Prédiction et validation des codons d'initiation :

approches croisées bioinformatique et protéomique

O. Lecompte, M. Argentini, C. Schaeffer, J-M Reyrat, O. Poch & A. Van Dorsselaer

- *Laboratoire de Bioinformatique et Génomique Intégratives, IGBMC, Strasbourg*
- *Unité de Pathogénie des Infections Systémiques, Paris*
- *Laboratoire de Spectrométrie de Masse Bio-organique, Strasbourg*

Bioinformatique et génomique intégratives

Etude de protéines informationnelles
Validation et analyse des données issues de la biologie à "haut-débit"

Développements algorithmiques

Plewniak *et al.* (2000) *Bioinformatics*.

Thompson *et al* (2000) *Nucl Acids Res.*

Thompson *et al*. (2003) *Bioinformatics.*

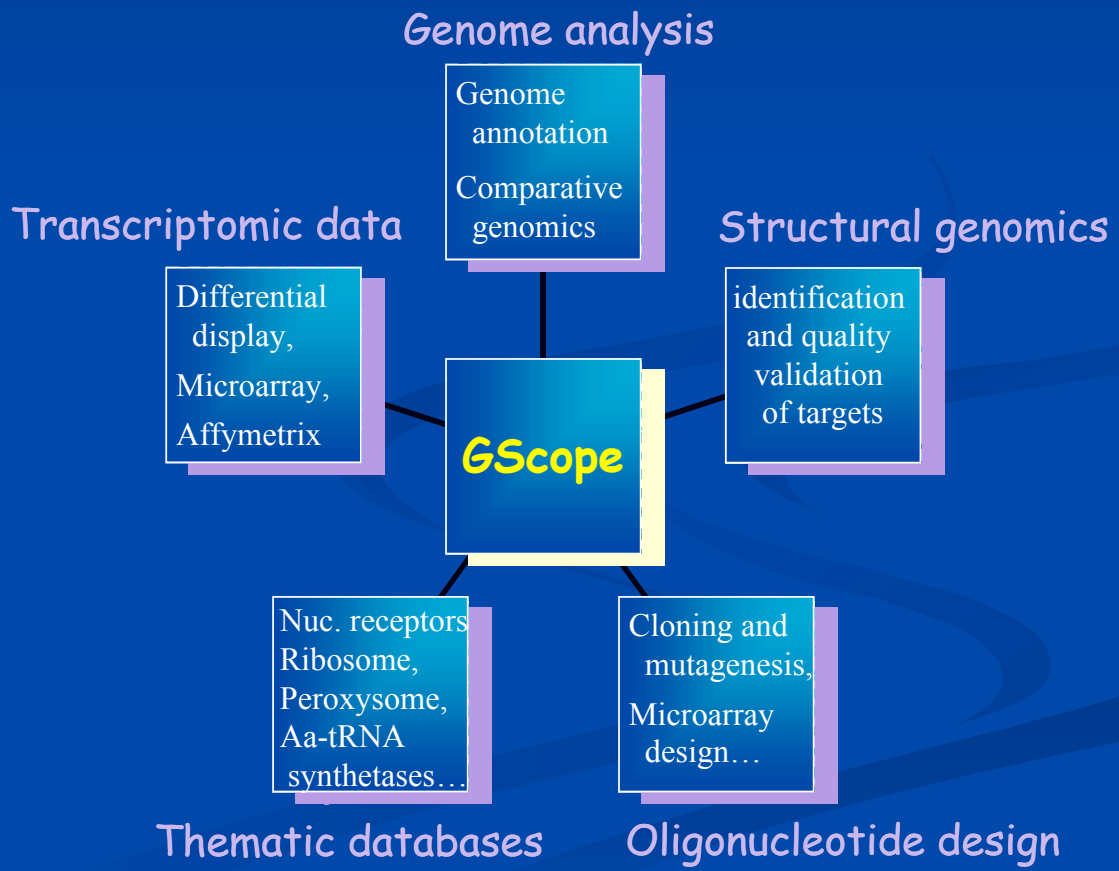
Thompson *et al* (2004) *Nucl Acids Res.*

