The cell factories:
what microbes can bring to us

http://www.pasteur.fr/recherche/unites/REG

adanchin@pasteur.fr

Background

- **Physics: *matter, energy, time***
- **Biology: Physics + *information, coding, control...***
- **Arithmetics: *strings of whole numbers, recursivity, coding…***
- **Computing: Arithmetics + *program* + *machine...***

A metaphore with practical consequences, the genetic program: we can manipulate genes, and hence their products
=> sources of genes are essential

adanchin@pasteur.fr

What is Life?

Three processes are needed for Life:

- **Information transfer (Living Computers?) => the goal of genomics is to decipher the blueprint of the "read-only" memory of the machine**

Driving force for a coupling between the genome structure and the structure of the cell:

- **Metabolism (eating and digesting)**
- **Compartmentalization**

adanchin@pasteur.fr

The cell factory

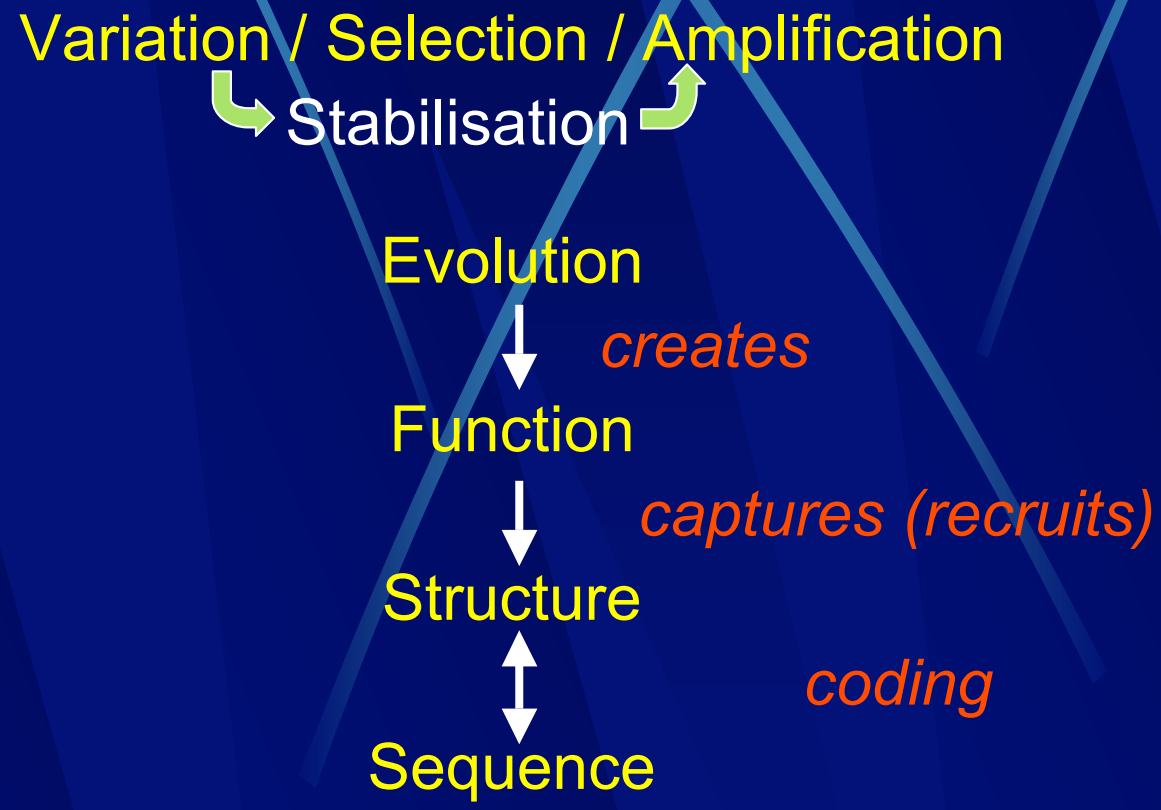
A cell behaves like a computer that would program the construction of similar computers

It has a magnetic tape, or hard disk (the « genetic program ») and reading devices which allow it to read the program and put it into action

The « cloning » of the ewe Dolly was exactly that: changing the program from a machine (an egg) to another one (an egg without a nucleus)

adanchin@pasteur.fr

The darwinian trio



adanchin@pasteur.fr

What functions for Life?
An extension of Cuvier's view….

