MDIBL talk July 14, 2005

The Evolution of Cytochrome P450 in animals.

Title slide (1) Tree of life 1891 Ernst Haeckel, Title on left

My opening slide is a collage (2) containing 35 eukaryotic species with a genome project, either completed or underway. Since I made this slide, many more species have been added to the list. The genomes we have now and the ones in the pipeline provide a treasure trove of data for comparative genomics. However, the genomes must be annotated and the information extracted. It has been my goal to cover the cytochrome P450 gene superfamily in great detail, for these organisms and those that do not have a genome project.

The information I gather is posted on my website and can be accessed by clicking on entries in the table shown here (3). The top of the table has links to 24 animal pages. The lower portion covers plants , lower eukaryotes and bacteria. I am constantly adding new data to these pages. Notice that Hydra was added in May, opossum in June and sponge was added Monday. The rat P450 collection was revised in Feb and is nearly complete. The P450 site receives between 1.2 and 1.4 million hits per year. You can find the page by doing a Google search for cytochrome P450.

This graph (4) shows the history of P450 sequencing, starting in 1982 with 3 sequences (CYP2B1 and 2B2 from rat, and CYP101A1 from *Pseudomonas putida*). Sequencing takes off in 1996 after the yeast genome was completed. The main eukaryotic genomes released are listed above the line. I cannot put all the genomes on this graph since there is not enough room. The total named P450 count is now greater than 4500 and about 35% are from animals, which is our focus today.

Cytochrome P450s perform many biosynthetic and degradative reactions . The evolution of biochemical defense mechanisms applies especially to the degradation of drugs and xenobiotics. Fred Guengerich (5) in a recent review of human CYPs states that three human P450s: CYP3A4, CYP2D6 and CYP2C9 metabolize about 75% of drugs and that inclusion of 2C19, 2E1, 1A2 and 2A6 will account for

90-95% of all drug metabolism. To understand the evolution of drug metabolism we have to study the evolution of these genes.

Before we look in some detail at evolution of mammalian genome CYPs, I will take you on a brief tour of animal P450 evolution. (6) The set of animals shown here would provide wide but not deep coverage of animal evolution. In the upper left is a sponge currently being sequenced at the Joint Genome Institute in California. The sequence data became public July 5 in the NCBI Trace Archive. The sponge represents the simplest animal. Hydra is a radial animal on the next branch in animal evolution. This branch includes jellyfish, corals and sea anemones. The rest of these are bilateral animals. They sort into two main groups, the protostomes, including the arthropods, many worms, and molluscs The second group is the deuterostomes that includes chordates and echinoderms. This is the juvenile form of *Ciona intestinalis*, a sea squirt or urochordate whose genome has already been published. Amphioxus is a cephalochordate and sister group to all vertebrates, represented here by zebrafish and opossum. The opossum is the first marsupial genome sequenced.

P450 or CYP genes can be sorted into a small number of gene clades called clans. These clans represent deep branching events in P450 evolution. There are 9 clans in mammals (7) but only five clans are found in protostomes, and only four clans are seen in arthropods and nematode worms. These last two groups in the ecdysozoans, or molting animals have lost the CYP20 clan seen in annelid worms and all other animals. The ecdysozoans have also lost CYP51, the sterol 14 alpha demethylase. The four conserved clans shown here (8) are the CYP2, CYP3, CYP4 and mitochondrial clans.

CYP51 is often lost, since its role in the synthesis of sterols can be replaced by eating sterols in the diet. However, CYP51 is found in the sponge genome, fungi, plants, protists and some bacteria. It was present in the last common animal ancestor. CYP20 is also seen in most animals(9), including sponges, hydra, leeches, sea urchins and all chordates. It's loss in ecdysozoans may have important consequences in the evolution of this group of animals.

