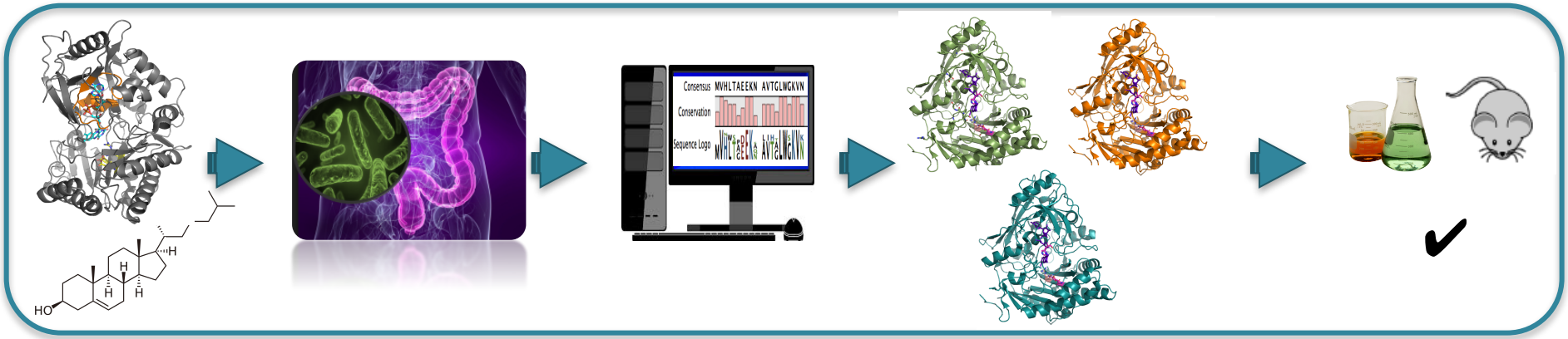
Towards the structural screening of mircobial ecosystems

Gwenaëlle André-Leroux

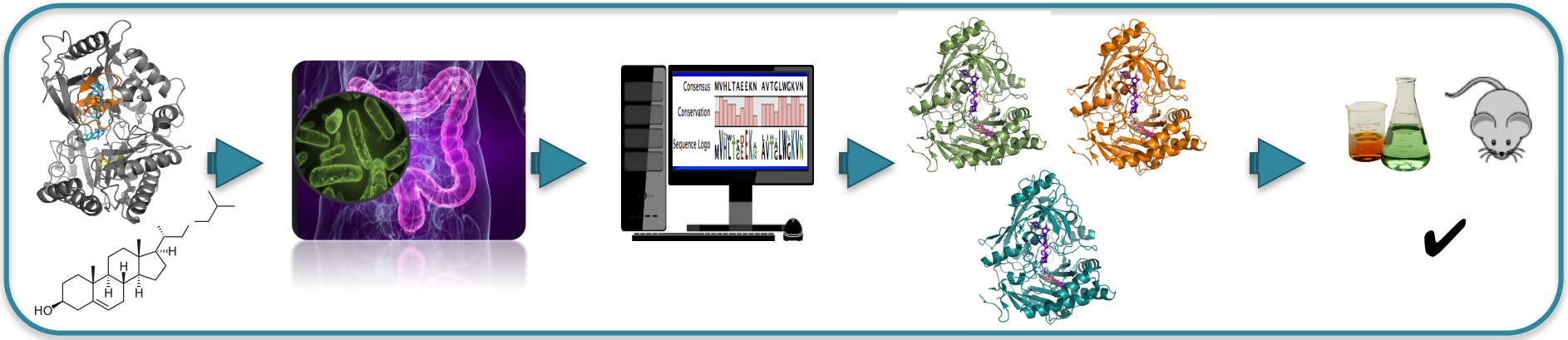
Towards the Structural Screening of Microbial Ecosystems



Gwenaëlle André-Leroux

Montevideo October 23rd

Towards the Structural Screening of Microbial Ecosystems



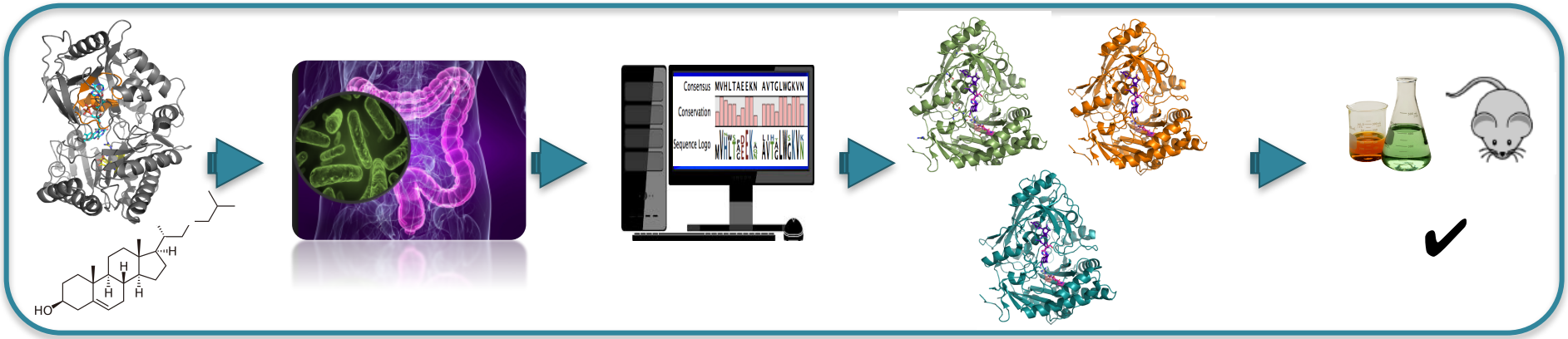
Outline

- 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

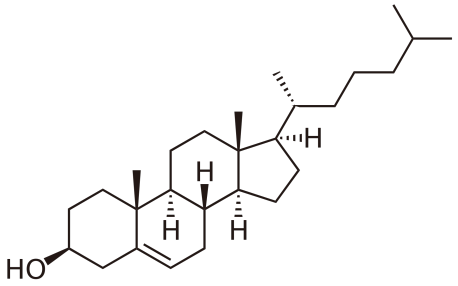
- Conclusions and perspectives

Metabolism of cholesterol by the gut microbiome

✪ Cholesterol :

- Is a sterol that plays a central role as cell-membrane agent
- Is precursor of steroïd 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 intestin, 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.

Two cholesterol-reducing bacterial strains have been isolated

✪ One strain from hog-sewage lagoon: Gram+ coccobacillus *Eubacterium coprostanoligenes* ATCC 51222T Freier *et al*, 1993

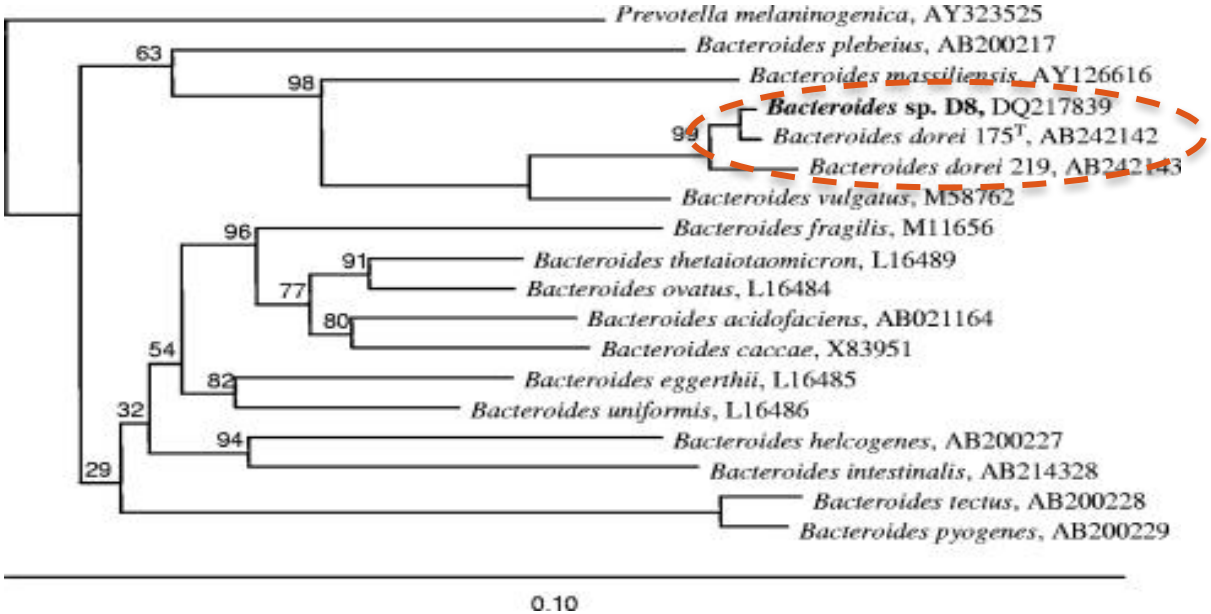
Oral administration of *E. copro* → significant decrease of plasma cholesterol in dietary induced hypercholesterolemic rabbits Li *et al*, 1995

✪ 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



Two cholesterol-reducing bacterial strains have been isolated

✪ 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

Valentin Loux

Two cholesterol-reducing bacterial strains have been isolated

Phylogenetically close

Phylogenetically remote

Strain \ Medium	SBM	BCM
B. Dorei 175^T	-	-
B. Dorei D8	+	-
E. coprostanoligenes	-	+

SBM : Standard Brain Medium
BCM: Basal Cholesterol Medium

Coll. C. Juste and P. Gérard, Micalis

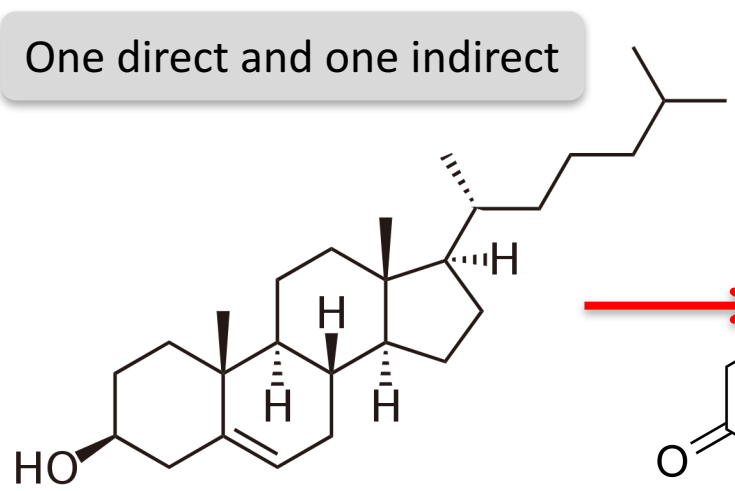
Metabolic pathway for cholesterol to coprostanol conversion in *B. dorei D8*?

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

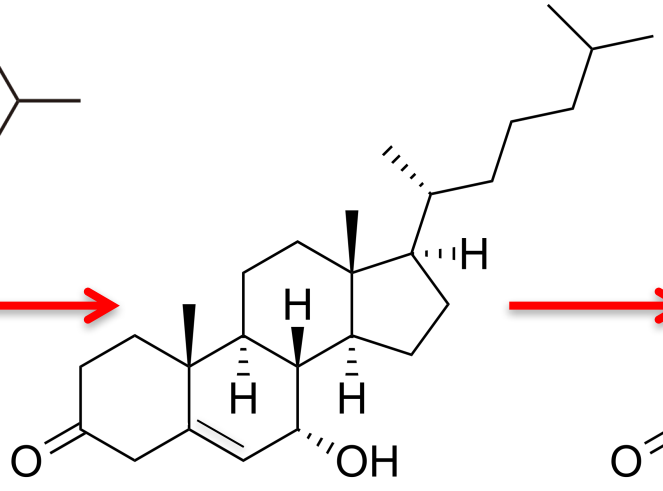
Cholesterol metabolic pathway -KEGG

One direct and one indirect

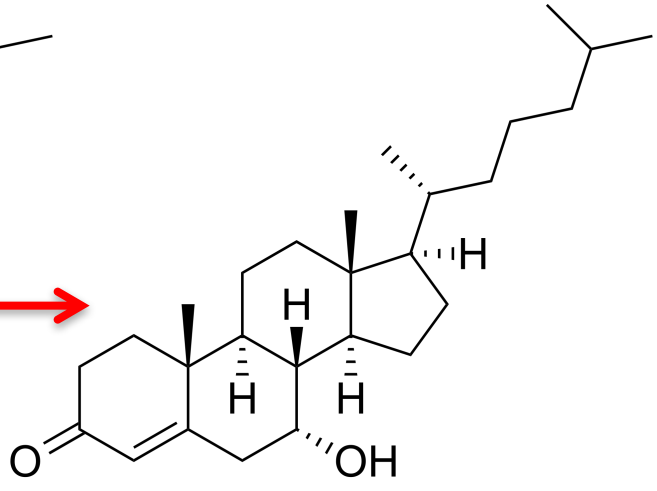
Aerobie bacteria



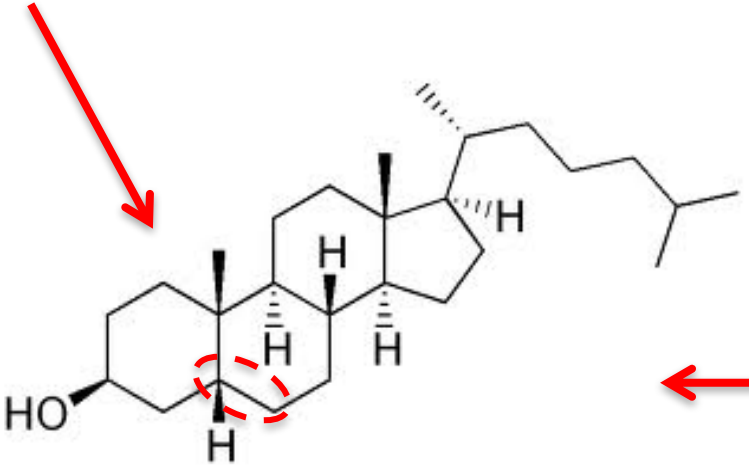
Cholesterol



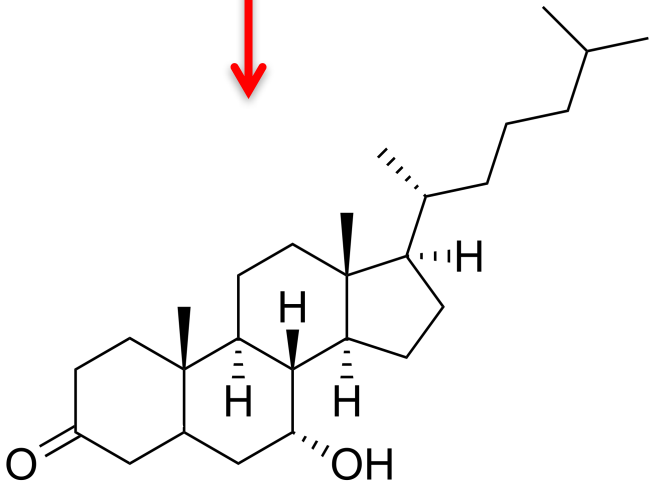
5-Cholesten-3-one transient



4-Cholesten-3-one



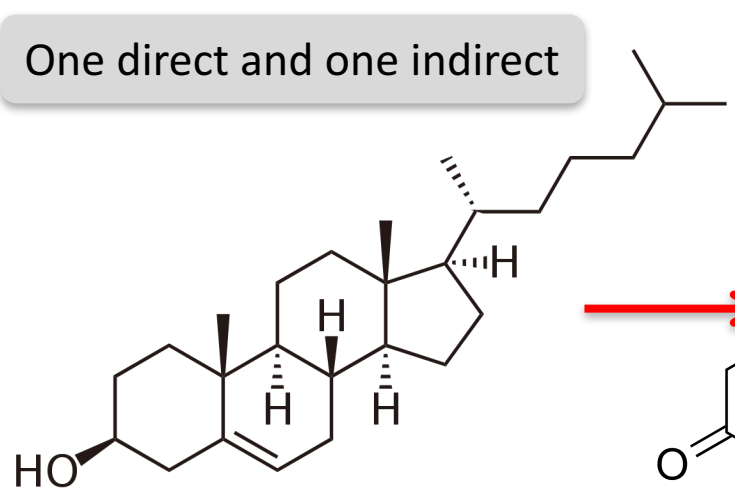
Coprostanol



Coprostanone

Cholesterol metabolic pathway -KEGG

One direct and one indirect

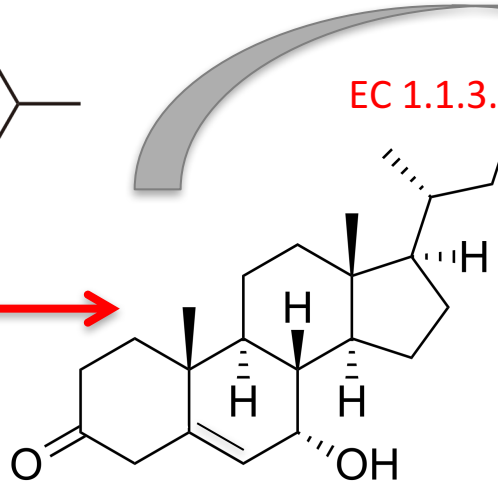


Cholesterol

KEGG: C00187

Enzyme A cholesterol oxidase

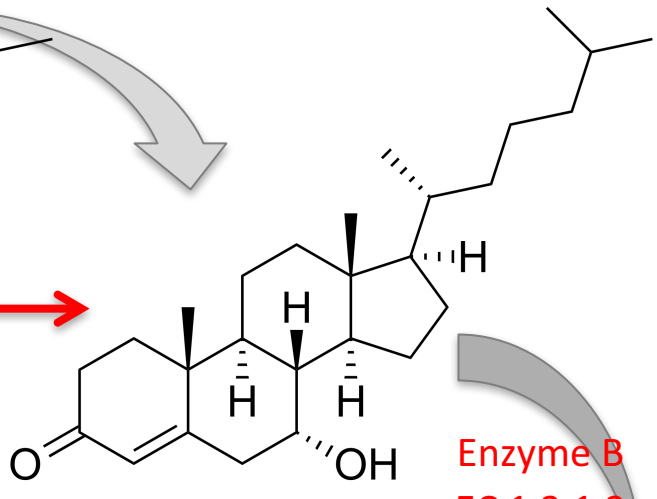
EC 1.1.3.6



5-Cholesten-3-one transient

KEGG: ?

