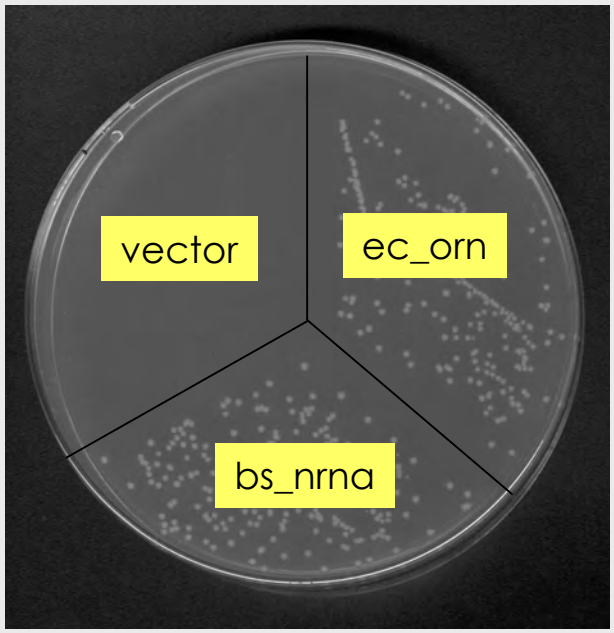
organisms such as *mycobacterium tuberculosis* have both *orn* and *nrna*, and also *cysq*

trashing is a required function

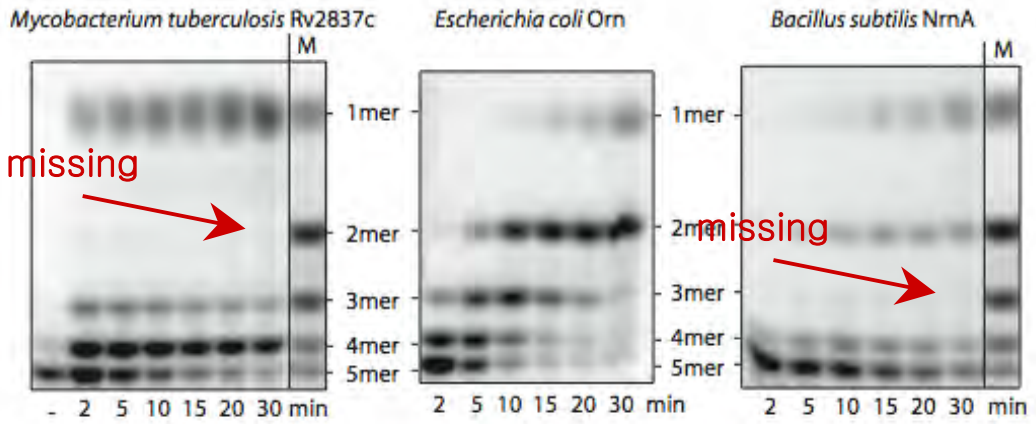
nanornase is an essential function

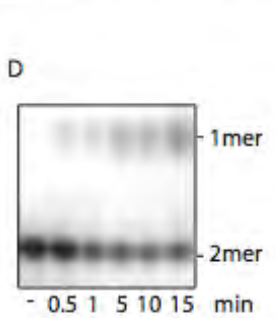


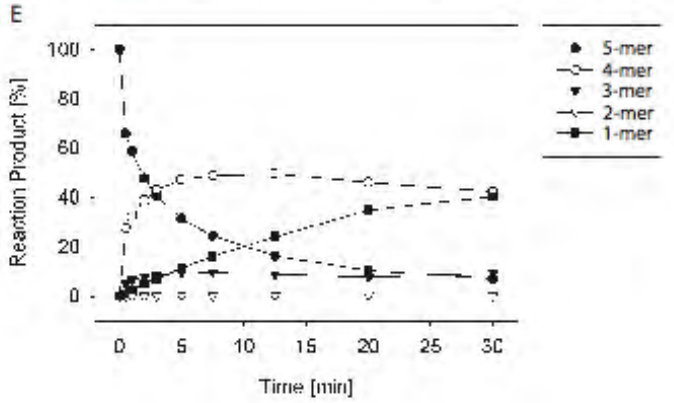
nano-rnases: functional, not structural ubiquity



