



**HAL**  
open science

## Towards the structural screening of microbial ecosystems : selected projects and MetaFoldScan presentation

Gwenaëlle André-Leroux

► **To cite this version:**

Gwenaëlle André-Leroux. Towards the structural screening of microbial ecosystems : selected projects and MetaFoldScan presentation. Séminaire, Dec 2018, Montevideo, Uruguay. hal-02785774

**HAL Id: hal-02785774**

**<https://hal.inrae.fr/hal-02785774>**

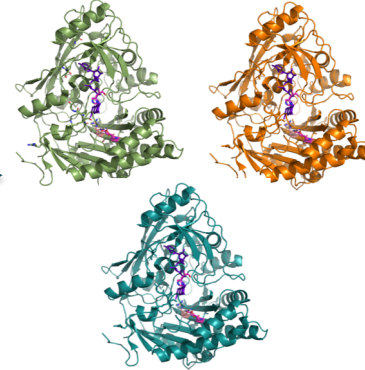
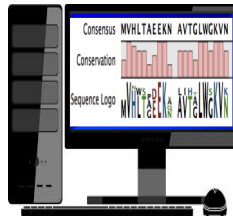
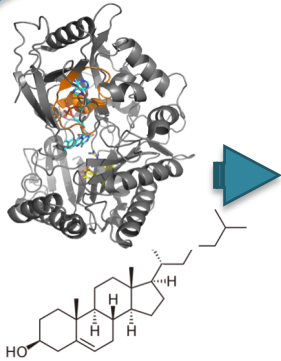
Submitted on 4 Jun 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Towards the Structural Screening of Microbial Ecosystems

Selected projets and MetaFoldScan presentation

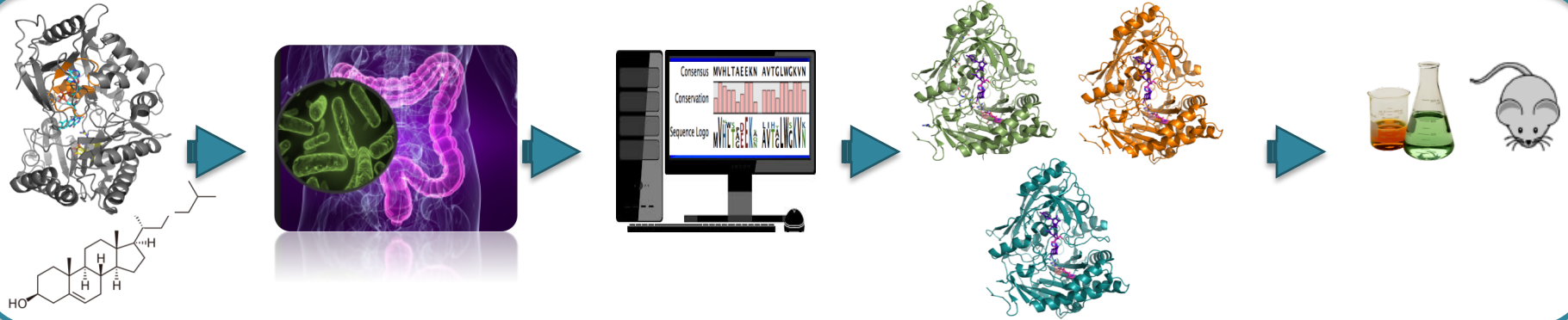


Gwenaëlle André-Leroux

Montevideo University of Sciences December 2018

# Towards the Structural Screening of Microbial Ecosystems

Selected projects and MetaFoldScan presentation

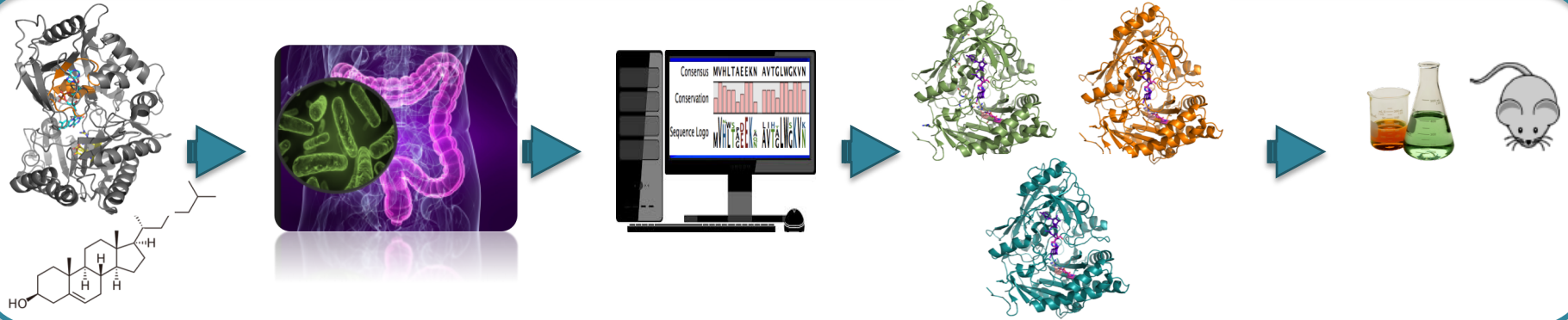


Outline

- Cholesterol conversion in the gut microbiota: the mystery enzyme(s) ?
- MetaFoldScan project : 3D screening of the gut microbiota
- Conclusions and perspectives

# Towards the Structural Screening of Microbial Ecosystems

Selected projects and MetaFoldScan presentation



Outline

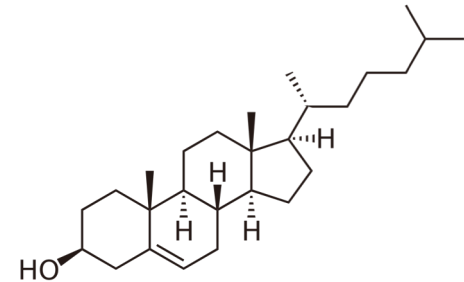
- Cholesterol conversion in the gut microbiota: the mystery enzyme(s) ?
- MetaFoldScan project : 3D screening of the gut microbiota
- Conclusions and perspectives

## ★ Cholesterol :

- Is a sterol that plays a central role as cell-membrane agent.
- Is precursor of steroid hormone (progesterone, testosterone, cortisone) and bile salt.
- Originates 30% from diet vs 70 % from bile and desquamated sterol of the gut epithelium.

## ★ Cholesterol level and gut microbiota :

- Can be metabolized by colonic bacteria.
- The gut microbiota reduces cholesterol to coprostanol.
- Neomycin impacts serum cholesterol and fecal sterol in hypercholesterolemic patients.



## ★ Cholesterol is absorbed by human intestine, not coprostanol :

- Cholesterol conversion relates to abundance of bacteria with cholesterol reducing activities.
- Gérard et al, 2004, Veiga et al, 2005
- Cholesterol conversion to coprostanol could result in lower cholesterolemia.

★ One strain from hog-sewage lagoon:

Gram+ coccobacillus *Eubacterium coprostanoligenes* ATCC 51222T Freier *et al*, 1993

Oral administration of *E. copro* → significant decrease of plasma cholesterol in dietary induced hypercholesterolemic rabbits Li *et al*, 1995

★ One from « high converter » human:

Isolate from the *Bacteroides* phylum, *Bacteroides* sp. Strain D8. Gram- anaerobie. Gérard *et al*, 2007

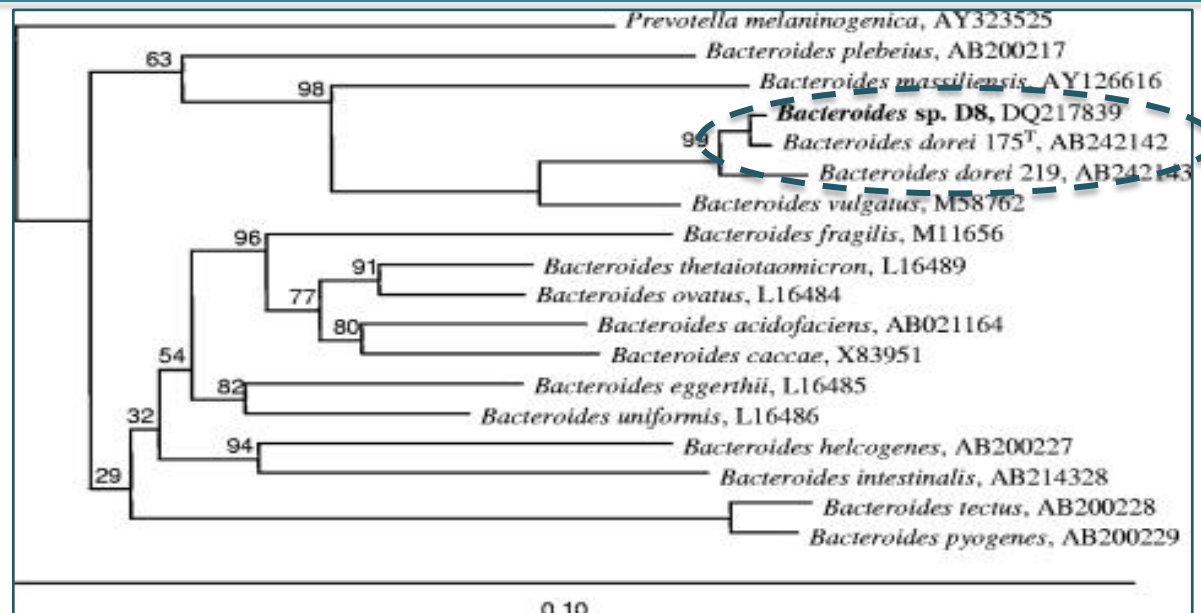
Phylogenetic tree construction → this strain clusters in an independant clade with the two isolates of *Bacteroides dorei* species.

**Intestinal microbiota**  
**>1,000 genera of bacteria.**

**2 phyla**

Bacteroidetes

Firmicutes



Phylogenetically close

Phylogenetically remote

Strain	Medium	SBM	BCM
<i>B. Dorei 175<sup>T</sup></i>		-	-
<i>B. Dorei D8</i>		+	-
<i>E. coprostanoligenes</i>		-	+

SBM : Standard Brain Medium  
 BCM: Basal Cholesterol Medium

Coll. C. Juste and P. Gérard, Micalis



## Metabolic pathway for cholesterol to coprostanol conversion in *B. dorei D8*?



### Which are the enzymes of the catabolic pathway that degrade cholesterol into coprostanol?

★ One strain from hog-sewage lagoon:

Gram+ coccobacillus *Eubacterium coprostanoligenes* ATCC 51222T Freier *et al*, 1993

Oral administration of *E. copro* → significant decrease of plasma cholesterol in dietary induced hypercholesterolemic rabbits Li *et al*, 1995

★ One from « high converter » human:

Isolate from the *Bacteroides phylum*, *Bacteroides sp. Strain D8*. Gram- anaerobie. Gérard *et al*, 2007

Phylogenetic tree construction → this strain clusters in an independant clade with the two isolates of *Bacteroides dorei* species.

Microbial genes or enzymes involved in cholesterol metabolism in the gut are unknown.

However ...

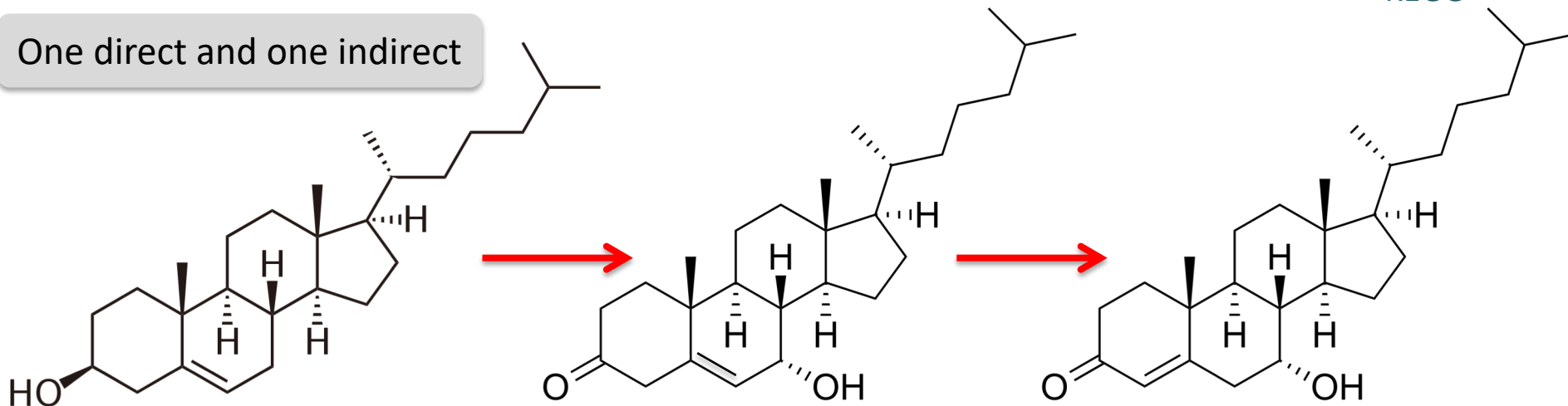
Genomes of *B. Dorei*D8 & *B. Dorei*175<sup>T</sup> sequenced and assembled at MaiAGE

Valentin Loux



KEGG

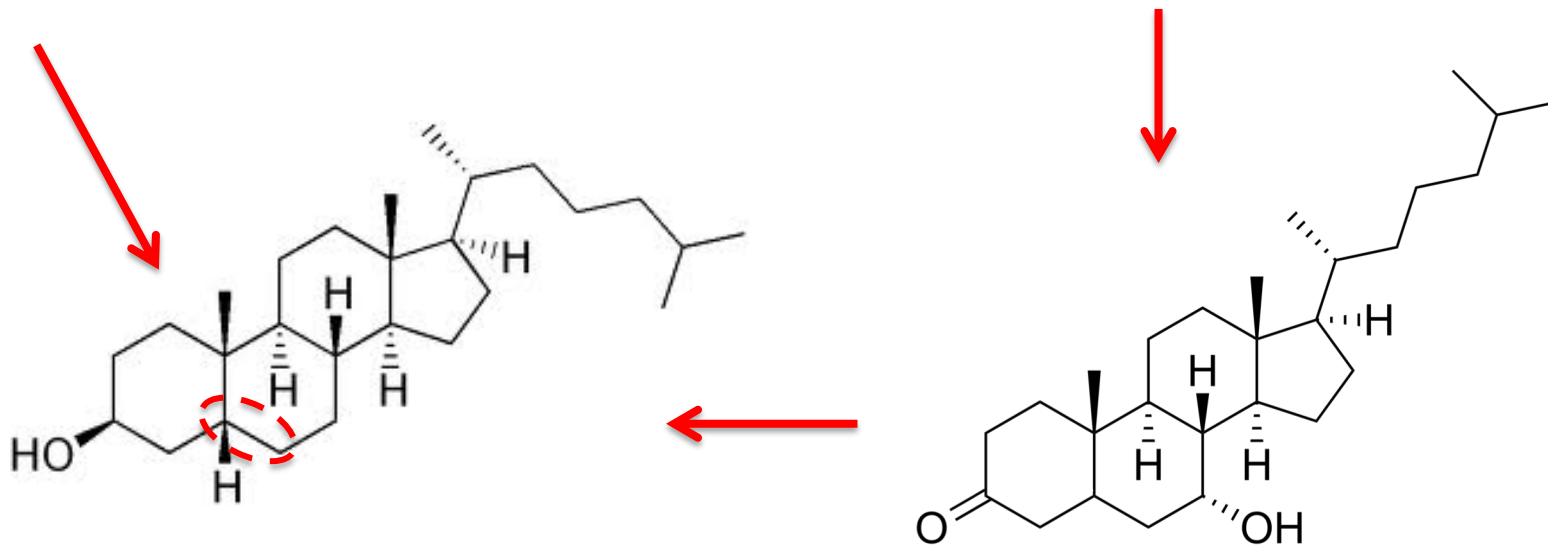
One direct and one indirect



Cholesterol

5-Cholesten-3-one transient

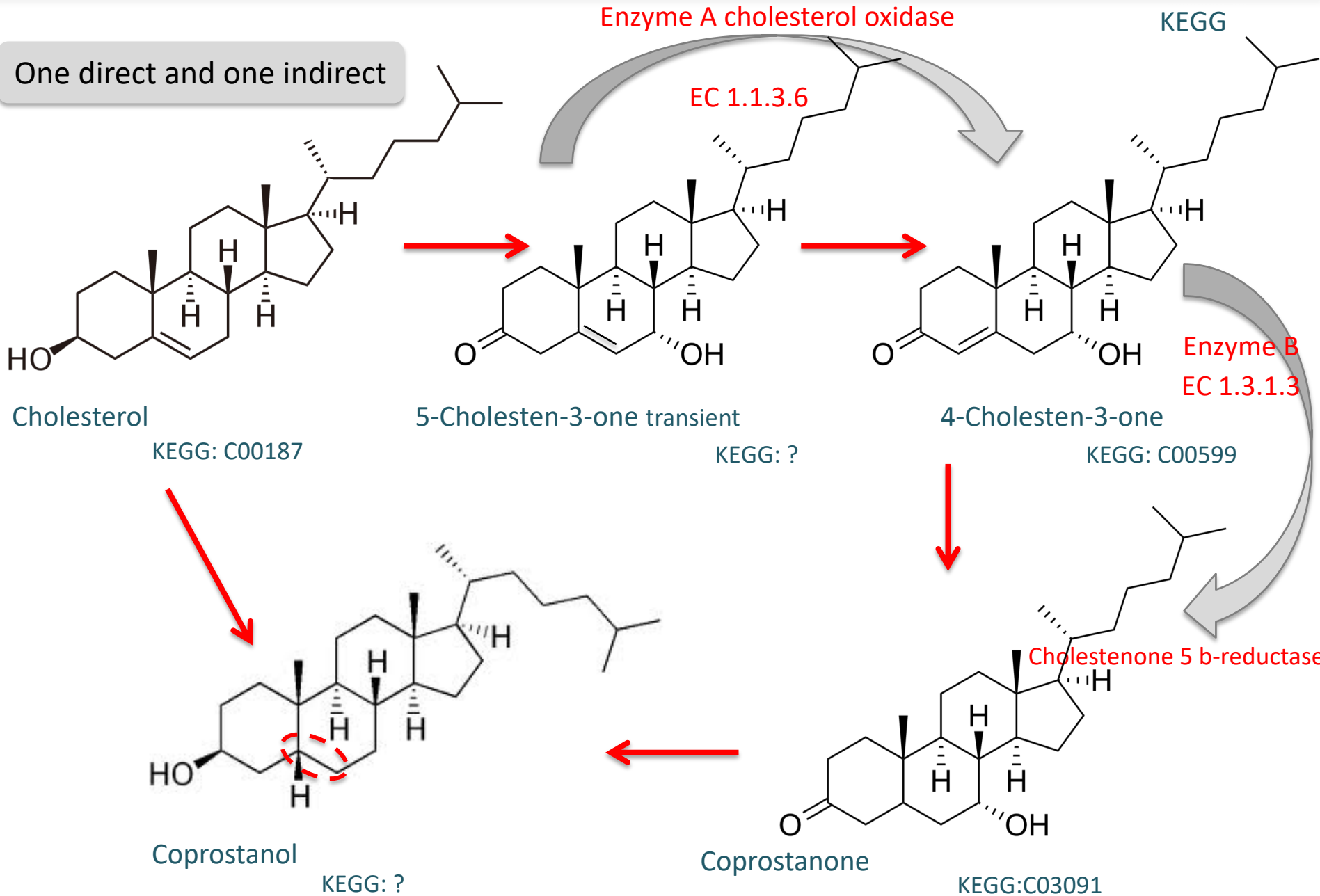
4-Cholesten-3-one



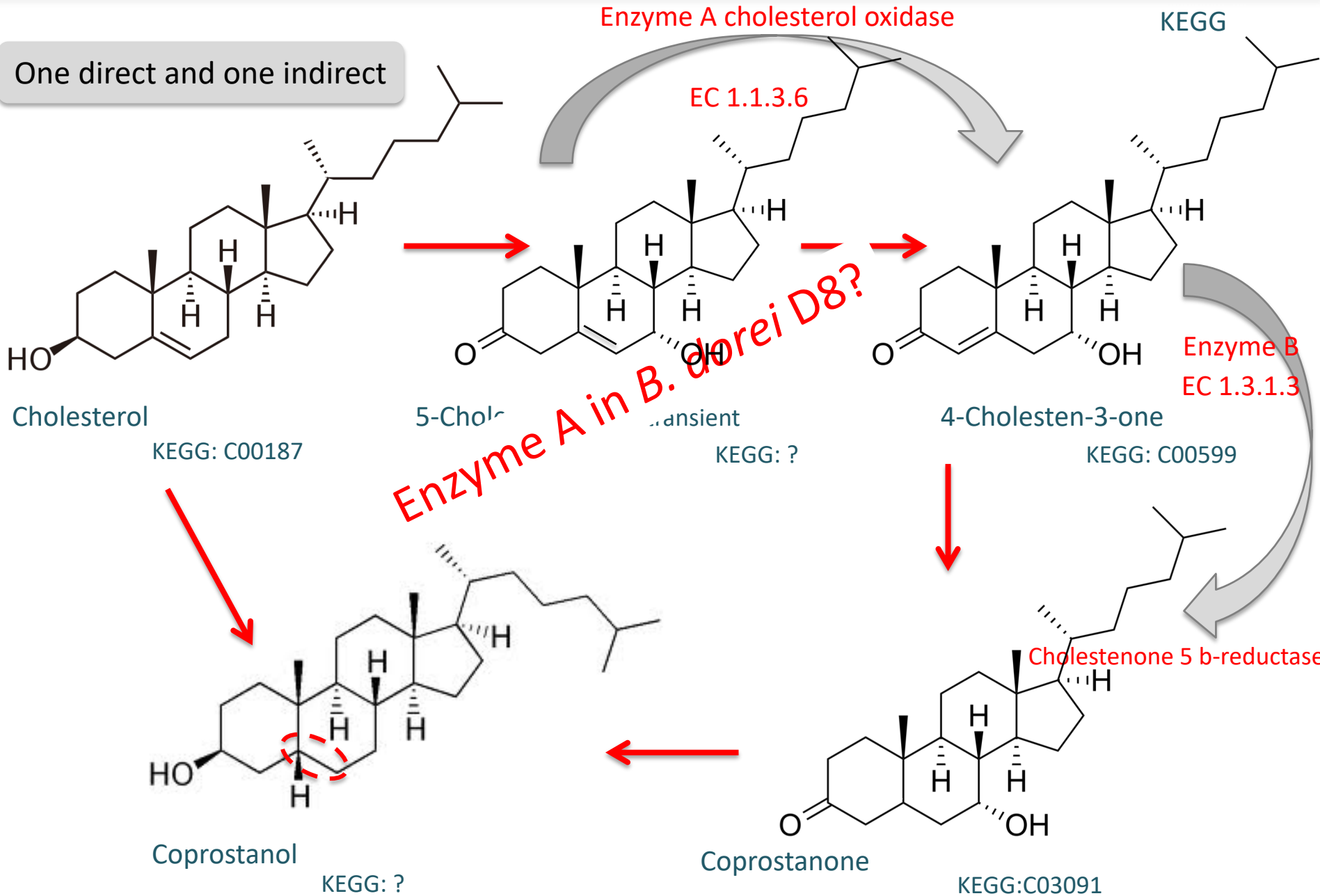
Coprostanol

Coprostanone

One direct and one indirect

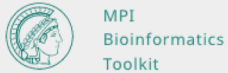


One direct and one indirect



Server of protein structure prediction

<http://toolkit.tuebingen.mpg.de/hhpred>



- Search
- Alignment
- Sequence Analysis
- 2ary Structure
- 3ary Structure
- Classification
- Utils

- HHblits
- HHpred
- HMMER
- PatternSearch
- ProtBLAST/PSI-BLAST

## HHpred

ID	Date	Tool	
7456410		HHPR	X
1681937		HHPR	X
5162681		HHPR	X
6311127		HHPR	X
8160192		HHPR	X

Input Parameters

Enter a protein sequence/multiple sequence alignment in FASTA/CLUSTAL format

To create a structural model of your query protein, run a HHpred search with the HH-suite database, select the top-scoring template(s) and click on 'Create model using top-scoring template(s)'. You can also generate a PIR file that can be subsequently submitted to MODELLER.