Wicker *et al*. (2001) *Mol Biol Evol.*
Wicker *et al*. (2002) *Nucl Acids Res.*

PipeAlign
Plewniak *et al*, 2003

http://www-igbmc.u-strasbg.fr/PipeAlign/

Plate-forme logicielle GScope



Contexte du projet

biologie à haut débit → Accumulation d'erreurs

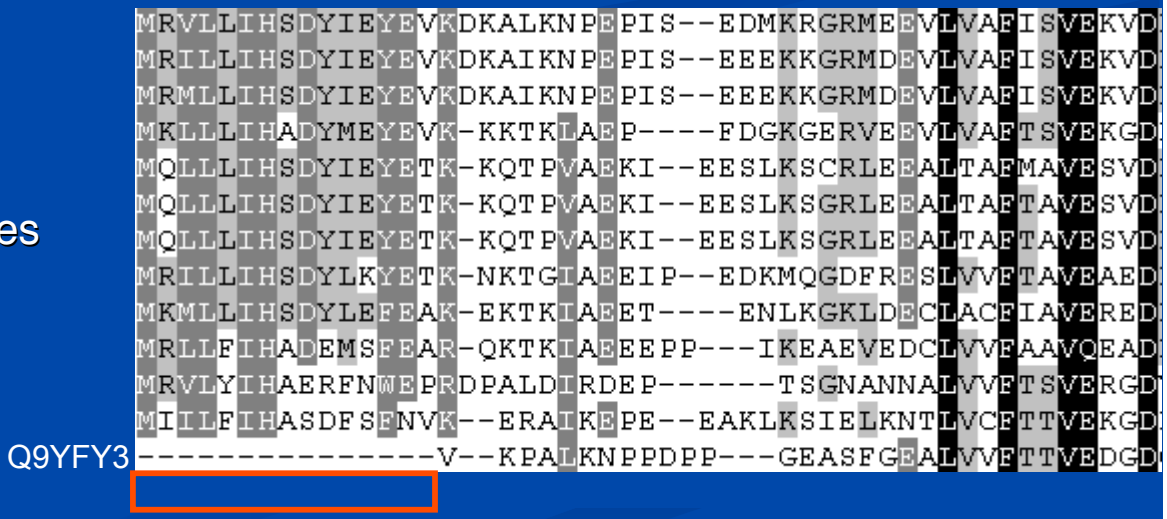
Erreurs de prédiction des gènes :

- 50 % des gènes de *Caenorhabditis elegans* (Reboul *et al.*, 2003)
- ~ 20 % des codons initiateurs prédits chez les bactéries

Conséquences :

- L'analyse des génomes *in silico*
  - régions régulatrices des gènes
  - signaux de localisation…
- Études expérimentales
  - Clonage et expression des gènes…

