



Towards the structural screening of mircobial ecosystems

Gwenaëlle André-Leroux

► To cite this version:

Gwenaëlle André-Leroux. Towards the structural screening of mircobial ecosystems. 2. Seminaire, Oct 2017, Montevideo, Uruguay. hal-02785773

HAL Id: hal-02785773

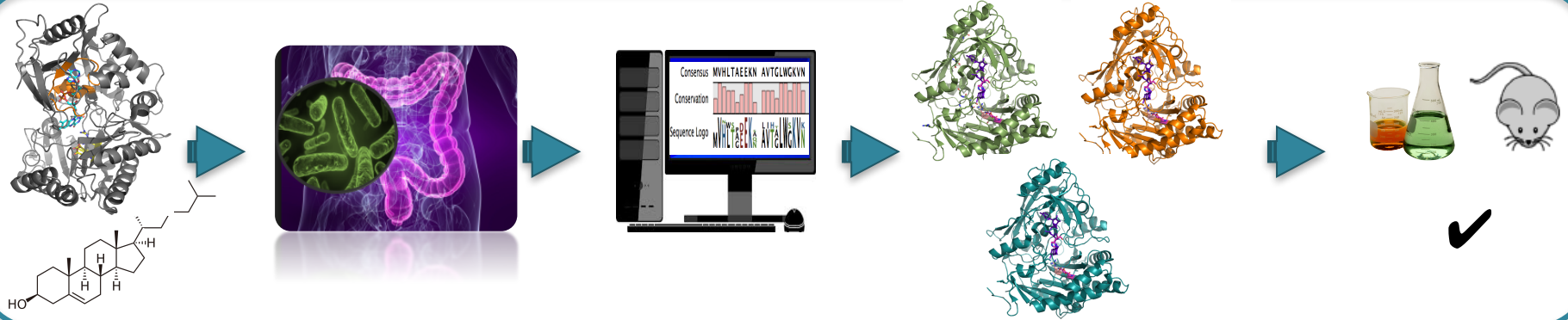
<https://hal.inrae.fr/hal-02785773>

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.

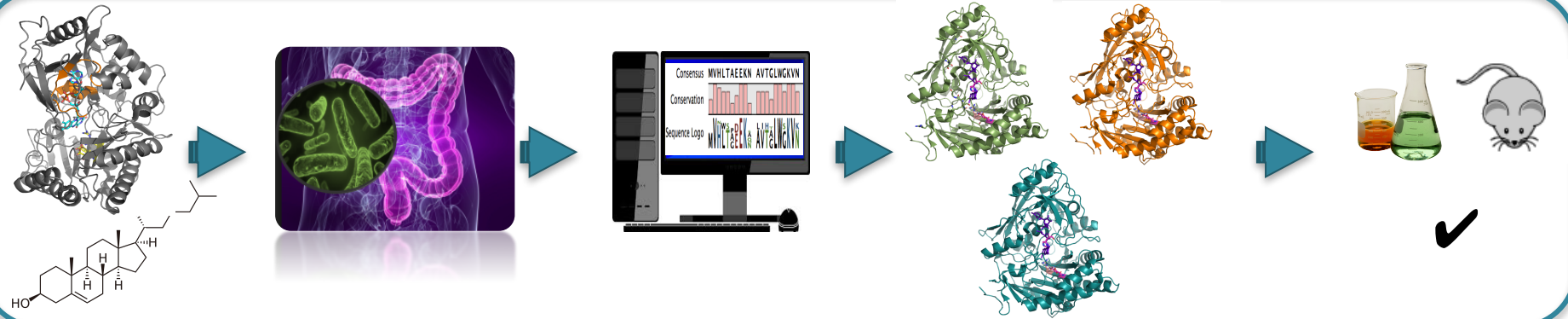
Towards the Structural Screening of Microbial Ecosystems



Gwenaëlle André-Leroux

Montevideo October 23rd

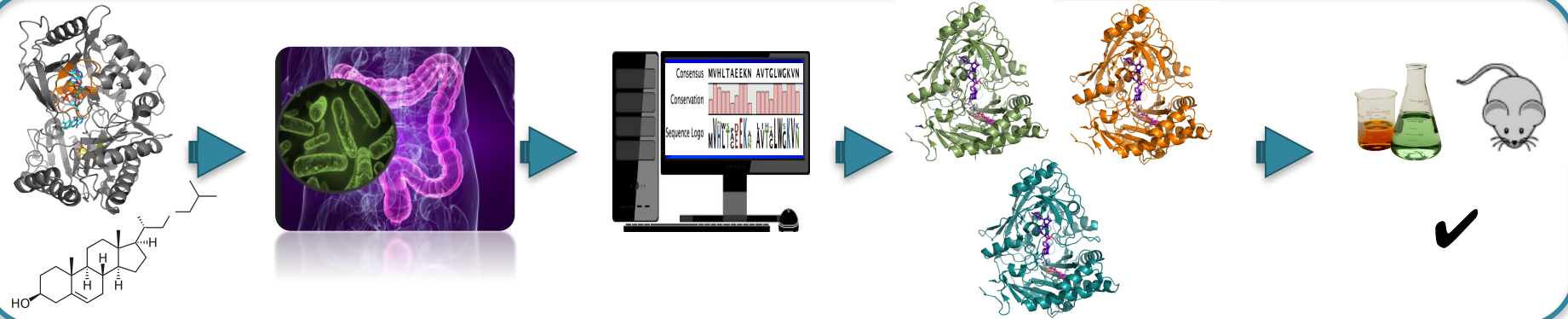
Towards the Structural Screening of Microbial Ecosystems



Outline

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

Towards the Structural Screening of Microbial Ecosystems



Outline

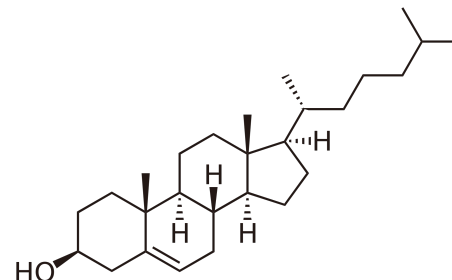
- Cholesterol conversion in the gut microbiota: the mystery enzyme(s) ?
- MetaFoldScan project : 3D screening of the gut microbiota
- Conclusions and perspectives

★ 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 desquamed 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

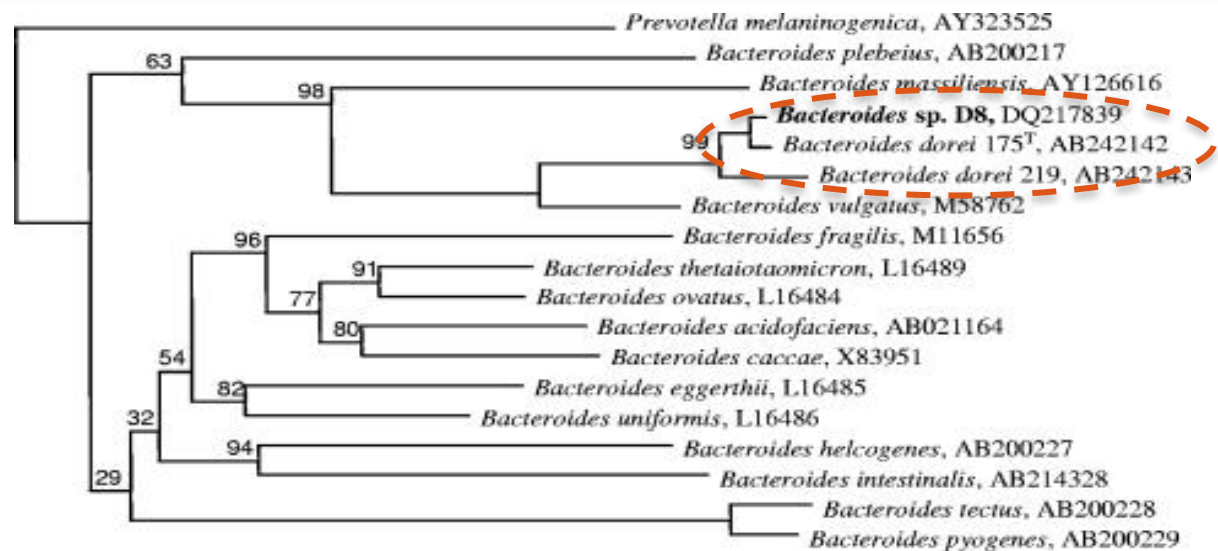
★ One from « high converter » human: isolate from the *Bacteroides* phylum, named *Bacteroides* sp. Strain D8. Gram- anaerobie. Gérard *et al*, 2007

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

Mammalian intestinal microbiota
>1,000 genera of bacteria.
2 phyla

Bacteroidetes

Firmicutes



★ 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, named *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^T sequenced and assembled at MaIAGE

Phylogenetically close

Phylogenetically remote



Strain \ Medium	SBM	BCM
<i>B. Dorei 175^T</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



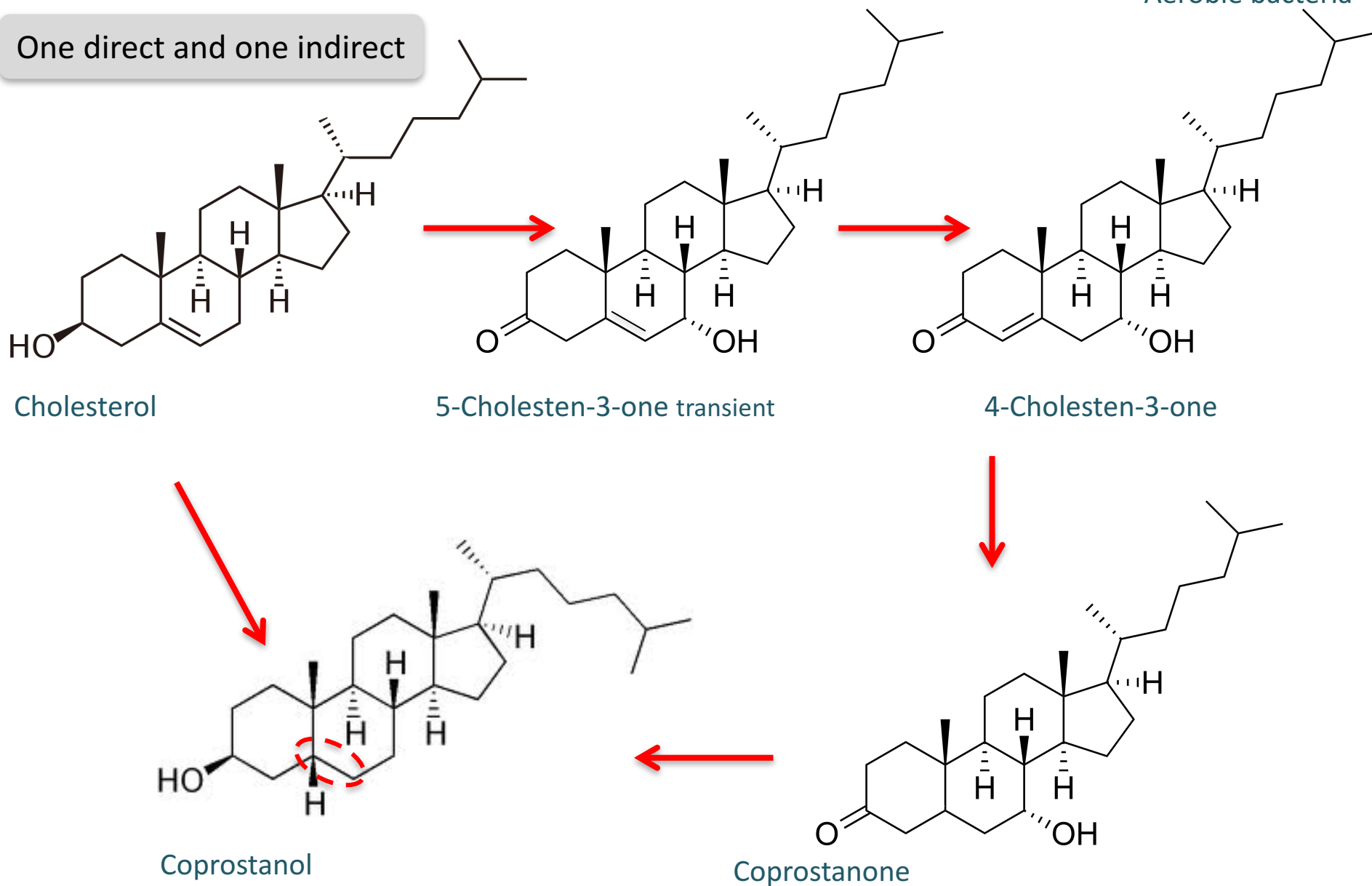
Metabolic pathway for cholesterol to coprostanol conversion in *B. dorei D8*?



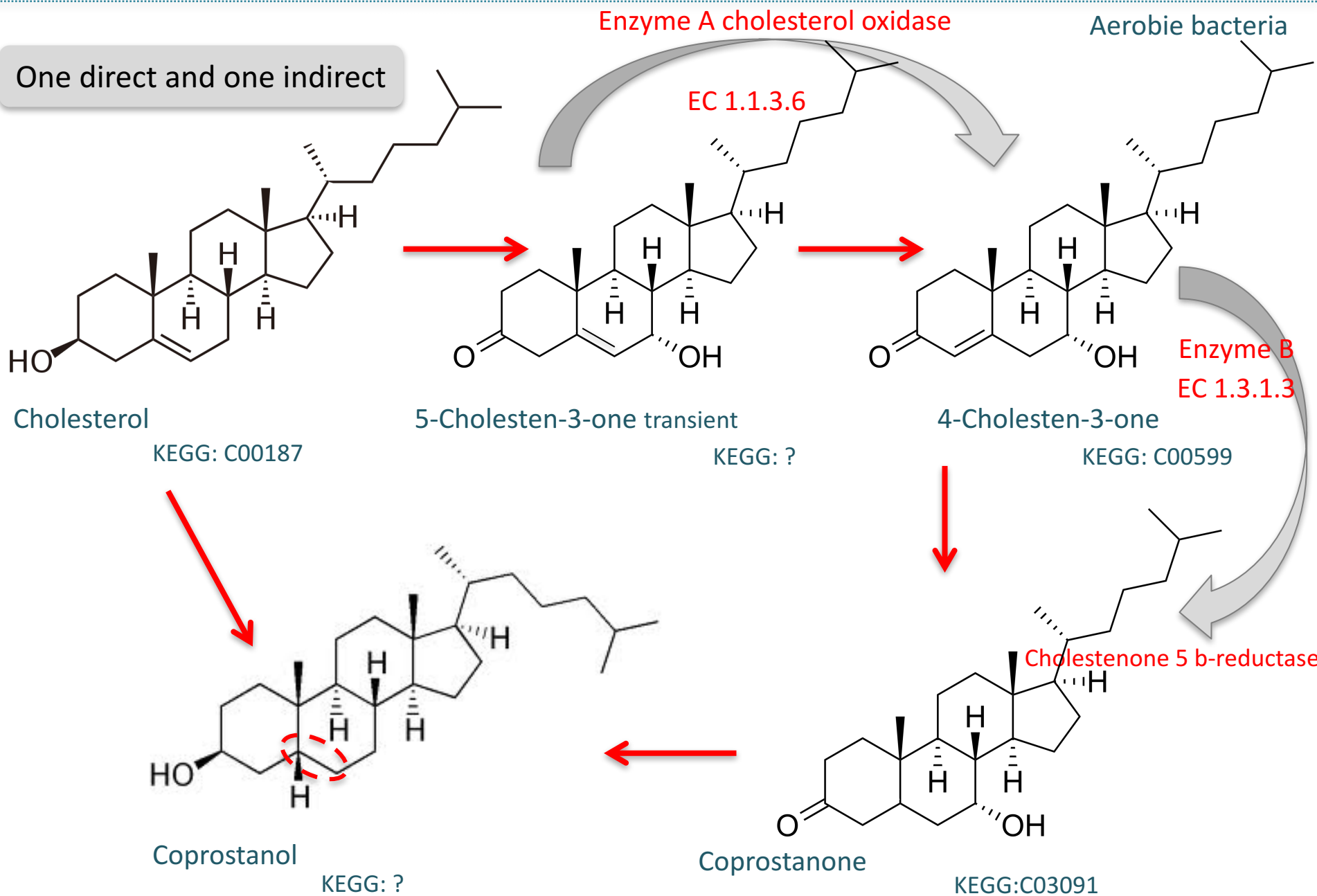
Which are the enzymes of the catabolic pathway that degrade cholesterol into coprostanol?