Query : Fasta sequence

[Paste Example](#) [Upload File](#)

Align two sequences or MSAs

Select HH-Suite Database

× PDB\_mmCIF70\_11\_May

Proteomes

Custom JobID

Submit Job

Protein data bank

### Profile-profile comparison tools:

- Calculated from a MSA of related sequences collected from Hhblits (>Psi-blast+).
- Matrix of similarity scores calculated from frequency of aa at the corresponding positions in the MSA.

Profile contains more information than a sequence.



Result



Hits: Rank of « matching » pdb proteins

- Prediction of MSA profile with SS for each protein of *B. Dorei 175T* & *B. dorei D8* genomes
- Compare each protein with a bank of cholesterol oxidase « Enzyme A » to be assessed

- 1- Split of each fasta sequence in the genome -> more than 3,400 for each genome of *B. dorei*
- 2- Multiple sequence alignment for each fasta: 'HHblits' > PSI-Blast
- 3- Prediction & addition of SS elements: 'adssl.pl'
- 4- Profiling & comparaison with the 3D template « bank »: 'HHBlitsdb.pl' & 'HHsearch'
- 5- Extraction, ranking and analyse of scores.
- 6- Analyse of selected sequences of proteins. Annotation? Fonction?
- 7- *In fine*, a list of cholesterol oxidase putative proteins to test *in vitro* / *in vivo*

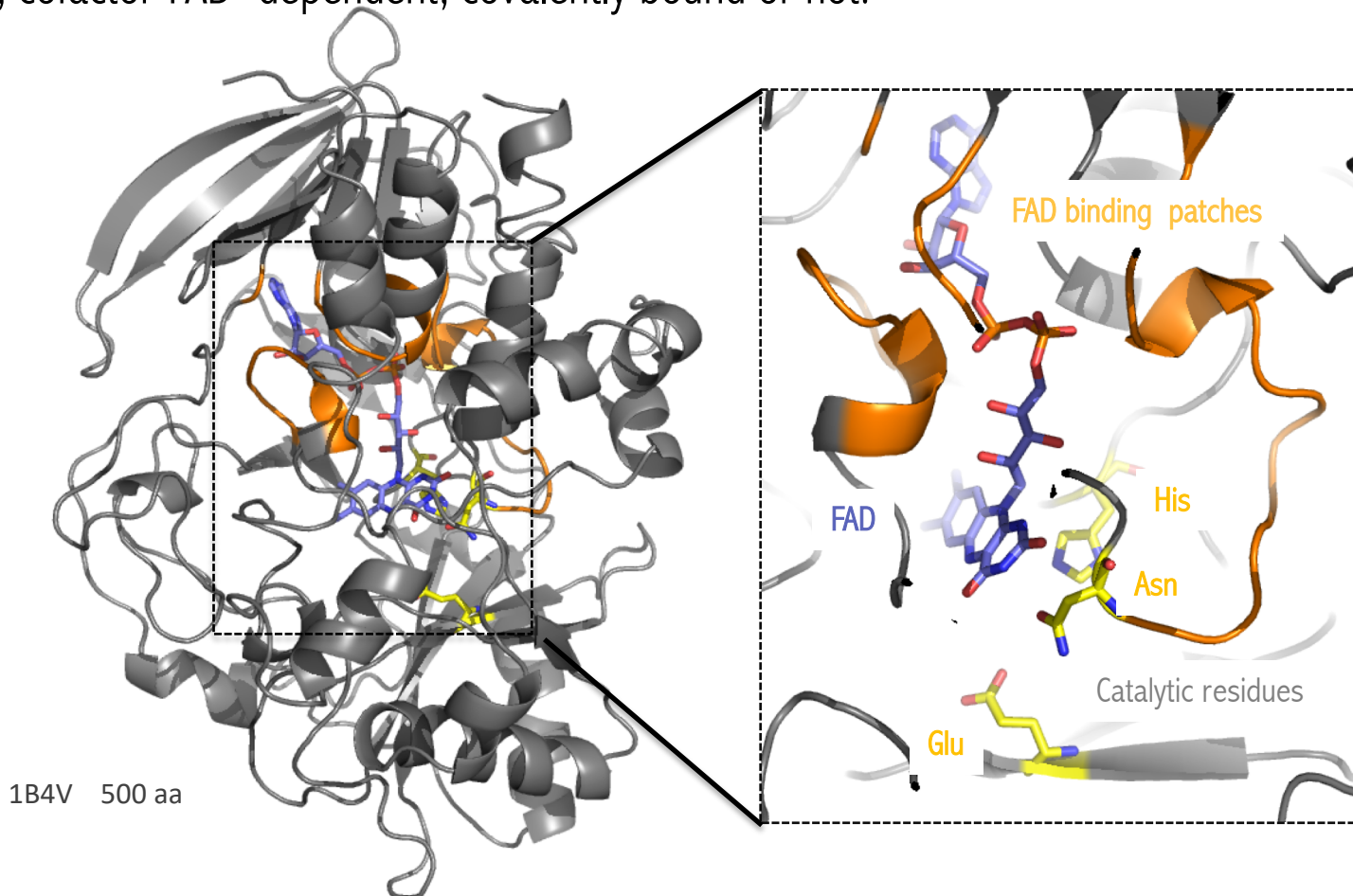


The 3D template « bank » is critical

## Cholesterol oxidase:

- monomeric flavoenzyme catalyzes oxidation and isomerization into cholest-4-en-3-one.
- two folds exist, cofactor-FAD- dependent, covalently bound or not.

Catalytic site  
Asn, Glu, His



1B4V. pdb & fasta

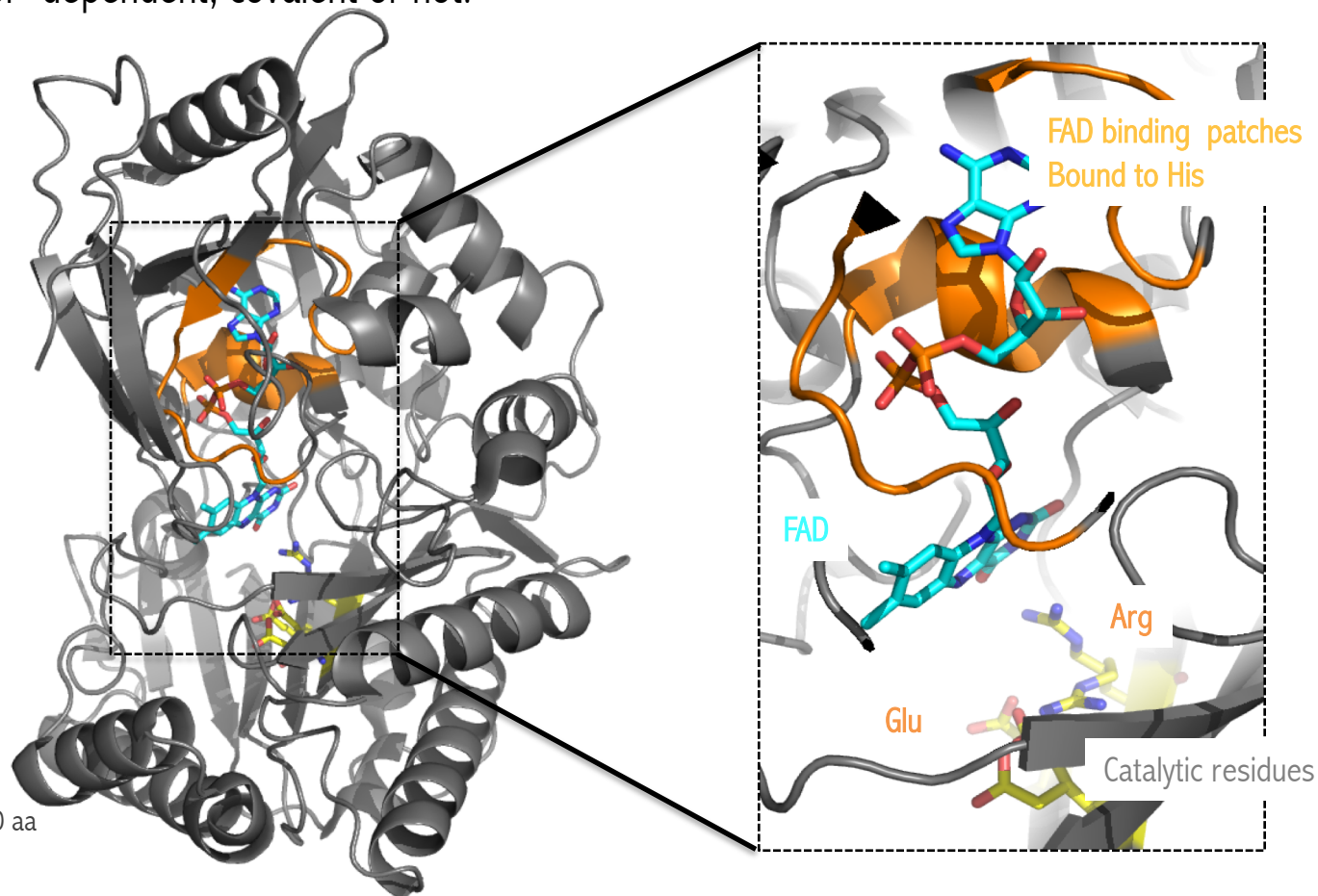
Cholesterol oxydase from *Streptomyces* Yue et al, 1999 - 1.5 Å

1COY.pdb & fasta

Cholesterol oxydase from *R. hoagii* ex *B. steroliticum* Li et al, 1993 - 1.8 Å

## Cholesterol oxidase:

- monomeric flavoenzyme catalyzes oxidation and isomerization into cholest-4-en-3-one.
- two folds exist, cofactor- dependent, covalent or not.



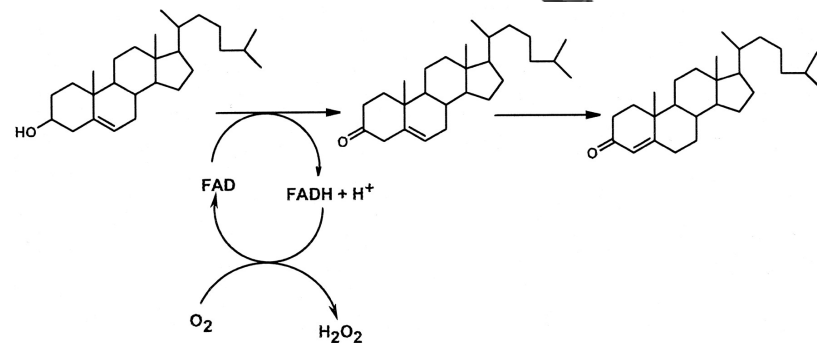
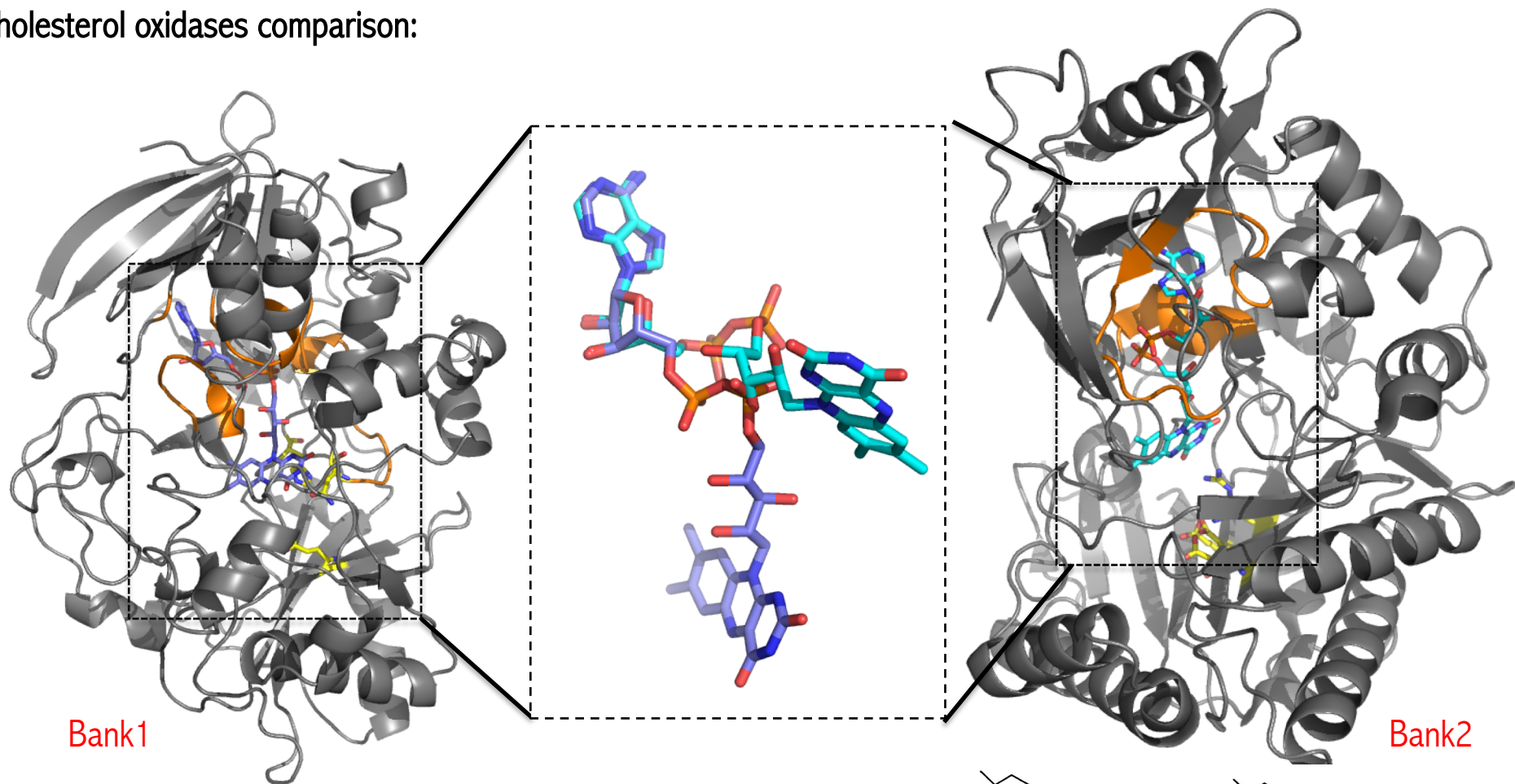
3JS8.pdb & fasta

1119.pdb & fasta

Cholesterol oxydase from *Chromobacterium sp. DS-1* Sagermann et al, 2010 - 1.5 Å

Cholesterol oxydase from *B. steroliticum* Coulombe et al, 2001 - 1.7 Å

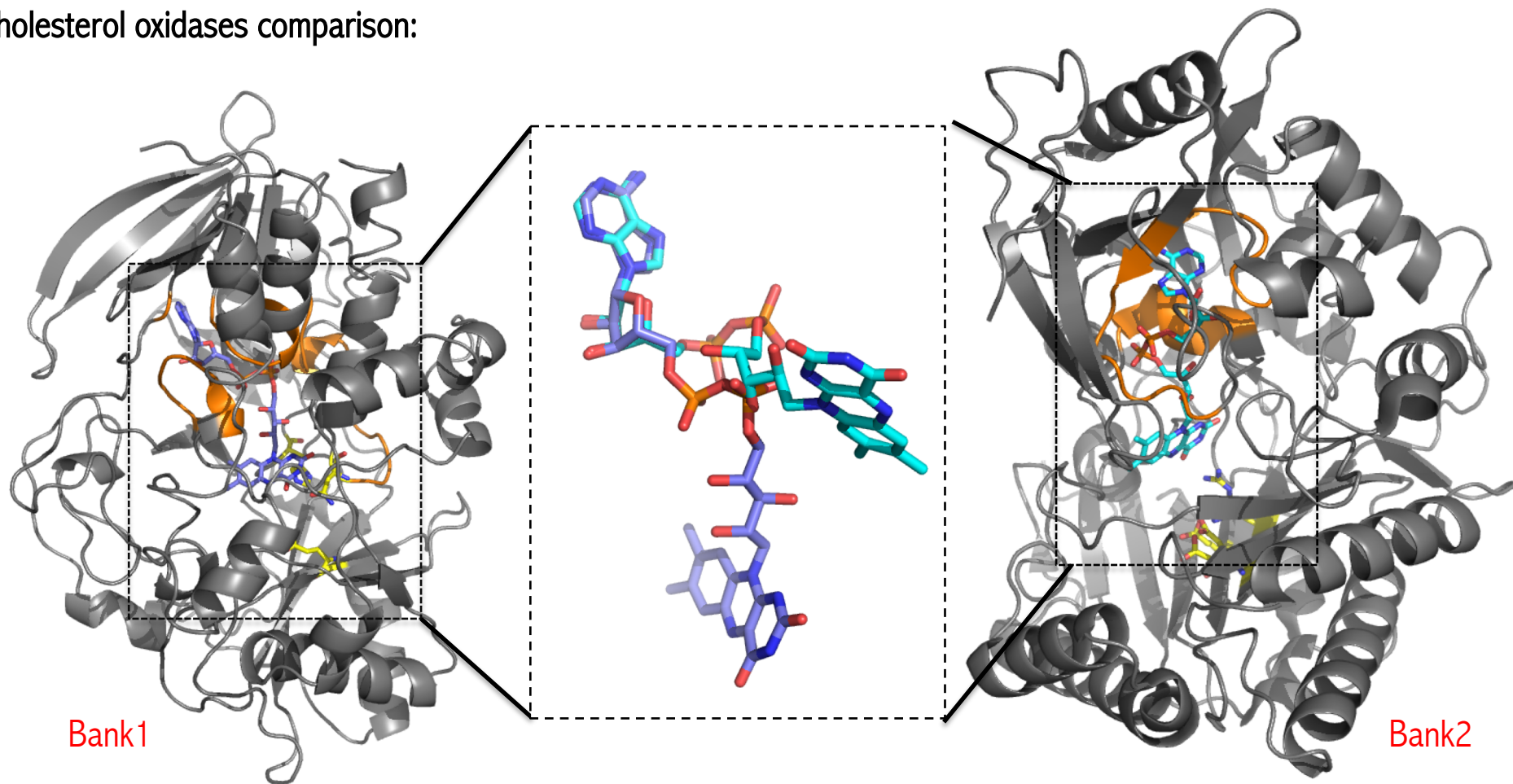
## Cholesterol oxidases comparison:



- ★ 3D structures distinct
- ★ Catalytic sites different
- ★ Binding patches divergent
- ★ FAD binding topologies contrasting despite same redox role



Cholesterol oxidases comparison:



- ★ 3D structures distinct
- ★ Catalytic sites different
- ★ Binding patches divergent
- ★ FAD binding topologies contrasting despite same redox role



