

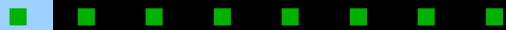


les gènes du démon de maxwell





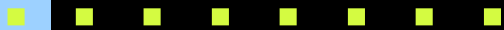
antoine danchin
唐善・安東



école normale supérieure
6 janvier 2010

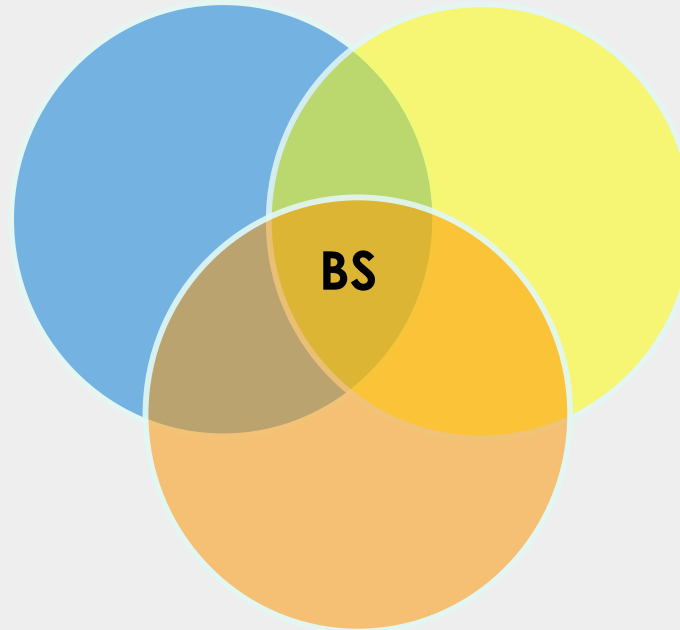


le projet de la biologie synthétique



position de la biologie synthétique

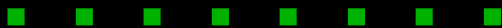
ingénierie
calcul
modélisation



biologie moléculaire
phylogénomique
biotechnologie

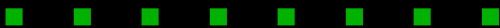
BS

origine de la vie
vie artificielle
vie orthogonale



les objectifs de la biologie synthétique

- un premier objectif de la BS est de **reconstruire la vie**, pour explorer si nous comprenons ce qu'elle est, en découvrant ce qui manque à partir de nos échecs
- dans un second effort, nous conservons ce que nous avons compris pour l'appliquer à des objets d'une **nature physico-chimique différente**
- un troisième objectif voit la vie du point de vue de l'ingénieur, et cherche à classer et normaliser des « biobriques » pour **construire une « usine cellulaire »**



une imprimante 3D auto-reproductrice

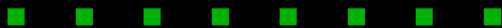
le projet RepRap (Replicating Rapid - prototyper, 2004) vise à créer une imprimante laser 3D auto - reproductrice :

la machine produit la plupart de ses composants (= "biobriques")

ce qui manque :

- o le programme
- o la ligne d'assemblage (gestion du temps et de l'espace, et fonctions spécifiques comme la lubrification)

<http://reprap.org/>



le paradigme électronique

1. abstraction

- retenir l'essentiel de ce qu'on veut (re)construire.

2. conception

- concevoir des circuits génétiques capables d'homéostasie ou de comportements plus compliqués comme des oscillations, de la multistabilité, de la résistance au bruit, etc.

3. standardisation

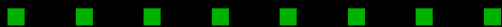
- étape importante pour le futur de la BS, la standardisation est à la base du projet "biobricks" au MIT.

4. modélisation

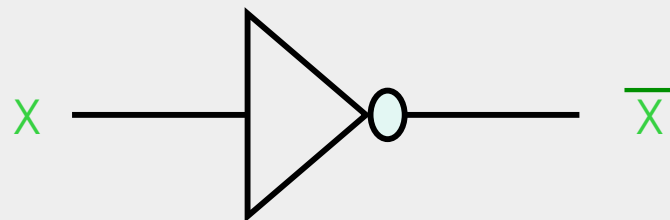
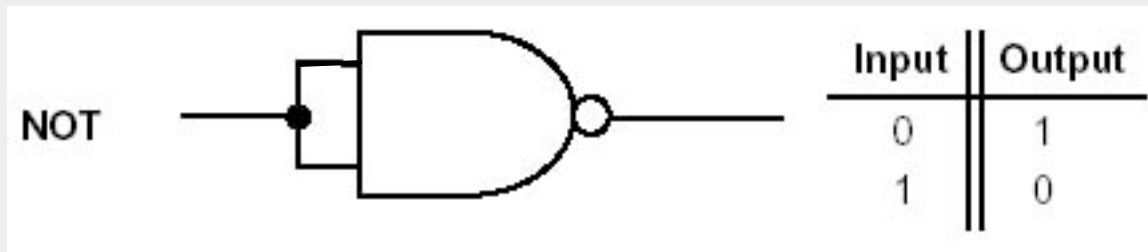
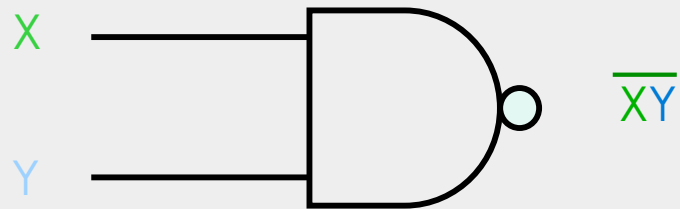
- simuler la construction et analyser son comportement dynamique et sa sensibilité au bruit, aux conditions initiales etc.

5. construction et test

- combinaison des diverses biobriques et introduction dans un "chassis". analyse du comportement de la construction et retour d'expérience. --
retour à 1



portes NAND et NOT



| X | Y | NAND |
|---|---|------|
| 0 | 0 | 1 |
| 0 | 1 | 1 |
| 1 | 0 | 1 |
| 1 | 1 | 0 |

| I | O (NOT) |
|---|---------|
| 0 | 1 |
| 1 | 0 |

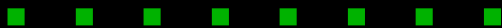
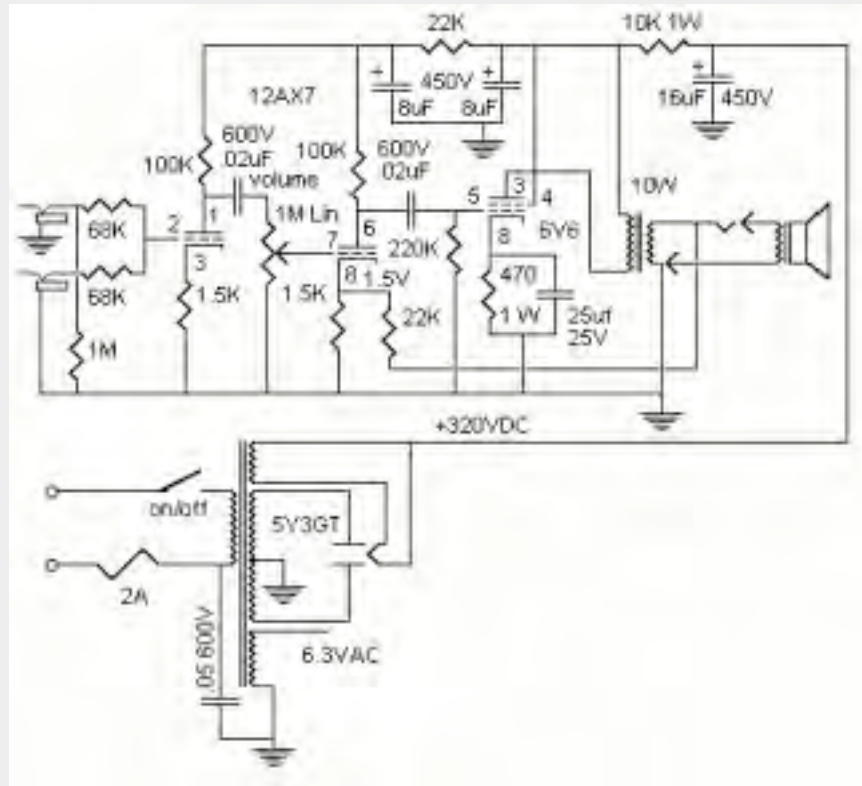
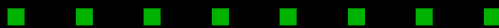


schéma électronique

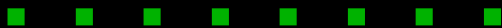


1. abstraction
2. standardisation
3. conception
4. construction
5. modélisation
6. test



circuits logiques à partir d'interrupteurs

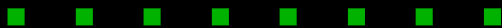
- les protéines jouent le rôle de conducteurs et de signaux
- l'existence de promoteurs et de la durée de vie limitée des messagers et des protéines permet de réaliser des portes logiques
- la porte NAND est un élément logique universel :
on peut construire n'importe quel circuit logique en l'utilisant...



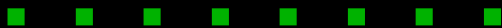
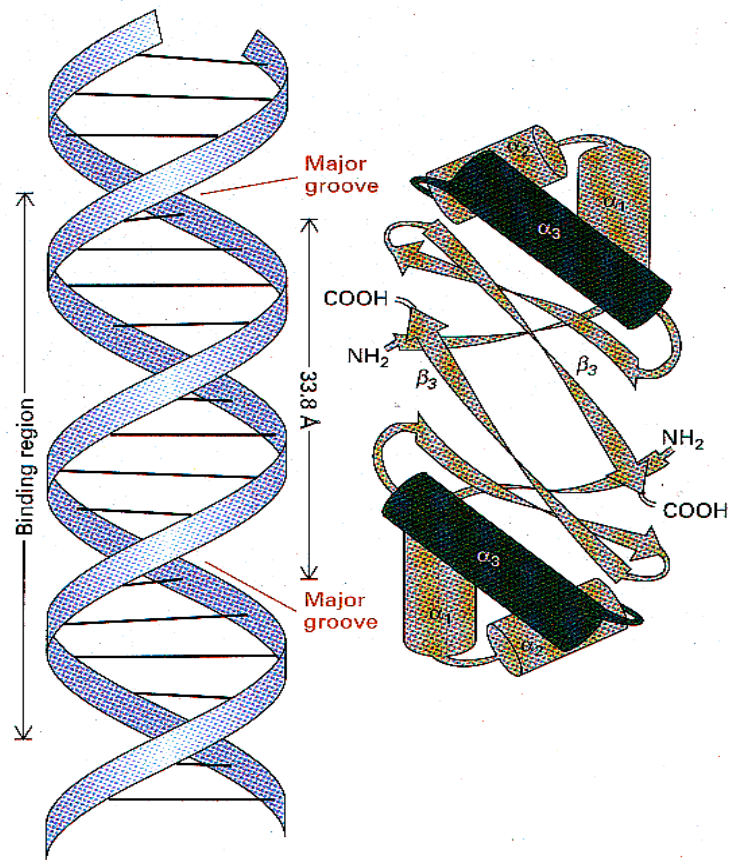
digital ou analogique ?

- digitaliser permet de programmer facilement (machine de turing)
 - traitement du signal + modularité = circuits robustes
- il existe bien des systèmes digitaux dans les cellules
 - lyse ou lysogénie ? le répresseur cl du phage λ
Mark Ptashne A genetic switch: phage lambda revisited Cold Spring Harbor Laboratory Press, 2004.
 - simulation logique du comportement du phage λ

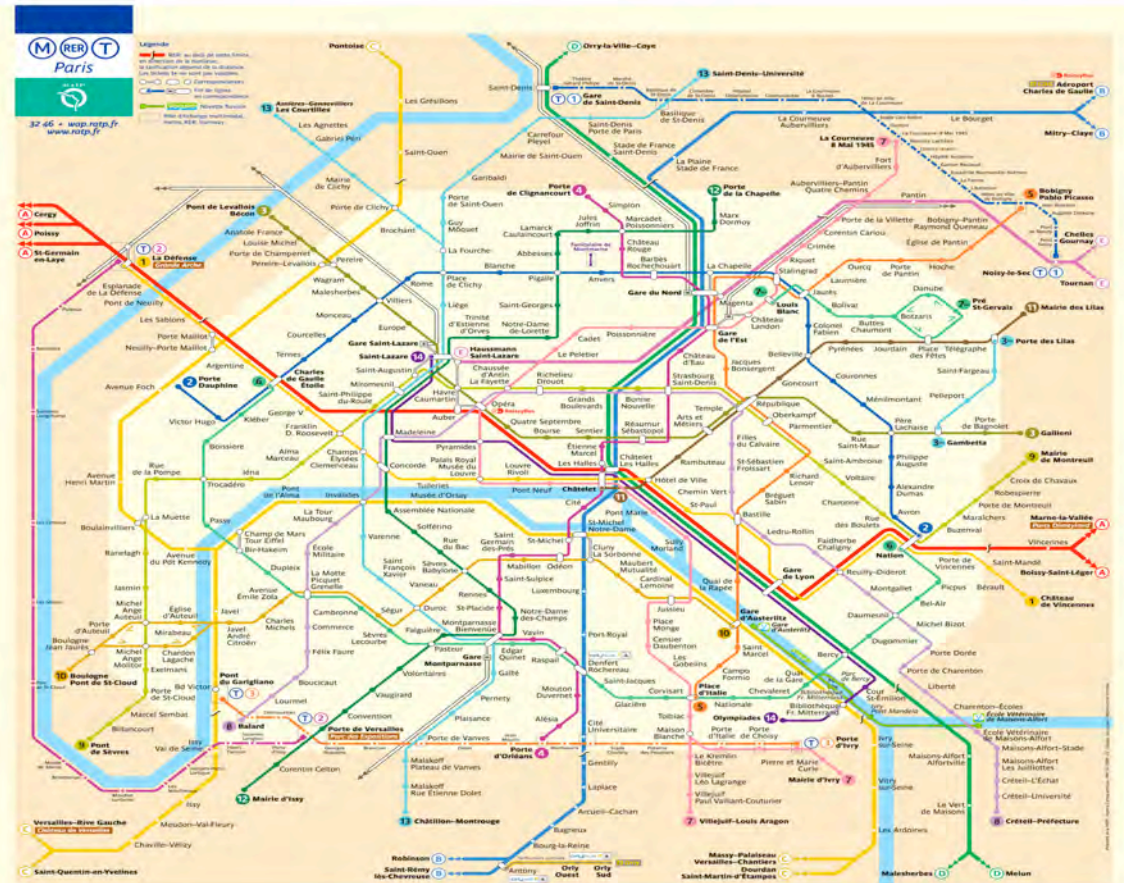
Thomas, R, "on the relation between the logical structure of systems and their ability to generate multiple steady states or sustained oscillations", *springer series in synergetics*, 9 1981 180-193
- il est possible de combiner digital et analogique (cf images et sons)



liaison physique présence / absence



mais attention aux flux de matière !



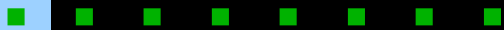
zones compartimentées

échangeurs

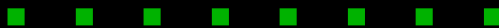
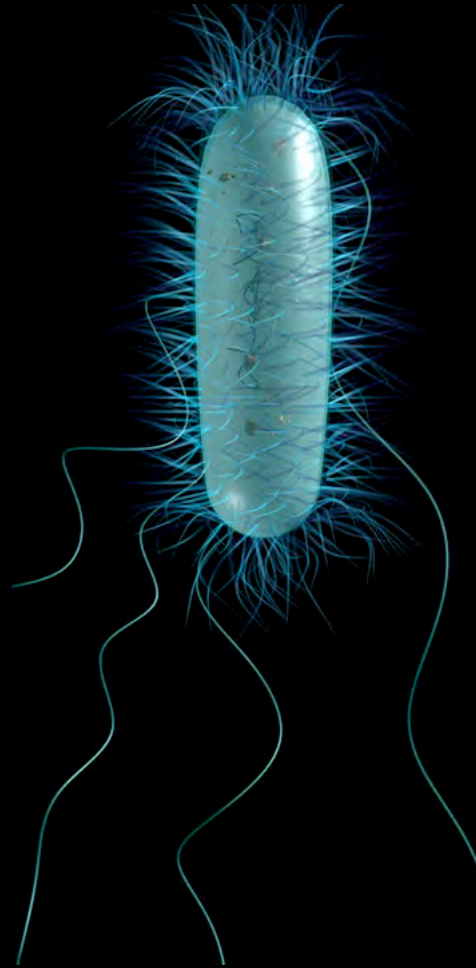
noter que beaucoup de composants biologiques ne sont pas isolables les uns des autres, au contraire de ce qui se passe dans les circuits électroniques



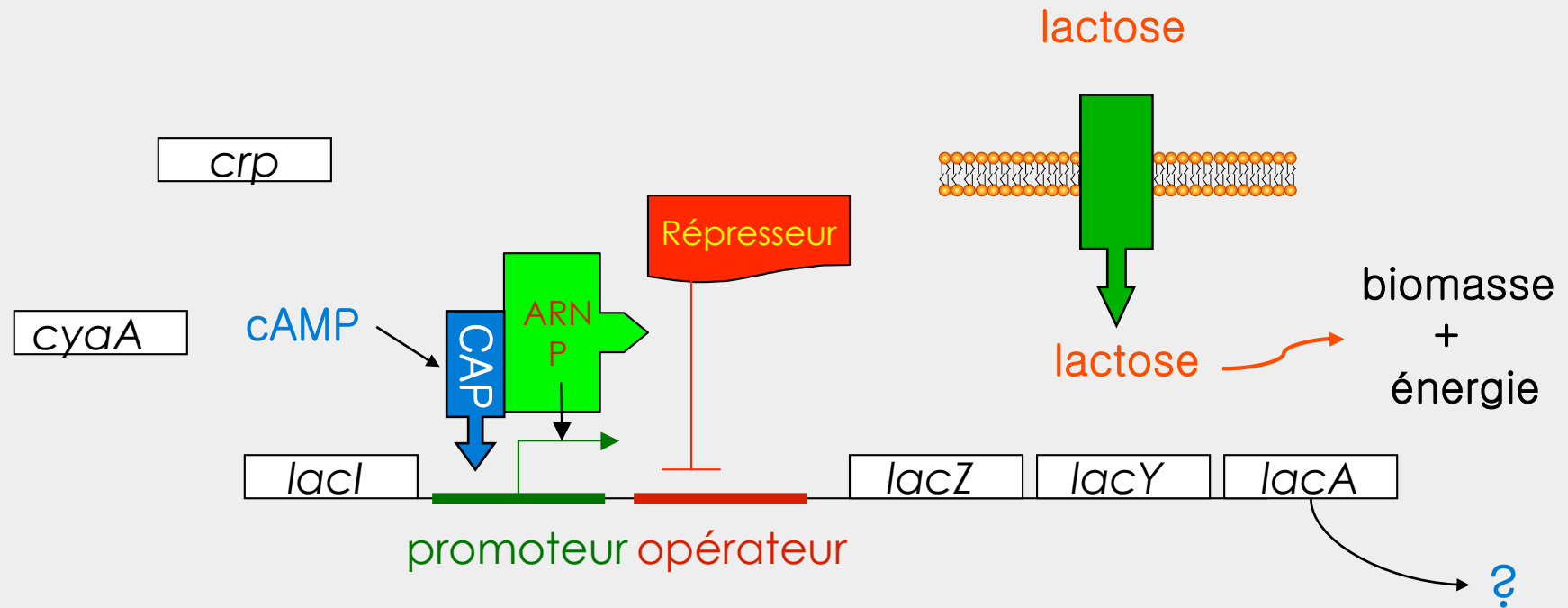
la logique du vivant



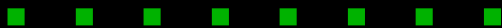
le chassis universel : *escherichia coli*