Aerobie bacteria



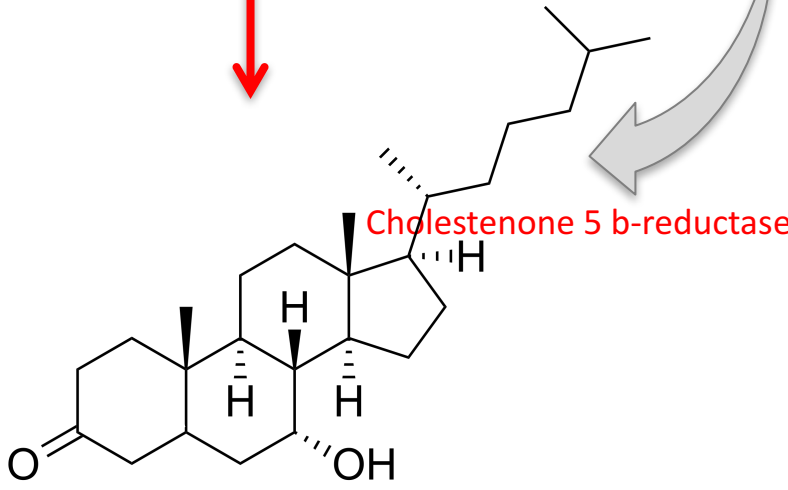
4-Cholesten-3-one

KEGG: C00599

Enzyme B

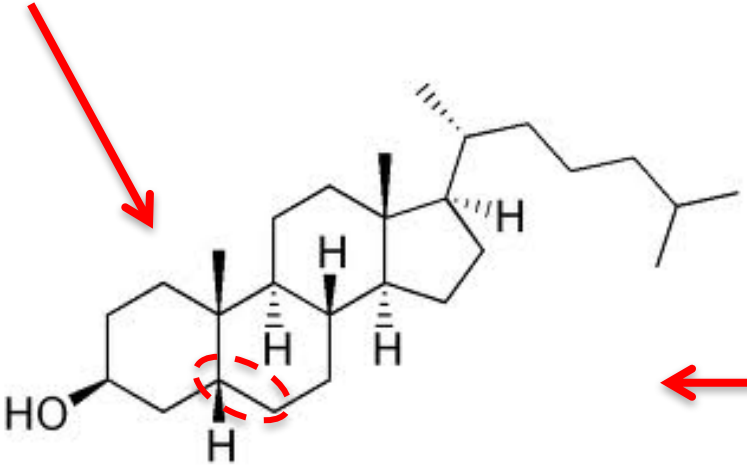
EC 1.3.1.3

Cholestenone 5 b-reductase



Coprostanone

KEGG:C03091



Coprostanol

KEGG: ?

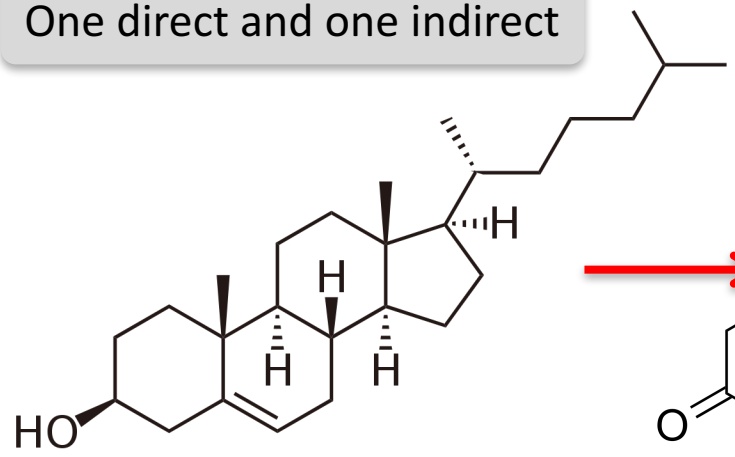
Cholesterol metabolic pathway -KEGG

One direct and one indirect

Enzyme A cholesterol oxidase

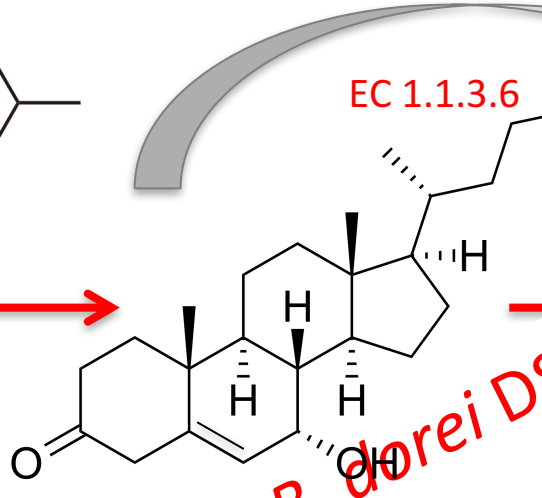
EC 1.1.3.6

Aerobie bacteria



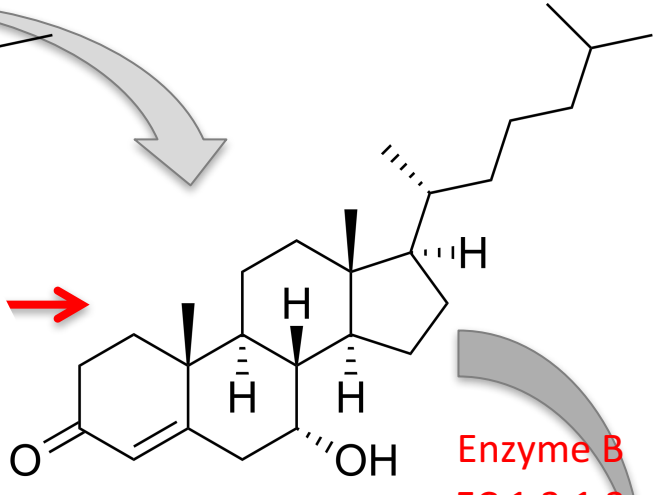
Cholesterol

KEGG: C00187



5-Chol… …ansient

KEGG: ?



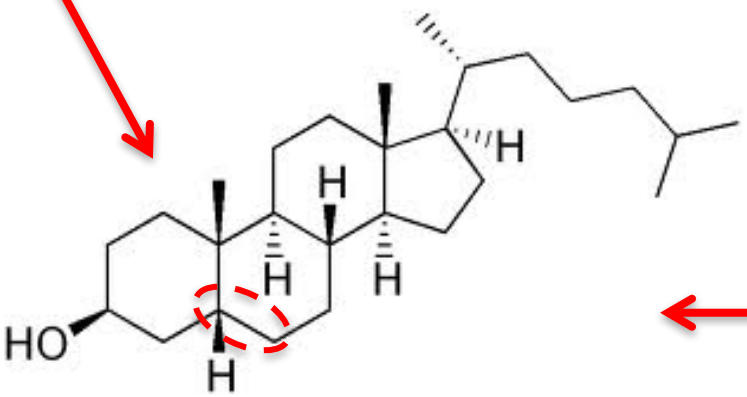
4-Cholesten-3-one

KEGG: C00599

Enzyme B

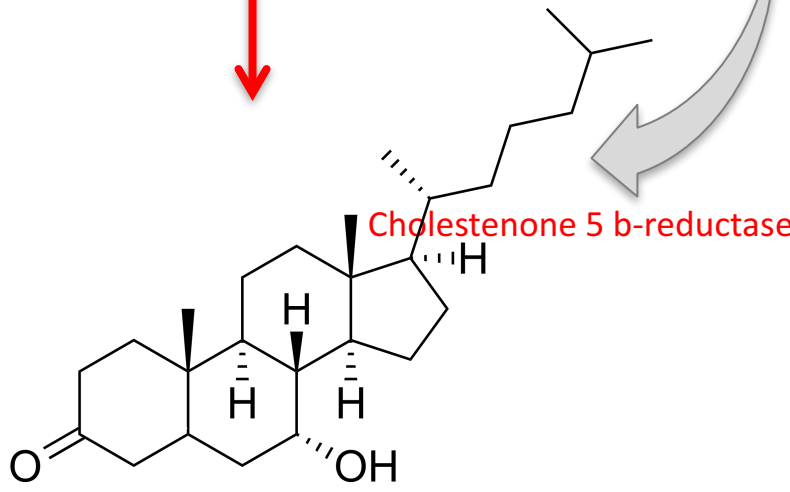
EC 1.3.1.3

Enzyme A in B. dorei D8?



Coprostanol

KEGG: ?



Cholestenone 5 b-reductase

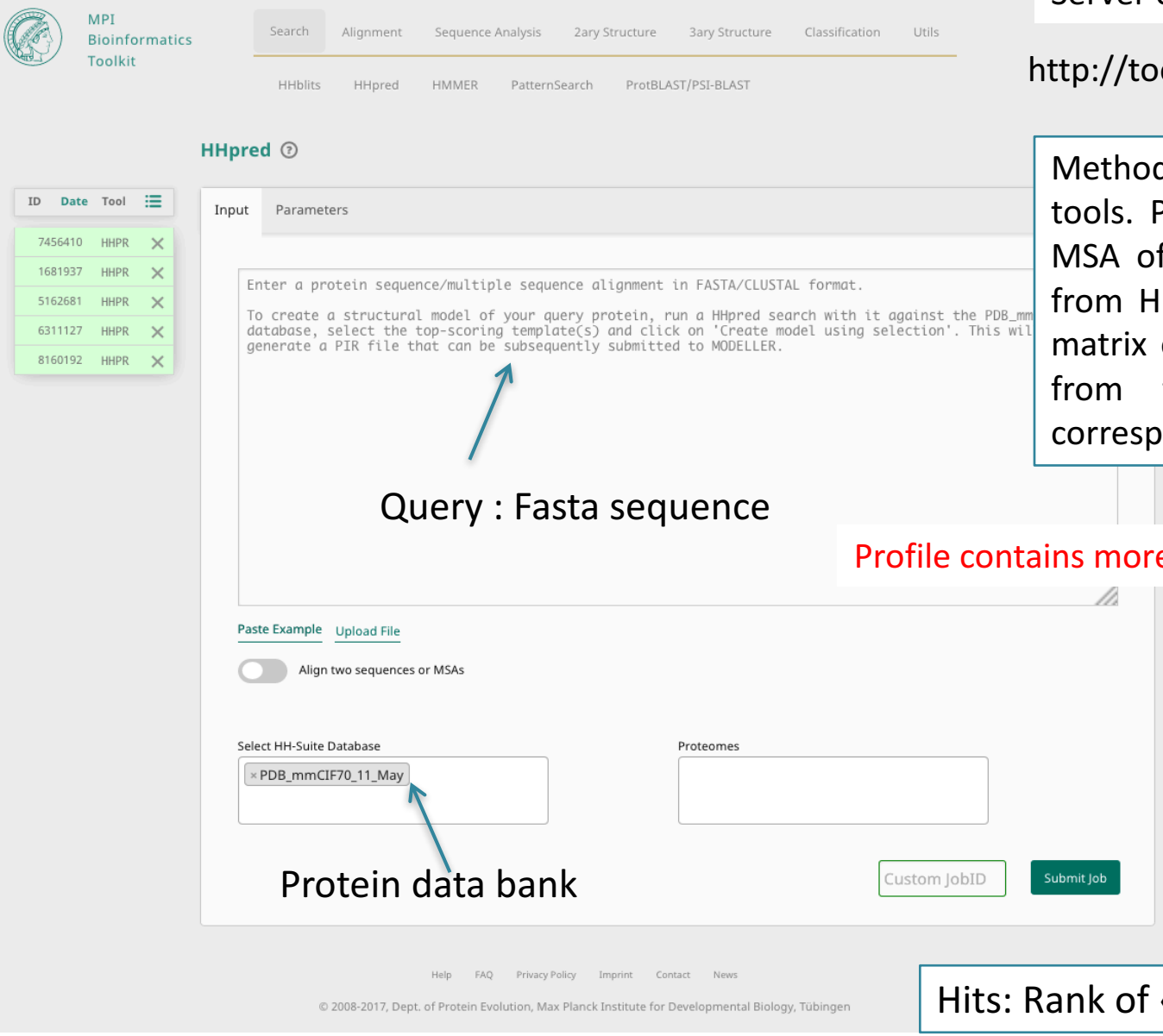
Coprostanone

KEGG:C03091

HHpred homology Detection & Structure Prediction

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

J. Söding, NAR 2005

Inverse strategy: setting up of standalone commands

Véronique Martin

- 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*

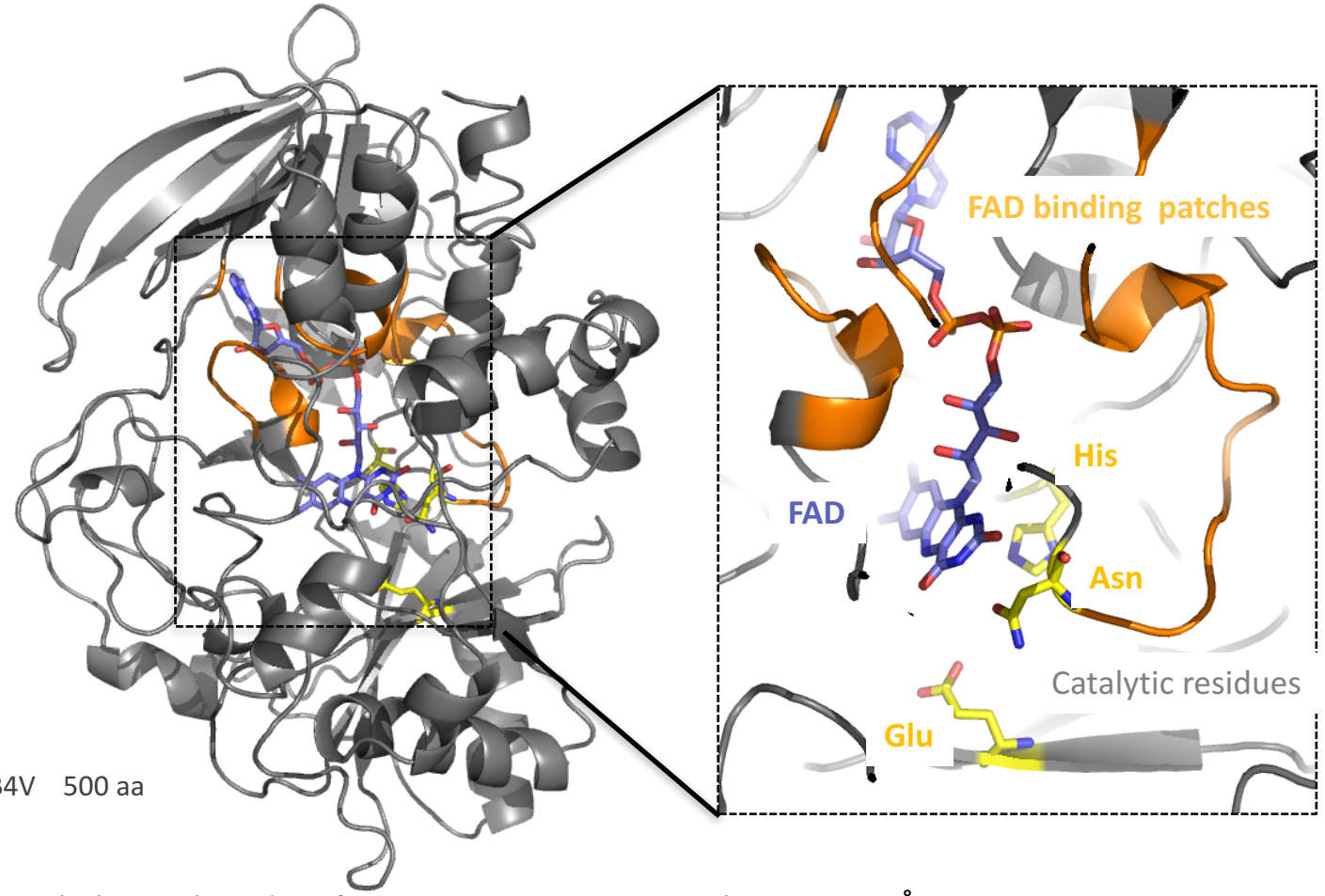
The 3D template « bank » is critical

Bank of cholesterol oxidases EC 1.1.3.6

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

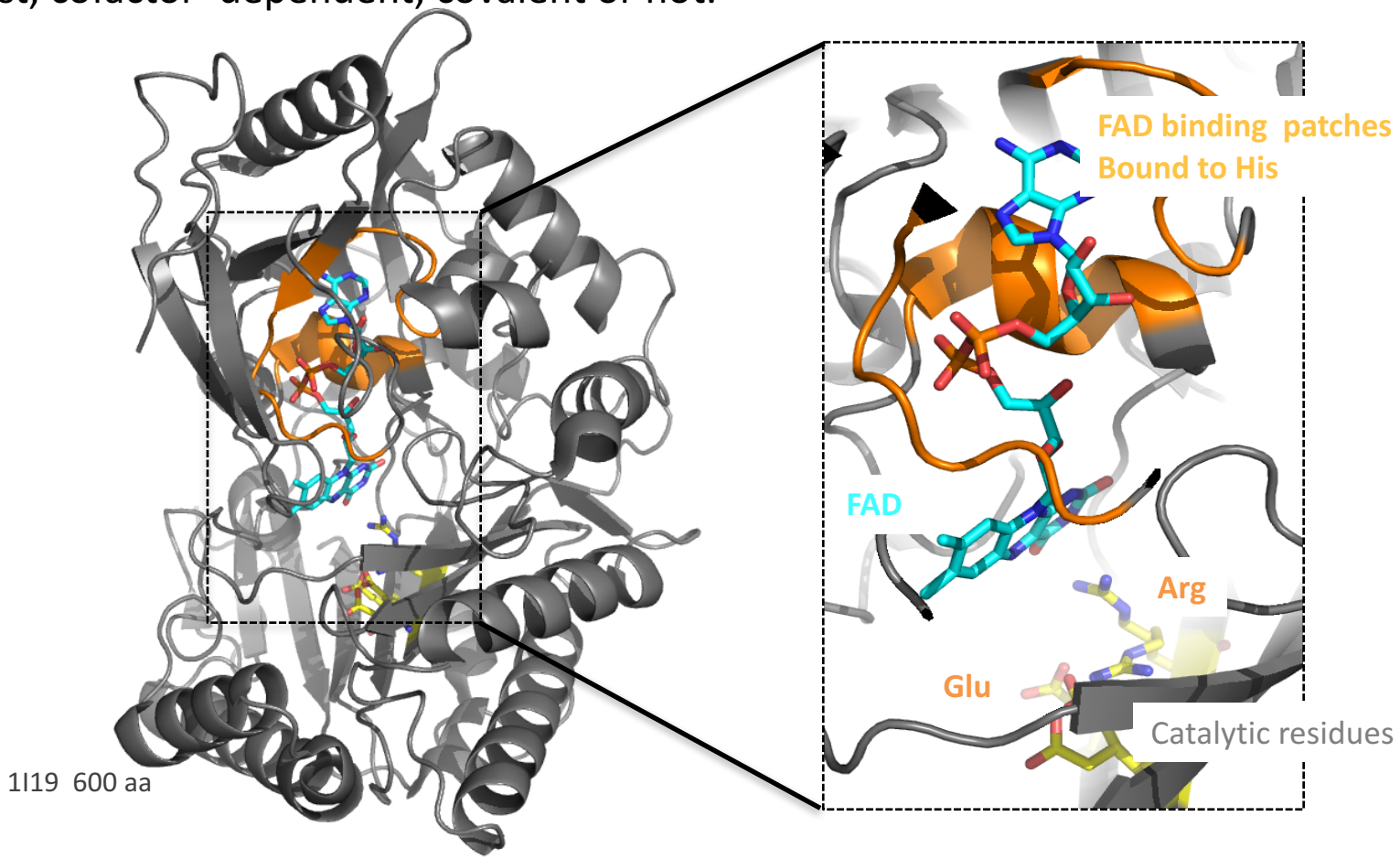
1COY.pdb & fasta — Cholesterol oxydase from *R. hoagii ex B. steroliticum* Li et al, 1993 - 1.8 Å