complementation
in vivo





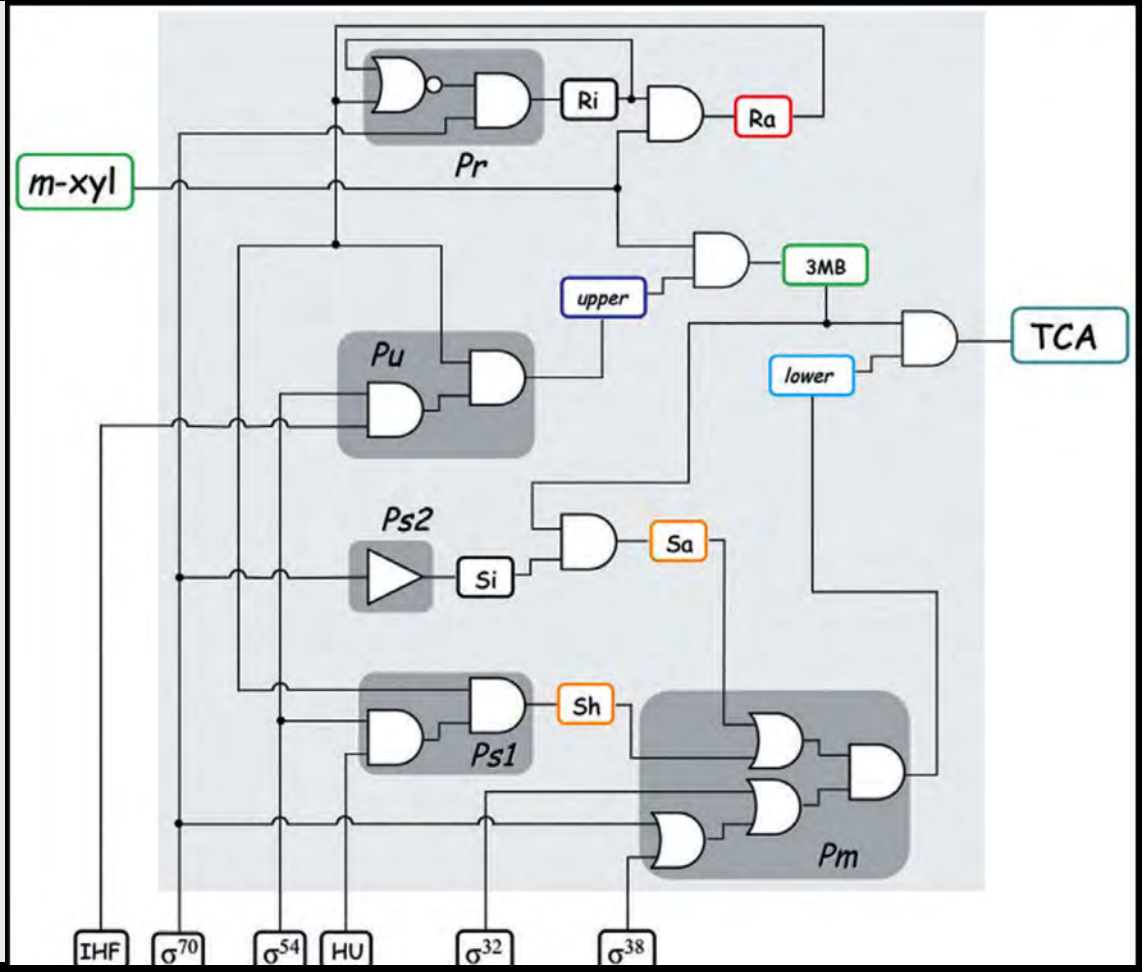



antifragility

the logicome

this is a non-linear behaviour

victor de lorenzo



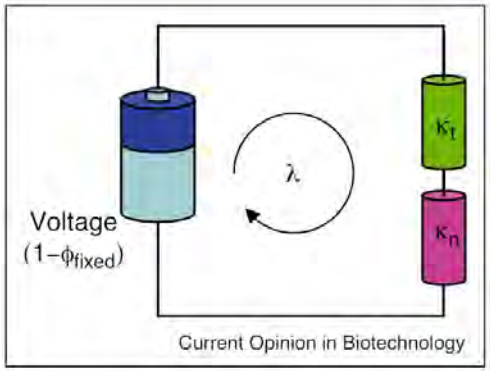
p h y s i o l o g y

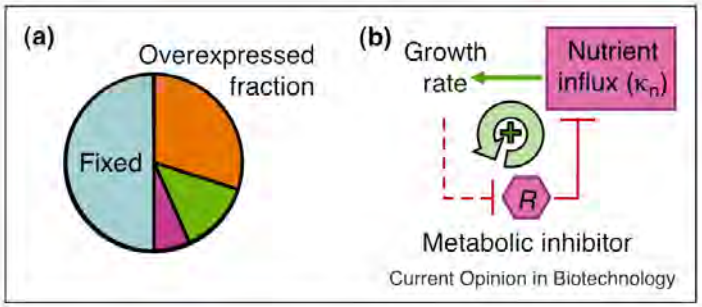
the cell expresses housekeeping genes (fixed), the translation machinery (mainly ribosomes, variable) and genes specific to the environment;

the growth rate is directly determined by the nutrient influx

this is a linear behaviour

terry hwa





t h e f l y w h e e l

a coupling device—the flywheel—is essential to smoothly link the non-linear behaviour of the engine with the linear movement of the overall machine
small untranslated rnas play a major role in this process

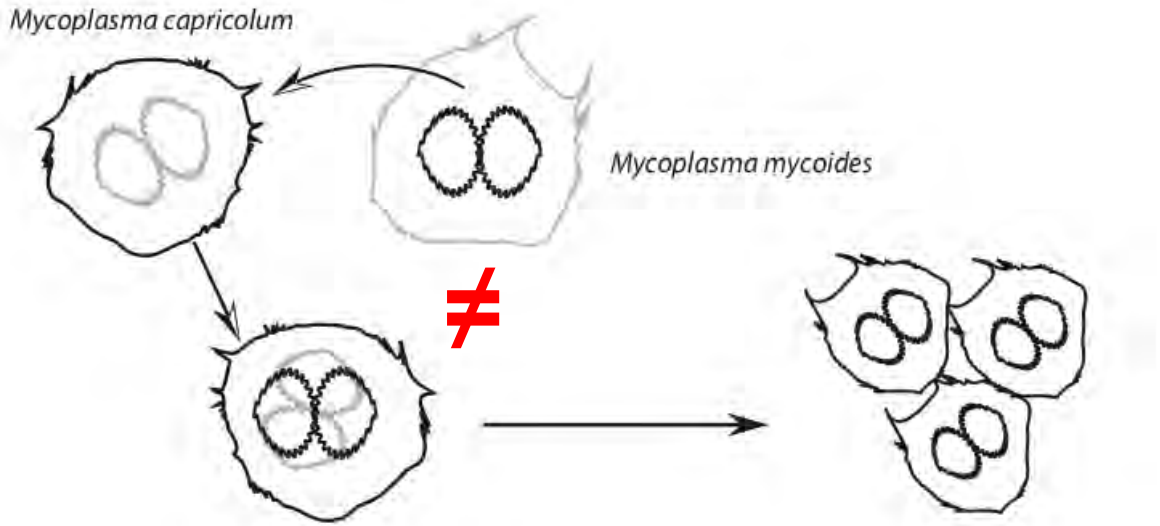
storage is a general flywheel, that is also advantageous in that it controls osmolarity as well

altering the program: the chassis changes!