- **Physical stability ([cyto]skeleton)**
- **Reproduction**
- **Respiration**
- **Locomotion**
- **Perception**
- **Transport (import / export)**
- **Circulation (internal fluxes)**
- **Digestion and recycling**
- **Assimilation**
- **Accommodation (regulation)**
- **Maintenance (repair)**
- **Etc…**

Microbial genome programs

1567 ongoing or finished programs, 294 completed, mostly of microbes

108,233,492,740 nucleotides at the International Nucleotide Sequence Database (INSD)

Microbes make 50% of the Earth's protoplasm

Still about 40-50% of the uncovered Coding DNA Sequences (CDSs) correspond to unknown functions; 10% only corresponds to the recognized core genome sequences (« persistent » genes)

adanchin@pasteur.fr

The first discovery of genomics: EEC genes

In 1991, at a EU meeting in Elounda, Greece, the presentation of the yeast chromosome III (André Goffeau) and 100 kb of the *B. subtilis* genome revealed that, contrary to expectation, at least half of the genes uncovered were totally unknown, whether in structure or in function

Because this was the result of a European collaboration, Piotr Slonimski proposed to call them Esoteric, Elusive, Conspicuous genes (EEC genes, in short...)

adanchin@pasteur.fr

From sequence to function

Combining genome sequence data and *in silico* prediction (bioinformatics) we test our hypotheses using large scale genomics techniques (transcriptome and proteome analysis) as well as other types of neighborhoods, such as common electric charge or codon usage bias

↓ Note that **regulation evolves much faster** than any other process: structural genes are the most important in the long term

adanchin@pasteur.fr

Concrete examples

- Microbial genes are of infinite diversity; only about 10% of those genes are of persistent and recognized function; we do not have yet a fair idea of the number of microbe species; the number of genes in a given species is highly variable (horizontal gene transfer)
- Example 1: growth under cold conditions
- Example 2: a new metabolic pathway
- ....

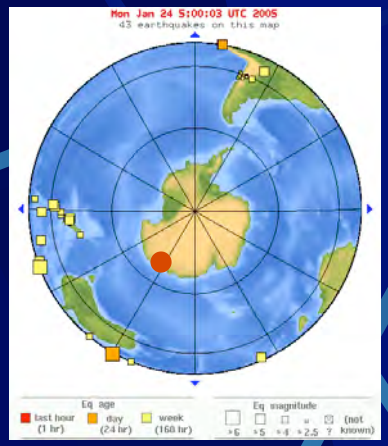
A. Sekowska

adanchin@pasteur.fr

Challenges posed by cold

- Protein folding (aging...)
- RNA folding
- Membrane fluidity
- Sensitivity towards oxygen and radicals
- ...