Bank of cholesterol oxidases EC 1.1.3.6

Cholesterol oxidase:

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

Catalytic site
Glu, Arg

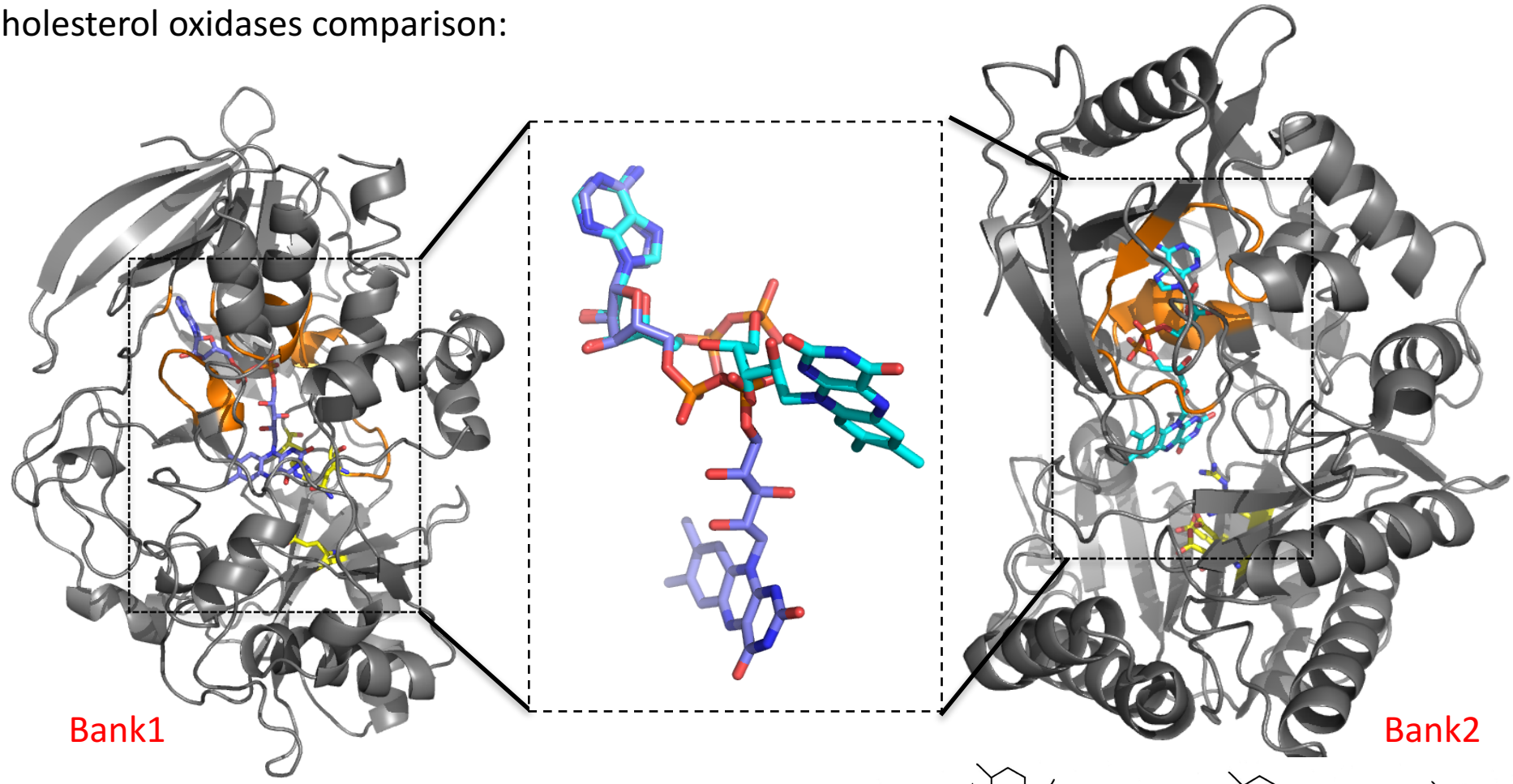


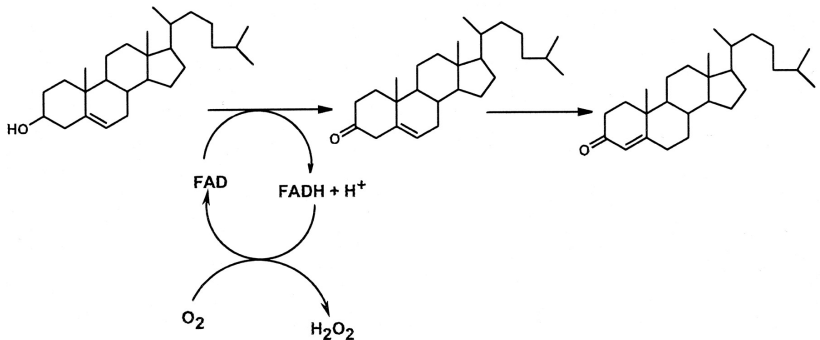
3JS8. pdb & fasta — Cholesterol oxydase from *Chromobacterium sp. DS-1* Sagermann et al, 2010 -1.5 Å

1I19.pdb & fasta — Cholesterol oxydase from *B. steroliticum* Coulombe et al, 2001 - 1.7 Å

Bank of cholesterol oxidases EC 1.1.3.6

Cholesterol oxidases comparison:

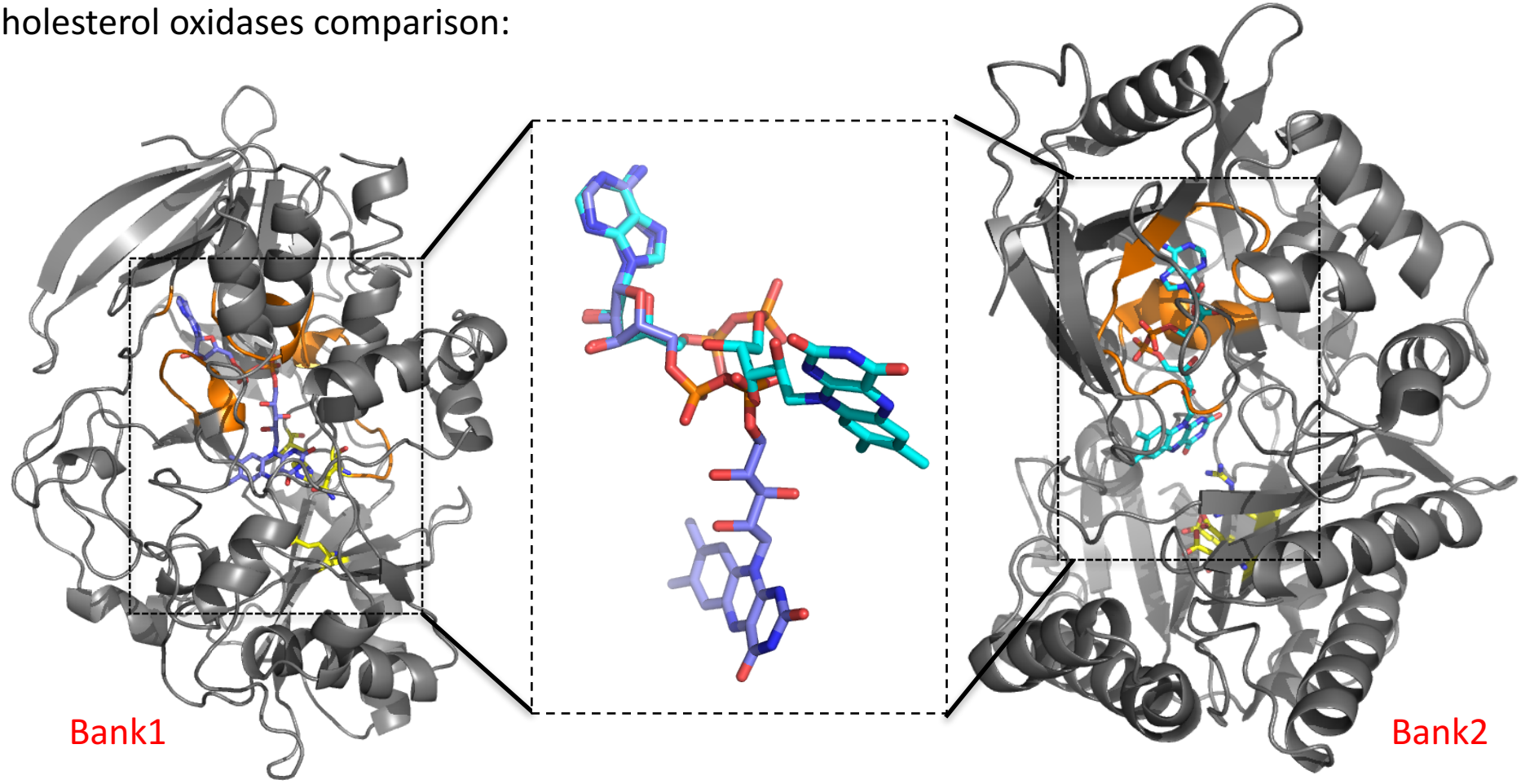




- 3D structures distinct
- Catalytic sites different
- Binding patches divergent
- FAD binding topologies contrasting despite same redox role

Bank of cholesterol oxidases EC 1.1.3.6

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

Results for *B. dorei* D8 with Bank1 and Bank2

✪ *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=69 _________  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-
```

Results for *B. dorei* D8 with Bank1 and Bank2

✪ *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

bdd8_00352.fasta.a3m.hhm.out.hhmakeBK1 | bdd8_00254.fasta.a3m.hhm.out.hhmakeBK1 | bdd8_03708.fasta.a3m.hhm.out.hhmakeBK1 | bdd8_03122.fasta.a3m.hhm.out.hhmakeBK1

```
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       hhsearch -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 1COY: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 1COY: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

Q ss_pred            HHHHHHHHHHcCCCeEEEcccEEEeeeecCccCccccccCCCCceEEEEEEEecCCC---eEEEEEeCeEEEecCCCCcc
Q BdD8_00352     133 IQTSLIEAVKAHPNITIFTDHYAVEIITQHHLGIIVTRHTPGIKCFGAYVLNEKTG---EVDTFLSKVTVMATGGCEAV  208 (526)
Q Consensus      133 ~~~~L~~~~~~~~gv~i~~~~~V~~l~~~~~~~~~g~~~~~~v~Gv~~~~~~g---~~~~i~Ak~VIlAtGg~~~~  208 (526)
                     ....+.+.+.+.++++|++++.|++|+.+++.    +      +++||.+.+. +|   +...++|+.||||+|++...
T Consensus      223 ~~~~~l~~a~~~g~v~i~~~~V~~I~~~~~~~~~~v~gV~~~~~~g~~~~~~i~A~~VVlaaGa~~t~  290 (504)
T 1B4V:A         223 LDKTYLAAALGTGKVTIQTLHQVKTIRQTKDG----G------YALTVEQKDT-DGKLLATKEISCRYLFLGAGSLGST  290 (504)
T ss_pred            hhHHHHHHHHhCCCeEEEeCCeEEEEEecCCC----C------eEEEEEeCC-CCCeeeeEEEEcCEEEEcccCcCcH
Confidence            34455666666646999999999999987520    2      7899987663 23   246789999999999753


No 2
>1COY:A
Probab=99.10  E-value=6.5e-14  Score=116.46  Aligned_cols=64  Identities=14%  Similarity=0.189  Sum_probs=47.5

Q ss_pred            HHHHHHHHHHcCCCeEEEcccEEEeeeecCccCccccccCCCCceEEEEEEEecCCCe---EEEEEeCeEEEecCCCCc
Q BdD8_00352     133 IQTSLIEAVKAHPNITIFTDHYAVEIITQHHLGIIVTRHTPGIKCFGAYVLNEKTGE---VDTFLSKVTVMATGGCEA  207 (526)
Q Consensus      133 ~~~~L~~~~~~~~gv~i~~~~~V~~l~~~~~~~~~g~~~~~~v~Gv~~~~~~g~---~~~~i~Ak~VIlAtGg~~~  207 (526)
                     ....++..+.+.++++|++++.|++|+.+++.    +      +++||.+.+. +|.   ...+.|+.||||+|++..
T Consensus      228 ~~~~~L~~a~~~g~~i~~~~~V~~I~~~~~~~~~~v~gV~~~~~~g~~~~~~v~A~~VVlaaGa~~t  294 (507)
T 1COY:A         228 LDKTYLAQAAATGKLTITTLHRVTKVAPATGS----G------YSVTMEQIDE-QGNVVATKVVTADRVFFAAGSVGT  294 (507)
T ss_pred            cCcchHHHHHHcCCcEEEEeCCEEEEEecCCC----c------eEEEEEEEcC-CCceeeeEEEEcCEEEEcccccccch
Confidence            33445566666458999999999999987520    3      7899987653 242   467999999999999865


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

Q ss_pred            ccccEEEECCcHHHHHHHHHhhc-CCcEEEEeccCCC
Q BdD8_00352         3 KKFDFLVIGSGLAGMSFALKVAH-KGKVALICKAGLE   38 (526)
Q Consensus          3 ~~~DVlIIG~G~AGl~AA~~a~~-g~~V~lvek~~~~   38 (526)
                      .++||||||+|++|+++|++|++ |.+|++||+...
T Consensus          4 ~~~DvIVVGaG~aG~~A~~La~~G~~VlvlE~G~~~   40 (504)
T 1B4V:A             4 GYVPAVVIGTGYGAAVSALRLGEAGVQTLMLEMGQLW   40 (504)
T ss_pred            cccEEEEEccChhHHHHHHHHhcCCcEEEEecCCCC
Confidence            46999999999999999999 8999999999533
```

✪ *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

bdd8_03708.fasta.a3m.hhm.out.hhmakeBK1 | bdd8_03122.fasta.a3m.hhm.out.hhmakeBK1 | HHM_le1erhit.hhmakeBK2.txt.final2 | bdd8_03618.fasta.a3m.hhm.out.hhmakeBK2

```
Query         BdD8_03618 UDP-N-acetylenolpyruvoylglucosamine reductase lcl|Agmial:UEPSD|D8#1381
Match_columns 337
No_of_seqs    264 out of 1760
Neff          7.5
Searched_HMMs 2
Date          Mon Feb 23 18:15:31 2015
Command       hhsearch -i /projet/mig/work/vmartin/D8/D8.batch4.20150203/fasta_split/hhm/bdd8_03618.fasta.a3m.hhm -d /projet/mig/work/vmartin/D8/BANK2/base/BANK2_hhm_db -cov 10 -v -cpu 8 -o /projet/mig/work/vmartin/D8/
D8.batch4.20150203/fasta_split/hhm/bdd8_03618.fasta.a3m.hhm.out.hhmakeBK2

 No Hit                             Prob E-value P-value  Score    SS Cols Query HMM  Template HMM
  1 1I19:A                          99.9 2.6E-25 1.3E-25  195.3  17.0  162   18-190    33-223 (561)
  2 3JS8:A                          99.8 6.6E-24 3.3E-24  184.2  18.0  165   17-190    26-219 (540)

No 1
>1I19:A
Probab=99.86  E-value=2.6e-25  Score=195.27  Aligned_cols=162  Identities=15%  Similarity=0.117  Sum_probs=134.7

Q ss_pred              CccccEEEEcCCHHHHHHHHHhhhcCCcEEEEcCCCC---CCccCCC---CceEEEe-C-CCceEEEecCCCeEEEEEe
Q BdD8_03618       18 DVKASLFVEYESVTELQSILSDKNLMAGNWLHIGGGSN---LLFKGDY---AGTVLHS-A-IRGYEVVAENEQEIEVRAG   89 (337)
Q Consensus       18 gg~a~~~v~p~~~edl~~~v~~a~~~~ip~~ilGgGsN---~~l~~d~~~~~GvvI~~~~~~~~i~~~~~~~~~~~~~v~v~   89 (337)
                    ...+..+++|+|++||++++++++|++++++|++++|+|||   +.+.++.   +|++|++ + |++|++ +.+++..+|+|+
T Consensus       33 ~~~p~~vv~P~s~eev~~~~v~~a~~~~~~~v~~~G~Ghs~~~~~~~~~~~~~~~gv~i~l~~~~l~~i~i~d~~~~~~~~v~v~  111 (561)
T 1I19:A         33 MLDATWVCSPKTPQDVVRLANWAHEHDYKIRPRGAMHGWTPLTVEKGANVEKVILADTMTHLNGITV-NTGGPVATVTAG  111 (561)
T ss_pred              ccCccEEEecCCHHHHHHHHHHHHHCCCcEEEEcCCCCCCCCccccCCcccccceEEEEcCCCCCceEE-cCCCCeeEEEEc
Confidence        4457899999999999999999999999999999999   4454432    6899999 6 999965 653223499999

