# Structural Genomics of Mycobacteria

Pedro M. Alzari

INSTITUT PASTEUR

# Structural Genomics

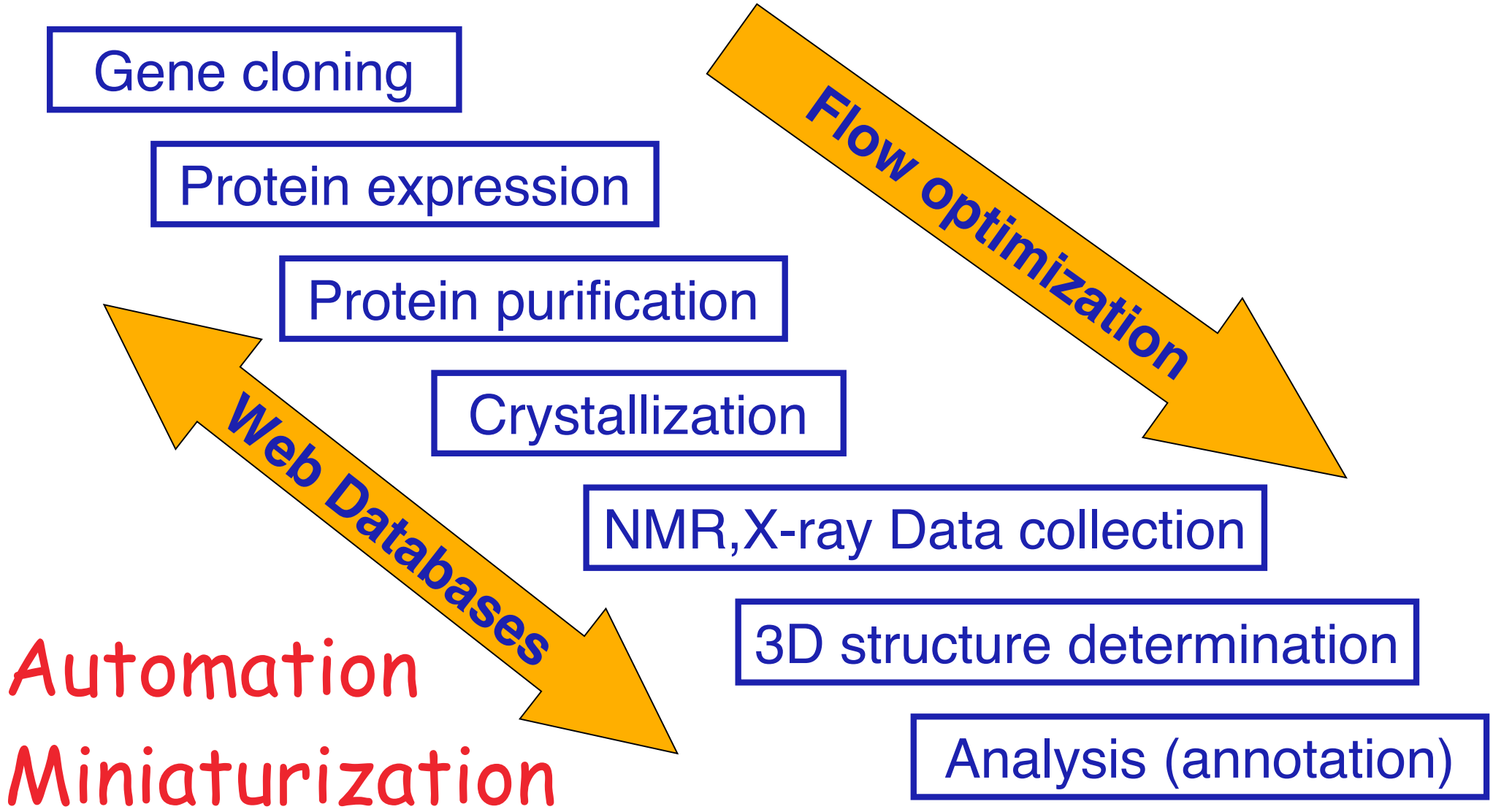
Why?

- Protein fold catalog
- Biomedical interest
- Function discovery

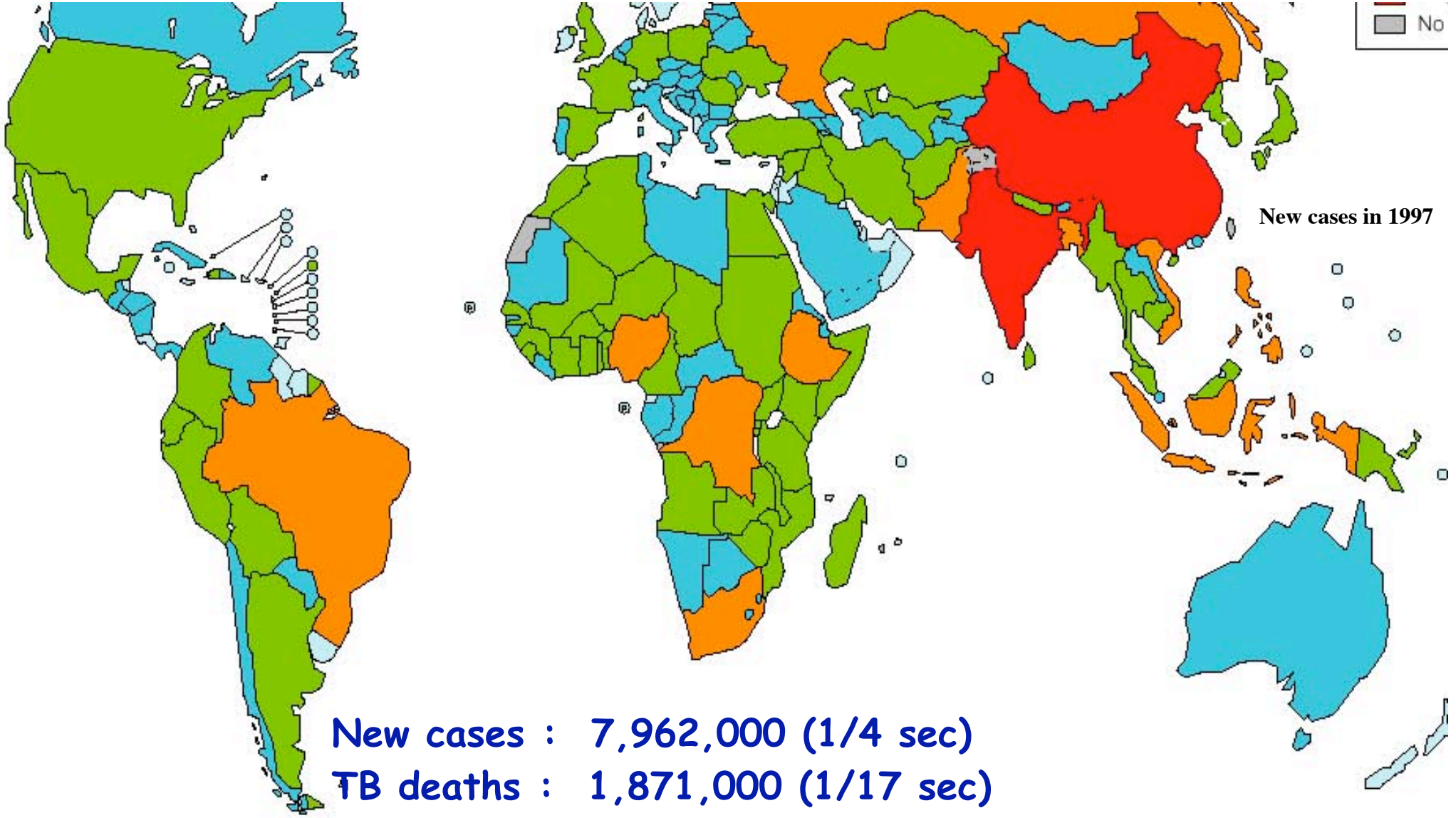
How?

- High-throughput methods

# The Structural Genomics Pipeline



# Estimates of TB burden (1997)

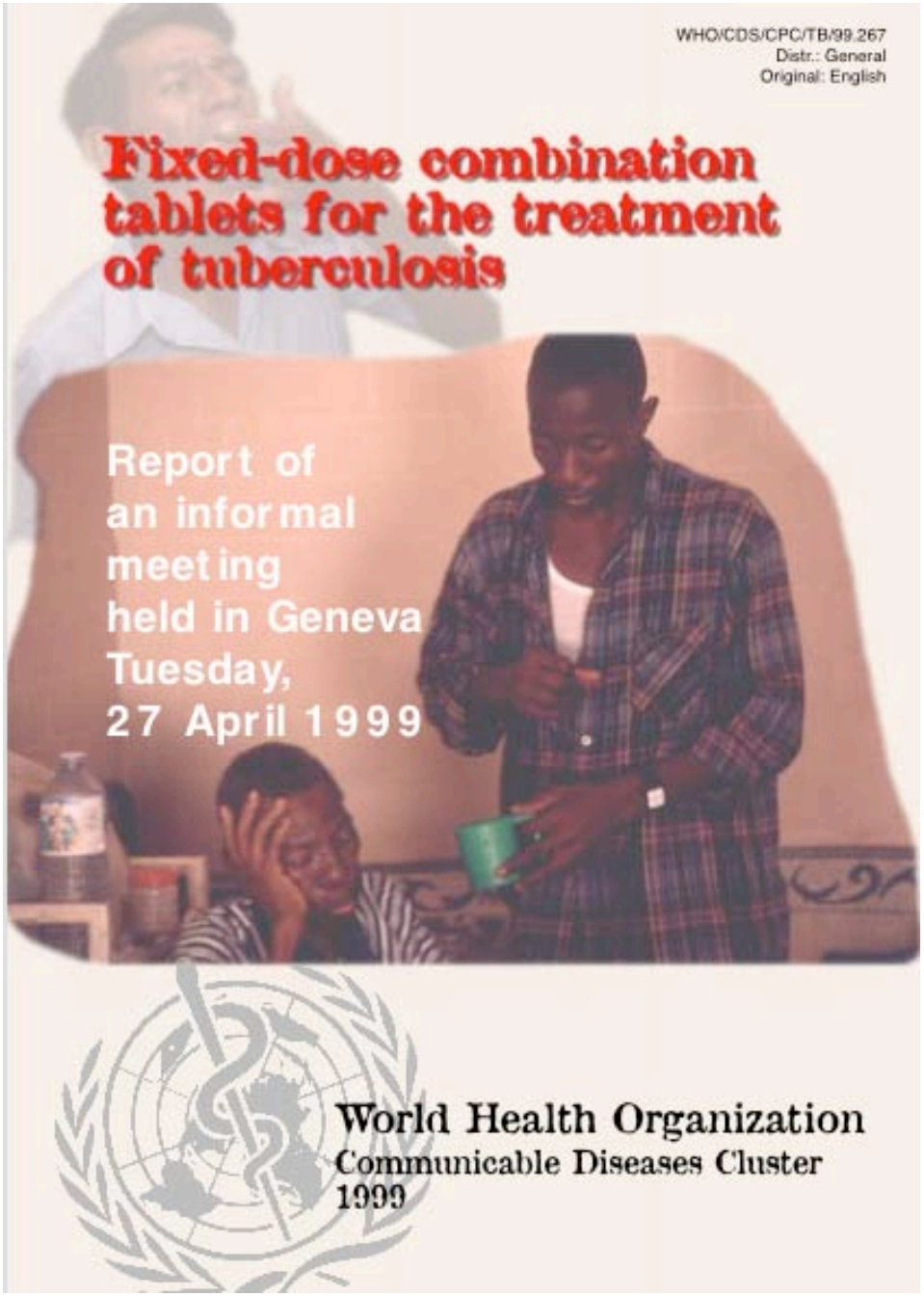


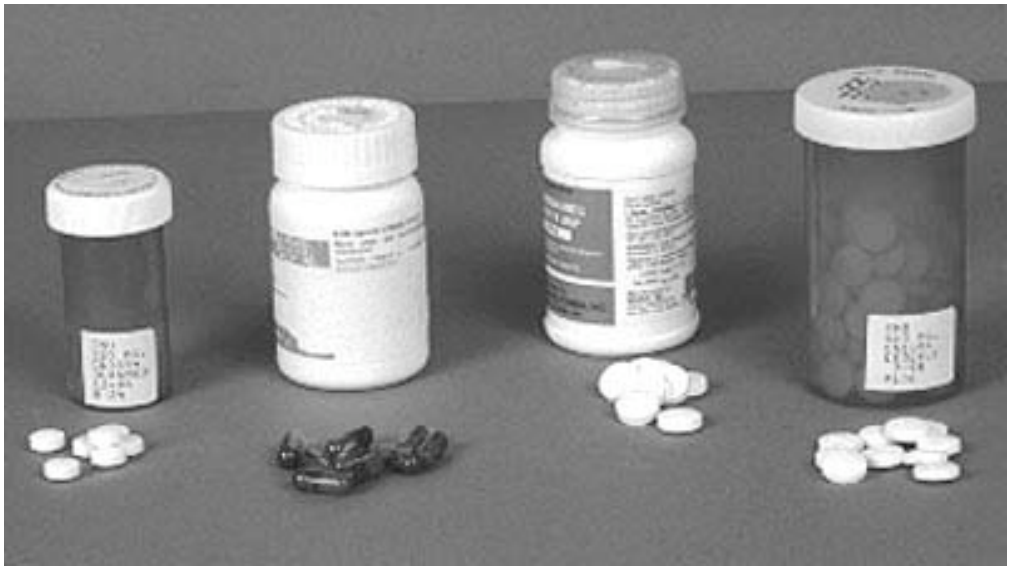

New cases : 7,962,000 (1/4 sec)

TB deaths : 1,871,000 (1/17 sec)

Infection prevalence : 1,855,880,000 (32%)

Dye *et al, JAMA*,1999



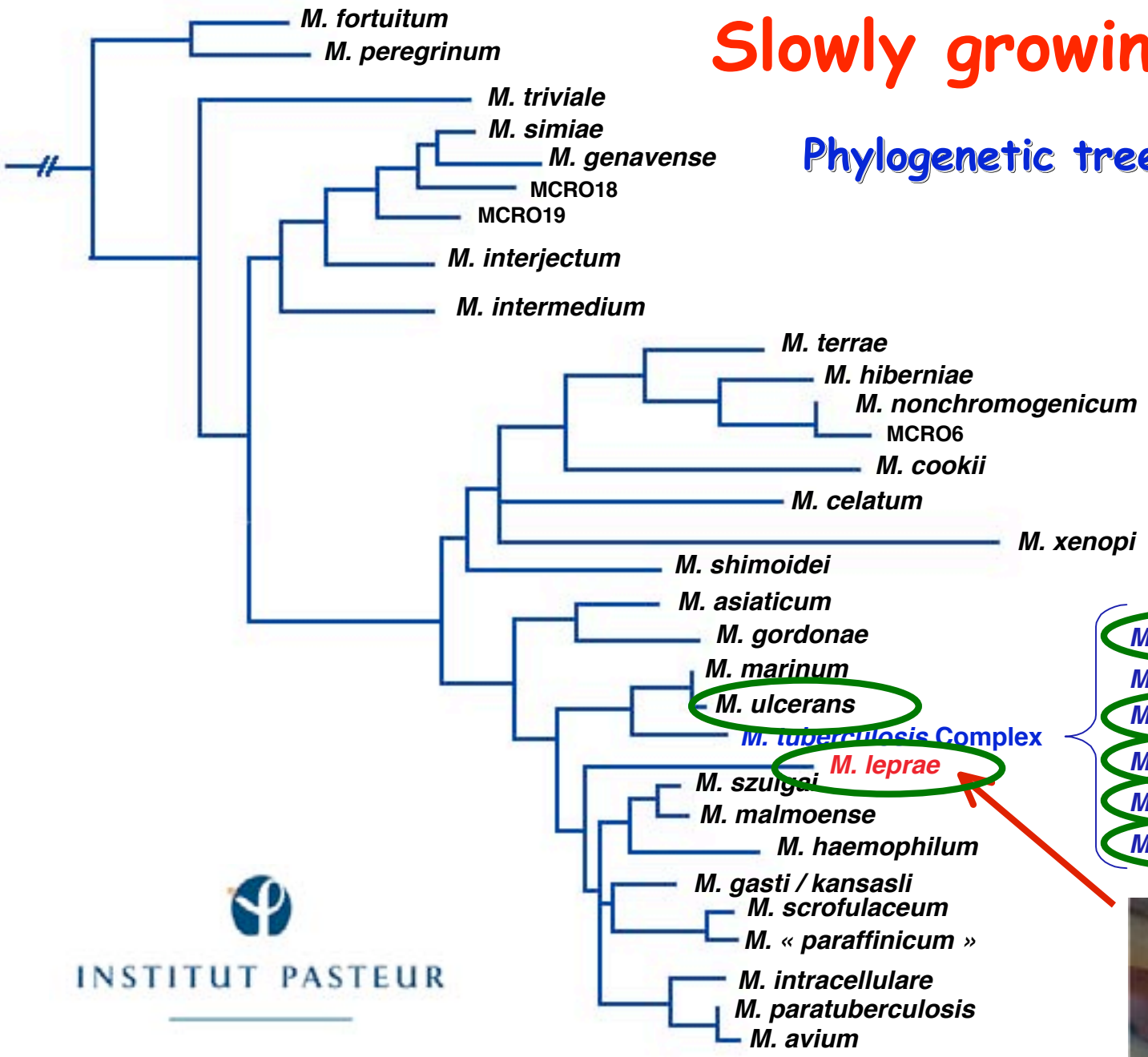



Isoniazid, Rifampicin,
Pyranizamide, Ethambutol

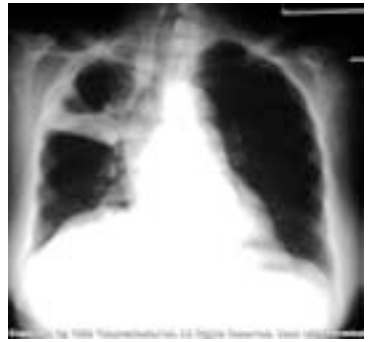
No new anti-TB drug
in more than 40 years