adanchin@pasteur.fr



INSTITUT PASTEUR

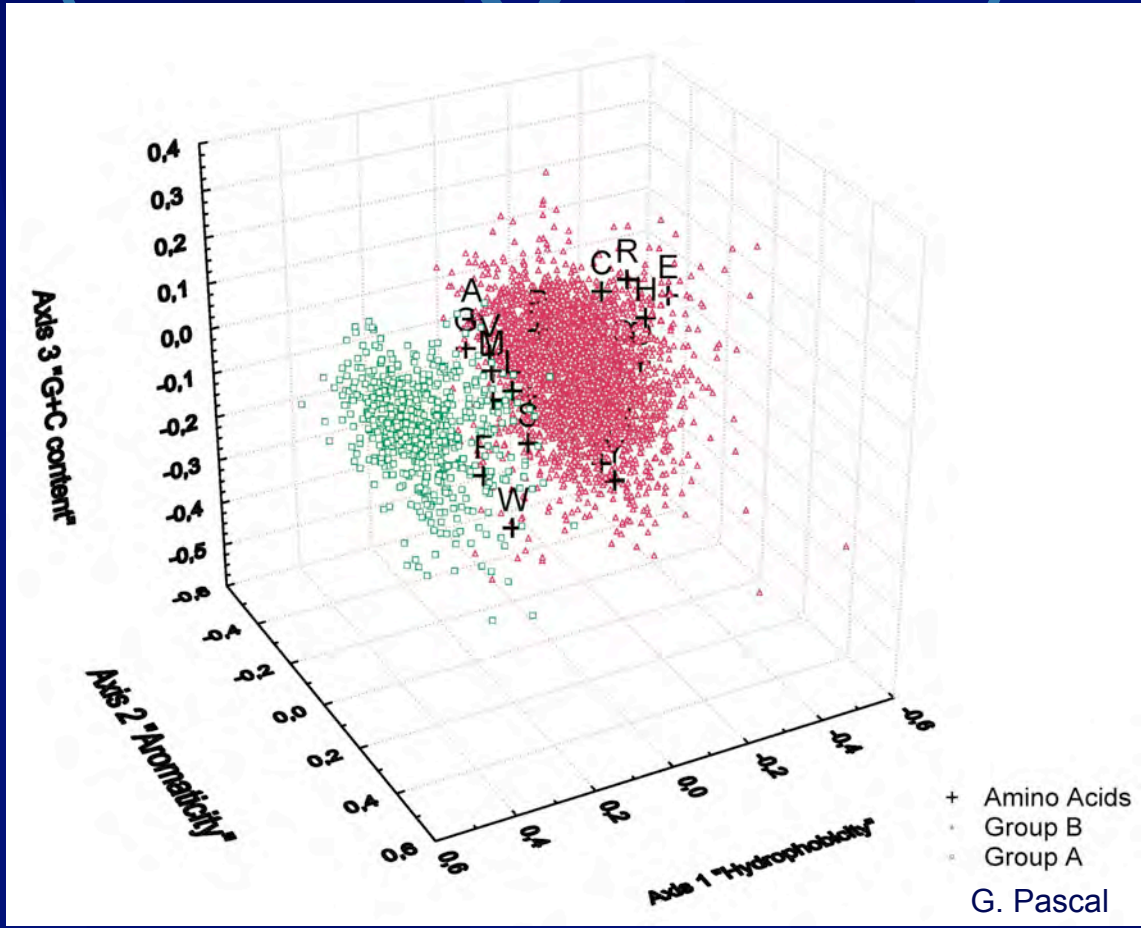
Pseudoalteromonas haloplanktis TAC125 has been isolated from an Antarctic costal sea water sample collected in the vicinity of the French Antarctic station Dumont d'Urville, Terre Adélie (66°40' S; 140° 01' E)

http://www.pasteur.fr/recherche/unites/REG

adanchin@pasteur.fr

Bias in amino acid distribution

Neighborhood:
distribution of aminoacids in the proteome
Integral Inner Membrane Proteins
IIMPs



G. Pascal

adanchin@pasteur.fr

Universal biases in amino acid composition

- **First axis:** separates Integral Inner Membrane Proteins (IIMP) from the rest; driven by opposition between charged and large hydrophobic residues

- **Second axis:** separates proteins according to an opposition driven by the G+C content of the *first* codon base

- **Third axis:** separates proteins by their content in aromatic amino acids; enriched in orphan proteins