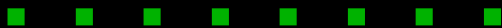
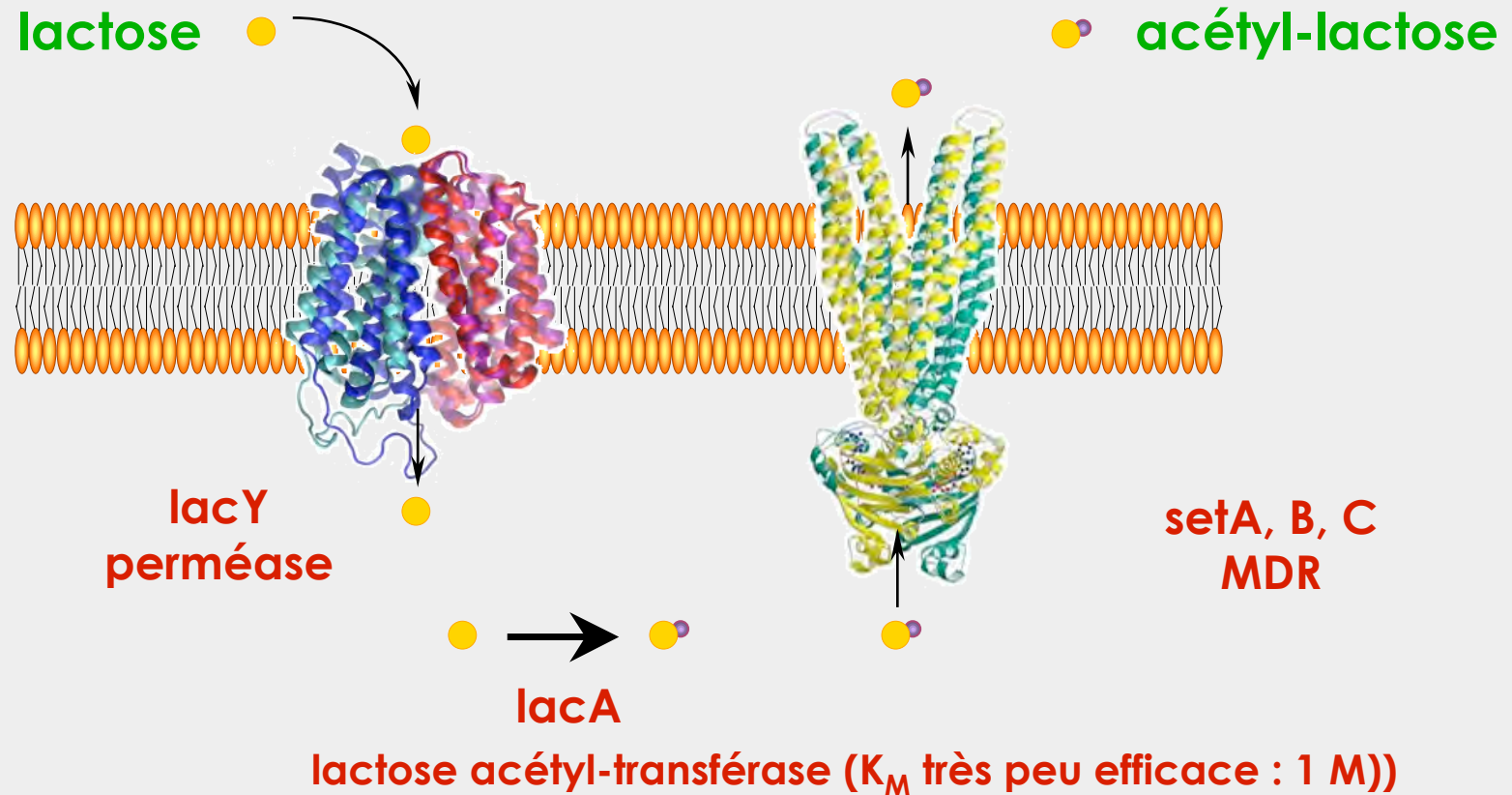
l'opéron lactose revisité



**quelle est la fonction de la lactose acétyl-transférase ?
pourquoi a-t-il fallu 60 ans pour poser la question ?**



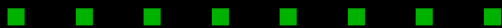
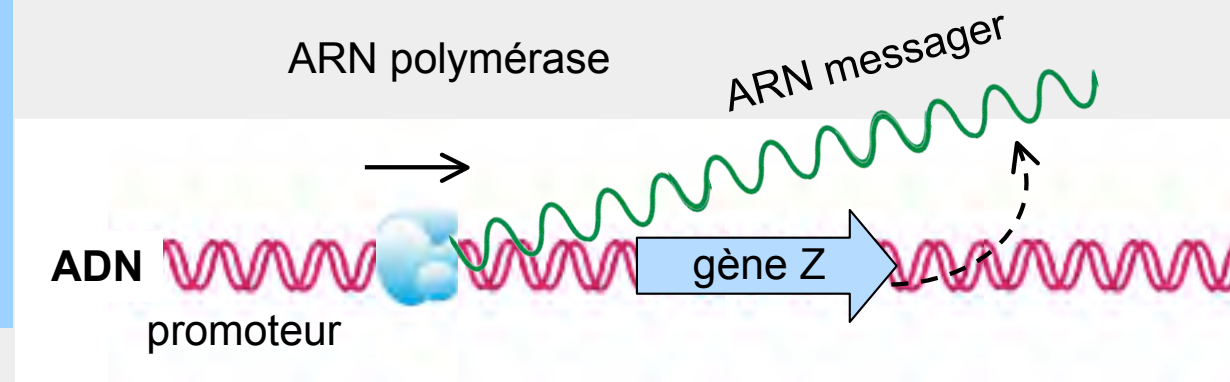
s o u p a p e d e s é c u r i t é



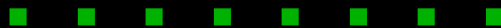
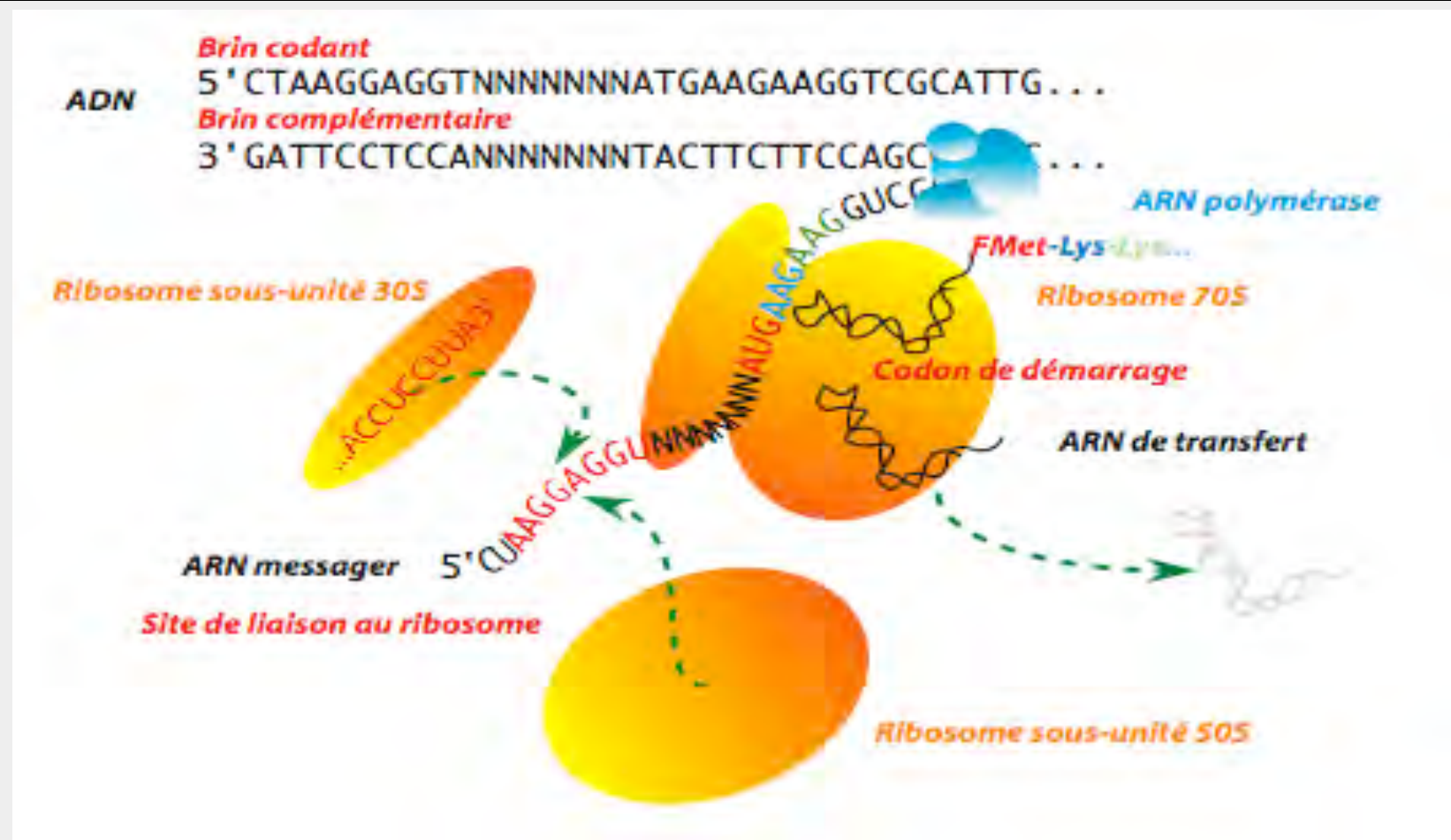
expression des protéines

- l'ARN polymérase se lie au promoteur
- l'ARNP transcrit le gène en ARN messenger
- le ribosome traduit l'ARN messenger en protéine

- reconnaissance
- démarrage
- allongement
- terminaison



synthèse des protéines



principes du génie génétique

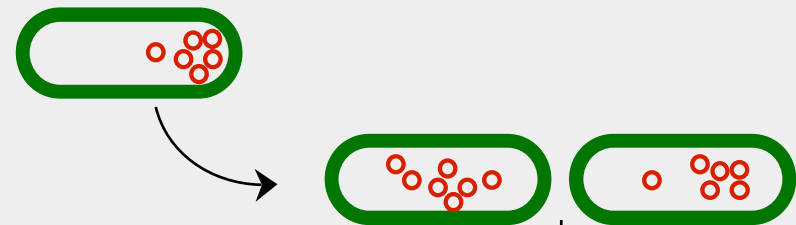
construction d'un plasmide chimérique



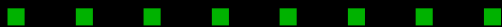
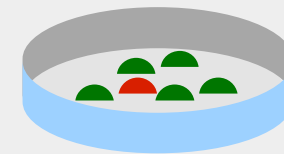
transformation de l'hôte



multiplication et division de l'hôte



criblage ou sélection identifiant un clone positif



régulation de l'expression génétique

contrôle transcriptionnel

démarrage et fin de la transcription

vitesse de transcription

durée de vie du messager (dégradosome)

contrôle traductionnel

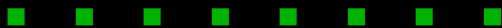
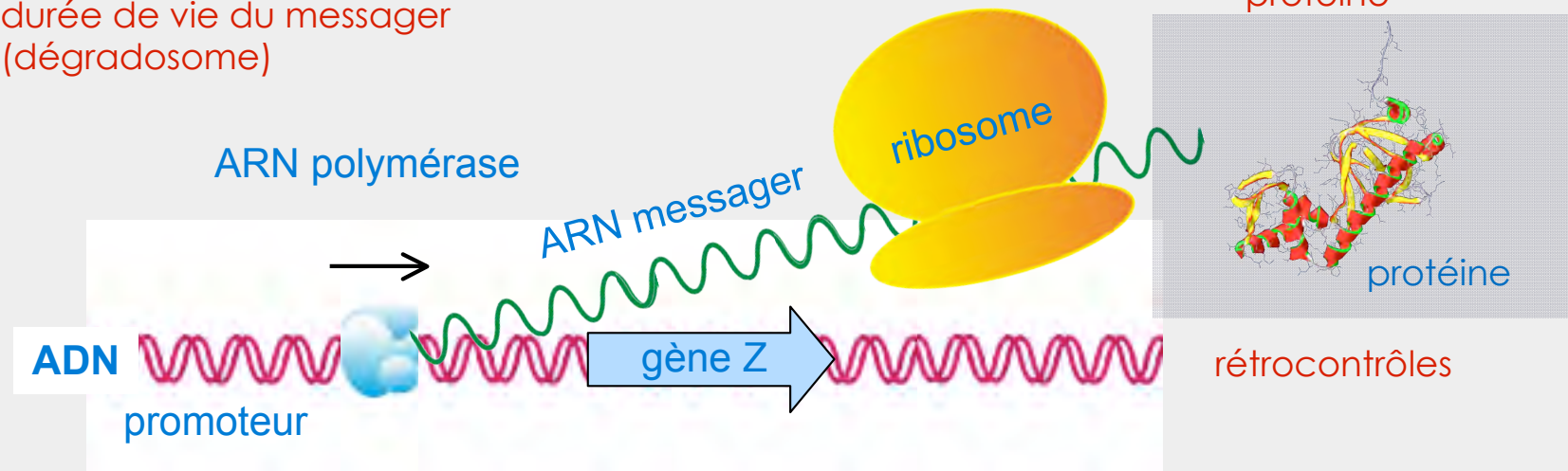
démarrage et fin de la traduction

vitesse de traduction

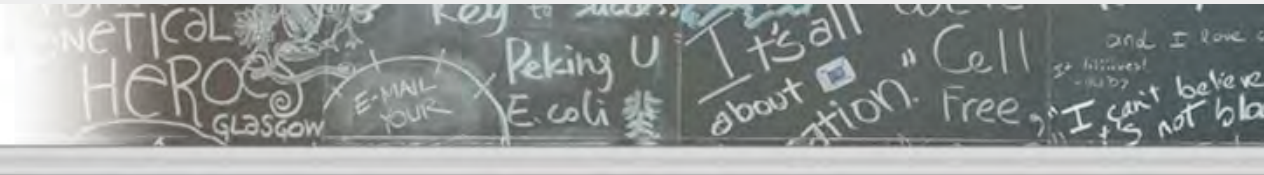
contrôle post-traductionnel

modification chimique de la protéine

durée de vie de la protéine

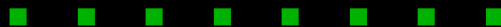
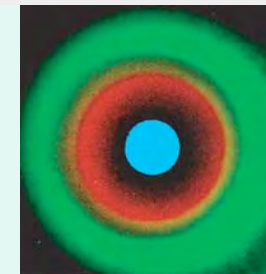
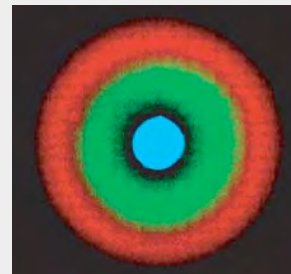


MIT : iGEM et ses « biobriques »



iGEM (International genetically engineered machines) **pose la question :**

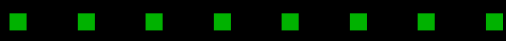
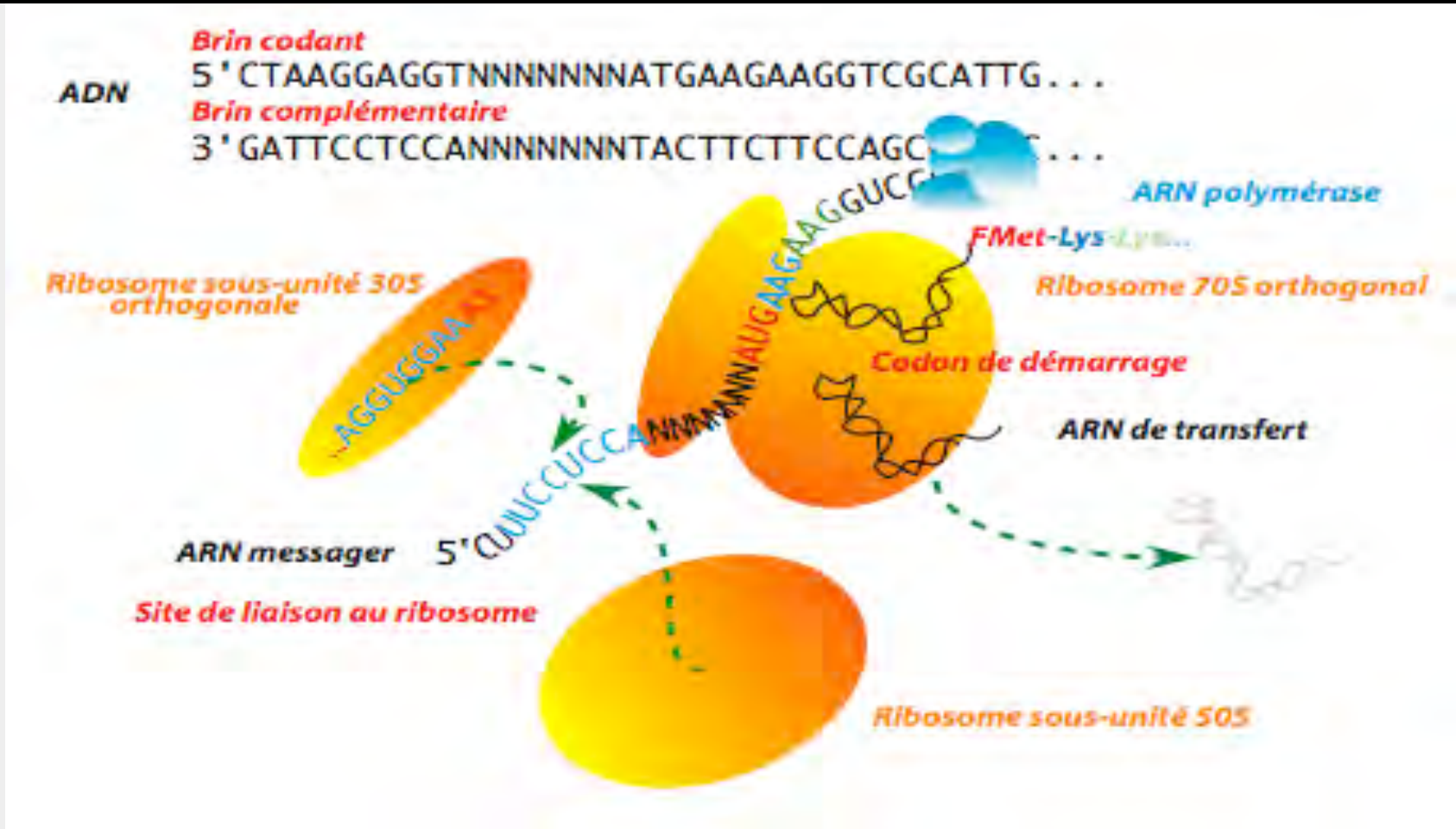
peut-on construire des systèmes biologiques simples à partir d'éléments standardisés, interchangeables et mis en œuvres dans des cellules vivantes ? ou la biologie est-elle trop compliquée pour être reconstruite ainsi ?



Basu S, Gerchman Y, Collins CH, Arnold FH, Weiss R.
A synthetic multicellular system for programmed pattern formation.