# Slowly growing mycobacteria

## Phylogenetic tree (16S rRNA gene seqs)




INSTITUT PASTEUR



Tuberculosis

M. tuberculosis
M. africanum
M. canettii
M. microti
M. bovis
M. bovis BCG

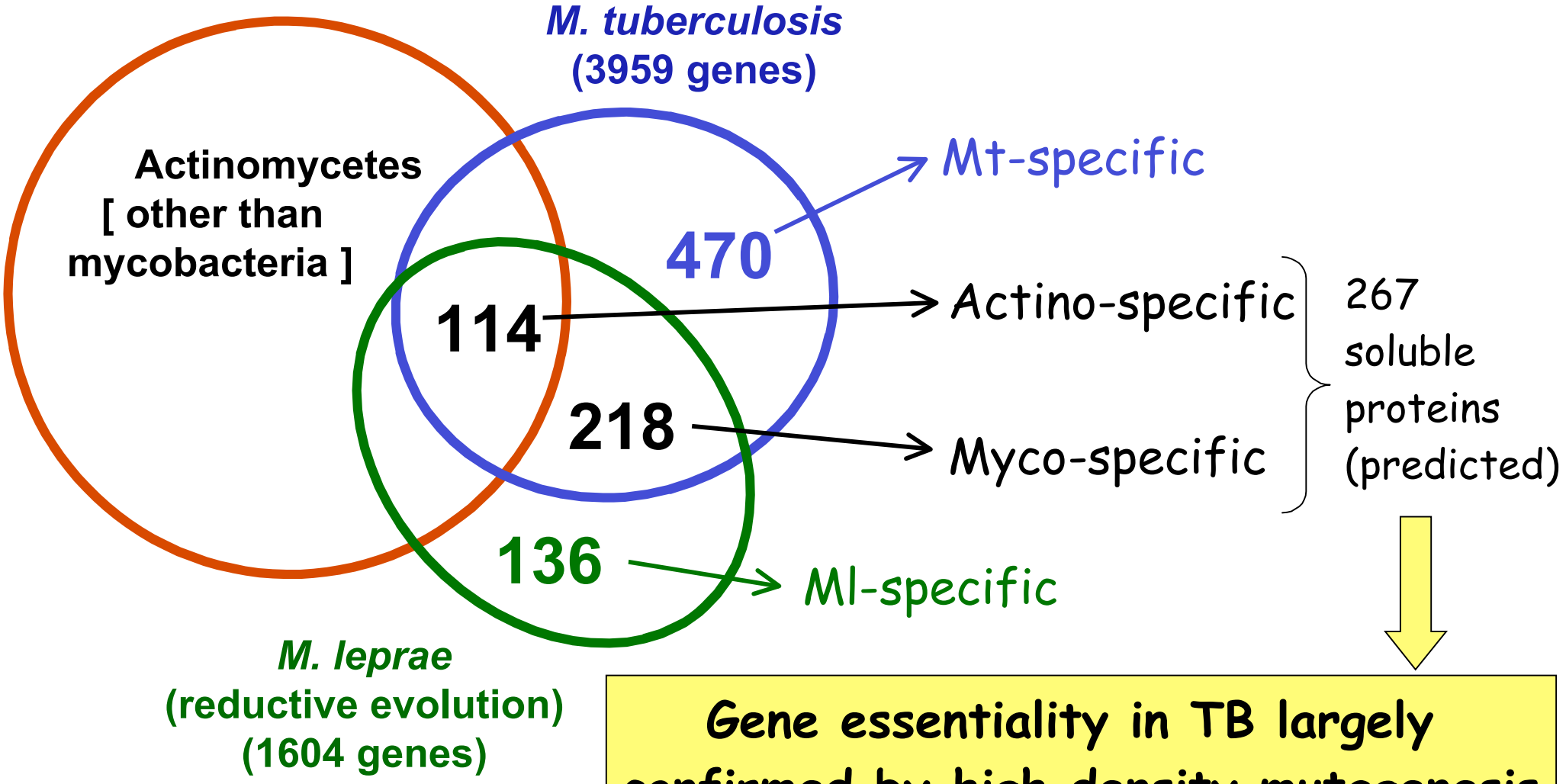
M. smegmatis

Leprosy

# Structural Genomics of Mycobacteria

- Target selection
- The pipeline
- Results

# Actinomycetes-restricted genes

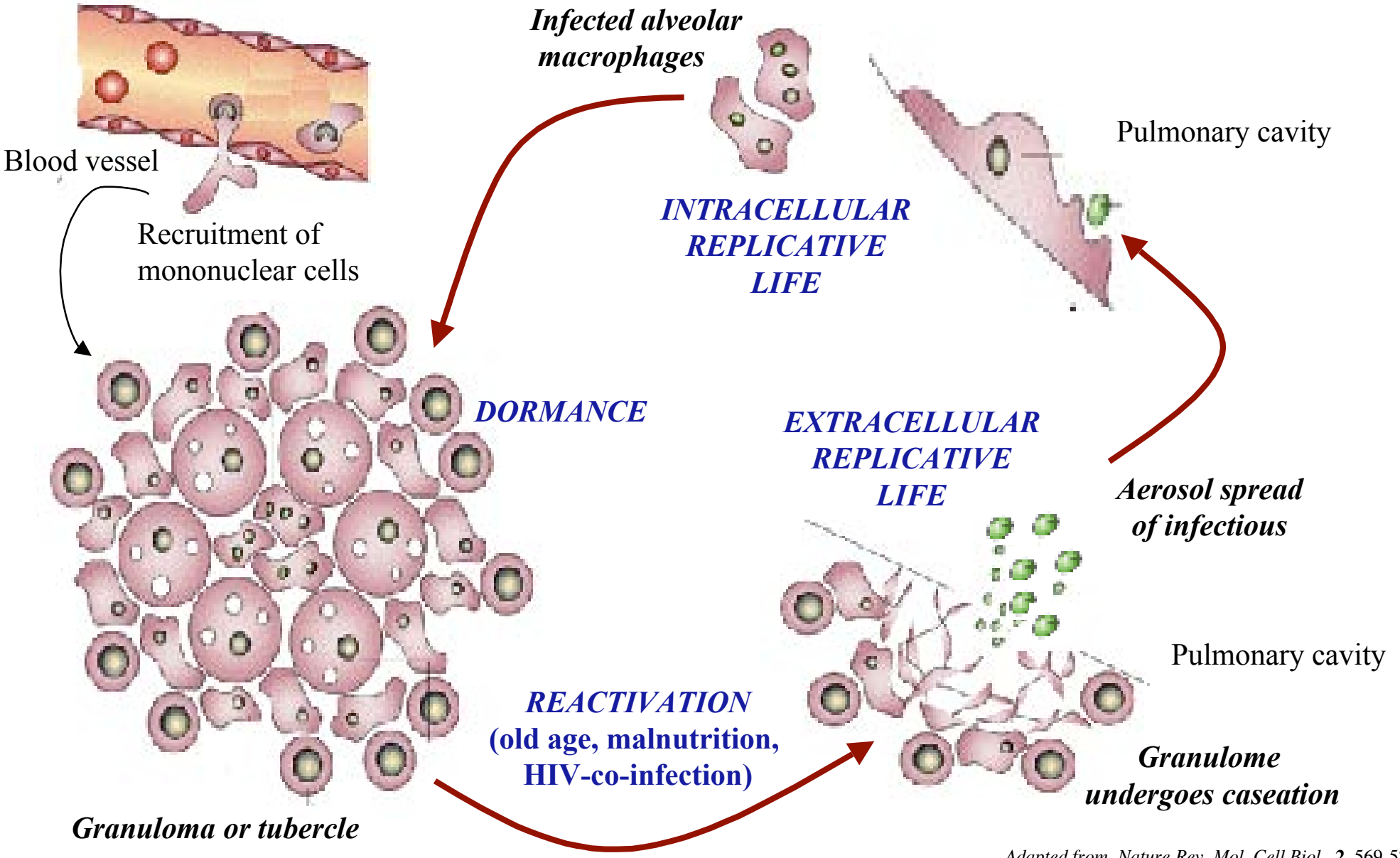



**Gene essentiality in TB largely confirmed by high density mutagenesis**

Sassetti *et al, Mol.Microbiol.*, 2003

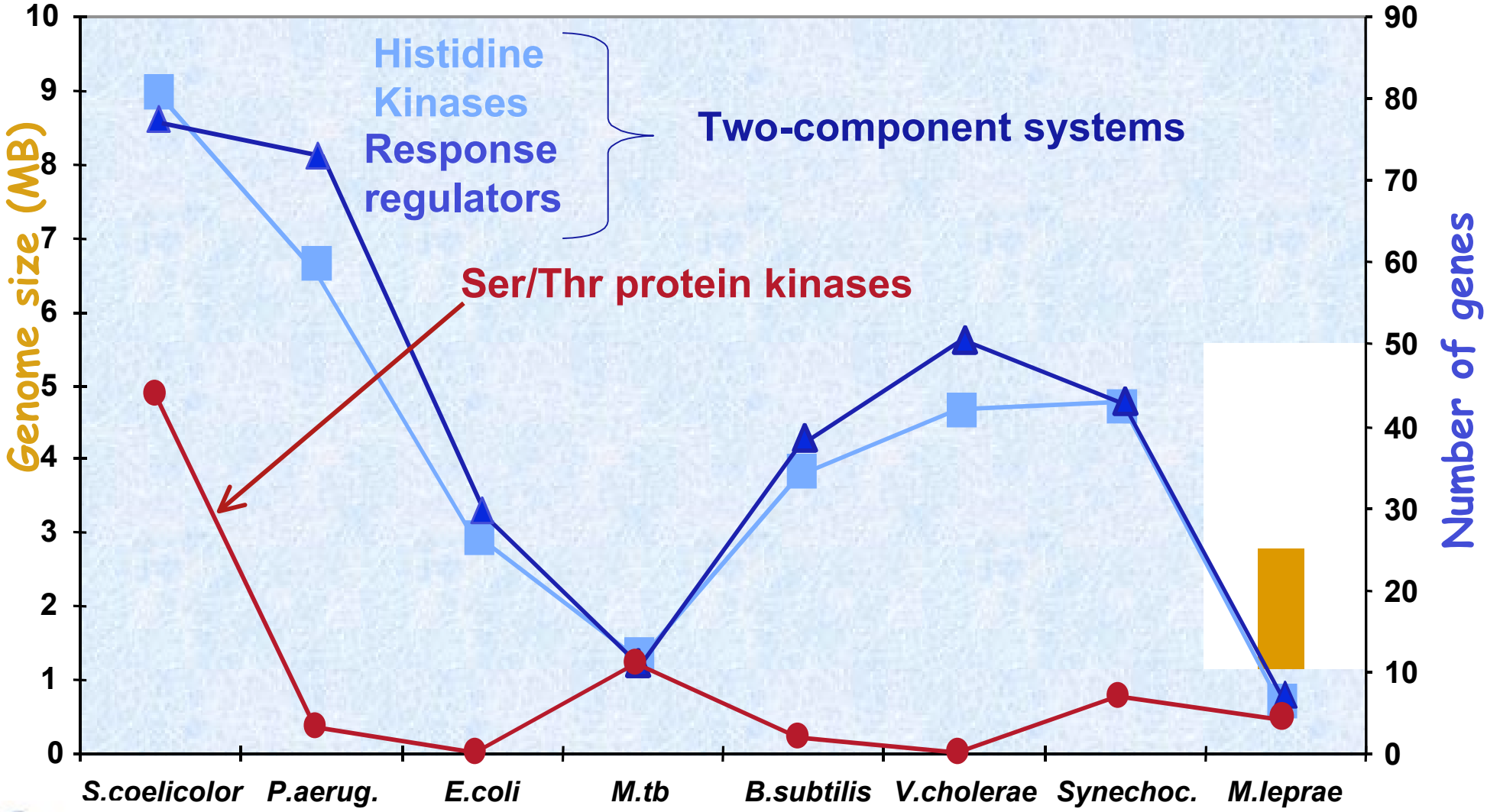
INSTITUT PASTEUR

# Life cycle of *M. tuberculosis*



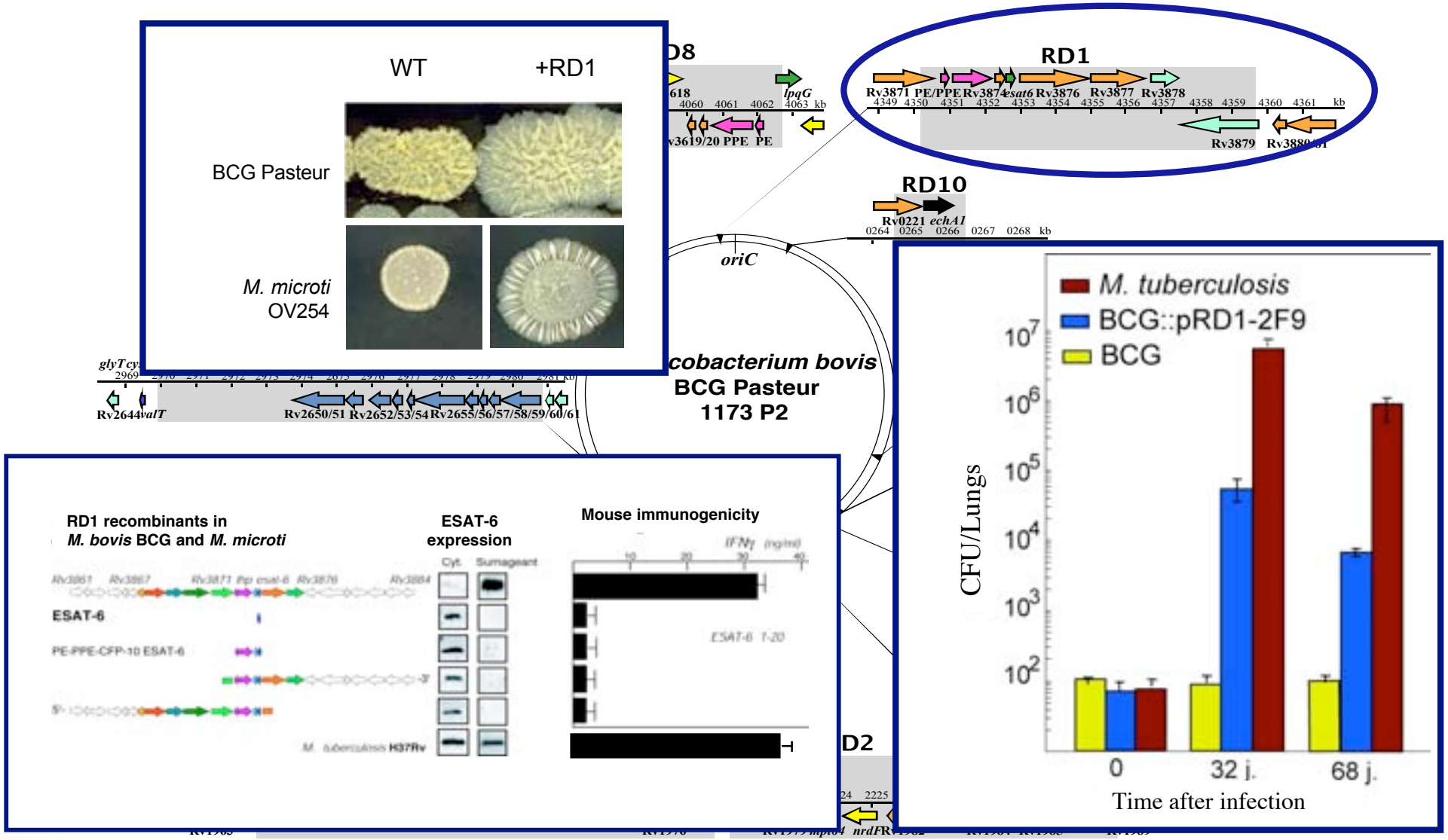

*Adapted from, Nature Rev. Mol. Cell Biol.*, **2**, 569-586 (2001)