Q ss_pred              cCCcHHHHHHHHHHCCCCCccCCCCCcc-cchHHHHHhhhc---------------cCcchhhheeEEEEEEcCCCc---
Q BdD8_03618       90 AGEVWDDFVAYTVKNSWYGAENLSLIPG-EVGASAVQNIGA---------------YGVEAKDLIVSVETVEVETGR---  150 (337)
Q Consensus       90 AG~~l~~l~~~~~~~~Gl~Gle~l~gIPG-TVGGav~~NaG~---------------yG~~i~d~v~~v~vv~~~~G~---  150 (337)
                    ||++|.+|.+++.++||. ++.+.++++ ||||++++|+++               || .++|+|++++++|++ +|+
T Consensus      112 aG~~~~~L~~~~l~~~Gl~~l~~~~~~~~~~TvGGaiat~~hG~g~~~~~~~~~~~l~~~~~G~~~~d~V~~~~~V~~~~G~~~~  188 (561)
T 1I19:A        112 AGASIEAIVTELQKHDLG-WANLPAPGVLSIGGALAVNAHGAALPAVGQTTLPGHTYG-SLSNLVTELTAVVW-NGTTYA  188 (561)
T ss_pred              CCccHHHHHHHHHhCCCe-eeeCCCCCcccccccccccCCCCCCCCccccCCCCCCccccc-CccccEEEEEEEee-CCcEee
Confidence        9999999999999999997 778889986 999999999755              55 49999999999999 599

Q ss_pred              EEEEecC--ccccccccccccccCceEEEEEEEEEeeeCCC
Q BdD8_03618      151 KRIFTKE--ECRYAYRESIFKKDLKGKYIVTYVTYRLSKQPV  190 (337)
Q Consensus      151 i~~~~~~~--e~~~~yR~s~f~~~~g~~~II~~a~~~l~~~~~  190 (337)
                    +++++++   |+++.+++.   +    ..  ||++++|++.|.+.
T Consensus      189 ~~~~~~~~~~dl~~~~~~g~---G---~~~~VIt~vtl~l~p~~~  223 (561)
T 1I19:A        189 LETYQRNDPRITPLLTNL---G---RC-FLTSVTMQAGPNFR  223 (561)
T ss_pred              EEEeccCCCChHHHHhcc---C---Cc-eeeEEEEEeEecCe
Confidence        5777765  565555543    2   47 99999999998765

No 2
>3JS8:A
Probab=99.83  E-value=6.6e-24  Score=184.16  Aligned_cols=165  Identities=12%  Similarity=0.046  Sum_probs=131.5