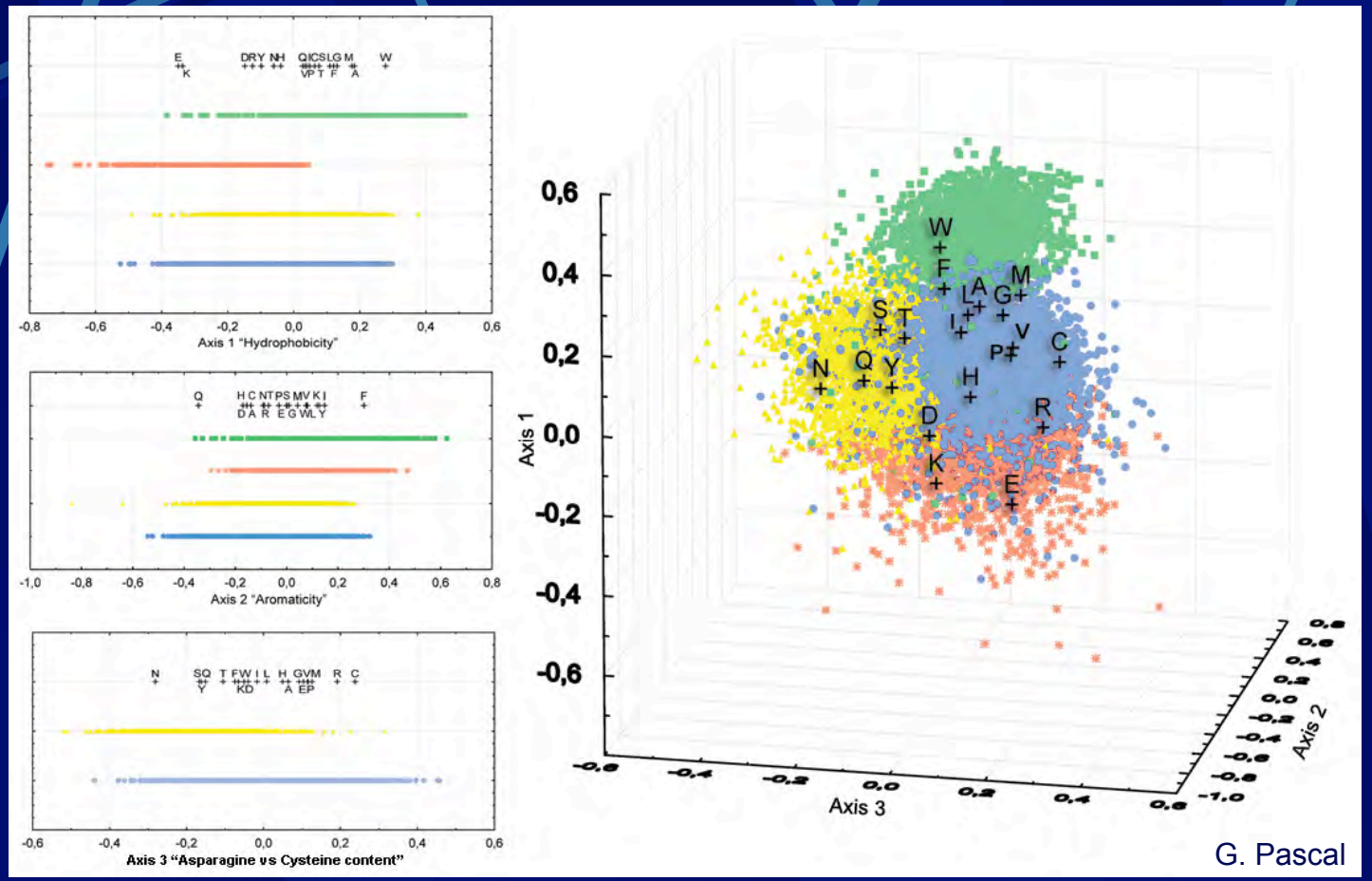
adanchin@pasteur.fr

Temperature-dependent biases in protein amino acid composition

- The amino acid composition of proteins depends heavily on the phylogeny => need to compare organisms related to each other
- The general trend of amino acid composition bias is to avoid some aminoacids at higher temperatures (associated to aging processes)
- Mesophilic bacteria belong to at least two different classes (in a 5-clusters analysis)
- Biases are always dominated by the IIMP clustering

adanchin@pasteur.fr

Temperature-dependent amino acid biases



G. Pascal

adanchin@pasteur.fr

Orphans: the gluons

A remarkable role of aromatic amino acids creates a universal bias. Expressed orphan proteins are enriched in these residues, suggesting that they might participate in a process of gain of function during evolution. We postulate that the majority is made of proteins — **gluons** — involved in stabilising complexes, thus defining the "self" of the species.

adanchin@pasteur.fr

From genes to metabolism

http://www.pasteur.fr/recherche/unites/REG

adanchin@pasteur.fr

Orthologs and Paralogs

Genes conserved throughout evolution often have orthologous function. Within cells, groups of conservation define paralogs. As a rule of thumb, one half of the genome is made of genes with no counterpart, one quarter, with two homologs, one eighth with three, etc... This « variation upon a theme » suggests that paralogs react with similar products, indicating that metabolism make errors which can be corrected using those paralogs.