# Eukaryotic-like signaling elements



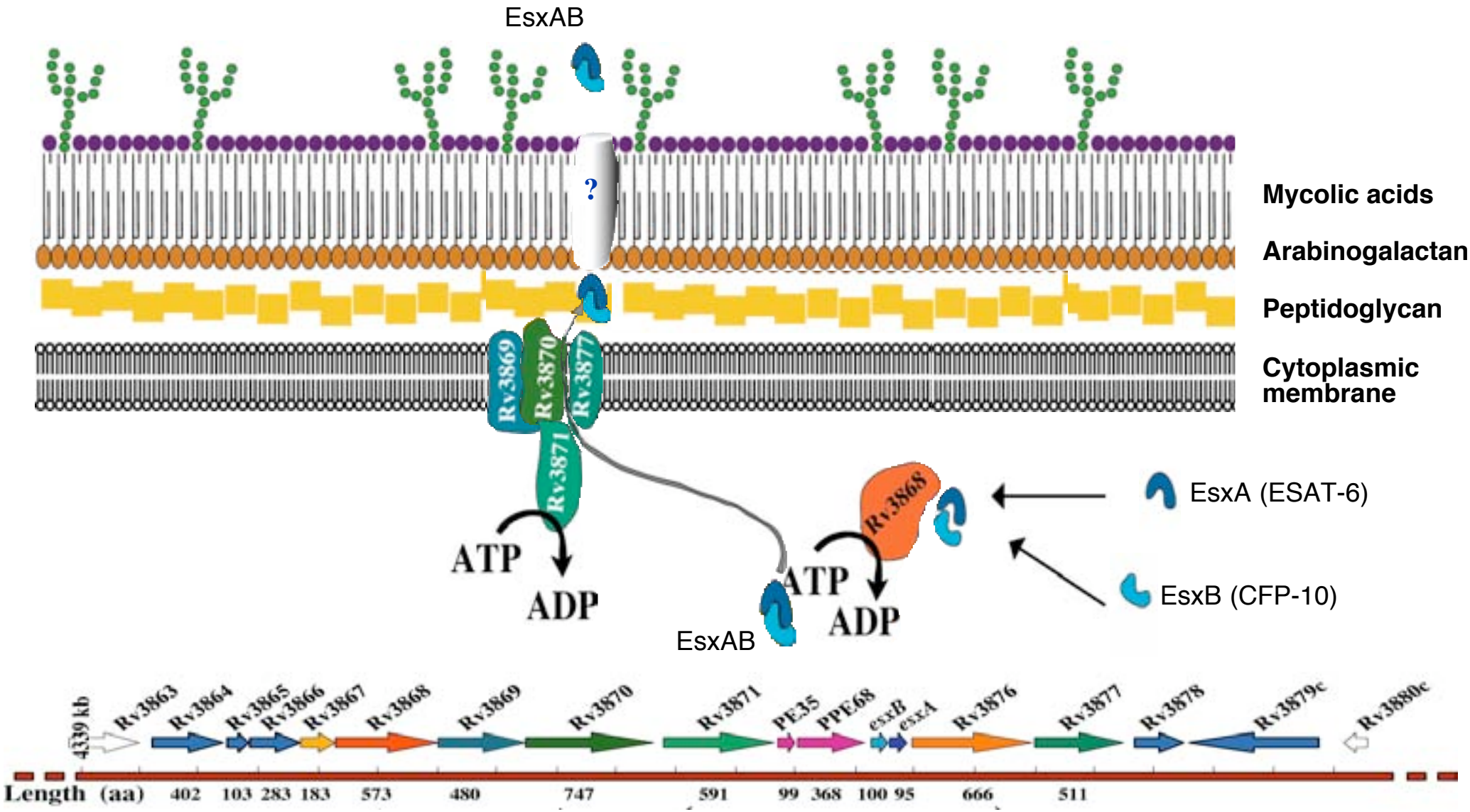

INSTITUT PASTEUR

# Virulence factors




Brodin *et al,* Infect Immun, 2002; Pym *et al*, Mol Microbiol, 2002; Nature Med, 2003.

INSTITUT PASTEUR



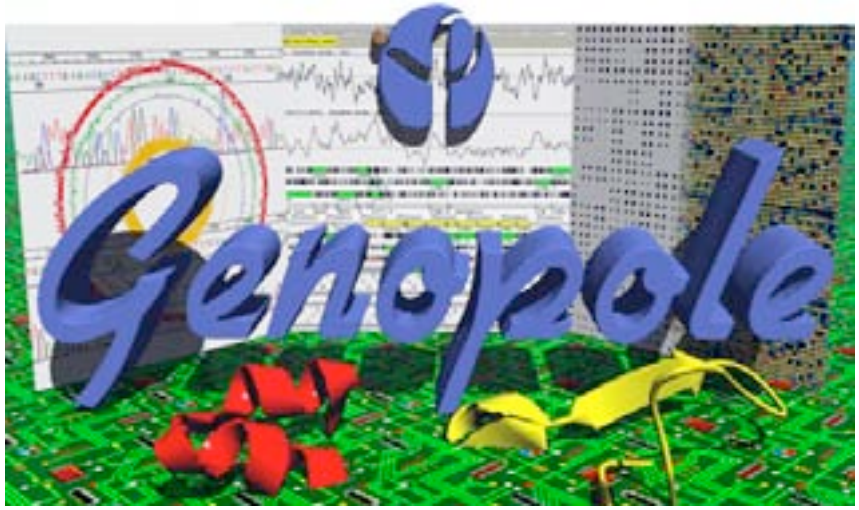

Systematic gene knock-out (secretion machinery)

Virulence of EsxA mutants (host-parasite interactions)

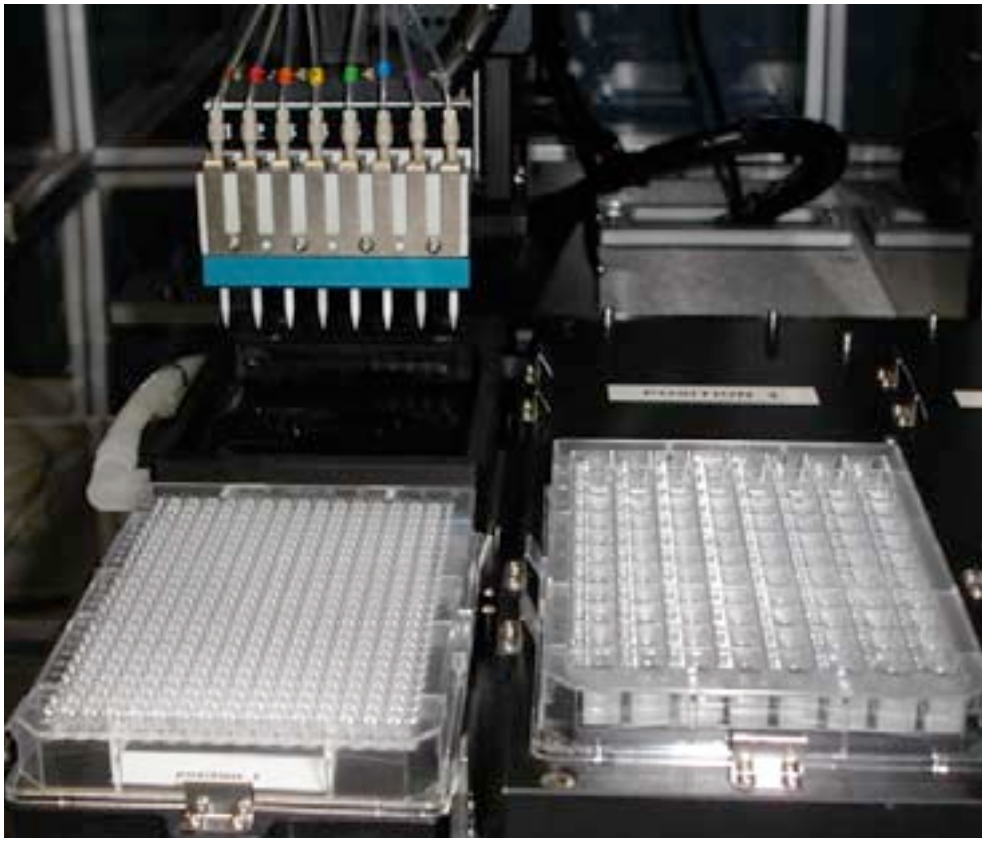
R.Brosch, IP

# Structural Genomics of Mycobacteria

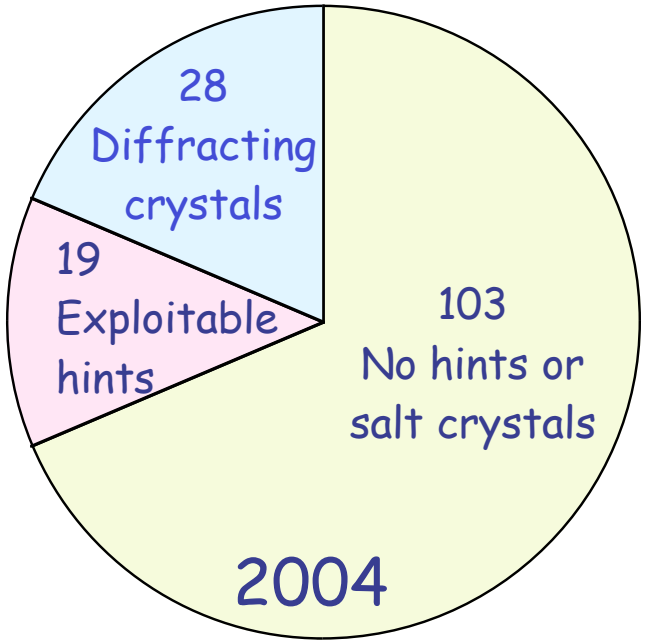
- Target selection
- **The pipeline**
- Results