One direct and one indirect

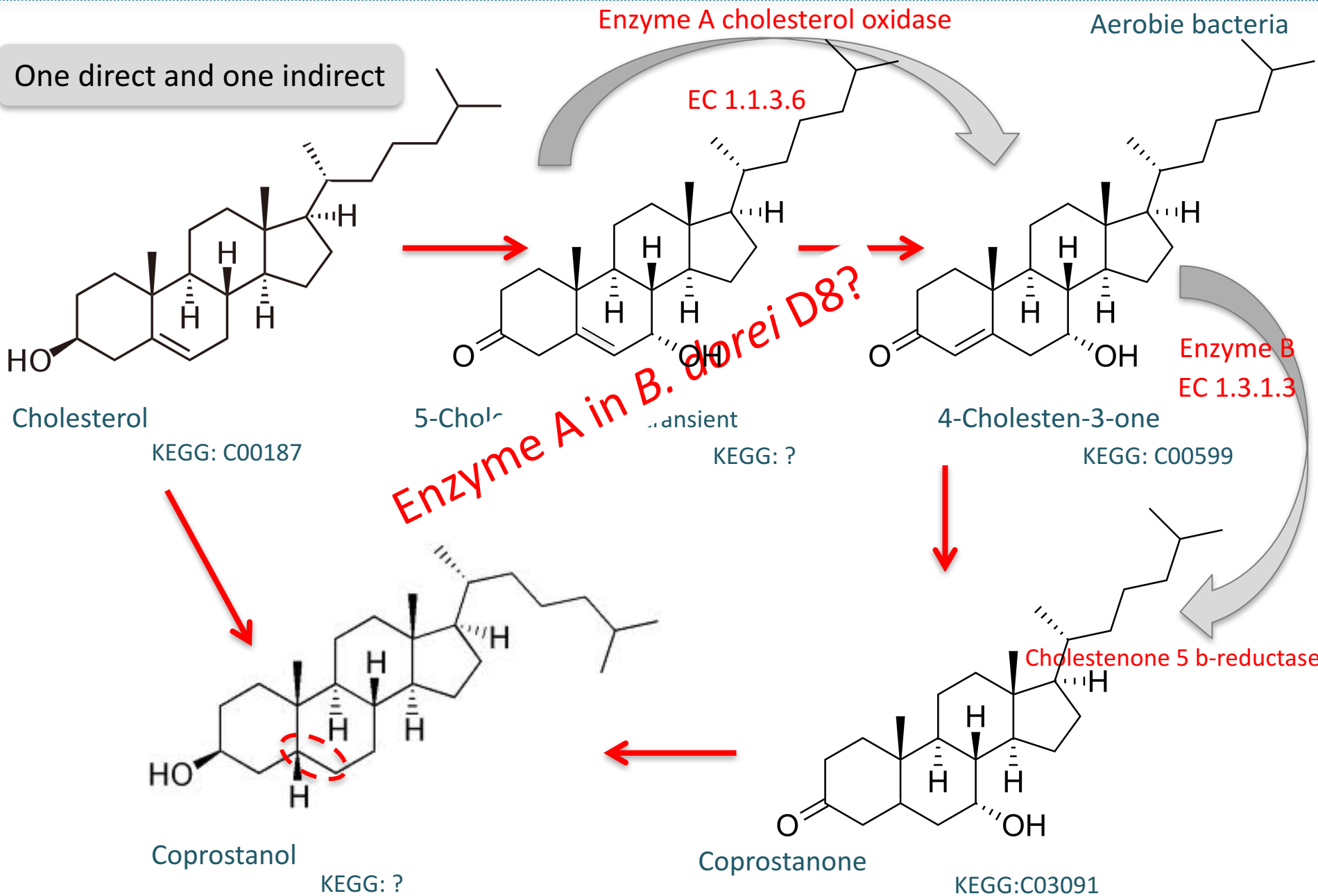
Aerobie bacteria



One direct and one indirect



One direct and one indirect



Server of protein structure prediction

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

Method: profile-profile comparison tools. Profiles are calculated from a MSA of related sequences collected from Hhblits (Psi-blast+). Profile is a 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



MPI
Bioinformatics
Toolkit

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

Hhblits HHpred HMMER PatternSearch ProtBLAST/PSI-BLAST

HHpred ?

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 it against the PDB_mm database, select the top-scoring template(s) and click on 'Create model using selection'. This will 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

Protein data bank

Custom JobID

Submit Job

[Help](#) [FAQ](#) [Privacy Policy](#) [Imprint](#) [Contact](#) [News](#)

© 2008-2017, Dept. of Protein Evolution, Max Planck Institute for Developmental Biology, Tübingen

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

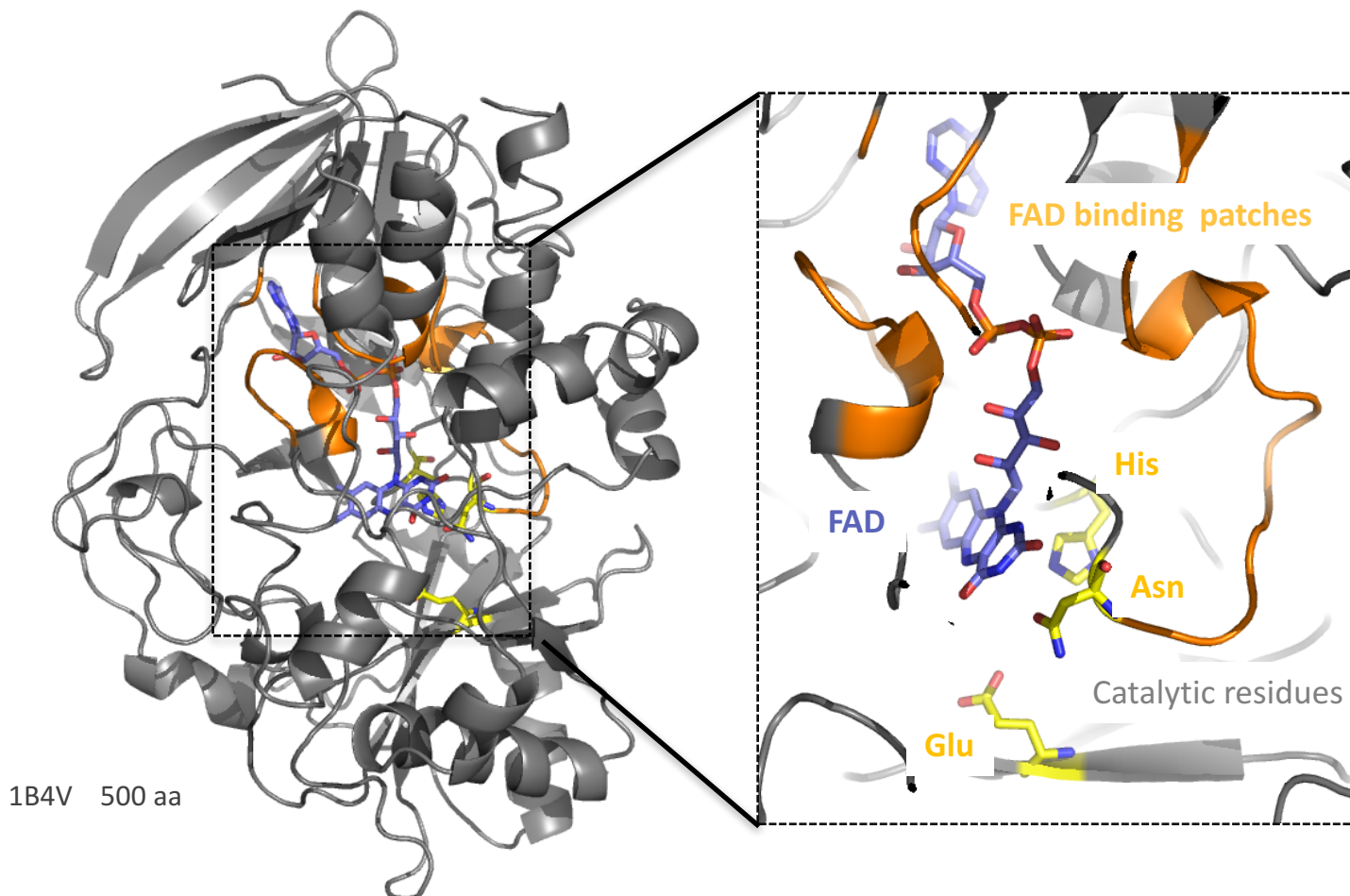
- 1- Split of each fasta sequence in the genome -> more than 3,400 for each genome of *B. dorei*
- 2- Multiple sequence alignment for each fasta: 'HHblits' > PSI-Blast
- 3- Prediction & addition of SS elements: 'adssl.pl'
- 4- Profiling & comparaison with the 3D template « bank »: 'HHBlitsdb.pl' & 'HHsearch'
- 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.



1B4V. pdb & fasta

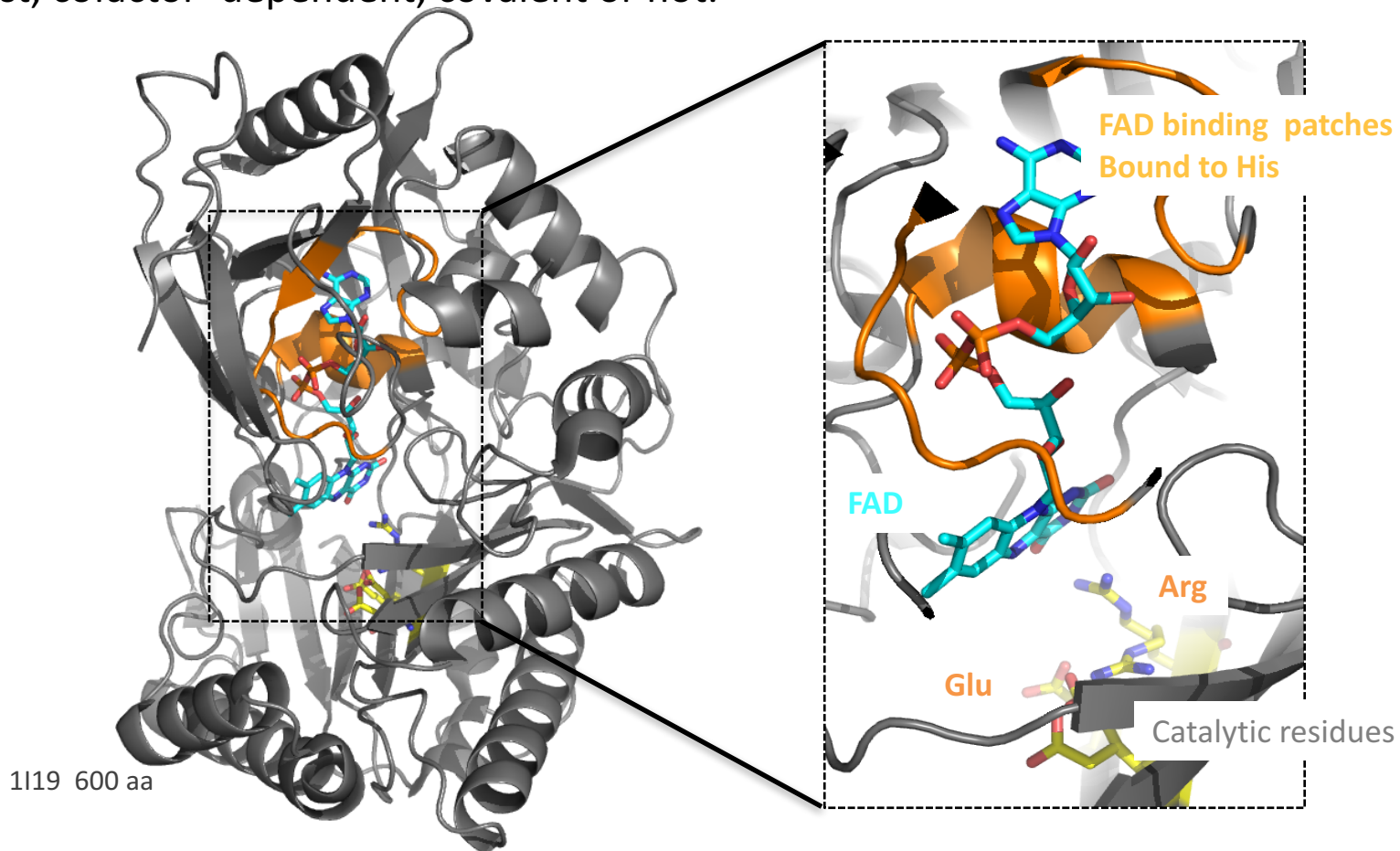
Cholesterol oxydase from *Streptomyces* Yue et al, 1999 - 1.5 Å

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.



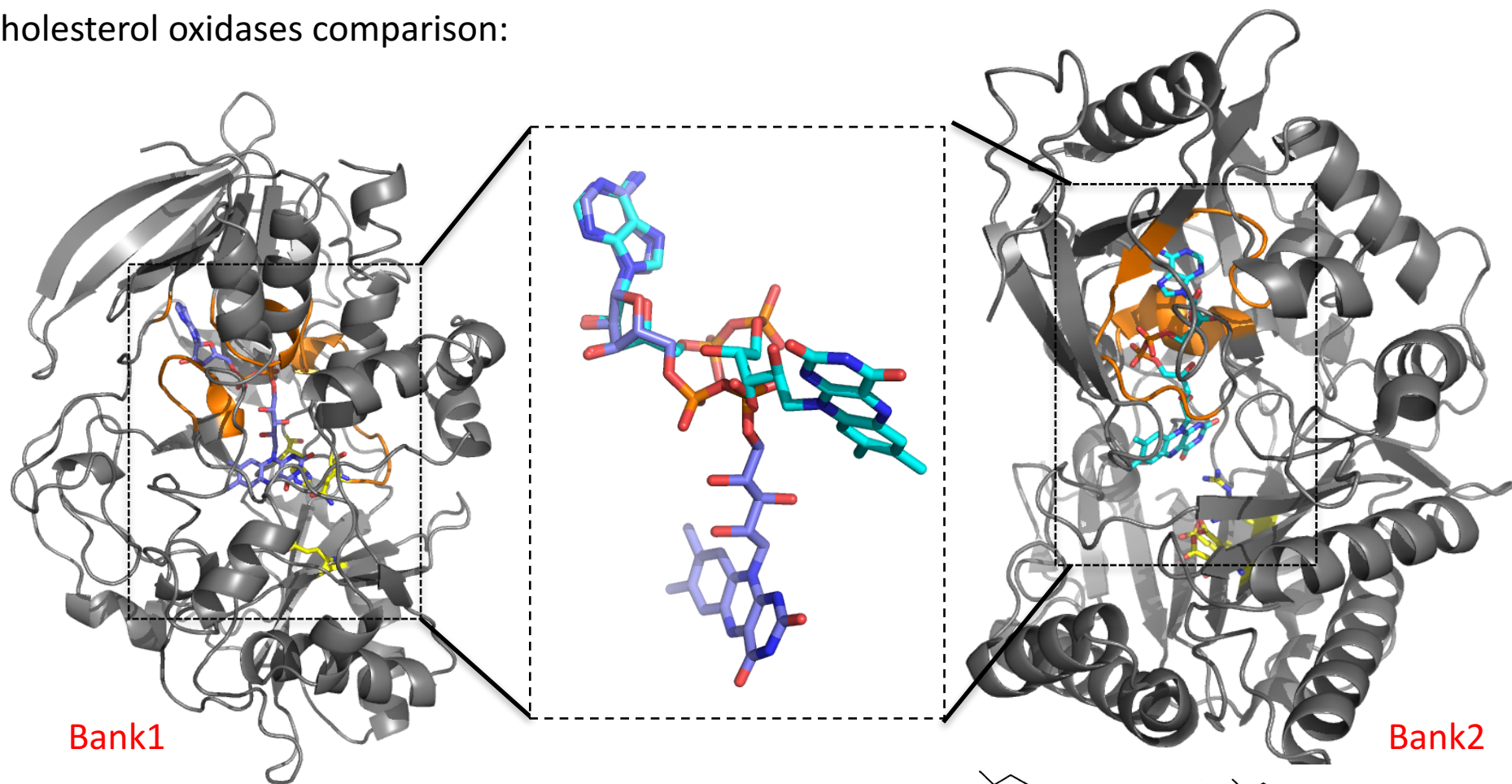
3JS8. pdb & fasta

1119.pdb & fasta

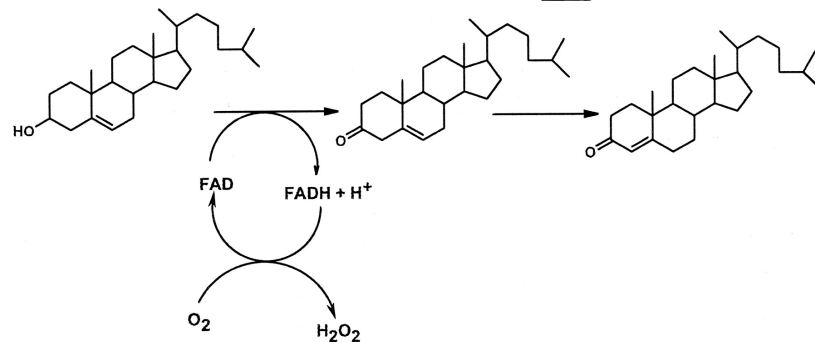
Cholesterol oxydase from *Chromobacterium* sp. DS-1 Sagermann et al, 2010 -1.5 Å

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

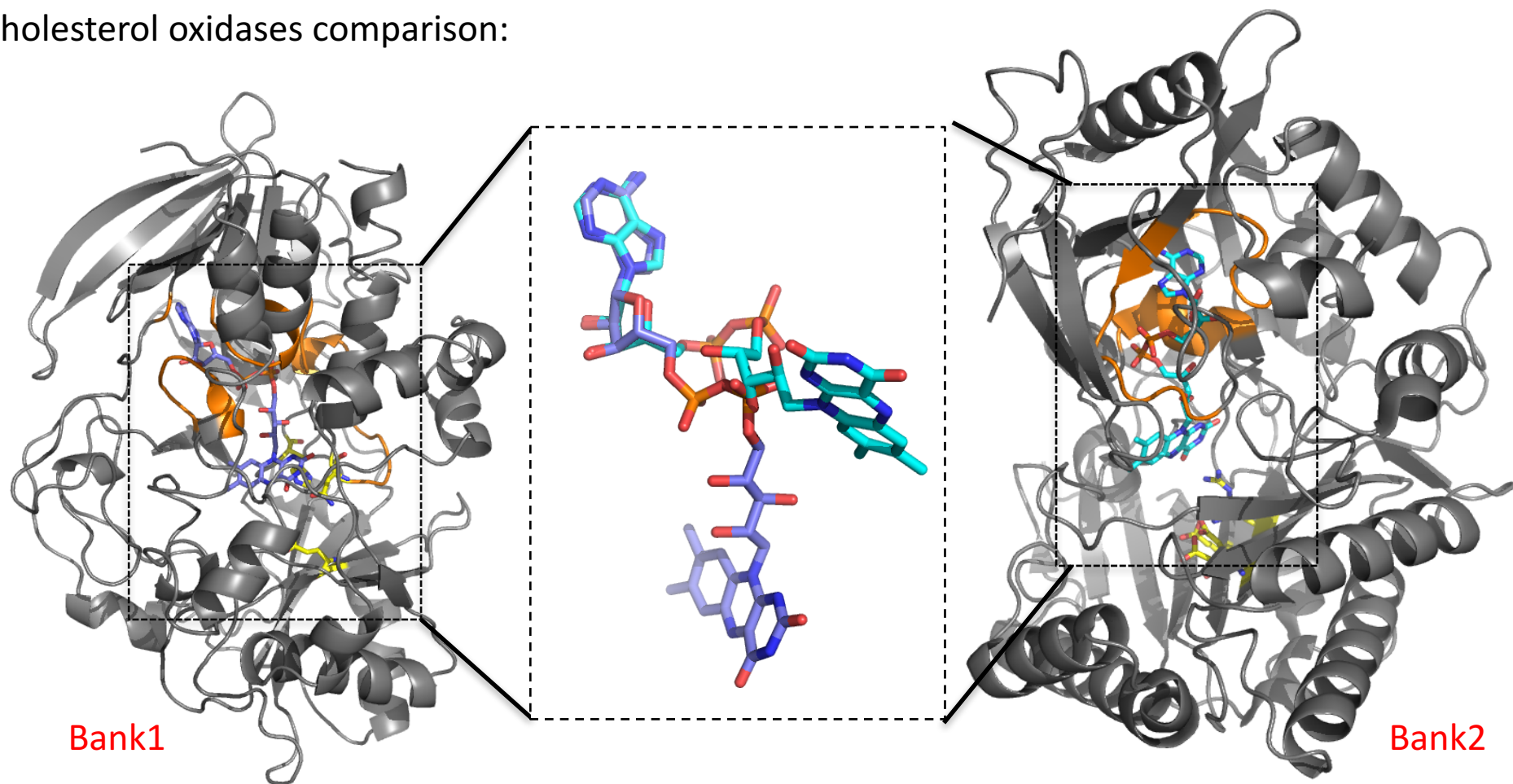
Cholesterol oxidases comparison:



