An *in silico* approach to model the assembly pathway of the respiratory complexes in *Saccharomyces cerevisiae*

Annie Glatigny[1], Viet-Dung Tran[1], Philippe Gambette[2], Marie-Hélène Mucchielli-Giorgi[1,3].
[1]CNRS, Centre de Génétique Moléculaire, UPR3404, Gif-sur-Yvette, F-91198; [2]LIGM, UMR8049, 5 bd Descartes, 77454 Champs-sur-Marne
[3]Université Pierre et Marie Curie, Paris 6, Paris, F-75005

The yeast *Saccharomyces cerevisiae* is a model for the analysis of the complexes of the oxidative phosphorylation chain. While the structure and the catalytic mechanisms of the complexes are well established, their biogenesis is far from understood. An *in silico* study of protein-protein interaction (PPI) networks can help to:

- identify new factors involved in the biogenesis of the complexes
- model the assembly pathway

Our method entailed using the PPI of APID and BioGRID databases, then merging all the physical interactions concerning the subunits and the assembly factors of complex III and constructing the resulting network *via* Cytoscape [1]. To find sub-complexes, the network was divided into highly interconnected sub-graphs with clusterONE, an algorithm that can produce overlapping classes and finally with MCODE that produces denser clusters.

This approach allowed us to identify a protein interacting with the complex III subunits, located within the mitochondria and involved in the biogenesis of the respiratory complexes III and IV [2].

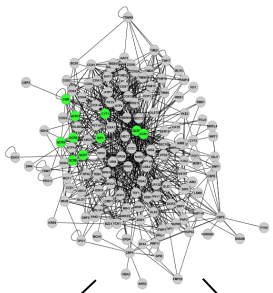
List of proteins involved in *S. cerevisiae* complex III

APID database — Protein Protein Interactions

BioGRID database — Interactions (physical methods)

ID unification (SGD Standard Name)
Looking for the functional environment of complex III proteins in BioGRID and APID
Merging and elimination of redundancy

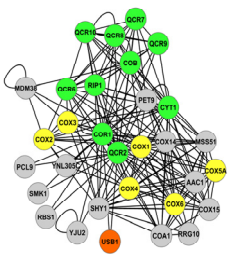
Unweighted PPI Network



1) Identification of factors involved in the complex biogenesis

2) Modeling the assembly pathway

1) Identification of factors involved in complex biogenesis



This approach has allowed us to identify a protein connected with complex III subunits Cor1p and Qcr2p: **Usb1p**.

We then tested its function by genetic and biochemical approaches

1) Usb1p is essential for viability in yeast
2) Ubs1p is located in mitochondria
3) Over-expression of USB1 (cloned on a multicopy plasmid) in wild type and mutants with defects in respiratory complexes biogenesis shows that it can compensate some defects.

Usb1p is involved in the biogenesis of complexes III and IV.

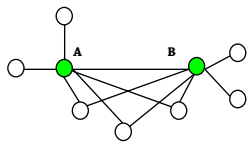
To model the succession of events leading to the formation of the complex, we adapted our methodology because using an un-weighted network results in cliques which cannot be clustered further and does not allow the detection of strongly interconnected sub-networks, corresponding to assembly intermediaries.

We used two methods to resolve this problem; a first consists of using a clustering based on the modularity of the weighted network. The weight of an edge between two nodes is computed with the Jaccard index: the number of nodes interacting with these two nodes divided by the number of nodes interacting with one or other node. The second, consists of a hierarchical clustering of the subunits of the complex, starting from the weights of the edges between subunits.

For complex III, we can find two clusters representing two possible assembly intermediaries. One with Cor1p, Qcr2p, Cyt1p, Qcr6p and the other with Qcr7p, Qcr8p, Qcr10p, Cobp, Rip1p. Moreover, the hierarchical clustering shows the order of assembly of the subunits leading to these clusters. Our results will be compared with models obtained by other methods [3].

2) Modeling the assembly pathway

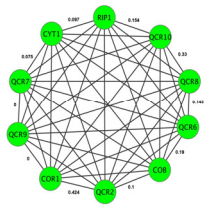
From the unweihgted PPI network we compute the Jaccard index for each link between two proteins of the complex A and B from their neighborhoods N(A) and N(B) (not taking into account the core proteins of the complex). So we obtain a weighted PPI network.



$$w(A,B) = |N(A) \cap N(B)| / |N(A) \cup N(B)| = 3/7$$

Results for the 10 subunits of complex III

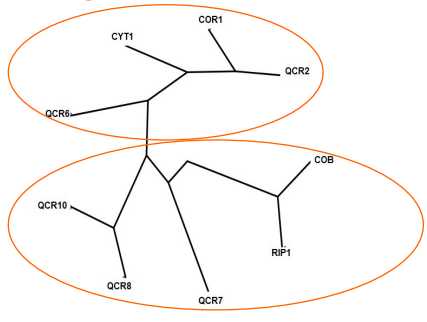
	COB	COR1	CYT1	QCR10	QCR2	QCR6	QCR7	QCR8	QCR9	RIP1
COB	0,000	0,096	0,132	0,182	0,100	0,190	0,125	0,133	0,000	0,250
COR1	0,096	0,000	0,344	0,083	0,424	0,164	0,056	0,077	0,000	0,073
CYT1	0,132	0,344	0,000	0,118	0,356	0,220	0,075	0,105	0,000	0,098
QCR10	0,182	0,083	0,118	0,000	0,064	0,111	0,300	0,333	0,000	0,154
QCR2	0,100	0,424	0,356	0,064	0,000	0,127	0,058	0,059	0,000	0,075
QCR6	0,190	0,164	0,220	0,111	0,127	0,000	0,087	0,143	0,045	0,174
QCR7	0,125	0,056	0,075	0,300	0,058	0,087	0,000	0,130	0,000	0,111
QCR8	0,133	0,077	0,105	0,333	0,059	0,143	0,133	0,000	0,000	0,118
QCR9	0,000	0,000	0,000	0,000	0,000	0,045	0,000	0,000	0,000	0,000
RIP1	0,250	0,073	0,098	0,154	0,075	0,174	0,111	0,118	0,000	0,000



a) Clustering of the weighted PPI network by modularity

With the algorithm TFitW provided by A. Guénoche we partition the weighted PPI network. We obtain 2 clusters:
Cluster 1: Cor1p, Qcr2p, Cyt1p, Qcr6p
Cluster 2: Cobp, Rip1p, Qcr7p, Qcr8p, Qcr8p
Qcr9p has not been included because experimental PPI results are missing for this protein.

b) Hierarchical clustering



Both methods suggest the existence of 2 assembly modules. Biological experiments are needed to confirm these results.

[1] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. (2003) *Genome Research* 13(11):2498-504
[2] Glatigny A, Mathieu L, Herbert CJ, Dujardin G, Meunier B, Mucchielli-Giorgi MH. (2011) *BMC Syst Biol*, 5 (1) 173.
[3] Zara V, Conte L, Trumpower BL. (2009) *Biochim Biophys Acta.* 1793 (1):89-96.