Q ss_pred              CCccccEEEEcCCHHHHHHHHHhhhcCCcEEEEcCCCCCC---ccCC---CCceEEEe-C-CCceEEEecCCCeEEEEE
Q BdD8_03618       17 MDVKASLFVEYESVTELQSILSDKNLMAGNWLHIGGGSNLL---FKGD---YAGTVLHS-A-IRGYEVVAENEQEIEVRA   88 (337)
Q Consensus       17 igg~a~~~v~p~~~edl~~~v~~a~~~~ip~~ilGgGsN~l---~~~d~~~~~~GvvI~~~~~~~~i~~~~~~~~~~~~~v~v   88 (337)
                    +...+..++.|+|++||++++++++|++++++|++++|+||..    ++++   .++++|++ + |++|++ |.+++..+|+|
T Consensus       26 ~~~~p~~v~~P~s~~ev~~~~v~~a~~~~~~~v~~~G~Ghs~~~~~~~~~~~~~~~~gv~i~l~~~~l~~i~i~d~~~~~~~~v~v  104 (540)
T 3JS8:A         26 IKVDDVWTCAPRSADEVVKVANWAKDNGYKVRARGMMHNWSPLTLAAGVSCPAVVLLDTTRYLTAMSI-DASGPVAKVTA  104 (540)
T ss_pred              cccCccEEEecCCHHHHHHHHHHHHHCCCcEEEEcCCCCCCCCcccCCCCcEEEEeccCCCCcEEe-CCCCCccEEEE
Confidence        3456789999999999999999999999999999999999953    3332    25799998 4 999954 65222238999
```

Results for *B. dorei 175*^{T} with Bank1 and Bank2

✪ *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

Results for *B. dorei 175*^{T} with Bank1 and Bank2

✪ *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

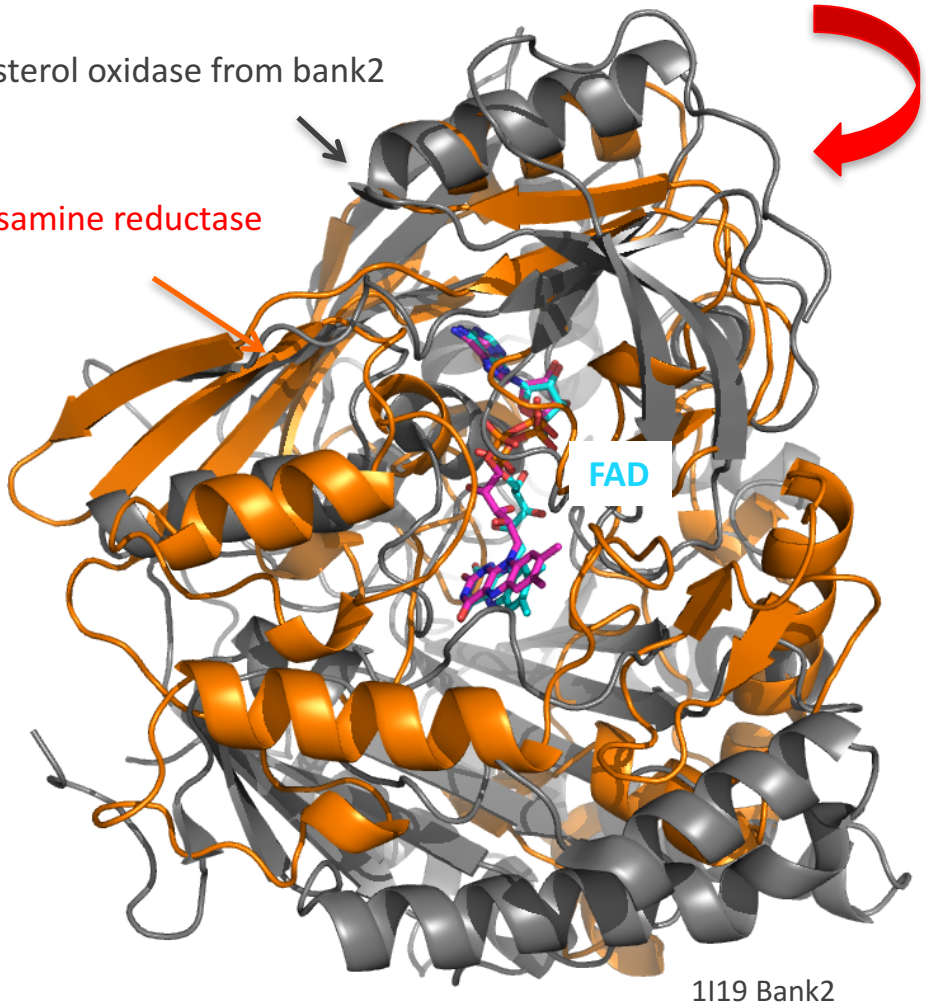
Results for *B. dorei* D8 with Bank2

✪ *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 Cholesterol oxidase from bank2

UDP-N-acetylenolpyruvoylglucosamine reductase



1I19 Bank2

Oxidoreductase MurB 1MBB E.coli Gram-

Results for *B. dorei* D8 with Bank2

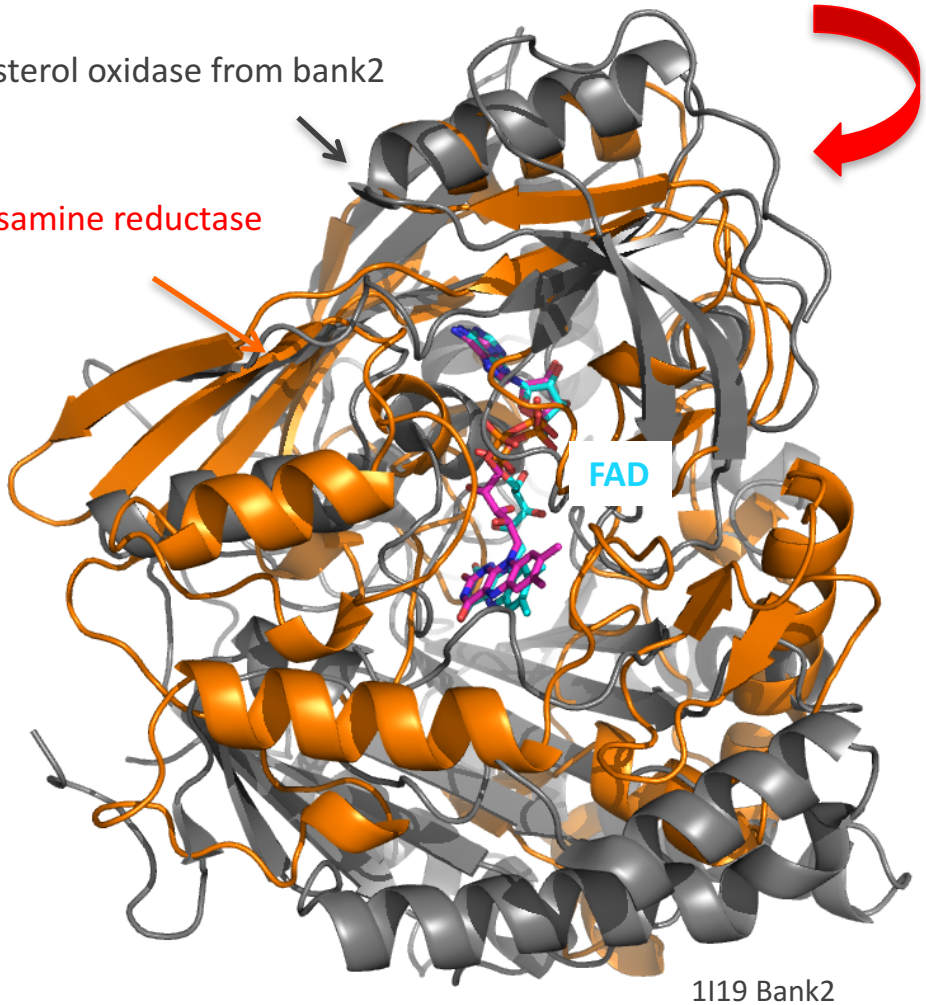
✪ *B. dorei* D8 Sorted by Prob > 95 and size cut off > 50 amino acid residues

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

1I19 Cholesterol oxidase from bank2

UDP-N-acetylenolpyruvoylglucosamine reductase

FAD



1I19 Bank2

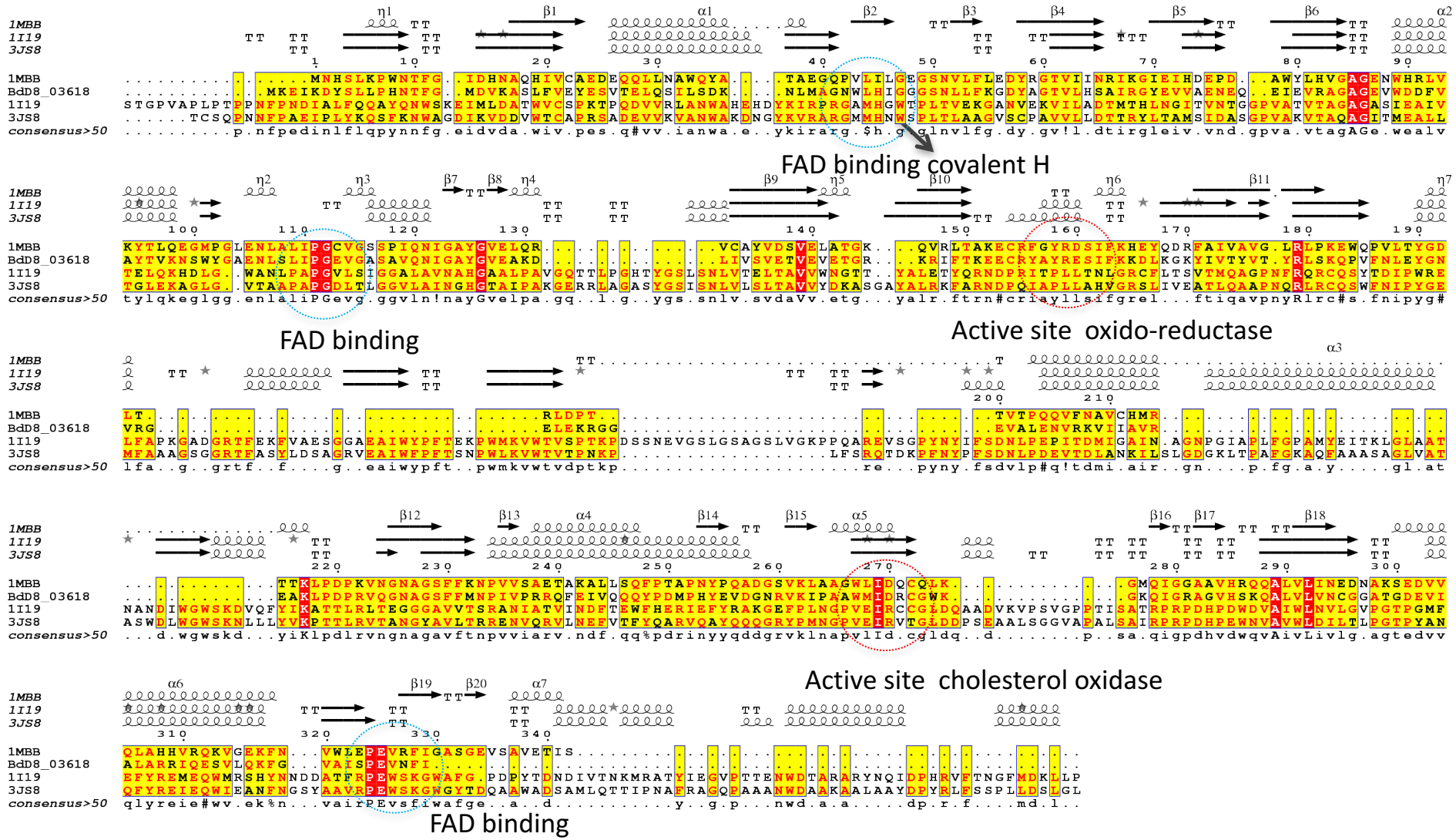
Oxidoreductase MurB

1MBB E.coli Gram-

Results for *B. dorei* D8 with Bank2

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

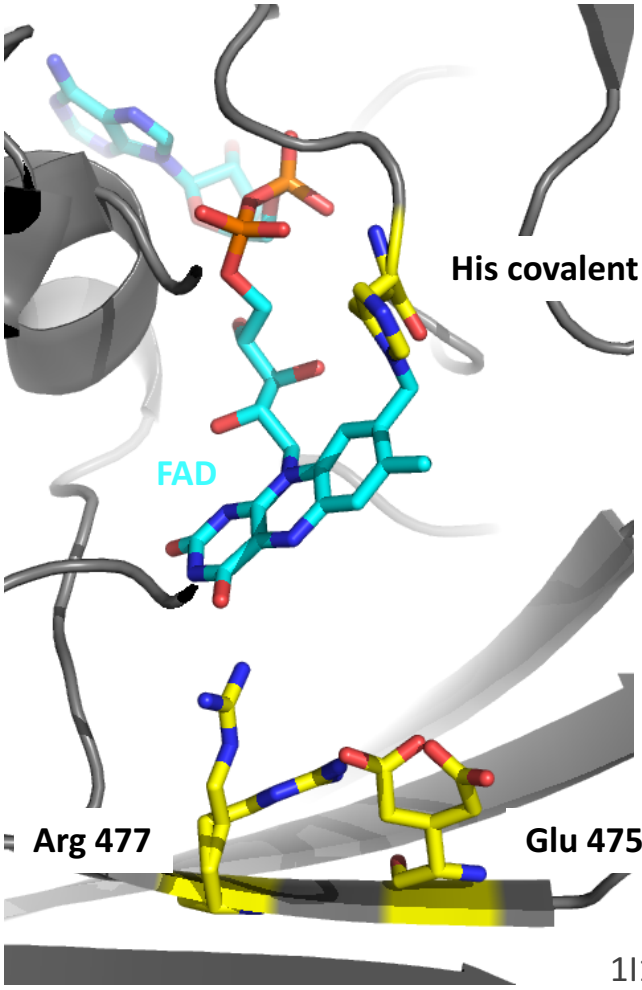
99.9 172 aa bdd8_003618.fasta.a3m.hhm.out.hhmakeBK2 → UDP-N-acetylenolpyruvoylglucosamine reductase



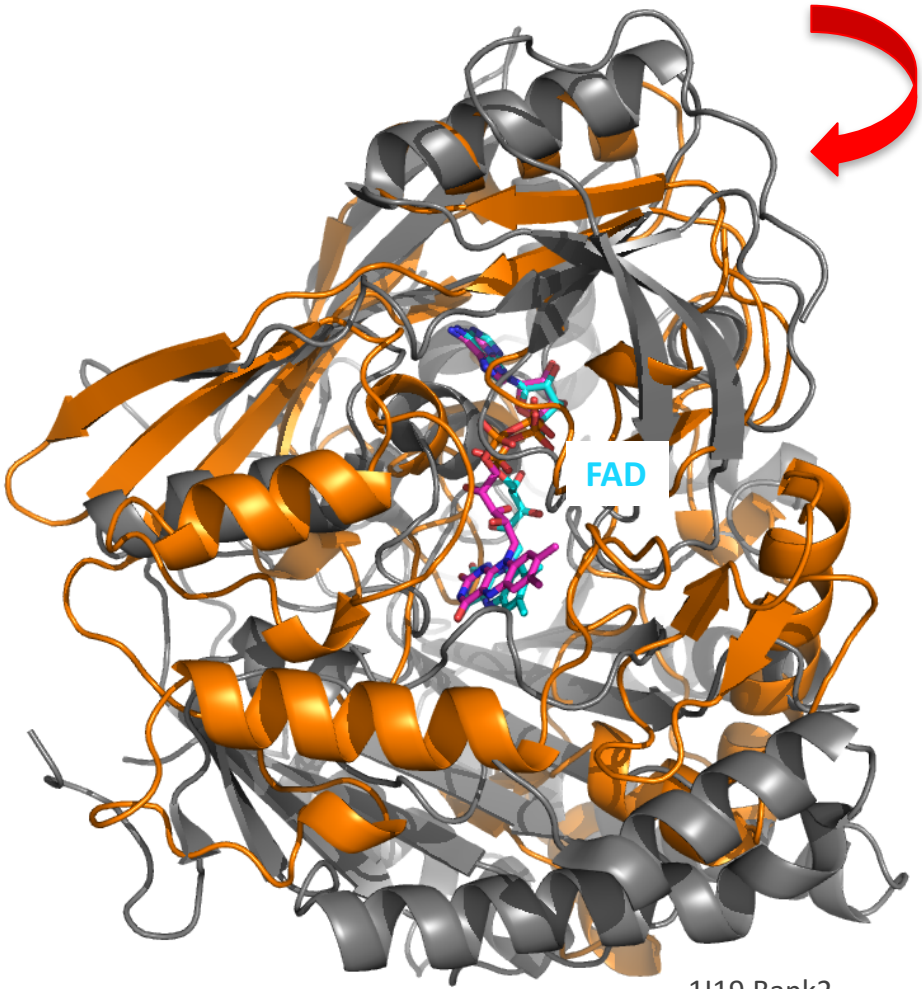
Results for *B. dorei* D8 with Bank2

✪ *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



1I19 Bank2

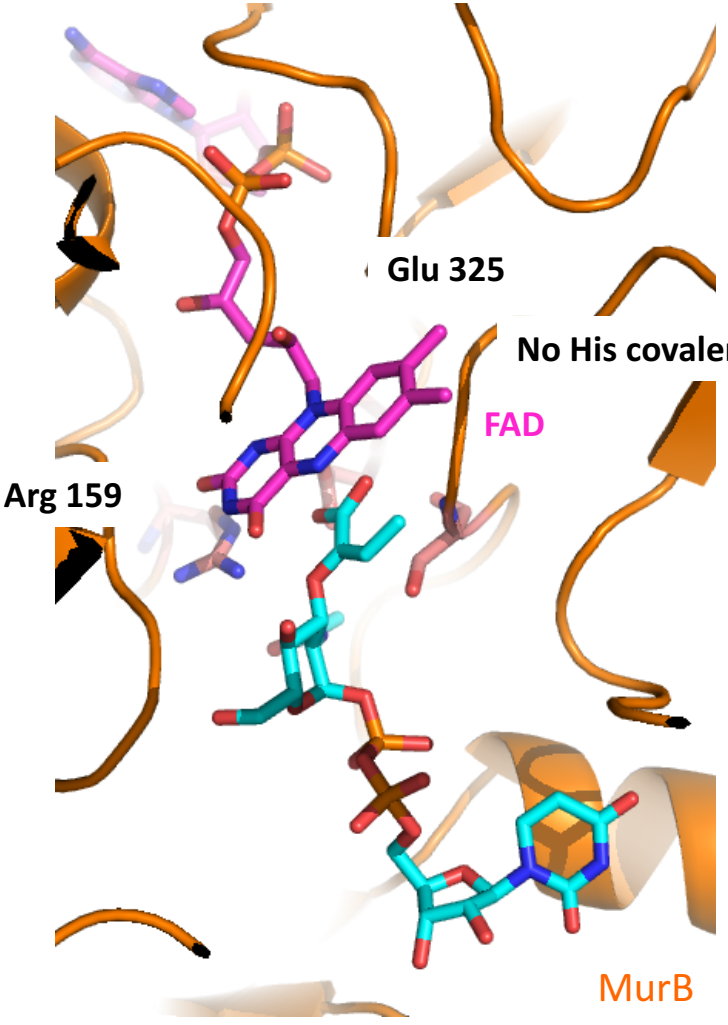
Oxidoreductase MurB

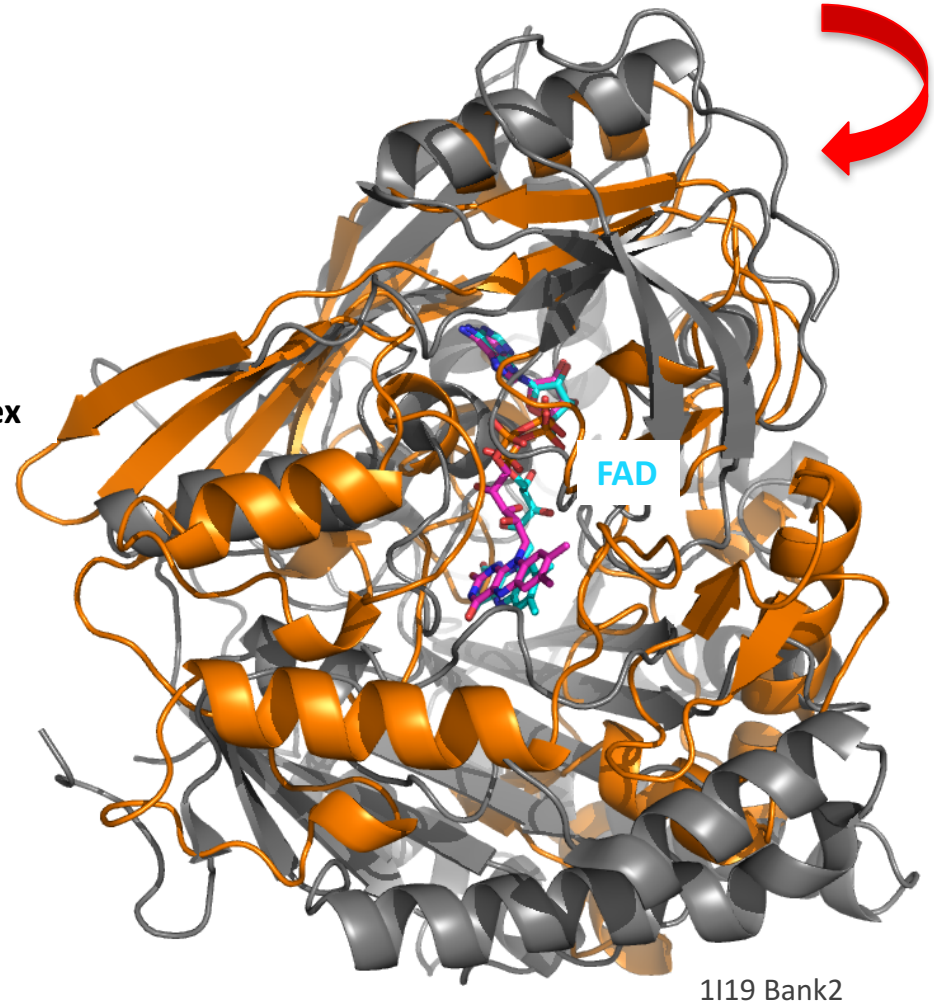
1MBB E.coli Gram-

Results for *B. dorei* D8 with Bank2

✪ *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

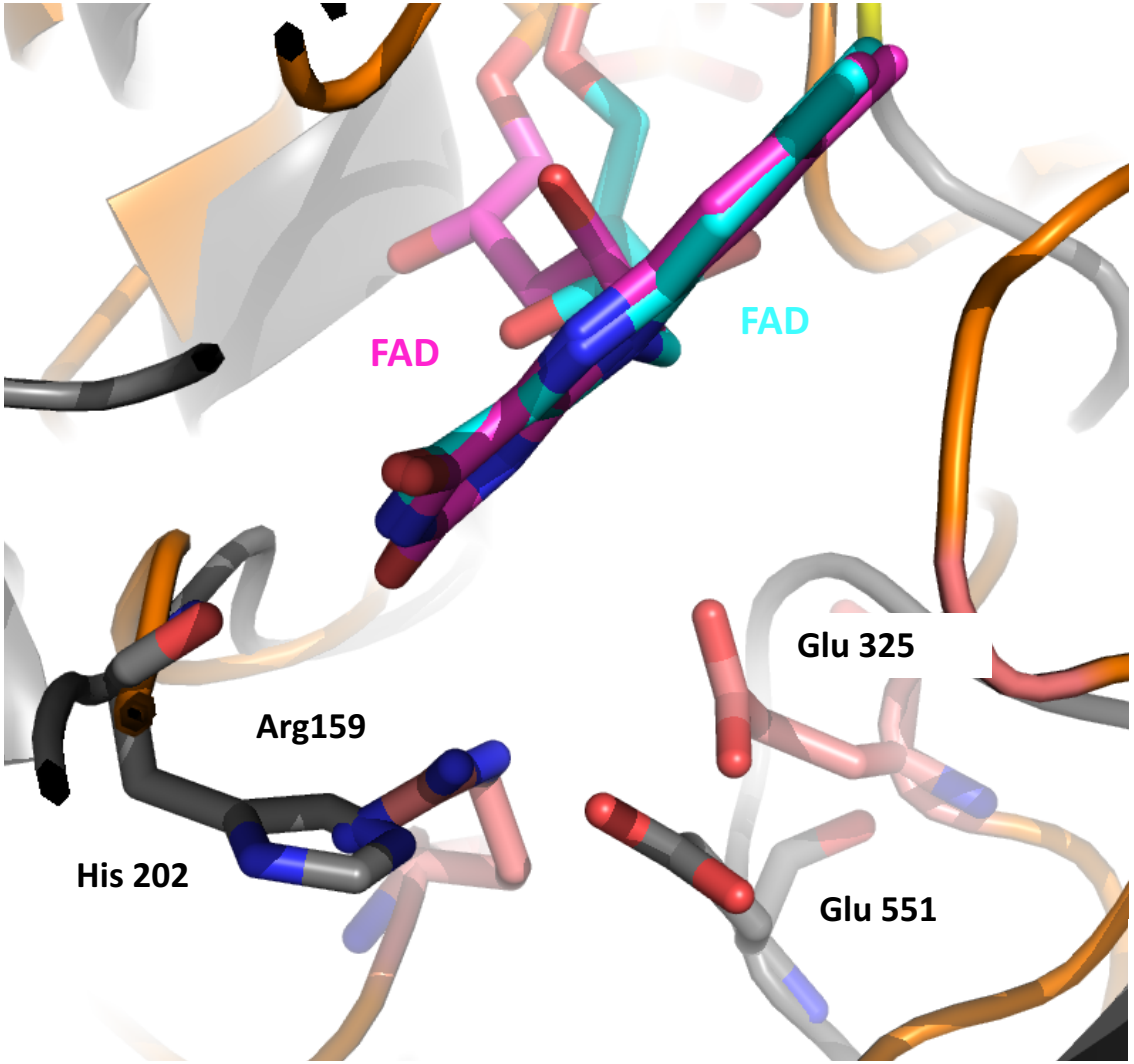
Oxidoreductase MurB

1MBB E.coli Gram-

Results for *B. dorei* D8 with Bank2

✪ *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

Results for *B. dorei 175*^{T} with Bank1 and Bank2

✪ *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.

Results for *B. dorei 175*^{T} with Bank1 and Bank2

✪ *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

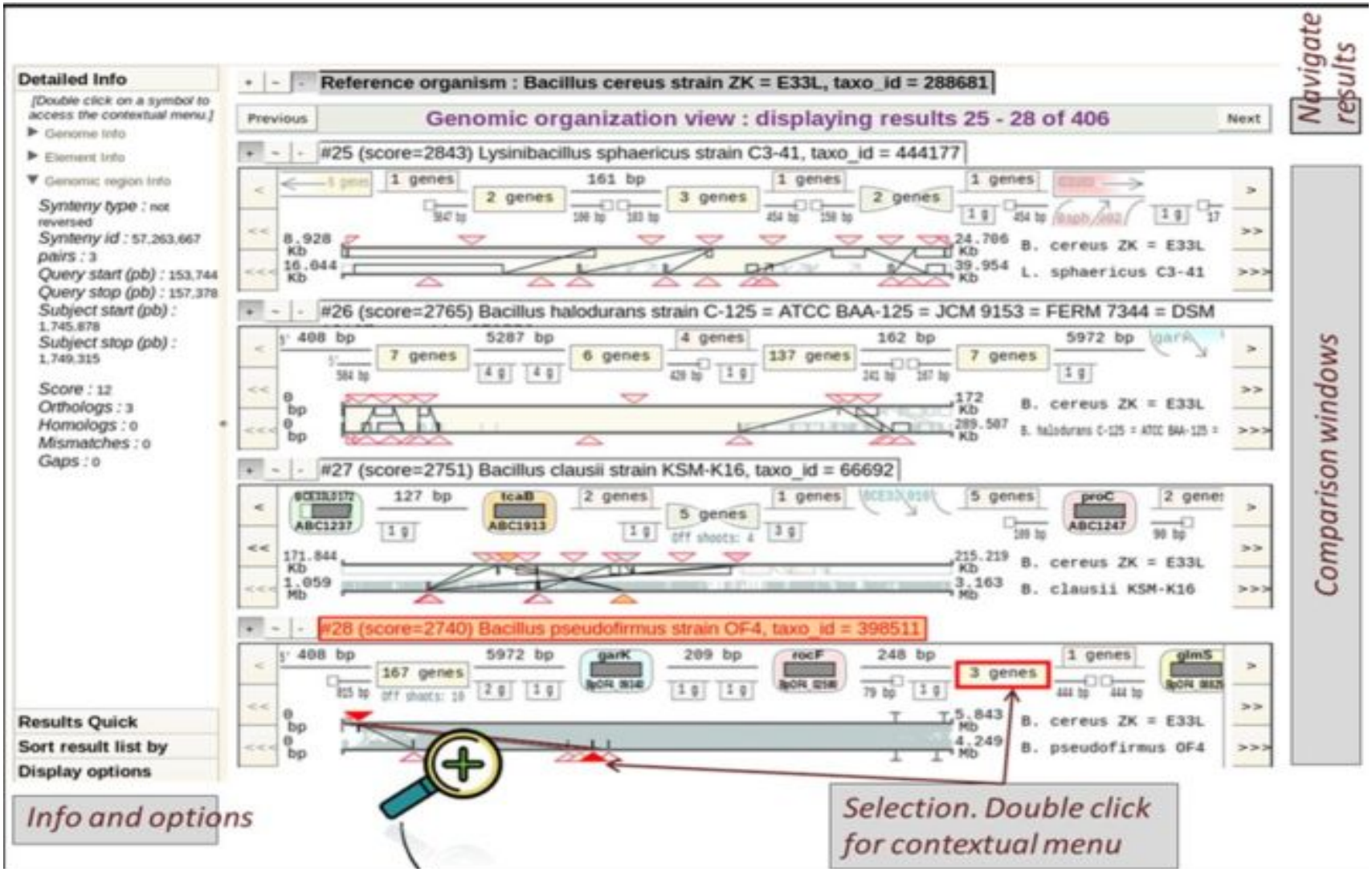
Analysis of synthenies by Thomas Lacroix using Insyght.

Analysis of the 5 genes from 3D, using Insyght

http://genome.jouy.inra.fr/Insyght

Thomas Lacroix

Powerful tool for genes comparison and syntheny inferences



Lacroix *et al*, Nar 2014

Analysis of the 5 genes from 3D, using Insyght

http://genome.jouy.inra.fr/Insyght

Thomas Lacroix

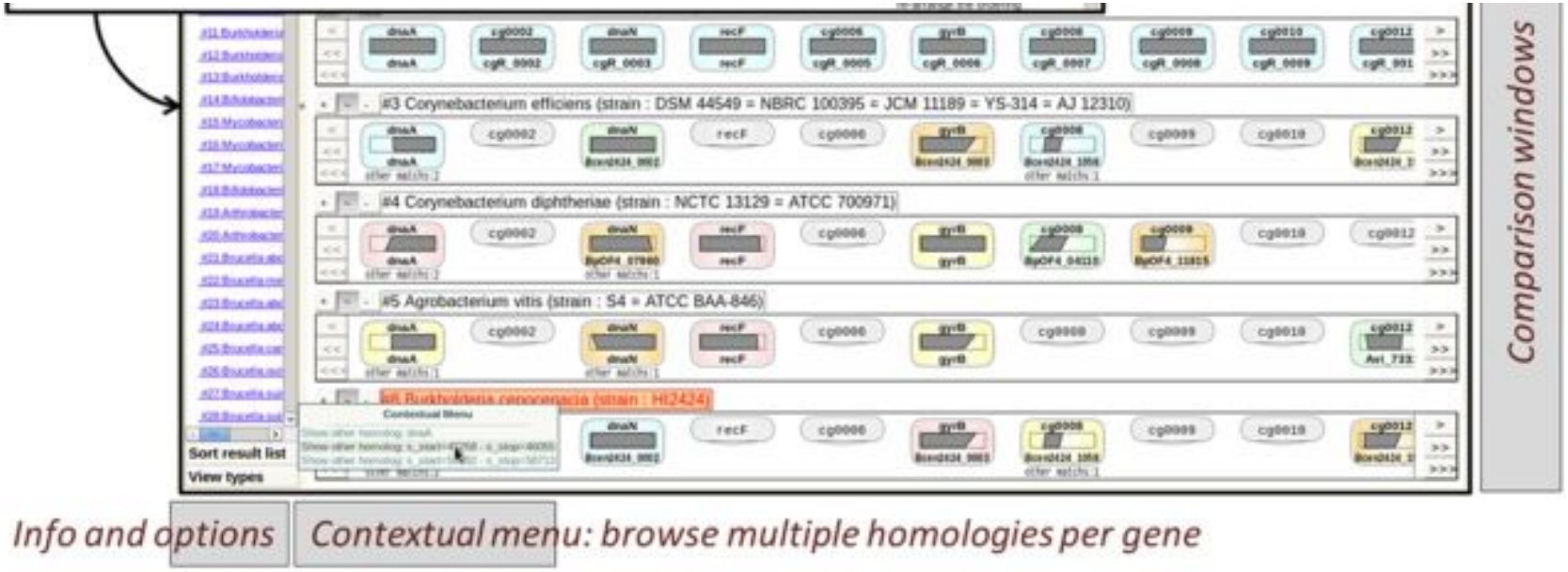
→ Powerful tool for genes comparison and syntheny inferences

Synthenies : co-localization of homologuous 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



Lacroix *et al*, Nar 2014

Analysis of the 5 genes from 3D, using Insyght

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

→ No significant hits with clear syntheny

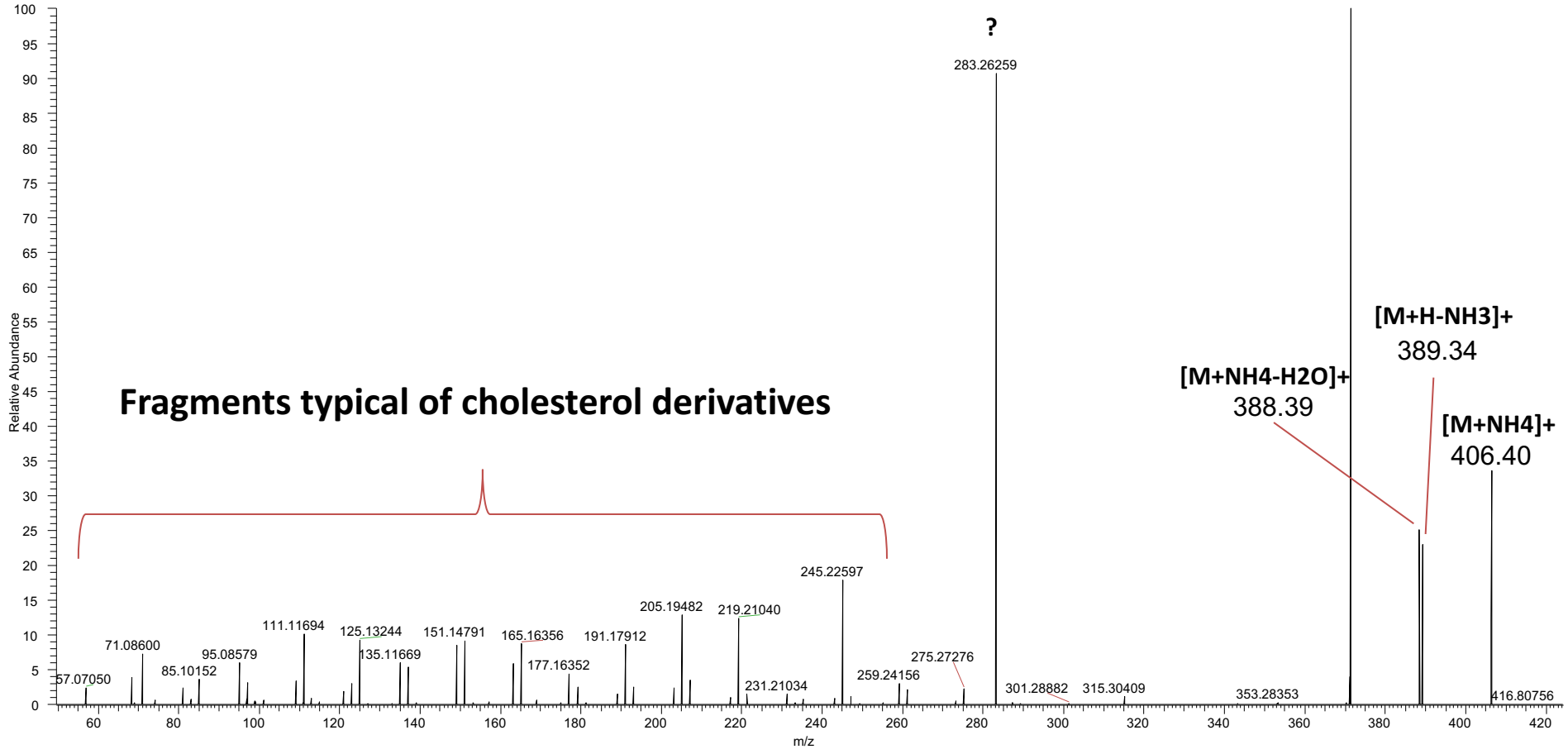
→ Need of transcriptomic analysis

Catherine. Juste
Fabienne Beguet-Crespel

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

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

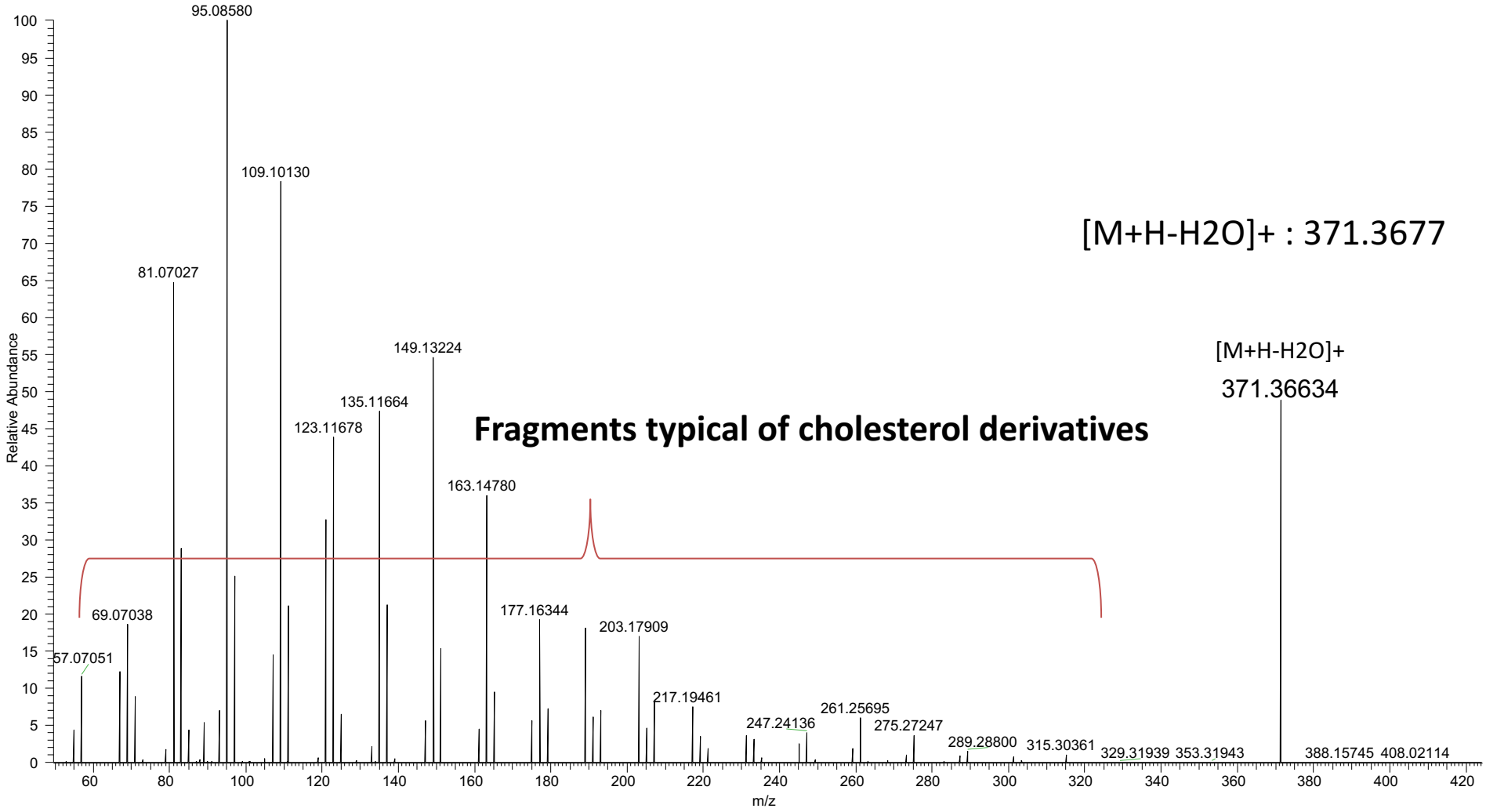
- Ions detection corresponding to COPROSTANOL
- DIMS-MS/MS method in positive ionisation mode



E. Coprostanoligenes samples

Catherine. Juste
Fabienne Beguet-Crespel

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

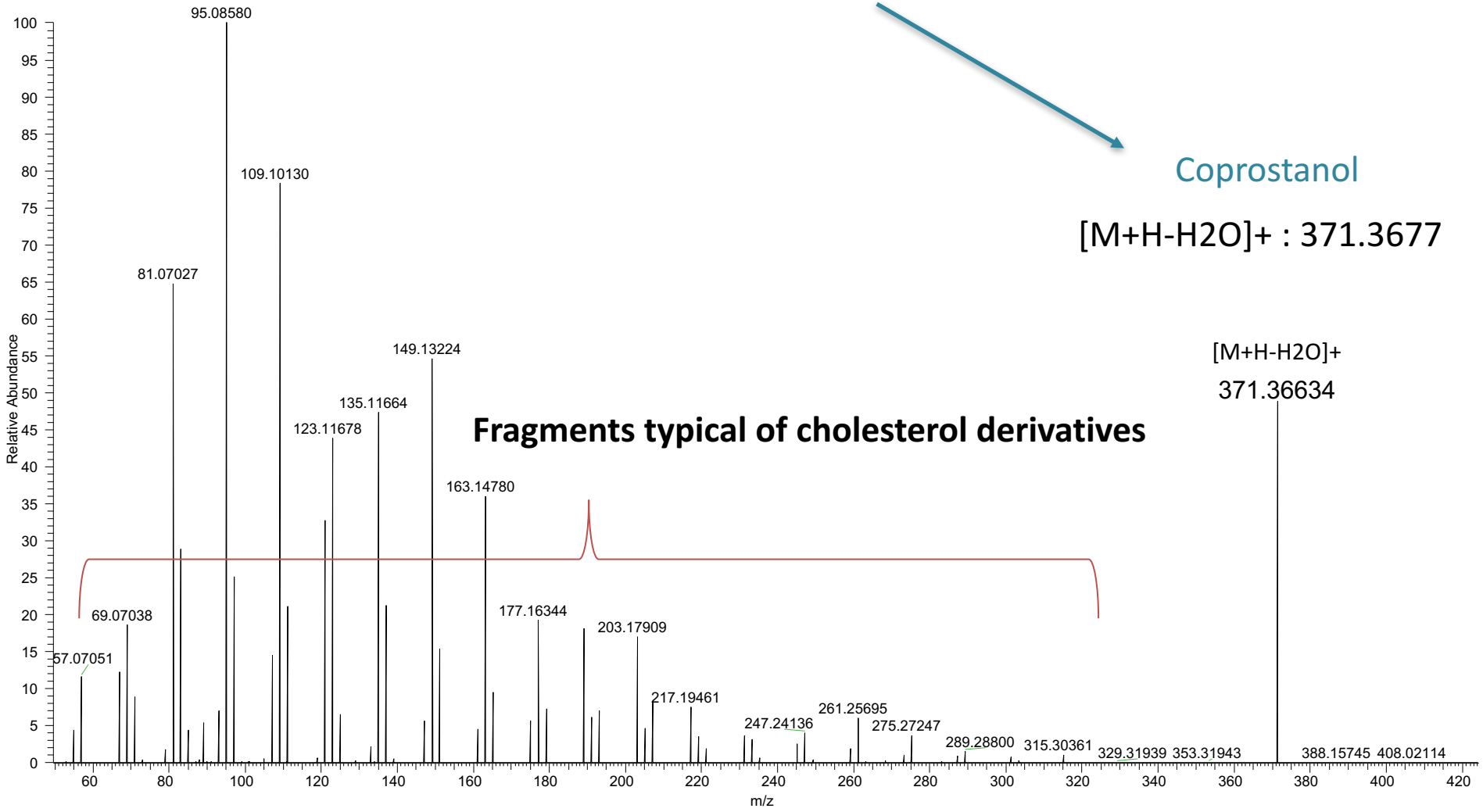


E. Coprostanoligenes samples

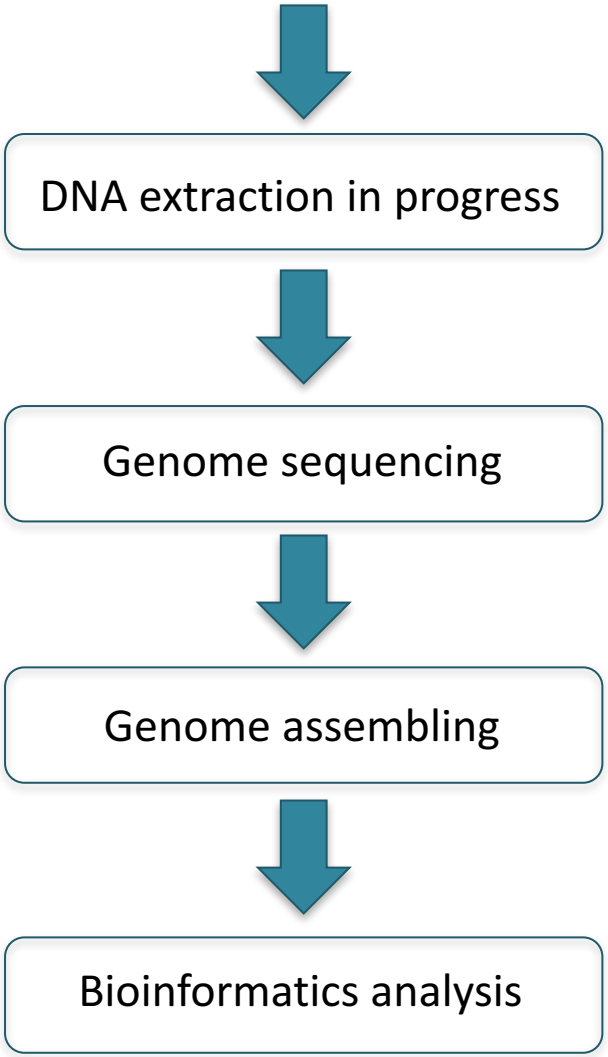
Catherine. Juste
Fabienne Beguet-Crespel

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

Coprostanol is present & preferentially detected



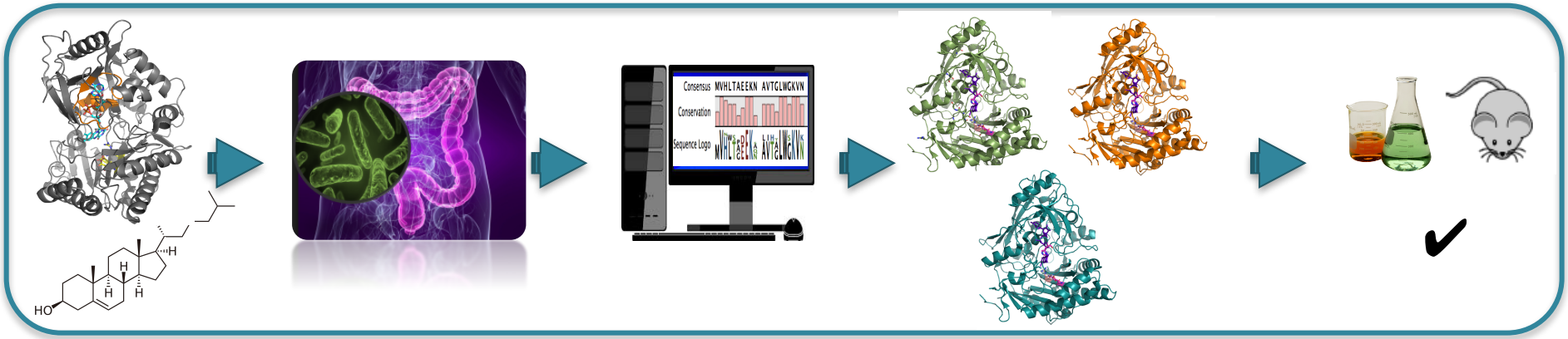
✪ E. Coprostanoligenes samples : Genome sequencing, assembling and analysis



Cholesterol conversion in the gut microbiota

- 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

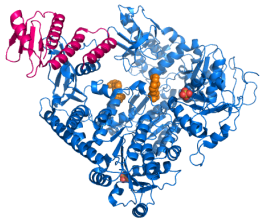
- Cholesterol conversion in the gut microbiota: the mystery enzyme(s) ?

- MetaFoldScan project : 3D screening of gut microbiota

- Conclusions and perspectives

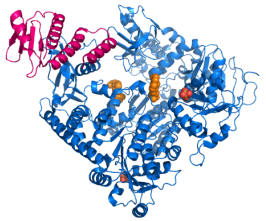