- ★ 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



Analysis 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* D8 Sorted by Prob > 95 and size cut off > 50 amino acid residues

99.9	75 aa	bdd8_00352.fasta.a3m.hhm.out.hhmakeBK1	→ L-aspartate oxidase
99.3	69 aa	bdd8_00254.fasta.a3m.hhm.out.hhmakeBK1	→ Putative uncharacterized protein
99.0	62 aa	bdd8_03708.fasta.a3m.hhm.out.hhmakeBK1	→ Succinate Dehydrogenase flavoprotein subunit
97.6	61 aa	bdd8_03122.fasta.a3m.hhm.out.hhmakeBK1	→ Pyridine nucleotide-disulfide oxidoreductase



```

Query      Bdd8_00352 L-aspartate oxidase lcl|Agmial:UEPSD|D8#5230
Match_columns 526
No of_seqs 337 out of 3311
Neff 9.0
Searched_HMMs 2
Date Mon Feb 23 14:43:03 2015
Command hsearch -i /projet/mig/work/vmartin/D8/D8.batch4.20150203/fasta_split/hhm/bdd8_00352.fasta.a3m.hhm -d /projet/mig/work/vmartin/D8/BANK1/base/BANK1_hhm_db -cov 10 -v -cpu 8 -o /projet/mig/work/vmartin/D8/D8.batch4.20150203/fasta_split/hhm/bdd8_00352.fasta.a3m.hhm.out.hhmakeBK1

```

No	Hit	Prob	E-value	P-value	Score	SS	Cols	Query HMM	Template HMM
1	1B4V:A	99.1	5E-14	2.5E-14	117.2	16.3	65	133-208	223-290 (504)
2	1C0Y:A	99.1	6.5E-14	3.2E-14	116.5	15.8	64	133-207	228-294 (507)
3	1B4V:A	10.6	0.017	0.0087	28.1	5.1	36	3-38	4-40 (504)
4	1C0Y:A	4.4	0.048	0.024	24.8	4.7	35	3-37	10-45 (507)

No 1
>1B4V:A
Probab=99.12 E-value=5e-14 Score=117.19 Aligned cols=65 Identities=12% Similarity=0.127 Sum probs=48.4

[illegible]

No 2
>IC0Y:A
Probab=99.10 E-value=6.5e-14 Score=116.46 Aligned cols=64 Identities=14% Similarity=0.189 Sum probs=47.5

[illegible]

No 3
>1B4V:A
Probab=10.56 E-value=0.017 Score=28.10 Aligned cols=36 Identities=19% Similarity=0.209 Sum probs=32.8

[illegible]

★ *B. dorei* D8 Sorted by Prob > 95 and size cut off > 50 amino acid residues

99.9 172 aa bdd8_03618.fasta.a3m.hhm.out.hhmakeBK2 → UDP-N-acetylenolpyruvoylglucosamine reductase

[illegible]

★ *B. dorei* D8 Sorted by Prob > 95 and size cut off > 50 amino acid residues

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^T Sorted by Prob > 95 and size > 50 amino acid residues

99.9 % 172 aa bacdor_03524.fasta.a3m.hhm.out.hhmakeBK2 → UDP-N-acetyl muramate dehydrogenase

★ *B. dorei* D8 Sorted by Prob > 95 and size cut off > 50 amino acid residues

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^T Sorted by Prob > 95 and size > 50 amino acid residues

99.9 % 172 aa bacdor_03524.fasta.a3m.hhm.out.hhmakeBK2 → UDP-N-acetyl muramate dehydrogenase

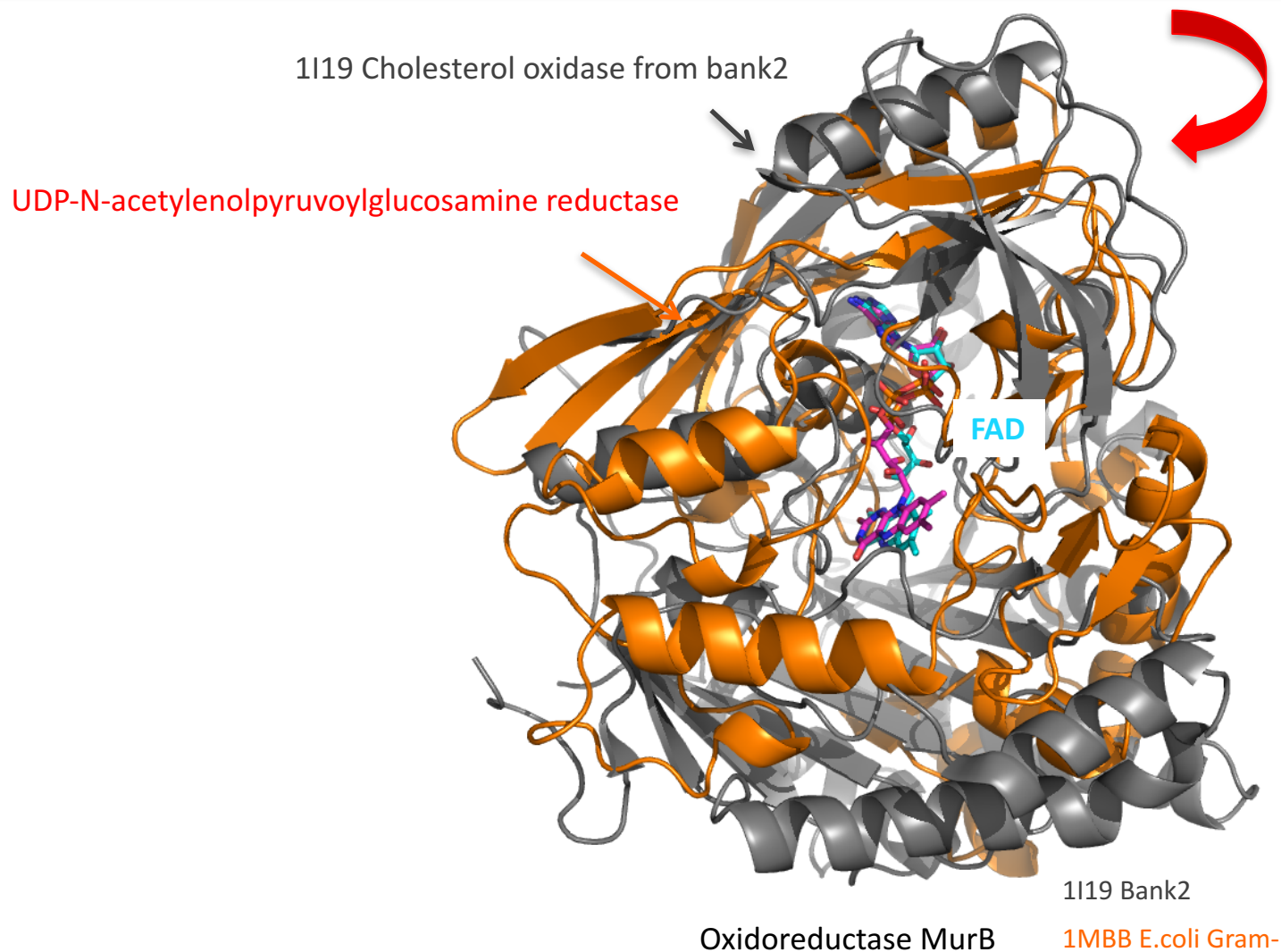
=

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

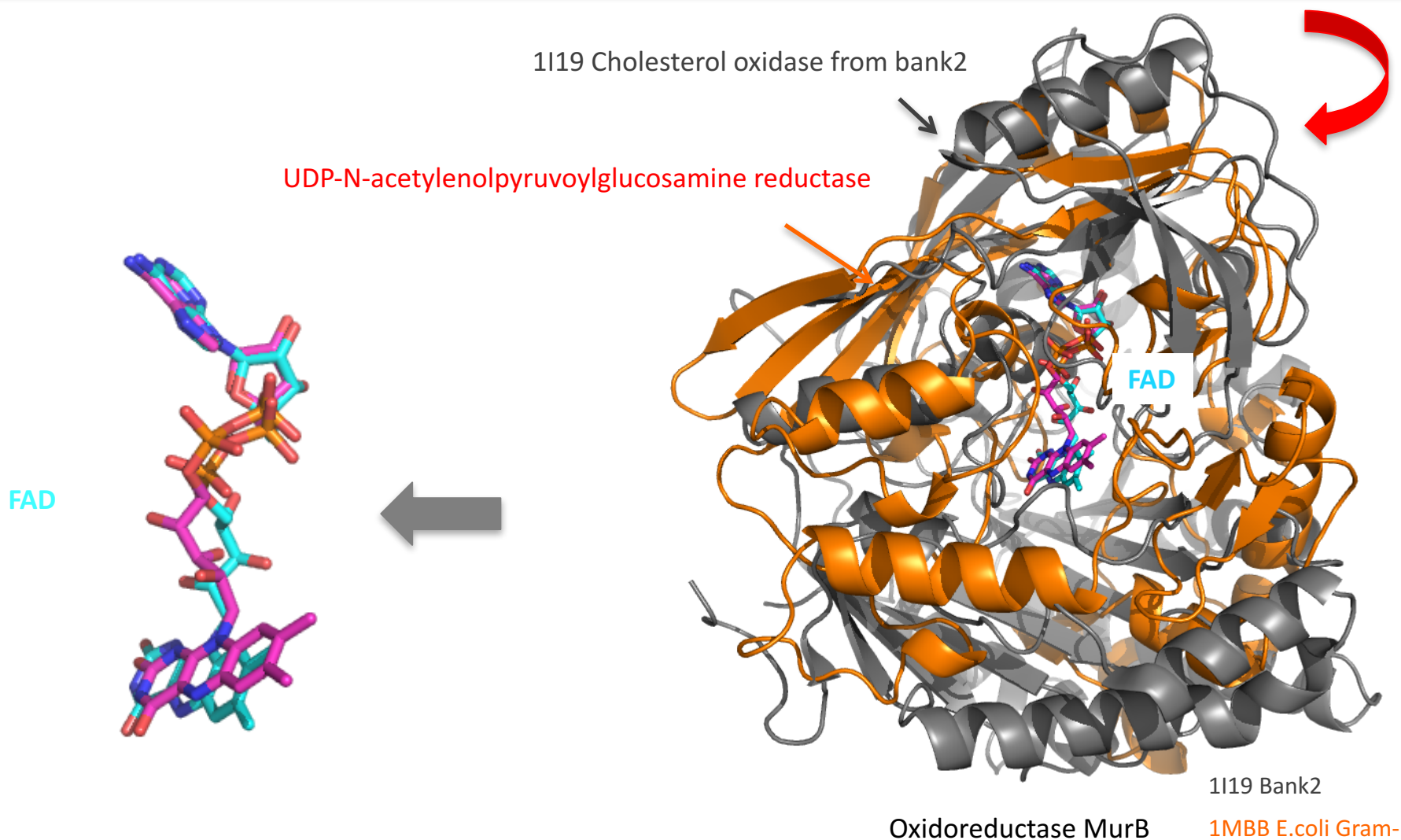
99.9 172 aa bdd8_003618.fasta.a3m.hhm.out.hhmakeBK2 → UDP-N-acetylenolpyruvoylglucosamine reductase



★ *B. dorei* D8

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

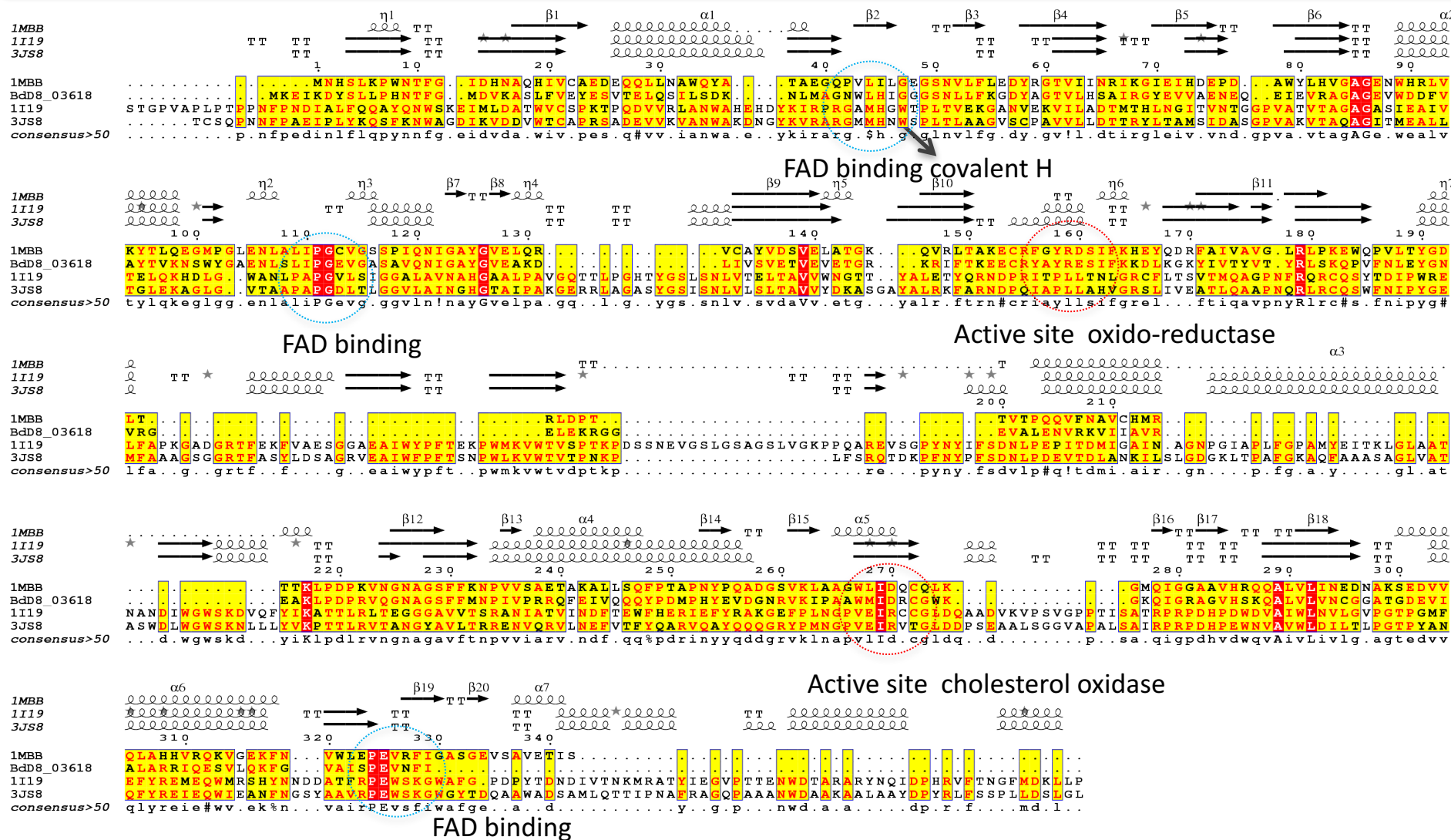
99.9 172 aa bdd8_003618.fasta.a3m.hhm.out.hhmakeBK2 → UDP-N-acetylenolpyruvoylglucosamine reductase