Nature. 2005 434:1130-1134

synthèse orthogonale

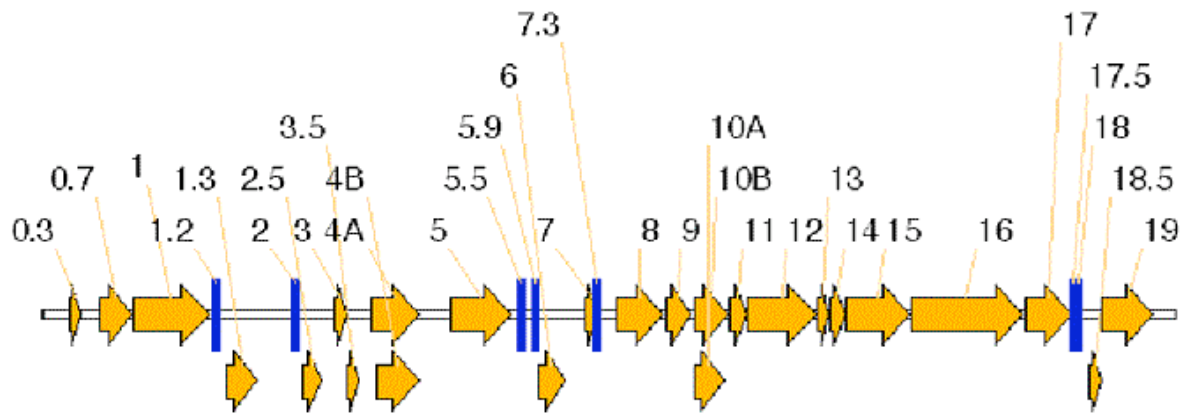


A network of orthogonal ribosome x mRNA pairs.
Rackham O, Chin JW. Nat Chem Biol 2005 1: 159-166.

reprogrammer le bactériophage T7



Les gènes connus du bactériophage T7



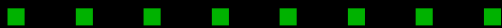
•Prise de contrôle•

- Destruction •
- Réplication •
- Genèse de la capside •
- Encapsidation •
- Préparation de la sortie de la cellule •
- Lyse •

reprogrammer le bactériophage T7



- les zones de contrôle des différentes phases sont partiellement superposées
- elles ne sont pas standardisées
- on sait les redessiner selon des principes d'ingénierie, et tester par des modèles mathématiques leurs propriétés
- les modèles permettent de prédire un comportement du phage synthétique semblable à celui du phage naturel
- l'expérience montre que le phage synthétique forme des plages de lyse, mais un peu plus petites que celles du phage naturel
- l'évolution de ce phage vers des formes spontanément virulentes efface la construction humaine

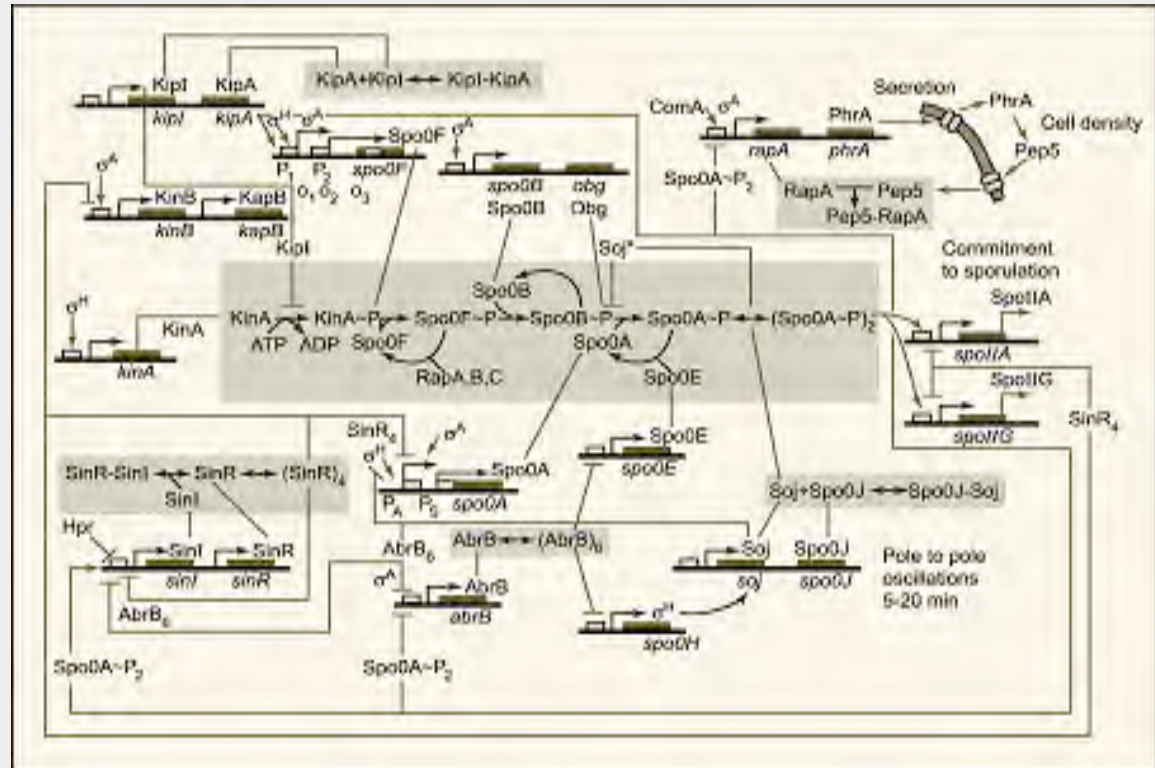


un exemple : bio-spice

Bio-SPICE analyse les contrôles de la sporulation chez *Bacillus subtilis*

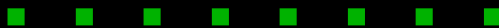
source: projects Bio-SPICE au LBNL

(PI: Adam Arkin)



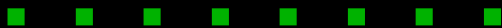
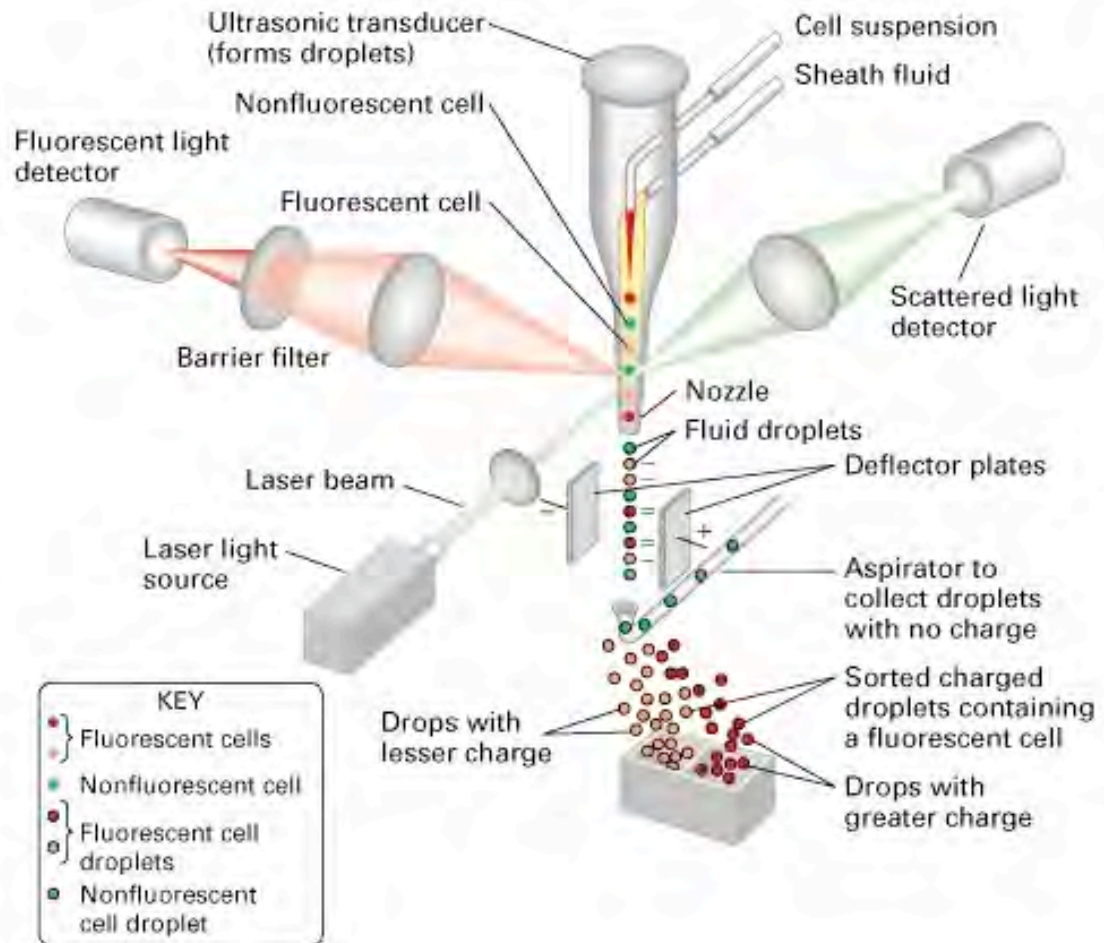
open source for systems biology

<http://biospice.sourceforge.net/>



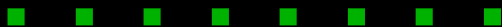
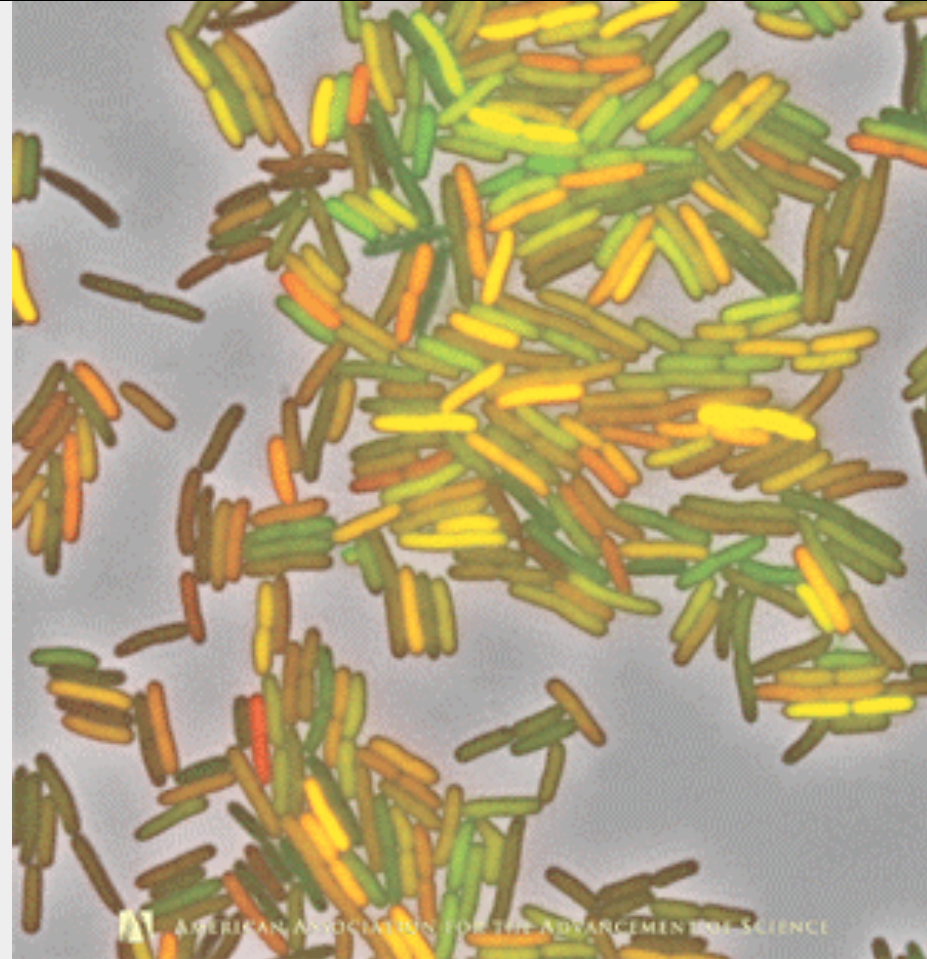
cytométrie de flux (FACS)

détection de la synthèse des protéines dans des cellules individuelles par fluorescence

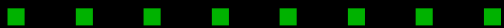
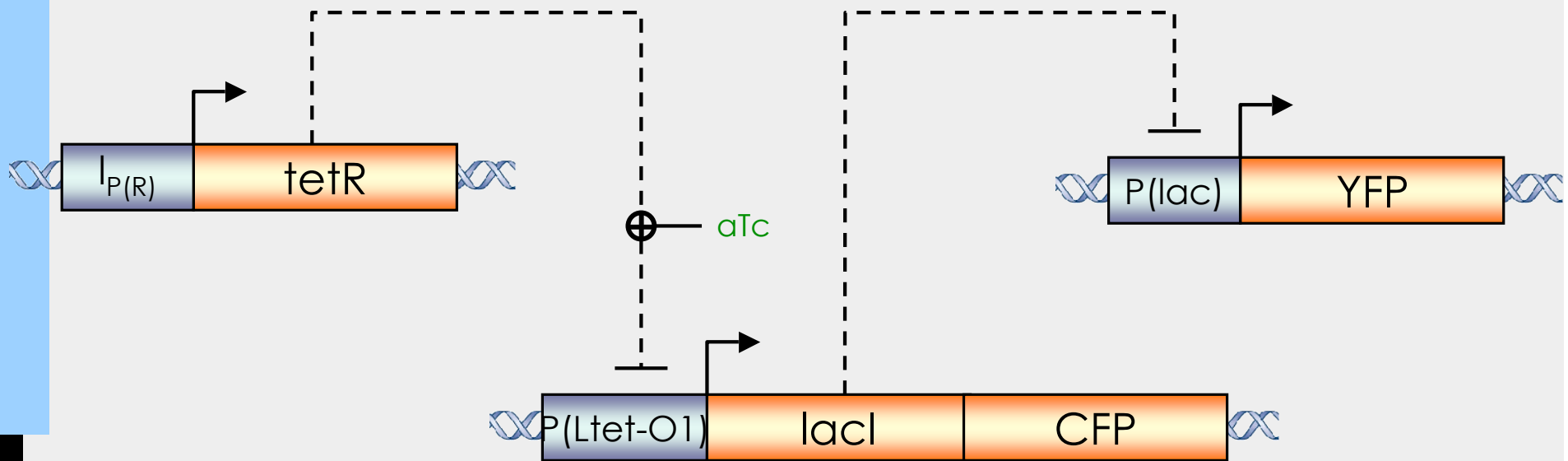
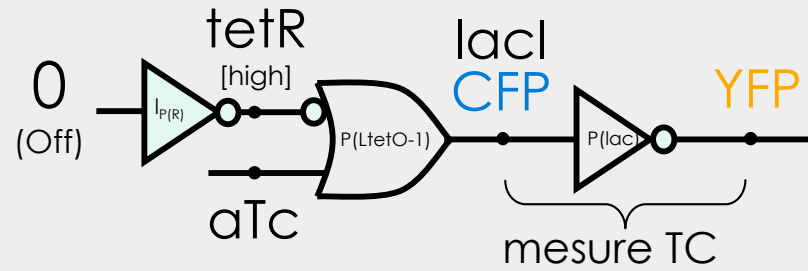


effet du bruit

l'expérience montre que la synthèse varie énormément d'une cellule à l'autre ; la plupart des modèles construits en biologie des systèmes est profondément irréaliste

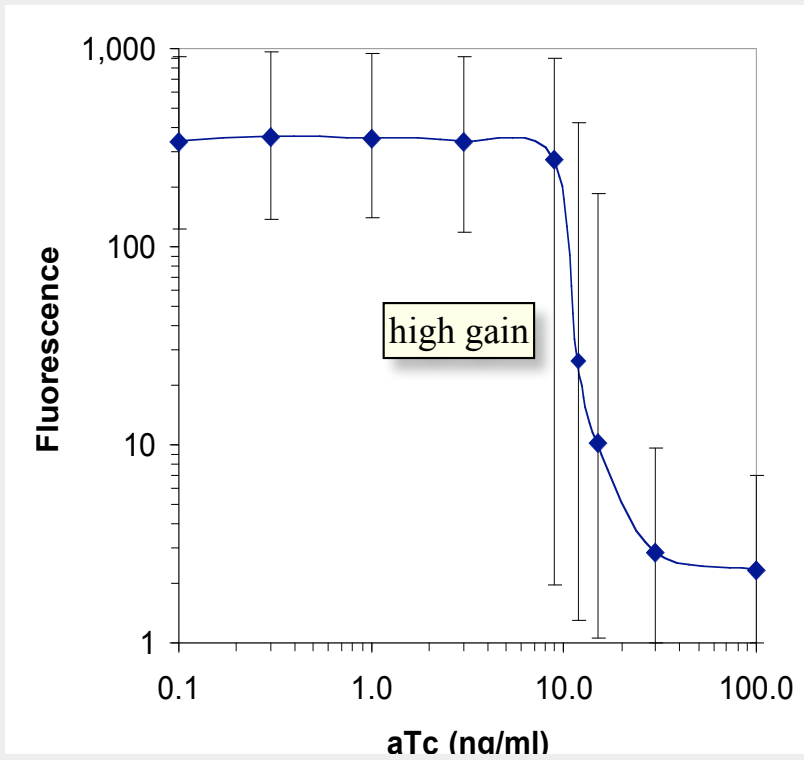


diagnostic fonctionnel

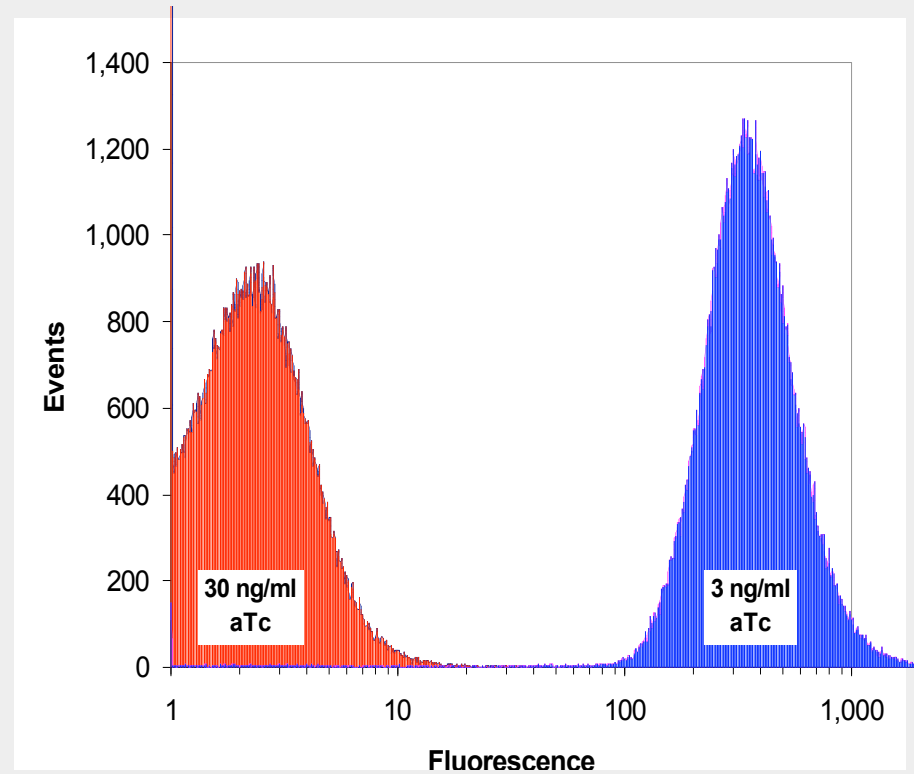


évaluation de la dispersion

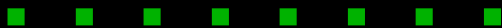
Gain / Signal restoration



bruit

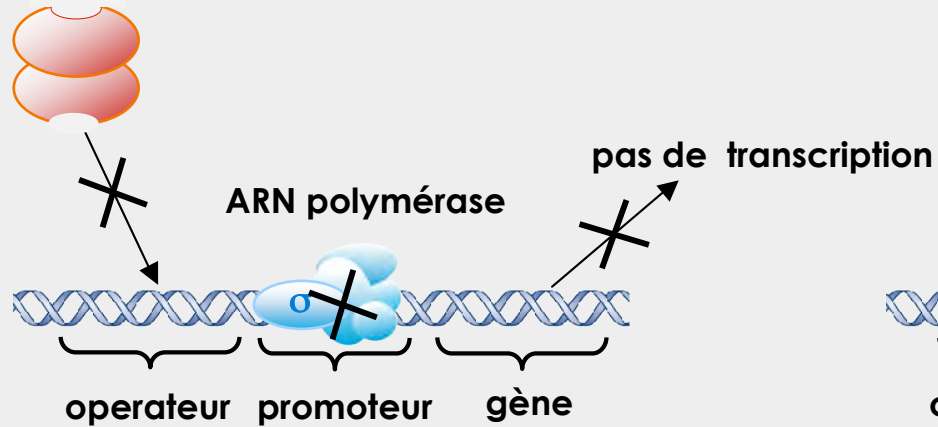


*note: graphing vs. aTc (i.e. transfer curve of 2 gates)