adanchin@pasteur.fr

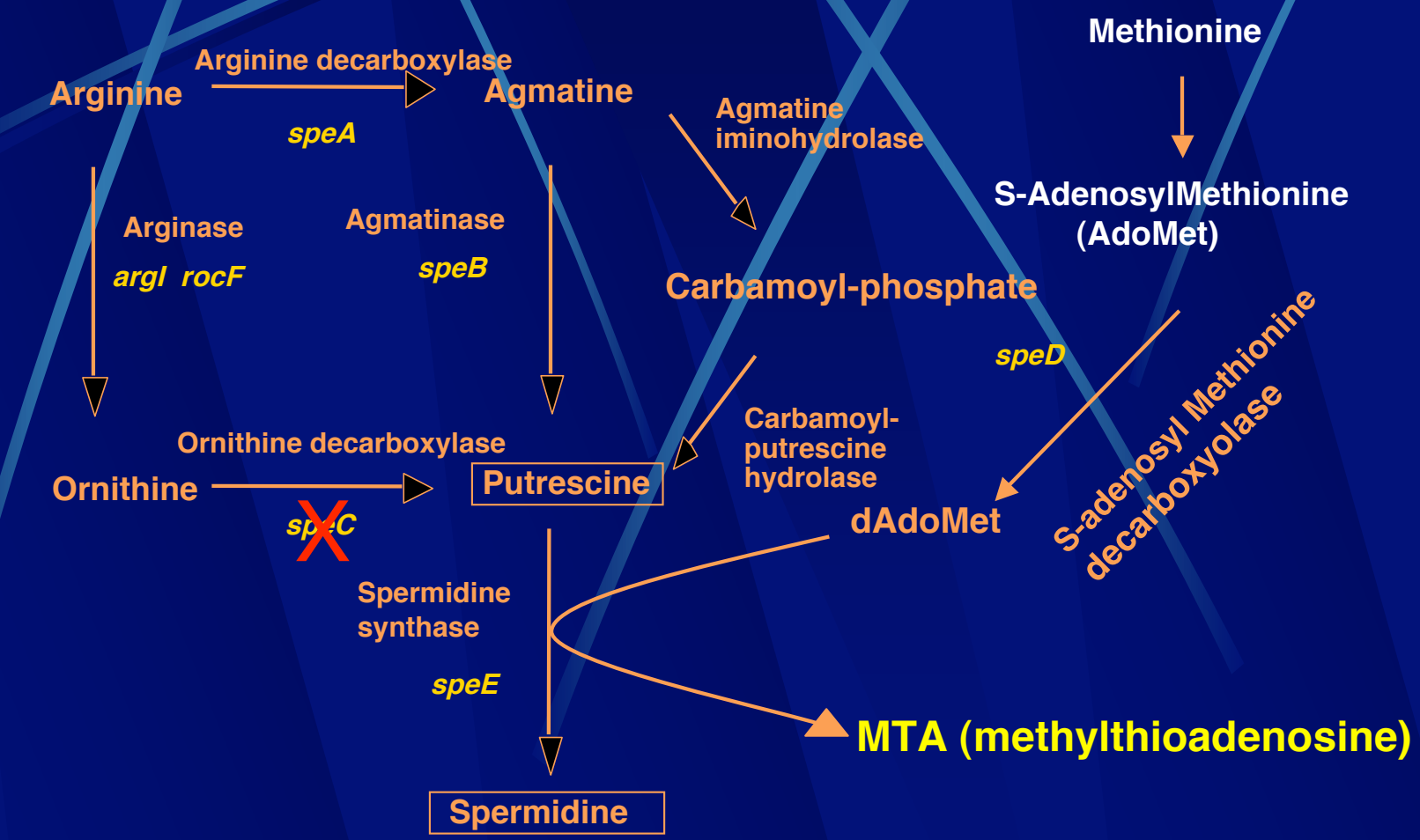
The error catastrophe

- Similarity in sequence leads to functional inference
- Because of recruitment of pre-existing structures, there is often no obvious link between a structure and a function (the book-paperweight)
- Hence a propagation of annotation errors
- *ykrS* annotated as « translation factor » is a component of sulfur metabolism!