MetaFoldScan : a pluridisciplinary challenge

- 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 !



MetaFoldScan : a pluridisciplinary challenge

- 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 !

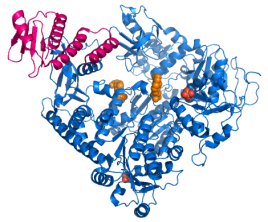


3D Fold

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

MetaFoldScan : a pluridisciplinary challenge

- 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***

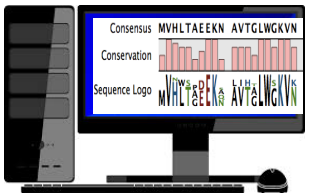
3- Identify hits from 3D Targets

4- Validate hits and pathways involved as pro- or antibiotic **markers**

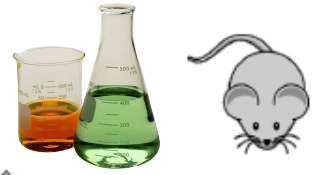
Beyond the description of Metagenomics
MetaFoldScan objectives

Structural screening of Microbial Ecosystems

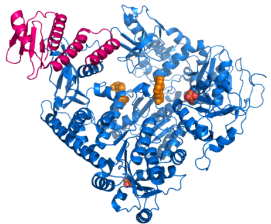
1- Develop & integrate cutting-edge tools in bioinformatics
- Identify Hits from 3D targets within MetaHit
- Increase the functional annotation
- Biologist friendly Web user



Gut microbiota



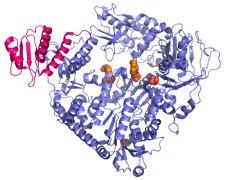
2- Achieve 3D-MetaOmics.
- Intensive modeling of entire proteome
- Distribution, diversity of folds
- Identify proteins with new functions/ fold



3- Validate Hits impacting Health
- Pro and anti-biotic biomarkers
- Universal antibiotic
- Identify homeostasis marker

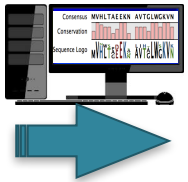
Strategy

Target protein

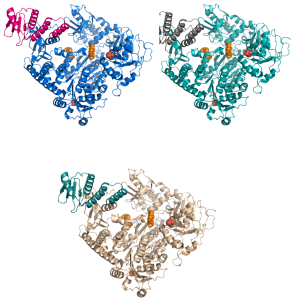


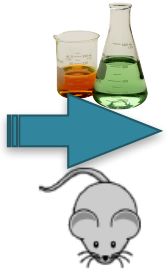
Relevance

Scan microbiota

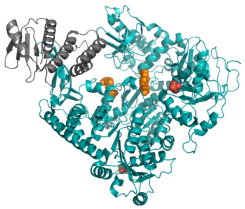


Hits identification





Validation



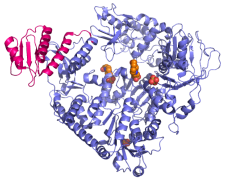
✔

Biomarker

- Biological property
- Therapeutical interest
- Structural challenge

Strategy

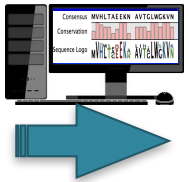
Target protein



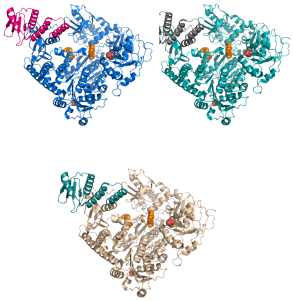
Relevance

Scan microbiota

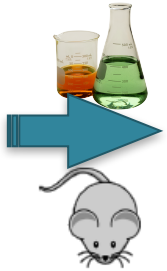
- Biological property
- Therapeutical interest
- Structural challenge

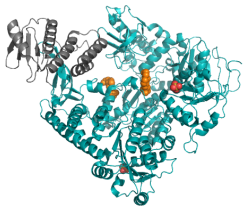


Hits identification



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

Targets of the microbiota

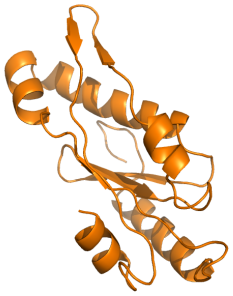
To tackle both methodological and biological issues

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**

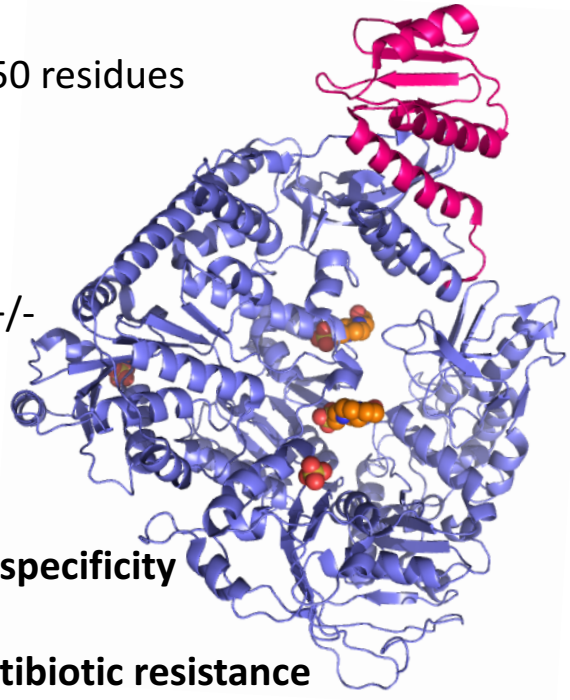


MFD

N. Rama Rao

Mutation Frequency Decline

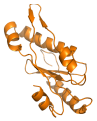
- Involved in bacterial DNA repair, in **virulence**
- **Multi-domains** 1150 residues
- Solved structure
- **Ubiquitous**, Gram+/-
- Promiscuity
- Molecular basis of **specificity**
- Associated with **antibiotic resistance**



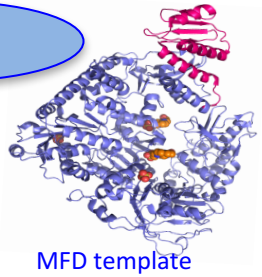
MetaFoldScan Pipeline

Target

3D Fold bank



MAM-template



MFD template

HMM profile of the target bank

Query

Genome x_i | Genome x_j | ... | Genome x_n

Coarse-grained filtering of sequence candidates.

Multiple Sequence Alignment MSA of filtered candidates.
Secondary structure addition

HMM profile of input MSA

Profile (candidate) *vs* profile (bank) Comparison

Scoring & ranking

Score analysis

Abundance classification...
Prevalence in MGS

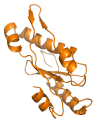
Hits prediction

Crystal structure ...

In vitro/in vivo testing...

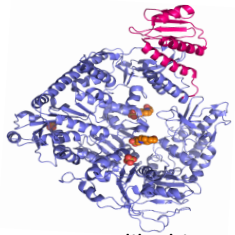
Biomarker ?

✔ ✖



MAM-like hits

+



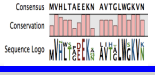
MFD-like hits

Collaborations

Collaborations

MetaFoldScan workflow

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

Experimental validation of targets

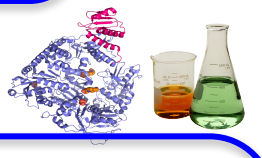
WP2

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

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

Li *et al*, Nat. Biotech. 2014

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

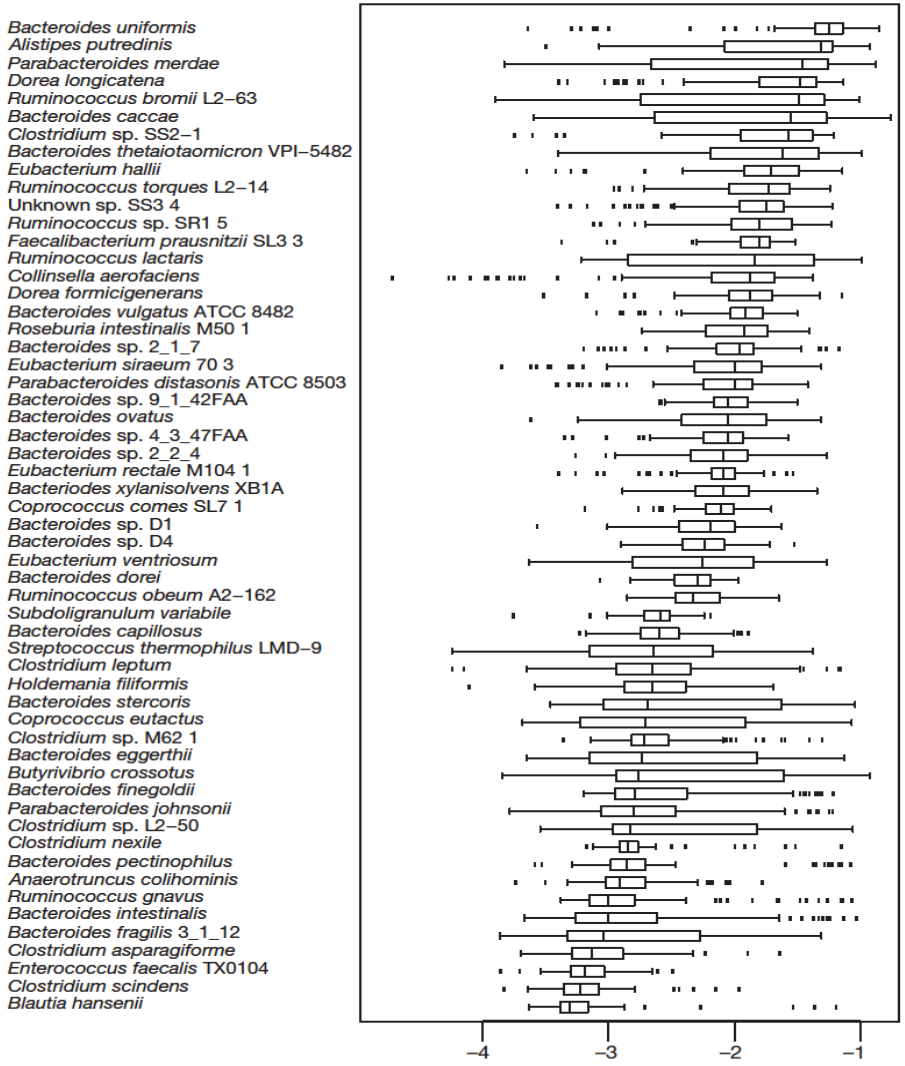
To understand the impact of gut microbes on human health and well-being it is crucial to assess their genetic potential. Here we describe the Illumina-based metagenomic sequencing, assembly and characterization of 3.3 million non-redundant microbial genes, derived from 576.7 gigabases of sequence, from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

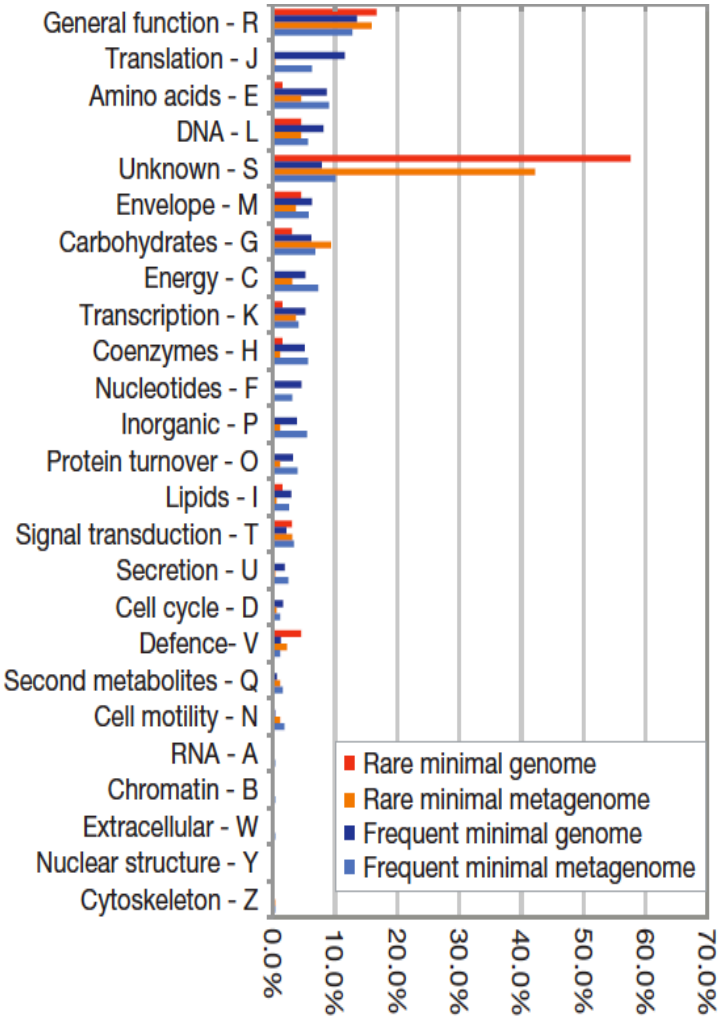
Gut core genome

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



Relative abundance (log_{10})



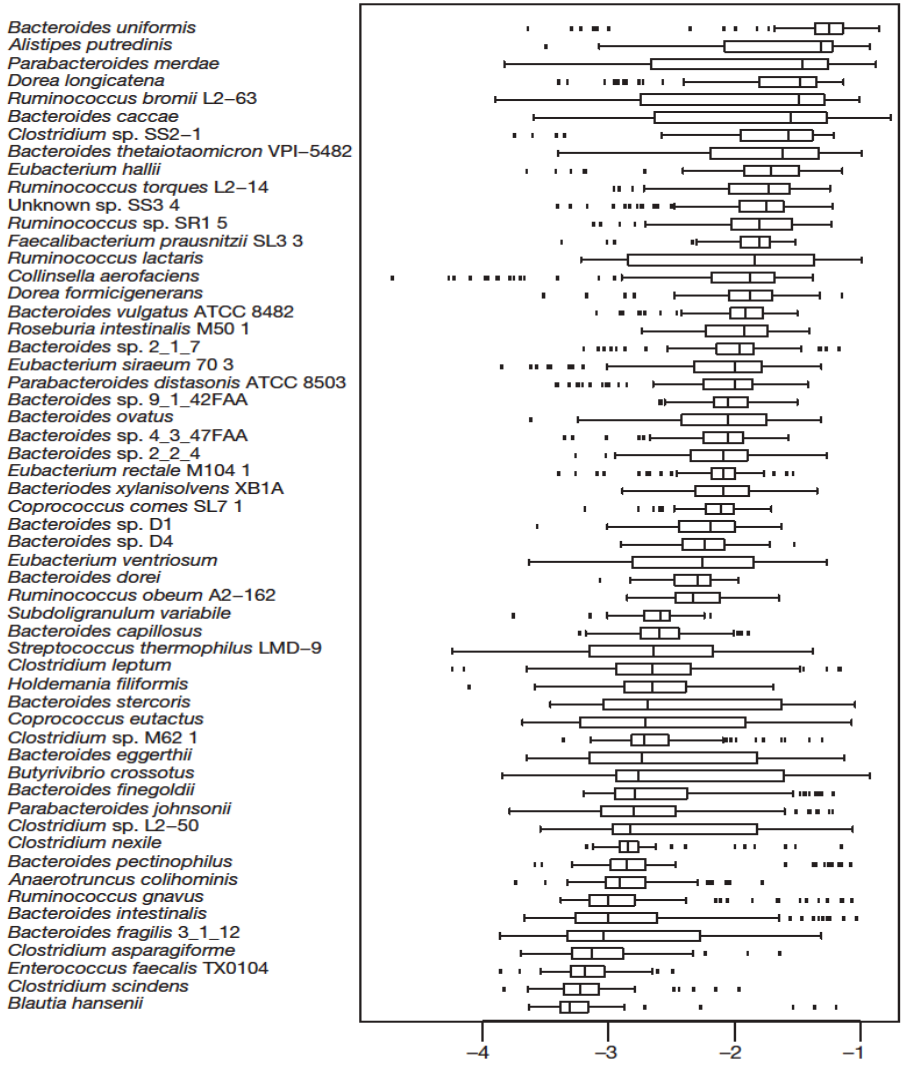
Functional composition

Qin et al, Nature 2010

Gut core genome

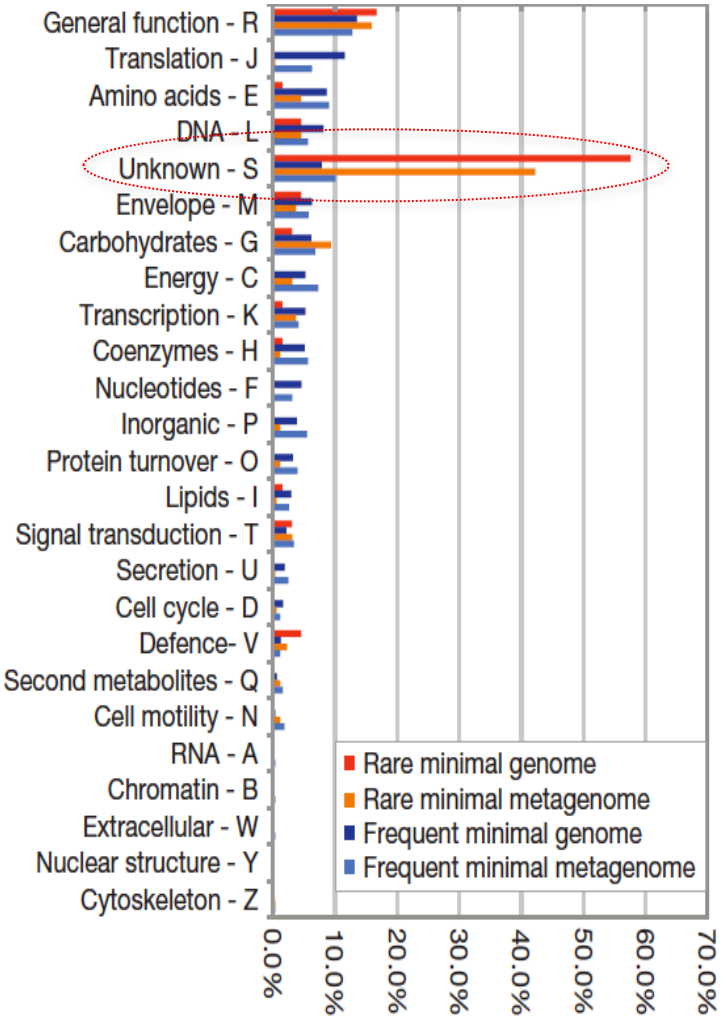
Core genome: 57 bacterial species present in > 90 of individuals (prevalence and consensus)

→ Mostly Bacteroidetes & Firmicutes



Qin et al, Nature 2010

Relative abundance (log_{10})

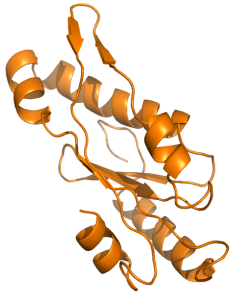


Functional composition

Structural hits of MAM in the core genome