the program replicates (makes an identical copy)

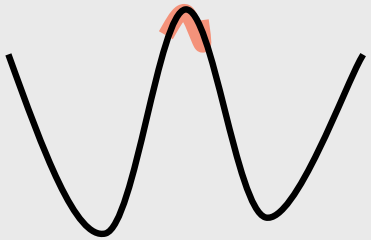
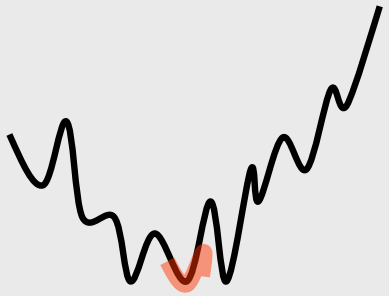
the cell reproduces (makes a similar copy)

this split is the basis of evolution; it is difficult to reconcile with efficient engineering



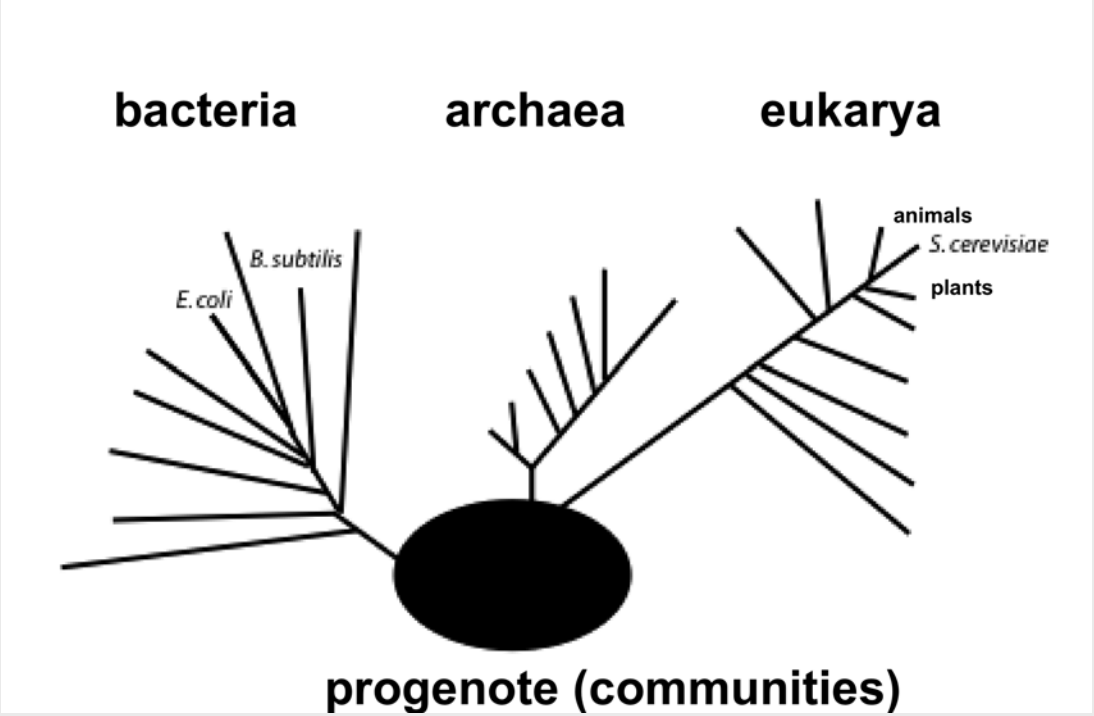
lartigue c, glass ji, alperovich n, pieper r, parmar pp, hutchison ca, 3rd, smith ho, venter jc: genome transplantation in bacteria: changing one species to another. *science* 2007, **317**: 632-638.

a n t i f r a g i l i t y

	fragile	**robust**	**antifragile**
greek mythology	sword of damocles	phoenix	hydra
mathematics			
lifestyle	corporate job	lifetime job	despise money
finance	debt	equity	venture capital
biology	prone to age	buffered	**information trap**

adapted from nassim taleb, antifragility, 2010

evolution is hydra-like



woese (1990)

kurland (2007)

the origin of functions is fuzzy, it splits between the machine and the program; challenges result in dichotomies