Thréonyl-ARNt synthétases

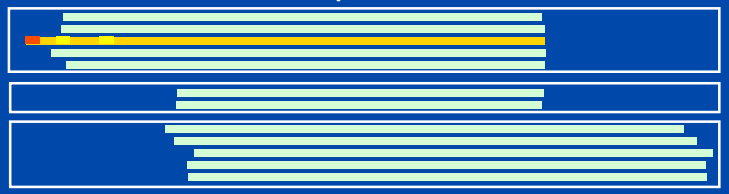


Validation/correction des codons initiateurs

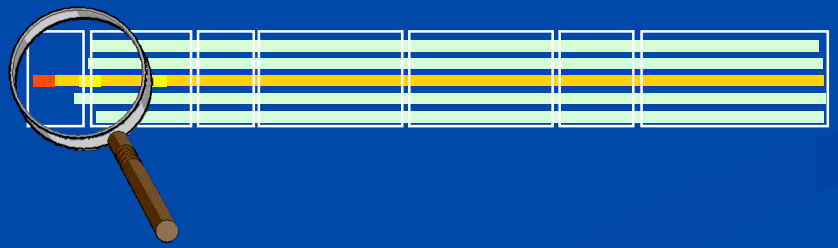
Conservation des séquences protéiques

Alignement multiple de séquences complètes

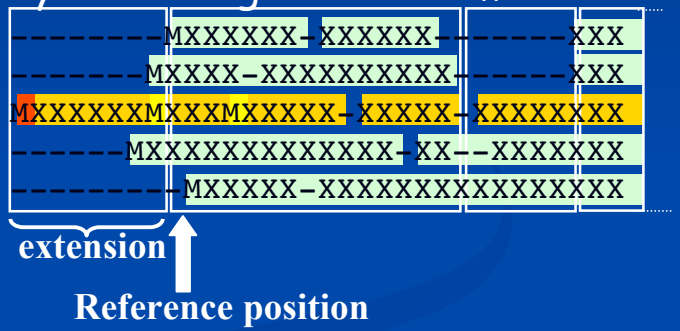
Classification des séquences



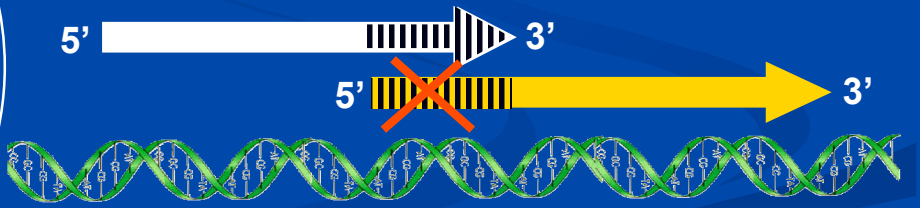
Clustering hiérarchique des positions
Définition de blocs



Analyse des régions N-terminales

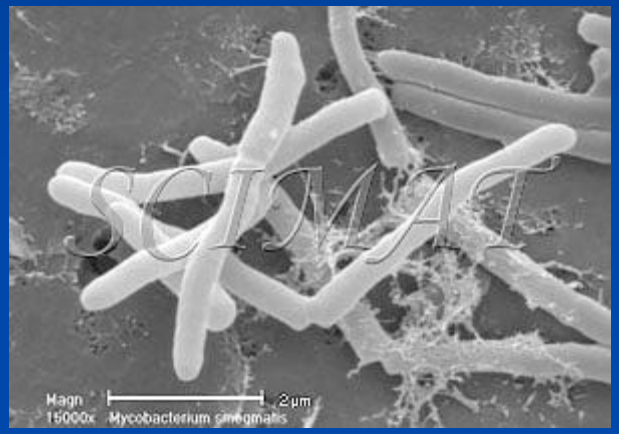


Etude du contexte génomique



Tests et optimisation

Mycobacterium smegmatis

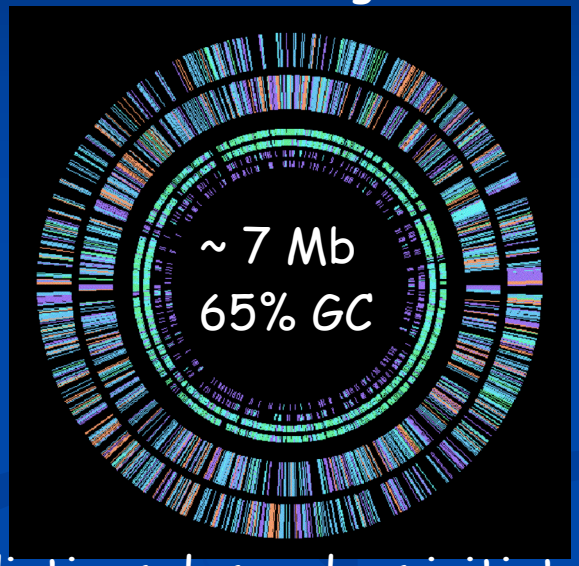


croissance rapide

Proche de bactéries pathogènes

- *M. tuberculosis*
- *M. leprae*
- *M. ulcerans*

Annotation du génome



prédictions des codons initiateurs

Détermination expérimentale des séquences N-terminales

DETERMINATION DES SEQUENCES N-TERMINALES DES PROTEINES DE *Mycobacterium Smegmatis*