1	**Faecalibacterium_prausnitzii_A2-165 NCBI**	WP_005932151.1	265	9,00E-93	100.00	100	100	135	0	0	1	135	1	135
2			bitscore	evalue	pident	qcovs	qcovhsp	length	mismatch	gapopen				
3	MAM_wt	tr|R6SBT0|R6SBT0_9BACE	23.5	6.4	47.826	17	17	23	10	1	14	36	624	644
4	MAM_wt	tr|D4JK97|D4JK97_9FIRM	28.5	0.013	45.000	15	15	20	11	0	1	20	9	28
5	MAM_wt	tr|R7EPB3|R7EPB3_9BACE	26.6	0.55	41.667	24	24	36	18	1	52	84	185	220
6	MAM_wt	tr|C3QBI3|C3QBI3_9BACE	29.3	0.063	40.625	24	24	32	19	0	88	119	95	126
7	MAM_wt	tr|D4C9H4|D4C9H4_9CLO	27.3	0.30	40.541	27	27	37	21	1	73	108	128	164
8	**Faecalibacterium_prausnitzii_SL3_3**	tr|D4KBR2|D4KBR2_9FIRM	75.9	1.02e-19	35.570	98	98	149	72	4	1	132	1	142
9	MAM_wt	tr|R7B146|R7B146_9BACE	25.0	1.2	34.694	36	36	49	31	1	15	63	119	166
10	MAM_wt	tr|D4L4M8|D4L4M8_9FIRM	24.3	1.9	34.483	21	21	29	19	0	75	103	516	544
11	MAM_wt	tr|C3QPR5|C3QPR5_9BAC	25.8	1.1	34.375	24	24	32	21	0	88	119	95	126
12	MAM_wt	tr|R5TKN8|R5TKN8_9FIRM	26.2	0.50	34.286	26	26	35	22	1	48	82	205	238
13	MAM_wt	tr|C0X1L4|C0X1L4_ENTFL	26.6	0.39	34.211	27	27	38	24	1	14	50	150	187
14	MAM_wt	tr|D1PS54|D1PS54_9FIRM	24.6	2.2	34.043	34	34	47	27	3	44	89	105	148
15	MAM_wt	tr|R6P689|R6P689_9CLOT	24.3	1.8	33.333	27	27	36	24	0	56	91	98	133
16	MAM_wt	tr|A6LD60|A6LD60_PARD8	24.3	4.4	33.333	36	36	57	26	2	31	78	382	435
17	MAM_wt	tr|B0G4V7|B0G4V7_9FIRM	25.4	1.1	32.500	29	29	40	26	1	68	106	150	189
18	MAM_wt	tr|C0D4Q5|C0D4Q5_9FIRM	25.0	3.3	32.394	53	53	71	42	3	16	86	189	253
19	MAM_wt	tr|R7JPH0|R7JPH0_9BACT	24.6	1.3	32.258	23	23	31	21	0	6	36	123	153
20	MAM_wt	tr|E4VUA2|E4VUA2_BACF	27.7	0.35	31.818	33	33	44	30	0	1	44	280	323
21	MAM_wt	tr|A7VJ13|A7VJ13_9CLOT	25.0	1.3	31.507	53	53	73	44	3	2	72	121	189
22	MAM_wt	tr|B0NGV9|B0NGV9_CLOS	26.2	0.68	31.373	38	38	51	33	1	4	54	61	109
23	MAM_wt	tr|Q5M488|Q5M488_STRT2	26.9	0.13	31.111	33	33	45	24	2	36	80	83	120
24	MAM_wt	tr|R5U2R3|R5U2R3_9BACE	28.5	0.12	30.909	41	41	55	36	1	76	130	47	99
25	MAM_wt	tr|A4EBV5|A4EBV5_9ACTN	27.3	0.21	30.612	34	34	49	30	2	57	102	292	339
26	**Clostridium_sp._SS2-1**	tr|B0P145|B0P145_9CLOT	32.7	0.004	30.337	66	66	89	49	5	9	97	1241	1316
27	MAM_wt	tr|D4LSX8|D4LSX8_9FIRM	24.3	2.2	30.303	24	24	33	23	0	21	53	78	110
28	MAM_wt	tr|R7FZ59|R7FZ59_9FIRM	25.0	0.99	30.159	41	41	63	37	1	23	78	160	222
29	MAM_wt	tr|B0PHX3|B0PHX3_9FIRM	23.5	4.7	29.787	34	34	47	31	2	3	48	173	218
30	MAM_wt	tr|R6X7N3|R6X7N3_9PORF	24.3	3.3	29.545	61	61	88	53	4	4	86	547	630
31	MAM_wt	tr|A5Z5Q6|A5Z5Q6_9FIRM	27.7	0.19	29.412	47	47	68	42	2	66	128	345	411
32	MAM_wt	tr|B9YAA0|B9YAA0_9FIRM	28.1	0.15	29.268	30	30	41	29	0	1	41	68	108
33	MAM_wt	tr|B7B871|B7B871_9PORP	25.0	1.9	28.846	32	32	52	24	3	54	96	204	251
34	MAM_wt	tr|P96214|P96214_MYCTU	25.4	1.4	28.571	38	38	56	35	1	52	102	61	116
35	MAM_wt	tr|R6LDT4|R6LDT4_9FIRM	24.3	1.5	28.070	39	39	57	36	1	49	100	17	73
36	MAM_wt	tr|C9L847|C9L847_BLAHA	24.3	1.9	27.869	41	41	61	38	2	5	59	79	139
37	MAM_wt	tr|D4JTH7|D4JTH7_9FIRM	23.9	2.3	27.778	40	40	54	39	0	15	68	1	54
38	MAM_wt	tr|D4M698|D4M698_9FIRM	26.6	0.46	27.692	47	47	65	45	1	72	134	187	251
39	MAM_wt	tr|B6VXT6|B6VXT6_9BACE	26.6	0.72	27.536	44	44	69	41	1	76	135	77	145

all_best_MAM_ident | all_best_MAM_cov | all_best_MAM_bitscore | +



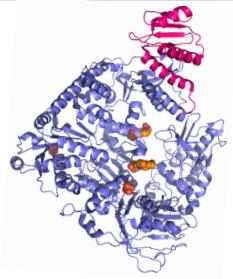
Analysis in progress

- No clear hit except in *Faecalibacterium Prausnitzii SL3_3*
- MAM possibly strictly restricted to *F. Praunitzii*

Structural hits of Mfd in the core genome

			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

all_best_MFD



Analysis in progress

- Mfd ubiquitarious as expected
- 10 best hits to be selected for experimental validation
- Homology modeling to assess the molecular basis of specificity

MetaFoldScan workflow

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

Experimental validation of targets

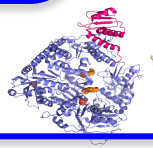
WP2

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

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

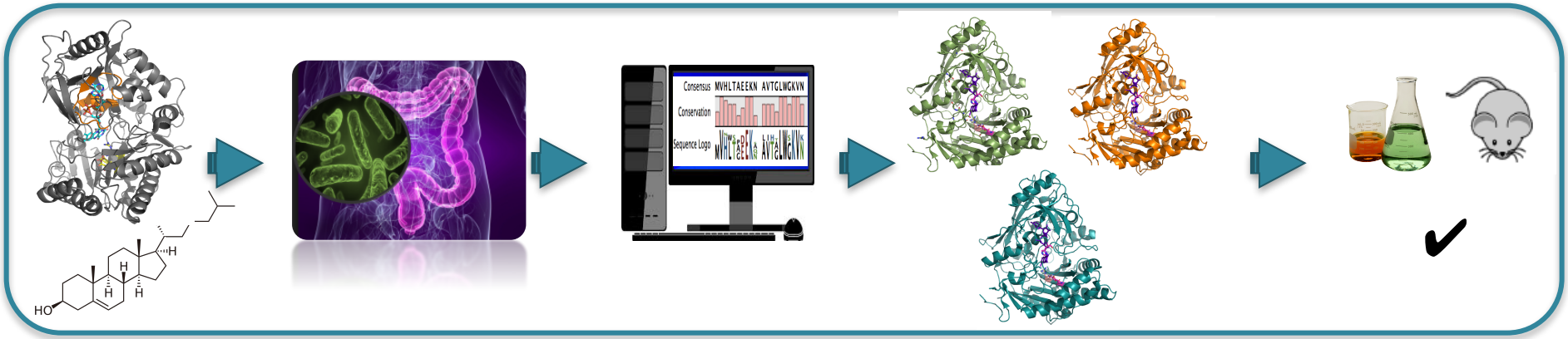
Biologist friendly web-tool interface

V. Loux

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

Li *et al*, Nat. Biotech. 2014

Towards the Structural Screening of Microbial Ecosystems



Outline

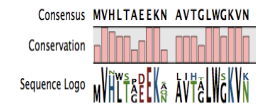
- Cholesterol conversion in the gut microbiota: the mystery enzyme(s) ?

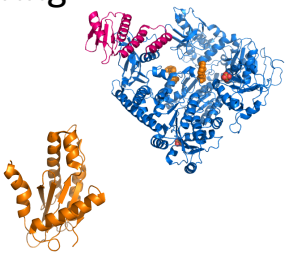
- MetaFoldScan project : 3D screening of the gut microbiota

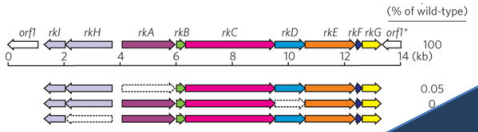
- Conclusions and perspectives

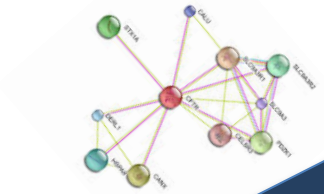
Conclusions and Perspectives

- **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

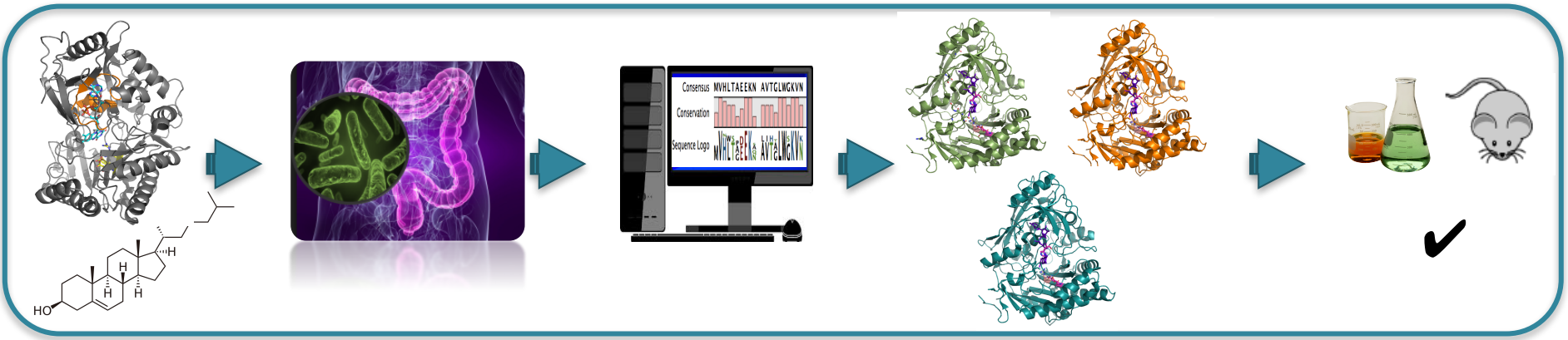
Fabienn Béguet-Crespel

Jean-Marc Chatel

Philippe Gérard

Catherine Juste

Nalini RamaRao



Mariano Martinez

Pedro Alzari

Acknowledgments

To my dear friends…

Andrea Villarino Mabel Berois Maria Natalia Lísa