Genopole

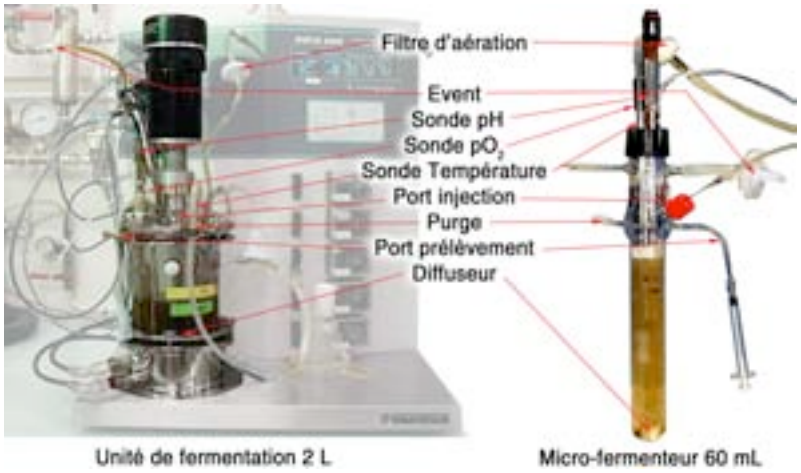
# Crystallization

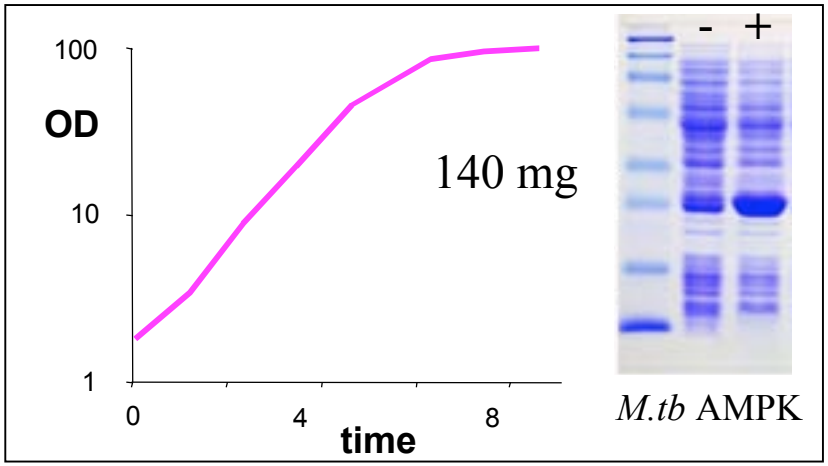


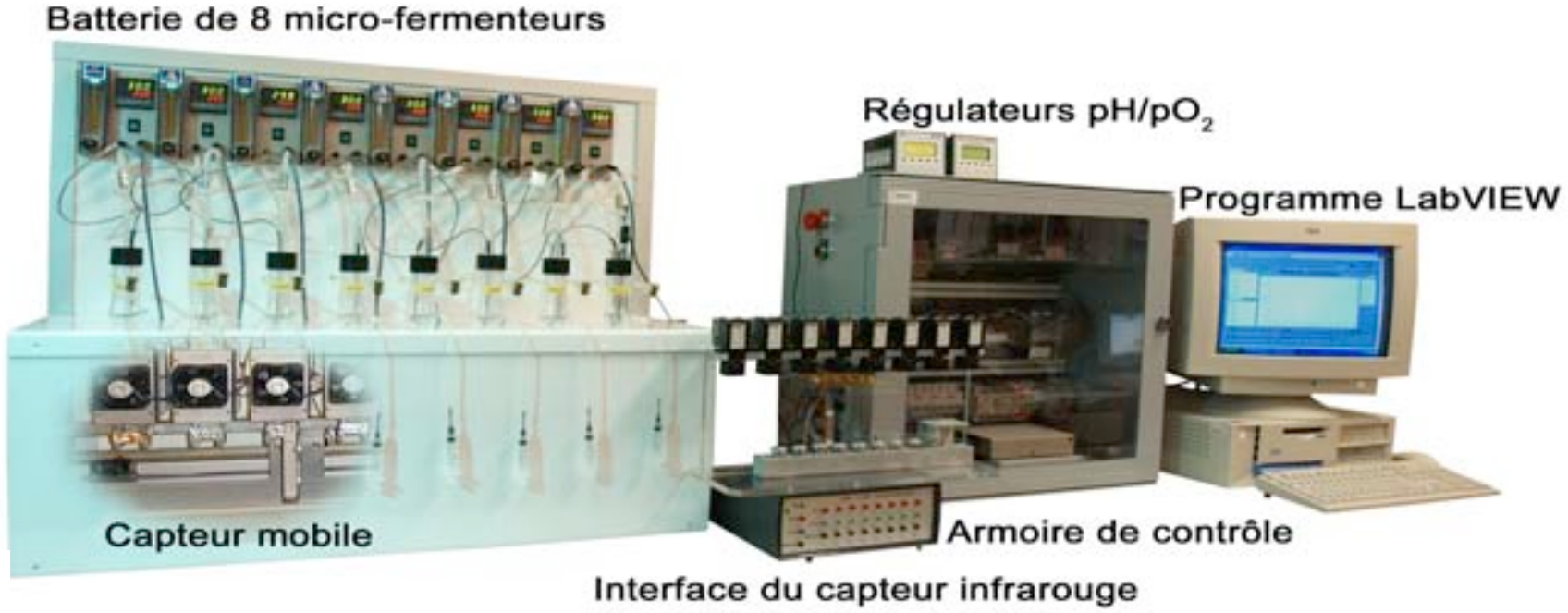
Cartesian Technologies nano-dispenser



# Parallel multi-microfermentors (MMF)



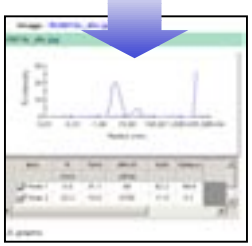



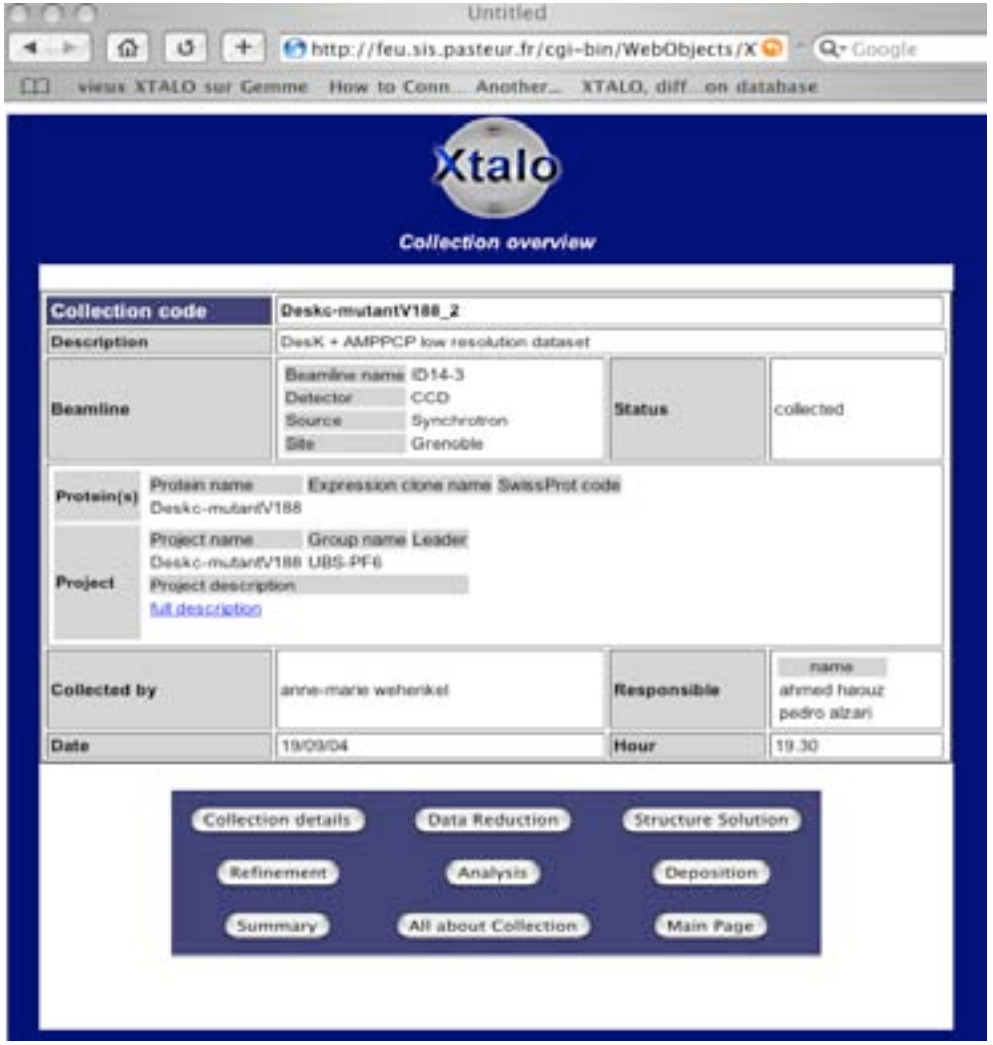




INSTITUT PASTEUR

J.Bellalou, P.Beguin, IP

# Web Databases


http://www.pasteur.fr/SGM

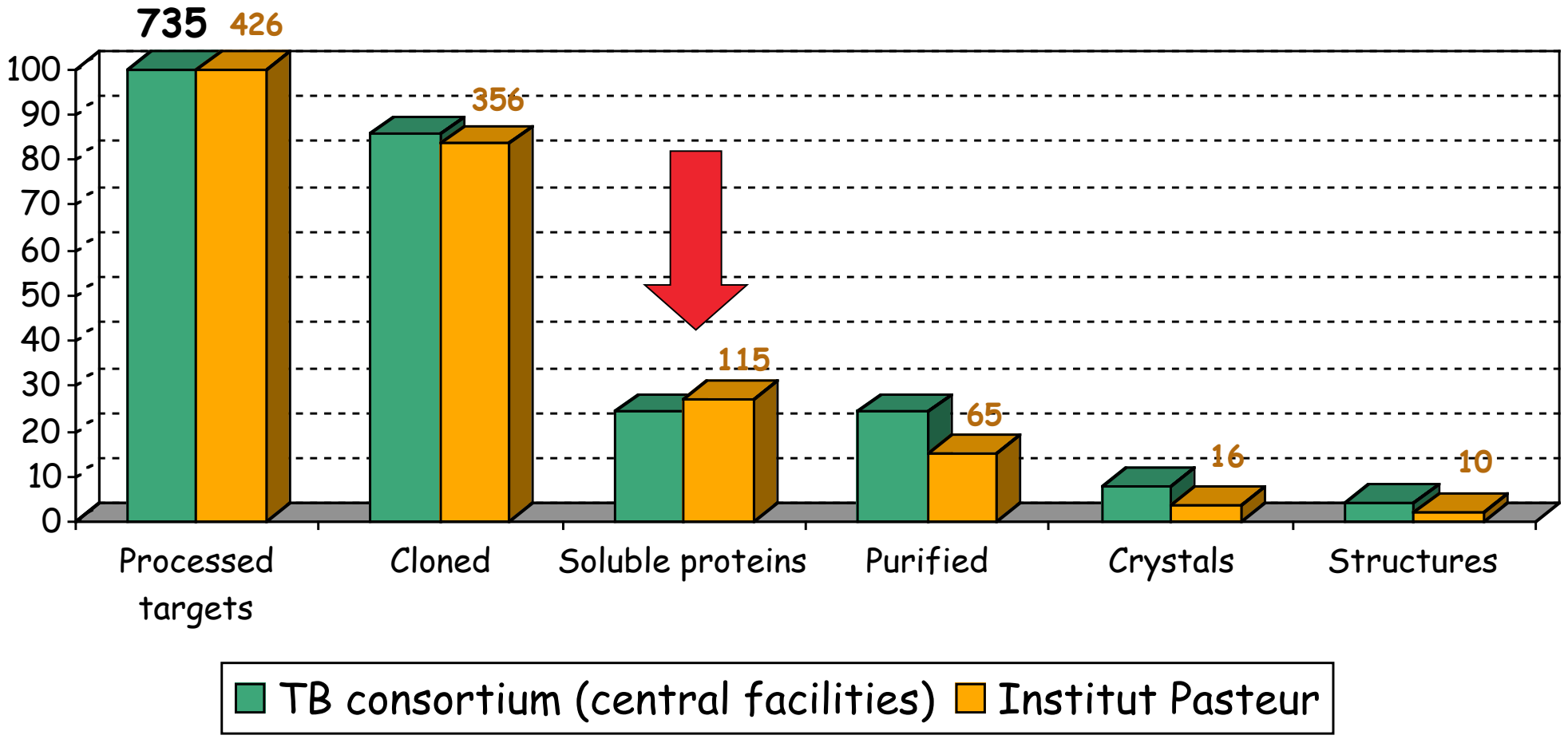
F. Guillemot, IP

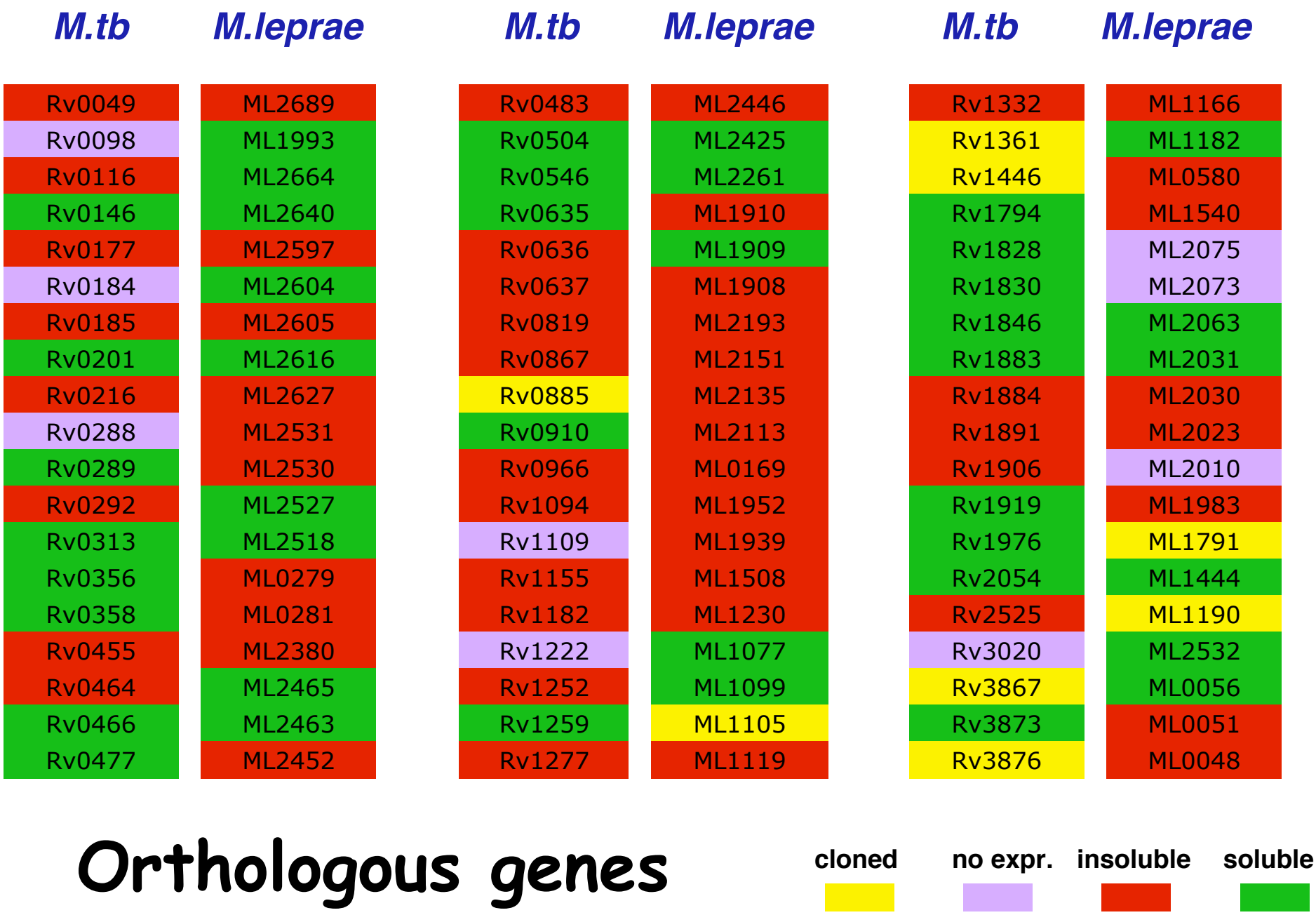
# Structural Genomics of Mycobacteria

- Target selection
- The pipeline
- **Results**

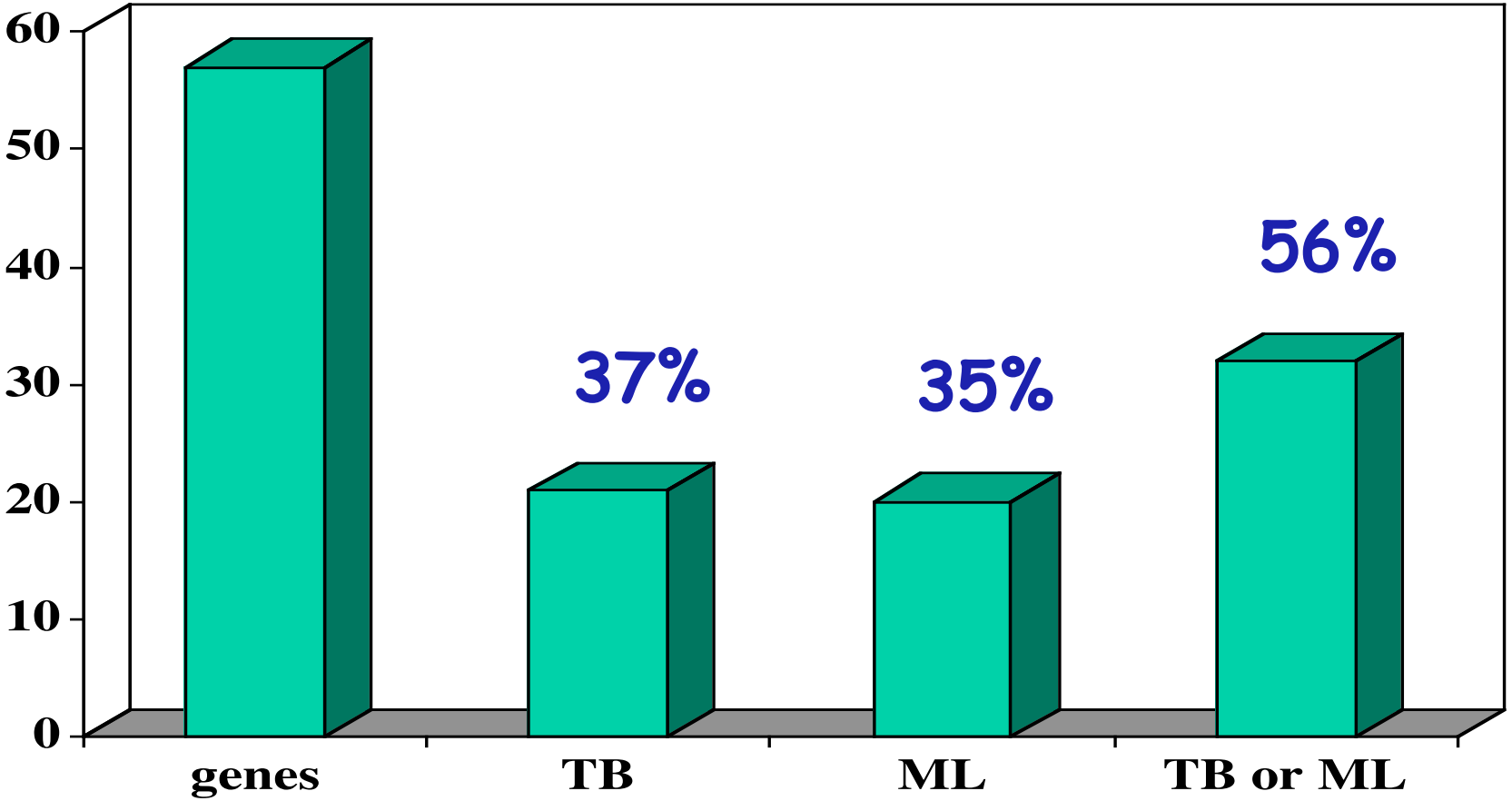
# The solubility bottleneck

(sept 2004)