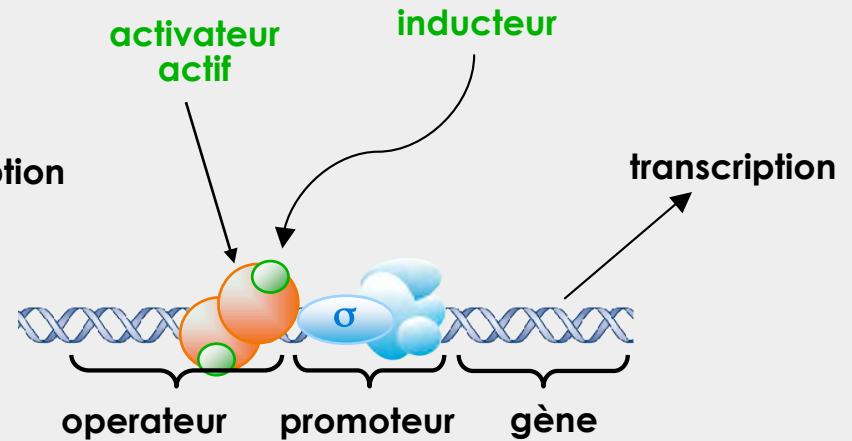
régulation par activation

activateur
inactif

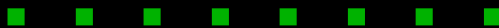
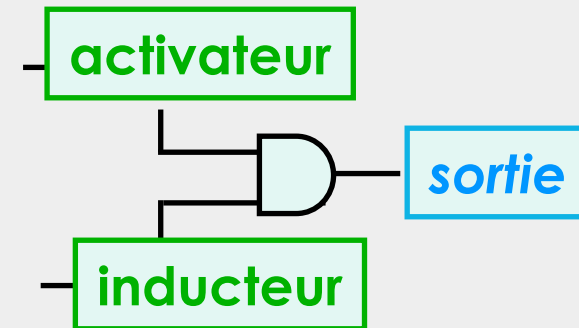


activateur
actif

inducteur

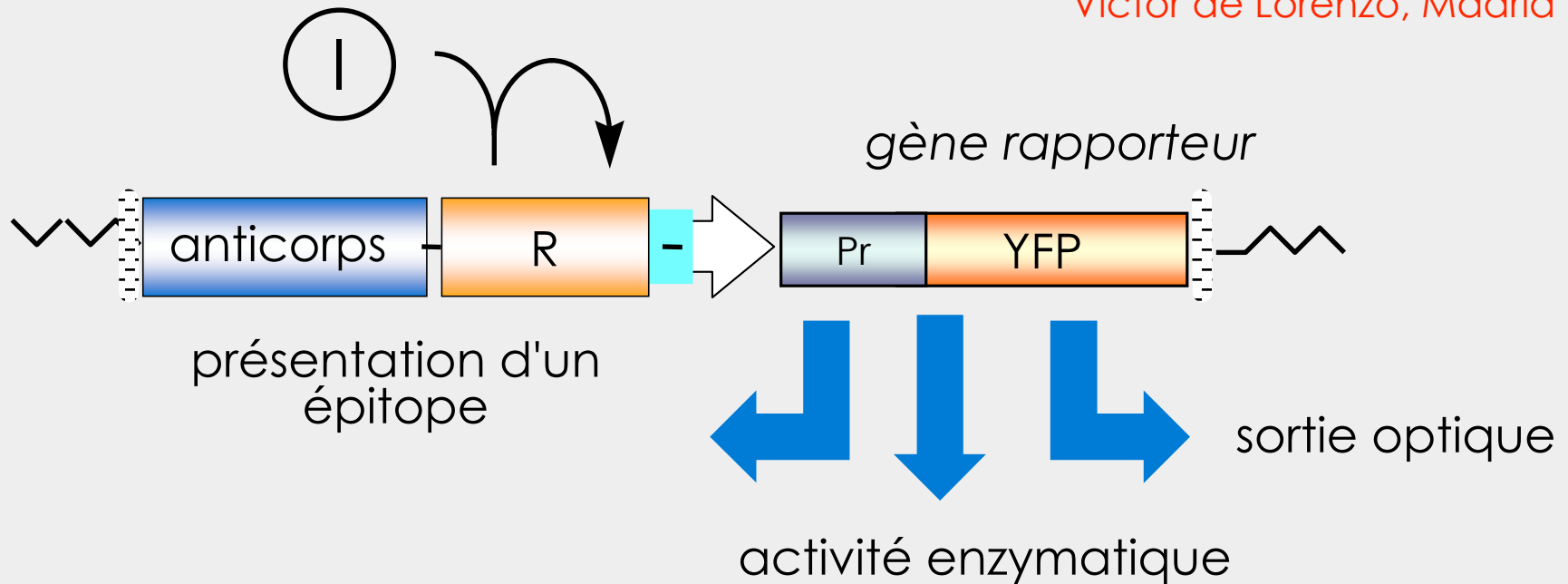


| activateur | inducteur | sortie |
|------------|-----------|--------|
| 0 | 0 | 0 |
| 0 | 1 | 0 |
| 1 | 0 | 0 |
| 1 | 1 | 1 |

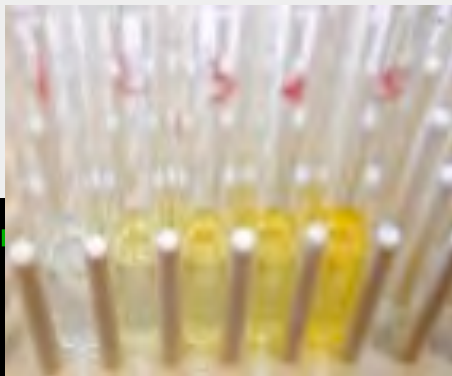


détecteurs associés à des promoteurs métaboliques

Victor de Lorenzo, Madrid



coloration



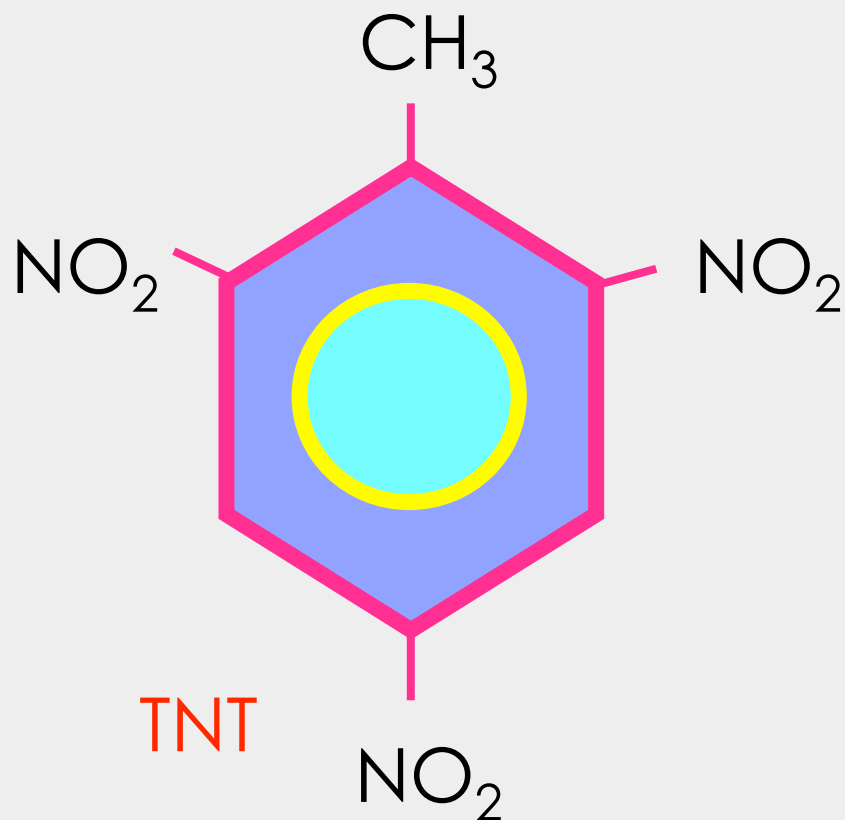
fluorescence



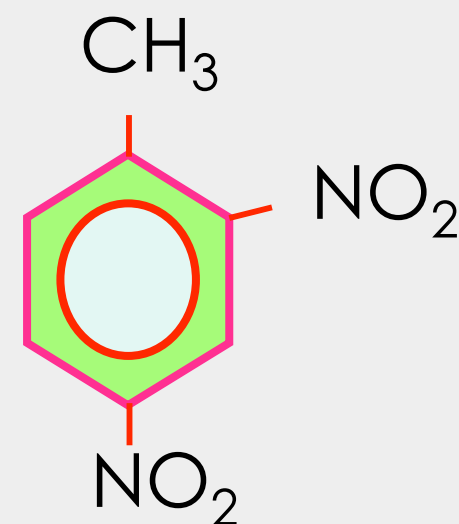
bioluminescence



vers un détecteur d'explosifs

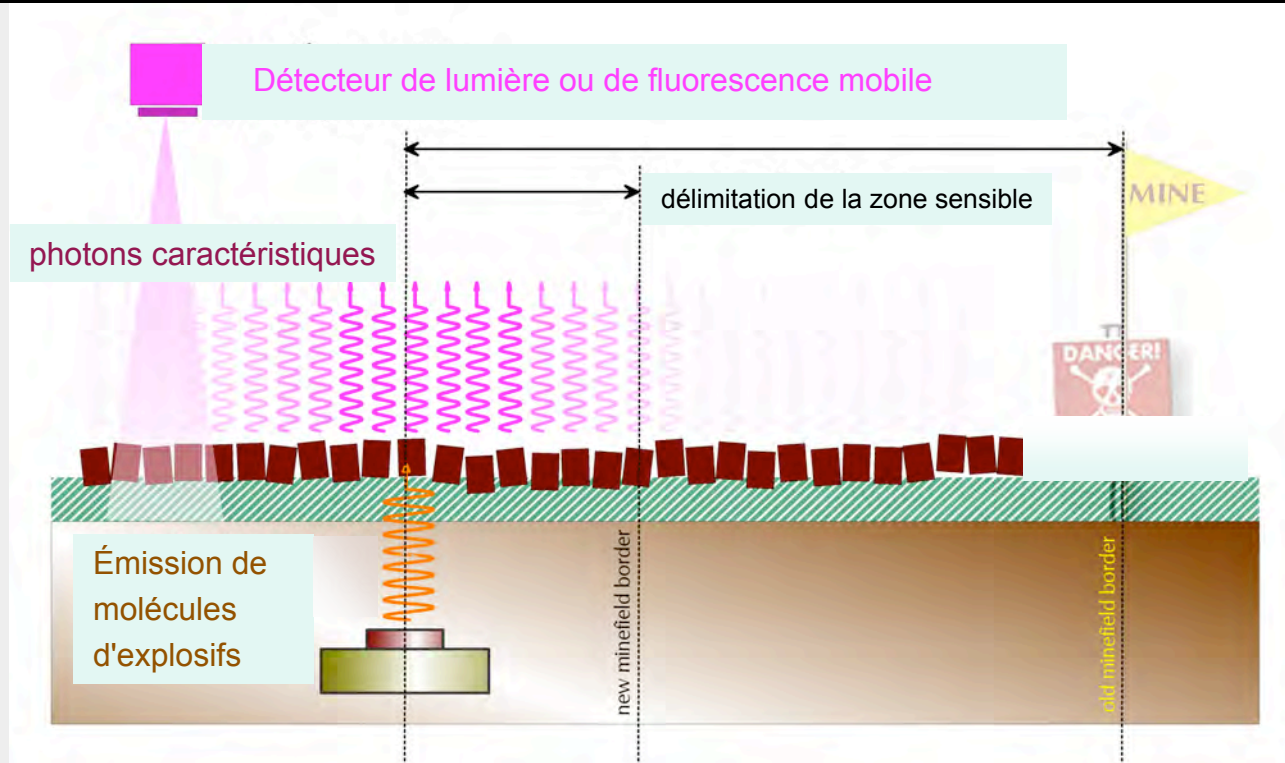


la molécule test
2,4-DNT

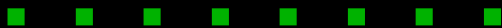


travaux du groupe de victor de lorenzo à madrid

construction d'un détecteur de mines



L'épandage de microbes sensibles aux émissions de molécules d'explosifs permet de délimiter la zone spécifique où se trouvent les mines par l'émission de lumière ou de fluorescence qui peut être détectée par un détecteur de lumière mobile (la détection des photons est très sensible)

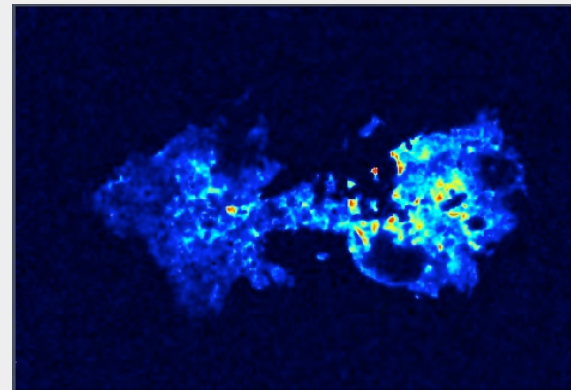
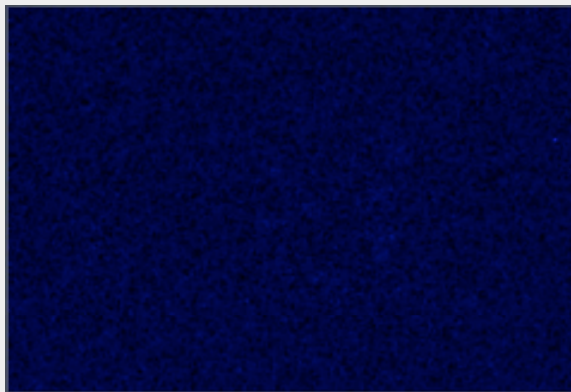


dispersion de bactéries sentinelles dans un sol modèle

wt XylR

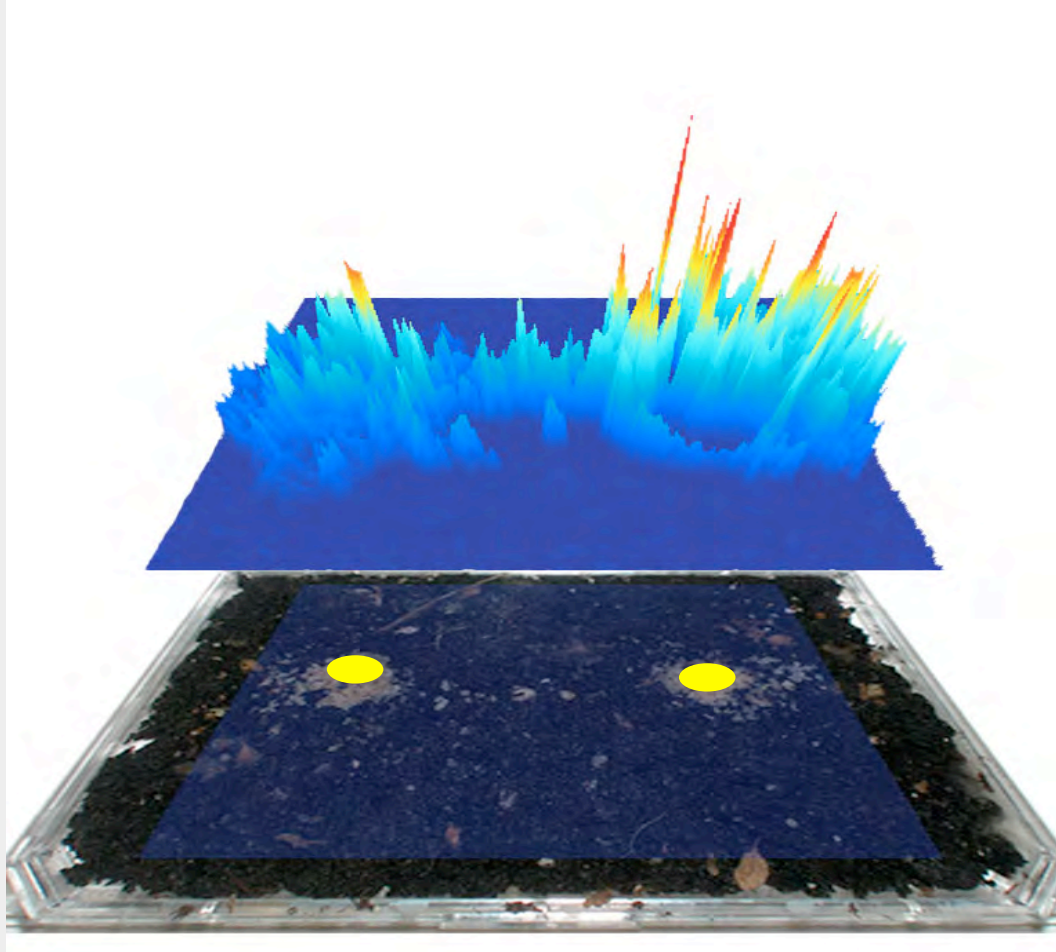


XylR5

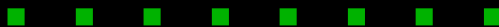


■ ■ ■ ■ ■ ■ ■ ■
Stable implantation of orthogonal sensor circuits in Gram-negative bacteria for environmental release
de las Heras A, Carreño CA, de Lorenzo V
Environ Microbiol (2008) **10**: 3305-3316