A Sekowska, V Dénervaud, H Ashida, K Michoud, D Haas, A Yokota, A Danchin Bacterial variations on the methionine salvage pathway *BMC Microbiol* (2004) **4:** 9

adanchin@pasteur.fr

Forgotten molecules

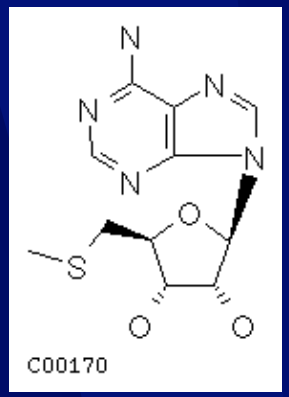


http://www.pasteur.fr/recherche/unites/REG

adanchin@pasteur.fr

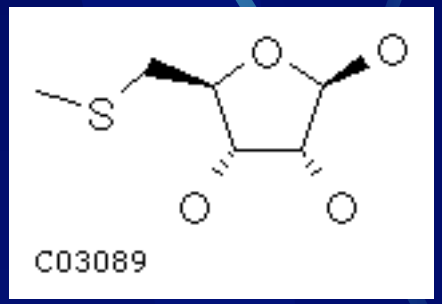
MTA degradation

MTA



mtnN (yrrU)

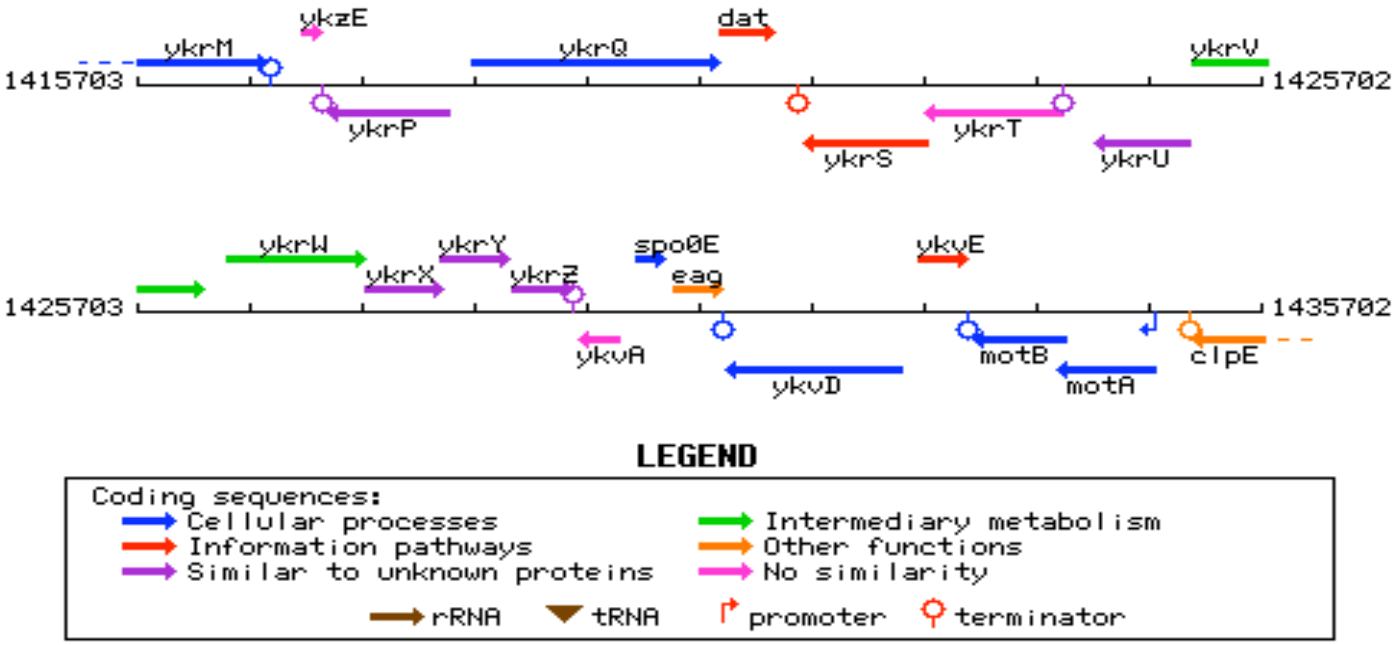
MTR



Expression profiling tells us what are the genes expressed in methylthioribose