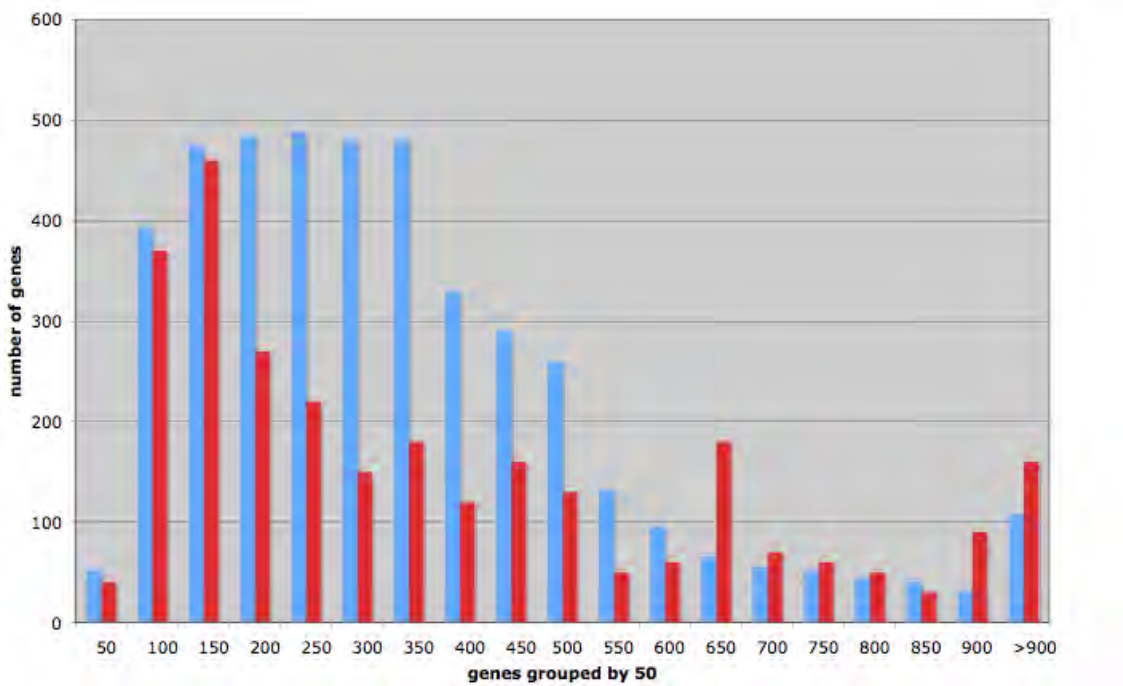
functions for steady-state life

process			nanomachine		*escherichia coli*		*bacillus subtilis*
maintenance							
	rna turnover		degradosome (exosome)		rne pnpa eno tpia orn pcnb		rnja pnpa eno tpia nrna nrnb
	protein turnover		proteasome		clpaxp lon hslvu ftsh...		clpxp lonab clpce clpq clpy ftsh
	repair	refolding			spy dnajk grpe grosl ...		dnajkgrpe grosl
		restoring			pcm frldb frlc msrab		frldb msra b
transcription			rna polymerase		rpoabc nusa nusg mfd sigmas		rpoabc nusag mfd sigmas
translation			ribosome and trnas		rps[a-u] rpl[a-y] rpm[a-j] 20 trna synthetases rmf(55) eftu efts efg modifications...		rps[b-u] rpl[a-y] rpm[a-j] 19 trna synthetases 1 amidotransferase eftu efts efg
		folding	chaperones		tig ppi dnajkgrpe grosl		tig dnajkgrpe grosl
metabolism		carbon			eno pyka pps acee flip ppa ...		eno tpi pyka pdhabc ppac...
		nitrogen			aminotransferases		
		phosphorus			adk ndk ppk...		adk ndk ppnka ppnkb
compartmenting		sensing transport			amino acid; nucleosides or bases; vitamins; carbohydrates or dicarboxylates; polyamines; ions		
replication	repair				chemical alterations, single and double strand breaks and recombination		
	initiation		primase		control of restart		

bias in antifragile proteins

blue: length of the proteins in the whole proteome

red: length of the proteins involved in steady-state life (X 10)



length is not an artefact

while essential during steady state life, rna polymerase subunits (rpob and rpoc are very long proteins; this is not an accident as in *helicobacter pylori*, they are fused in a gigantic protein, that cannot be split into two with keeping resistance to environmental cues

a need for clocks

cells are computer making computers

they work in a highly parallel fashion

this requires clocks to synchronize activities

is there a structural property in proteins that may be related to length and be used for measuring time?