détection de dinitrotoluène dans un sol modèle

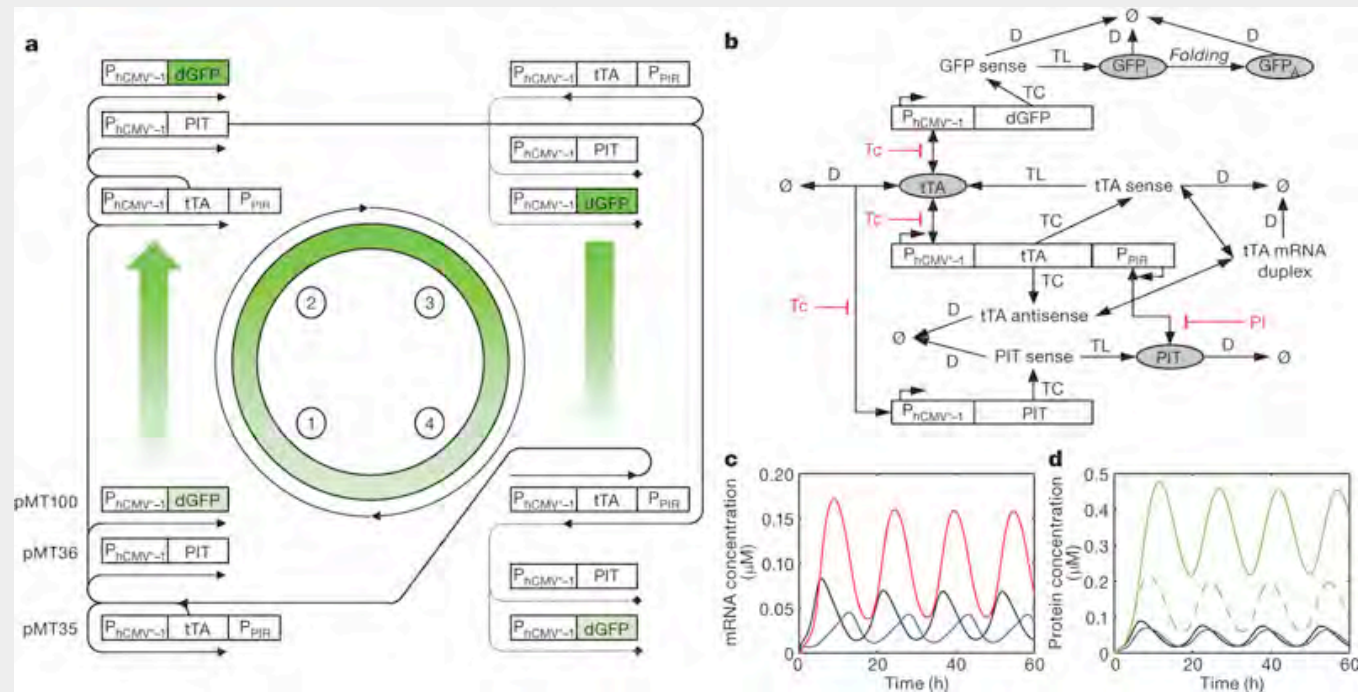


Stable implantation of orthogonal sensor circuits in Gram-negative bacteria for environmental release.
de las Heras A, Carreño CA, de Lorenzo V
Environ Microbiol (2008) **10**: 3305-3316



biologie synthétique dans les cellules de mammifères

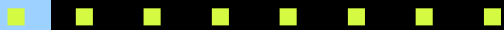
horloge moléculaire dans des cellules de mammifères et prédiction des oscillations produites à partir d'ARN régulateurs



nature



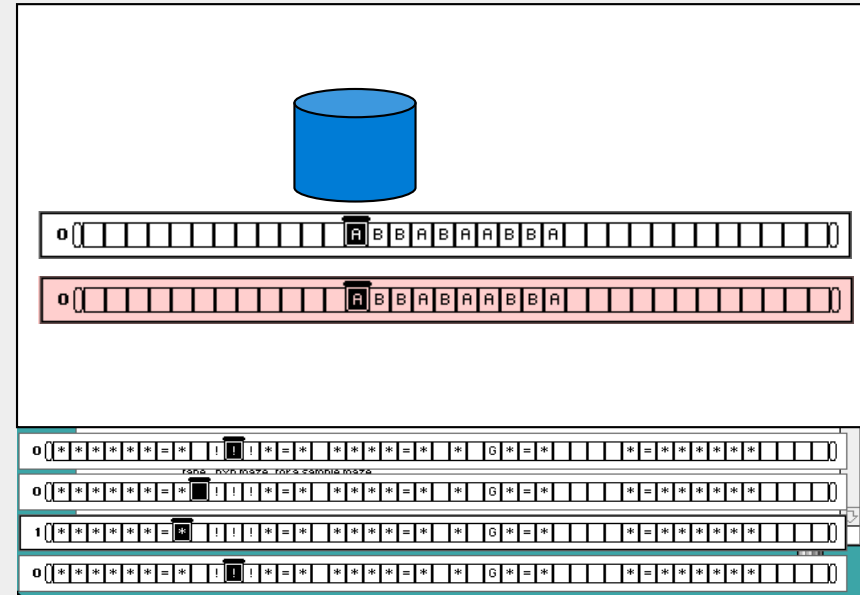
du programme à la cellule



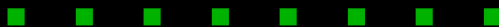
ce qu'est le calcul

deux entités permettent le calcul :

- une machine capable de lire et d'écrire
- un programme (sur un support physique), séparé par l'esprit humain, pas conceptuellement en deux entités :
 - programme (définissant l'"objectif")
 - données (définissant le contexte)



la machine est distincte de l'entité données/programme



cellules et ordinateurs

la génétique repose sur la description des génomes comme des textes écrits au moyen d'un alphabet : mais les cellules se comportent-elles comme des ordinateurs ?

→ transfert génétique horizontal

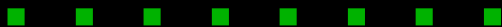
→ virus

→ génie génétique

→ transplantation d'un génome nu dans une cellule réceptrice suivie du changement de la machine réceptrice en une nouvelle (2007)

tout concourt à séparer

"machine" (usine cellulaire) et "données/programme" (le génome)

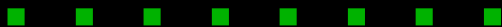


v i r u s e t p r o g r a m m e

comme un virus informatique, un virus biologique est fait d'un ou plusieurs acides nucléiques (son génome), codant une ou plusieurs protéines permettant sa réplication, sa protection dans la cellule et l'environnement, et sa propagation.

souvent il code aussi des activités qui interfèrent avec le métabolisme de la cellule hôte, le détournant au profit de sa propre construction ; comme dans les ordinateurs il peut piéger des fonctions de la machine et les transmettre à sa descendance.

cependant il ne code pas la totalité des objets permettant l'expression de son génome : **il dépend entièrement de l'usine cellulaire pour se perpétuer.**

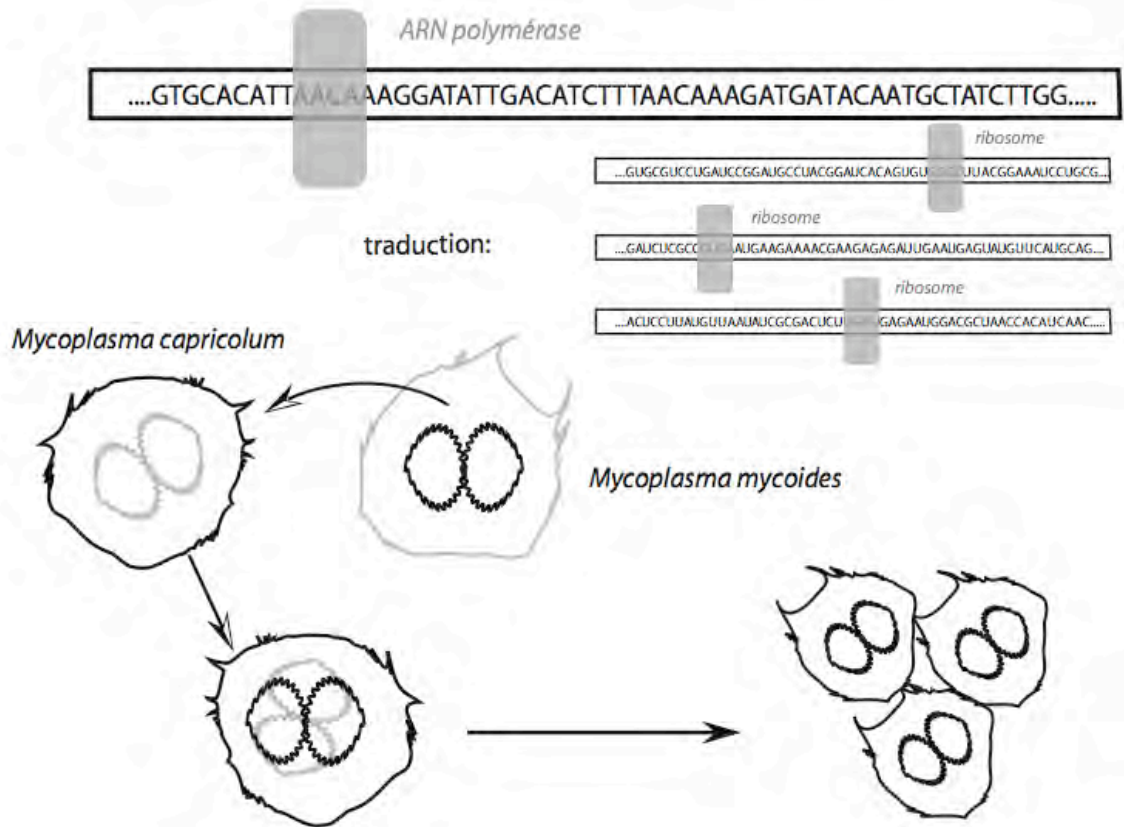


la démonstration de l'artigue-venter

la machine de turing

peut être parallélisée

transplantation d'un génome

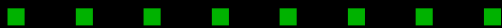


■ ■ ■ ■ ■ ■ ■ ■

genome transplantation in bacteria: changing one species to another
Lartigue C, Glass JI, Alperovich N, Pieper R, Parmar PP, Hutchison CA 3rd, Smith HO, Venter JC
Science (2007) 317: 632-638

u n o r d i n a t e u r g é n é t i q u e

- dans un ordinateur, la machine est séparée des données et du programme
- données et programme jouent le même rôle (càd on peut les considérer comme des ' déclarations ') ; ils peuvent être modifiés par la machine
- une réflexion générale due à alan turing et de nombreux autres penseurs considère les actions de la machine, mais pas la façon dont elle est construite

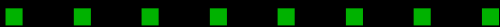


objection au modèle de la cellule-ordinateur

“En plus du programme génétique, la cellule comporte une quantité considérable d'information...”

Mais c'est exactement la même chose pour un ordinateur et son système d'exploitation !

Cette observation — valide — signifie simplement que notre réflexion sur le concept d'information est incomplète



dans un ordinateur, n'oublions pas le support physique



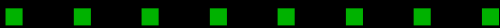
il ne suffit pas d'avoir une molécule d'ADN avec la bonne séquence, il doit encore être correctement replié !

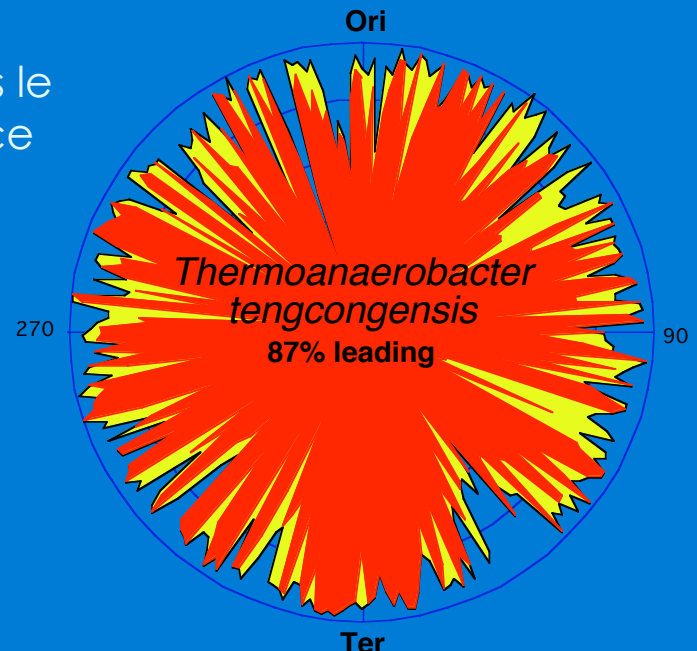
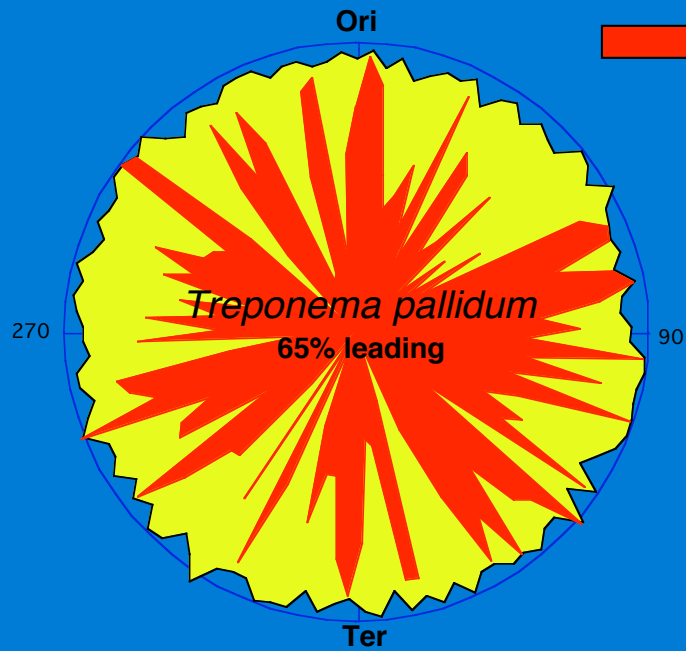
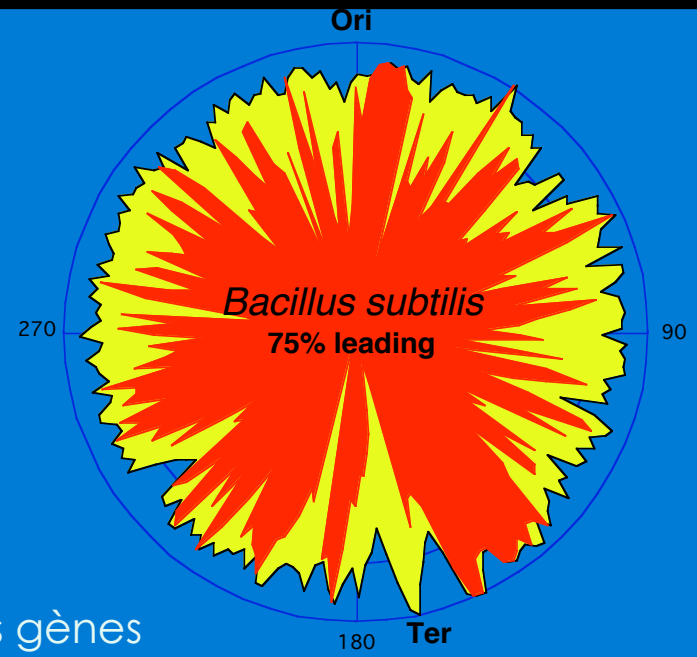
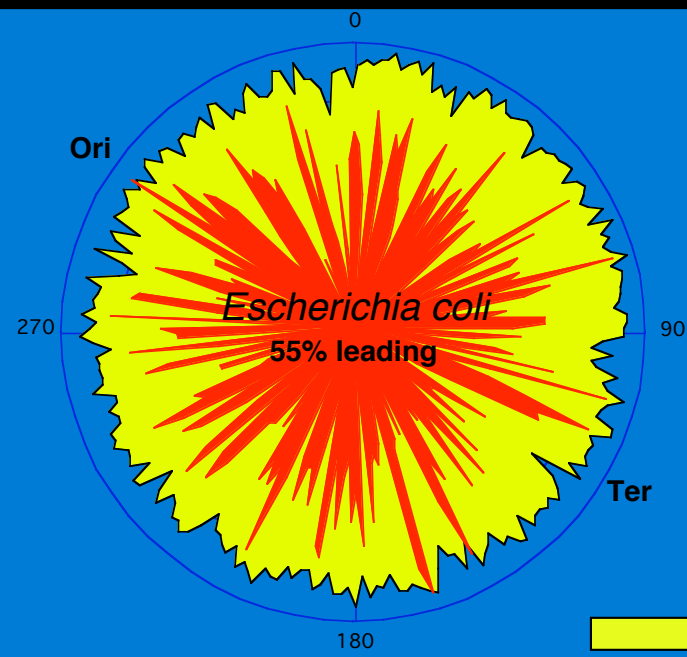
■ ■ ■ ■ ■ ■ ■ ■ ■ ■
Gibson DG, Benders GA, Axelrod KC, Zaveri J, Algire MA, Moodie M, Montague MG, Venter JC, Smith HO, Hutchison CA 3rd. One-step assembly in yeast of 25 overlapping ADN fragments to form a complete synthetic *Mycoplasma genitalium* genome. *Proc Natl Acad Sci U S A.* (2008) 105:20404-20409



organisation des génomes

l'ordre des gènes est-il aléatoire dans les chromosomes ?