Structural detection with 2 banks required

## ★ *B. dorei*D8 Sorted by Probability – Bank1

37	92.9	5=32	1COY:A	/projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS/BANK1/bdd8_01957.fasta.a3m.hhm.out.hhmakeBK1
33	93.3	2=31	1B4V:A	/projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS/BANK1/bdd8_03587.fasta.a3m.hhm.out.hhmakeBK1
32	94.2	3=29	1B4V:A	/projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS/BANK1/bdd8_00872.fasta.a3m.hhm.out.hhmakeBK1
241	97.1	209=32	1B4V:A	/projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS/BANK1/bdd8_00845.fasta.a3m.hhm.out.hhmakeBK1
479	97.2	441=38	1COY:A	/projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS/BANK1/bdd8_04063.fasta.a3m.hhm.out.hhmakeBK1
179	97.5	142=37	1COY:A	/projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS/BANK1/bdd8_01842.fasta.a3m.hhm.out.hhmakeBK1
49	97.6	11=38	1COY:A	/projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS/BANK1/bdd8_02168.fasta.a3m.hhm.out.hhmakeBK1
256	97.6	195=61	1B4V:A	/projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS/BANK1/bdd8_03122.fasta.a3m.hhm.out.hhmakeBK1
35	97.6	2=33	1B4V:A	/projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS/BANK1/bdd8_01099.fasta.a3m.hhm.out.hhmakeBK1
35	97.8	2=33	1B4V:A	/projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS/BANK1/bdd8_00348.fasta.a3m.hhm.out.hhmakeBK1
37	97.9	4=33	1COY:A	/projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS/BANK1/bdd8_04051.fasta.a3m.hhm.out.hhmakeBK1
37	98.0	2=35	1COY:A	/projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS/BANK1/bdd8_00992.fasta.a3m.hhm.out.hhmakeBK1
66	98.1	29=37	1COY:A	/projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS/BANK1/bdd8_04457.fasta.a3m.hhm.out.hhmakeBK1
37	98.2	1=36	1COY:A	/projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS/BANK1/bdd8_01766.fasta.a3m.hhm.out.hhmakeBK1
40	98.2	3=37	1COY:A	/projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS/BANK1/bdd8_00293.fasta.a3m.hhm.out.hhmakeBK1
41	98.2	7=34	1COY:A	/projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS/BANK1/bdd8_01761.fasta.a3m.hhm.out.hhmakeBK1
163	98.7	104=59	1B4V:A	/projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS/BANK1/bdd8_00882.fasta.a3m.hhm.out.hhmakeBK1
118	98.9	81=37	1B4V:A	/projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS/BANK1/bdd8_01104.fasta.a3m.hhm.out.hhmakeBK1
208	99.1	133=75	1B4V:A	/projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS/BANK1/bdd8_00352.fasta.a3m.hhm.out.hhmakeBK1
237	99.1	175=62	1B4V:A	/projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS/BANK1/bdd8_03708.fasta.a3m.hhm.out.hhmakeBK1
280	99.3	211=60	1COY:A	/projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS/BANK1/bdd8_00254.fasta.a3m.hhm.out.hhmakeBK1

## ★ *B. dorei*D8 Sorted by Probability – Bank2

5.9	-	212-241	- 3JS8:A	- /projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS-BANK2/bdd8_01995.fasta.a3m.hhm.out.hhmakeBK2-
6.8	-	5-19	- 1I19:A	- /projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS-BANK2/bdd8_01730.fasta.a3m.hhm.out.hhmakeBK2-
7.0	-	26-52	- 3JS8:A	- /projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS-BANK2/bdd8_00608.fasta.a3m.hhm.out.hhmakeBK2-
7.2	-	237-266	- 3JS8:A	- /projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS-BANK2/bdd8_01989.fasta.a3m.hhm.out.hhmakeBK2-
7.3	-	5-40	- 3JS8:A	- /projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS-BANK2/bdd8_01278.fasta.a3m.hhm.out.hhmakeBK2-
7.7	-	135-164	- 3JS8:A	- /projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS-BANK2/bdd8_03492.fasta.a3m.hhm.out.hhmakeBK2-
7.7	-	9-28	- 3JS8:A	- /projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS-BANK2/bdd8_00561.fasta.a3m.hhm.out.hhmakeBK2-
8.2	-	233-262	- 3JS8:A	- /projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS-BANK2/bdd8_04010.fasta.a3m.hhm.out.hhmakeBK2-
8.3	-	9-28	- 3JS8:A	- /projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS-BANK2/bdd8_03809.fasta.a3m.hhm.out.hhmakeBK2-
9.1	-	9-26	- 3JS8:A	- /projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS-BANK2/bdd8_02053.fasta.a3m.hhm.out.hhmakeBK2-
9.9	-	9-28	- 3JS8:A	- /projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS-BANK2/bdd8_00931.fasta.a3m.hhm.out.hhmakeBK2-
15.5	-	364-406	- 3JS8:A	- /projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS-BANK2/bdd8_02744.fasta.a3m.hhm.out.hhmakeBK2-
18.0	-	6-23	- 3JS8:A	- /projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS-BANK2/bdd8_02160.fasta.a3m.hhm.out.hhmakeBK2-
99.9	-	18-190	- 1I19:A	- /projet/mig/work/vmartin/D8/D8.batch4.20150203/RESULTS-BANK2/bdd8_03618.fasta.a3m.hhm.out.hhmakeBK2-





★ *B. dorei* 175<sup>T</sup>

Sorted by Prob &gt; 95 and size cut off &gt; 50 amino acid residues

Véronique Martin

99.4 % 70 aa bacdor\_01751.fasta.a3m.hhm.out.hhmakeBK1 → Hypothetical protein  
99.1 % 75 aa bacdor\_01928.fasta.a3m.hhm.out.hhmakeBK1 → L-aspartate oxidase  
99.1 % 62 aa bacdor\_02089.fasta.a3m.hhm.out.hhmakeBK1 → Succinate Dehydrogenase flavoprotein  
97.5 % 61 aa bacdor\_03886.fasta.a3m.hhm.out.hhmakeBK1 → Pyridine nucleotide oxido reductase

★ *B. dorei* 175<sup>T</sup>

Sorted by Prob &gt; 95 and size &gt; 50 amino acid residues

99.9 % 172 aa bacdor\_03524.fasta.a3m.hhm.out.hhmakeBK2 → UDP-N-acetyl muramate dehydrogenase

★ *B. dorei* 175<sup>T</sup> Sorted by Prob > 95 and size cut off > 50 amino acid residues

99.4 % 70 aa baccor\_01751.fasta.a3m.hhm.out.hhmakeBK1 → Hypothetical protein  
99.1 % 75 aa baccor\_01928.fasta.a3m.hhm.out.hhmakeBK1 → L-aspartate oxidase  
99.1 % 62 aa baccor\_02089.fasta.a3m.hhm.out.hhmakeBK1 → Succinate Dehydrogenase flavoprotein  
97.5 % 61 aa baccor\_03886.fasta.a3m.hhm.out.hhmakeBK1 → Pyridine nucleotide oxido reductase

★ *B. dorei* 175<sup>T</sup> Sorted by Prob > 95 and size > 50 amino acid residues

99.9 % 172 aa baccor\_03524.fasta.a3m.hhm.out.hhmakeBK2 → UDP-N-acetyl muramate dehydrogenase

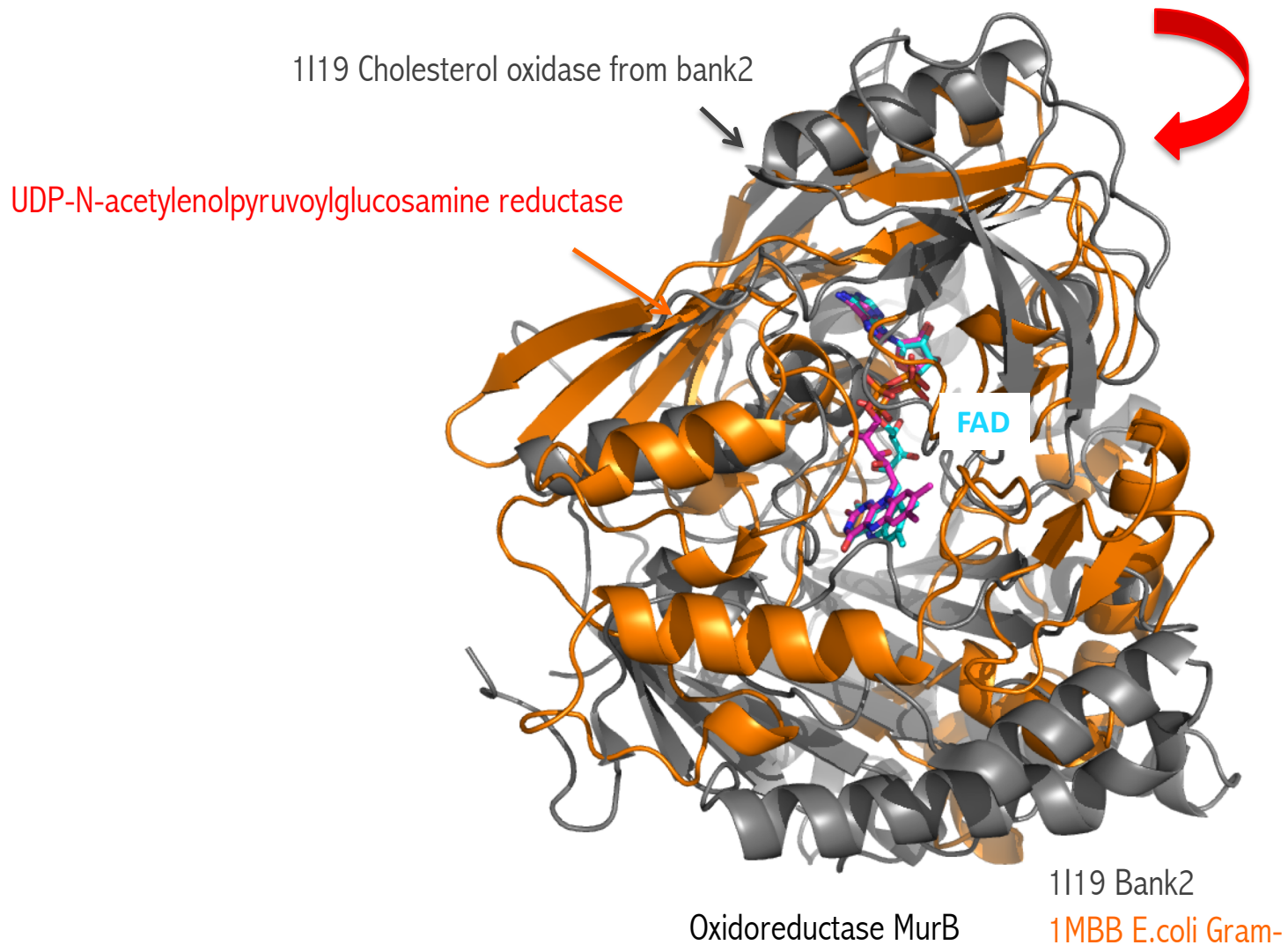
=

99.9 % 172 aa bdd8\_003618.fasta.a3m.hhm.out.hhmakeBK2 → UDP-N-acetylenolpyruvoylglucosamine reductase

★ *B. dorei*D8

Sorted by Prob > 95 and size cut off > 50 amino acid residues

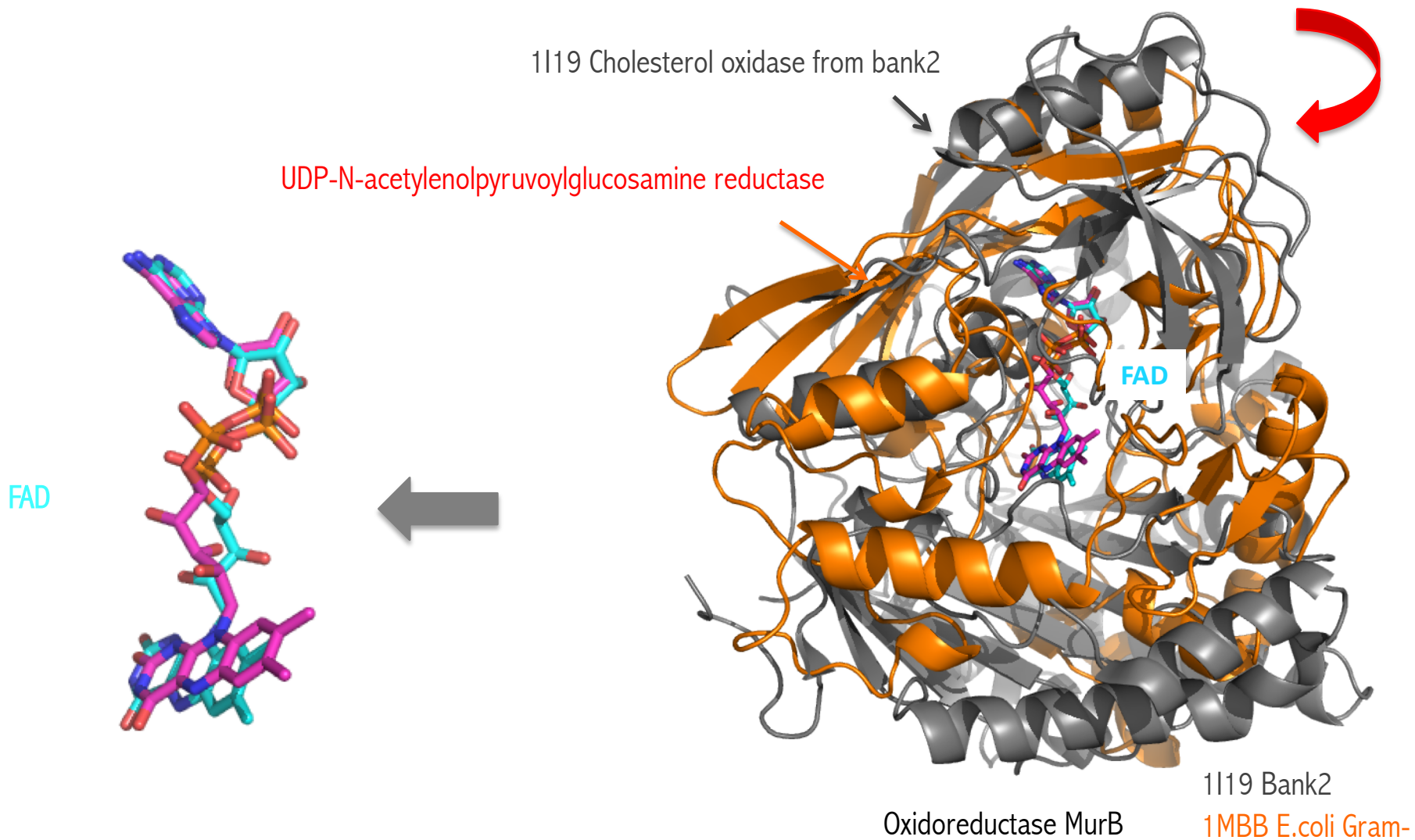
99.9 172 aa bdd8\_003618.fasta.a3m.hhm.out.hhmakeBK2 → UDP-N-acetylenolpyruvoylglucosamine reductase



★ *B. dorei*D8

Sorted by Prob > 95 and size cut off > 50 amino acid residues

99.9 172 aa bdd8\_003618.fasta.a3m.hhm.out.hhmakeBK2 → UDP-N-acetylenolpyruvoylglucosamine reductase

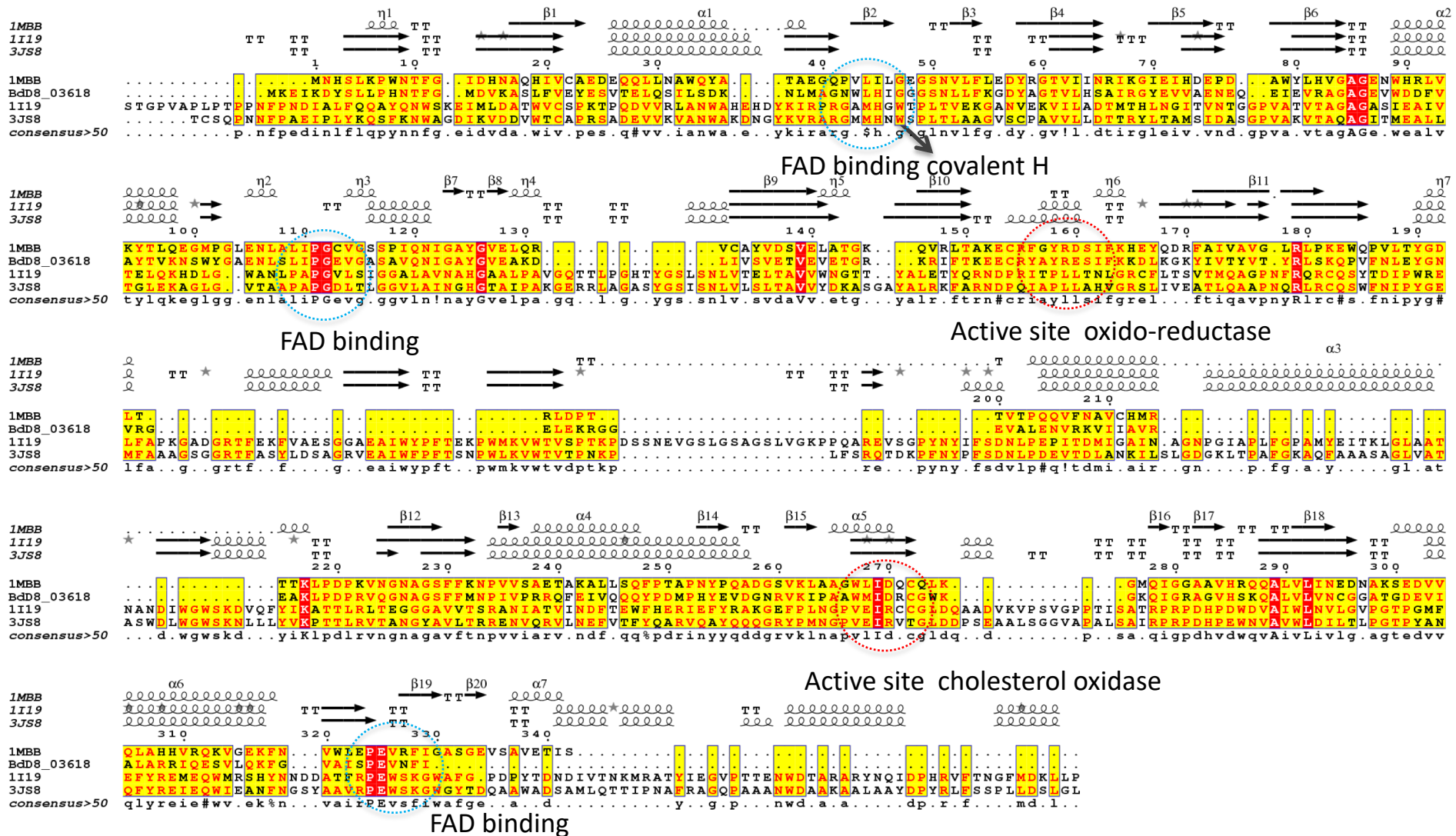




★ *B. dorei* D8

Sorted by Prob > 95 and size cut off > 50 amino acid residues

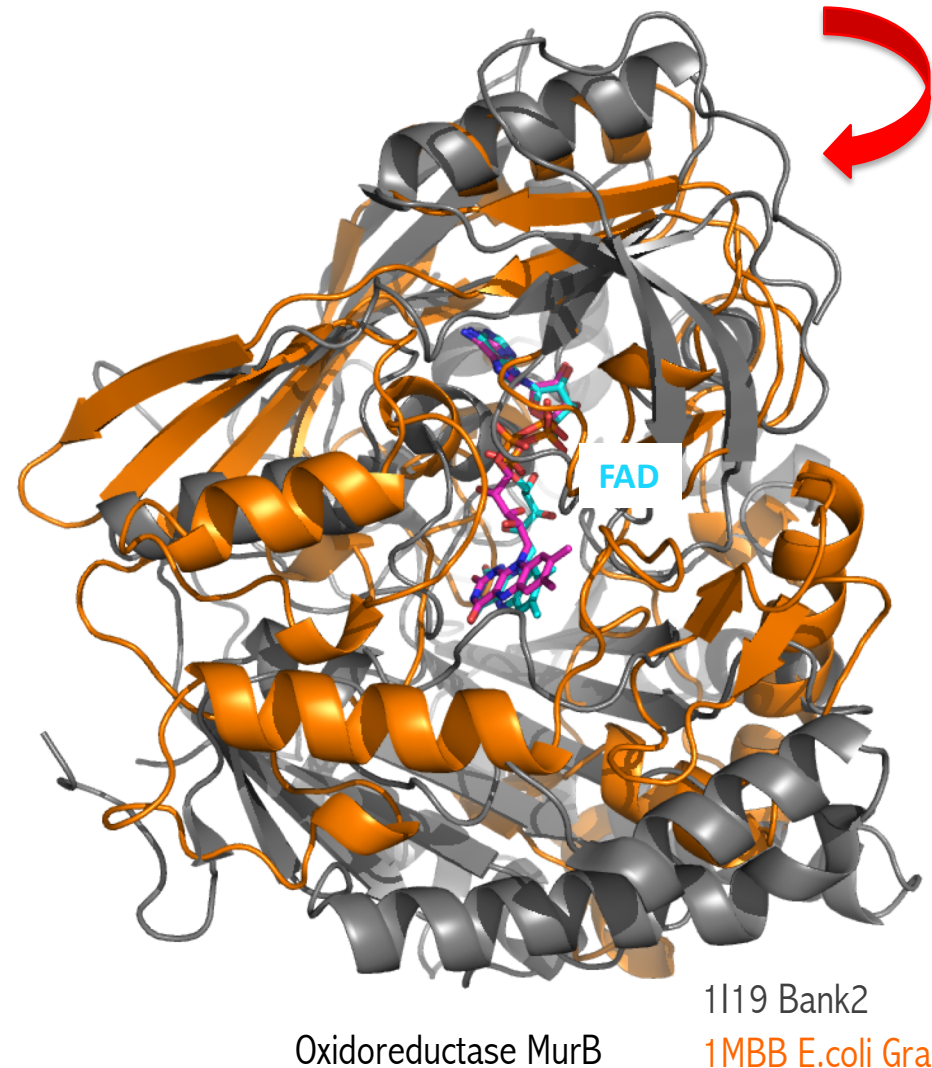
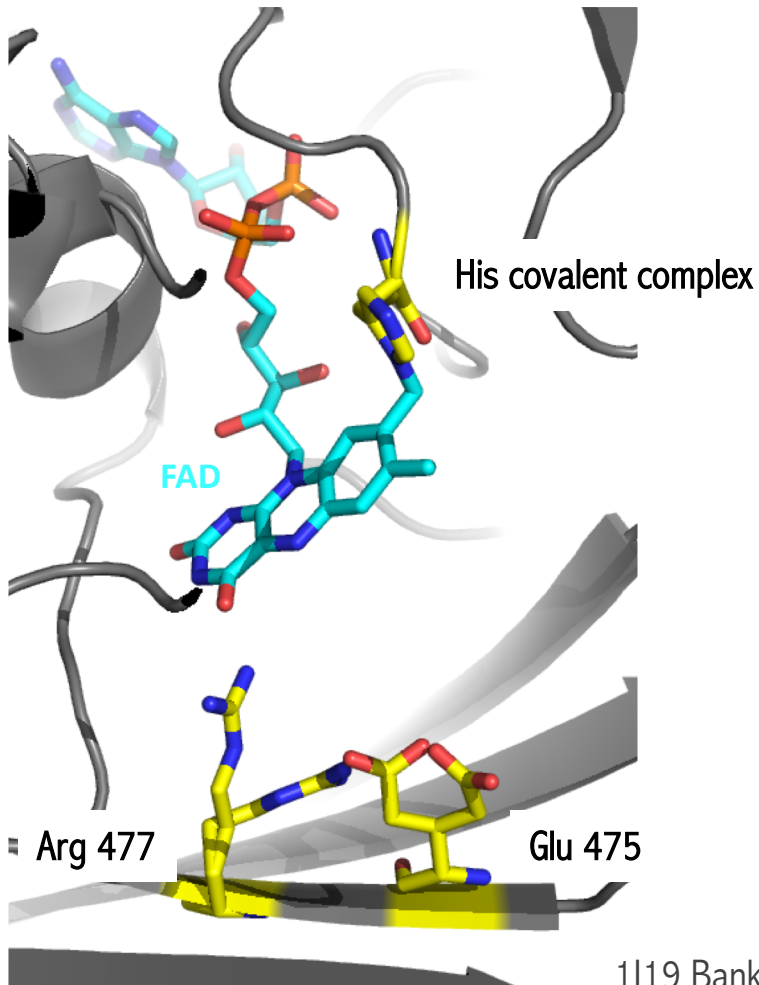
99.9 172 aa bdd8\_003618.fasta.a3m.hhm.out.hhmakBK2 → UDP-N-acetylenolpyruvoylglucosamine reductase