Since CYP51 and CYP20 were present at the beginning, probably in a choanoflagellate ancestor(10), the question then becomes, when did the

four clans CYP2, 3, 4 and mitochondrial arise. We can address that question by examining the hydra genome. The hydra genome is not assembled yet, but nearly all the exons can be found and a gene count can be made. I estimate there are 31 CYP genes in Hydra(11). They include one CYP20, 13 CYP2 clan genes and 17 CYP4 clan genes. There are no CYP3 clan genes or no mitochondrial CYPs. This immediately suggests that the CYP3 clan and the mitochondrial clan arose in the bilateral animals after they diverged from the radial animals. One would also predict that the sponge genome would also be missing CYP3 clan and mitochondrial clan sequences.

Imagine my surprise then when I looked at the first sequence data from sponge this past weekend. There were clear mitochondrial and CYP3 clan sequences present along with CYP20, CYP51 and CYP2 and CYP4 clan members. This did not make sense. Did the hydra lose two clans of CYPs? Did the sponge genome have contamination with a bilateral animal genome? It will take some additional genomes to clear this up.

What happens at the chordate end of the tree?
Ciona (the sea squirt) is a key organism since it branched off early in chordate evolution, before steroid and vitamin D receptors evolved. Therefore, Ciona should be lacking the P450s required for steroid and vitamin D metabolism. This slide (12) shows a tree of Ciona intestinalis CYPs along with some from Ciona savigny. Most are in the CYP2 clan. There are 7 clans present. The CYP26 clan (retinoic acid metabolism) and CYP7 (7 alpha hydroxylation) are new. The CYP19 Clan and CYP51 are missing.

Amphioxus is the next organism after Ciona with a genome sequence available. Biochemical assays show that amphioxus has estrogen, therefore, it must have this pathway (13) to make estrogen. That includes CYP11A, CYP17 and CYP19. This tree of amphioxus CYPs (14) shows that CYP17 and CYP19 are clearly present, but these sequences are not very sequence similar to their vertebrate orthologs. CYP19 is 37-38% identical to zebrafish CYP19s. CYP17 is 35% identical to Xenopus CYP17 and 36% identical to CYP1A6 and CYP1A7 from Xenopus, so these early CYP17 and CYP19 sequences are barely recognizable as belonging in these families.

Also present in amphioxus but missing in Ciona are CYP51, CYP39 and CYP46, both involved in cholesterol metabolism in vertebrate brain. Results from David Russell's lab presented in Dallas showed CYP46 was required for certain memory and learning tasks in knockout mice. There are two CYP7 sequences but no CYP8, so prostacyclin synthase may be absent in amphioxus. CYP20 is here and there are two CYP26 sequences. There is no obvious CYP1 sequence in amphioxus. Instead the CYP17 sequence is the best match to CYP1. This suggests that the true CYP1s we know from vertebrates evolved after amphioxus developed the first CYP17 sequence and steroid biosynthesis.

This tree does not include the mitochondrial clan which includes CYP11, CYP24 and CYP27 sequences. These enzymes are essential for the first step in steroid biosynthesis and for vitamin D metabolism. (15). A vitamin D receptor has been cloned from lamprey, but it was not found in Ciona. Amphioxus may be the point of origin of these critical signaling pathways that control calcium homeostasis and calcification. Four genes of the amphioxus mitochondrial clan are shown here (16) with representative sequences of vertebrates. Two sequences cluster with CYP24s. One sorts with the CYP11s, and one with CYP27s. These CYPs probably include the first side chain cleavage enzyme and early vitamin D metabolizing enzymes.

The rat, mouse and human genomes.

Having more than one mammalian genome is essential for understanding how our genome has evolved. My colleagues Dan Nebert, Darryl Zeldin, Susan Hoffman and two genome nomenclature experts Hester Wain and Lois Maltais and I published a comprehensive comparison of mouse and human P450s last year in Pharmacogenetics. This is a tree of the mouse and human P450s, with the ortholog pairs in red(17). These results showed an expansion of the mouse P450s in seven gene clusters. I will now show these seven gene clusters including the rat CYPs. In general, the rat is intermediate in P450 cluster size and complexity. The mouse has gone to extremes in amplifying the P450 genes in these clusters. These figures emphasize the point that a rodent is not a human especially when it comes to drug metabolism.