★ *B. dorei* D8

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

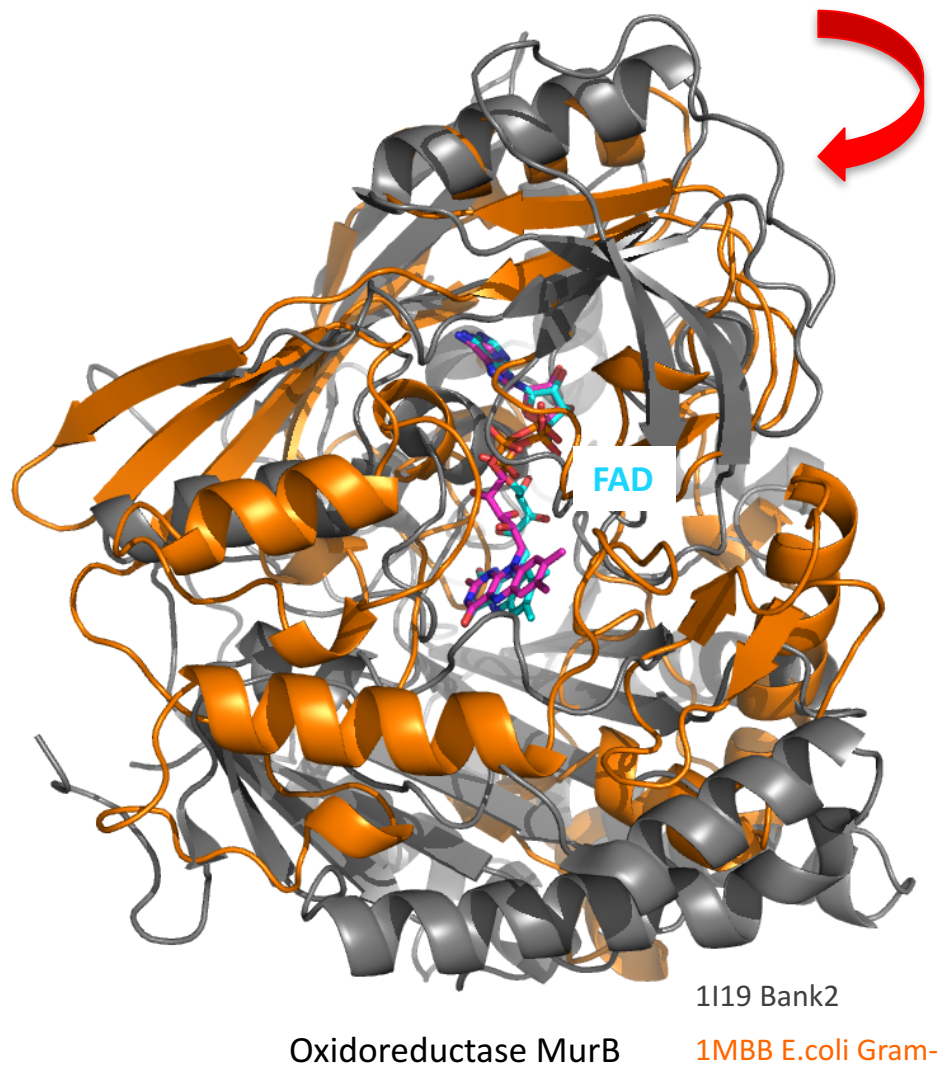
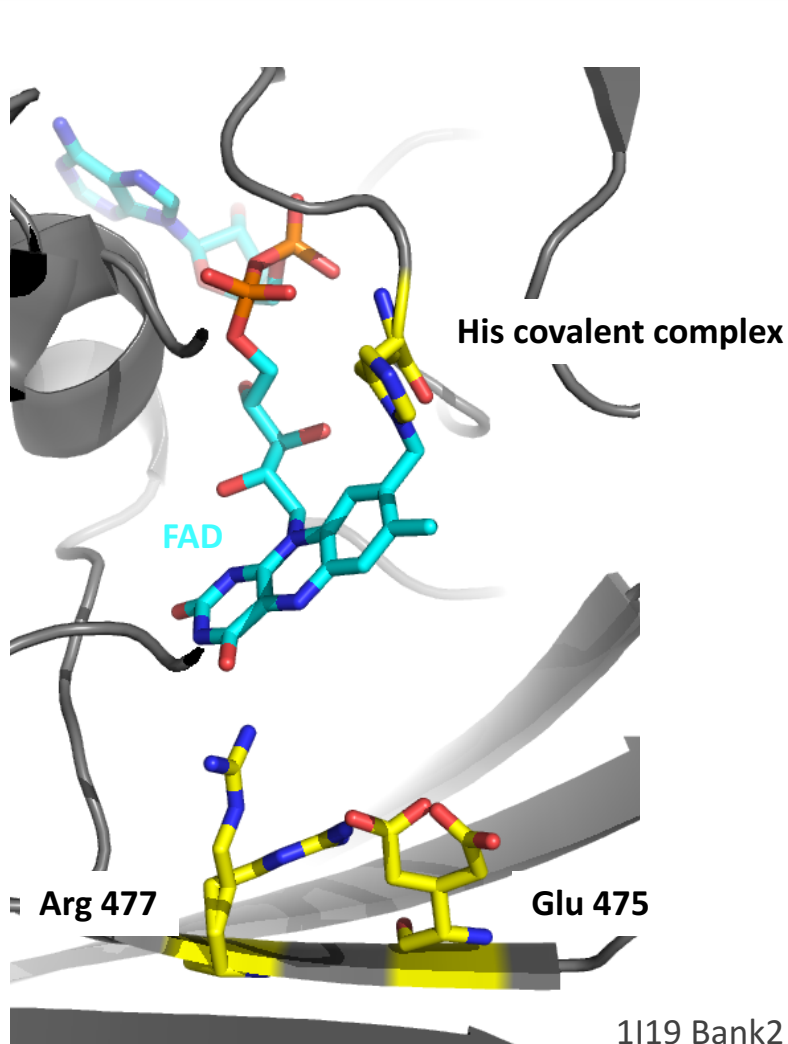
99.9 172 aa bdd8_003618.fasta.a3m.hhm.out.hhmakeBK2 → UDP-N-acetylenolpyruvoylglucosamine reductase



★ *B. dorei* D8

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

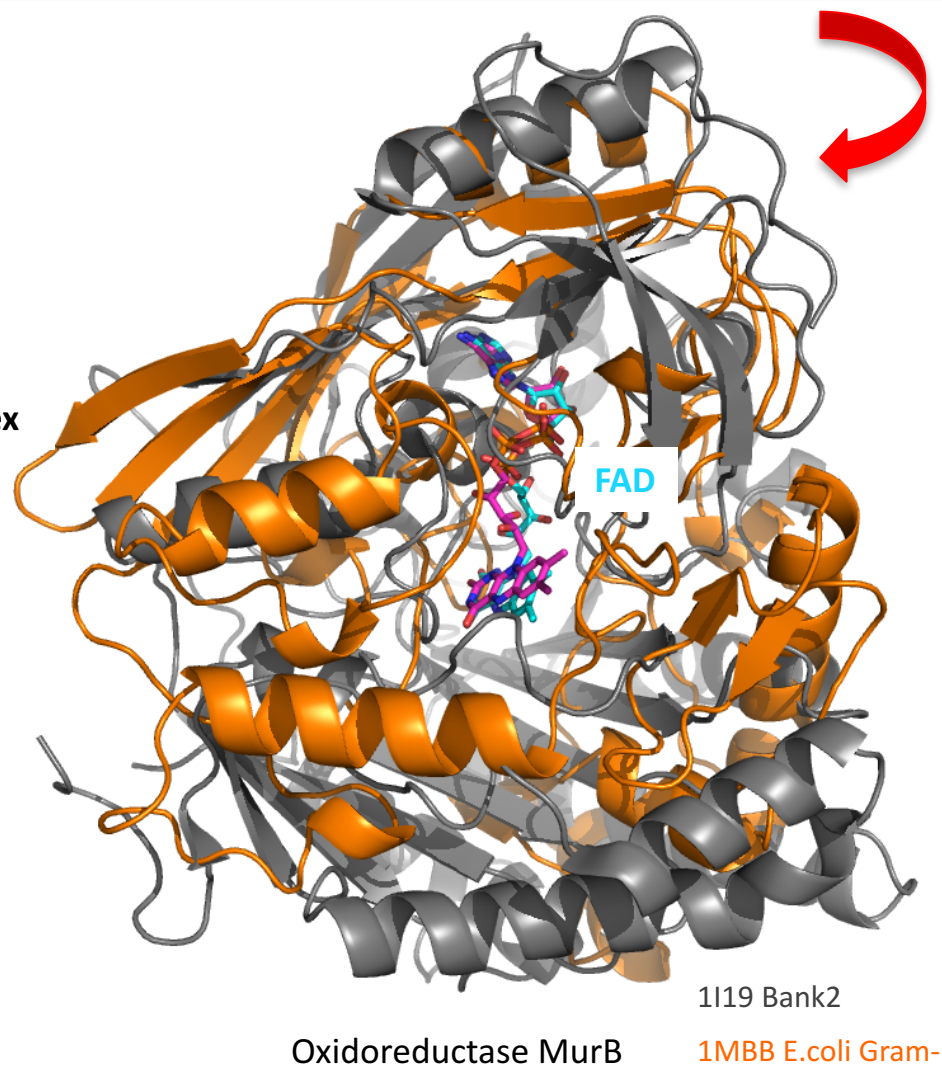
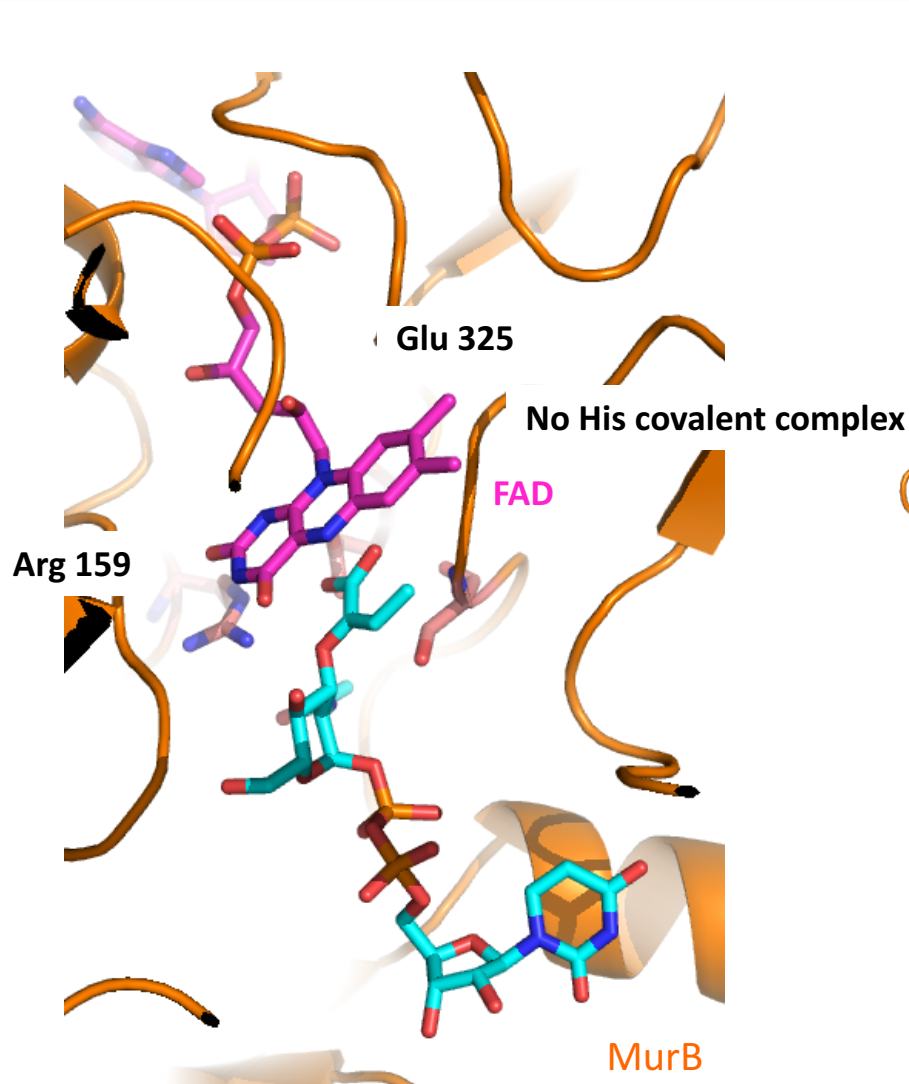
99.9 172 aa bdd8_003618.fasta.a3m.hhm.out.hhmakeBK2 → UDP-N-acetylenolpyruvoylglucosamine reductase



★ *B. dorei* D8

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

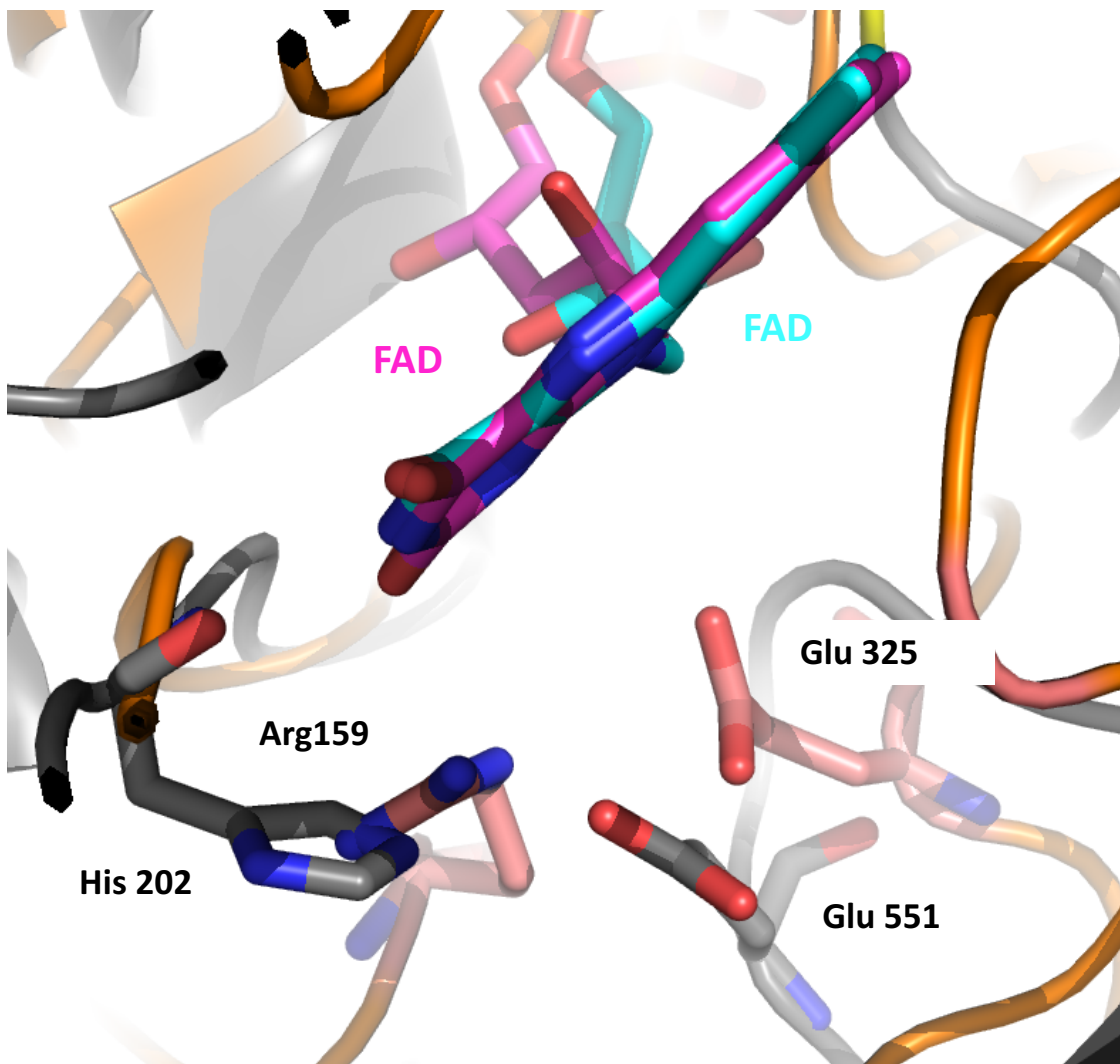
99.9 172 aa bdd8_003618.fasta.a3m.hhm.out.hhmakeBK2 → UDP-N-acetylenolpyruvoylglucosamine reductase



★ *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.hhmakeBK2 → UDP-N-acetylenolpyruvoylglucosamine reductase



1I19 Bank2

1MBB MurB

★ *B. dorei* D8 Sorted by Prob > 95 and size cut off > 50 amino acid residues

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^T Sorted by Prob > 95 and size > 50 amino acid residues

99.9 % 172 aa bacdor_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



Gene ordered at Genscript. 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 > 95 and size cut off > 50 amino acid residues

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^T Sorted by Prob > 95 and size > 50 amino acid residues

99.9 % 172 aa bacdor_03524.fasta.a3m.hhm.out.hhmakeBK2 → UDP-N-acetyl muramate dehydrogenase

=

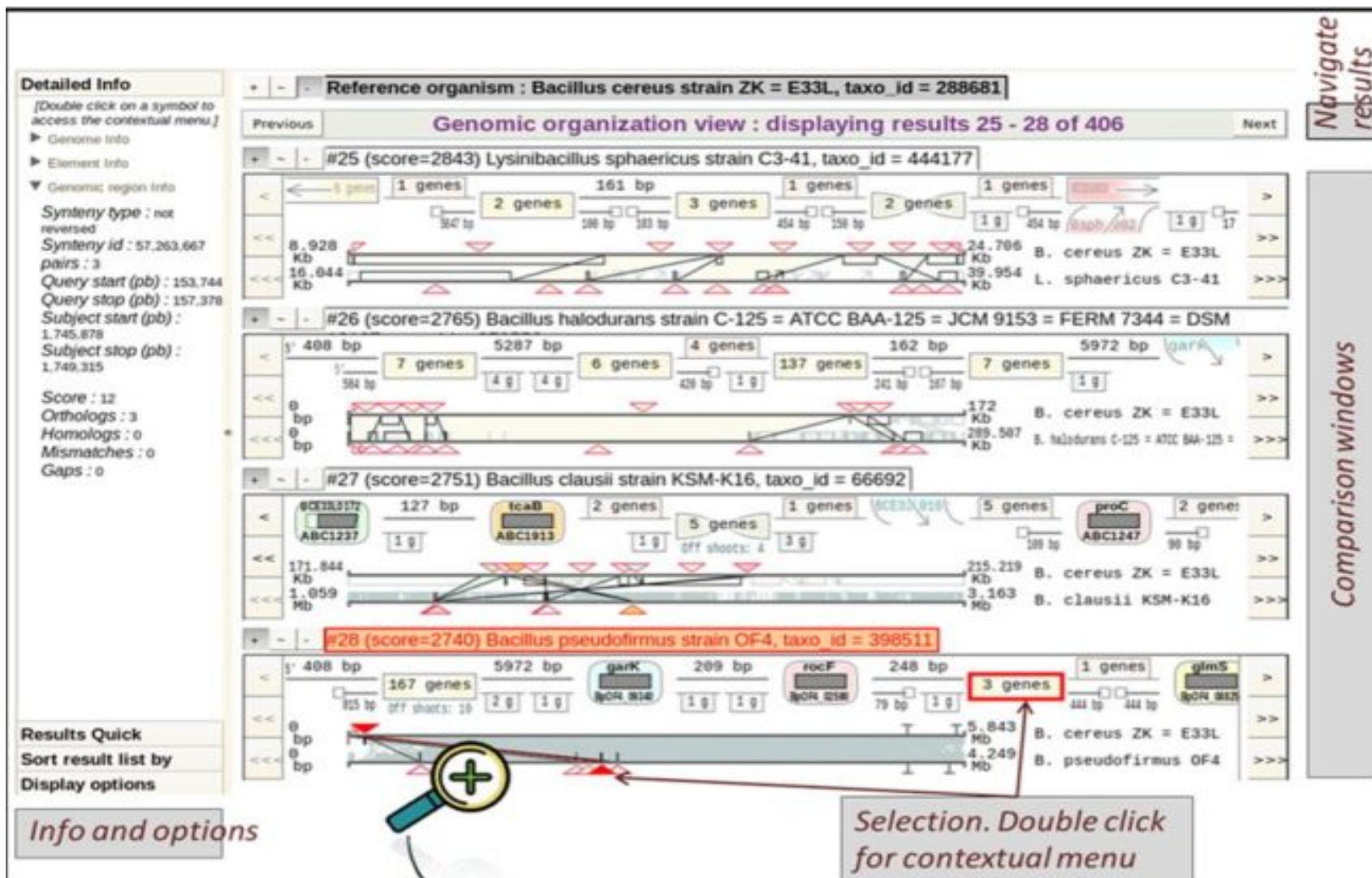
99.9 % 172 aa bdd8_003618.fasta.a3m.hhm.out.hhmakeBK2 → UDP-N-acetylenolpyruvoylglucosamine reductase



Analysis of synthenies by Thomas Lacroix using Insyght.