M.tb
M.leprae
M.tb
M.leprae
M.tb
M.leprae
Rv0049 ML2689
Rv0098 ML1993
Rv0116 ML2664
Rv0146 ML2640
Rv0177 ML2597
Rv0184 ML2604
Rv0185 ML2605
Rv0201 ML2616
Rv0216 ML2627
Rv0288 ML2531
Rv0289 ML2530
Rv0292 ML2527
Rv0313 ML2518
Rv0356 ML0279
Rv0358 ML0281
Rv0455 ML2380
Rv0464 ML2465
Rv0466 ML2463
Rv0477 ML2452
Rv0483 ML2446
Rv0504 ML2425
Rv0546 ML2261
Rv0635 ML1910
Rv0636 ML1909
Rv0637 ML1908
Rv0819 ML2193
Rv0867 ML2151
Rv0885 ML2135
Rv0910 ML2113
Rv0966 ML0169
Rv1094 ML1952
Rv1109 ML1939
Rv1155 ML1508
Rv1182 ML1230
Rv1222 ML1077
Rv1252 ML1099
Rv1259 ML1105
Rv1277 ML1119
Rv1332 ML1166
Rv1361 ML1182
Rv1446 ML0580
Rv1794 ML1540
Rv1828 ML2075
Rv1830 ML2073
Rv1846 ML2063
Rv1883 ML2031
Rv1884 ML2030
Rv1891 ML2023
Rv1906 ML2010
Rv1919 ML1983
Rv1976 ML1791
Rv2054 ML1444
Rv2525 ML1190
Rv3020 ML2532
Rv3867 ML0056
Rv3873 ML0051
Rv3876 ML0048
Orthologous genes
cloned
no expr.
insoluble
soluble

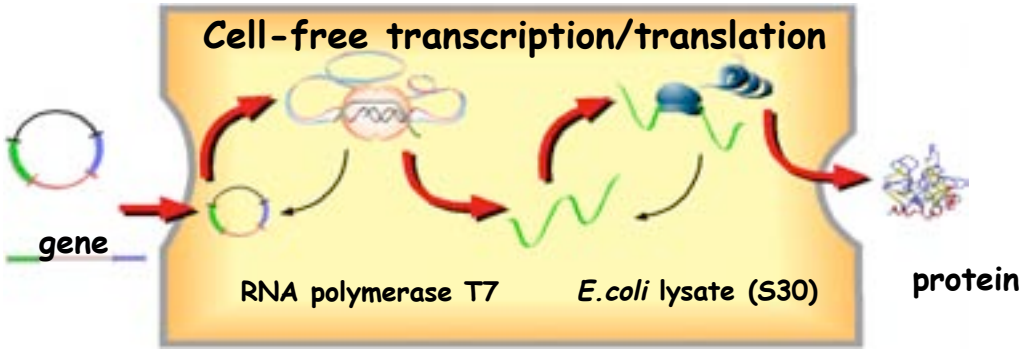
# Soluble proteins



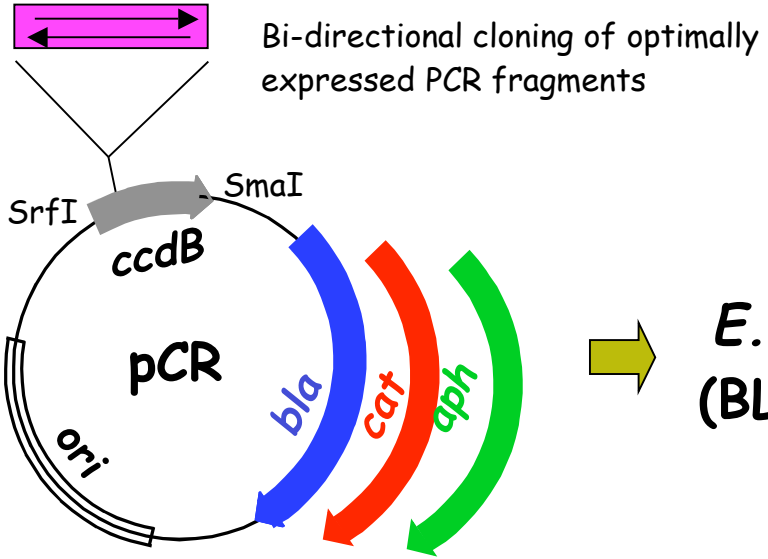
# Optimizing protein expression parameters


**Testing protein expression before bacterial cloning**

PCR:

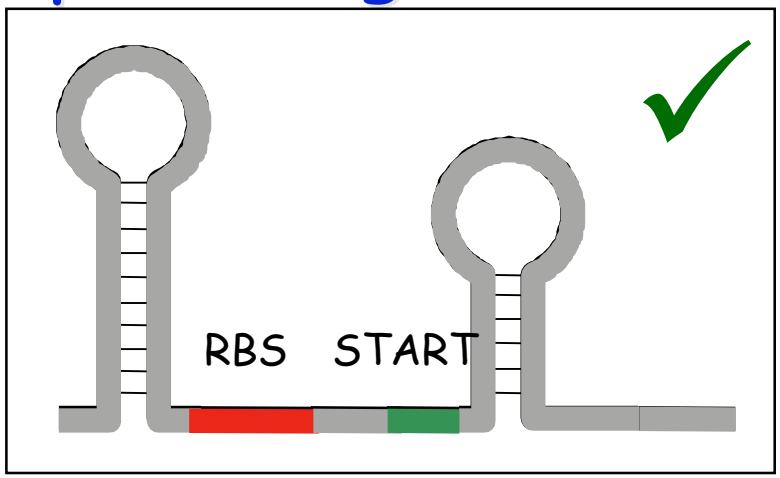

Parameter optimization

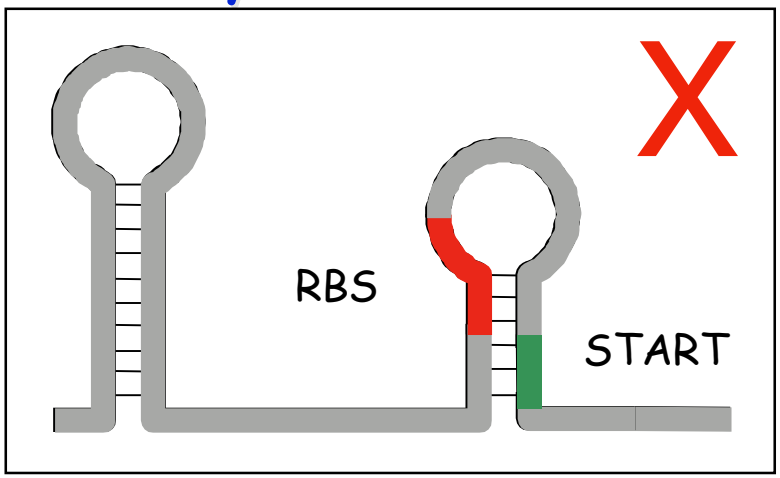



*E.coli* (BL21)

J.M.Betton, IP

# Optimizing for mRNA secondary structure

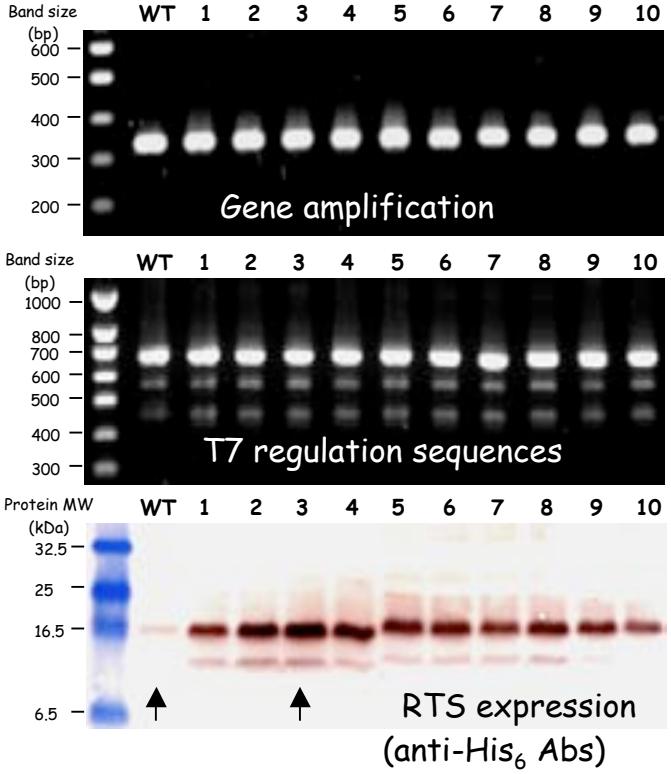


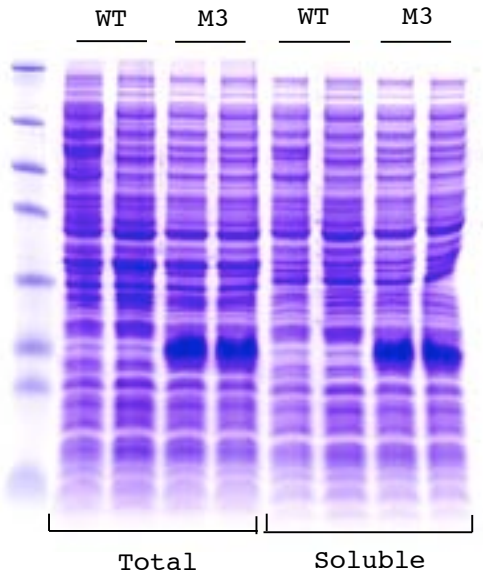

## ML0180

| | Met | Pro | Thr | Tyr | Ser | Tyr | Glu |
|---|---|---|---|---|---|---|---|
| WT | ATG | CCG | ACC | TAC | AGC | TAC | GAG |
| M1 | ATG | CCA | ACT | TAT | TCA | TAT | GAA |
| M2 | ATG | CCA | ACT | TAT | TCA | TAT | GAG |
| M3 | ATG | CCA | ACA | TAT | TCA | TAT | GAG |
| M4 | ATG | CCA | ACC | TAT | TCA | TAT | GAA |
| M5 | ATG | CCA | ACT | TAC | TCA | TAT | GAA |
| M6 | ATG | CCA | ACT | TAC | TCT | TAT | GAA |
| M7 | ATG | CCA | ACC | TAT | TCA | TAT | GAG |
| M8 | ATG | CCA | ACT | TAT | TCA | TAC | GAG |
| M9 | ATG | CCA | ACT | TAT | TCA | TAC | GAA |
| M10 | ATG | CCA | ACA | TAT | TCA | TAC | GAG |

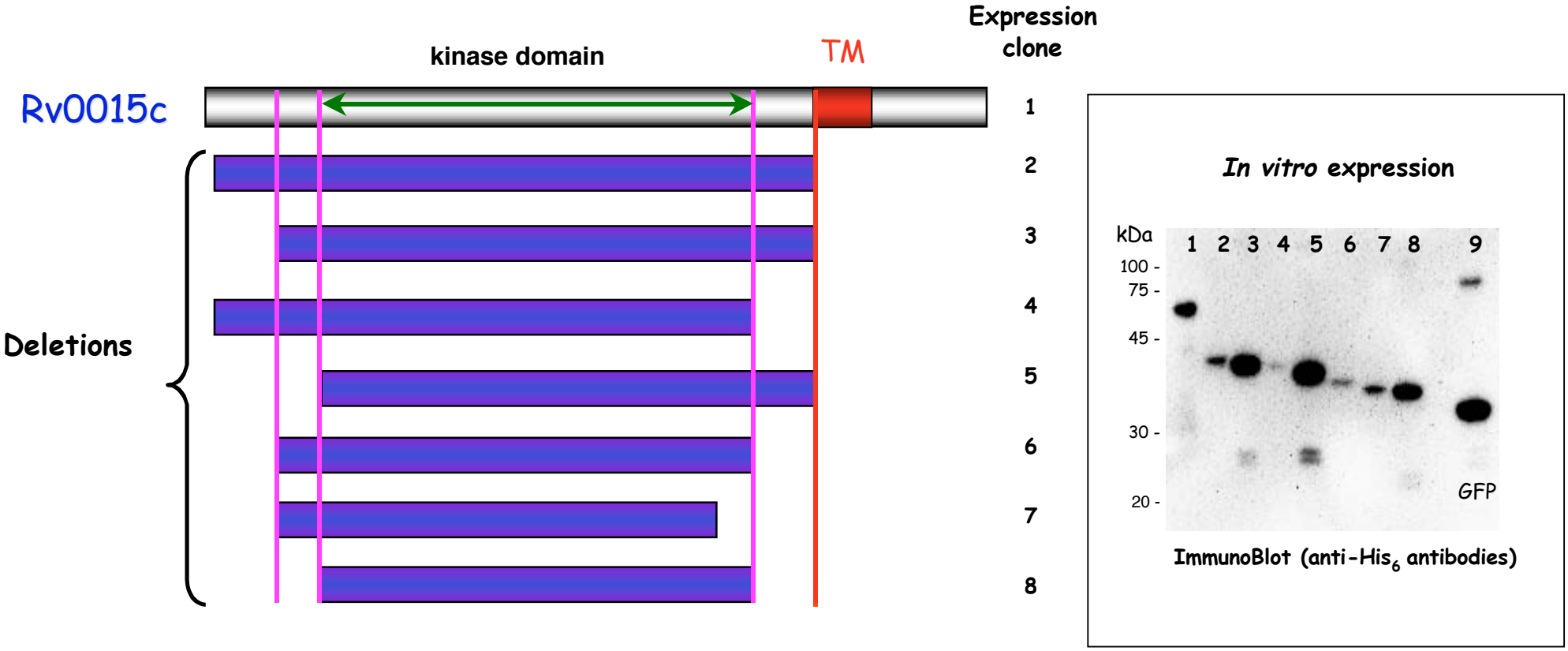
Silent mutations (ProteoExpert)