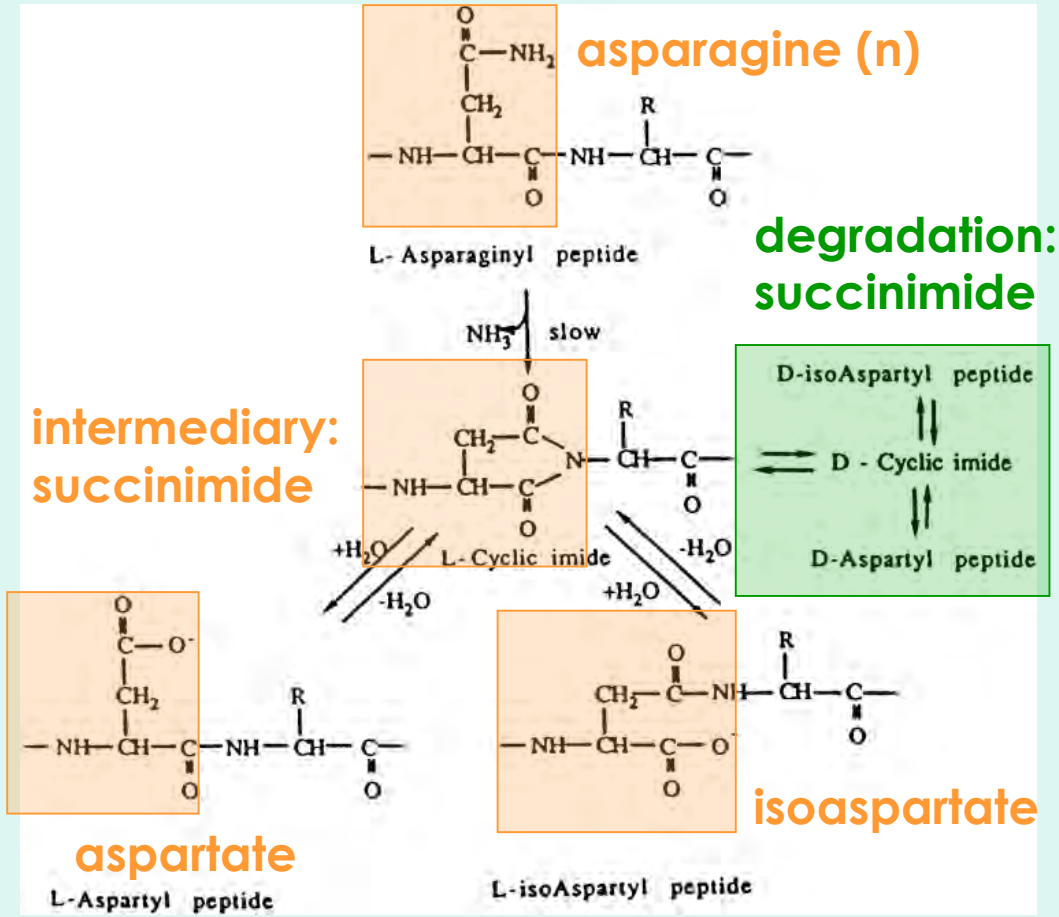
time and antifragility

many steady-state proteins have disordered, flexible, regions

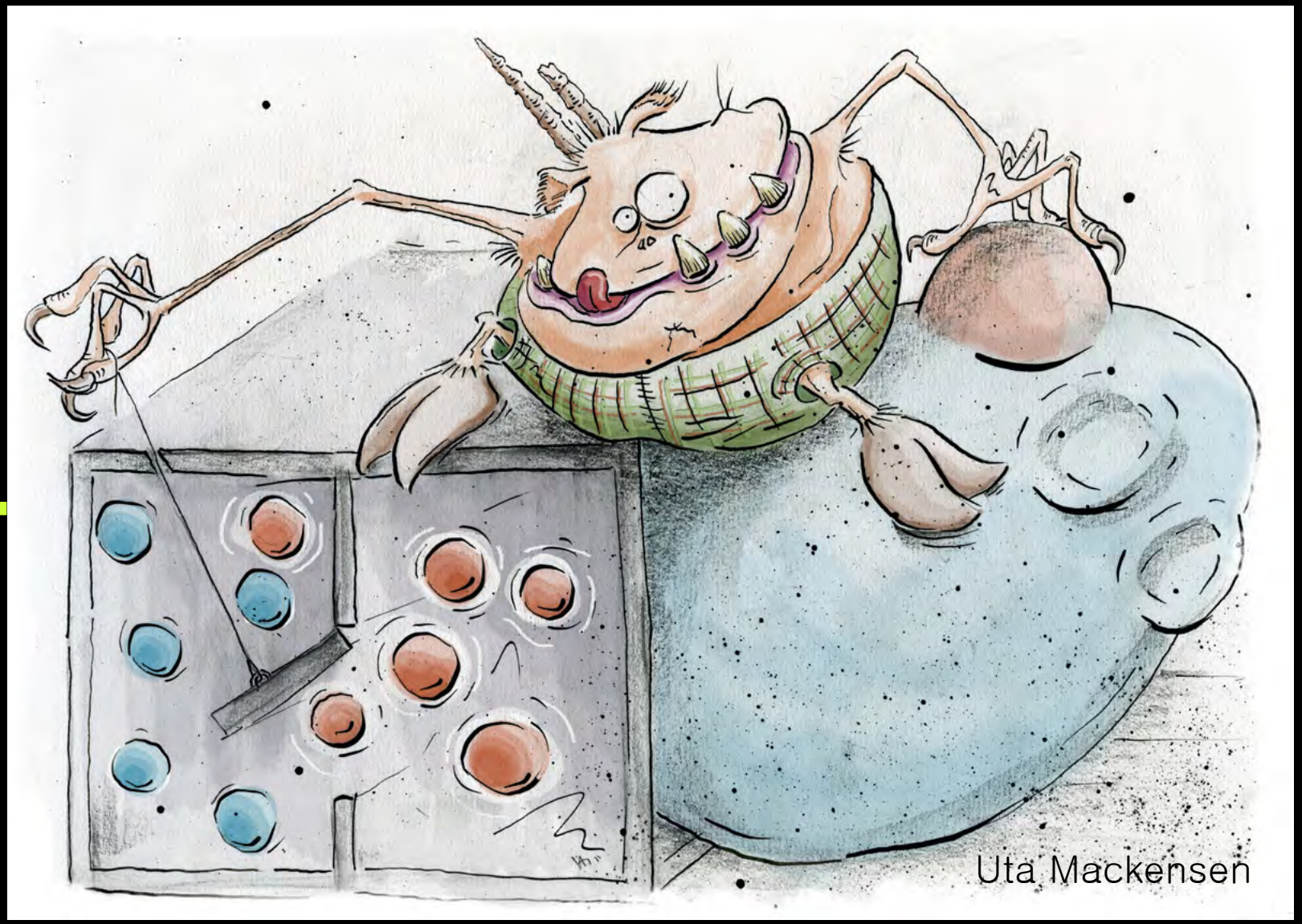
these regions are prone to change spontaneously, at aspartate and asparagine residues

asparagine-glycine di-peptides evolve fast towards l-succinimide l-aspartate, then d-succinimide and finally d-asparate

aging is also a change in information



maxwell's demon's genes



pm binder a danchin (2011) life's demons: information and order in biology embo reports (in press)

a standing enigma

- phage t7 has been redesigned according to engineering rules, and tested using mathematical models