★ *B. dorei*D8

Sorted by Prob > 95 and size cut off > 50 amino acid residues

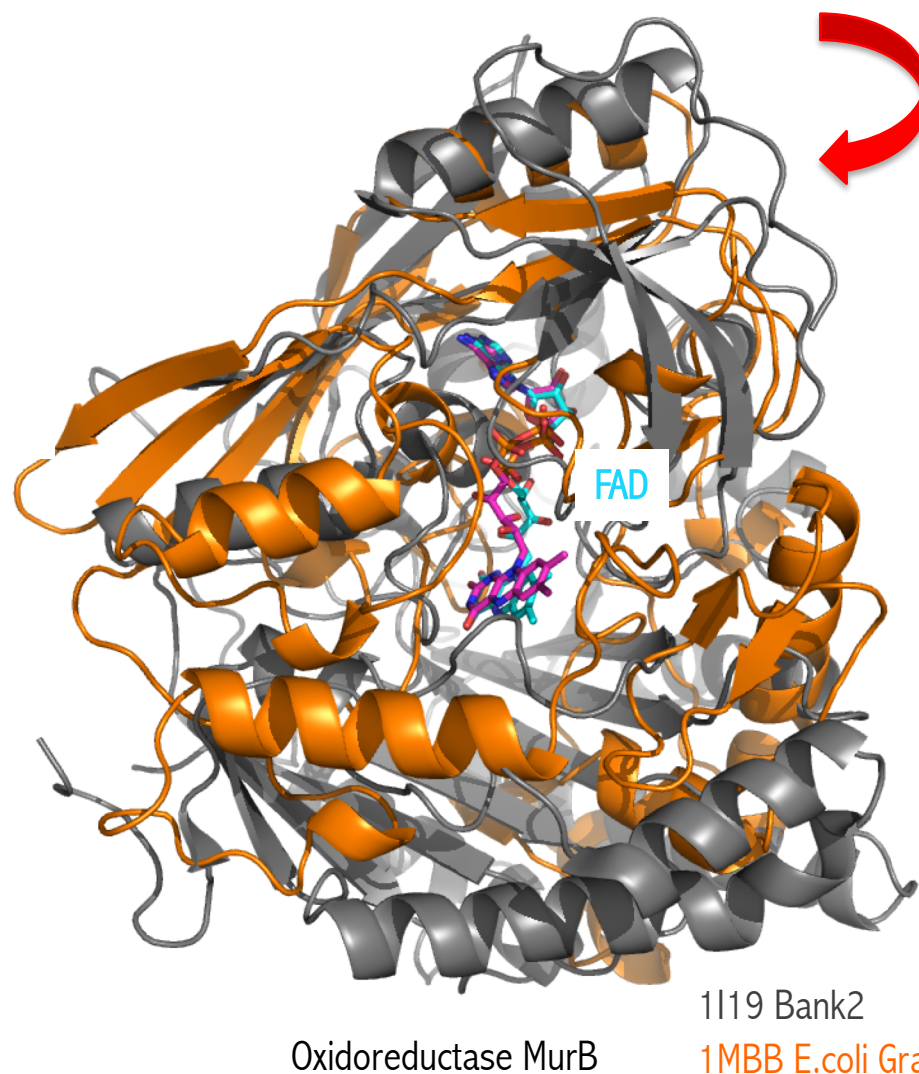
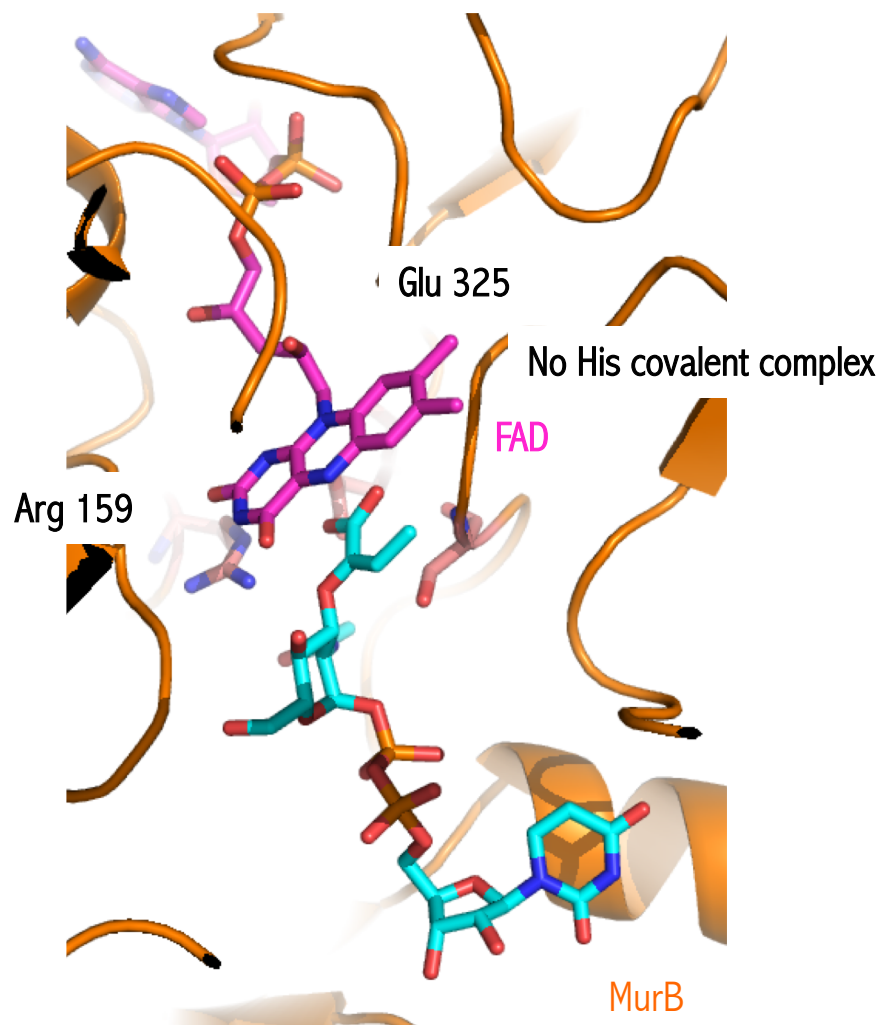
99.9 172 aa bdd8\_003618.fasta.a3m.hhm.out.hhmakeBK2 → UDP-N-acetylenolpyruvoylglucosamine reductase



★ *B. dorei*D8

Sorted by Prob > 95 and size cut off > 50 amino acid residues

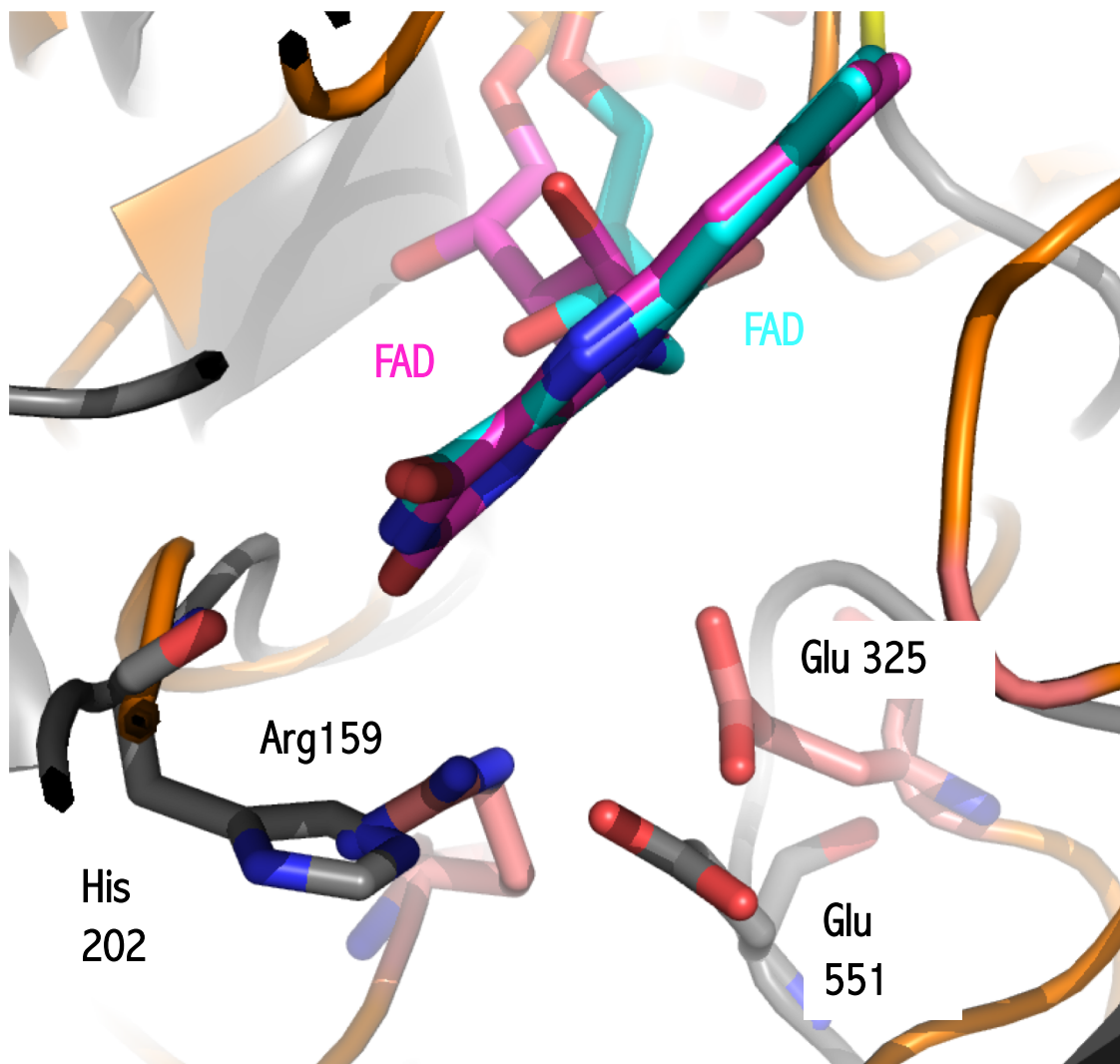
99.9 172 aa bdd8\_003618.fasta.a3m.hhm.out.hhmakeBK2 → UDP-N-acetylenolpyruvoylglucosamine reductase



★ *B. dorei*D8

Sorted by Prob > 95 and size cut off > 50 amino acid residues

99.9 172 aa bdd8\_003618.fasta.a3m.hhm.out.hhmakeBK2 → UDP-N-acetylenolpyruvoylglucosamine reductase



1119 Bank2  
1MBB MurB

★ *B. dorei*D8

Sorted by Prob &gt; 95 and size cut off &gt; 50 amino acid residues

- 99.4 % 70 aa baccor\_01751.fasta.a3m.hhm.out.hhmakeBK1 → Hypothetical protein
- 99.1 % 75 aa baccor\_01928.fasta.a3m.hhm.out.hhmakeBK1 → L-aspartate oxidase
- 99.1 % 62 aa baccor\_02089.fasta.a3m.hhm.out.hhmakeBK1 → Succinate Dehydrogenase flavoprotein
- 97.5 % 61 aa baccor\_03886.fasta.a3m.hhm.out.hhmakeBK1 → Pyridine nucleotide oxido reductase

★ *B. dorei* 175<sup>T</sup>

Sorted by Prob &gt; 95 and size &gt; 50 amino acid residues

- 99.9 % 172 aa baccor\_03524.fasta.a3m.hhm.out.hhmakeBK2 → UDP-N-acetyl muramate dehydrogenase

=

- 99.9 % 172 aa bdd8\_003618.fasta.a3m.hhm.out.hhmakeBK2 → UDP-N-acetylenolpyruvoylglucosamine reductase



Cloning and expression tests in progress.

Then, purification, activity test and crystallization assays

Cholesterol oxidases are secreted. This contains a signal peptide.

★ *B. dorei*D8

Sorted by Prob &gt; 95 and size cut off &gt; 50 amino acid residues

- 99.4 % 70 aa baccor\_01751.fasta.a3m.hhm.out.hhmakeBK1 → Hypothetical protein
- 99.1 % 75 aa baccor\_01928.fasta.a3m.hhm.out.hhmakeBK1 → L-aspartate oxidase
- 99.1 % 62 aa baccor\_02089.fasta.a3m.hhm.out.hhmakeBK1 → Succinate Dehydrogenase flavoprotein
- 97.5 % 61 aa baccor\_03886.fasta.a3m.hhm.out.hhmakeBK1 → Pyridine nucleotide oxido reductase

★ *B. dorei* 175<sup>T</sup>

Sorted by Prob &gt; 95 and size &gt; 50 amino acid residues

- 99.9 % 172 aa baccor\_03524.fasta.a3m.hhm.out.hhmakeBK2 → UDP-N-acetyl muramate dehydrogenase

=

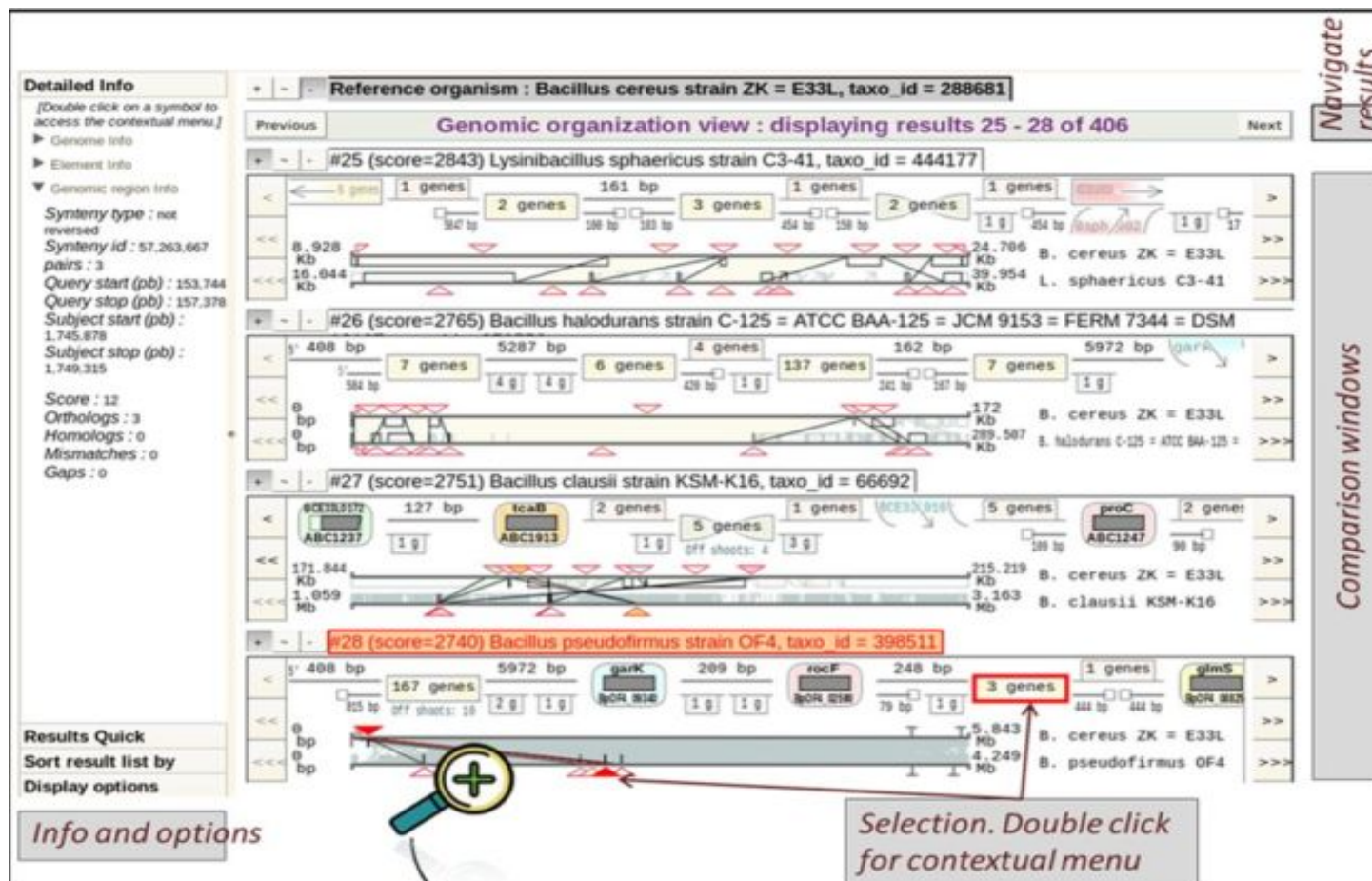
- 99.9 % 172 aa bdd8\_003618.fasta.a3m.hhm.out.hhmakeBK2 → UDP-N-acetylenolpyruvoylglucosamine reductase



Analysis of synthenies by Thomas Lacroix using Insyght.

★ <http://genome.jouy.inra.fr/Insyght>

➡ Powerful tool for genes comparison and synteny inferences



**Detailed Info**

[Double click on a symbol to access the contextual menu.]

- Genome info
- Element info
- Genomic region info

**Synteny type** : not reversed  
**Synteny id** : 57.263.667  
**pairs** : 3  
**Query start (pb)** : 153.744  
**Query stop (pb)** : 157.378  
**Subject start (pb)** : 1.745.878  
**Subject stop (pb)** : 1.749.315

**Score** : 12  
**Orthologs** : 3  
**Homologs** : 0  
**Mismatches** : 0  
**Gaps** : 0

**Results Quick**

- Sort result list by
- Display options

**Reference organism : Bacillus cereus strain ZK = E33L, taxo\_id = 288681**

**Genomic organization view : displaying results 25 - 28 of 406**

**#25 (score=2843) Lysinibacillus sphaericus strain C3-41, taxo\_id = 444177**

**#26 (score=2765) Bacillus halodurans strain C-125 = ATCC BAA-125 = JCM 9153 = FERM 7344 = DSM**

**#27 (score=2751) Bacillus clausii strain KSM-K16, taxo\_id = 66692**

**#28 (score=2740) Bacillus pseudofirmus strain OF4, taxo\_id = 398511**

**Navigation:** Previous, Next, <<, >>, <, >

**Comparison windows** (vertical label on the right)

**Info and options** (bottom left callout)

**Selection. Double click for contextual menu** (bottom right callout with magnifying glass icon)

★ <http://genome.jouy.inra.fr/Insyght>

➡ Powerful tool for genes comparison and syntenic inferences

Syntenies : co-localization of homologous loci across species

Comparison 2 vs 2 with a reference genome.

No synteny. No clear information

The screenshot shows the Insyght web interface with a comparison of five bacterial strains. The strains listed are:

- #1 Burkholderia
- #3 Corynebacterium efficiens (strain : DSM 44549 = NBRC 100395 = JCM 11189 = YS-314 = AJ 12310)
- #4 Corynebacterium diphtheriae (strain : NCTC 13129 = ATCC 700971)
- #5 Agrobacterium vitis (strain : S4 = ATCC BAA-846)
- #6 Burkholderia cenocepacia (strain : H12424)

Gene models for dnaA, recF, gyrB, and other genes are shown for each strain. A vertical label on the right reads "Comparison windows". A "Contextual menu" is open at the bottom, showing options like "Show other homologs".