- en plus d'une gestion différente de l'ADN dans les divers organismes les gènes semblent répartis plus au moins au hasard dans le génomes
- pourtant, des groupes de gènes comme les **opérons** ou les **îlots de pathogénicité** se regroupent, et codent des protéines dont les fonctions sont reliées

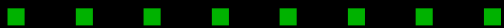




 densité des gènes
 densité des gènes dans le brin précoce

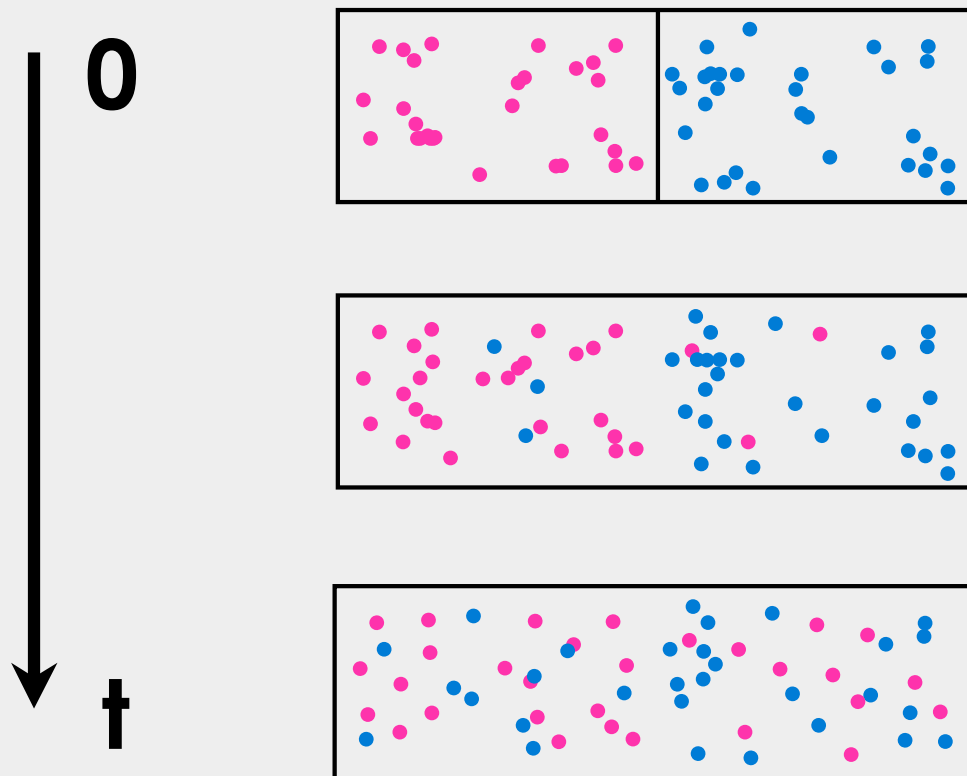
organisation des génomes

quelles forces physiques permettent la séparation des chromosomes, l'organisation des gènes, leur regroupement ?



une vue familière de l'entropie

l'évolution spontanée des particules isolées va de l'hétérogène vers l'homogène



$$S = k \log \Omega$$



Benjamin Crowell, licensed under the Creative Commons Attribution-ShareAlike license

analyses multifactorielles

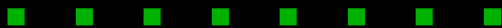
à l'opposé de la génétique habituelle, la génomique analyse de grandes collections de gènes et de produits de gènes

les analyses multivariées cherchent à extraire l'information en réduisant le plus possible les descripteurs des objets étudiés

l'analyse en composantes principales qui utilise la valeur centrée réduite des caractères étudiés est une méthode de base

l'analyse factorielle des correspondances est de la même famille, mais utilise la mesure du χ^2 comme distance ; cela permet non seulement d'étudier des objets dont les descripteurs sont hétérogènes, mais aussi de travailler sur l'espace des descripteurs et sur l'espace des objets simultanément

l'analyse en composantes indépendantes utilise le caractère non gaussien de la répartition des valeurs, etc



les biais de l'usage du code sont liés aux fonctions

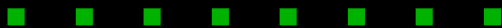
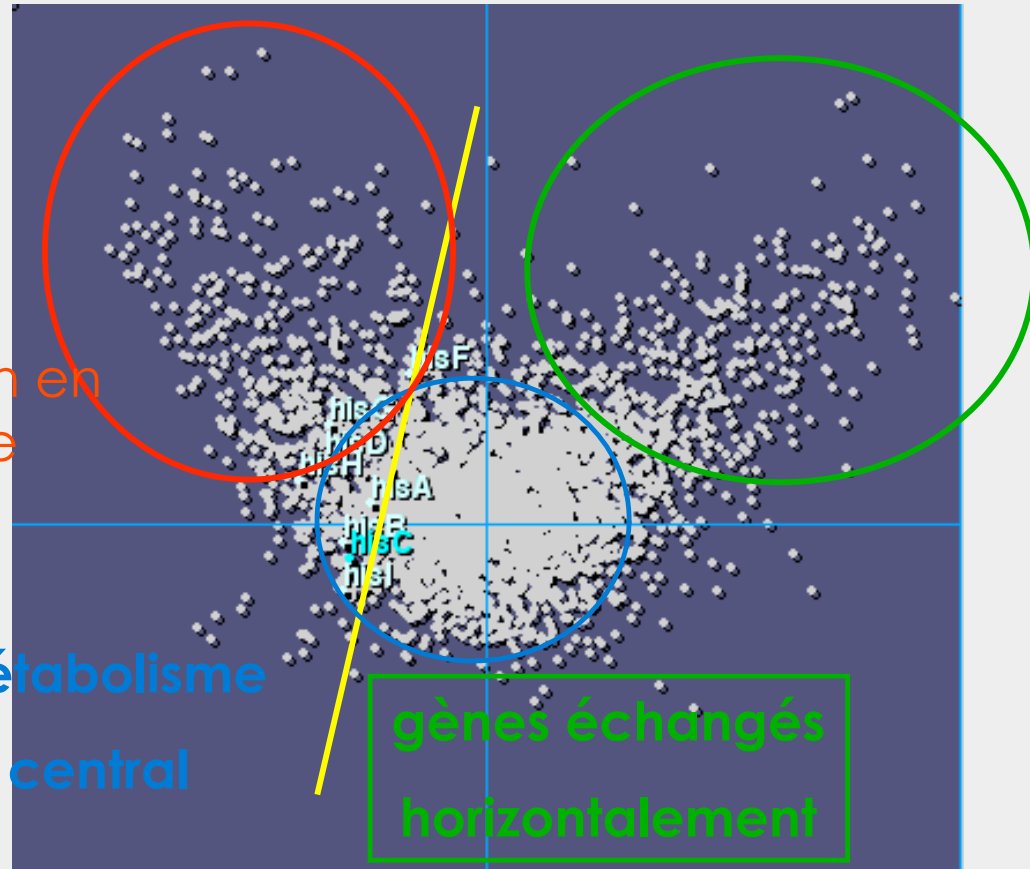
classe I :
métabolisme
central

classe II : forte expression en
croissance exponentielle

classe III :
transfert
horizontal

métabolisme
central

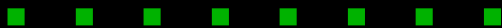
gènes échangés
horizontalement



la traduction organise le chromosome

l'analyse factorielle des correspondances montre que les gènes qui ont un biais similaire dans l'usage des codons sont apparentés fonctionnellement ; comment cela se répartit-il dans le chromosome ?

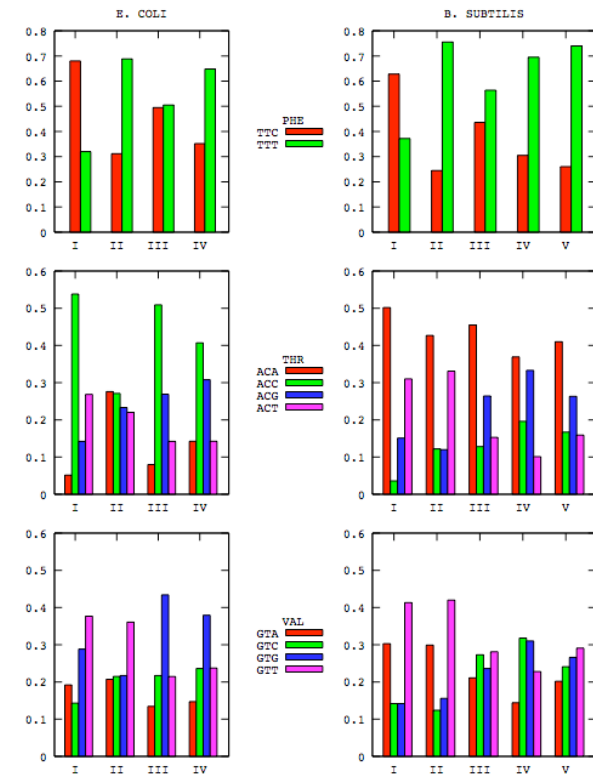
une méthode de groupage fondée sur l'analyse des biais d'**usage des codons** au moyen d'une théorie de l'information regroupe les gènes en familles homogènes, qui ne se répartissent pas au hasard dans le chromosome ; la méthode permet à la fois d'identifier des biais cohérents et de trouver le nombre pertinent de classes à considérer (4 pour *e. coli* et 5 pour *b. subtilis*)



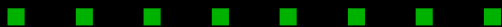
îlots génomiques

les gènes qui ont un biais similaire sont organisés en régions plus étendues que les opérons (30 kb), ce qui démontre un rôle de la traduction dans la structuration du chromosome bactérien

une part importante de la contribution à cet effet vient du recyclage des ARN de transfert rares



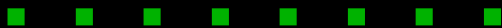
M. Bailly-Béchet



une image de la cellule dans le chromosome ?

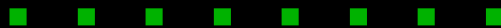
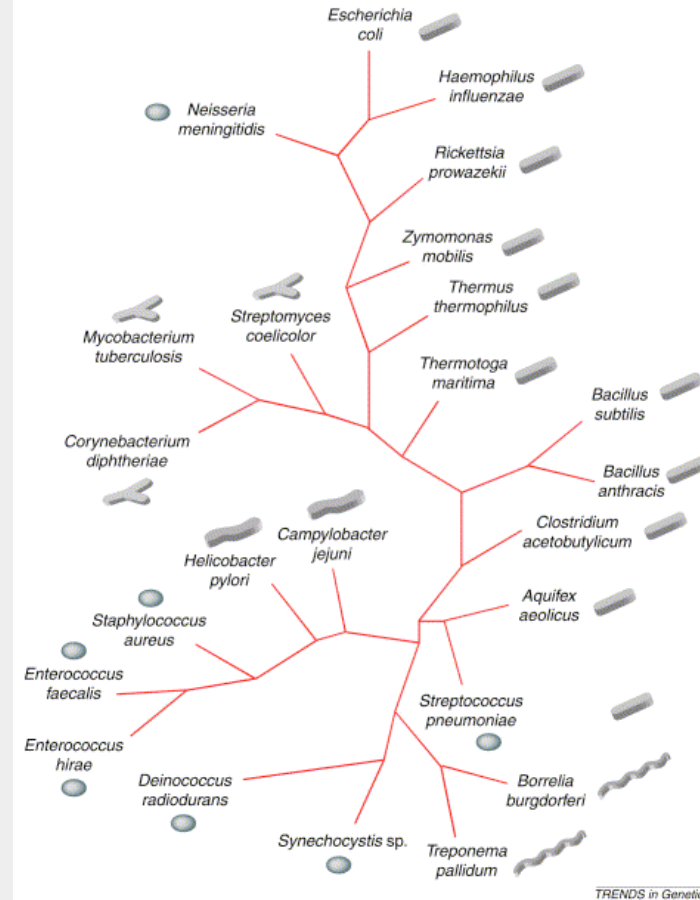
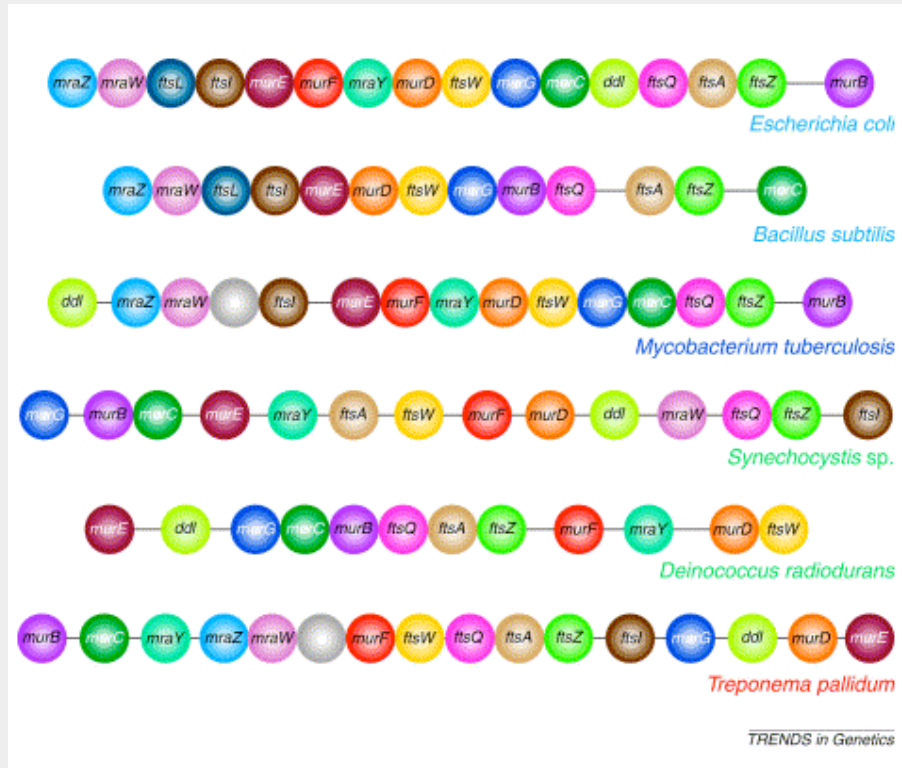
si la machine doit non seulement se comporter comme un ordinateur mais doit aussi construire la machine, il faut trouver quelque part une image de la machine (John von Neumann)

y a-t-il une image de la cellule (de l'organisme) dans le génome ?
l'ordre des gènes est-il aléatoire ?



ordre des gènes et forme de la cellule

les îlots *mur-fts*



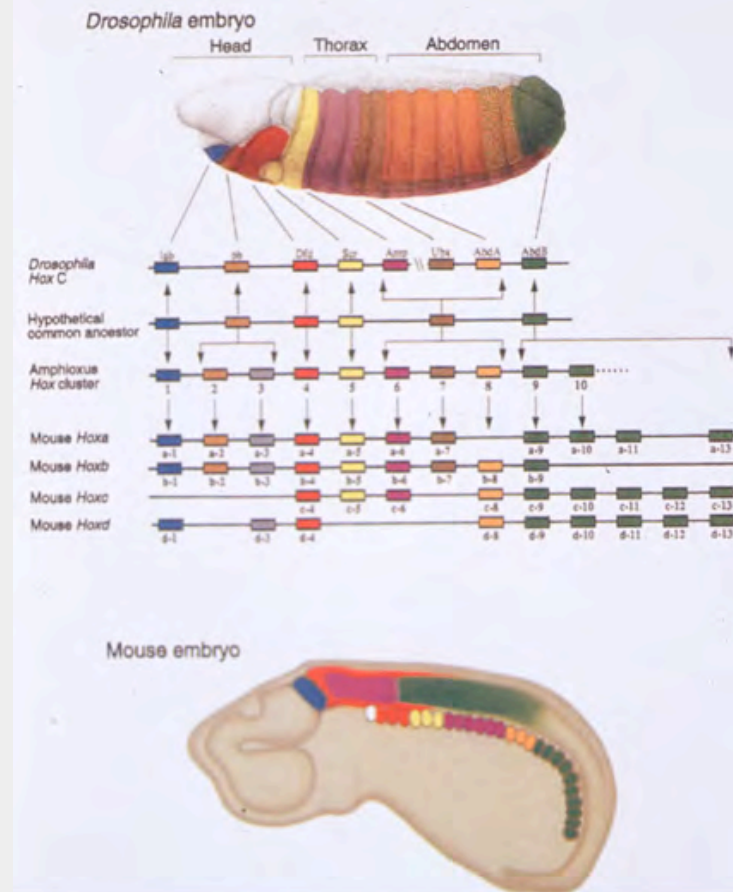
Tamames J, Gonzalez-Moreno M, Mingorance J, Valencia A, Vicente M
Bringing gene order into bacterial shape
Trends in génétiques (2001) 17: 124-126

les homéogènes

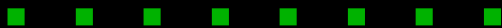
drosophilocus ?

homunculus ?

cellulocus ?

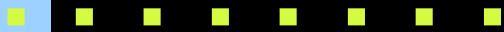


From Carroll, S. B. (1995) Nature 376, 479



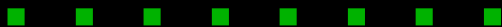


l'information

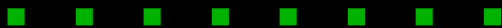
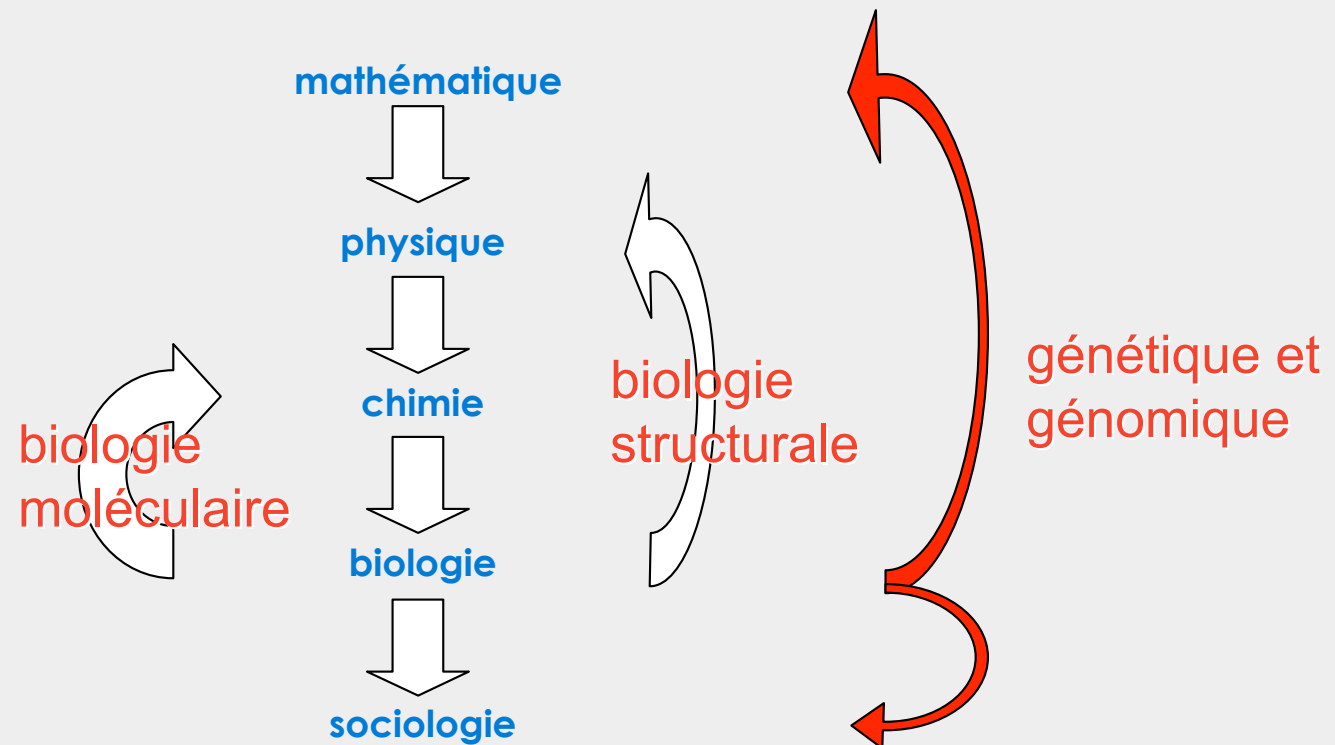


évolution et biologie synthétique

- ♣ la bs repose sur le modèle du programme génétique, où un **algorithme de construction** est **répliqué** au cours des générations
- ♣ une **machine (« chassis »)**, qui se **reproduit** au cours des générations, exécute l'algorithme, dont elle est physiquement séparée du support, comme dans les ordinateurs
- ♣ mais, si les ordinateurs font des ordinateurs, comment font-ils pour gérer le vieillissement ?
- ♣ l'analyse comparative des programmes génétiques met au jour des gènes (**gènes du démon de maxwell**) nécessaires pour **faire des organismes jeunes à partir d'organismes âgés** ; elle montre que les organismes vivants sont des **pièges à information**



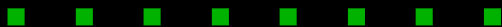
une idéologie sous-jacente



« complexité » : un terme à éviter

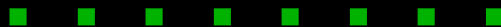
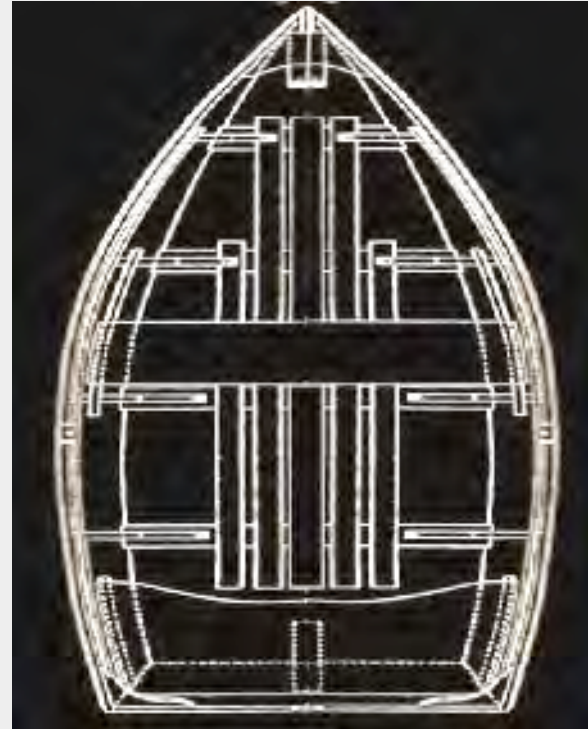
ce mot signifie en effet des choses diamétralement opposées

il signifie ou bien qu'un phénomène est d'un ordre très riche et très compliqué parce qu'il est *hautement organisé* mais comporte de multiples éléments (d'un nombre globalement supposé dépasser l'entendement), ou tout au contraire, parce qu'il est très embrouillé, très *désordonné*



l a b a r q u e d e d e l p h e s

- la biologie est une science des relations entre objets
- elle est **symplectique** (de $\sigma\upsilon\nu$ ensemble, $\pi\lambda\epsilon\kappa\tau\epsilon\iota\nu$, tisser), le même mot que « complexe » en Latin ; utilisé ici pour éviter les connotations malheureuses associées au mot « complexité » ; une connotation avec la géométrie n'aura pas d'inconvénient...
- c'est une information qui exprime ce qui est conservé dans la barque et non la matière de ses planches !