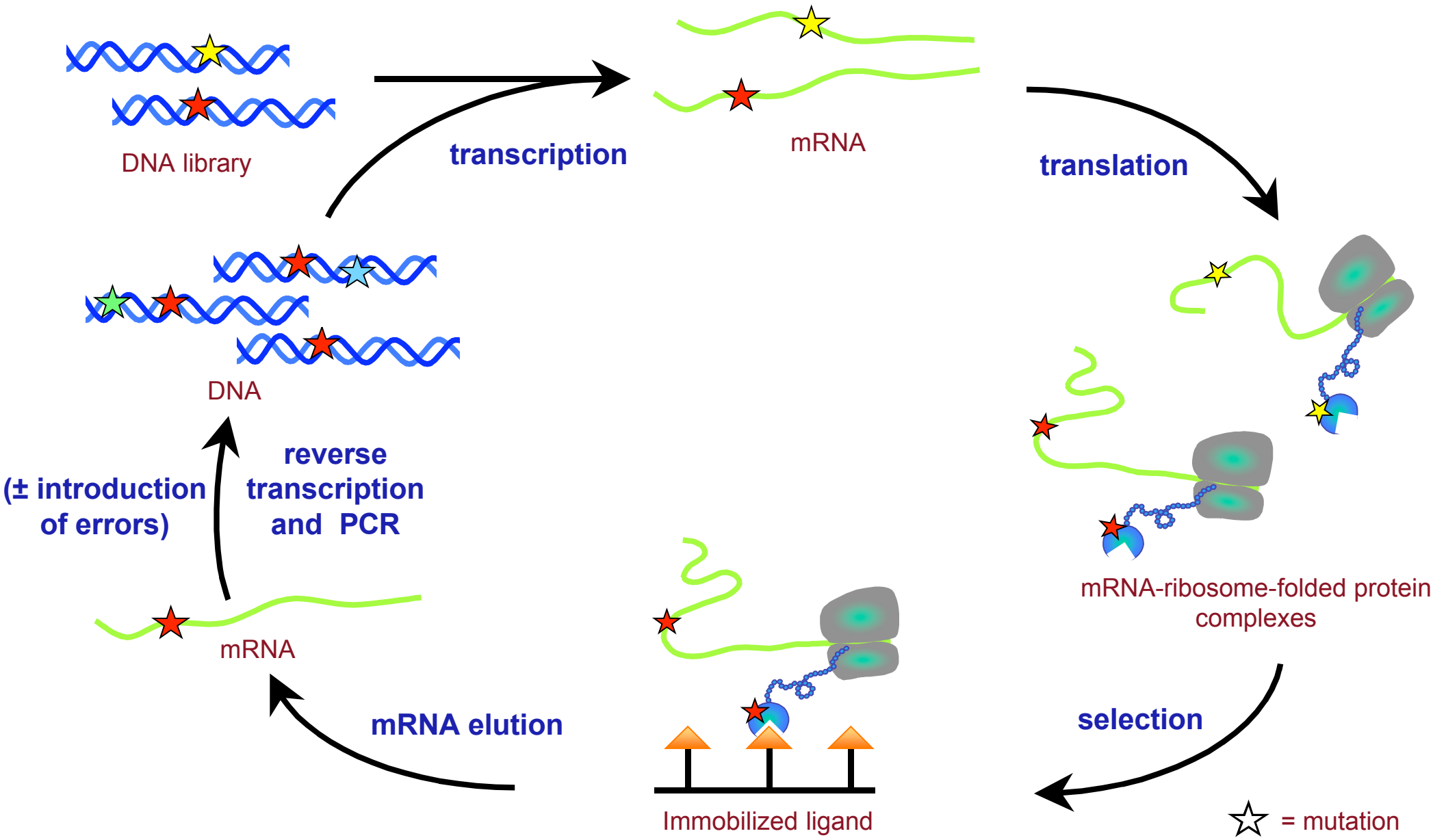

## BL21 expression



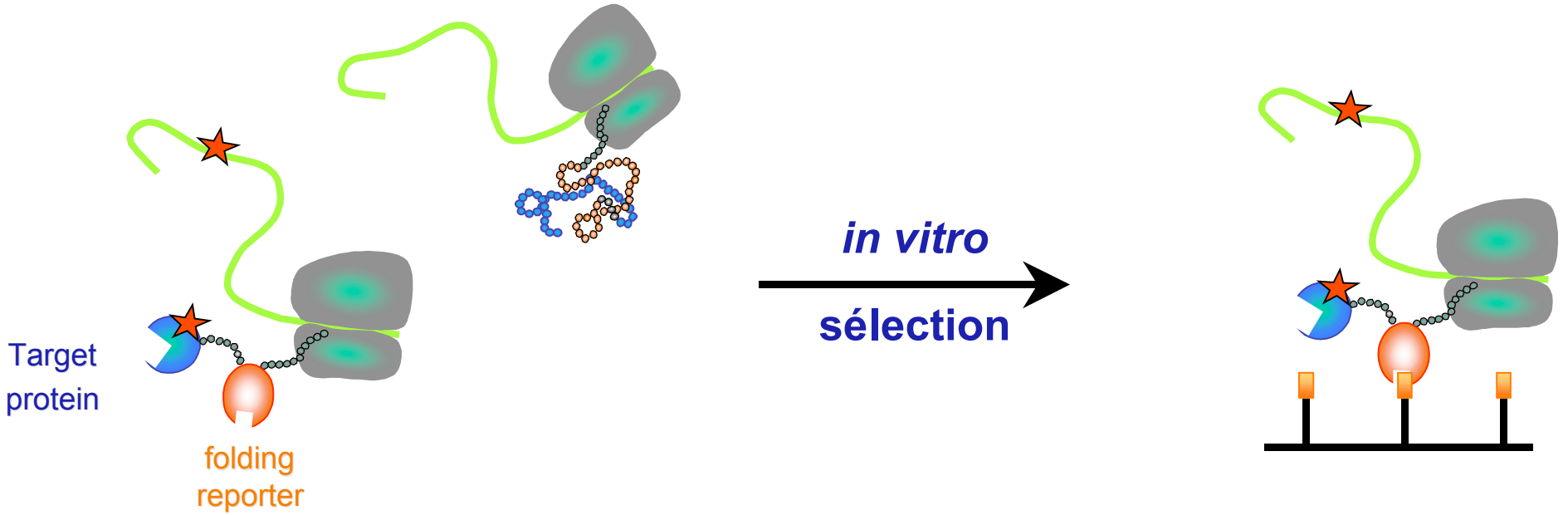
# Optimizing the expression of soluble domains



# *In vitro evolution (ribosome display)*



# Selection for solubility




library of variants of
the target protein(s)

**Esx-1 system**

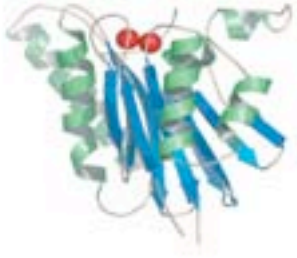
immobilized
ligand

☆ = mutation

F. Pecorari, IP


Rv0014c$_{331}$
Rv0014c$_{279}$
Rv0018c
Rv0733
Rv0813c
Rv0877
Rv1846c
Rv1908c
Rv2238
Rv2276
Rv2428
Rv2461c
Rv2543
Rv2610c
Rv2667
Rv2714
Rv2883c
Rv2991
Rv3628
ML2640
Rv1155
Rv1208
Rv1399c
Rv2125
Rv2171
Rv2945c
Rv3013
Rv3849

Rv0014c



Rv0018c

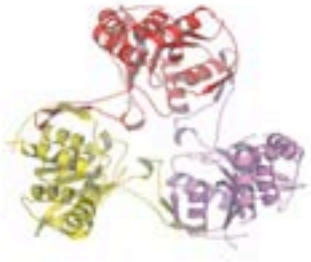
Rv0733

Rv0813c

Rv2238

Rv2428

Rv2461c



Rv2714

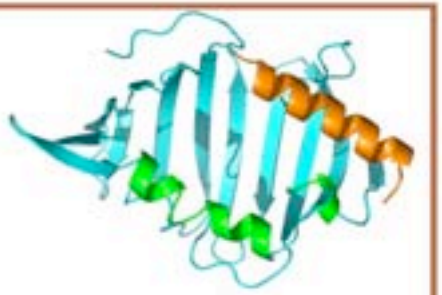
Rv2991

ML2640

Rv2276
(coll. A.Munro)

Rv3628
(coll. S.Benini)

Rv1155
(Y.Bourne)



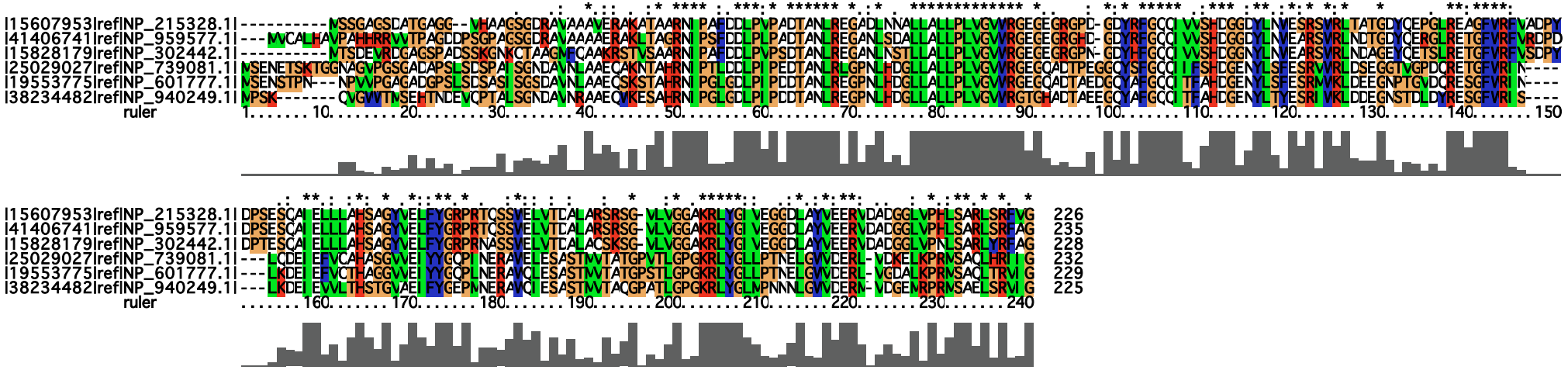
Rv2945c
(Y.Bourne)

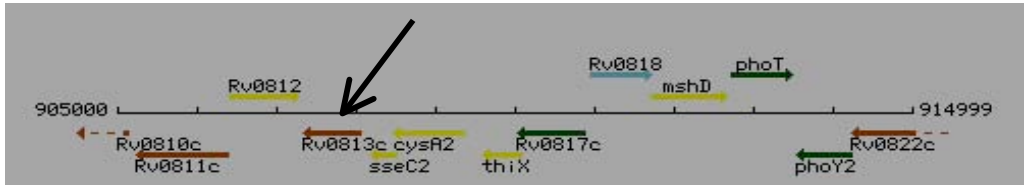
**Diffraction data:** Rv0410c, Rv1846c, Rv2438c, Rv2543, Rv2610c, Rv2883c

# Structural Genomics of Mycobacteria

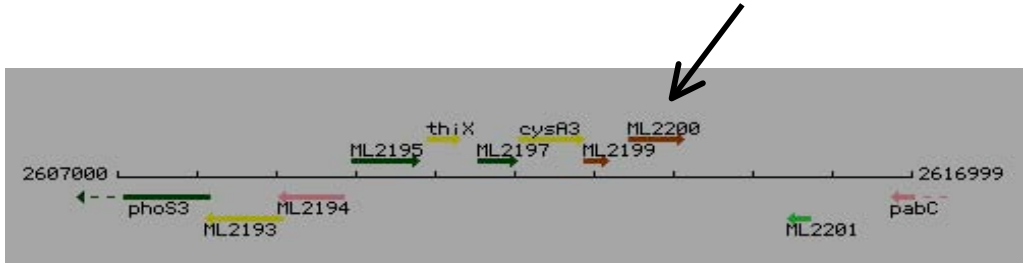
## From structure to function: structure-based functional annotation

# Rv0813c, a hypothetical protein conserved in mycobacteria and corynebacteria



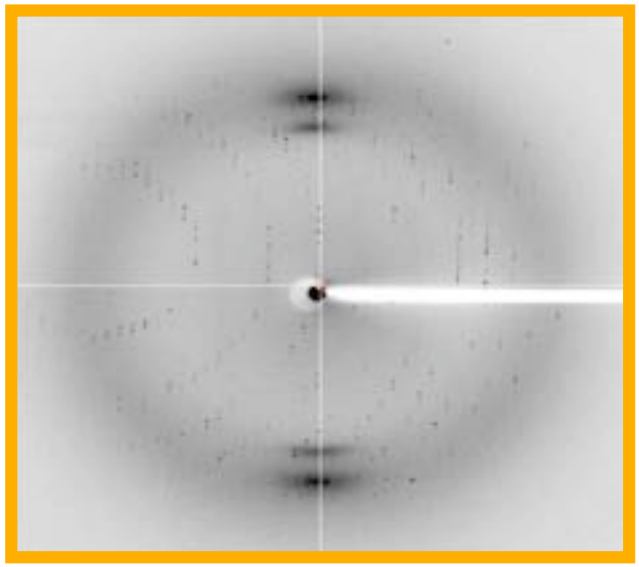




*M. tuberculosis*

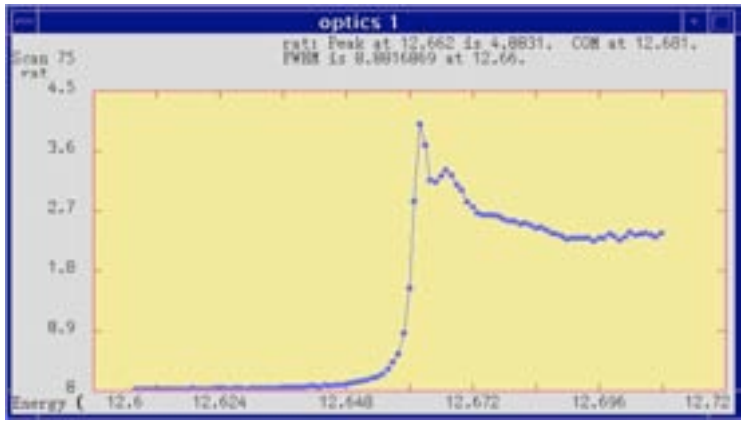


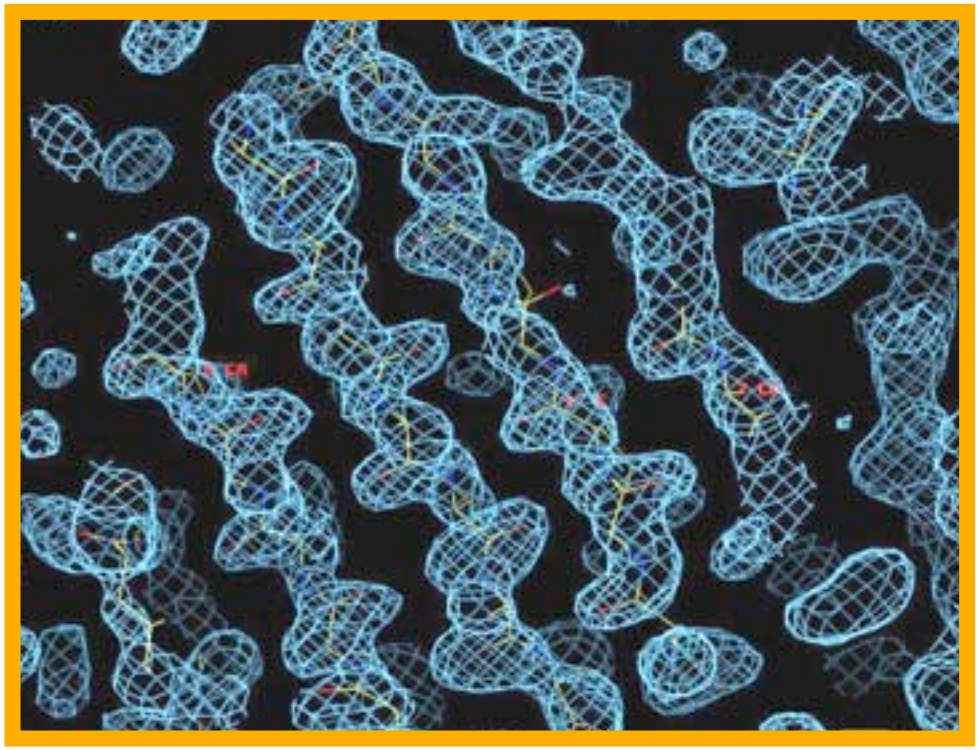

*M. leprae*

# 3D Structure Resolution

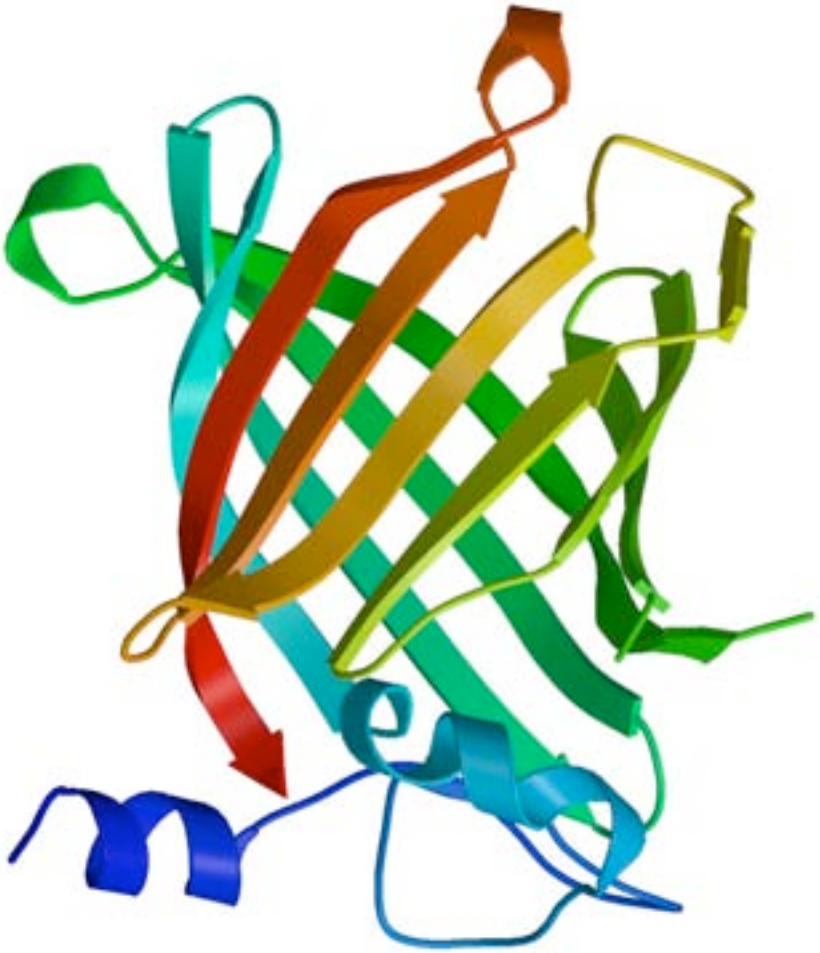


- Strategies
  - Absence of methionines
    - Double mutant Ile->Met
  - Anomalous diffraction experiments
    - SAD & MAD on ID29
    - Selenium K-edge
    - Se(Met) crystal size < 40µm
  - Phase extension to 1.7Å
  - Automatic tracing
    - 80% amino acids