adanchin@pasteur.fr

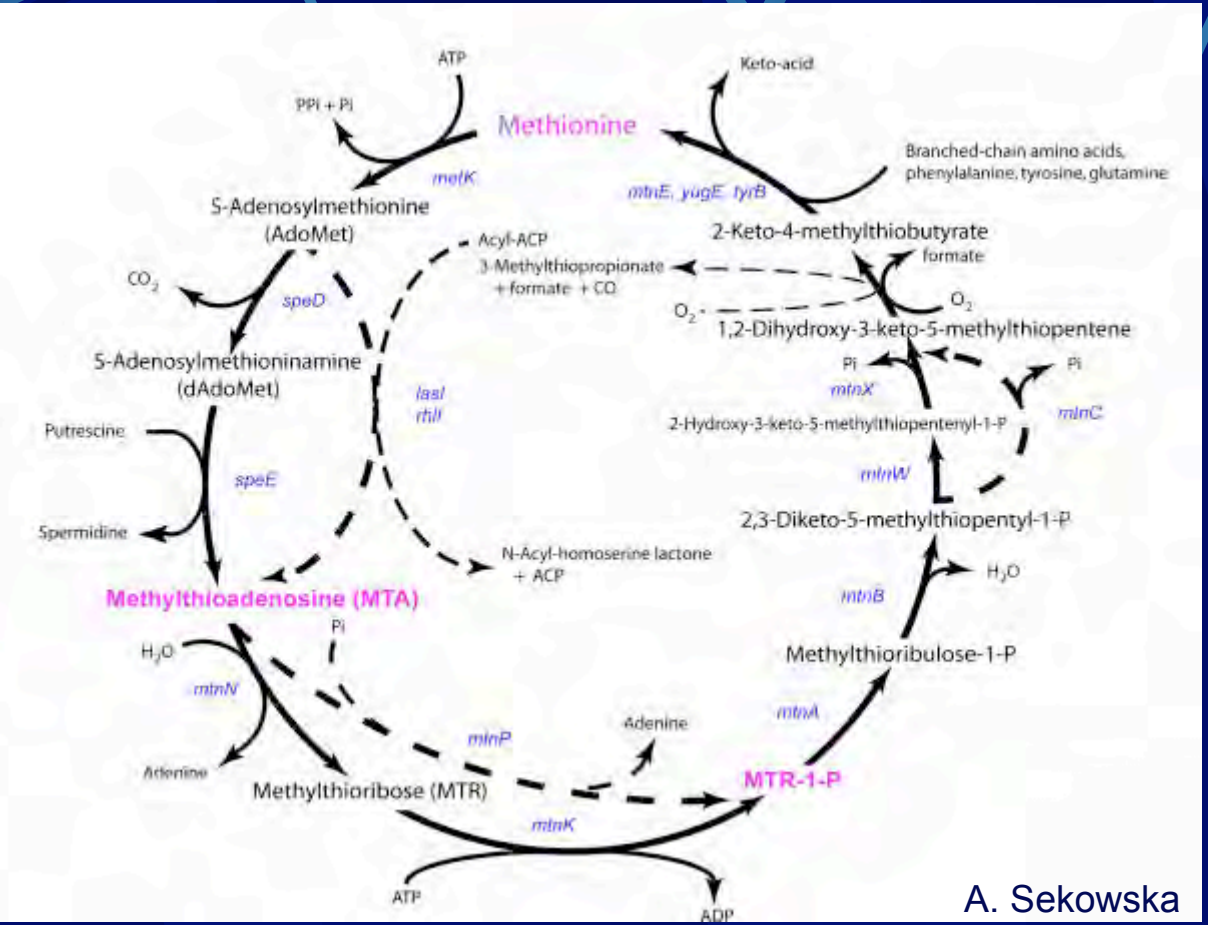
MTR Degradation



http://www.pasteur.fr/recherche/unites/REG

adanchin@pasteur.fr

A new metabolic pathway



A. Sekowska

adanchin@pasteur.fr

A new metabolic pathway

- The ancestor of RuBisCO, the most abundant, but one of the less efficient protein in the world...
- New unexpected intermediates in sulfur and carbohydrate chemistry
- A way to synthesize carbon monoxyde? a new gaseous mediator?
- ....

adanchin@pasteur.fr