A. Danchin The Delphic Boat, Harvard University Press, 2003
La barque de Delphes, Odile Jacob, 1998

V. de Lorenzo, A. Danchin Synthetic Biology: discovering new worlds and new words 9: 822-827. EMBO Reports, 2008

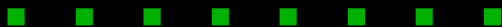
au delà des catégories usuelles

fait remarquable, la durée de l'exploration mentale varie de manière linéaire avec les distances réelles des points marqués par le sujet sur la carte, de la plage au cocotier, à la hutte, au trésor. la carte mentale contient donc la même information sur les distances que la carte réelle

jean-pierre changeux (1983) collège de france

des systèmes génétiques reconstruits ont été utilisés pour manipuler l'information, construire des matériaux, retraiter des produits chimiques, produire de l'énergie, produire de la nourriture, et aider à conserver et améliorer la santé humaine et notre environnement

drew endy (2005) mit



explorer l'information : l'infotaxie



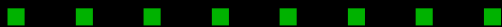
grand paon de nuit saturnia pyri
<http://pdubois.free.fr/>

comment un papillon peut-il trouver un partenaire 1 000 mètres plus loin ?

remonter un gradient chimique est impossible à une telle distance (turbulence de l'air, obstacles...)

vergassola et collègues ont montré que maximiser la collecte d'information permet d'atteindre cet objectif...

'Infotaxis' as a strategy for searching without gradients
Vergassola M, Villermaux E, Shraiman BI
Nature (2007) 445: 406-409



l'information, cinquième catégorie de la nature ?

physique classique

$$E = mc^2$$

énergie

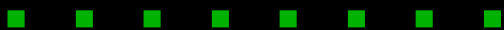
matière

espace/temps

physique quantique

$$\Delta x \Delta p \geq h/4\pi$$

indétermination = manque d'information

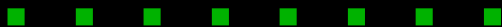


heuristique nouvelle

matière / énergie / espace / temps

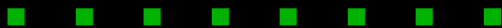
- physique classique
- physique quantique
- chimie
- biologie
 - développement
 - neurobiologie
 - linguistique
- mathématique (informatique)

information



l'information comme nouvelle catégorie du réel

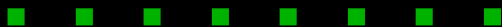
- 1929 **leo szilard**, analysant le démon de maxwell, a l'intuition (fausse) de la relation entre énergie et information: la création d'1 bit consomme $1/2 kT$
- 1949 **claudes shannon** crée une théorie de la communication
- 1961 **rolf landauer**, prouve que le calcul est réversible (aucune énergie n'est requise pour la création d'information) ; l'énergie est requise pour effacer la mémoire du processus de création
- ~1974 **andrey kolmogorov**, **gregory chaitin**, **ray solomonoff** définissent la complexité algorithmique
- 1988 **charles bennett** définit la profondeur logique (lien entre temps et complexité algorithmique) pour prendre en compte la valeur d'une information et illustre concrètement le théorème de landauer
- 1989 **wojciech zurek** lie la complexité algorithmique et l'énergie, en réfléchissant à ce que serait le démon de maxwell
- 2007 **scott muller** définit l'information comme attribut quelconque permettant de déterminer l'état d'un système, via une asymétrie
- 2009 **takahiro sagawa** et **masahito ueda** reprennent le théorème de landauer liant énergie et effacement de la mémoire



différents niveaux de l'information

- ce qui est vu par la réplication : aucune « signification », information de Shannon
- ce qui est vu par la machinerie de l'expression des gènes et par l'évolution
 - **complexité algorithmique (espace)**
 - **profondeur logique (temps)**
 - (profondeur critique) (compaction)

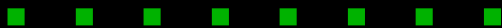
Danchin, A. 1996. On genomes and cosmologies. In *Integrative Approaches to Molecular Biology* (eds. J. Collado-Vides B. Magasanik, and T. Smith), pp. 91-111. The MIT Press, Cambridge (USA)



ce que voit la réplication

au cours de la réplication, la séquence de l'ADN est recopiée exactement ; la copie ne se soucie nullement de la signification de la séquence

l'enveloppe génétique (la collection des séquences recopiées) peut être étudiée à un premier niveau de l'information qui lui correspond ; typiquement on utilise à ce stade l'« information » ou « entropie » de shannon



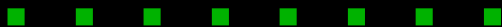
l'entropie de shannon (1)

une collection de messages (suite de symboles dans un alphabet fini) communiqués au travers d'un canal bruité a la forme probabiliste suivante :

$$H(p_i) = - \sum \{p_i \log_2 p_i \mid i \in I\}$$

cette formule est si générale que de nombreux modèles d'autres phénomènes que la communication de suite de symboles peuvent être ainsi représentés ; **danger des extrapolations** : la sphère est caractéristique des étoiles, des oranges, et des bulles de savon...

la **validité** de cette formule repose sur des hypothèses très fortes sur la nature des signaux en cause (en particulier ils doivent satisfaire aux lois des probabilité de laplace-gauss)

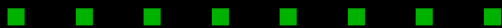


l' entropie de shannon (2)

mise en garde : myron tribus raconte que von neumann, à qui shannon avait demandé de l'aide pour trouver un nom pour sa fonction définissant l'information, lui proposa de façon prophétique :

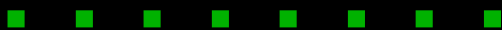
*“vous pouvez l'appeler **entropie** pour deux raisons ; d'abord votre fonction d'incertitude a bien été utilisée en mécanique statistique sous ce nom, ce qui fait qu'elle a déjà un nom ; ensuite, et c'est plus important, personne ne sait réellement ce qu'est l'entropie, ce qui fait que dans le moindre débat vous aurez l'avantage”*

créant ainsi la boîte de pandore de la plus grande confusion intellectuelle



complexité algorithmique

- l'entropie de shannon ne se soucie pas de la signification (réplication)
- kolmogorov et d'autres ont proposé de définir l'information par la **longueur du plus petit programme définissant exactement une suite de symboles** : la complexité algorithmique ; on a pensé que cela permet de définir le hasard: **ce qui est aléatoire est incompressible**
- connaissant une séquence de symboles, on ne peut en connaître la complexité... mais cela nous fournit un **programme de recherche** : afin d'approcher la complexité algorithmique d'une séquence, on s'intéresse à la façon dont elle a été engendrée (au cours de l'évolution, des synthèses, etc)

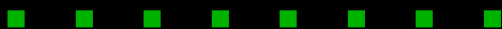


procaryotes et eucaryotes

→ **procaryotes** : une seule enveloppe, plus ou moins compliquée, génome superficiellement « aléatoire » ; superposition des signaux (combinatoire pour l'interaction avec des protéines impossible)

→ **eucaryotes** : multiplication des membranes et des peaux; génome riche en répétitions, juxtaposition des signaux (riche combinatoire des protéines qui contrôlent la transcription)

dans les deux cas cela semble inintéressant ; n'est-ce pas la marque d'une faille dans notre description ?

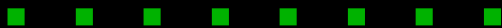


le contexte : "valeur" d'une information

la complexité algorithmique est insuffisante ; elle ne suffit nullement à définir le hasard ; tout dépend du **contexte** ; une séquence quelconque doit se trouver quelque part dans le développement des décimales de π : si l'on sait en donner la position, alors la complexité algorithmique de la séquence diminue !

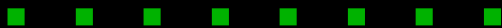
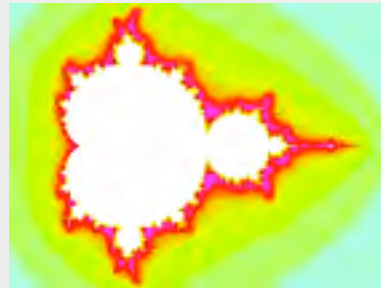
quel sens a **main sale**
pour un locuteur anglais, un locuteur français, un ordinateur ?

de même, les répétitions ont un sens, une valeur :
les poules du couvent couvent



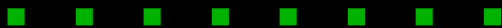
la profondeur logique (1)

un programme très court (**faible complexité algorithmique**) peut décrire une simple séquence répétée, mais aussi une figure fractale comme le flocon de koch ou l'ensemble de mandelbrot



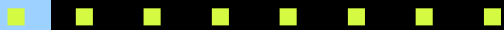
la profondeur logique (2)

- la différence entre ces deux cas de même complexité algorithmique concerne la **valeur** de l'information associée à la $n^{\text{ième}}$ étape
- les programmes récursifs impliquent qu'il est nécessaire d'exécuter le programme pour obtenir l'information
- bennett a nommé « profondeur logique » le temps nécessaire pour avoir accès à cette information



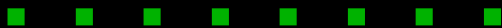


l'évolution au centre de la vie



vision phylogénétique du monde

- les programmes génétiques évoluent par mutations (parfois de simples réarrangements, sans changement du « texte » génomique)
- ceux qui ne sont pas totalement incapables de survivre et de se multiplier ont une descendance : survivre implique que les conditions locales de l'environnement **créent une information qui sculpte une image du monde** dans les génomes qui perdurent
- ce processus crée, via la mise à l'épreuve du fonctionnement concret de chaque organisme sa stabilité dans l'environnement où il est né

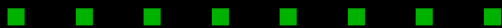


trois niveaux reliés sculptent une image du monde

trois processus se superposent pour constituer l'individu final :

- *phylogenèse* (evolution)
- *ontogenèse* (différentiation: les homéogènes)
- *epigenèse* (individuation)

chacun est stabilisé par la pression de sélection des événements qui agissent sur les fonctions en cause

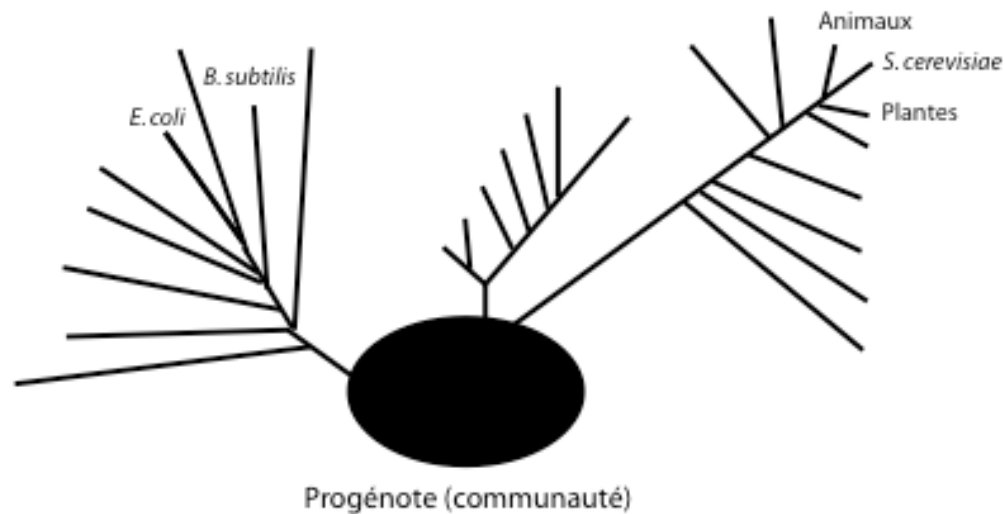


phylogénèse

Bactéries

Archées

Eucaryotes

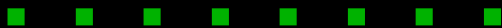


woese (1990)

kurland (2007)

tout organisme est le fruit d'une longue évolution, la profondeur logique implique qu'il n'y a pas d'ADN « poubelle » (« junk » DNA)

l'origine des fonctions est floue, elle distingue la machine du programme : ne soyons pas « adamistes » !



à la recherche des fonctions ubiquistes

variation / sélection / amplification

↪ stabilisation ↩

évolution



crée (l'information apparaît ici)

fonction



capture ("recrute")

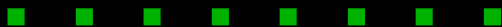
structure



code

séquence

l'ubiquité fonctionnelle n'implique pas l'ubiquité structurale



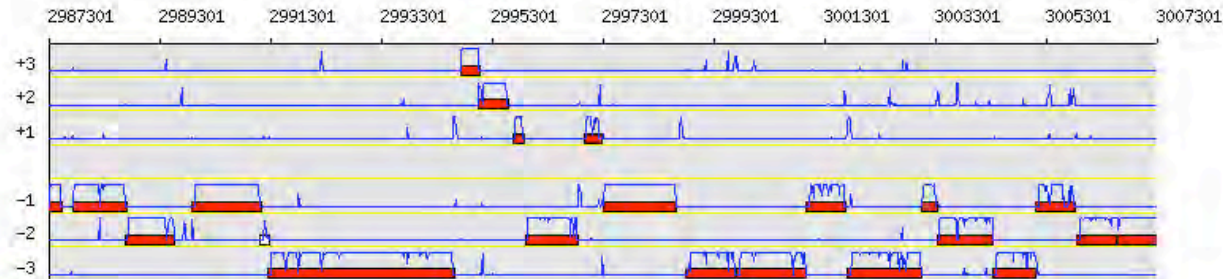
la pierre de rosette de la génomique comparative



Bacillus subtilis 168 chromosome BSU 843

2987301 -- 3007301

(sequence length : 4215606 bases)



▼ Synteny results



C. Médigue

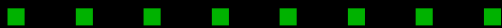
de l'ubiquité fonctionnelle à la persistance des gènes

l'ubiquité fonctionnelle n'implique pas l'ubiquité des gènes
les objets efficaces tendent à persister au cours des générations :

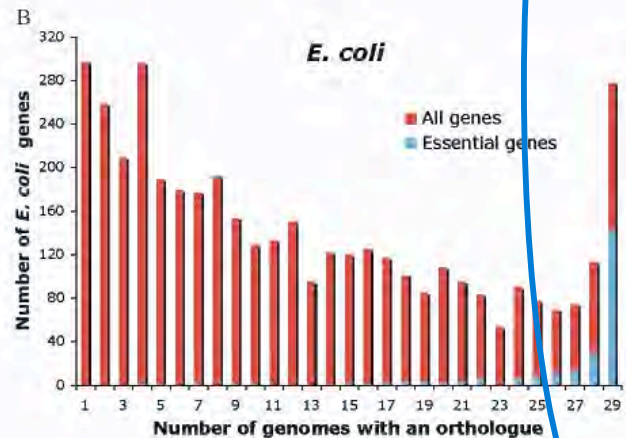
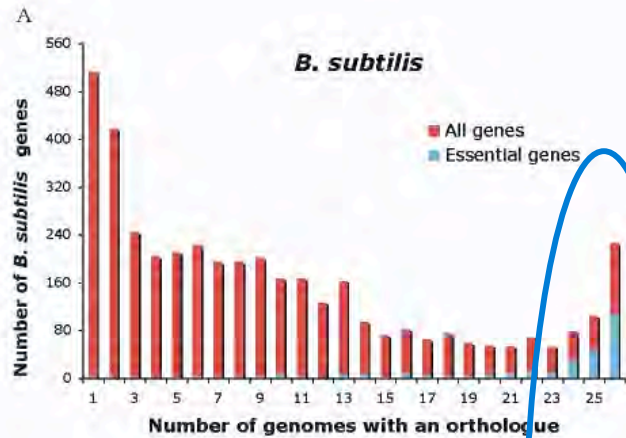
→ rechercher la « persistance » permet d'identifier la plupart des fonctions ubiquistes

→ « ubiquiste » est-il un synonyme d'« essentiel »? Les gènes « essentiels en laboratoire » sont situés dans le brin direct de l'ADN

~ 500 gènes persistent dans les génomes bactériens ; ils sont impliqués non seulement dans les trois processus nécessaires à la vie, mais aussi dans la **maintenance** et l'**adaptation** aux **phénomènes transitoires** ; une fraction gère l'**évolution** de l'organisme



persistance : des gènes en trop !



gènes persistants

gènes essentiels et

stress, maintenance et réparation

dégradation utilisant de l'énergie

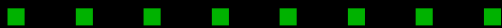
rustines métaboliques (effet sérine)

une origine minérale de la vie

la nécessité de sélectionner des molécules dans un mélange complexe, et de la polymérisation de ces molécules conduit à penser à une origine du métabolisme **à la surface de solides**. Ce scénario met l'accent sur la séparation du métabolisme de la réplication, cette dernière étant une invention **tardive** de l'évolution prébiotique :

maillons élémentaires => **nucléotides**

=> ARN^t => ribosome => ADN

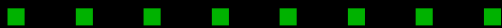


le monde des génomes à ARN

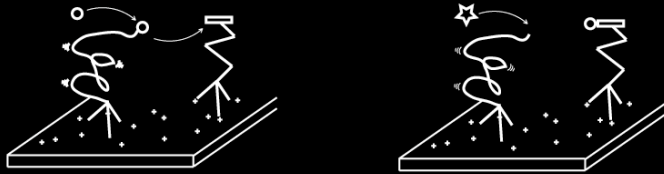
le "monde ARN", si populaire, est impossible sans la préexistence de **molécules remarquablement instables**, les nucléotides

les ARN ont deux rôles : ce sont des **substrats et des catalyseurs** de réactions chimiques, et ce sont des **matrices**, qui permettent leur propre réplication

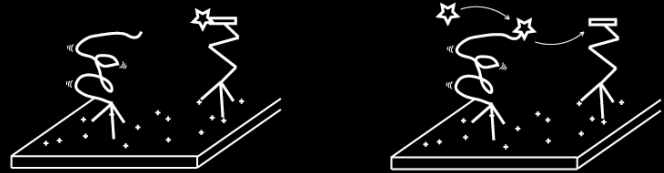
le vrai "monde ARN" est le "**monde des génomes à ARN**", qui apparaît avec leur fonction de matrice répliquatives



Variations sur un thème



Transformation homéotopique



Origin

Origine

ARNt hors de la traduction

Formation d'un lien peptidique

Gly~ARNTgly murein
 Phe~ARNTphe N-terminal
 Leu~ARNTLeu N-terminal
 Ara~ARNTAra Ubiquitine

Homéotopie

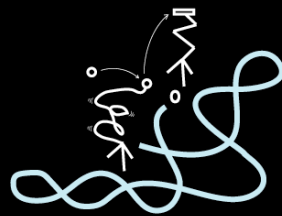
Met~ARNT^{FMet} → FMet~ARNT^{FMet}
 Glu~ARNT^{Gln} → GlnARNT^{Gln}
 Ser~ARNT^{SeCys} → SeCys~ARNT^{SeCys}

Lipides et coenzymes

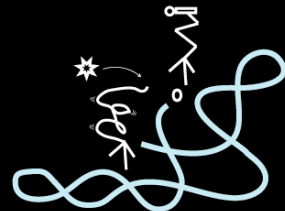
Lys~ARNT^{Lys} Lipids
 Glu~ARNT^{Glu} Aminolevullinate

Autres modifications

Base modifications



Les ARN de transfert
 se substituent aux surfaces
 minérales



Présence d'un groupement phosphate sans raison chimique

SERINE (SERINE~PHOSPHATE)
 PYRIDOXAL PHOSPHATE
 ISOPRENYL ~ PYROPHOSPHATE
 4-PHOSPHOPANTÉTÉINE
 THIAMINE ~ PYROPHOSPHATE

Réactions utilisant un nucléotide

CYTIDYLATE (LIPIDES)
 ADENYLATE, GUANYLATE, URIDYLATE (SUGARS)
 COENZYMES (NAD, CoA...)