Info and options    Contextual menu: browse multiple homologies per gene



## ★ Direct introduction of the COPROSTANOL standard by DIMS-MS/MS

Samples resuspended in 500 mL of Chloroform/ Methanol (50/50)

Diluted to 1/100eme before analysis

- Ions detection corresponding to COPROSTANOL
- DIMS-MS/MS method in positive ionisation mode

Catherine. Juste  
Fabienne Beguet-Crespel

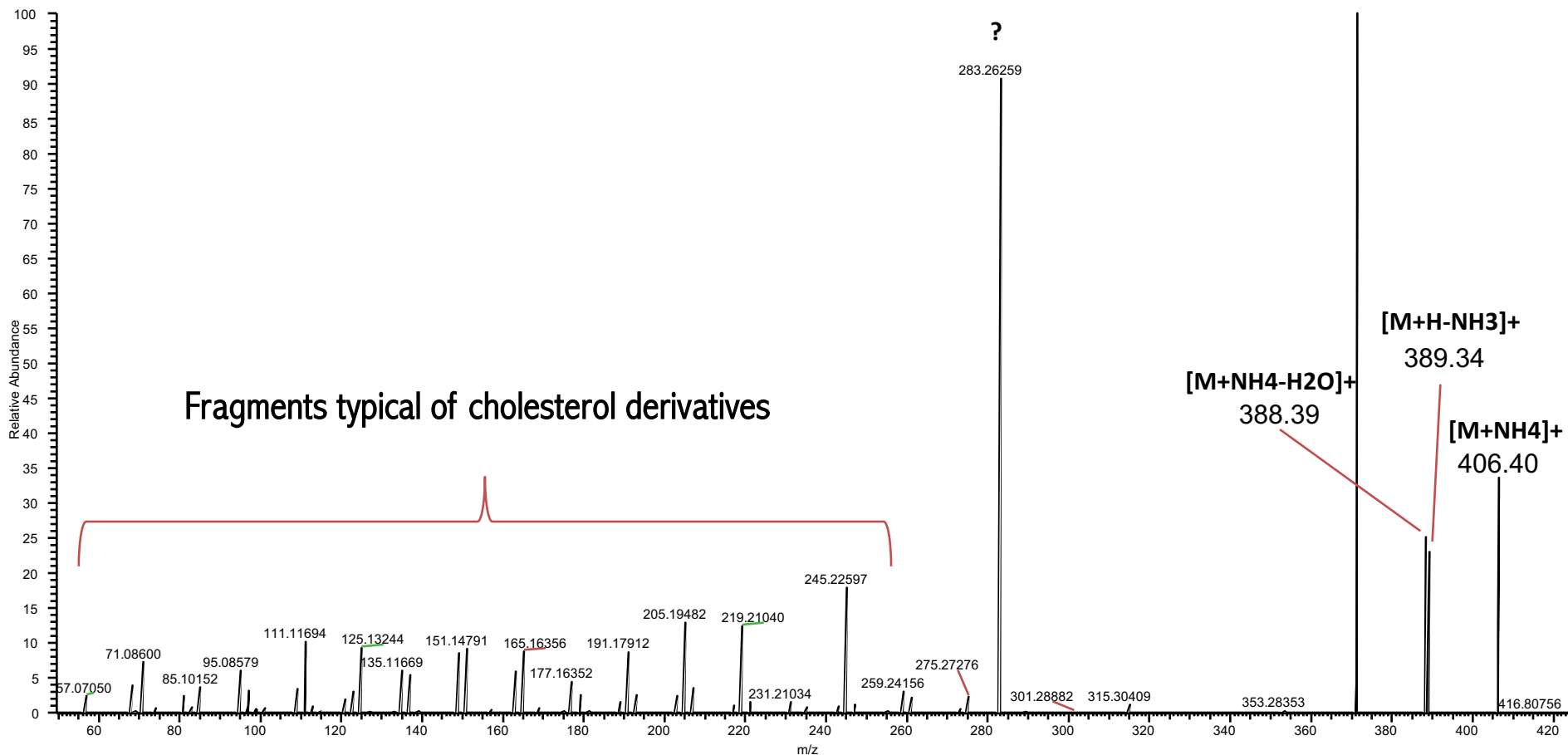
### Coprostanol

$[M+H-H_2O]^+$   
371.36

$[M+H-NH_3]^+$   
389.34

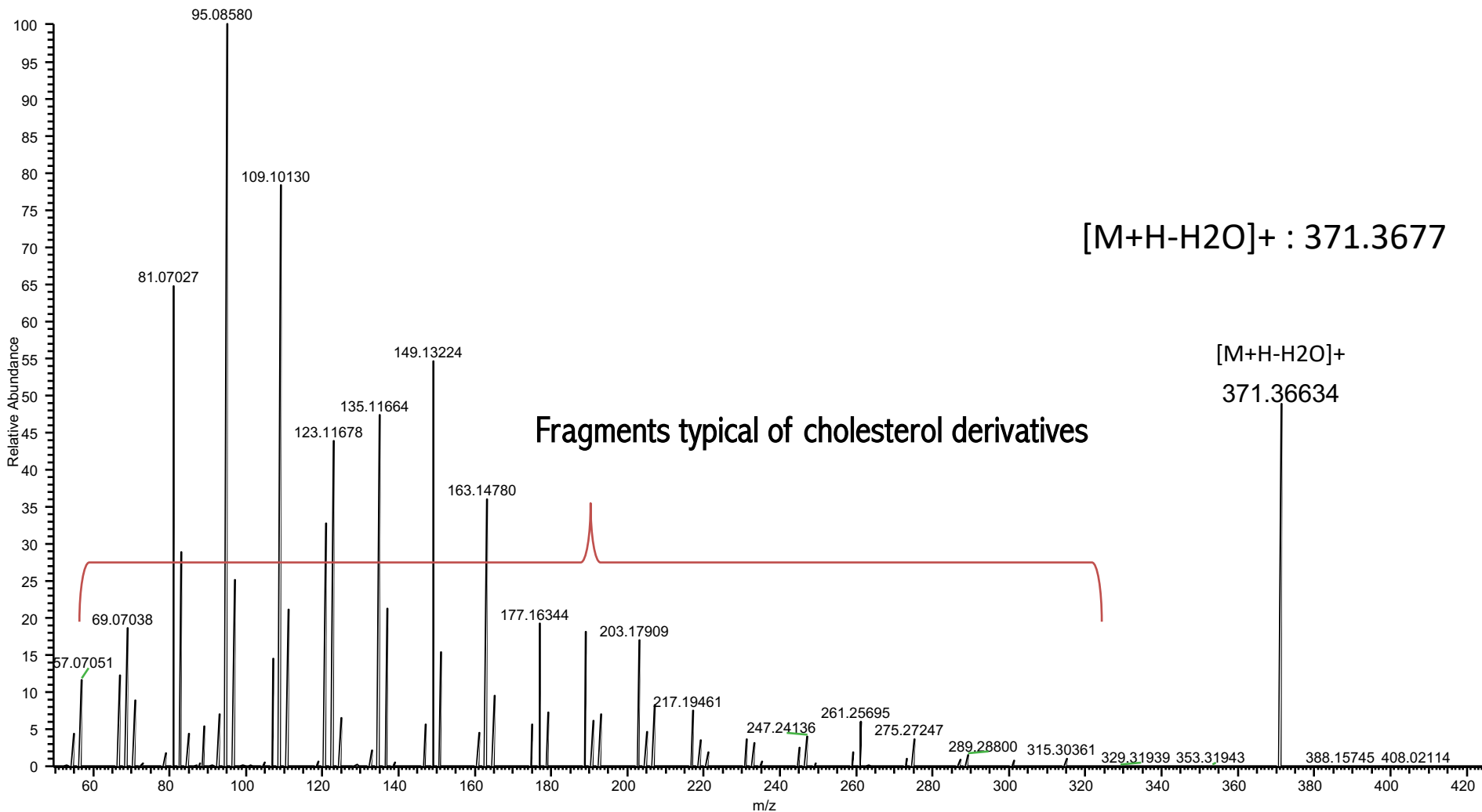
$[M+NH_4-H_2O]^+$   
388.39

$[M+NH_4]^+$   
406.40



- Ions detection corresponding to COPROSTANOL
- DIMS-MS/MS method in positive ionisation mode

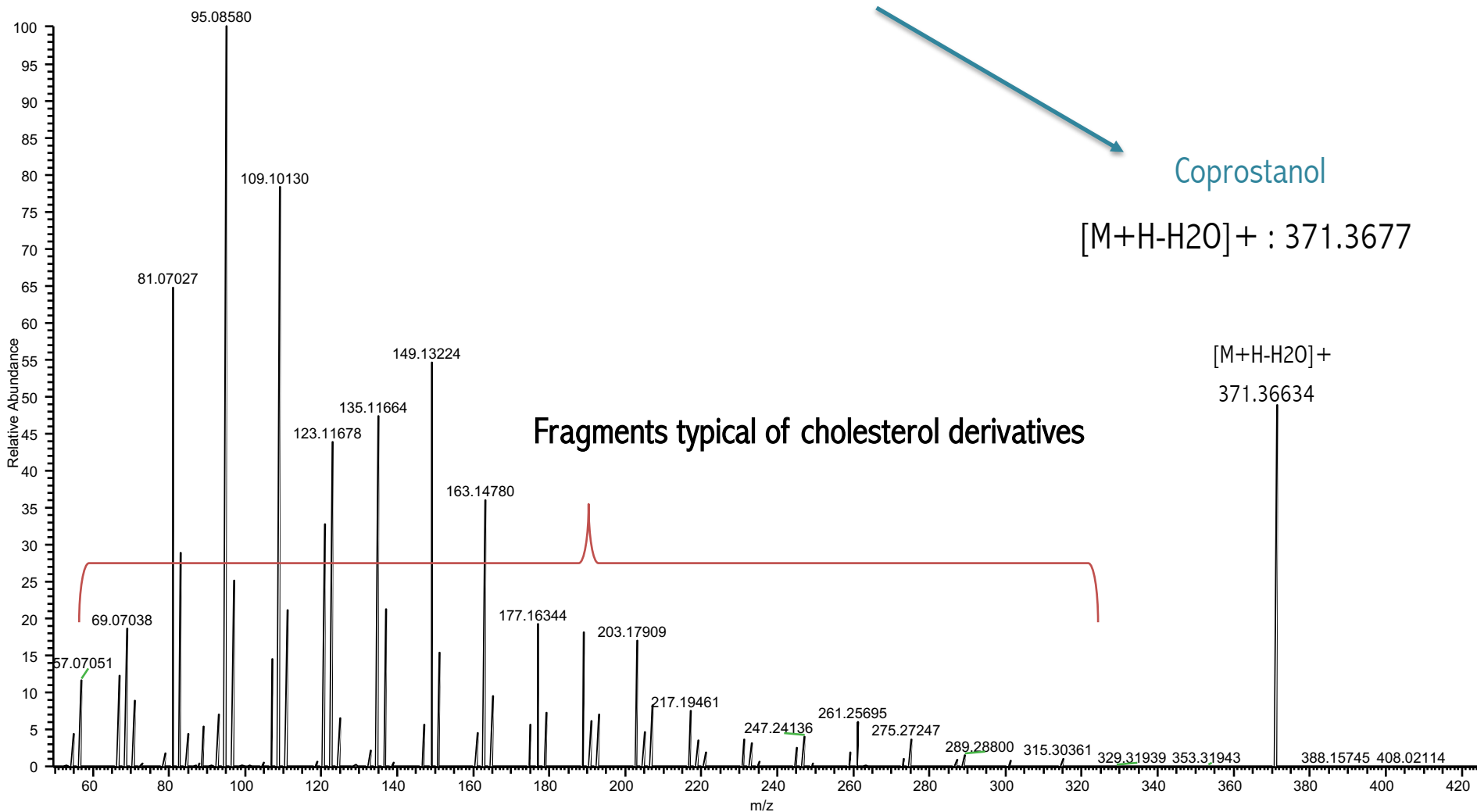
Catherine. Juste  
Fabienne Beguet-Crespel



- Ions detection corresponding to COPROSTANOL
- DIMS-MS/MS method in positive ionisation mode

Catherine. Juste  
Fabienne Beguet-Crespel

Coprostanol is present & preferentially detected



- ★ Analysis of candidates- Comparison within cholesterol degrading genomes and others  
Comparison between *E. coprostanoligenes* (2017), *B. Dorei D8*, *Parabacteroides distasonis* shown to lower cholesterolemia while producing coprostanol. PhD thesis Aline Potiron.
- ★ Annotation comparator's analysis
- ★ Genomic organisation's analysis



One hit Interesting: Analysis in progress with Thomas Lacroix.  
Present in degrading bacteria, absent in non degrading bacteria



Annotated as lipolytic protein G-D-S-L family, phospholipase, major facilitator superfamily

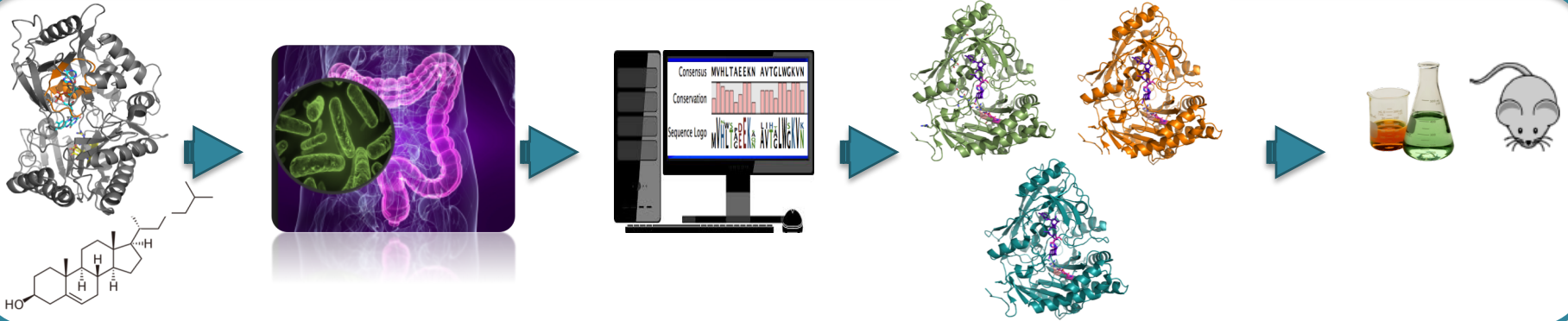


Cholesterol oxidase with a Ser His Asp conserved catalytic site plus the oxyanionic hole identified

Homology modeling and docking of cholesterol in progress

- ➔ No significant difference between the two genomes. Similar enzymes are found.
- ➔ Transcriptomic analysis of both chromosome and plasmids of *B. dorei* 175<sup>T</sup> & *B. dorei* D8?
- ➔ Activity test and 3D structure validation of our cholesterol-oxidase hit in bank2.
- ➔ Genomic & transcriptomic analysis of both chromosome & plasmids of *E. coprostanoligenes*.
- ➔ Can we rule out some synergy among bacteria of the microbiota?
- ➔ Promising hit coming from entire genomes comparison: cholesterol esterase function and pathway.

# Towards the Structural Screening of Microbial Ecosystems



## Outline

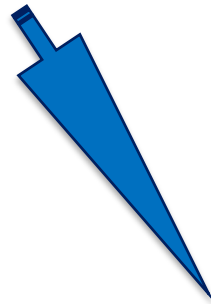
- Cholesterol conversion in the gut microbiota: the mystery enzyme(s) ?
- **MetaFoldScan project : 3D screening of gut microbiota**
- Conclusions and perspectives

# The gut microbiota and our cells: intimate interaction

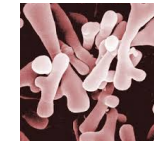
- 100 000 billions of bacteria
- 4 phyla
- ~1000 species



Pionneur



*Lactobacille*



*Bifidobacteria*



*Faecalibacterium prausnitzii*

« Symbiontes »



*Clostridium difficile*

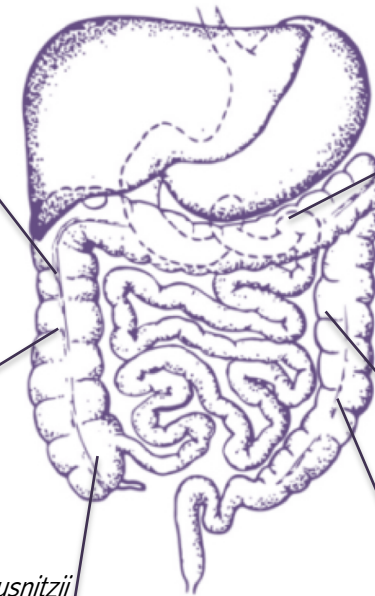


*Enterococcus faecalis*



*Campylobacter*

« Pathobiontes »



Catalog of 10 millions genes



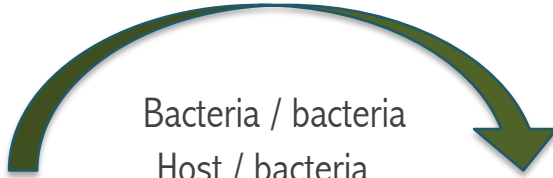
Universel : core genome

Identity card

Bidirectional dialog: language based on signaling

# The gut microbiota and our cells: intimate interaction

Dialog

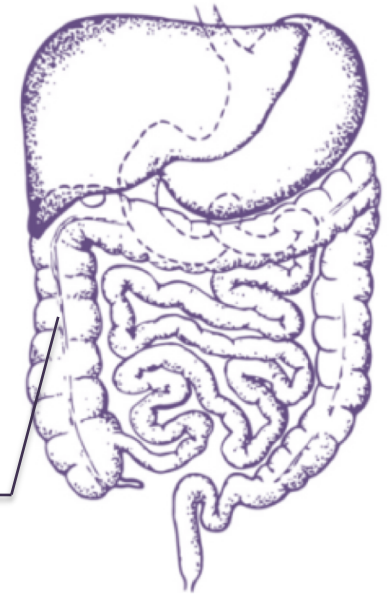


Gene diversity  
Composition in bacteria

Symbiose

Health markers

Homeostasis

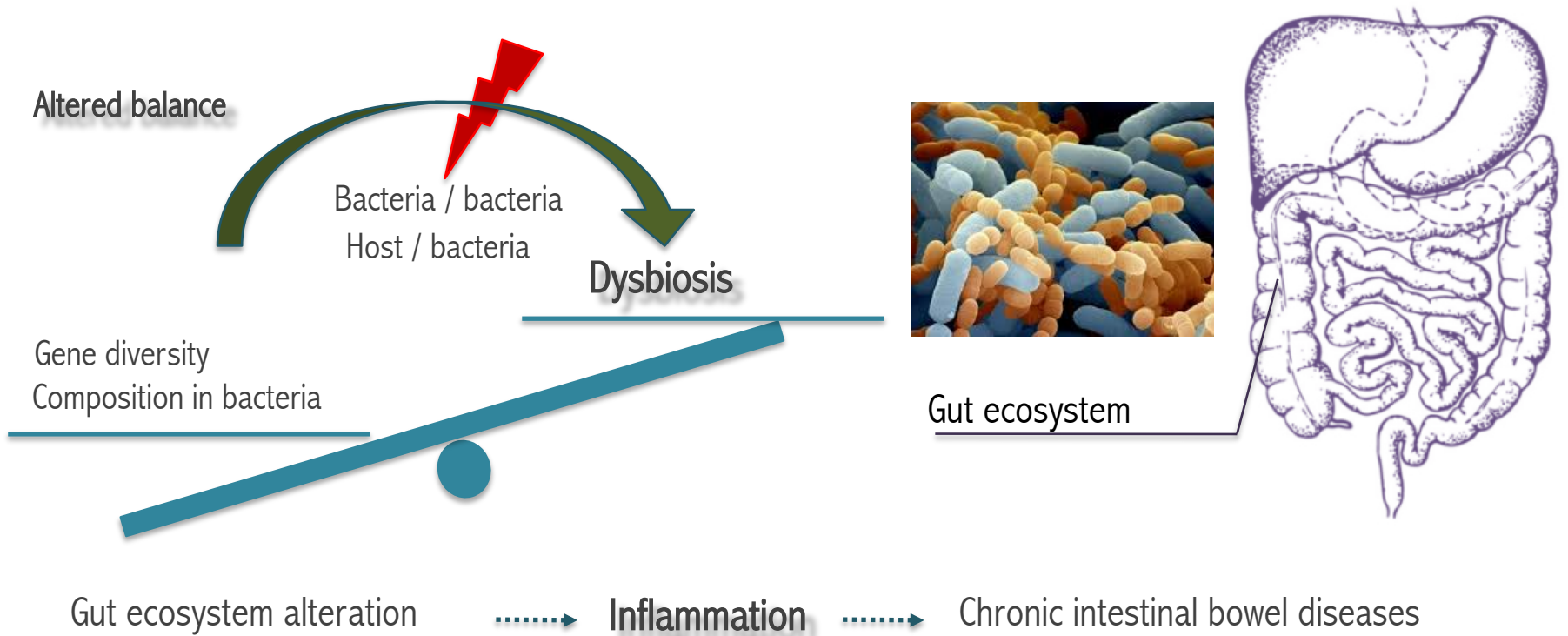


Gut ecosystem

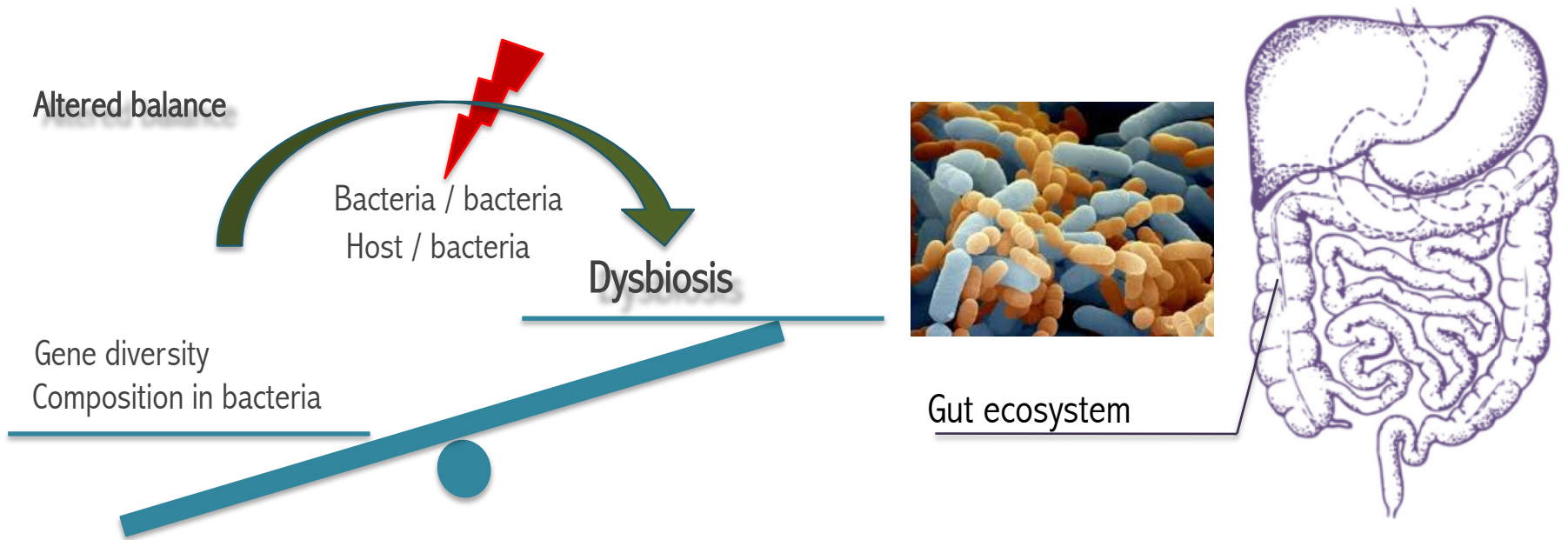
Key interface cells/food/bacteria



# Decipher the molecular mechanisms responsible for this dialog



# Decipher the molecular mechanisms responsible for this dialog



Gut ecosystem alteration ..... Inflammation ..... Chronic intestinal bowel diseases

Catalogue of 10 millions of genes

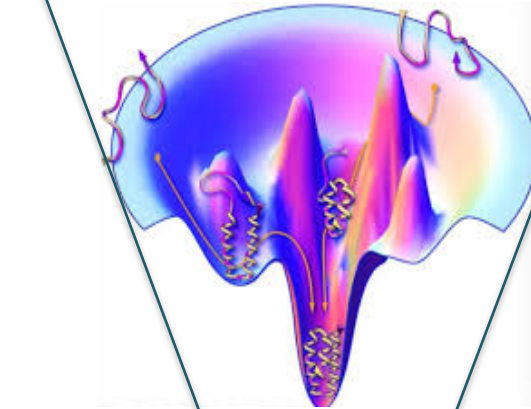
Metadata to explore and target



Investigate the molecular mechanisms of this balance

- - More than 10 millions prokaryotic genes
- - 40% protein sequences of unknown function
- - Function can be inferred from structure
- - Assessing the fold is possible
- - Fishing out the right fold is crucial !