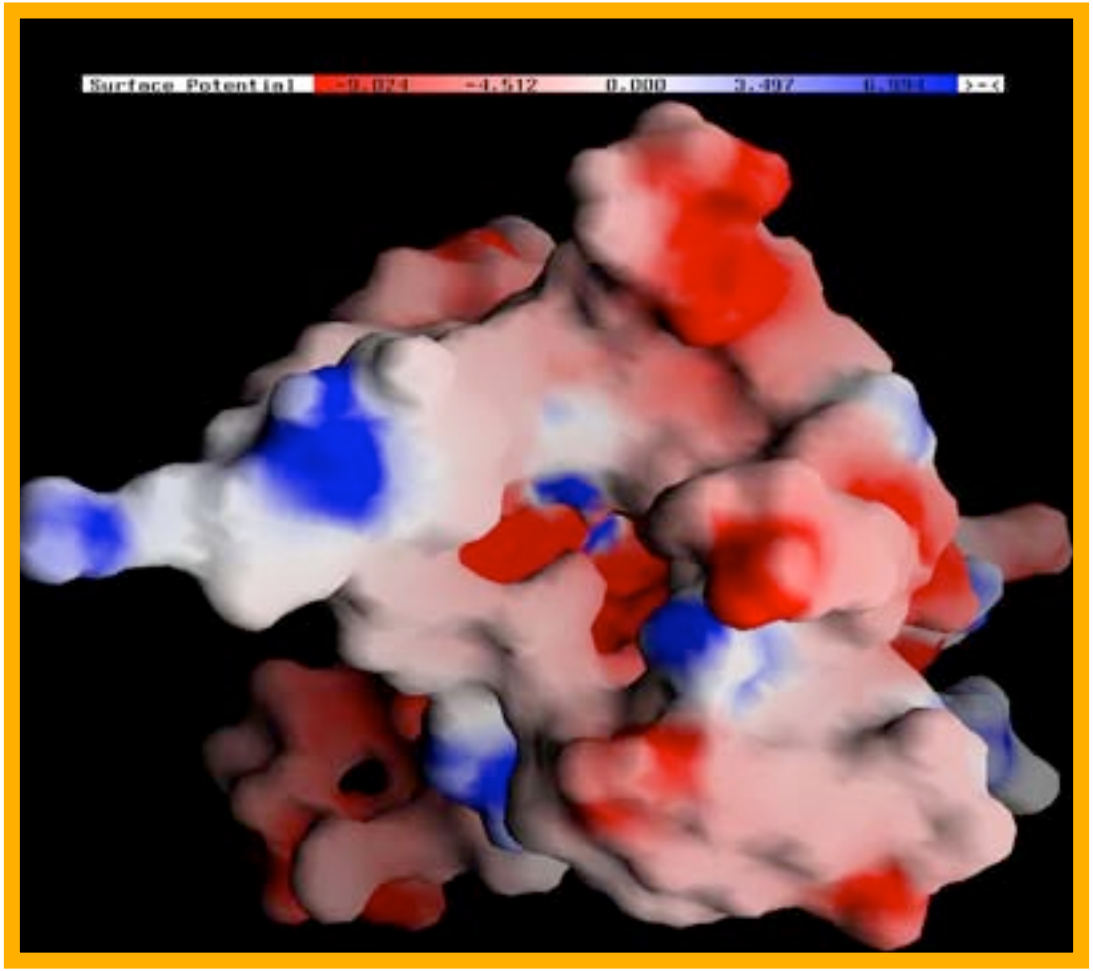


# Rv0813c, a FABP-like fold

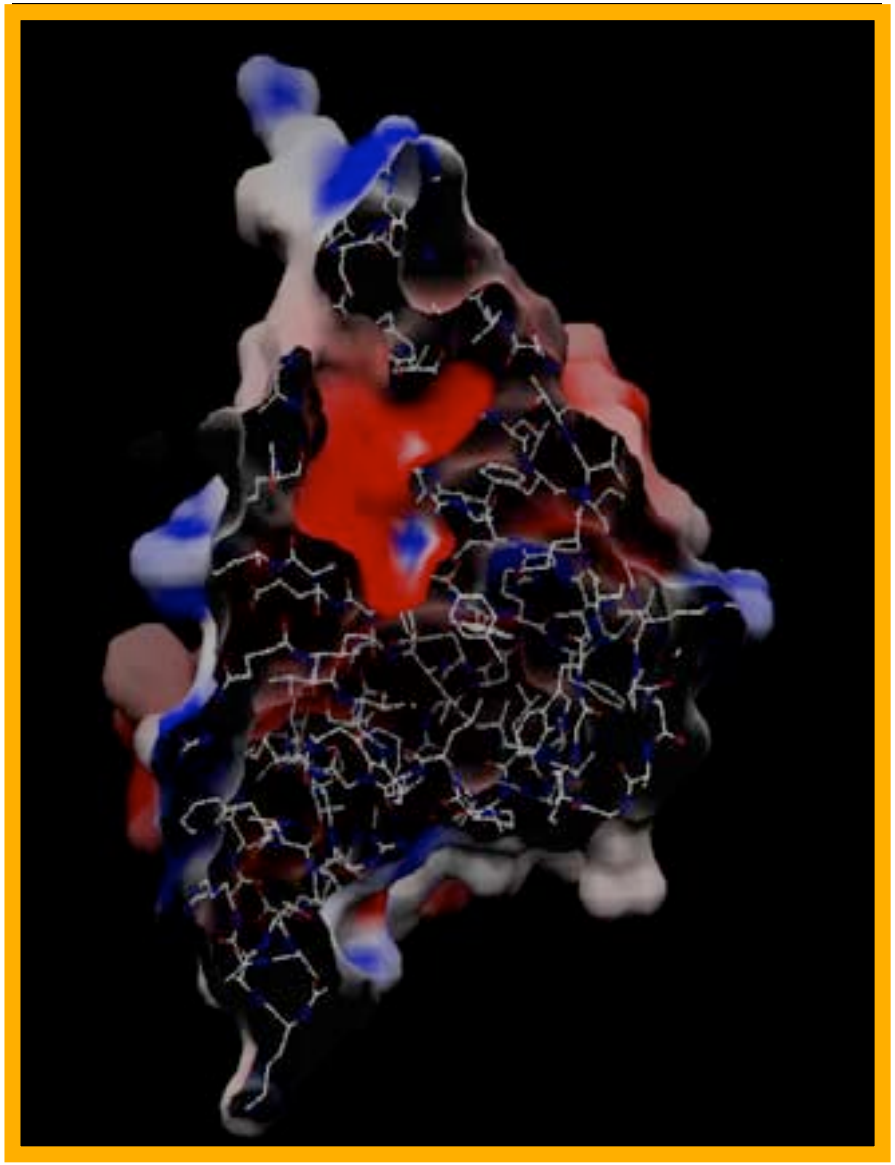


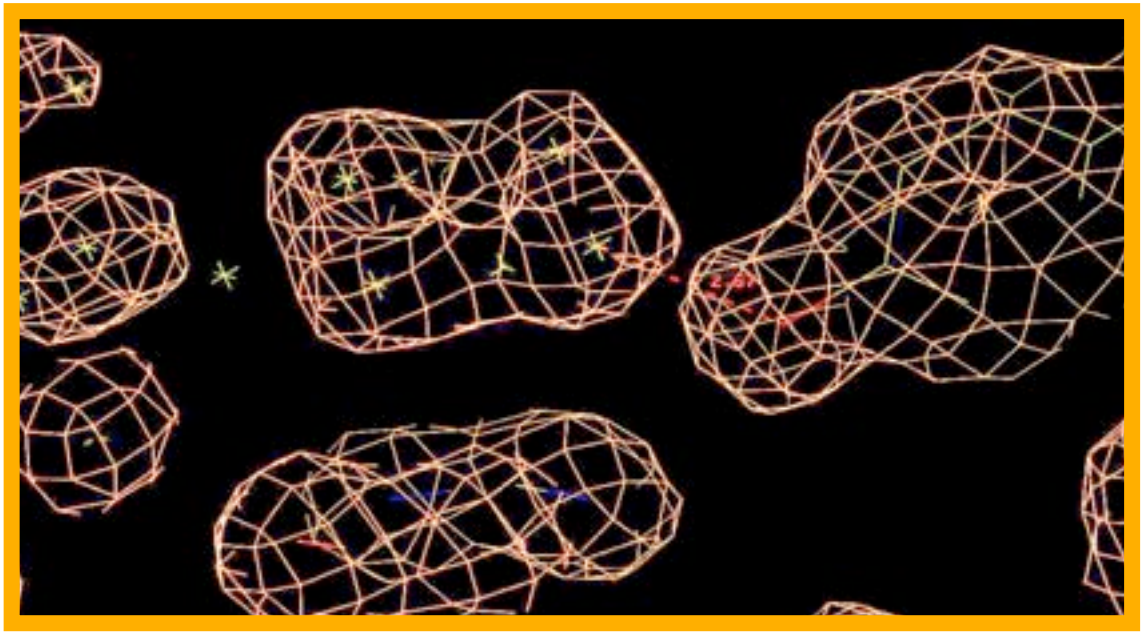
Side view

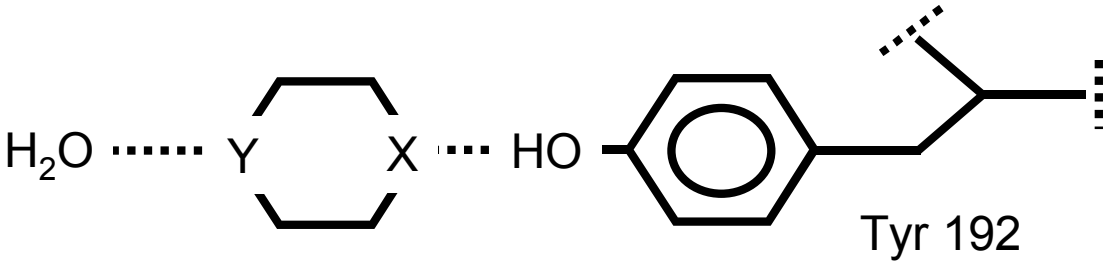



Top view

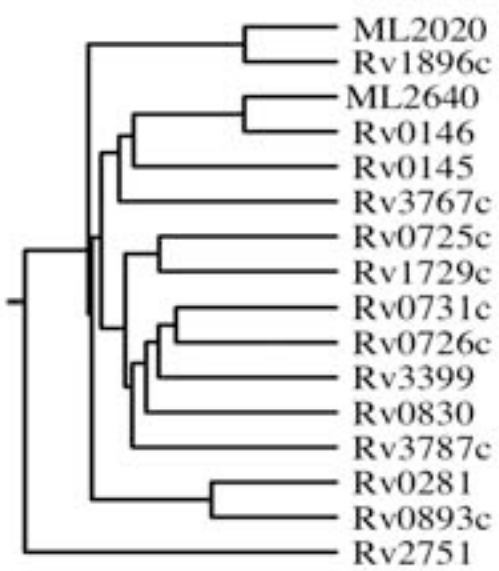
# Ligand Binding Pocket

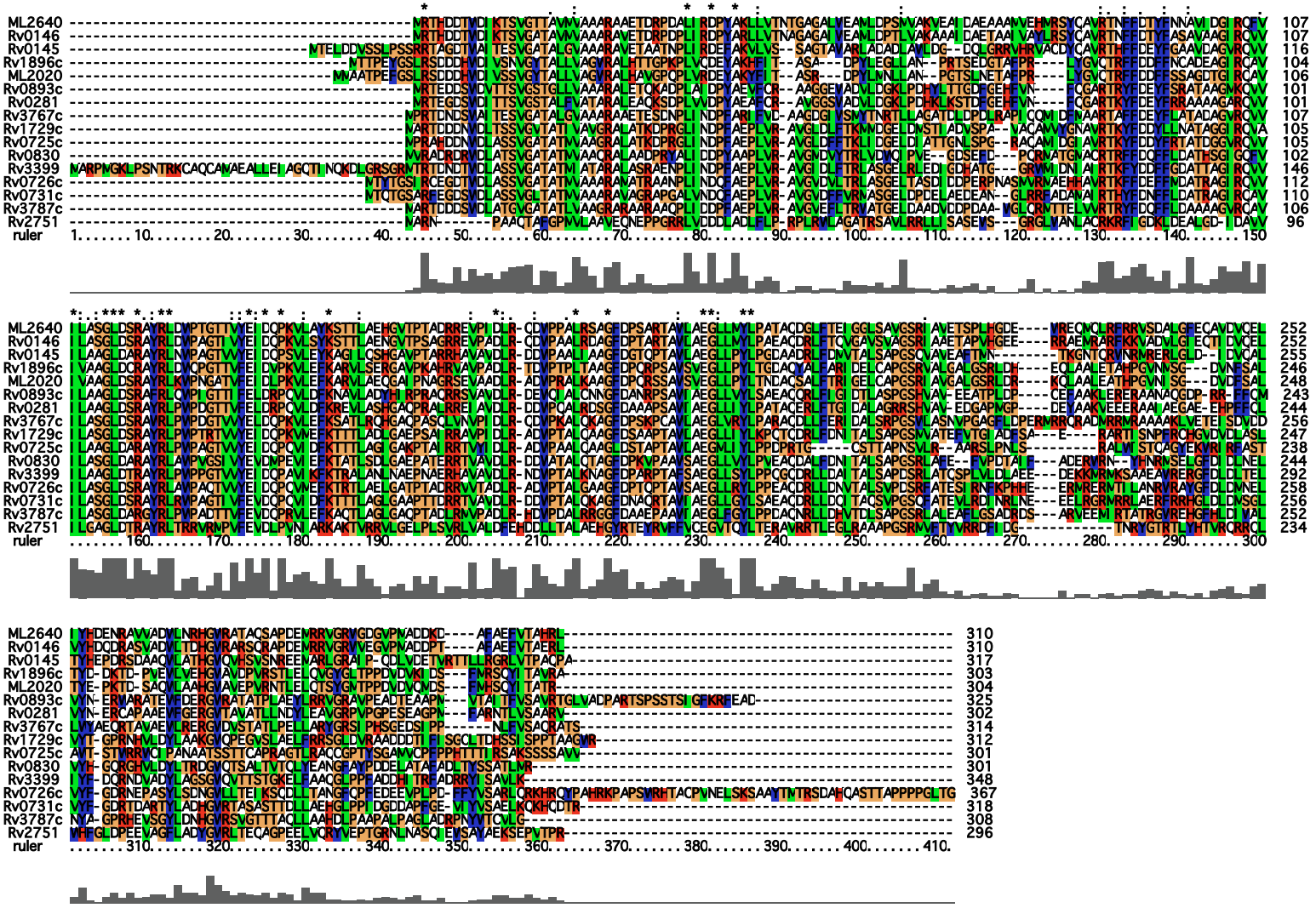




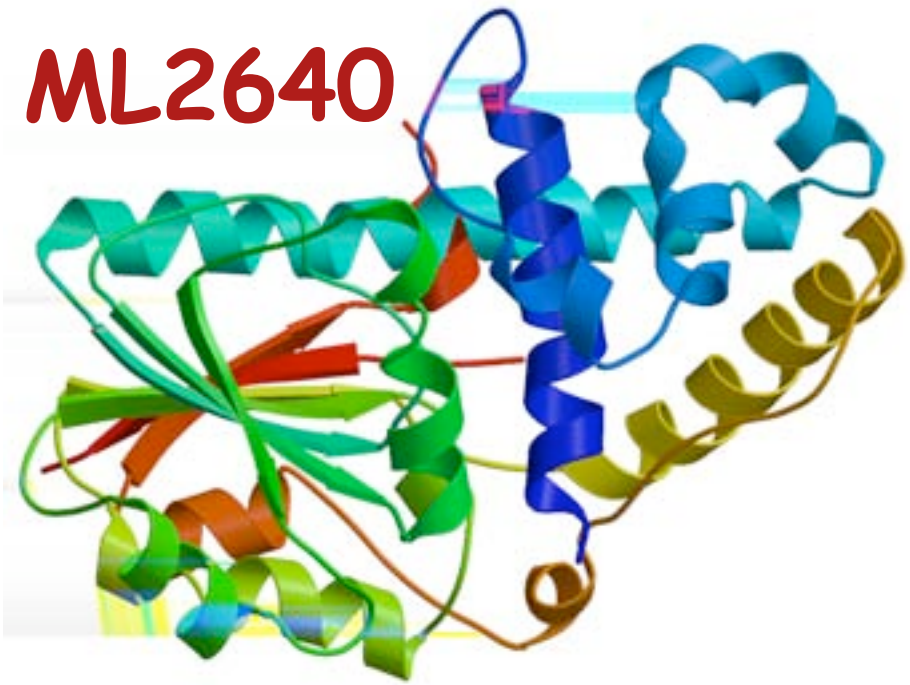


# ML2640 gene family

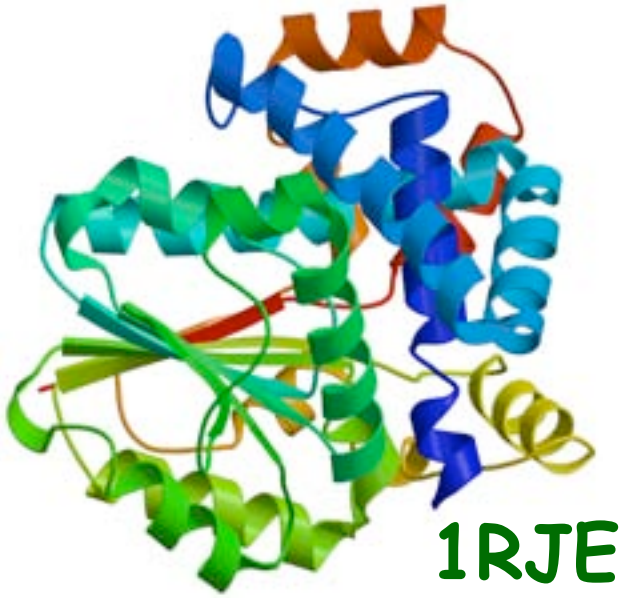


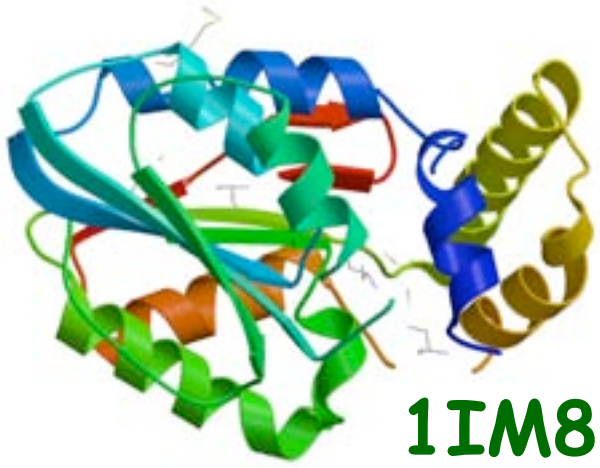
# ML2640



| PDB | Equiv. residues | r.m.s. deviation | SSE |
|---|---|---|---|
| 1RJE | 218 | 2.2 Å | ~70% |
| 1IM8 | 173 | 3.1 Å | ~50% |
| 1H1D | 153 | 2.8 Å | ~50% |



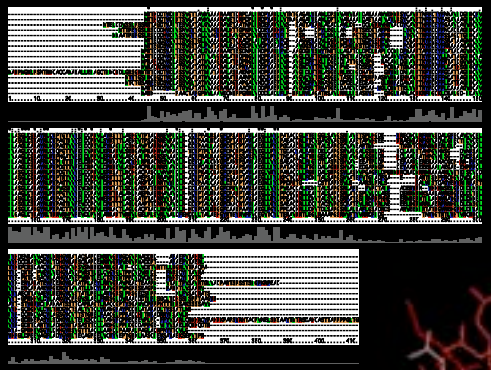

Leu C-methyltransferase (PPM1)

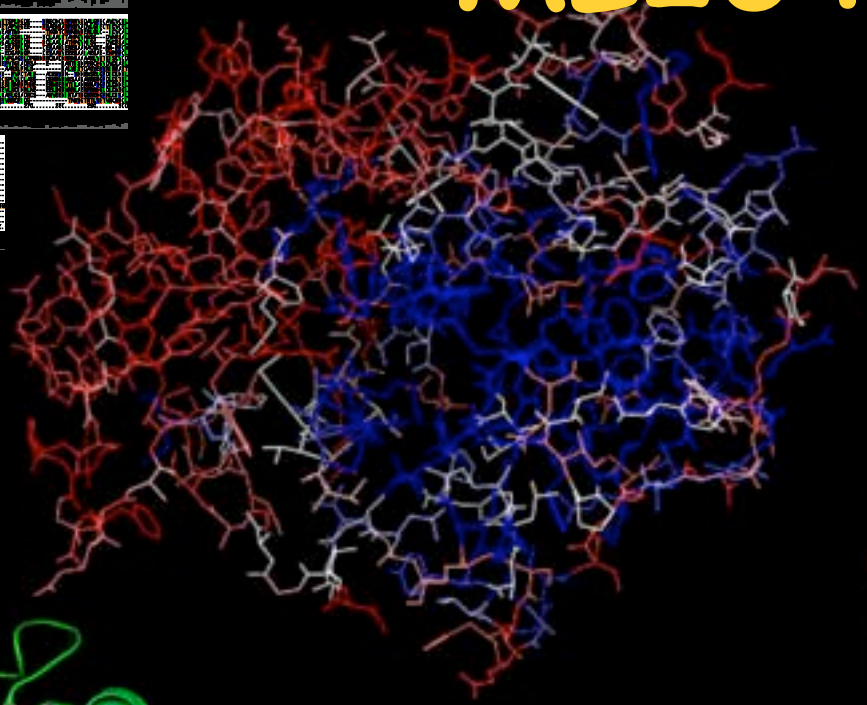


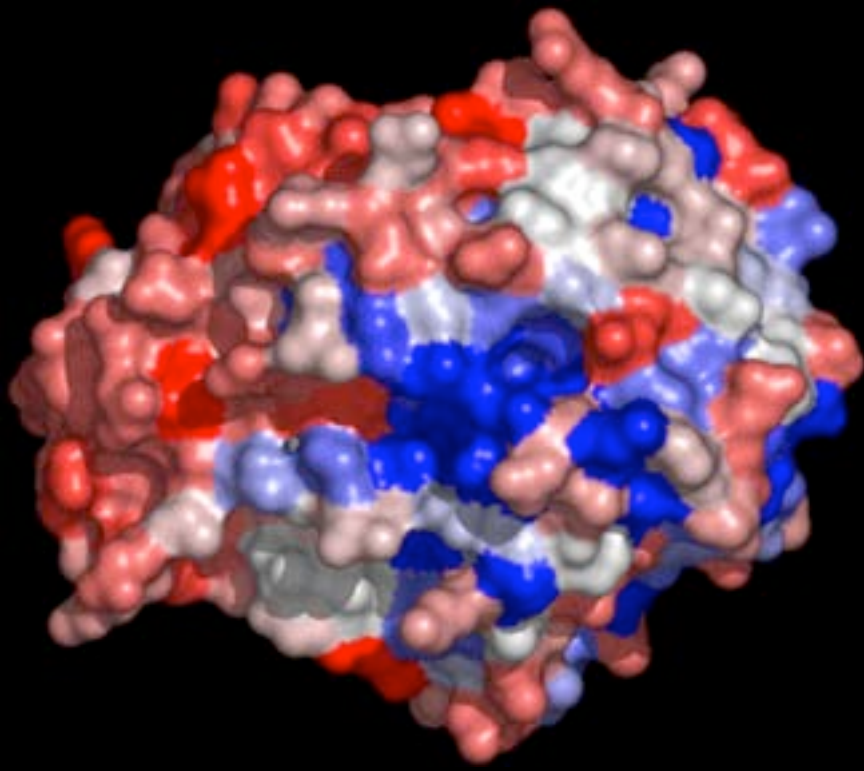

YecO *H. influenzae* (AdoMet)

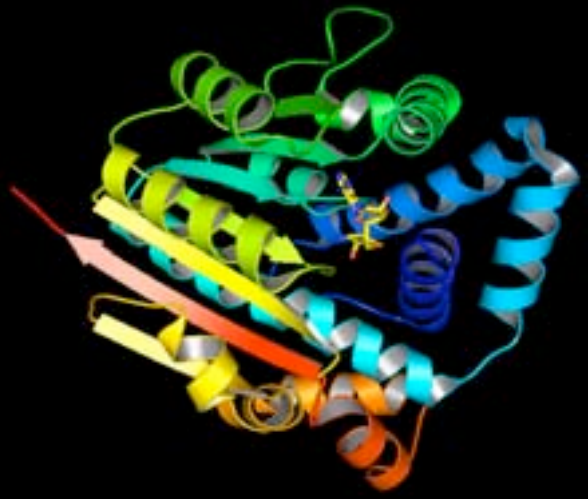

Catechol O-methyltransferase (rat)

