Short abstract:

We observed a similar conservation rate of experimentally verified and predicted binding sites for four yeast transcriptional regulators. We identified a candidate transporter for alpha-isopropylmalate, an intermediate in the leucine biosynthesis pathway, using analysis of candidate regulatory sites, microarray data and protein-DNA interaction.

Long abstract:

Binding sites of eukaryotic transcription factor are usually short and localized in long regulatory regions, compared to prokaryotes. Thus eukaryotic candidate binding sites can not be identified by standard approaches.

We investigated the conservation rates for four transcription regulators/regulatory complexes of the yeast *Saccharomyces cerevisiae*: Gcn4p, the global regulator of the amino acids biosynthesis; Met31/Met32 and Cbf1/Met4/Met28, the complexes regulating the methionine biosynthesis; and Leu3p, the pathway-specific regulator of the leucine biosynthesis. All analyzed binding sites for these regulators were divided into three groups: “known” (experimentally verified sites), “strong predicted sites” (sites scoring higher of at least as highly as the majority of “known” sites), and “weak predicted sites” (scoring lower the majority of experimentally verified sites, but higher then the threshold, the minimal score observed for experimentally verified sites).

Analysis of conservation rates in each group of sites for each regulator lead to two conclusions: 1) conservation of binding sites in more closely related genomes is much higher then in more distant ones; 2) conservation rates for known and strong predicted binding sites are similar for all studied regulators except the Met31/Met32 complex where it is much lower. We see no obvious explanation for such sharp difference, as the molecular function of this regulatory complex is still unclear. These observations make it possible to predict candidate binding sites for eukaryotic transcriptional regulators based on their conservation rates.

Systematic analysis of conserved (high-scoring) sites lead to identification of a candidate transporter for alpha-isopropylmalate, an intermediate of leucine biosynthesis, YOR271c. this function is necessary, but no gene encoding such protein has been identified yet. The base for this prediction is the following:

1. The protein product of this gene is known to localize in the mitochondrion. It contains several predicted transmembrane segments. Thus, this protein is localized in the mitochondrial membrane. The outer mitochondrial membrane is permeable to small metabolites, whereas the permeability of the inner membrane is tightly controlled. We predict that the YOR271cp protein resides in the inner mitochondrial membrane.

2. A high-scoring candidate binding site for the specific transcriptional regulator, Leu3p, is identical to the consensus and is strongly conserved in all studied genomes.
3. Protein-DNA ChIP data show that the upstream region of the YOR271c gene is bound by Leu3p along with other genes involved in leucine biosynthesis. Half of the genes of the leucine pathway had sites bound by Leu3p in this experiment. Among the experimentally identified genes with unknown molecular function only two contained potential binding sites for Leu3p: YOR271c and YDL228c. The predicted site upstream of the YDL228c gene was weaker and was not conserved even in the closest genomes. Summarizing all these data we see that the product of YOR271c gene might be involved in the leucine biosynthesis, localized in the inner mitochondrial membrane, and it seems to be the only candidate for the alpha-isopropylmalate transporter — with such properties.
4. The closest homologs of the predicted transporter are tricarboxylate transporter of *Rattus norvegicus* and sideroflexins. The literature data on sideroflexins are extremely conflicting. Currently, these proteins are usually characterized as potential tricarboxylate/iron transporters, which is consistent with the alpha-isopropylmalate carrier function.

To validate this prediction we considered related genomes. Among fungi, leucine biosynthesis has been extensively studied only in *Aspergillus nidulans* and *Neurospora crassa*. However, there is a significant difference in the alpha-isopropylmalate synthesis system in yeasts and higher fungi. Alpha-isopropylmalate synthases of *N.crassa* and *A.nidulans* are localized in the cytosole and, thus, these fungi do not need a mitochondrial alpha-isopropylmalate transporter at all. Consistent with that, these genomes do not contain orthologs of YOR271c. However, this is a very indirect evidence and the suggested functional assignment remains hypothetical, although we believe that it is sufficiently strong to warrant experimental verification.