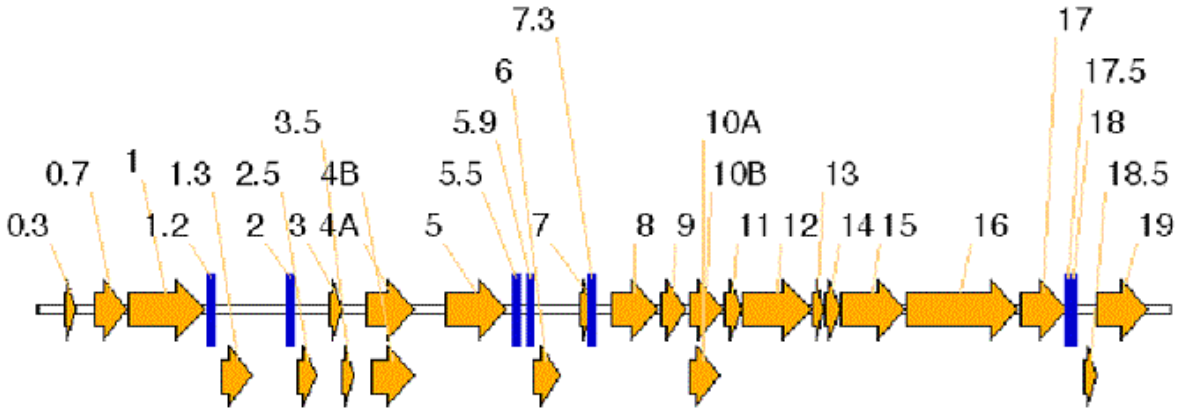
- the synthetic phage forms lysis plaques, but they are smaller than those of its natural counterpart

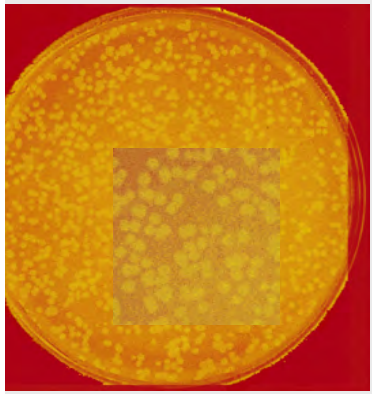
- the evolution the synthetic phage to more virulent forms **erases the human construct**

- **what does this imply for the future of metabolic engineering?**

why does synthetic t7 evolve large plaques?

known genes of bacteriophage t7





•taking control•

• destruction •

• replication •

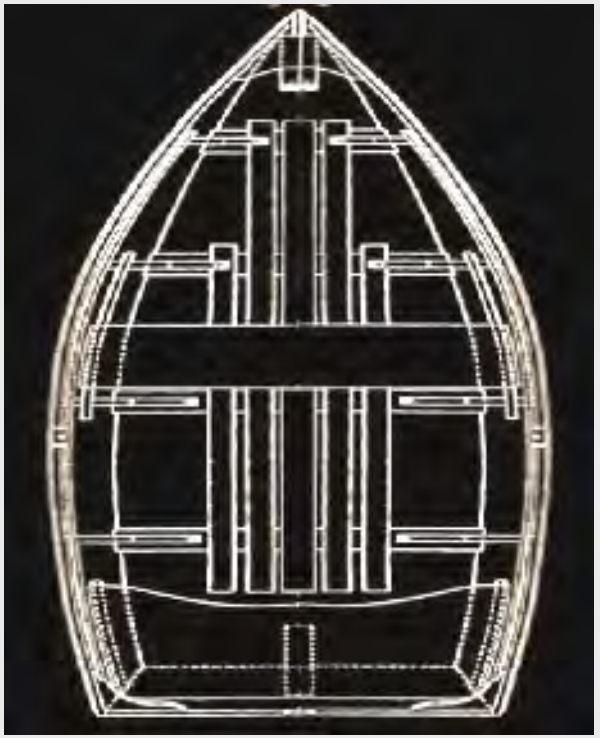
• synthesis of the capsid •

•encapsidation•

• getting out of the cell lysis •

the ship of theseus

- biology is a science of relationships between objects
- it is symplectic (συν together, πλεκτειν, to weave), same word as « complex »
- it is an information that expresses what is conserved in the boat, not the matter of its planks !