We will begin with the most straight forward clusters and move to more complex clusters. The CYP4F cluster (18) has complete conservation of orthology between mouse and rat. In all of these figures, each dot equals

one exon. In human, the only clear ortholog is CYP4F22. This suggests that 4F39 was the ancestral sequence in this cluster and its function probably is the original function. There is only one 4F sequence in fish, so it is probable that the ancestor of humans and rodents had only one 4F that has diverged independently for 75 million years. Of course, the new sequences may take on new functions.

(19) This is the Cyp2d locus. There is only one CYP2D6 functional gene in humans, though a report has appeared showing a functional polymorphic CYP2D7 in human brain. The mouse has 9 Cyp2d sequences, while the rat has 5. The red lines represent orthologous genes or groups of orthologous genes. Cyp2d22 and CYP2D4 are most similar to CYP2D6 and these may represent the common ancestral sequence. If this is correct, then there is a pattern emerging that an outside member of a cluster is the ancestor of the cluster, like we saw in the 4F cluster. The cow has two CYP2D genes 94% identical so this is a recent duplication and the ancestor of cow and human probably had only one CYP2D. Opossum has only one CYP2D.

(20) This is the Cyp2j locus. There is one human CYP2J2 and 8 mouse Cyp2js. There are 5 2Js in the rat. Once again there is strong orthology between mouse and rat, with some expansion in the mouse. The 2js in rodents don't have a clear ortholog to human. The cow genome has 5 CYP2Js. There seems to be selective pressure to expand the 2J cluster.

The next slide (21) shows the CYP2ABFGST cluster. This is the most complex P450 cluster in mammals with six subfamilies present. All six subfamilies were in the common ancestor, so this may be the oldest P450 gene cluster in mammals. All three species have CYP2S1 one one side and CYP2T on the opposite side. This continues the pattern that the edges of clusters are well conserved. The mouse and rat have strong similarity with clear orthologs all across the cluster (shown by the red lines). The human cluster seems to have a mirror symmetry indicating a possible inverted duplication event that makes it different from the rodent pattern. The opossum cluster has only one member of each subfamily except for 2 CYP2Bs. The opossum gene order and orientation is preserved.

The CYP2C cluster (22) is small in humans (only 4 genes) very large in mice (15 genes) and slightly smaller in rats (12 genes). The orthology

across the cluster is broken by an inversion shown by the crossed red lines. There is no obvious orthology to the human genes. Note that on the right side there is a 4.1Mb space between 2C22 and 2C23. This is conserved in mouse. Also interesting is the orthology between male specific 2C11 and a pseudogene 2c52-ps in mouse, which has one stop codon and two frameshifts. This must be a very recent pseudogene to retain 82% sequence identity. The genes flanking this CYP cluster are the same in all three species.

The 4ABXZ cluster (23) has some interesting features. There is no 4Z gene in the rodents. In fact, this gene is only seen in humans and chimpanzees, making it a rare hominid specific gene. The human and chimp clusters are highly similar. Three of the four pseudogenes in the middle are 100% identical and the fourth has only 1 aa difference. This preservation has lasted for 6 million years. Oddly, there are expansions in the human that are outside the cluster boundaries and we have not seen that before. Opossum has one 4A gene, 2 4B genes and 2 4X-like genes. The 4ABX cluster seems to arise from the fish 4T sequence. CYP4B is most like CYP4T and probably is ancestral to 4A and 4A is ancestral to 4X.

The CYP3A cluster (24) is the last of the seven CYP clusters. The CYP3A9 gene is the ortholog of Cyp3a13 in mouse. These genes are outside the main cluster by 1 million bp in mouse and 7.5 million bp in rat. 3A1 and 3A23 are now considered to be the same rat gene. There is expansion in the mouse and some duplications are moved to new locations as with the 3a57 gene. The human genes do not have clear orthologs to the rodent genes, suggesting a single ancestral sequence evolved independently in human and rodent lines.

These figures show that rapid change in gene cluster size and organization can permit the development of new xenobiotic and drug metabolism profiles within mammals. They also warn us that drug metabolism studies may not be extrapolated with confidence from rodents to humans. Finally, let us remember (25) that what we have said about these P450s is limited to this single branch on the tree of life and that the true breadth and depth of cytochrome P450 diversity has yet to be explored.