Origin

Origine

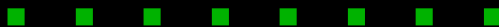
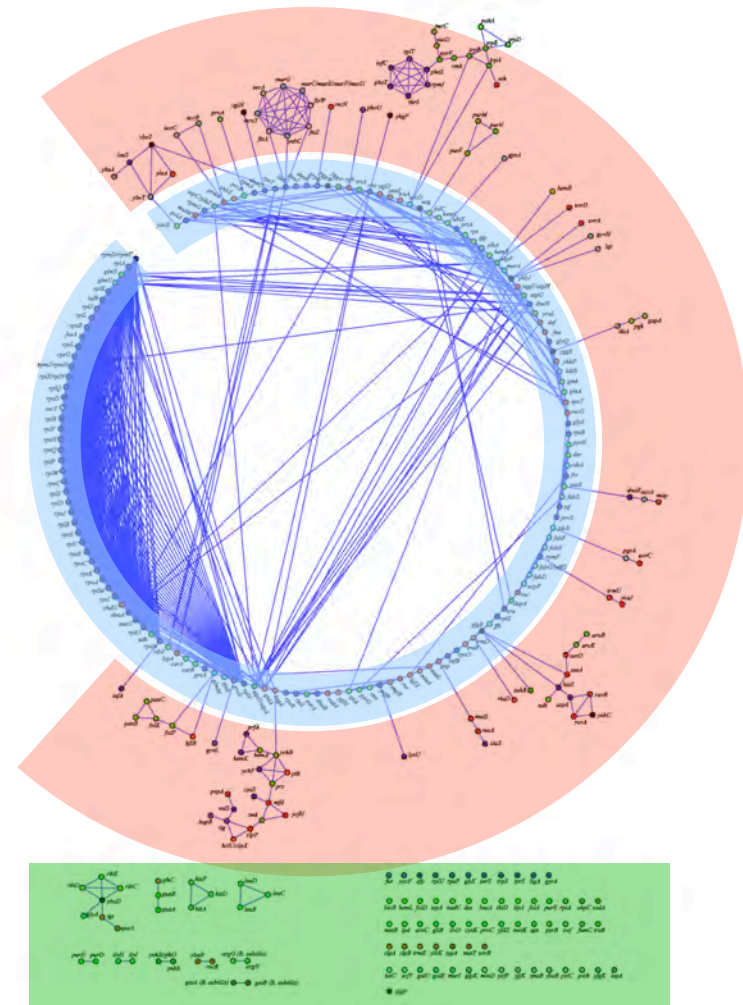
les gènes persistants récapitulent l'origine de la vie

Le **réseau externe**, composé de gènes du métabolisme (nucléotides et coenzymes, lipides), est très fragmenté ; le **réseau intermédiaire** est construit autour des ARNt synthétases ; et le **réseau central**, presque continu, organisé autour du ribosome, de la transcription et de la réplication, gère les transferts d'information

A Danchin, G Fang, S Noria

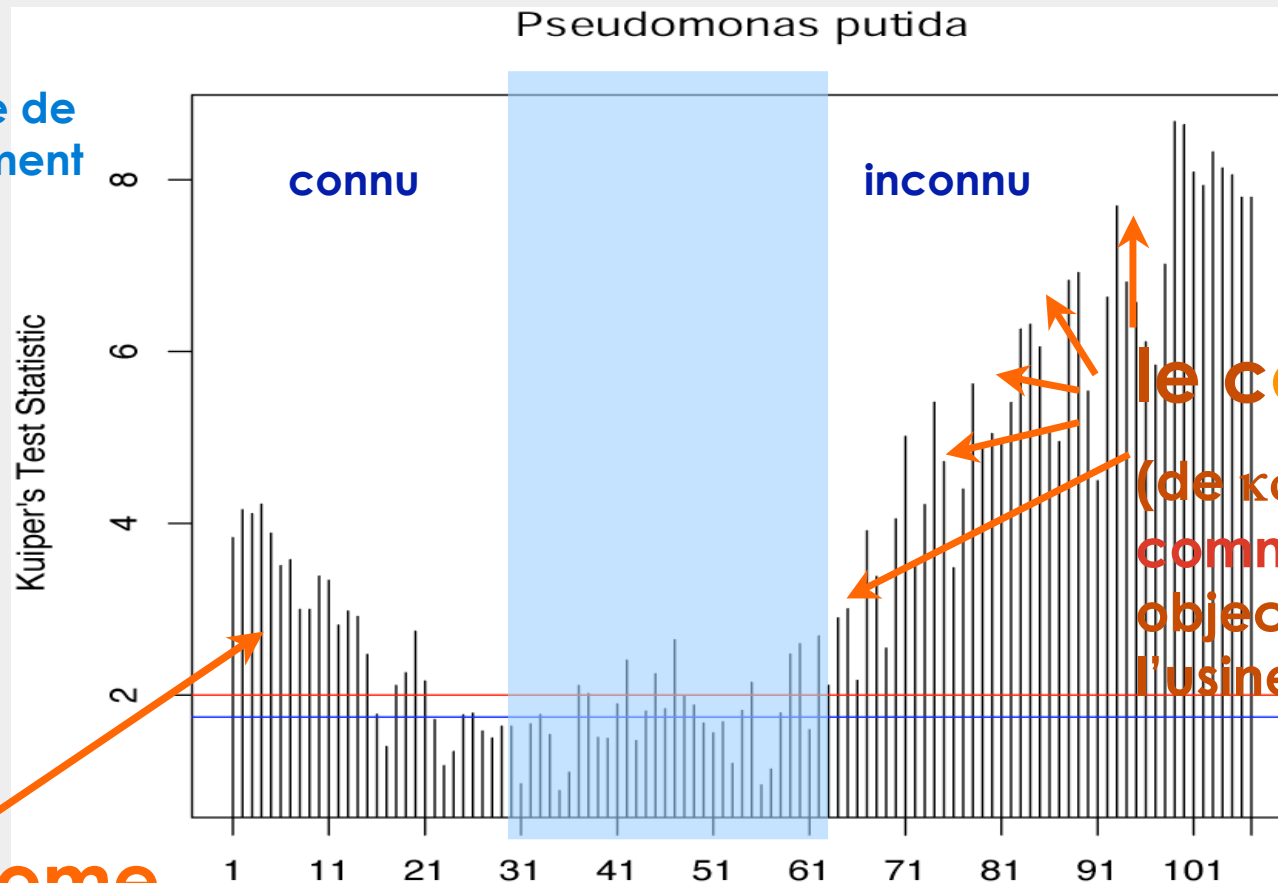
The extant core bacterial proteome is an archive of the origin of life

Proteomics. (2007) 7:875-889



organisation des génomes bactériens

fréquence de regroupement



le cénome

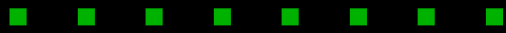
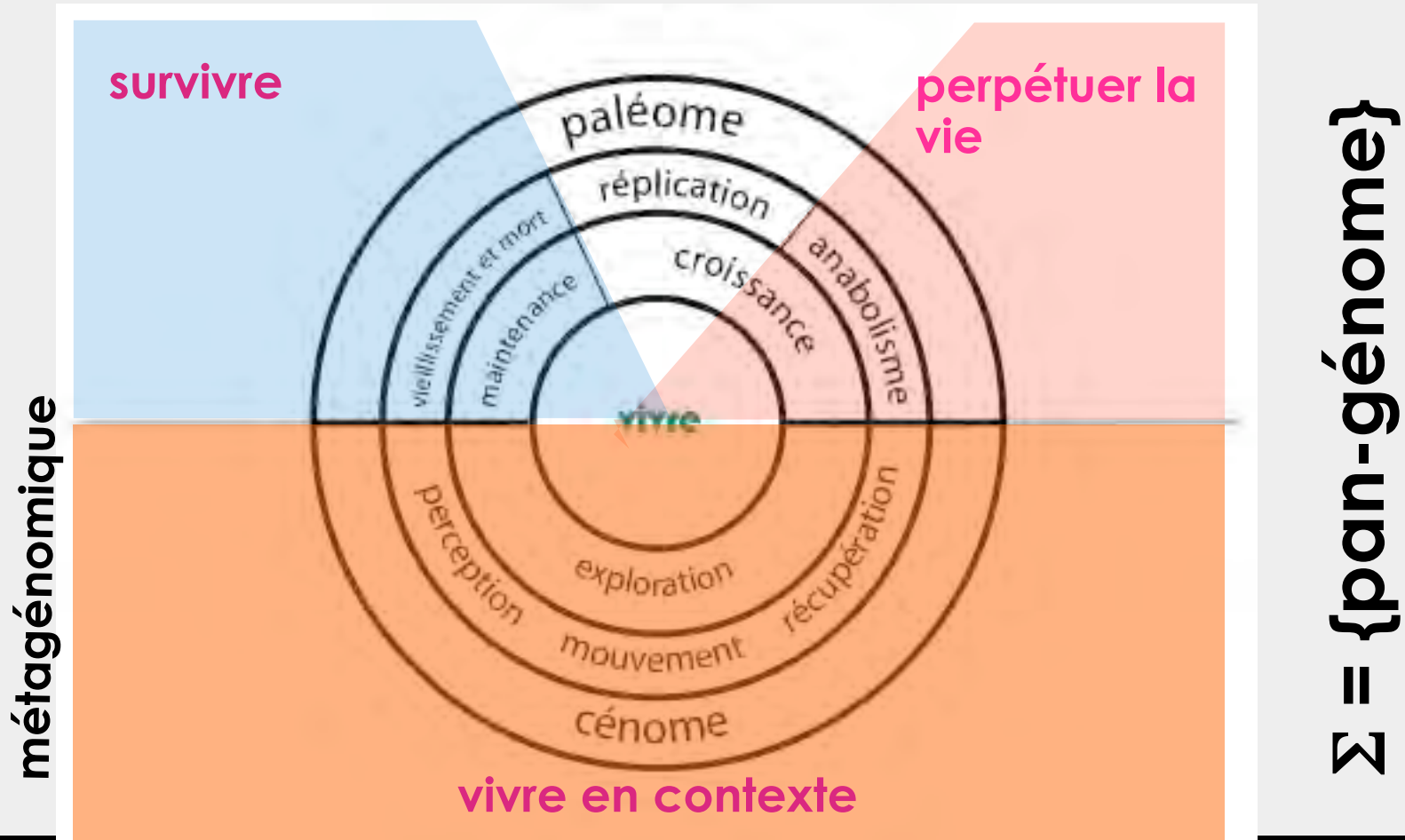
(de κοινος, commun): l'objectif de l'usine cellulaire

fréquence dans les génomes

le paléome

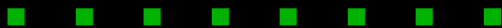
(de παλαιος, ancien): l'usine cellulaire

histoire de deux génomes



le paléome et le cénome

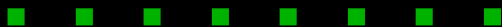
- la structure du paléome
 - fonctions essentielles ; évolution de la machinerie de l'expression des gènes
 - dégradation utilisant de l'énergie
 - métabolisme du soufre (anabolisme, récupération, catabolisme)
 - « frustration » chimique (« rustines » métaboliques)
- le cénome : occupation d'une niche écologique



origines de la vie

- le physicien freeman dyson a montré qu'il ne peut y avoir **une seule** origine à la vie
- la **réplication** conduit presque inmanquablement à une avalanche d'erreurs (sauf en cas d'échanges avec l'extérieur qui permet de revenir en arrière)
- la **reproduction**, qui concerne des relations entre objets, peut s'améliorer au cours du temps

la reproduction a donc dû précéder la réplication



les bébés naissent très jeunes !

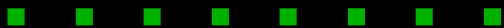
- **la machine se reproduit**

- la reproduction peut s'améliorer au cours du temps :
c'est toujours un organisme âgé qui donne naissance
à un jeune (cela implique création d'information)

- **le programme se réplique**

- la réplication accumule progressivement des erreurs

quels gènes permettent l'accumulation d'information ?

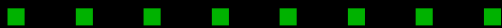


revisiter l'information

l'intuition dit que la création d'information demande de l'énergie

faux : la **création d'information est réversible** (landauer, 1961 ; bennett, 1982, 1988, zurek, 1989); accumuler l'information requiert un processus **énergivore pour "faire de la place"**

question ouverte : "faire de la place" est nécessaire pour accumuler l'information ; comment est-ce réalisé ? pouvons-nous identifier dans les génomes les gènes codant les fonctions qui mettent en œuvre ce processus ? pouvons-nous trouver une source ubiquiste et stable d'énergie ?



un paléome diversifié

paléome 1 (gènes essentiels)

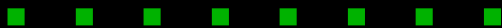
constructeur : l'ADN spécifie les protéines qui forment la machine qui construit la cellule (reproduction)

réplicateur : l'ADN spécifie les protéines qui répliquent l'ADN (réplication)

paléome 2 (gènes persistants non-essentiels)

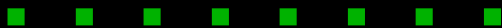
pérennisation de la vie, **dégradation dépendant de l'énergie**

rustines métaboliques (« frustration » chimique)



i n f o r m a t i o n

- les processus dégradatifs dépendant de l'énergie font de la place pour des entités nouvellement synthétisées ; de l'énergie est consommée pour **prévenir** la dégradation des entités fonctionnelles
- ce processus accumule de l'information, quelle que soit son origine, par effet de cliquet
- comme ce processus est ubiquiste, on attend que les fonctions correspondantes soient codées dans le paléome, y compris la gestion des sources d'énergie postulées ici

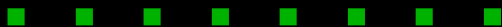


valeur de l'information

dans les modèles classiques de l'information on ne se soucie pas de la signification, ou de la **valeur** de l'information

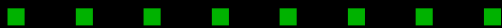
l'information du programme est transmise telle quelle par la réplication, sans valeur, mais d'où vient donc l'information de la machine ? L'idée est qu'il faut reconsidérer le concept dès qu'on a affaire à des objets individualisés :

peut-on **imaginer les gènes d'un démon de maxwell** qui ferai le tri entre ce qui est fonctionnel (localement) et ce qui ne l'est pas ?

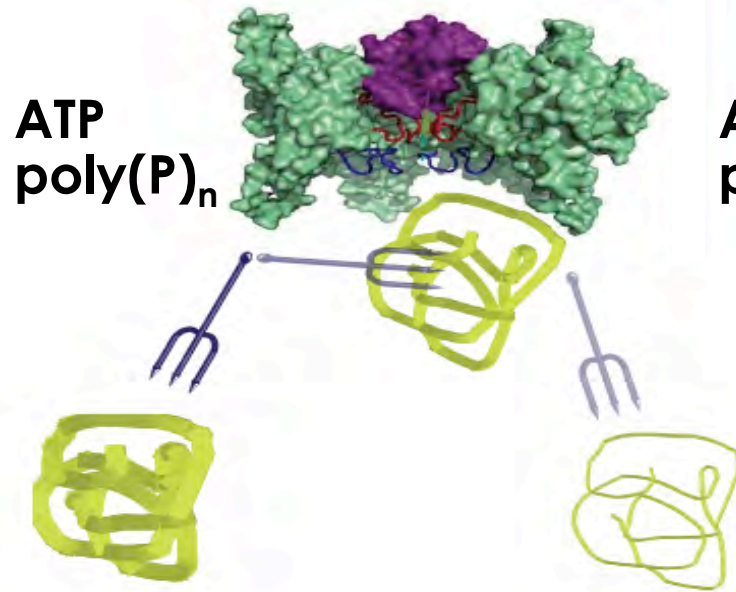


conjecture : les polyphosphates

- la synthèse et la dégradation des poly-P se retrouve codée dans les gènes non-essentiels persistants ; ce processus est encore peu connu et associé à la dégradation de l'ARN
- le poly-P est un **minéral**, donc extrêmement stable ; il est présent de façon ubiquiste dans les cellules
- les NTPs peuvent être régénérés à partir des NMP et de poly-P ; la protéase Lon peut utiliser le poly-P au lieu d'ATP ; le NADP (anabolisme) peut être engendré à partir de NAD et de poly-P...



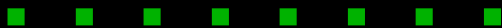
les gènes du démon de maxwell



la machinerie de dégradation
utilise de l'énergie pour rejeter
intacte une entité fonctionnelle

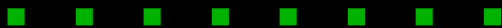


les entités non fonctionnelles
sont reconnues et dégradées



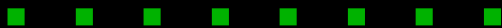
vivre et perpétuer la vie

- le paléome comprend à la fois les gènes permettant de construire l'usine cellulaire, et les gènes permettant de contrecarrer le vieillissement par l'accumulation d'une information nouvelle dans la descendance
- les processus dégradatifs dépendant de l'énergie font de la place pour des entités nouvellement synthétisées ; de l'énergie est consommée pour **prévenir** la dégradation des entités fonctionnelles
- ce processus accumule de l'information, quelle que soit son origine, par effet de cliquet
- ce processus est **myope**: il ne peut avoir un projet, d'où l'aspect "bricolé" de la vie et de son évolution



innovation : les mutations adaptatives

- l'accumulation d'information dépendant de l'énergie est aveugle ; elle ignore d'où l'information proviendra
- l'information peut provenir d'une mémoire, celle du génome pré-existant ; elle peut aussi être créée de novo
- les mutations adaptatives sont des créations de novo d'information ; elles dépendent donc des gènes impliqués dans l'accumulation d'information



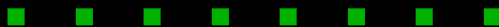
mutations adaptatives

construction de bactéries "intelligentes"

mises en croissance sur un milieu limitant en éléments nutritifs, elles forment des colonies de 10^7 bactéries environ. Le milieu contient aussi des nutriments **qu'elles ne savent pas utiliser**.

au bout de quelques semaines apparaissent des papilles qui se mettent à croître et qui utilisent les éléments ajoutés ; ce sont des **mutations adaptatives**.

elles ne préexistaient pas et supposent donc la **création d'information**



des mutations apparaissent au cours du vieillissement



Séquençage de 14 génomes + 30 PCRs

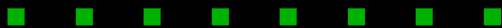
Le nombre total des mutations croît avec l'âge des colonies

Les mutations sont réparties dans tout le chromosome, et concentrées dans un gène => PCR de plusieurs colonies

Dans ce gène on trouve différentes mutations dans les différentes papilles, 2 mutations in 30% des cas

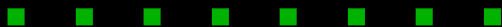
Dans certains cas une des deux mutations est silencieuse

Avec une source de carbone particulière, c'est au moins un autre gène qui est impliqué



prédictions diverses

- la persistance bactérienne dans un hôte dépend de gènes persistants non-essentiels
- l'initiation du cancer provient de cellules (souches) qui ont découvert des mutations adaptives leur permettant de créer une descendance immortelle
- l'accumulation d'information dans le cerveau (mémoire et apprentissage) implique des processus pour faire de la place tout en préservant les connexions fonctionnelles, d'une façon qui doit dépendre de l'énergie



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

in silico

gang fang, eduardo rocha, tingzhang wang

in vivo

agnieszka sekowska, evelyne turlin, andrew martens

collaborations

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