126,000 known structures



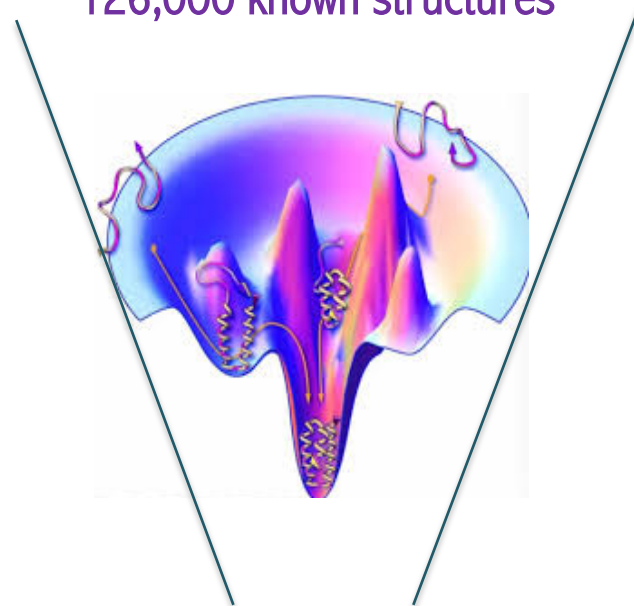
Only 1400 3D folds

- - More than 10 millions prokaryotic genes
- - 40% protein sequences of unknown function
- - Function can be inferred from structure
- - Assessing the fold is possible
- - Fishing out the right fold is crucial !



**Fold prediction informs on the function**

126,000 known structures



Only 1400 3D folds

- - More than 10 millions prokaryotic genes
- - 40% protein sequences of unknown function
- - Function can be inferred from structure
- - Assessing the fold is possible
- - Fishing out the right fold is crucial !



**Fold prediction informs on the function**

Metagenomic



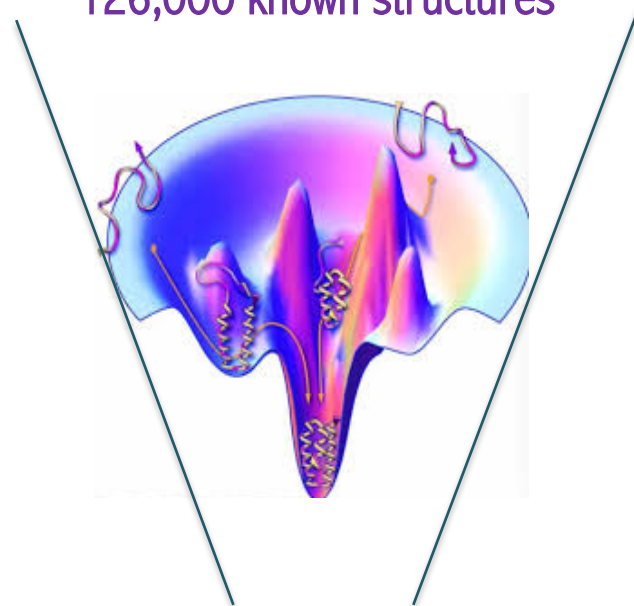
Functional metagenomic

Catalogues of sequences

Fold detection and prediction accurate

Computer compatible with big data

126,000 known structures



Only 1400 3D folds

- - More than 10 millions prokaryotic genes
- - 40% protein sequences of unknown function
- - Function can be inferred from structure
- - Assessing the fold is possible
- - Fishing out the right fold is crucial !



**Fold prediction informs on the function**

Metagenomic



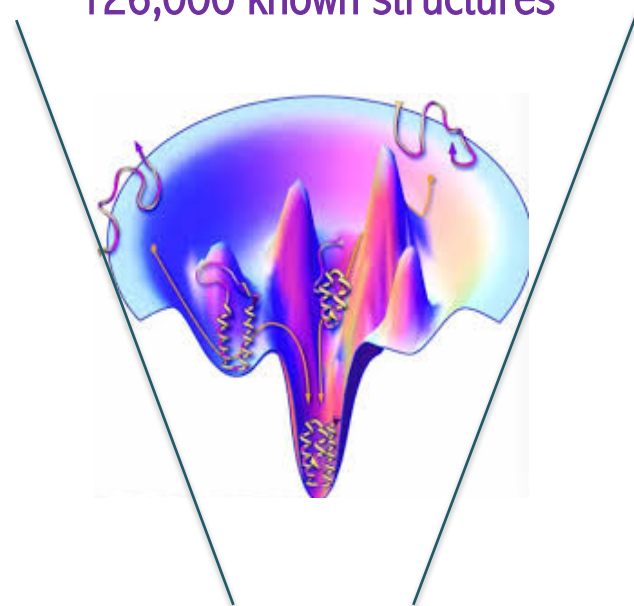
Functional metagenomic

Catalogues of sequences

Fold detection and prediction accurate

Computer compatible with big data

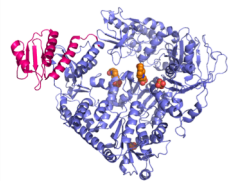
126,000 known structures



Only 1400 3D folds

**Comprehensive detection and 3D modeling = StructurOmics**

Target protein

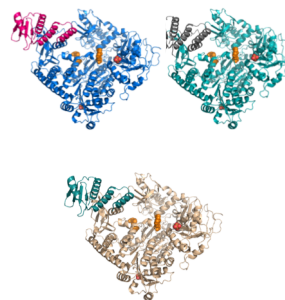
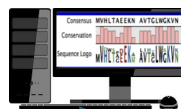


Relevance

- Biological property
- Therapeutical interest
- Structural challenge



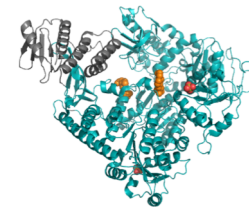
Scan microbiota



Hits identification

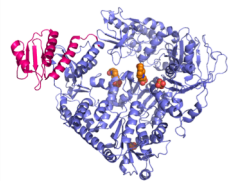


Validation



Biomarker

Target protein

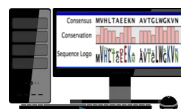


Relevance

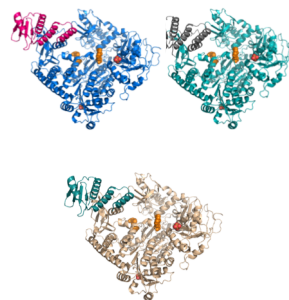
- Biological property
- Therapeutical interest
- Structural challenge



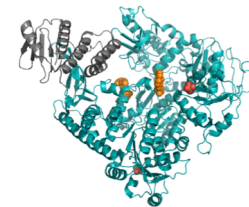
Scan microbiota



Hits identification



Validation



Biomarker

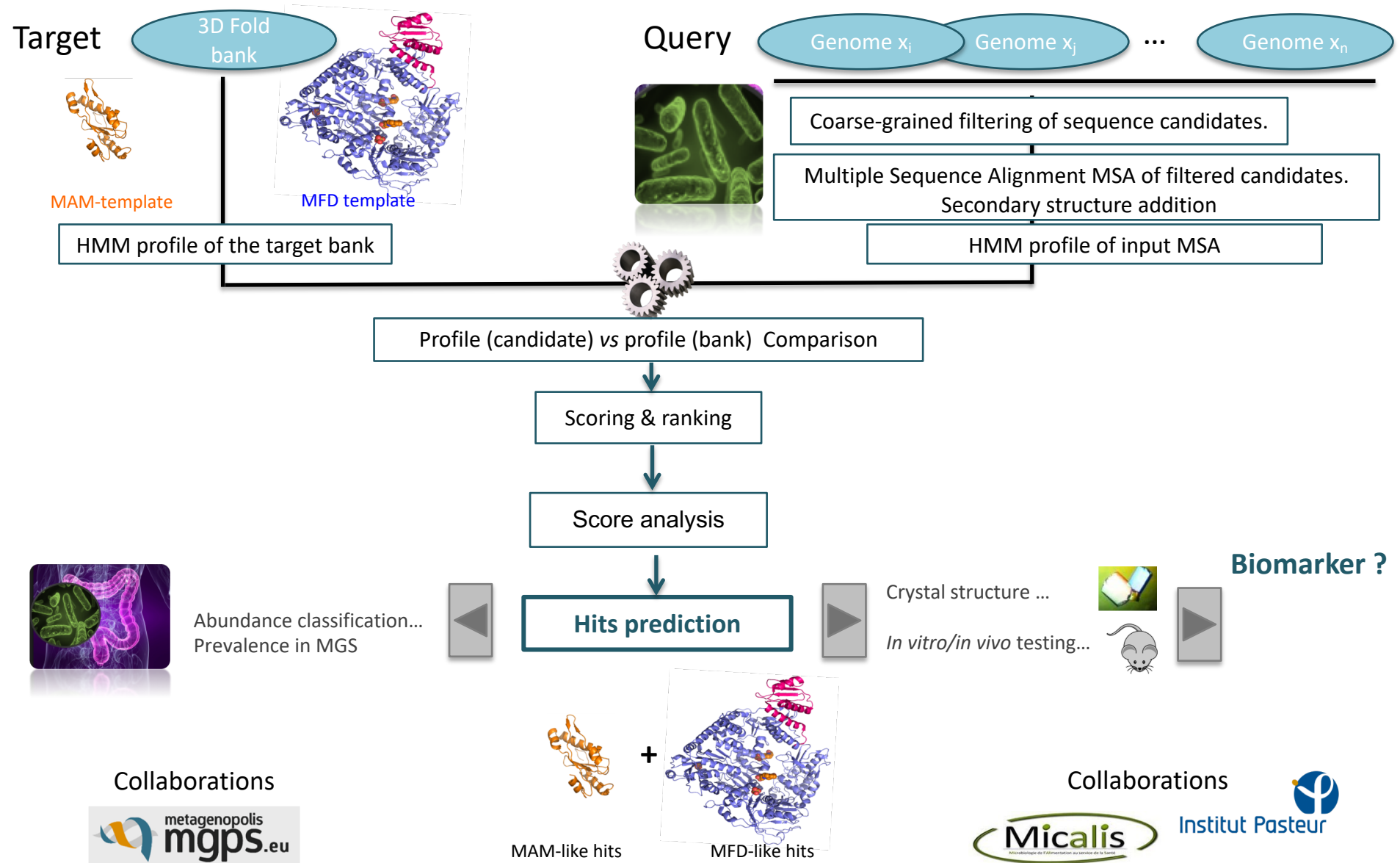
Preliminary data:

- Histone-like Nucleotide Structuring protein in *Enterococcus faecalis*
- Penicillin binding protein in *Bacillus subtilis*
- Cholesterol oxidase for *Bacteroides dorei* and *dorei D8*



- ❖ Experimental data in progress
- ❖ Scaling up to full microbiota requested





To tackle both methodological and biological issues

MAM

JM. Chatel

## Microbial Anti-inflammatory Molecule

- Unknown biological function
- Small 180 residues
- Structurally not characterized
- Rare, found in *F. prausnitzii*
- *F. Prausnitzii* 5% of the microbiota, IBD  
Crohn Disease, Ulcerative Colitis,
- Fishing from homology modeling
- Associated with probiotic properties

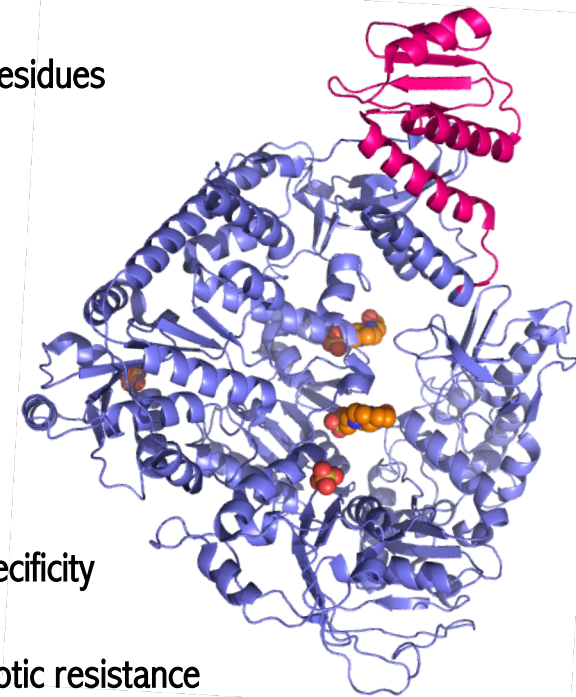


MFD

N. Rama Rao

## Mutation Frequency Decline

- Involved in bacterial DNA repair, in virulence
- Multi-domains 1150 residues
- Solved structure
- Ubiquitous, Gram+/-
- Promiscuity
- Molecular basis of specificity
- Associated with antibiotic resistance



Cohort of 124 European individuals, healthy and obese, danish and spanish  
 Nordic & Mediterranean faecal DNA

Vol 464 | 4 March 2010 | doi:10.1038/nature08821

nature

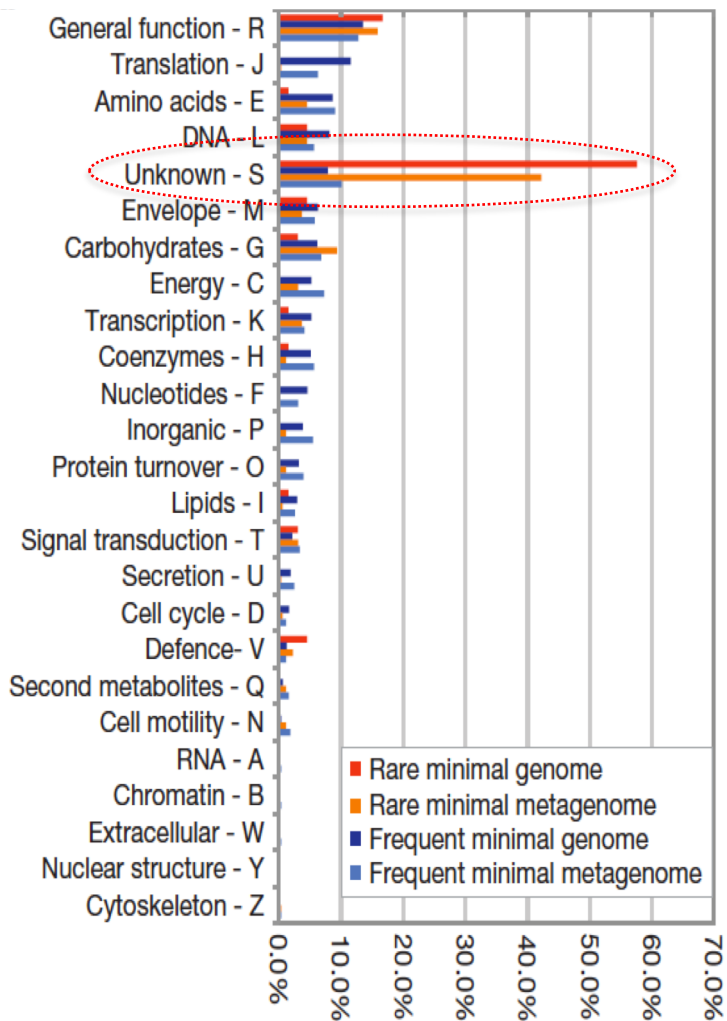
## A human gut microbial gene catalogue established by metagenomic sequencing

Junjie Qin<sup>1\*</sup>, Ruiqiang Li<sup>1\*</sup>, Jeroen Raes<sup>2,3</sup>, Manimozhiyan Arumugam<sup>2</sup>, Kristoffer Solvsten Burgdorf<sup>4</sup>, Chaysavanh Manichanh<sup>5</sup>, Trine Nielsen<sup>4</sup>, Nicolas Pons<sup>6</sup>, Florence Levenez<sup>6</sup>, Takuji Yamada<sup>2</sup>, Daniel R. Mende<sup>2</sup>, Junhua Li<sup>1,7</sup>, Junming Xu<sup>1</sup>, Shaochuan Li<sup>1</sup>, Dongfang Li<sup>1,8</sup>, Jianjun Cao<sup>1</sup>, Bo Wang<sup>1</sup>, Huiqing Liang<sup>1</sup>, Huisong Zheng<sup>1</sup>, Yinlong Xie<sup>1,7</sup>, Julien Tap<sup>6</sup>, Patricia Lepage<sup>6</sup>, Marcelo Bertalan<sup>9</sup>, Jean-Michel Batto<sup>6</sup>, Torben Hansen<sup>4</sup>, Denis Le Paslier<sup>10</sup>, Allan Linneberg<sup>11</sup>, H. Bjørn Nielsen<sup>9</sup>, Eric Pelletier<sup>10</sup>, Pierre Renault<sup>6</sup>, Thomas Sicheritz-Ponten<sup>9</sup>, Keith Turner<sup>12</sup>, Hongmei Zhu<sup>1</sup>, Chang Yu<sup>1</sup>, Shengting Li<sup>1</sup>, Min Jian<sup>1</sup>, Yan Zhou<sup>1</sup>, Yingrui Li<sup>1</sup>, Xiuqing Zhang<sup>1</sup>, Songgang Li<sup>1</sup>, Nan Qin<sup>1</sup>, Huanming Yang<sup>1</sup>, Jian Wang<sup>1</sup>, Søren Brunak<sup>9</sup>, Joel Doré<sup>6</sup>, Francisco Guarner<sup>5</sup>, Karsten Kristiansen<sup>13</sup>, Oluf Pedersen<sup>4,14</sup>, Julian Parkhill<sup>12</sup>, Jean Weissenbach<sup>10</sup>, MetaHIT Consortium†, Peer Bork<sup>2</sup>, S. Dusko Ehrlich<sup>6</sup> & Jun Wang<sup>1,13</sup>

To understand the impact of gut microbes on human health and well-being it is crucial to assess their genetic potential. Here we describe the Illumina-based metagenomic sequencing, assembly and characterization of 3.3 million non-redundant microbial genes, derived from 576.7 gigabases of sequence, from faecal samples of 124 European individuals. The gene set, ~150 times larger than the human gene complement, contains an overwhelming majority of the prevalent (more frequent) microbial genes of the cohort and probably includes a large proportion of the prevalent human intestinal microbial genes. The genes are largely shared among individuals of the cohort. Over 99% of the genes are bacterial, indicating that the entire cohort harbours between 1,000 and 1,150 prevalent bacterial species and each individual at least 160 such species, which are also largely shared. **We define and describe the minimal gut metagenome and the minimal gut bacterial genome in terms of functions present in all individuals and most bacteria, respectively.**

- - 3,3 millions of genes , 99,1 % of bacterial origin
- - Minimal gut bacterial genome : 1,000 to 1,150 prevalent bacterial species
- - Each individual harbors at least 160 such species