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

➡ Powerful tool for genes comparison and syntenic inferences



The screenshot displays the Insyght web interface for genomic analysis. The main window shows a 'Genomic organization view' for a reference organism, *Bacillus cereus* strain ZK = E33L (taxo_id = 288681). The view displays results 25 - 28 of 406. The interface includes several panels:

- Detailed Info**: A sidebar on the left providing genomic region information, including syntenic type, syntenic id, query start/stop coordinates, subject start/stop coordinates, score, orthologs, homologs, mismatches, and gaps.
- Genomic organization view**: The main display area showing genomic maps for four different strains:
 - #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)
- Comparison windows**: A vertical panel on the right side of the interface.
- Navigation and Selection**: A magnifying glass icon with a plus sign is shown over the genomic maps, indicating a selection action. A text box below it states: 'Selection. Double click for contextual menu'.

Annotations on the screenshot include:

- 'Navigate results' (vertical text on the right)
- 'Comparison windows' (vertical text on the right)
- 'Info and options' (text box at the bottom left)
- 'Selection. Double click for contextual menu' (text box at the bottom center)

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

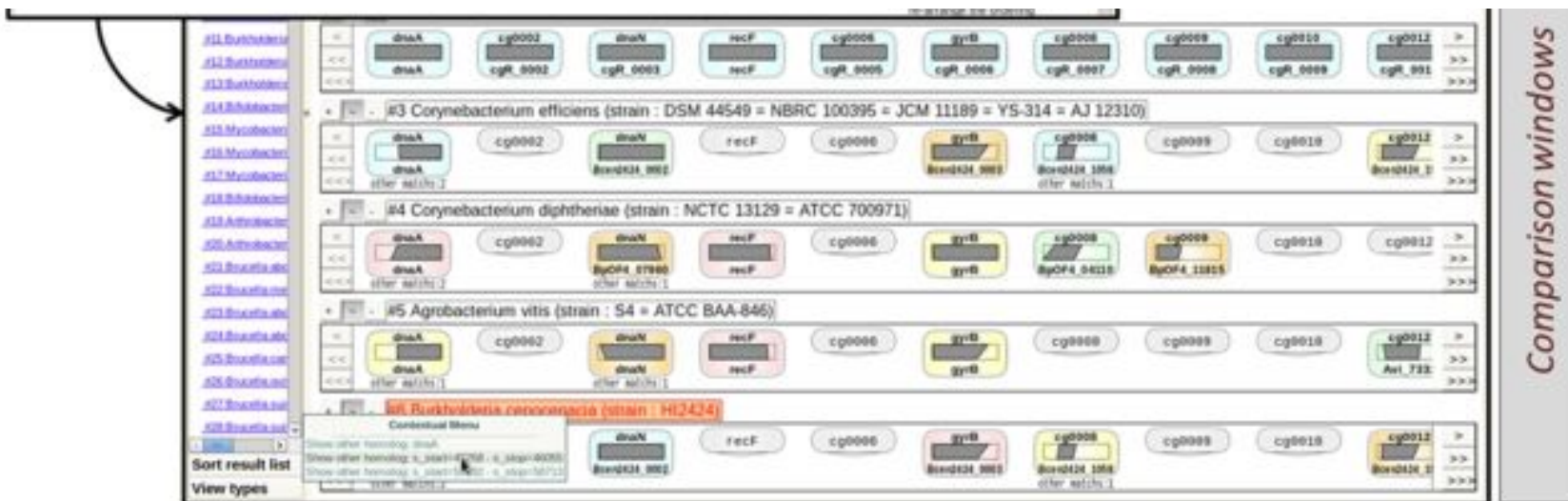
➡ Powerful tool for genes comparison and syntenic inferences

Synthenies : 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. Synthenies visible



The screenshot shows the Insyght web interface for gene comparison. It displays a list of genes and their syntenic blocks across different species. The genes shown are dnaA, dnaN, recF, dnaK, dnaL, dnaP, dnaQ, dnaR, dnaS, and dnaT. The syntenic blocks are represented by colored bars (blue, green, yellow, orange, red) indicating different levels of conservation. The interface includes a 'Contextual menu' at the bottom left with options like 'Show other homologs', 'Show other homologs', and 'Show other homologs'. A 'Comparison windows' sidebar is visible on the right side of the interface.

Info and options Contextual menu: browse multiple homologies per gene

- ★ Analysis of candidates- Comparison within cholesterol degrading genomes and others
- ★ Annotation comparator's analysis
- ★ Genomic organisation's analysis



No significant hits with clear sytheny



Need of transcriptomic analysis

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

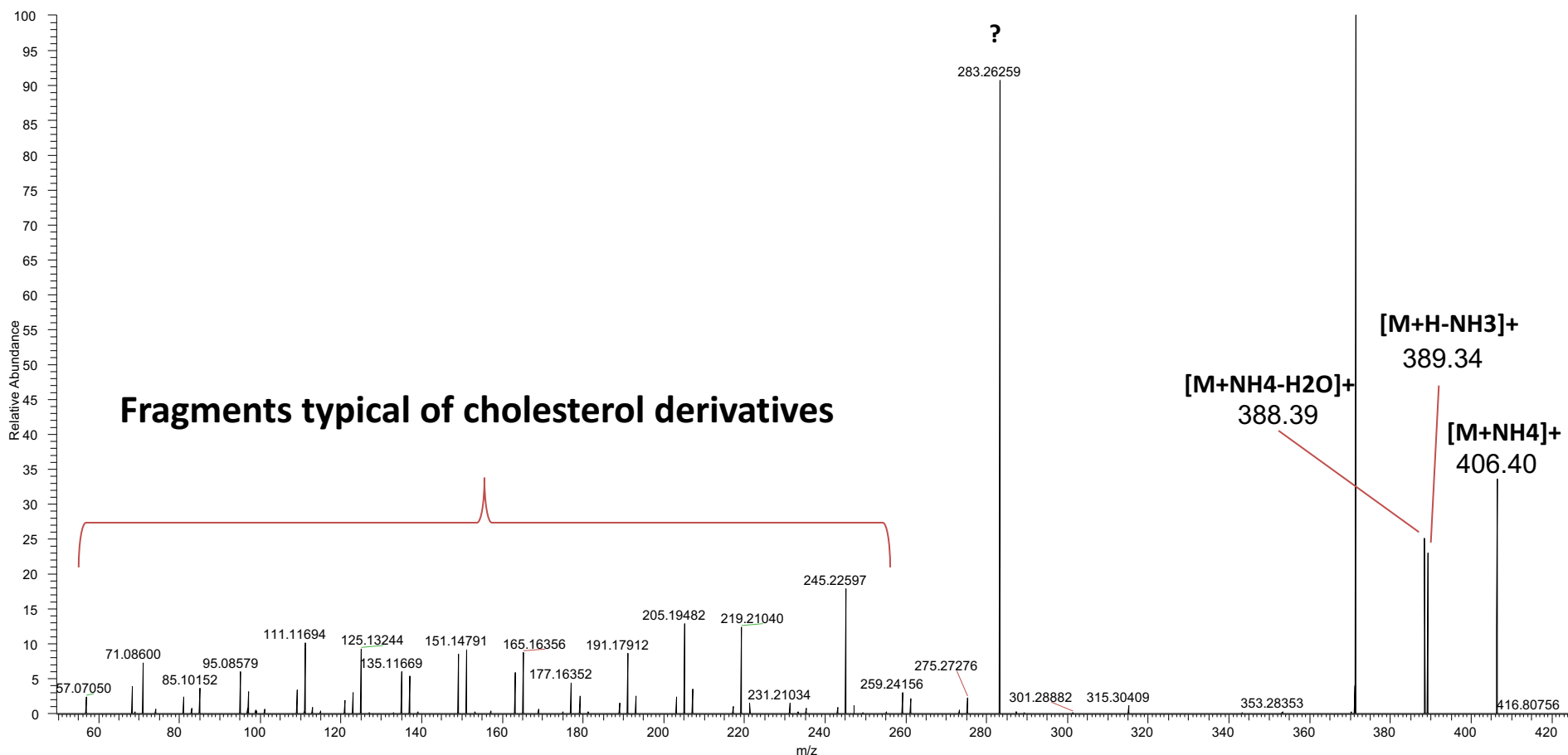
Catherine. Juste
Fabienne Beguet-Crespel

Samples resuspended in 500 μ L of Chloroform/ Methanol (50/50)
Diluted to 1/100eme before analysis

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

Coprostanol

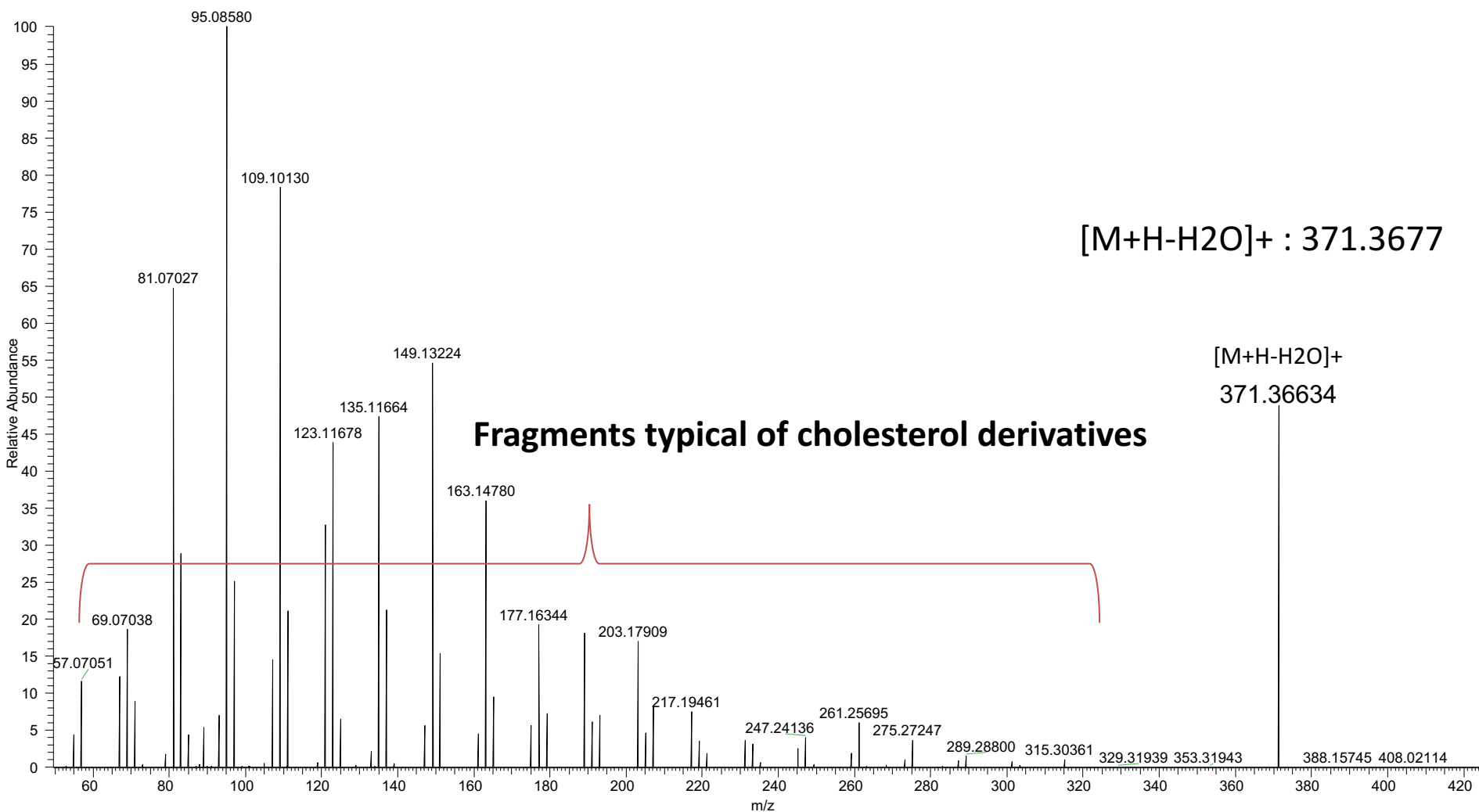
$[M+H-H_2O]^+$
371.36



Catherine. Juste

Fabienne Beguet-Crespel

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

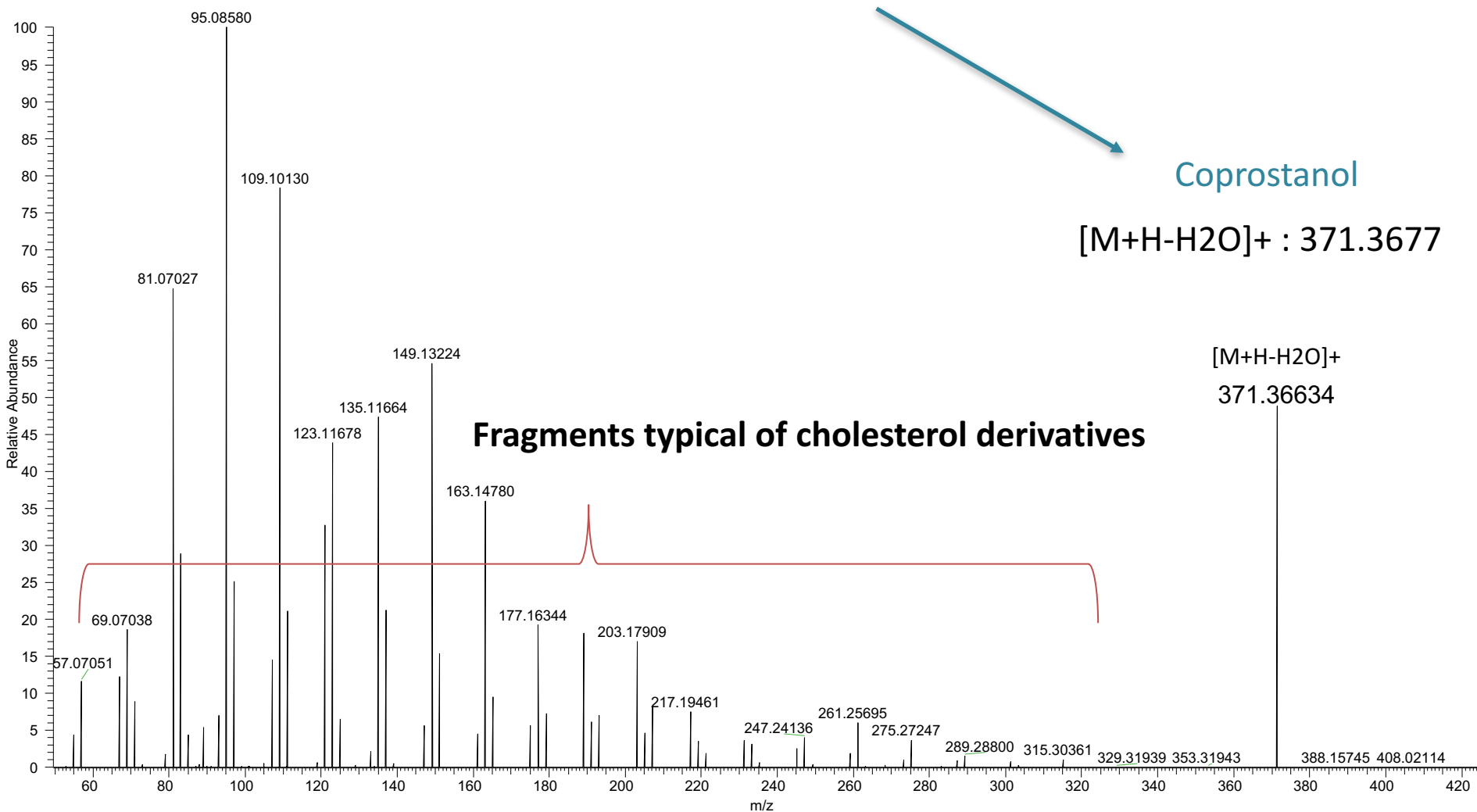


Catherine. Juste

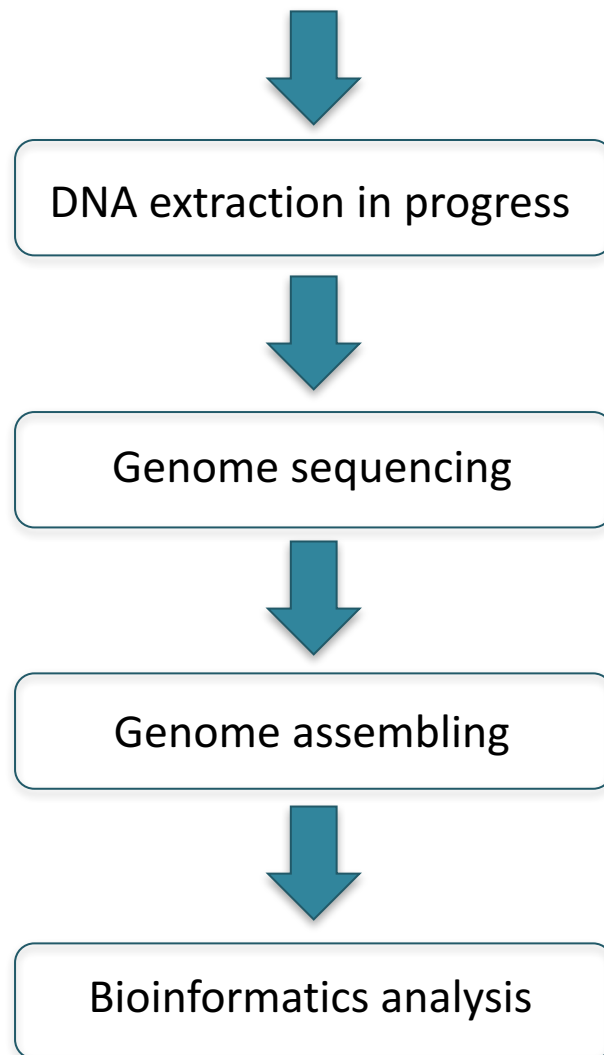
Fabienne Beguet-Crespel

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

Coprostanol is present & preferentially detected

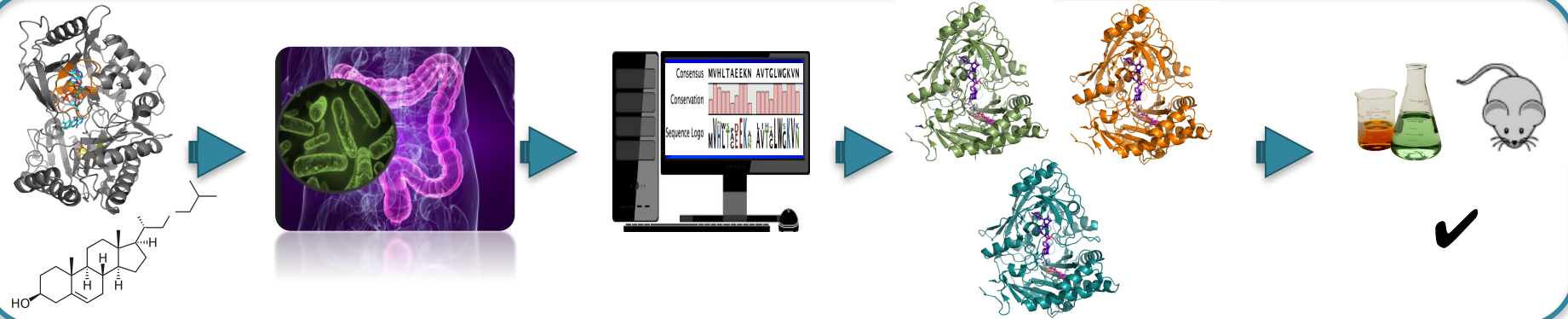


★ E. Coprostanoligenes samples : Genome sequencing, assembling and analysis



- ➡ No significant difference between the two genomes. Similar enzymes are found.
- ➡ Transcriptomic analysis of both chromosome and plasmids of *B. dorei* 175^T & *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?