5000-10000 PROTEINES EXPRIMEES

DIFFERENTES CULTURES BACTERIENNES

- PLANCTONIQUE SOUS AGITATION
- STATIQUE EN BIOFILM
- EN CONDITION DE STRESS

DIFFERENTES FRACTIONS PROTEIQUES

- PROTEINES SOLUBLES
- PROTEINES MEMBRANAIRES

(Unité de Pathogénie des Infections Systémiques, UMR570, Responsable J.M. Reyrat)

STRATEGIES D'ANALYSE DES SOUS-PROTEOMES DE *Mycobacterium Smegmatis*

APPROCHE CLASSIQUE

- SEPARATION DES PROTEINES SOLUBLES ET MEMBRANAIRES PAR GEL DE POLYACRILAMIDE
- DIGESTION ENZYMATIQUE « IN GEL »
- ANALYSE SPECTROMETRIQUES MALDI ET/OU LC-MS-MS

RAPIDITE ET SIMPLICITE TECHNIQUE

PERTE SIGNIFICATIVE D'INFORMATIONS LIEE A L'UTILISATION DES GELS

OBTENTION D'UN NOMBRE ELEVE ET REDONDANT DE PEPTIDES POUR CHAQUE PROTEINE ANALYSEE (5-10 peptides = 1 proteine)

BUT: CONNAITRE LES SEQUENCES DES PEPTIDES N-TERMINAUX DE L'ENSEMBLE DES PROTEINES BACTERIENNES

STRATEGIES D'ANALYSE DES SOUS-PROTEOMES DE *Mycobacterium Smegmatis*

APPROCHE COFRADIC
(Combined FRActional Diagonal Chomatography)

Nature biotechnology - volume 21 - may 2003

Exploring proteomes and analyzing protein processing by mass spectrometric identification of sorted N-terminal peptides