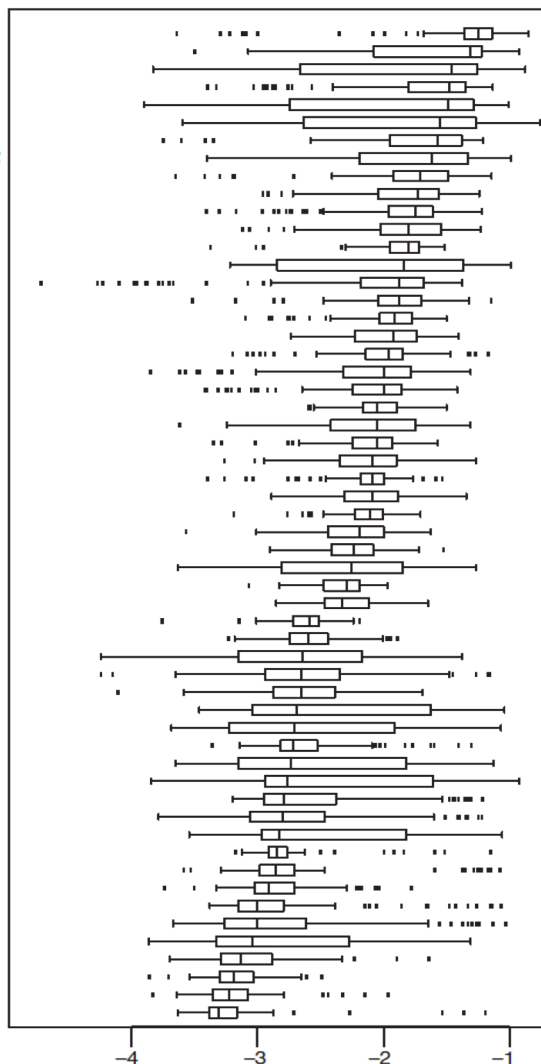
## Minimal gut genome: functions necessary for a bacterium to thrive in a gut context



Core genome: 57 bacterial species present in > 90 of individuals (prevalence and consensus)

➔ Mostly Bacteroidetes & Firmicutes

- Bacteroides uniformis*
- Alistipes putredinis*
- Parabacteroides merdae*
- Dorea longicatena*
- Ruminococcus bromii* L2-63
- Bacteroides caccae*
- Clostridium* sp. SS2-1
- Bacteroides thetaiotaomicron* VPI-5482
- Eubacterium hallii*
- Ruminococcus torques* L2-14
- Unknown sp. SS3 4
- Ruminococcus* sp. SR1 5
- Faecalibacterium prausnitzii* SL3 3
- Ruminococcus lactaris*
- Collinsella aerofaciens*
- Dorea formicigenerans*
- Bacteroides vulgatus* ATCC 8482
- Roseburia intestinalis* M50 1
- Bacteroides* sp. 2\_1\_7
- Eubacterium siraeum* 70 3
- Parabacteroides distasonis* ATCC 8503
- Bacteroides* sp. 9\_1\_42FAA
- Bacteroides ovatus*
- Bacteroides* sp. 4\_3\_47FAA
- Bacteroides* sp. 2\_2\_4
- Eubacterium rectale* M104 1
- Bacteroides xylanisolvens* XB1A
- Coprococcus comes* SL7 1
- Bacteroides* sp. D1
- Bacteroides* sp. D4
- Eubacterium ventriosum*
- Bacteroides dorei*
- Ruminococcus obeum* A2-162
- Subdoligranulum variabile*
- Bacteroides capillosus*
- Streptococcus thermophilus* LMD-9
- Clostridium leptum*
- Holdemanina filiformis*
- Bacteroides stercoris*
- Coprococcus eutactus*
- Clostridium* sp. M62 1
- Bacteroides eggerthii*
- Butyrivibrio crossotus*
- Bacteroides fingoldii*
- Parabacteroides johnsonii*
- Clostridium* sp. L2-50
- Clostridium nexile*
- Bacteroides pectinophilus*
- Anaerotruncus colihominis*
- Ruminococcus gnavus*
- Bacteroides intestinalis*
- Bacteroides fragilis* 3\_1\_12
- Clostridium asparagiforme*
- Enterococcus faecalis* TX0104
- Clostridium scindens*
- Blautia hansenii*



Relative abundance (log<sub>10</sub>)

		265	9,00E-93	100.00	100	100	135	0	0	1	135	1	135	
		bitscore	evalue	pident	qcovs	qcovhsp	length	mismatch	gapopen					
1	Faecalibacterium_prausnitzii_A2-165 NCBI	WP_005932151.1												
2	MAM_wt	tr R6SBT0 R6SBT0_9BACE	23.5	6.4	47.826	17	17	23	10	1	14	36	624	644
3	MAM_wt	tr D4JK97 D4JK97_9FIRM	28.5	0.013	45.000	15	15	20	11	0	1	20	9	28
4	MAM_wt	tr R7EPB3 R7EPB3_9BACE	26.6	0.55	41.667	24	24	36	18	1	52	84	185	220
5	MAM_wt	tr C3QB13 C3QB13_9BACE	29.3	0.063	40.625	24	24	32	19	0	88	119	95	126
6	MAM_wt	tr D4C9H4 D4C9H4_9CLOT	27.3	0.30	40.541	27	27	37	21	1	73	108	128	164
7	MAM_wt	tr D4KBR2 D4KBR2_9FIRM	75.9	1.02e-19	35.570	98	98	149	72	4	1	132	1	142
8	Faecalibacterium_prausnitzii_SL3_3	tr R7B146 R7B146_9BACE	25.0	1.2	34.694	36	36	49	31	1	15	63	119	166
9	MAM_wt	tr D4L4M8 D4L4M8_9FIRM	24.3	1.9	34.483	21	21	29	19	0	75	103	516	544
10	MAM_wt	tr C3QPR5 C3QPR5_9BACE	25.8	1.1	34.375	24	24	32	21	0	88	119	95	126
11	MAM_wt	tr R5TKN8 R5TKN8_9FIRM	26.2	0.50	34.286	26	26	35	22	1	48	82	205	238
12	MAM_wt	tr C0X1L4 C0X1L4_ENTFL	26.6	0.39	34.211	27	27	38	24	1	14	50	150	187
13	MAM_wt	tr D1PS54 D1PS54_9FIRM	24.6	2.2	34.043	34	34	47	27	3	44	89	105	148
14	MAM_wt	tr R6P689 R6P689_9CLOT	24.3	1.8	33.333	27	27	36	24	0	56	91	98	133
15	MAM_wt	tr A6LD60 A6LD60_PARD8	24.3	4.4	33.333	36	36	57	26	2	31	78	382	435
16	MAM_wt	tr B0G4V7 B0G4V7_9FIRM	25.4	1.1	32.500	29	29	40	26	1	68	106	150	189
17	MAM_wt	tr C0D4Q5 C0D4Q5_9FIRM	25.0	3.3	32.394	53	53	71	42	3	16	86	189	253
18	MAM_wt	tr R7JPH0 R7JPH0_9BACT	24.6	1.3	32.258	23	23	31	21	0	6	36	123	153
19	MAM_wt	tr E4VUA2 E4VUA2_BACT	27.7	0.35	31.818	33	33	44	30	0	1	44	280	323
20	MAM_wt	tr A7VJ13 A7VJ13_9CLOT	25.0	1.3	31.507	53	53	73	44	3	2	72	121	189
21	MAM_wt	tr B0NGV9 B0NGV9_CLOS	26.2	0.68	31.373	38	38	51	33	1	4	54	61	109
22	MAM_wt	tr Q5M488 Q5M488_STRT2	26.9	0.13	31.111	33	33	45	24	2	36	80	83	120
23	MAM_wt	tr R5U2R3 R5U2R3_9BACE	28.5	0.12	30.909	41	41	55	36	1	76	130	47	99
24	MAM_wt	tr A4EBV5 A4EBV5_9ACTN	27.3	0.21	30.612	34	34	49	30	2	57	102	292	339
25	MAM_wt	tr B0P145 B0P145_9CLOT	32.7	0.004	30.337	66	66	89	49	5	9	97	1241	1316
26	Clostridium_sp._SS2-1	tr D4LSX8 D4LSX8_9FIRM	24.3	2.2	30.303	24	24	33	23	0	21	53	78	110
27	MAM_wt	tr R7FZ59 R7FZ59_9FIRM	25.0	0.99	30.159	41	41	63	37	1	23	78	160	222
28	MAM_wt	tr B0PHX3 B0PHX3_9FIRM	23.5	4.7	29.787	34	34	47	31	2	3	48	173	218
29	MAM_wt	tr R6X7N3 R6X7N3_9PORP	24.3	3.3	29.545	61	61	88	53	4	4	86	547	630
30	MAM_wt	tr A5Z5Q6 A5Z5Q6_9FIRM	27.7	0.19	29.412	47	47	68	42	2	66	128	345	411
31	MAM_wt	tr B9YAA0 B9YAA0_9FIRM	28.1	0.15	29.268	30	30	41	29	0	1	41	68	108
32	MAM_wt	tr B7B871 B7B871_9PORP	25.0	1.9	28.846	32	32	52	24	3	54	96	204	251
33	MAM_wt	tr P96214 P96214_MYCTU	25.4	1.4	28.571	38	38	56	35	1	52	102	61	116
34	MAM_wt	tr R6LDT4 R6LDT4_9FIRM	24.3	1.5	28.070	39	39	57	36	1	49	100	17	73
35	MAM_wt	tr C9L847 C9L847_BLAHA	24.3	1.9	27.869	41	41	61	38	2	5	59	79	139
36	MAM_wt	tr D4JTH7 D4JTH7_9FIRM	23.9	2.3	27.778	40	40	54	39	0	15	68	1	54
37	MAM_wt	tr D4M698 D4M698_9FIRM	26.6	0.46	27.692	47	47	65	45	1	72	134	187	251
38	MAM_wt	tr B6VXT6 B6VXT6_9BACE	26.6	0.72	27.536	44	44	69	41	1	76	135	77	145
39	MAM_wt	tr C0X1L4 C0X1L4_ENTFL	26.6	0.39	34.211	27	27	38	24	1	14	50	150	187



## Analysis



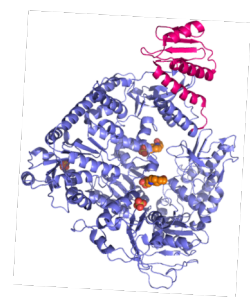
No other hit except in *Faecalibacterium Prausnitzii SL3\_3*

MAM strictly restricted to a niche of *F. Prausnitzii species*

			bitscore	evalue	pident	qcovs	qcovhsp	length	mismatch	gapopen
2	2EYQ:A PDBID CHAIN SEQUENCE	tr A0A0D6MVR3 A0A0D6MVR3_A	773	0.0	39.464	95	95	1120	642	15
3	2EYQ:A PDBID CHAIN SEQUENCE	tr R7JPZ2 R7JPZ2_9BACT	572	0.0	44.162	78	57	668	361	4
4	2EYQ:A PDBID CHAIN SEQUENCE	tr B0PG14 B0PG14_9FIRM	660	0.0	36.837	92	92	1094	641	18
5	2EYQ:A PDBID CHAIN SEQUENCE	tr R5UYH1 R5UYH1_9BACE	542	9.11e-175	41.740	74	57	678	374	5
6	2EYQ:A PDBID CHAIN SEQUENCE	tr B6W1Q4 B6W1Q4_9BACE	561	0.0	40.387	80	61	723	405	6
7	2EYQ:A PDBID CHAIN SEQUENCE	tr E5WUF0 E5WUF0_9BACE	538	1.75e-173	41.654	74	57	677	376	5
8	2EYQ:A PDBID CHAIN SEQUENCE	tr R6S1S3 R6S1S3_9BACE	544	6.91e-176	41.802	74	57	677	375	5
9	2EYQ:A PDBID CHAIN SEQUENCE	tr E4VYD7 E4VYD7_BACFG	544	1.05e-175	42.097	73	57	677	373	5
10	2EYQ:A PDBID CHAIN SEQUENCE	tr B3CGR2 B3CGR2_9BACE	543	2.44e-175	41.740	74	57	678	375	5
11	2EYQ:A PDBID CHAIN SEQUENCE	tr R7B293 R7B293_9BACE	650	0.0	37.827	83	83	994	552	14
12	2EYQ:A PDBID CHAIN SEQUENCE	tr C3QPJ5 C3QPJ5_9BACE	539	1.63e-173	41.445	74	57	678	376	5
13	2EYQ:A PDBID CHAIN SEQUENCE	tr C6Z4G4 C6Z4G4_9BACE	558	0.0	40.249	80	61	723	406	6
14	2EYQ:A PDBID CHAIN SEQUENCE	tr A0A0M1W4E3 A0A0M1W4E3_9	560	0.0	40.387	80	61	723	405	6
15	2EYQ:A PDBID CHAIN SEQUENCE	tr C3QBP8 C3QBP8_9BACE	537	7.85e-173	41.298	74	57	678	377	5
16	2EYQ:A PDBID CHAIN SEQUENCE	tr A0A108TBR9 A0A108TBR9_BA	534	5.30e-172	41.298	74	57	678	377	5
17	2EYQ:A PDBID CHAIN SEQUENCE	tr Q8AB59 Q8AB59_BACTN	536	1.41e-172	40.855	74	57	678	380	5
18	2EYQ:A PDBID CHAIN SEQUENCE	tr R7EH57 R7EH57_9BACE	539	2.31e-173	41.003	74	57	678	380	5
19	2EYQ:A PDBID CHAIN SEQUENCE	tr A6L2L5 A6L2L5_BACV8	558	0.0	40.249	80	61	723	406	6
20	2EYQ:A PDBID CHAIN SEQUENCE	tr C9LCA3 C9LCA3_BLAHA	576	0.0	44.250	72	60	687	375	2
21	2EYQ:A PDBID CHAIN SEQUENCE	tr R5L9D1 R5L9D1_9FIRM	593	0.0	46.311	72	55	637	337	2
22	2EYQ:A PDBID CHAIN SEQUENCE	tr C0CUY2 C0CUY2_9FIRM	601	0.0	46.330	78	57	654	345	2
23	2EYQ:A PDBID CHAIN SEQUENCE	tr R6P1H2 R6P1H2_9CLOT	578	0.0	46.271	80	57	657	339	5
24	2EYQ:A PDBID CHAIN SEQUENCE	tr R6PNN2 R6PNN2_9CLOT	578	0.0	45.723	69	56	643	340	3
25	2EYQ:A PDBID CHAIN SEQUENCE	tr B0NCF3 B0NCF3_CLOSV	570	0.0	44.428	71	59	682	364	3
26	2EYQ:A PDBID CHAIN SEQUENCE	tr A7VI89 A7VI89_9CLOT	656	0.0	35.992	88	88	1053	619	11
27	2EYQ:A PDBID CHAIN SEQUENCE	tr D4C8Z1 D4C8Z1_9CLOT	674	0.0	37.367	87	87	1033	597	12
28	2EYQ:A PDBID CHAIN SEQUENCE	tr B0NYB3 B0NYB3_9CLOT	596	0.0	47.088	73	52	601	314	1
29	2EYQ:A PDBID CHAIN SEQUENCE	tr A4E9V3 A4E9V3_9ACTN	534	1.92e-171	42.857	76	58	672	370	6
30	2EYQ:A PDBID CHAIN SEQUENCE	tr R6LB70 R6LB70_9FIRM	634	0.0	37.151	93	93	1074	612	16
31	2EYQ:A PDBID CHAIN SEQUENCE	tr R5WLH5 R5WLH5_9FIRM	477	7.11e-158	50.316	41	41	475	229	3
32	2EYQ:A PDBID CHAIN SEQUENCE	tr B0G3N8 B0G3N8_9FIRM	583	0.0	43.629	82	63	722	387	6
33	2EYQ:A PDBID CHAIN SEQUENCE	tr R7FNN1 R7FNN1_9FIRM	637	0.0	37.681	89	89	1035	568	14
34	2EYQ:A PDBID CHAIN SEQUENCE	tr C0X941 C0X941_ENTFL	684	0.0	36.750	92	92	1083	636	15
35	2EYQ:A PDBID CHAIN SEQUENCE	tr R6G373 R6G373_9FIRM	654	0.0	36.794	91	91	1098	600	22
36	2EYQ:A PDBID CHAIN SEQUENCE	tr D4JLK0 D4JLK0_9FIRM	226	1.79e-63	34.483	34	34	406	247	6
37	2EYQ:A PDBID CHAIN SEQUENCE	tr D4JVX9 D4JVX9_9FIRM	562	0.0	46.026	75	52	604	320	3
38	2EYQ:A PDBID CHAIN SEQUENCE	tr A5Z5K5 A5Z5K5_9FIRM	672	0.0	36.886	90	90	1079	612	21
39	2EYQ:A PDBID CHAIN SEQUENCE	tr D4KAB8 D4KAB8_9FIRM	564	0.0	46.384	71	55	636	333	4

## Post-analysis

- ➔ Mfd ubiquitous as expected
- ➔ Modeling of Mfd complex functional cycle that associates distinct functions with conformations
- ➔ Focus on *B. cereus* strains clinical and not to identify some mark of pathogeny-NO stress



Galaxy is a scientific workflow, data integration and data analysis platform to render computational biology accessible to research scientists who do not have computer programming or system administration knowledge.

Galaxy is an open, web-based platform:

- **Accessible:** users without programming experience can easily specify parameters and run tools and workflows.
- **Reproducible:** galaxy captures information so that any user can reproduce a complete computational analysis.
- **Transparent:** users share and publish analyses *via* the web, compute interactive, web-based a complete analysis.

The screenshot displays the Galaxy web interface with several key components highlighted:

- Tools available:** A blue arrow points to the left-hand sidebar containing a search bar and a list of tools, including 'Migale Tools', 'ProteoRE', and multiple instances of 'MetaFoldScan'.
- Submission form or view of input/output files:** An orange arrow points to the central configuration panel for 'MetaFoldScan 5'. This panel includes:
  - Protein(s) of interest:** A list of FASTA files: '91: MFD Chaîne A.fasta', '76: MFD.fasta', '56: Bacteroides\_caccae.fasta', '55: Acetobacter\_aceti\_NBRC\_14818.fasta', and '53: .fasta'.
  - Query File (genome):** A section with a 'Select/Unselect all' checkbox and an empty input field.
  - Minimum sequence size:** A text input field containing '100'.
  - Maximum sequence size:** A text input field containing '700'.
  - Execute button:** A blue button with a checkmark icon.
- History of results:** A purple arrow points to the right-hand 'History' panel, which shows a list of past analyses. The most recent entry is '93: Results archive' (208.01 GB), followed by '92: MFD Chaîne A / Acetobacter\_aceti\_NBR C\_14818.faa' and '91: MFD Chaîne A.fast a' (1 sequence, FASTA format).



**Galaxy / Dev-Migale** | Analyse de données | Workflow | Données partagées | Visualisation | Aide | Utilisateur | Using 15.8 GB

**Tools** | search tools | ProteoRE | MiRDeep2 | SNIPlay | **MetaFoldScan** | MetaFoldScan | MetaFoldScan 2 | MetaFoldScan 3 | MetaFoldScan 4 | MetaFoldScan 4 (1 CPU) | MetaFoldScan 5 | MetaFoldScan 4 (test) | MetaFoldScan 6 | MetaFoldScan Results Filtering

**Protein(s) of interest** | 20: ORF\_VH1\_dani\_wt.faa | 1: dorei-2prot.fasta | **Reference protein**

Will you select a genome from your history or use a built-in index? | Use a built-in genome index | **Query genomes**

**Select genome** |  Select/Unselect all

- Test\_Genome
- Test2\_Genome
- Acetobacter\_aceti\_NBRC\_14818
- Alistipes\_putredinis
- Anaerotruncus\_colihominis
- Bacteroides\_caccae
- Bacteroides\_dorei\_5\_1\_36\_D4
- Bacteroides\_doreiDSM17855
- Bacteroides\_eggerthii
- Bacteroides\_finegoldii
- Bacteroides\_fragilis\_3\_1\_12
- Bacteroides\_intestinalis
- Bacteroides\_ovatus\_ATCC\_8483
- Bacteroides\_pectinophilus
- Bacteroides\_sp.\_2\_2\_4
- Bacteroides\_sp.\_4\_3\_47FAA
- Bacteroides\_sp.\_9\_1\_42FAA
- Bacteroides\_sp.D1
- Bacteroides\_stercoris
- Bacteroides\_thetaiotaomicron\_VPI-5482
- Bacteroides\_uniformis
- Bacteroides\_vulgatus\_ATCC\_8482
- Bacteroides\_xylanisolvans\_XB1A
- Blautia\_hansenii
- butyrate-producing\_bacterium\_SS3-4
- Butyrivibrio\_crossotus
- Clostridium\_asparagiforme
- Clostridium\_leptum
- Clostridium\_nexile
- Clostridium\_scindens
- Clostridium\_sp.\_L2-50
- Clostridium\_sp.\_M62\_1
- Clostridium\_sp.\_SS2-1
- Collinsella\_aerofaciens
- Coprococcus\_comes
- Coprococcus\_eutactus
- Dorea\_formicigenerans\_ATCC\_27755
- Dorea\_longicatena
- Enterococcus\_faecalis\_TX010
- Eubacterium\_hallii
- Eubacterium\_rectale\_M104\_1
- Eubacterium\_siraeum\_70\_3
- Eubacterium\_ventriosum
- Faecalibacterium\_prausnitzii\_SL3\_3
- Holdemania\_filiformis
- Mycobacterium\_tuberculosis

**History** | Rechercher des données | **MetaFoldScan** | 10 shown, 6 deleted | 15.83 GB

- 30: Log
- 29: ORF\_VH1\_dani\_wt / Bacteroides\_thetaiotaomicron\_VPI-5482 & butyrate-producing\_bacterium\_SS3-4
- 28: Log
- 27: ORF\_VH1\_dani\_wt / Faecalibacterium\_prausnitzii\_SL3\_3
- 22: ORF\_VH1\_dani\_wt / Bacteroides\_thetaiotaomicron\_VPI-5482 & butyrate-producing\_bacterium\_SS3-4
- 21: ORF\_VH1\_dani\_wt / Faecalibacterium\_prausnitzii\_SL3\_3

**MetaFoldScan 5 MetaFoldScan, .... (Galaxy Version 1)** Options

**Protein(s) of interest**

- 91: MFD Chaîne A.fasta
- 76: MFD.fasta
- 56: Bacteroides\_caccae.fasta
- 55: Acetobacter\_aceti\_NBRC\_14818.fasta
- 53: .fasta

**Query File (genome)**

Select/Unselect all

ace

**Acetobacter\_aceti\_NBRC\_14818**

100

**Maximum sequence size**

700

Execute

METAFOLDSCAN - ...  
DESCRIPTION

*Work in progress*

**Promising as a web tool to screen a couple of genomes or more**

Associates accuracy and autonomy

**MetaFoldScan 5 MetaFoldScan, .... (Galaxy Version 1)**
Options

**Protein(s) of interest**

91: MFD Chaîne A.fasta

76: MFD.fasta

56: Bacteroides\_caccae.fasta

55: Acetobacter\_aceti\_NBRC\_14818.fasta

53: .fasta

**Query File (genome)**

Select/Unselect all

Acetobacter\_aceti\_NBRC\_14818

**Maximum sequence size**

Execute

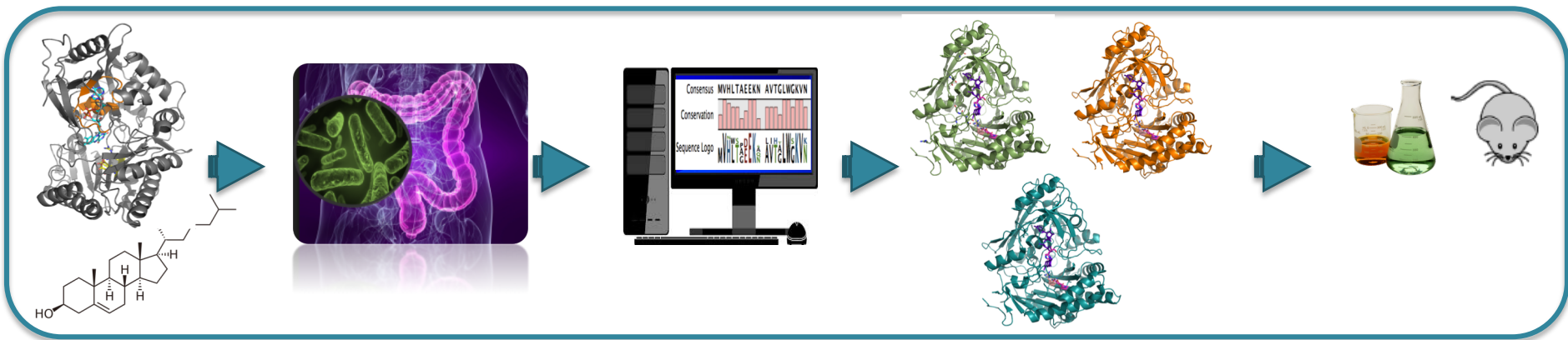
METAFOLDSCAN - ... *Work in progress*

## DESCRIPTION

**Main hurdle** : scaling up from core genome to 10 millions of genes

→ **Filtering**: size (number of residues), signal peptide, secretome, etc.