# ML2640







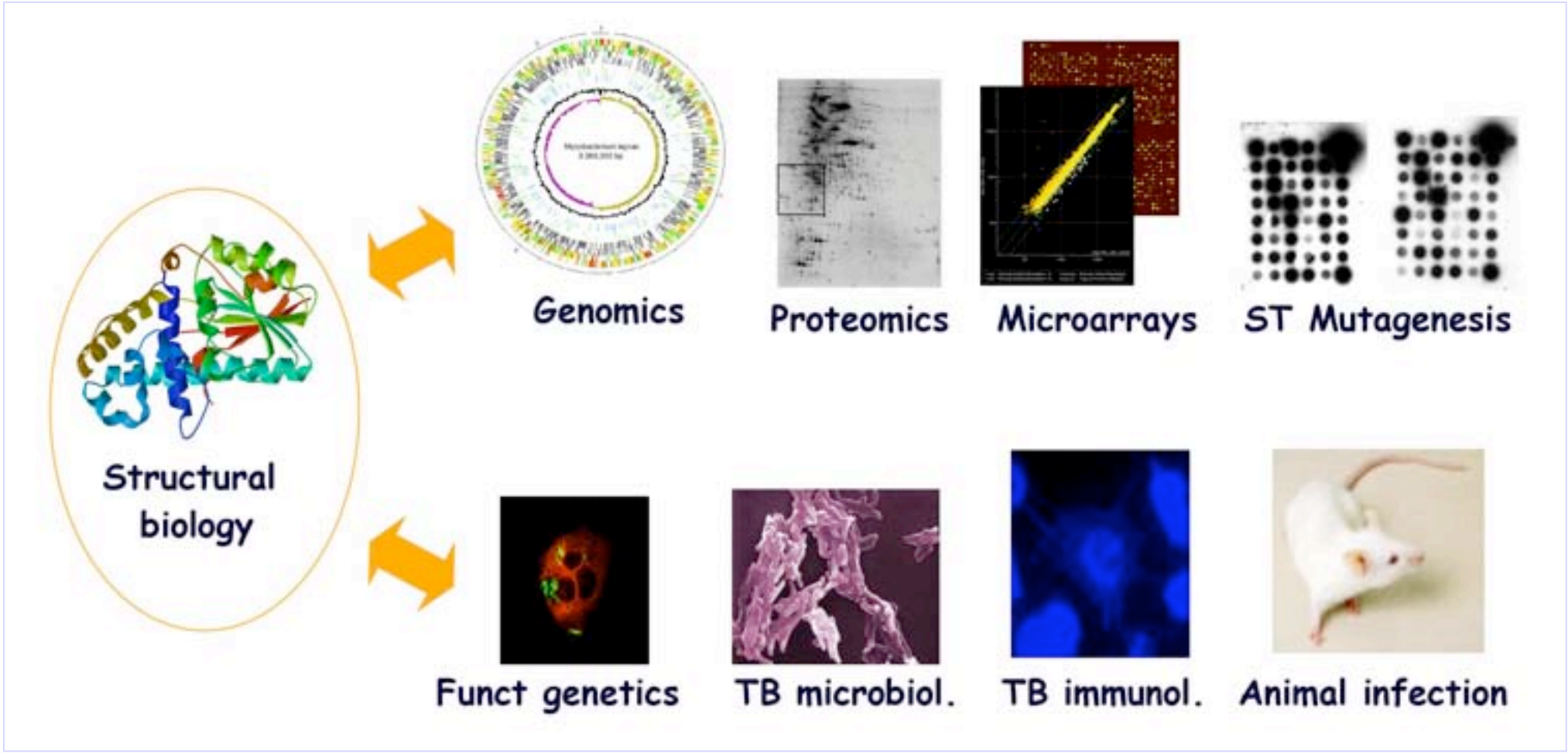


**1RJE**

Leulliot *et al, J.Biol.Chem.*, 2004

In progress: search for TB substrate(s)

# GPH - Tuberculose (IP)



# Structural Genomics

New unique folds

Reduced costs

Technology developments

Functional annotation

Drug discovery

Opportunistic approach

Biological relevance of structures

Protein complexes

Methodology bottlenecks (membrane proteins)

# Acknowledgements


**The pipeline**

Jacques Bellalou
Vincent Bondet
Cedric Fiez-Vandal
Fabrice Guillemot
Ahmed Haouz
Nadine Honoré
Stéphane Petres
Florence Proux

Bill Shepard (ESRF)

**Research labs**

Stewart Cole (UGMB, IP) Mycobacterial genomics

Brigitte Gicquel (UGM, IP) Gene essentiality

Jean M. Betton (URMP, IP) Protein expression

Muriel Delepierre (URMN, IP) NMR

Michael Nilges (UBS, IP) Structural bioinformatics

Pedro M. Alzari (UBS, IP) Protein crystallography

**Funding: Inst. Pasteur, MENRT, French Genopole, X-TB, SPINE**