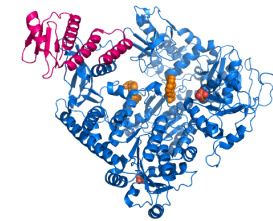
Towards the Structural Screening of Microbial Ecosystems



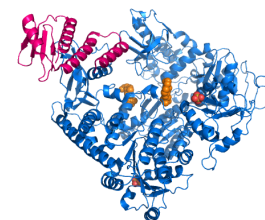
Outline

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

- More than 10 millions prokaryotic genes
 - 40% protein sequences of unknown function
 - Function can be inferred from structure, only 1393 unique 3D fold
 - Assessing the fold is possible
 - Fishing out the right fold is crucial !



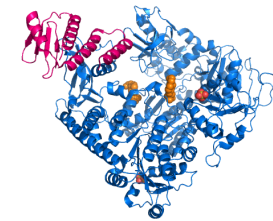
- More than 10 millions prokaryotic genes
- 40% protein sequences of unknown function
- Function can be inferred from structure, only 1393 unique 3D fold
- Assessing the fold is possible
- Fishing out the right fold is crucial !



3D Fold {

- ★ Reveals evolutionary relationship
- ★ Helps to understand the mechanisms of function
- ★ Deciphers its biological role within the organism

- More than 10 millions prokaryotic genes
- 40% protein sequences of unknown function
 - Function can be inferred from structure, only 1393 unique 3D fold
 - Assessing the fold is possible
 - Fishing out the right fold is crucial !



Objectives of structural metagenomics

- 1- **Capitalize** on MetaOmic datasets that are **underexploited**
- 2- **Develop** a versatile bioinformatics tool to **scan 3D fold in *MetaOmics***
- 3- **Identify** hits from 3D Targets
- 4- **Validate hits and pathways** involved as pro- or antibiotic markers

Structural screening of Microbial Ecosystems

1- Develop & integrate cutting-edge tools in bioinformatics

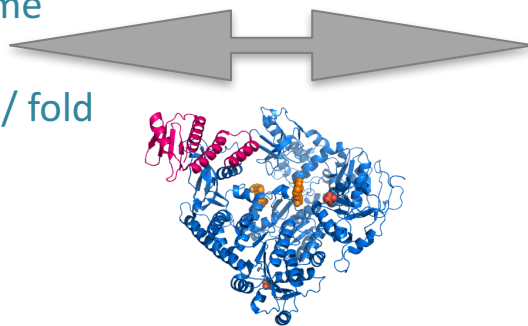
- Identify Hits from 3D targets within MetaHit
- Increase the functional annotation
- Biologist friendly Web user



Gut microbiota

2- Achieve 3D-MetaOmics.

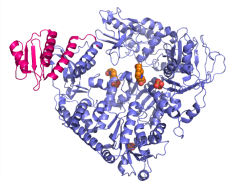
- Intensive modeling of entire proteome
- Distribution, diversity of folds
- Identify proteins with new functions/ fold



3- Validate Hits impacting Health

- Pro and anti-biotic biomarkers
- Universal antibiotic
- Identify homeostasis marker

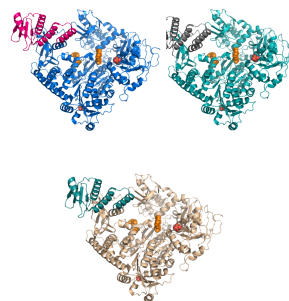
Target protein



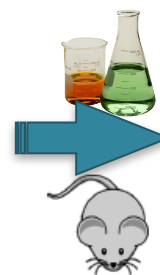
Relevance



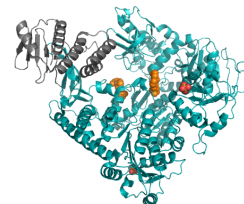
Scan microbiota



Hits identification



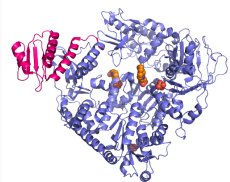
Validation



Biomarker

- Biological property
- Therapeutical interest
- Structural challenge

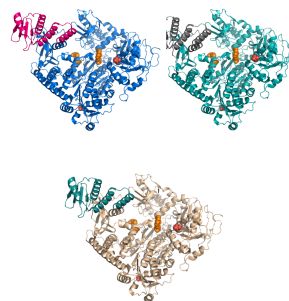
Target protein



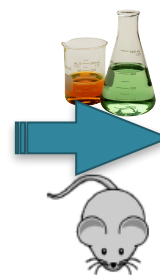
Relevance



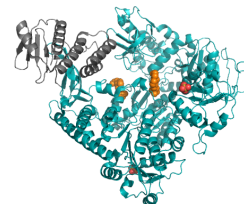
Scan microbiota



Hits identification



Validation



Biomarker

- Biological property
- Therapeutical interest
- Structural challenge

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**

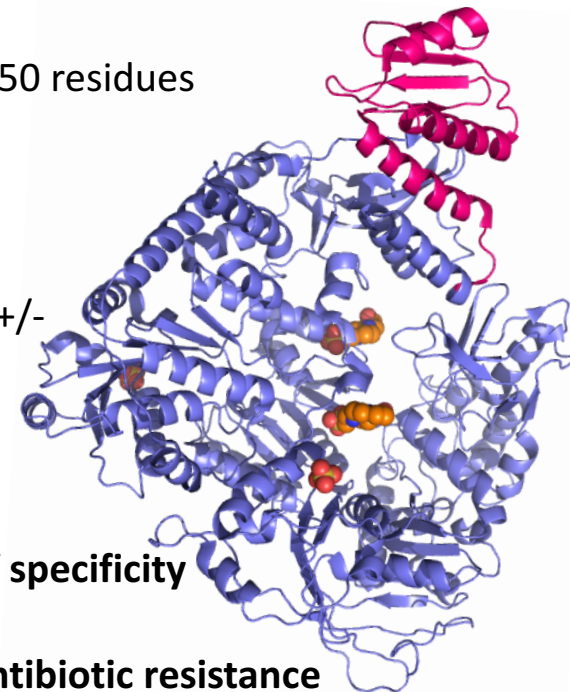


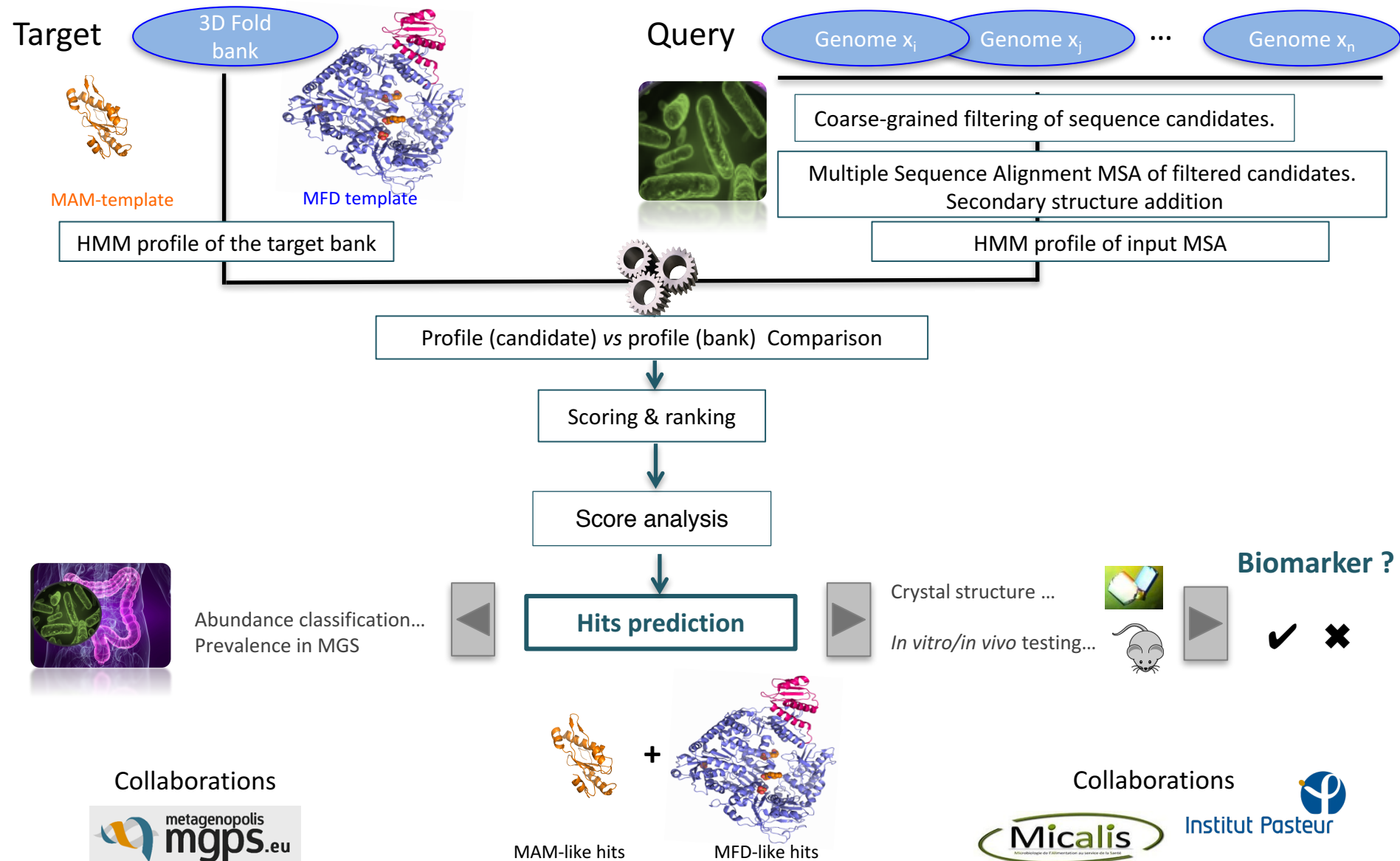
MFD

N. Rama Rao

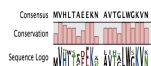
Mutation Frequency Decline

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





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

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^{1*}, Ruiqiang Li^{1*}, Jeroen Raes^{2,3}, Manimozhiyan Arumugam², Kristoffer Solvsten Burgdorf⁴, Chaysavanh Manichanh⁵, Trine Nielsen⁴, Nicolas Pons⁶, Florence Levenez⁶, Takuji Yamada², Daniel R. Mende², Junhua Li^{1,7}, Junming Xu¹, Shaochuan Li¹, Dongfang Li^{1,8}, Jianjun Cao¹, Bo Wang¹, Huiqing Liang¹, Huisong Zheng¹, Yinlong Xie^{1,7}, Julien Tap⁶, Patricia Lepage⁶, Marcelo Bertalan⁹, Jean-Michel Batto⁶, Torben Hansen⁴, Denis Le Paslier¹⁰, Allan Linneberg¹¹, H. Bjørn Nielsen⁹, Eric Pelletier¹⁰, Pierre Renault⁶, Thomas Sicheritz-Ponten⁹, Keith Turner¹², Hongmei Zhu¹, Chang Yu¹, Shengting Li¹, Min Jian¹, Yan Zhou¹, Yingrui Li¹, Xiuqing Zhang¹, Songgang Li¹, Nan Qin¹, Huanming Yang¹, Jian Wang¹, Søren Brunak⁹, Joel Doré⁶, Francisco Guarner⁵, Karsten Kristiansen¹³, Oluf Pedersen^{4,14}, Julian Parkhill¹², Jean Weissenbach¹⁰, MetaHIT Consortium†, Peer Bork², S. Dusko Ehrlich⁶ & Jun Wang^{1,13}

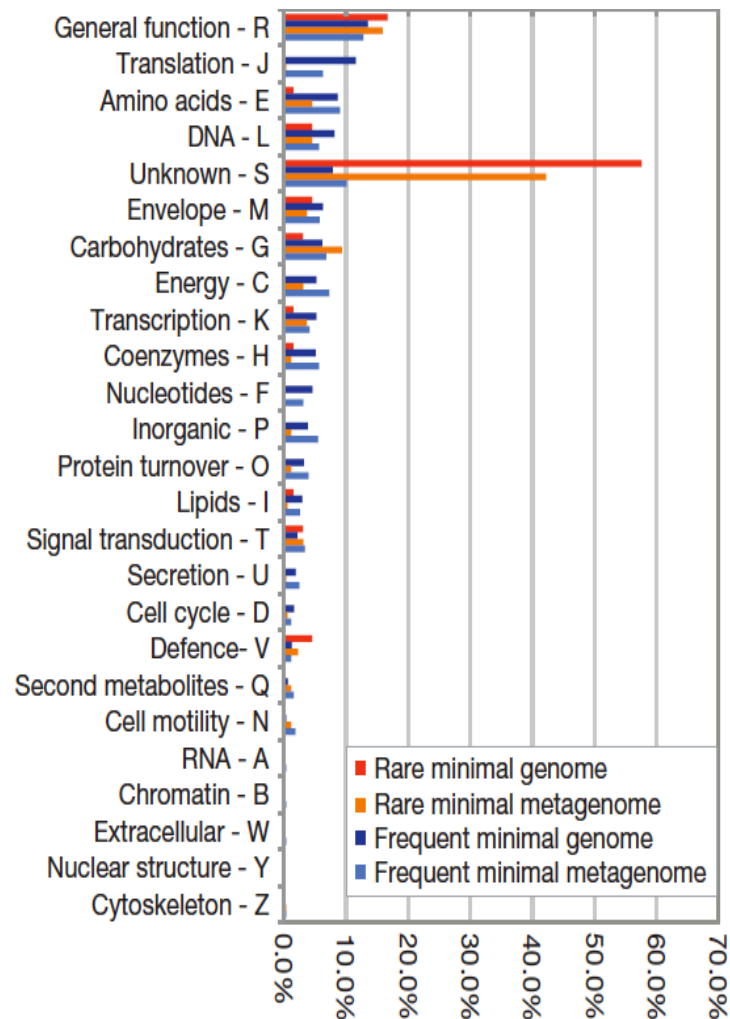
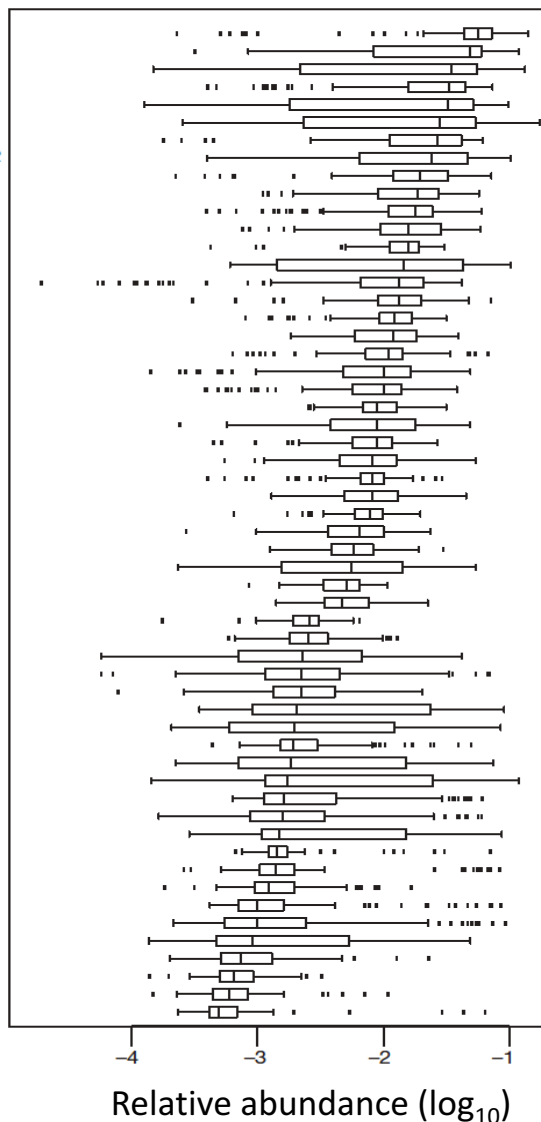
To understand the impact of gut microbes on human health and well-being it is crucial to assess their genetic potential. Here we describe the Illumina-based metagenomic sequencing, assembly and characterization of 3.3 million non-redundant microbial genes, derived from 576.7 gigabases of sequence, from 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

Minimal gut metagenome: functions involved in the homeostasis of the whole ecosystem