And you for your kind attention …

INRA SCIENCE & IMPACT

MATHÉMATIQUES ET INFORMATIQUE APPLIQUÉES MaIAGE DU GÉNOME À L'ENVIRONNEMENT

Analysis of the 5 genes from 3D, using Insyght

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, thetaiotaomicron, helcogenes : well conserved (id 60-80% lenght 100%) ; no synteny

=> 1-3/5 in other, less well conserved (id 20-40% lenght 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 neighbhor CDS ; synteny of 14-26 with other B. dorei, 3-11 with other B no synteny with further species

Zoom out analysis using Insyght

Thomas Lacroix

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, taxo_id = 83332 [NC_000962]
  - 1 * cholesterol oxidase :

choD, Rv3409c: cholesterol oxidase
=> Close to oxidoreductase, in conserved synteny with all bacteria cholesterol catabolism except Schizophyllum 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

Zoom out analysis using Insyght

Thomas Lacroix

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

KstR ….TetR family studied in *Mycobacterium tuberculosis*

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 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, 19

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

KstR = TetR type transcriptional regulator
supAB / mce4 + oxydase to look for

e cluster encoding cholesterol catabolism oil actinomycete provides insight into *bacterium tuberculosis* survival in macrophages

er Geize*, Katherine Yam†, Thomas Heuser†, Maarten H. Wilbrink*, Hirofumi Hara†, nderton‡, Edith Sim‡, Lubbert Dijkhuizen*, Julian E. Davies†, William W. Mohn†§, . Eltis†§

licrobiology, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, 9751 NN, Groningen, The artment 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

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

train RHA1, a soil bacterium related to *Mycobacte-*, 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 Δ1-dehydrogenase (KSTD) and 3-ke-
alyze the opening of ring B and aro-
d 3-HSA (3, 7–9). The subsequent
17-dioxo-1,2,3,4,10,19-hexanorandro-
via oxygenolytic cleavage of ring A is
n the testosterone-degrading strain
41 (10, 11). In *Rhodococcus equi*, the
AA is removed via β-oxidation (12).
n steroid catabolism have yet to be
athway enzymes are poorly charac-
volved in degrading the bicycloal-
s C and D. Detailed knowledge of
to engineering strains for the bio-

ealed that rhodococci may be useful
processes: ≈60% of the 3,999 genes
H37Rv are conserved in RHA1,
unction (2). *M. tuberculosis* is the
bacterial infection, killing 2 million
e each year, and extensive drug
TB are now emerging (ref. 13 and
sheets/fs104/en/index.html). One
nycobacterial physiology that con-

MICROBIOLOGY

Microbiology (2010), 156, 1362–1371
DOI 10.1099/mic.0.034538-0

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

Sharon L. Kendall,[1] Philippa Burgess,[1] Ricardo Balhana,[1] Mike Withers,[1] Annemieke ten Bokum,[1] J. Shaun Lott,[2] Chen Gao,[2] Iria Uhia-Castro[3] and Neil G. Stoker[1]

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

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

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

Correspondence
Sharon L. Kendall
skendall@rvc.ac.uk

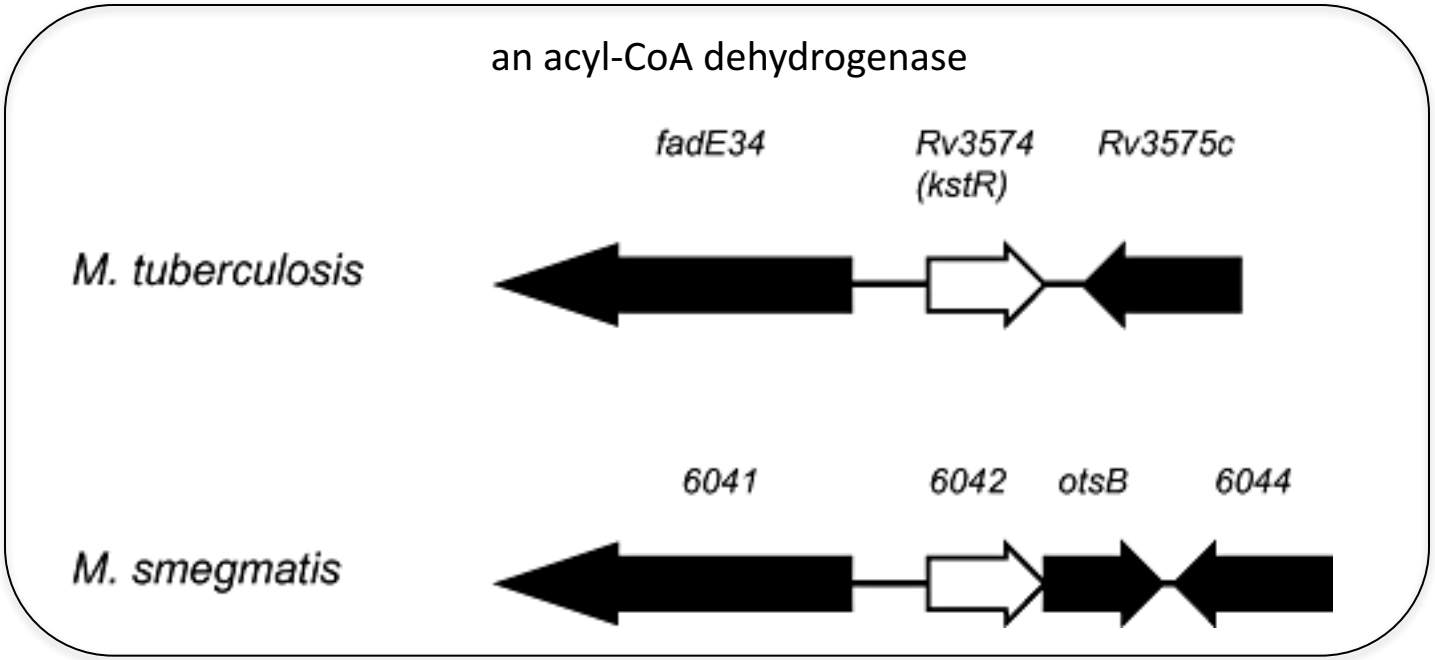
Mycobacterium tuberculosis is able to use a variety of carbon sources *in vivo* and current knowledge suggests that cholesterol is used as a carbon source during infection. The catabolized cholesterol is used both as an energy source (ATP generation) and as a source of precursor molecules for the synthesis 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 ….TetR family studied in *Mycobacterium tuberculosis*

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*.



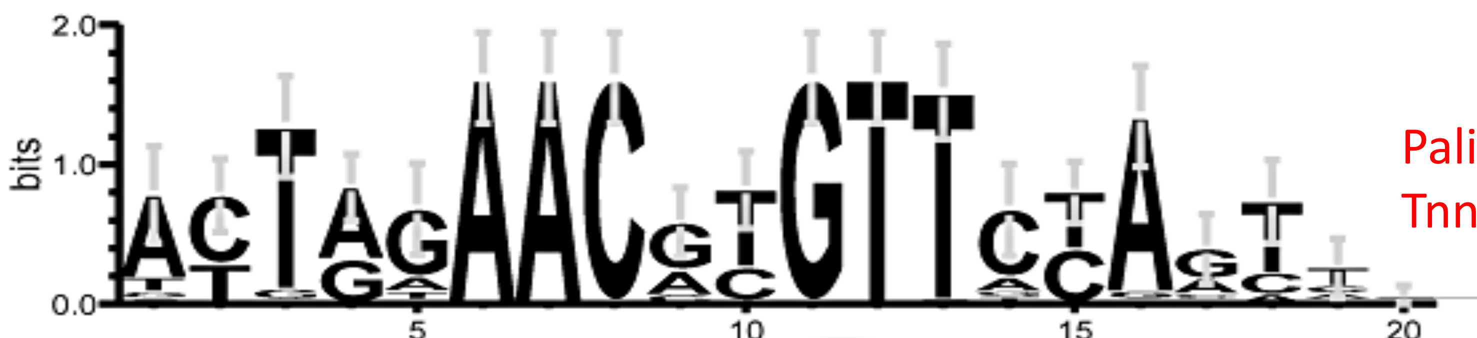
Molecular Microbiology
Volume 65, Issue 3, pages 684-699, 16 JUL 2007 DOI: 10.1111/j.1365-2958.2007.05827.x
http://onlinelibrary.wiley.com/doi/10.1111/j.1365-2958.2007.05827.x/full#f3

Kendall , 2007

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    TCGACTTGACGGTTGATCGTTAGCACTATTAGAACACGTTCTAGTGGGC-AAGAACGCTC 237
MB     TCGACTTGACGGTTGATCGTTAGCACTATTAGAACACGTTCTAGTGGGC-AAGAACGCTC 237
MM     CCAACTTGACG-TTGGTCGTCTGCACTATTAGAACACGTTCTAGTAAGT-ATGAACGGTT 236
MSM    TCGACTTGACGGTCGAACGCTGTCACTATTAGAACACGTTCTAGTGAG--AAGCACGGAC 172
MAP    TCGACTTGACGGTCGATCACTGGCACTATTAGAACACGTTCTAGTCTCCGGTGAGCGAGT 107
NFA    TTGTCTTGACTTCCGCCCGGCAGGGATATTAGAACATGTTCTAG------GTGTGGGAGT 208
           ******    *  *       **********  *******           *    *
```



Sequence logo

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?

In search of TnnAACnnGTTnnA pattern

Véronique Martin

✪ *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.