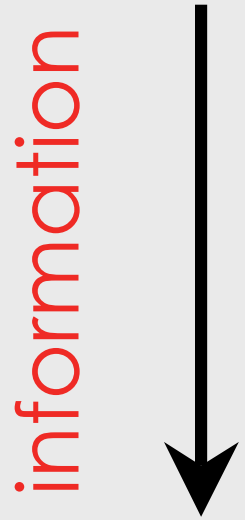
a. danchin the delphic boat, harvard university press, 2003

v. de lorenzo, a. danchin synthetic biology: discovering new worlds and new words 9: 822-827. embo reports, 2008

information is a novel currency of reality

matter / energy / space / time

- classic physics
- quantum physics
- chemistry
- biology
  - development
  - neurobiology
  - linguistics
- mathematics (informatics)



"information is physical" (rolf landauer, 1992)

f u n c t i o n s

functions are actions performing on flows

flows are "tubes" of spatio-temporal manifolds

functions come into three flavours acting on:

- flows of matter
- flows of energy
- flows of information

many types of information

shannon's information (1949) **does not take meaning into account:** this is what replication takes into account

algorithmic complexity (1975): kolmogorov, chaitin, solomonoff

logical depth (1988): bennett (ibm)

further developments (landauer, 1961, ibm): contextual information and links between information and energy: toyabe and colleagues recently (2010) claimed to have converted information directly into energy

r e v i s i t i n g i n f o r m a t i o n

intuition tells us that you need energy to create of information:
szilard 1929, von Neuman 1956, but this is wrong

creation of information is reversible (landauer, 1961; bennett, 1982, 1988, zurek, 1989); to accumulate information requires an energy-dependent process to reset the process and start again

open question:

"to make room" is necessary to accumulate information; how is this performed? can we identify in genomes the genes coding for the functions that permit this process? can we find a ubiquitous and stable energy source?

"useless" reactions

hopfield stated that in order to identify important unexpected functions, we should explore reactions that use energy in an apparently expletive way: « *known reactions which otherwise appear to be useless or deleterious complications* »; this is the case observed with eftu, efts, gtp and translation accurracy

in particular, degradation is exothermic, why should degradation processes use energy?

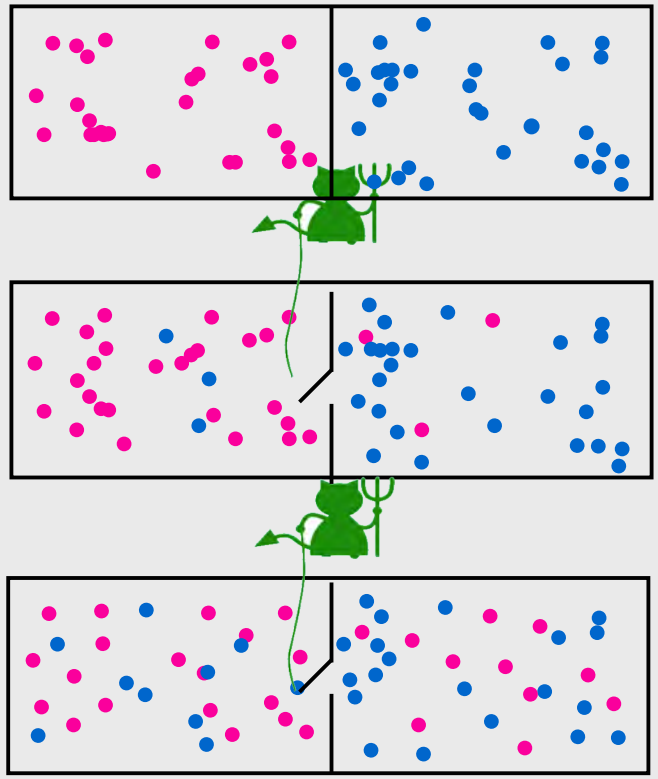
second-kind perpetual motion

hot

cold

0

t



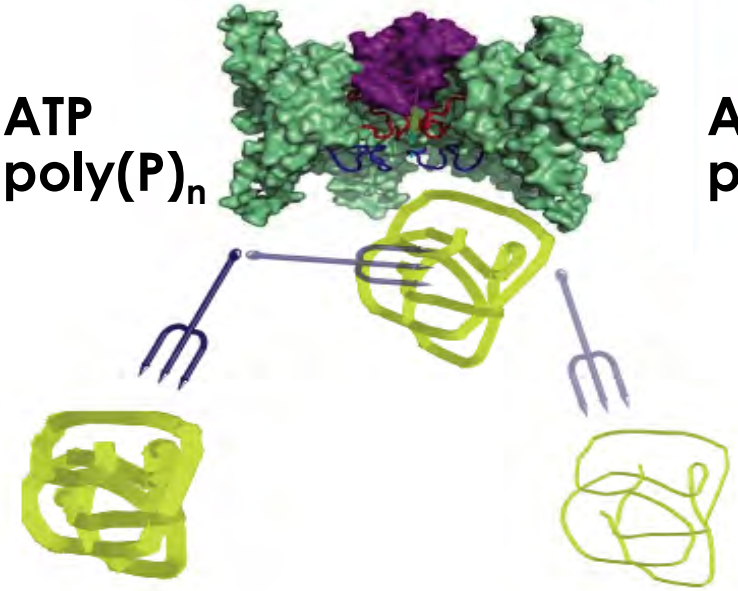
the demon reverses time while measuring the speed of the atoms of gas, recording an information to calculate when it must close the trap, it needs to erase its memory to make a further measurement