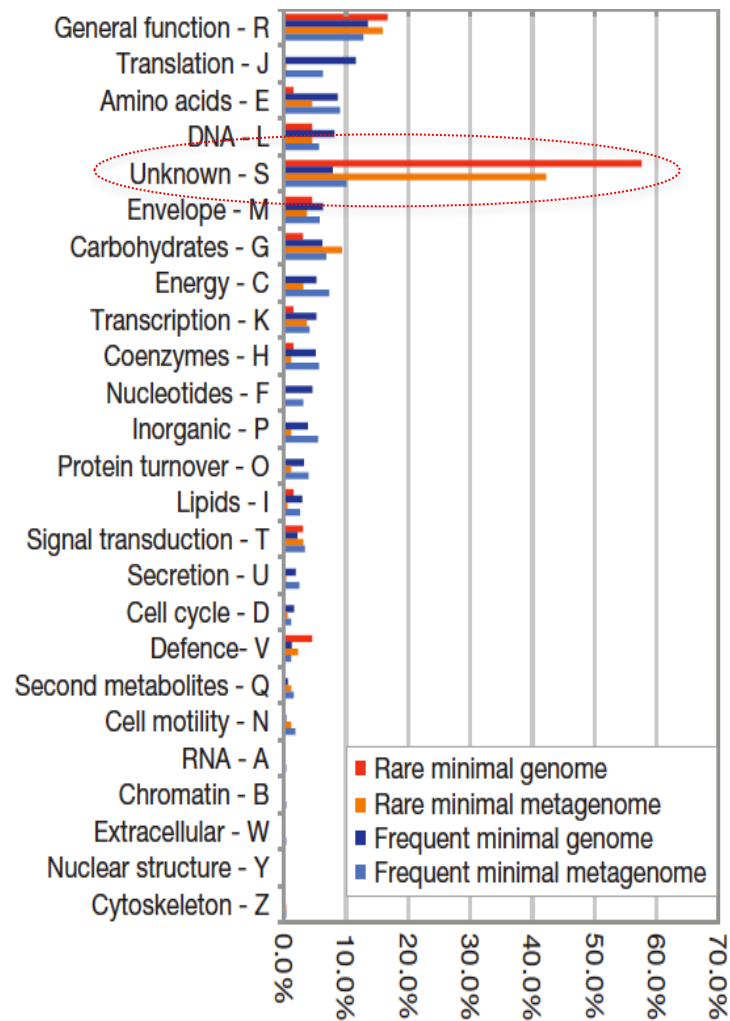
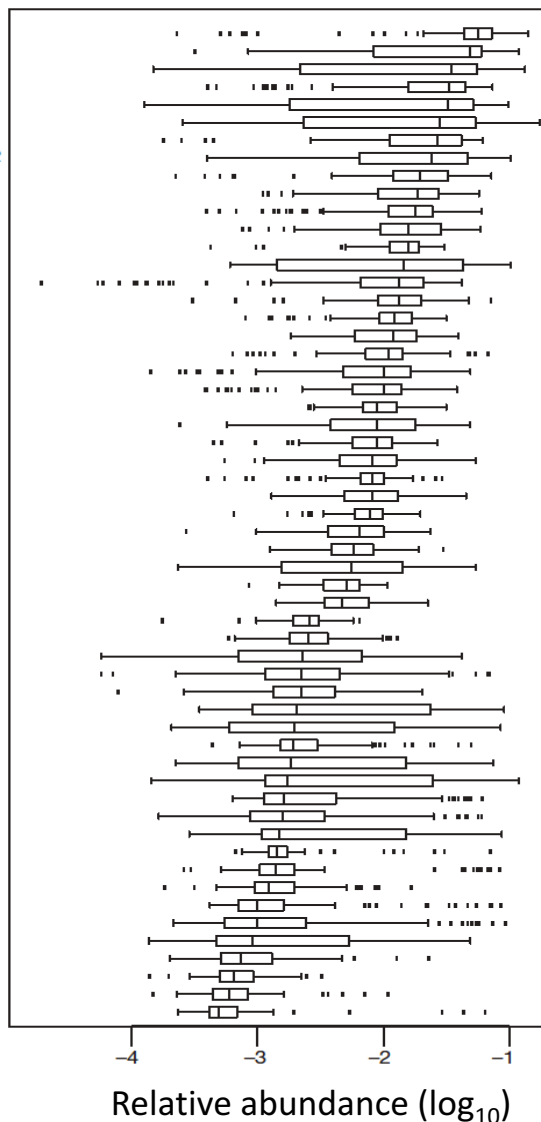
Bacteroides uniformis
Alistipes putredinis
Parabacteroides merdae
Dorea longicatena
Ruminococcus bromii L2-63
Bacteroides caccae
Clostridium sp. SS2-1
Bacteroides thetaiotaomicron VPI-5482
Eubacterium hallii
Ruminococcus torques L2-14
Unknown sp. SS3 4
Ruminococcus sp. SR1 5
Faecalibacterium prausnitzii 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
Bacterioides 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
Holdemania filiformis
Bacteroides stercoris
Coprococcus eutactus
Clostridium sp. M62 1
Bacteroides eggerthii
Butyrivibrio crossotus
Bacteroides finegoldii
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



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
Bacterioides 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
Holdemania filiformis
Bacteroides stercoris
Coprococcus eutactus
Clostridium sp. M62 1
Bacteroides eggerthii
Butyrivibrio crossotus
Bacteroides finegoldii
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



	Faecalibacterium_prausnitzii_A2-165 NCBI	WP_005932151.1	265	9,00E-93	100.00	100	100	135	0	0	1	135	1	135
			bitscore	eval	pid	qcovs	qcovhsp	length	mismatch	gapopen				
1	MAM_wt	tr R6SBT0 R6SBT0_9BACE	23.5	6.4	47.826	17	17	23	10	1	14	36	624	644
2	MAM_wt	tr D4JK97 D4JK97_9FIRM	28.5	0.013	45.000	15	15	20	11	0	1	20	9	28
3	MAM_wt	tr R7EPB3 R7EPB3_9BACE	26.6	0.55	41.667	24	24	36	18	1	52	84	185	220
4	MAM_wt	tr C3QB13 C3QB13_9BACE	29.3	0.063	40.625	24	24	32	19	0	88	119	95	126
5	MAM_wt	tr D4C9H4 D4C9H4_9CLOT	27.3	0.30	40.541	27	27	37	21	1	73	108	128	164
6	Faecalibacterium_prausnitzii_SL3_3	tr D4KBR2 D4KBR2_9FIRM	75.9	1.02e-19	35.570	98	98	149	72	4	1	132	1	142
7	MAM_wt	tr R7B146 R7B146_9BACE	25.0	1.2	34.694	36	36	49	31	1	15	63	119	166
8	MAM_wt	tr D4L4M8 D4L4M8_9FIRM	24.3	1.9	34.483	21	21	29	19	0	75	103	516	544
9	MAM_wt	tr C3QPR5 C3QPR5_9BACE	25.8	1.1	34.375	24	24	32	21	0	88	119	95	126
10	MAM_wt	tr R5TKN8 R5TKN8_9FIRM	26.2	0.50	34.286	26	26	35	22	1	48	82	205	238
11	MAM_wt	tr C0X1L4 C0X1L4_ENTFL	26.6	0.39	34.211	27	27	38	24	1	14	50	150	187
12	MAM_wt	tr D1PS54 D1PS54_9FIRM	24.6	2.2	34.043	34	34	47	27	3	44	89	105	148
13	MAM_wt	tr R6P689 R6P689_9CLOT	24.3	1.8	33.333	27	27	36	24	0	56	91	98	133
14	MAM_wt	tr A6LD60 A6LD60_PARD8	24.3	4.4	33.333	36	36	57	26	2	31	78	382	435
15	MAM_wt	tr B0G4V7 B0G4V7_9FIRM	25.4	1.1	32.500	29	29	40	26	1	68	106	150	189
16	MAM_wt	tr C0D4Q5 C0D4Q5_9FIRM	25.0	3.3	32.394	53	53	71	42	3	16	86	189	253
17	MAM_wt	tr R7JPH0 R7JPH0_9BACT	24.6	1.3	32.258	23	23	31	21	0	6	36	123	153
18	MAM_wt	tr E4VUA2 E4VUA2_BACF	27.7	0.35	31.818	33	33	44	30	0	1	44	280	323
19	MAM_wt	tr A7VJ13 A7VJ13_9CLOT	25.0	1.3	31.507	53	53	73	44	3	2	72	121	189
20	MAM_wt	tr B0NGV9 B0NGV9_CLOS	26.2	0.68	31.373	38	38	51	33	1	4	54	61	109
21	MAM_wt	tr Q5M488 Q5M488_STRT	26.9	0.13	31.111	33	33	45	24	2	36	80	83	120
22	MAM_wt	tr R5U2R3 R5U2R3_9BACE	28.5	0.12	30.909	41	41	55	36	1	76	130	47	99
23	MAM_wt	tr A4EBV5 A4EBV5_9ACTN	27.3	0.21	30.612	34	34	49	30	2	57	102	292	339
24	Clostridium_sp._SS2-1	tr B0P145 B0P145_9CLOT	32.7	0.004	30.337	66	66	89	49	5	9	97	1241	1316
25	MAM_wt	tr D4LSX8 D4LSX8_9FIRM	24.3	2.2	30.303	24	24	33	23	0	21	53	78	110
26	MAM_wt	tr R7FZ59 R7FZ59_9FIRM	25.0	0.99	30.159	41	41	63	37	1	23	78	160	222
27	MAM_wt	tr B0PHX3 B0PHX3_9FIRM	23.5	4.7	29.787	34	34	47	31	2	3	48	173	218
28	MAM_wt	tr R6X7N3 R6X7N3_9PORP	24.3	3.3	29.545	61	61	88	53	4	4	86	547	630
29	MAM_wt	tr A5Z5Q6 A5Z5Q6_9FIRM	27.7	0.19	29.412	47	47	68	42	2	66	128	345	411
30	MAM_wt	tr B9YAA0 B9YAA0_9FIRM	28.1	0.15	29.268	30	30	41	29	0	1	41	68	108
31	MAM_wt	tr B7B871 B7B871_9PORP	25.0	1.9	28.846	32	32	52	24	3	54	96	204	251
32	MAM_wt	tr P96214 P96214_MYCTU	25.4	1.4	28.571	38	38	56	35	1	52	102	61	116
33	MAM_wt	tr R6LDT4 R6LDT4_9FIRM	24.3	1.5	28.070	39	39	57	36	1	49	100	17	73
34	MAM_wt	tr C9L847 C9L847_BLAHA	24.3	1.9	27.869	41	41	61	38	2	5	59	79	139
35	MAM_wt	tr D4JTH7 D4JTH7_9FIRM	23.9	2.3	27.778	40	40	54	39	0	15	68	1	54
36	MAM_wt	tr D4M698 D4M698_9FIRM	26.6	0.46	27.692	47	47	65	45	1	72	134	187	251
37	MAM_wt	tr B6VXT6 B6VXT6_9BACE	26.6	0.72	27.536	44	44	69	41	1	76	135	77	145
38	MAM_wt	tr C0X1L4 C0X1L4_ENTFL	26.6	0.39	34.211	27	27	38	24	1	14	50	150	187



Analysis in progress



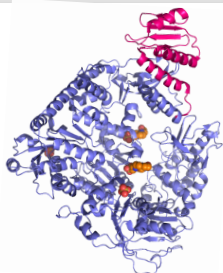
No clear hit except in *Faecalibacterium Prausnitzii* SL3_3



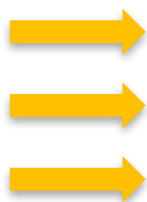
MAM possibly strictly restricted to *F. Prausnitzii*

			bitscore	evalue	pident	qcovs	qcovhsp	length	mismatch	gapopen
1										
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

all_best_MFD +



Analysis in progress

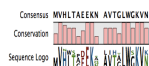


Mfd ubiquitous as expected

10 best hits to be selected for experimental validation

Homology modeling to assess the molecular basis of specificity

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



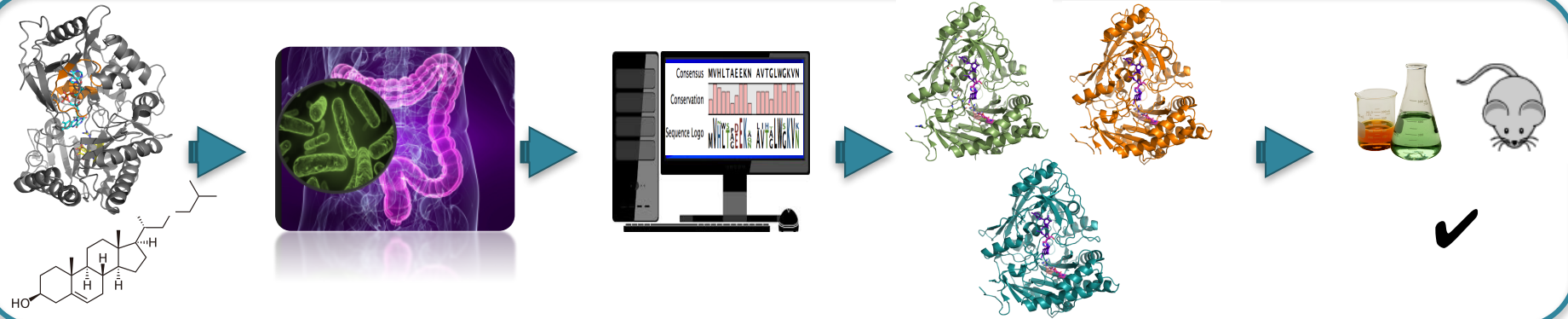
Applications - Impacts

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

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

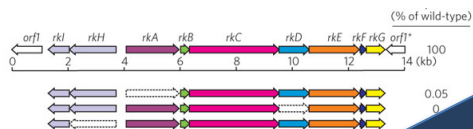
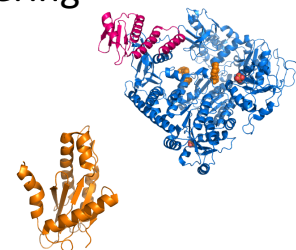
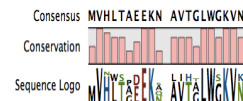
❖ **Innovative webserver** to achieve gut ecosystem exploration. Needs some scaling up & filtering

Needs some scaling up & filtering

Experimental validation of the structural hits

❖ **Strong positioning** at the interface structural bioinformatics & Biology

❖ **High valorisation potential**



HTS of structures

Structure –Function –Annotation

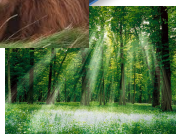
Functional clusters and Pathways

Originality /prevalence of folds

Disease/health -associated patterns of diversity

... Towards new targets with Health impact

... Towards other microbiotas



Acknowledgments



Sandra Derozier

Jean-François Gibrat

Thomas Lacroix

Valentin Loux

Véronique Martin

MEM Metaprogram



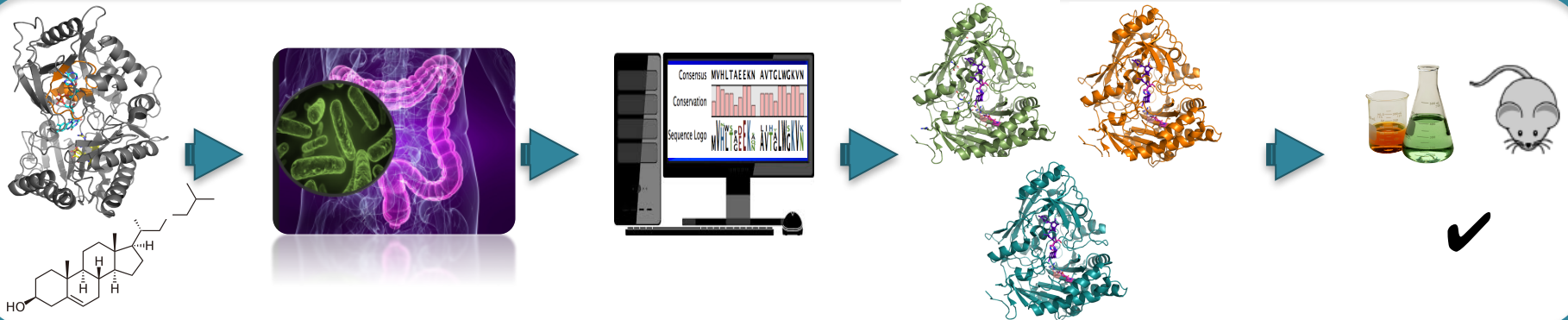
Fabienn Béguet-Crespel

Jean-Marc Chatel

Philippe Gérard

Catherine Juste

Nalini RamaRao



Institut Pasteur

Mariano Martinez

Pedro Alzari

Acknowledgments

To my dear friends...

Andrea Villarino



Mabel Berois

Maria Natalia Lía



And you for your kind attention ...



Analysis of candidates- Comparison within cholesterol degrading genomes and other

=> all 5 present in other *B. dorei* : very well conserved (id 100% length 100%) , synteny BdD8_00352 and BdD8_00254 (only one on same contig) but far apart

=> + *B. vulgatus*, *intestinalis*, *ovatus*, *plebeius* : well conserved (id 60-80% length 100%) ; no synteny

=> 4/5 (mostly BdD8_00254 missing) in *B. massiliensis*, *acidifaciens*, *fragilis*, *uniformis*, *eggerthii*, *caccae*, *pyogenes*, *thetaitaomicron*, *helcogenes* : well conserved (id 60-80% length 100%) ; no synteny

=> 1-3/5 in other, less well conserved (id 20-40% length 60-100%) ; no synteny

-- Annotation comparator's analysis

- nadB : L-aspartate oxidase also in 28 other orthologs (some well conserved)
- BdD8_00254 : FAD dependant oxidoreductase also in 6 other orthologs (some well conserved)
- sdhA : succinate dehydrogenase also in 19 other orthologs (some well conserved)
- BdD8_03122 : pyridine nucleotide-disulfide oxidoreductase also in 17 other orthologs (some well conserved)
- murB : UDP-N-acetylenolpyruvoylglucosamine reductase also in 30 other orthologs (some well conserved)

-- Genomic organisation's analysis

- nadB : 1 interesting genes in 5' ? : lpdA (Dihydrolipoyl dehydrogenase) ; synteny of ~192 in other *B. dorei*, ~6 in other *B.*, no synteny with further species :
- BdD8_00254 : 2 interesting genes in 3' ? : choloylglycine hydrolase, phenylacetate--CoA ligase ; synteny of ~192 in other *B. dorei*, ~6 in other *B.*, no synteny with further species
- sdhA : nothing interesting
- BdD8_03122 : nothing interesting
- murB : no interesting keyword in neighbor CDS ; synteny of 14-26 with other *B. dorei*, 3-11 with other *B.* no synteny with further species

Analyse of interesting genes from other species known to reduce cholesterol

Do they have orthologs in D8?

keyword related to cholesterol degradation :

5-cholesten-3-one; 4cholesten-3-one; cholestenone beta reductase; EC1.1.3.6; EC1.3.1.3; cholesterol oxidase and FAD dependent oxidoreductase (POSIX Regex : .*FAD.*oxidoreductase.*)

Set of

- 20 bacteroides anaerobie (included 6 B dorei) which do not degrade cholesterol
- 3 Mycobacterium aerobie lifestyle which degrade cholesterol
- 4 Rhodococcus aerobie lifestyle which degrade cholesterol



- Mycobacterium tuberculosis H37Rv, taxon_id = 83332 [NC_000962]

- 1 * cholesterol oxidase :

choD, Rv3409c: cholesterol oxidase

=> Close to oxidoreductase, in conserved synteny with all bacteria cholesterol catabolism except Schizosaccharomyces commune (eukaryotes)

=> no interest for D8

- 0 * .*FAD.*oxidoreductase.* - 0 * EC1.3.1.3

- Mycobacterium smegmatis str. MC2

- 0 * cholest - 3 * EC1.3.1.3

Analyse of interesting genes from other species known to reduce cholesterol

Do they have orthologs in D8?

keyword related to cholesterol degradation :

5-cholesten-3-one; 4cholesten-3-one; cholestenone beta reductase; EC1.1.3.6; EC1.3.1.3; cholesterol oxidase and FAD dependent oxidoreductase (POSIX Regex : .*FAD.*oxidoreductase.*)

Set of

- 20 bacteroides anaerobie (included 6 B dorei) which do not degrade cholesterol
- 3 Mycobacterium aerobie which degrade cholesterol
- 4 Rhodococcus aerobie which degrade cholesterol



Analysis using Insyght.