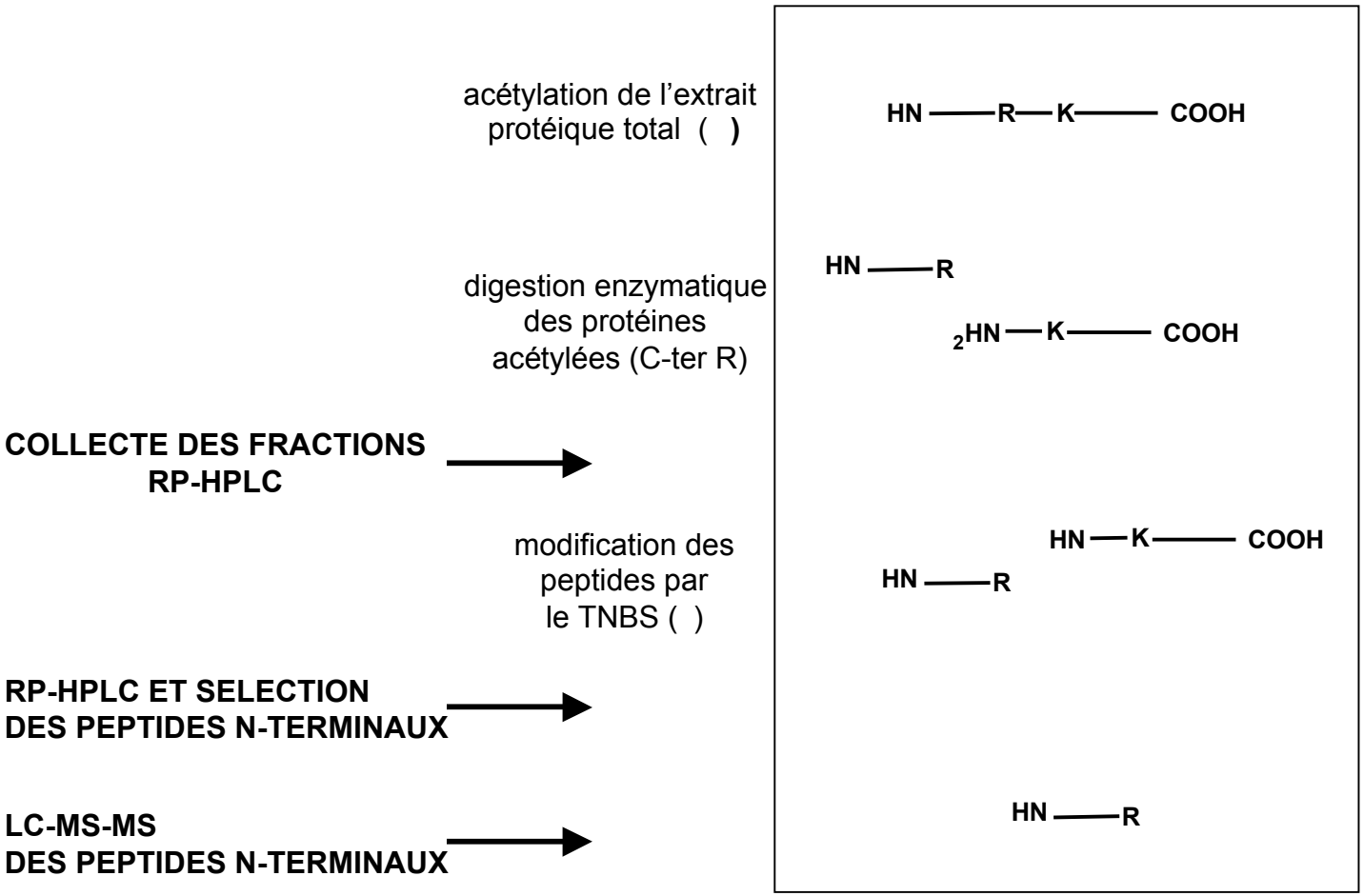
Kris Gevaert, Marc Goethals, Lennart Martens, Jozef Van Damme, An Staes, Gr goire R. Thomas and Jo' l Vandekerckhove

ABSENCE DE GEL

SELECTION DES PEPTIDES N-TERMINAUX (1peptide=1 proteine)

COMPLEXITE TECHNIQUE

APPROCHE COFRADIC
(Combined FRActional Diagonal Chomatography)



L'APPROCHE COFRADIC

↓

NOMBRE SATISFAISANT DE SEQUENCES N-TERMINALES DES PROTEINES DU *Mycobacterium Smegmatis*

↓

LA VALIDATION DU PROGRAMME DE PREDICTION/DIAGNOSTIC DES CODONS D'INITIATION

PLATE-FORME MULTI-SITES RIO de STRASBOURG

(Recouvrement avec la Plate-forme Cancéropôle et Génopôle)

Responsable scientifique A. Van Dorsselaer

Site Partenaire 1
IGBMC Illkirch
Pôle de protéomique
Génopôle-Cancéropôle
Responsable scientifique:
A. Van Dorsselaer et
D. Moras

Site Principal
LSMBO Cronenbourg
– UMR 7509
Pôle Analytique ECPM
Responsable
scientifique: A. Van
Dorsselaer

Site Partenaire 3
IML Strasbourg
Institut de
Médicine Légale
Responsable scientifique:
B. Ludes

Site Partenaire 2
CHU Hautepierre
Plateforme cancéropôle
U381 Inserm
Responsible scientifique:
Jean-François Launay

LABORATOIRE DE SPECTROMETRIE DE MASSE BIO-ORGANIQUE (A. Van Dorsselaer)

RP-HPLC OFF-LINE

LC-MS-MS

Christine SCHAEFFER