→ **Data storage**, including the temporary data generated crash after 22 genomes and 2 months of calculations



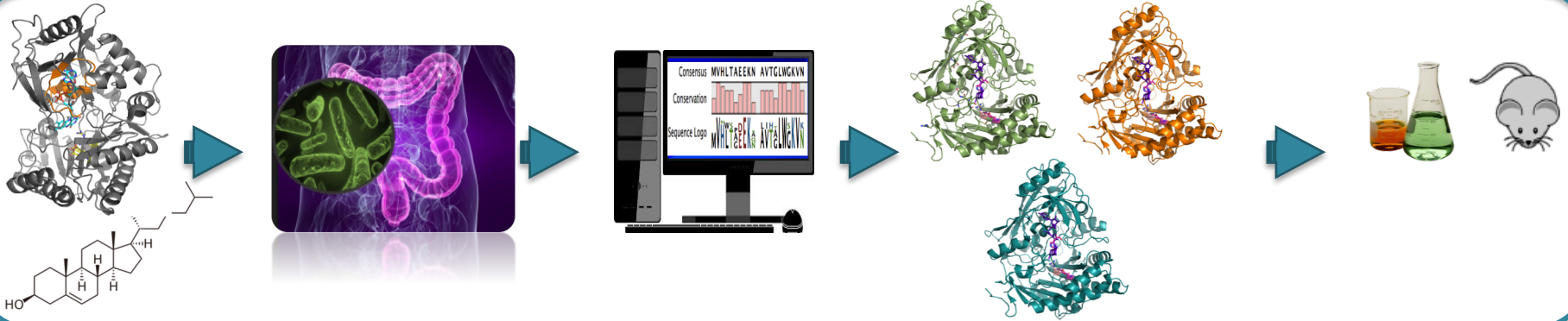
- Collaboration with B. Offmann –University of Nantes, France, for FORSA
- Structural annotation of protein sequences: from Hidden Markov Model profiling to alphabet protein
- FORSA scores the compatibility of any amino acid sequence with any fold represented in the form of protein blocks sequences (1D representation of 3D information).
- FORSA is highly scalable to large datasets
- FORSA uses both local and global threading through parallelising large amount of analyses.

➡ **One genome one protein one hour on a laptop**



Deployable, scalable to 1,000 genomes and more, early 2019

# Towards the Structural Screening of Microbial Ecosystems



## Outline

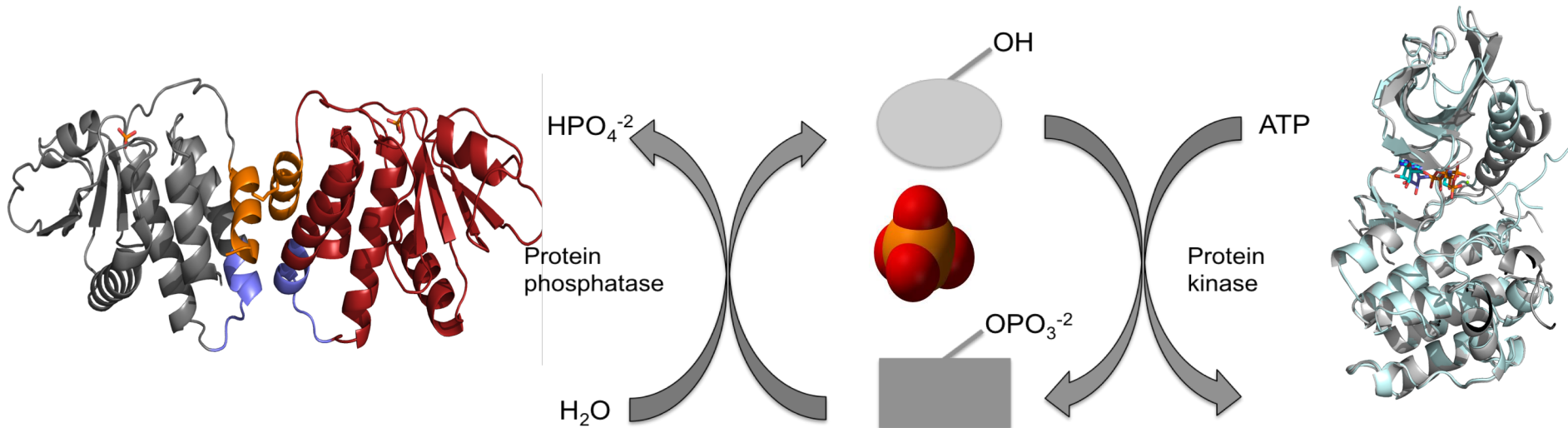
- Cholesterol conversion in the gut microbiota: the mystery enzyme(s) ?
- MetaFoldScan project : 3D screening of the gut microbiota
- **Conclusions and perspectives**

## Collaborative projects: actors of the dialog gut /host

Cell signaling proteins



Responsible for the dialog  
bacteria/bacteria et host/bacteria

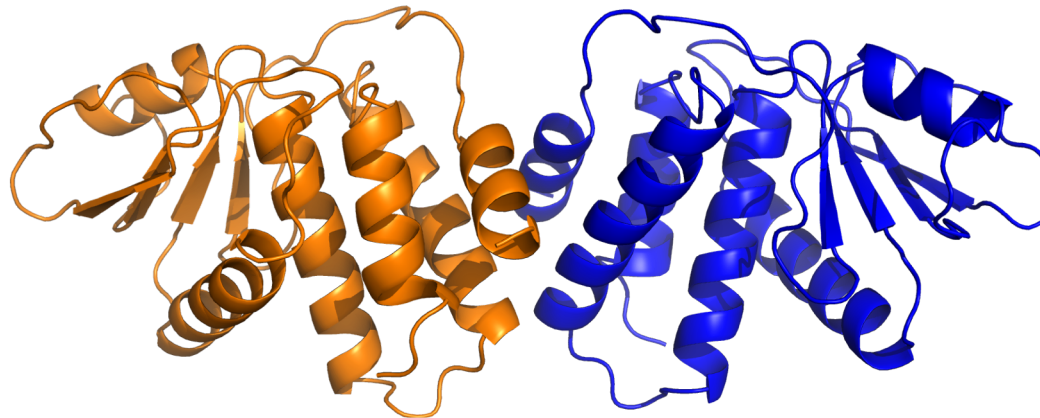


Dialog between in and outside: Tyr-phosphatase & kinases amplify a signal from the membrane

# Collaborative project 1: identify and characterize Tyr-phosphatases

Dr. A. Villarino & Dr. M. Berois, Universidad de Montevideo, Uruguay

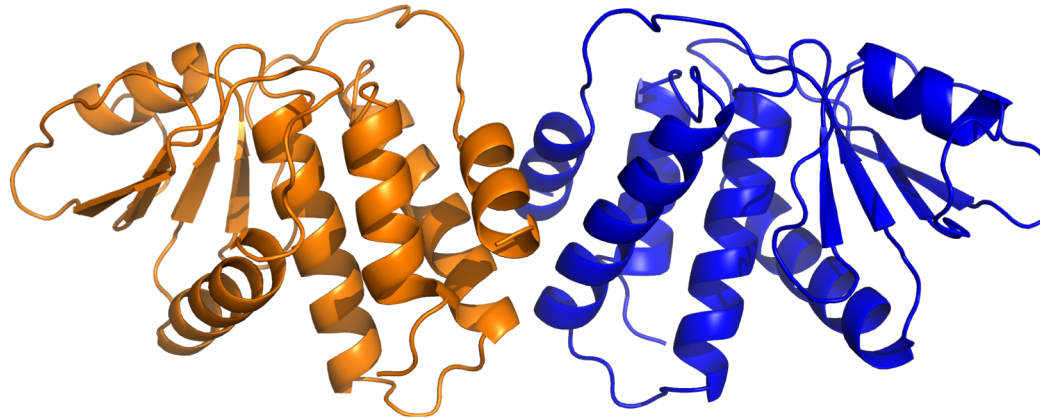
## Tyr-Phosphatase






# Collaborative project 1: identify and characterize Tyr-phosphatases



Dr. A. Villarino & Dr. M. Berois, Universidad de Montevideo, Uruguay



## Tyr-Phosphatase



### Intracellulaire pathogen ORF

-  Actif as a dimer
-  Hijacks the immune system
-  Dephosphorylates an essential cell mediator cellulaire essentiel

 2 CR1, 1AI, 2 Master  
 MaIAGE, Migale

 2 PR, 2 Master  
 Facultad de Ciencias, Montevideo, Uruguay

 Ecos-Sud Uruguay  
 grants

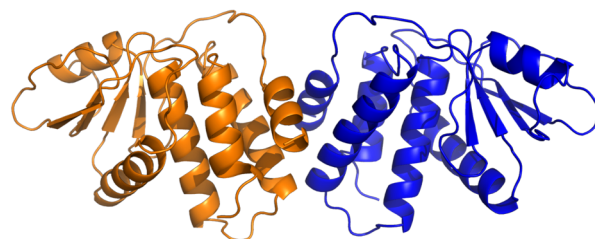


# Screen the gut microbiota to target Tyr-phosphatases

Dr. A. Villarino & Dr. M. Berois, Universidad de Montevideo, Uruguay

Antigène in the new Crohn detection kit: Tyr-phosphatase PtpA

*Mycobacterium avium ssp paratuberculosis*



viral Tyr-phosphatase

Bacteria



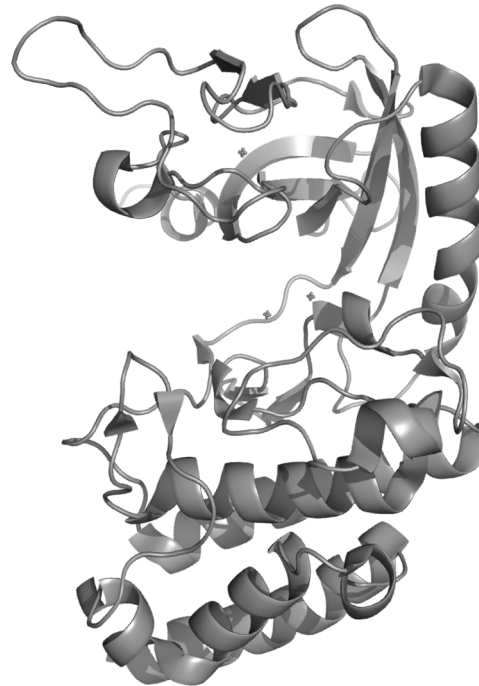
- Conservation
- Distribution
- Structure/fonction

**Structural homologs of Tyr-phosphatases**




## Collaborative project 2: identify a pathway involved in metabolism



Dr. M.N. Lisa, IBR-Conicet, Rosario Argentina

### Ser/Thr Protein Kinase



*Mycobacterium tuberculosis*

-  Central metabolism
-  Operon of 7 proteins
-  Detects nutrients availability

 1DR2,1 CR1, 1AI  
 MalAGe, Migale

 1PI  
 IBR, Rosario, Argentine

 **MEM**  
CONICET  


Grants

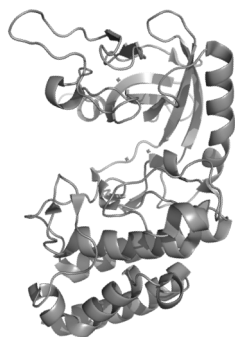
Deposited

# Screen and target kinase PknG and its upstream and downstream partners

Dr. M.N. Lisa, IBR-Conicet, Rosario Argentina

Antigène in the new Crohn detection kit: Ser/Thr Protein Kinase PknG

*Mycobacterium avium ssp paratuberculosis*



*Mycobacterium tuberculosis*

Actinobactéries



- Conservation
- Distribution
- Structure/fonction

## Structural homologs of Ser/Thr kinases and the 6 partners of the signaling pathway

Actinobacteria: 250 genomes, 3 months computation time, heat-map with R analysis

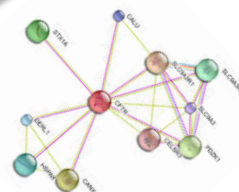
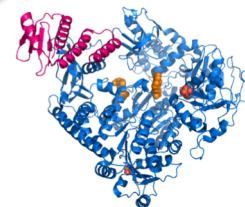
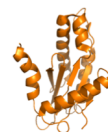
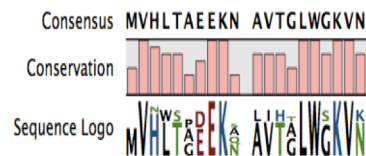
❖ Strong positioning at the interface structural bioinformatics & Biology

❖ Innovative webserver to achieve gut ecosystem exploration.

Needs scaling up & tuning (filters)

Experimental validation of the structural hits

❖ High valorisation potential



HTS of structures  
Structure –Function –Annotation  
Functional clusters and Pathways

Originality /prevalence of folds  
Disease/health -associated  
patterns of diversity

... Towards new targets with Health impact

... Towards other microbiotas



# Acknowledgments



Véronique Martin

Jean-François Gibrat

Thomas Lacroix



Plateforme de BioInformatique - INRA Jouy en Josas

Sandra Derozier

Valentin Loux



Fabienne Béguet-Crespel

Jean-Marc Chatel

Philippe Gérard

Catherine Juste

Nalini Rama Rao



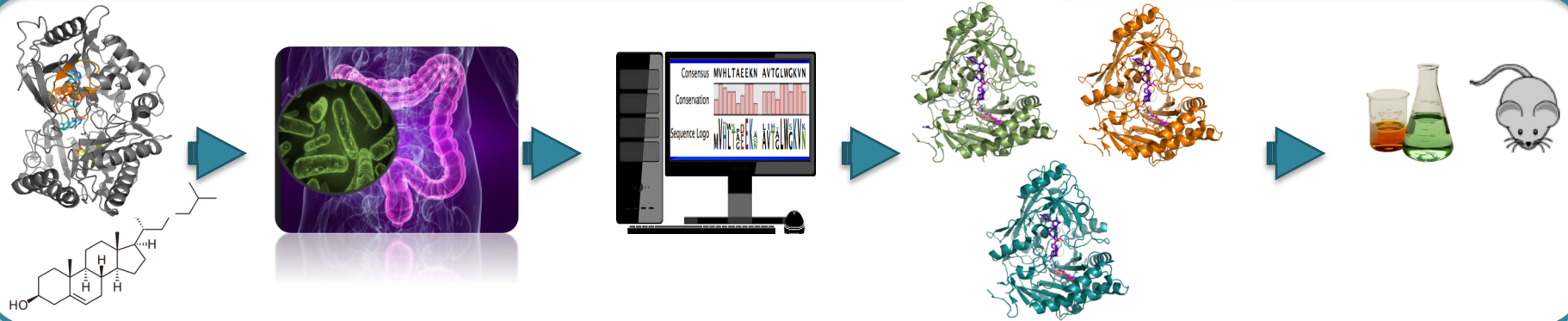
FACULTAD DE  
**CIENCIAS**  
UDELAR | [fcienc.edu.uy](http://fcienc.edu.uy)

Andrea Villarino

Mabel Berois

Danilo Segovia

Dario Porley



UNIVERSITÉ DE NANTES

Bernard Offmann

Surbhi Dingra

MEM Metaprogram



Institut Pasteur

Mariano Martinez

Pedro Alzari

CONICET



María Natalia Lisa

# Acknowledgments



Véronique Martin

Jean-François Gibrat

Thomas Lacroix



Plateforme de BioInformatique - INRA Jouy en Josas

Sandra Derozier

Valentin Loux



Fabienne Béguet-Crespel

Jean-Marc Chatel

Philippe Gérard

Catherine Juste

Nalini Rama Rao



FACULTAD DE CIENCIAS  
UDELAR | [fcienc.edu.uy](http://fcienc.edu.uy)

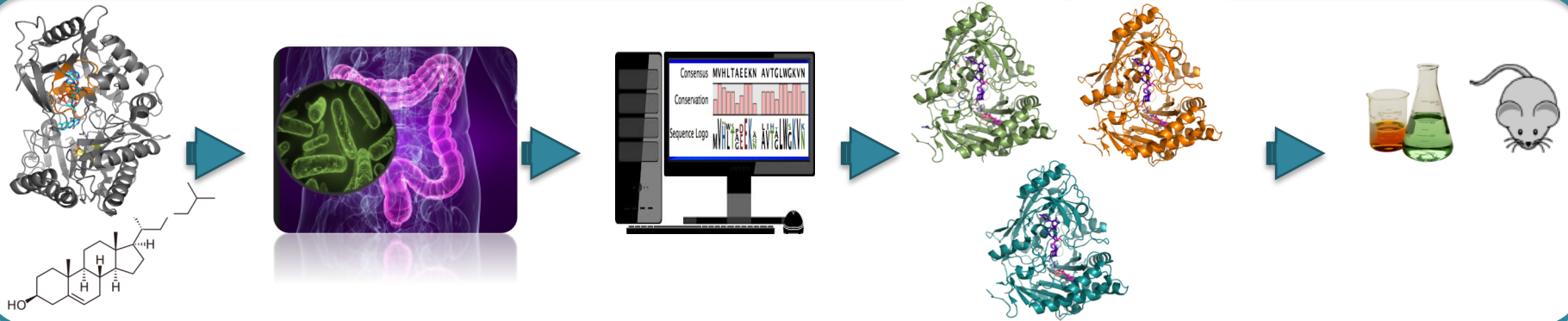
Andrea Villarino

Mabel Berois

Danilo Segovia

Dario Porley

And you for your kind attention ...



UNIVERSITÉ DE NANTES

Bernard Offmann

Surbhi Dingra

MEM Metaprogram



Institut Pasteur

Mariano Martinez

Pedro Alzari

CONICET



María Natalia Lisa

## WP1



## Bioinformatics development

**Integration of MetaHit datasets, filtering, scanning and 3D fold detection**

**G. André-Leroux**

- ✧ Set up of Metahit\* browsing. Splitting of genomes. Filtering. Connection to pdb
- ✧ HMM profiling of MAM and MFD targets 3D candidates: from hits to leads
- ✧ Analysis of specificity, promiscuity, identification of associated pathways

## WP2



### MAM hits

**JM. Chatel**

- ✧ *In vitro* validation of MAM –like leads using HEK293 NFκB luciferase reporter assay
- ✧ *In vivo* validation of MAM-like leads using IBD colitis in mouse model
- ✧ MAM production, purification, crystallization, structural characterization

## Experimental validation of targets

## WP3



### MFD hits

**N. Rama Rao**

- ✧ Clone MFD-like 10 best hits in *E.coli* deficient MFD mutant
- ✧ Test resistance to NO for “Complemented *E.coli* strains”
- ✧ Assess the function and specificity of MFD leads in prevalent species such as *Bacteroidetes* & *Firmicutes*

## WP4



## Applications - Impacts

### Biologist friendly web-tool interface

**V. Loux**

- ✧ Production of command line pipeline distribution.
- ✧ Integration into Galaxy.
- ✧ Integration into training sessions within Migale Bioinformatics cycle
- ✧ Extension to other microbiotas

★ <http://genome.jouy.inra.fr/Insyght>

➡ Powerful tool for genes comparison and sytheny inferences

Sythenies : co-localization of homologous loci accross species

Comparison 2 vs 2 with a reference genome.

BDBH method Bi-Directional Best Hit i.e best reciprocal match between two proteomes.

Gene comparison. Sythenies visible

Info and options    Contextual menu: browse multiple homologies per gene