Conclusions: no significant hits with clear synthenie, genes linked to lipid metabolism, etc ...

Could cholesterol-degrading-enzymes present in aerobie bacteria be different from anaerobie bacteria? -> Need of transcriptomic analysis

environmental
microbiology

Environmental Microbiology (2011) 13(4), 943–959



doi:10.1111/j.1462-2920.2010.02398.x

Initial step in the catabolism of cholesterol by *Mycobacterium smegmatis* mc²155

I. Uhía, B. Galán,* V. Morales and J. L. García
Environmental Biology Department, Centro de
Investigaciones Biológicas, CSIC, Ramiro de Maeztu,
9, 28040 Madrid, Spain.

Summary

The first step in the catabolism of cholesterol, i.e. the transformation of cholesterol into cholestenone, has been investigated in *Mycobacterium smegmatis*. *In silico* analysis identified the MSMEG_1604 gene encoding a putative protein similar to the ChoD cholesterol oxidase of *M. tuberculosis* H37Rv (Rv3409c) and the MSMEG_5228 gene coding for a protein similar to the NAD(P)-dependent cholesterol dehydrogenase/isomerase of *Nocardia* sp. The expression of the MSMEG_5228 gene was inducible by cholesterol whereas the expression of MSMEG_1604 gene was constitutive. When both genes were expressed in *Escherichia coli* only the MSMEG_5228 protein was active on cholesterol. The function of ChoD-like MSMEG_1604 protein remains to be elucidated, but it does not appear to play a critical role in the mineralization of cholesterol as a MSMEG_1604 mutant was not affected in the production of cholestenone. However, a MSMEG_5228 mutant showed a drastic reduction in the synthesis of cholestenone. The finding that this mutant was still able to grow in cholesterol, allowed us to demonstrate that the cholesterol-inducible MSMEG_5233 gene encodes an additional cholesterol dehydrogenase/isomerase similar to the AcmA dehydrogenase of *Sterolibacterium denitrificans*. The observation that the double MSMEG_5228-5233 mutant was able to grow in cholesterol suggests that in addition to these enzymes other dehydrogenase/isomerases can also catalyse the first reaction of the cholesterol degradation pathway in *M. smegmatis*, which is not the limiting step of the process.

Introduction

Cholesterol and related steroid compounds are ubiquitous and very abundant in the environment as part of cytoplasmic

mic membranes and as precursors of vitamin D, bile acids and several sexual hormones. In addition, many synthetic steroids (e.g. sexual hormones as estrogens and androgens) frequently appear in municipal and industrial wastewaters as environmental pollutants (Routledge *et al.*, 1998; Poole and Cord-Ruwisch, 2004; Gagné *et al.*, 2006). Although the abundance and ubiquity of steroids have made them a common carbon source for saprophytic microorganisms belonging to various physiological groups their complete degradative pathways remain challenging (Tak, 1942; Brown and Peterson, 1957).

et

ba
ch
ch
the
two
da:
(FA
usi
del
as
Kis
200
diffi
thei
enz
200
oxic
ind
Apa
A
gro
culti
whic
for
teru
cida
acq
zatic
seer
the
Sass

Microbiology (2010), 156, 1362–1371

DOI 10.1099/mic.0.034536-0

Cholesterol utilization in mycobacteria is controlled by two TetR-type transcriptional regulators: *kstR* and *kstR2*

Sharon L. Kendal¹, Philippa Burgess¹, Ricardo Balhana¹, Mike Withers¹, Annemieke ten Bokum¹, J. Shaun Lott², Chen Gao², Iria Uhia-Castro³ and Neil G. Stoker¹

¹Department of Pathology and Infectious Diseases, The Royal Veterinary College, Centre for Emerging, Endemic and Exotic Disease, Hawkshead Lane, Hertfordshire, AL9 7TA, UK

²Laboratory of Structural Biology and Maurice Wilkins Centre for Molecular Biodiscovery, School of Biological Sciences, University of Auckland, New Zealand

³Department of Environmental Biology, Centro de Investigaciones Biológicas-Consejo Superior de Investigaciones Científicas, 28040 Madrid, Spain

Mycobacterium tuberculosis is able to use a variety of carbon sources *in vivo* and current knowledge suggests that cholesterol is used as a carbon source during infection. The catabolized cholesterol is used both as an energy source (ATP generation) and as a source of precursor molecules for the synthesis of complex methyl-branched fatty acids. In previous studies, we described a TetR-type transcriptional repressor, *kstR*, that controls the expression of a number of genes involved in cholesterol catabolism. In this study, we describe a second TetR-type repressor, which we call *kstR2*. We knocked this gene out in *Mycobacterium smegmatis* and used microarrays and quantitative RT-PCR to examine the effects on gene expression. We identified a palindromic regulatory motif for KstR2, showed that this motif is present in three promoter regions in mycobacteria and rhodococcus, and demonstrated binding of purified KstR2 to the motif. Using a combination of motif location analysis, gene expression analysis and the examination of gene conservation, we suggest that *kstR2* controls the expression of a 15 gene regulon. Like *kstR*, *kstR2* and the *kstR2* regulon are highly conserved among the actinomycetes and studies in rhodococcus suggest a role for these genes in cholesterol catabolism. The functional significance of the regulon and implications for the control of cholesterol utilization are discussed.

Received 13 October 2009
Revised 21 January 2010
Accepted 12 February 2010

KstR = Tet-R type transcriptional regulator
supAB / mae + oxylase to look for
e cluster encoding cholesterol catabolism
oil actinomycete provides insight into
bacterium tuberculosis survival in macrophages

Geize*, Katherine Yam¹, Thomas Heuser¹, Maarten H. Wilbrink*, Hirofumi Hara¹,
nderton¹, Edith Sim¹, Lubbert Dijkhuizen*, Julian E. Davies¹, William W. Mohn¹,
Eltis¹

¹Microbiology, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, 9751 NN, Groningen, The
Department of Microbiology and Immunology, Life Sciences Institute, University of British Columbia, Vancouver, BC, Canada V6T 1Z3; and
armacology, University of Oxford, Oxford OX1 3QT, United Kingdom

J. Jacobs, Jr., Albert Einstein College of Medicine, Bronx, NY, and accepted by the Editorial Board December 6, 2006 (received for review
October 10, 2006)

strain RHA1, a soil bacterium related to *Mycobacterium tuberculosis*, degrades an exceptionally broad range of organic
scriptomic analysis of cholesterol-grown RHA1 re-
: pathway predicted to proceed via 4-androstene-
3,4-dihydroxy-9,10-seconandrost-1,3,5(10)-triene.

(Fig. 1). In some *Mycobacterium* (4) and *Rhodococcus* (5, 6) species, the aliphatic side chain at C17 is removed via a process similar to β -oxidation involving progressively shorter carboxylic acids. In these strains, 3-ketosteroid Δ 1-dehydrogenase (KSTD) and 3-ke-

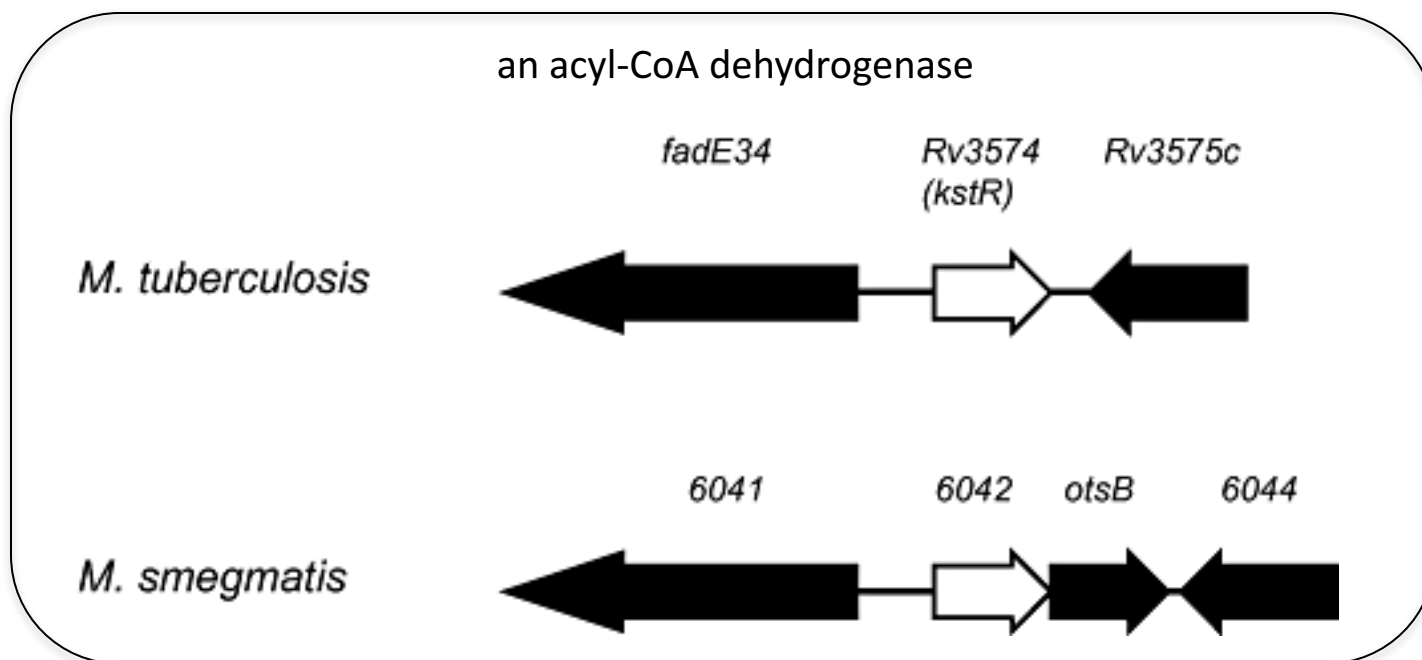
alyze the opening of ring B and around 3-HSA (3, 7–9). The subsequent 17-dioxo-1,2,3,4,10,19-hexanorandrost-17-one is cleaved via oxygenolytic cleavage of ring A in the testosterone-degrading strain R1 (10, 11). In *Rhodococcus equi*, the AA is removed via β -oxidation (12). In steroid catabolism have yet to be pathway enzymes are poorly characterized in degrading the bicyclic C and D. Detailed knowledge of to engineering strains for the bio-

realed that rhodococci may be useful processes: ~60% of the 3,999 genes H37Rv are conserved in RHA1, function (2). *M. tuberculosis* is the bacterial infection, killing 2 million each year, and extensive drug TB are now emerging (ref. 13 and sheets/15104/en/index.html). One mycobacterial physiology that con-

ENVIRONMENTAL
MICROBIOLOGY

Received 16 July, 2010; accepted 9 November, 2010. *For correspondence. E-mail: bgalan@icb.csic.es; Tel.: (+34) 918 373 112; Fax: (+34) 915 360 432.

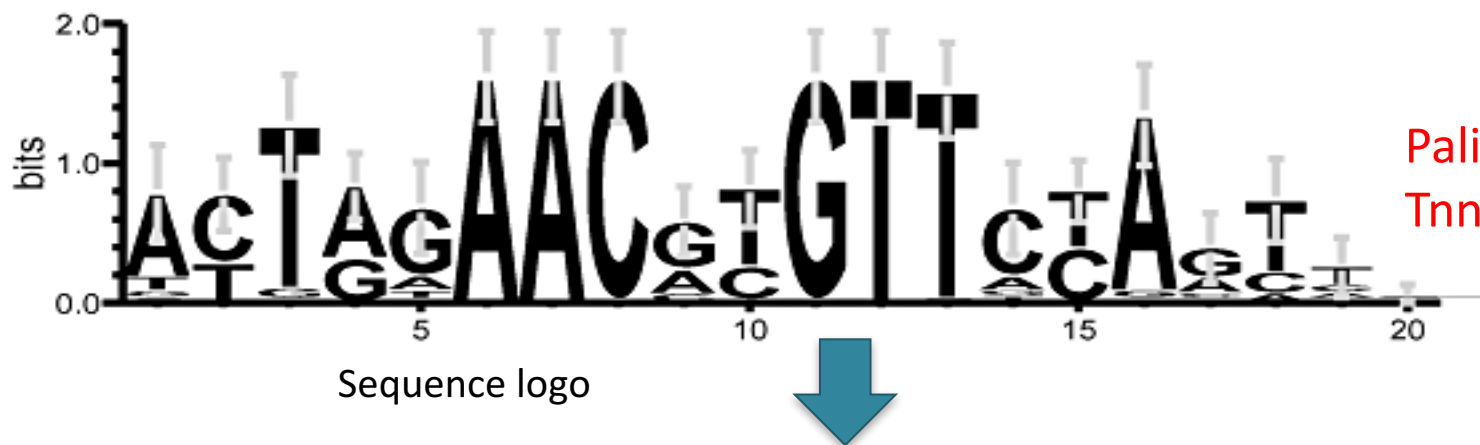
- ★ A highly conserved transcriptional repressor controls a large regulon involved in lipid degradation in *Mycobacterium smegmatis* and *Mycobacterium tuberculosis*
- *M.Tb* TetR-type regulator Rv3574 is implicated in pathogenesis as it is induced *in vivo*
- Genome-wide essentiality studies show Rv3574 is required for infection.
- Many of these genes are known to be induced by growth on cholesterol in *Rhodococci*, and palmitate in *M. tb*.



KstR regulates genes cluster encoding cholesterol catabolism in actinomycetes

- Transcriptomic analysis:
 - cluster of genes necessary for microbial steroid degradation.
 - more than 83 genes in *M.smeg* and 74 *M.tb*, 18 are essential.
- Identification of a conserved motif within its own promoter (TnnAACnnGTTnnA)

MTB	TCGACTTGACGGTTGATCGTTAGCACTAT TAGAACACGTTCTAG TGGGC-AAGAACGCTC	237
MB	TCGACTTGACGGTTGATCGTTAGCACTAT TAGAACACGTTCTAG TGGGC-AAGAACGCTC	237
MM	CCAAC TTGACG -TTGGTCGTCTGCAC TATAGAACACGTTCTAG TAAGT-ATGAACGGTT	236
MSM	TCGACTTGACGGT CGAACGCTGTCAC TAT TAGAACACGTTCTAG TGAG--AAGCACGGAC	172
MAP	TCGACTTGACGGT CGATCACTGGCACTATAGAACACGTTCTAG TCTCCGGTGAGCGAGT	107
NFA	TTGTCTTGACTTCCGCCCGGCAGGGATAT TAGAACATGTTCTAG -----GTGTGGGAGT	208
	***** * * ***** *	



Palindromic motif
TnnAACnnGTTnnA

Kendall , 2007& 2010
Van der Geize, 2007

Do we have such motif in *B. dorei* D8 and *B. dorei* 175^T?
What are the genes induced by this motif?

★ *B. dorei* D8


Fuzznuc

 ★ *B. dorei* 175^T


Fuzznuc

Bd. D8

202 - 3== Taille 199 _____ Score 99.9 /bdd8_00467.fasta.a3m.D8.batch4.20150203.outBK4 ==> Transcriptional regulator
 192 - 1== Taille 191 _____ Score 100.0 /bdd8_01282.fasta.a3m.D8.batch4.20150203.outBK4 ==> TetR/AcrR family transcriptional regulator
 198 - 3== Taille 195 _____ Score 100.0 /bdd8_00441.fasta.a3m.D8.batch4.20150203.outBK4 ==> Putative uncharacterized protein
 192 - 4== Taille 188 _____ Score 100.0 /bdd8_01261.fasta.a3m.D8.batch4.20150203.outBK4 ==> Putative transcriptional regulator
 203- 4== Taille 199 _____ Score 100.0 /bdd8_01039.fasta.a3m.D8.batch4.20150203.outBK4 ==> Putative transcriptional regulator

5 putative -or annotated- transcriptional regulators

B. Dorei

215 - 46== Taille 169 _____ dorei-17855/bacdor_01099.fasta.a3m.dorei.outBK4 ==> transcriptional regulator, TetR family
 192 - 1== Taille 191 _____ dorei-17855/bacdor_00223.fasta.a3m.dorei.outBK4 ==> transcriptional regulator, TetR family
 192 - 4== Taille 188 _____ dorei-17855/bacdor_00052.fasta.a3m.dorei.outBK4 ==> transcriptional regulator, TetR family
 198 - 4== Taille 194 _____ dorei-17855/bacdor_03957.fasta.a3m.dorei.outBK4 ==> transcriptional regulator, TetR family
 203 - 4== Taille 199 _____ dorei-17855/bacdor_02664.fasta.a3m.dorei.outBK4 ==> transcriptional regulator, TetR family
 204- 4== Taille 200 _____ dorei-17855/bacdor_01097.fasta.a3m.dorei.outBK4 ==> transcriptional regulator, TetR family

6 putative -or annotated- transcriptional regulator

Analysis in progress to identify genes involved in cholesterol or lipid catabolism.