0

t

maxwell's demon's genes

ATP
poly(P)$_n$



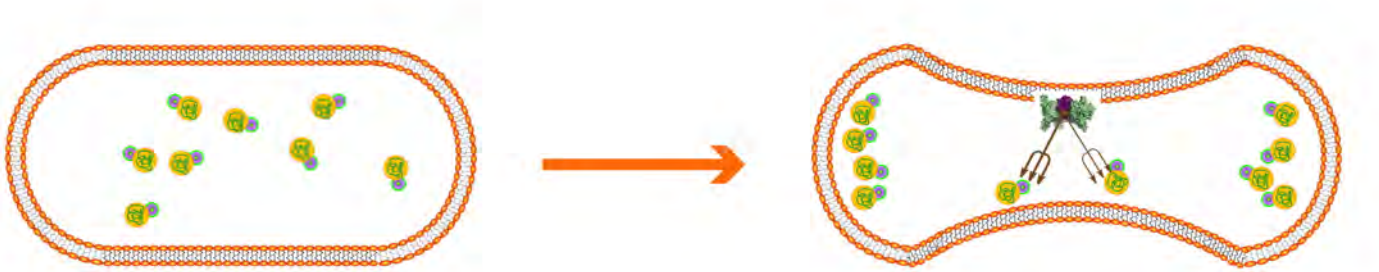
ADP + Pi
poly(P)$_{n-1}$ + Pi **<= in the paleome**

the degradation machinery uses energy to reject unaltered a functional entity; acyldepsipeptides antibiotics uncouple degradation from energy consumption

non functional entities are recognised and degraded

the demon and aggregates

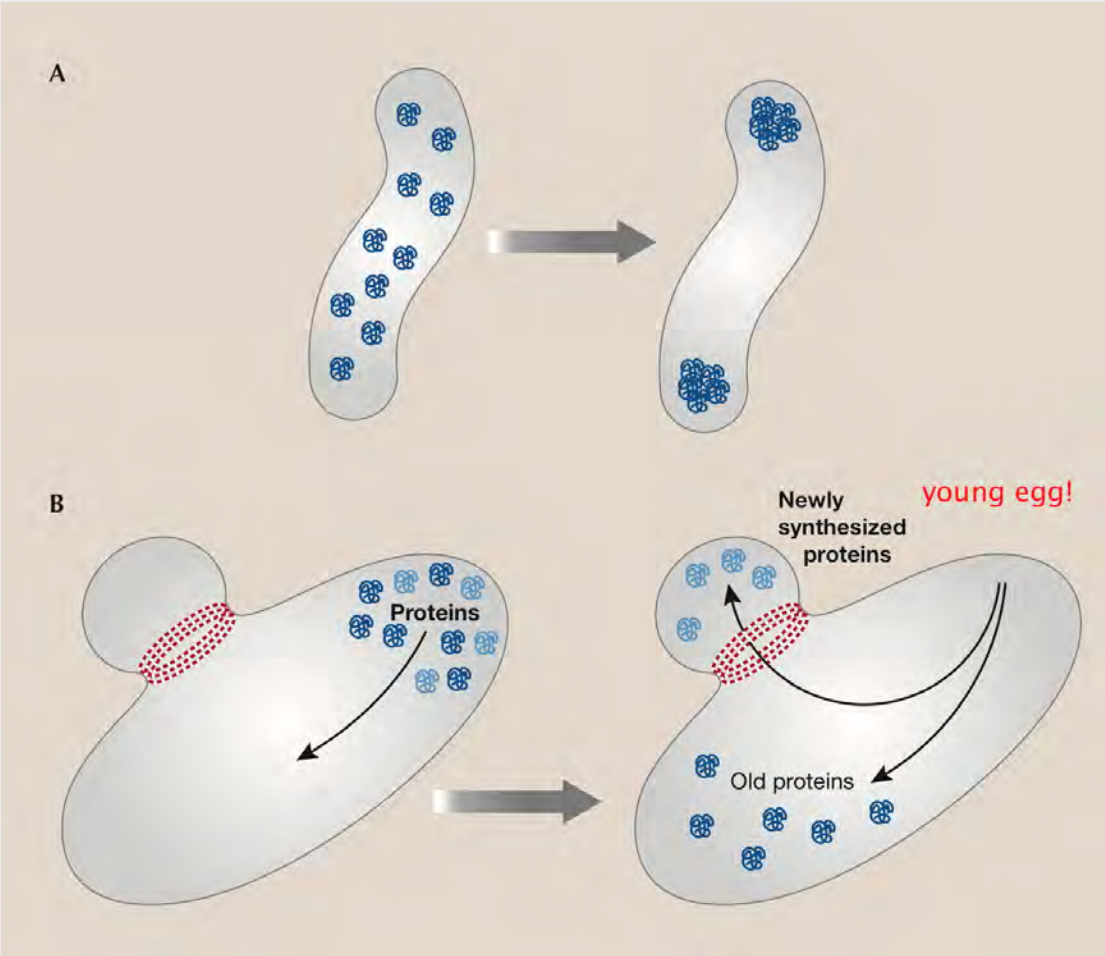
using energy, cells can use their poles as garbage bins, or a specialised cell, such as the mother cell in brewer's yeast, or in formation of a "clean" egg in animals



eggs are very young

the way to create a young progeny is to create cells that only contains newly synthesized proteins, with all the aged ones in the parental cells

a maxell's demon is required in the process; this accu-mulates information



a synthetic cell?

- the engineering view of **sb** precludes that artificial cells be innovative
- it is possible to exclude the genes permetting accumulation of information
- the consequence is that, as all factories, the cell factory will age and will need to be systematically rebuilt
- but this poses problems when applications require that industrial processes are scaled-up: this may not be possible, unless we can harness the function of the maxwell's demon's genes to the human goals

c o n t r i b u t i o n s

in silico
gang fang eduardo rocha

in vivo
agnieszka sekowska undine mechold

collaborations
carlos acevedo-rocha philippe binder (hawai'i)
david ussery markus schmidt

institutions
genoscope, beijing genome institute, fudan university, the university of hong kong, hong kong university of science and technology

AMAbiotics
FONDATION SCIENTIFIQUE FOURMENTIN-GUILBERT
2007 - 2013
microme