Evolution of three multi-gene families (GSTs, CCEs & CYPs) in primates

Jade Westlake

Ollscoil na Gaillimhe
University of Galway

Introduction

The idea of evolutionary divergence from a shared ancestor aligns with the concept of clusters of orthologous protein groups, suggesting that most of today's genes are likely derived from a "core" set of 7,000-12,000 genes that existed more than 500 mya. It is energetically more efficient for new genes to evolve through duplication and crossover events than being created from scratch. These processes led to the formation of gene families with new genes sharing homology with their ancestral genes (Nebert amd Wain, 2003). Enzymes belonging to the cytochrome P450 (CYP), glutathione S-transferase (GST), and carboxylesterase (CCE) families, are found across a wide range of species, from bacteria to plants, fungi, and animals & are estimated to have originated 1.5-2 bya. These gene families play critical roles in the detoxification of endogenous compounds and foreign substances including drugs, toxins and pharmaceuticals.

Pipeline

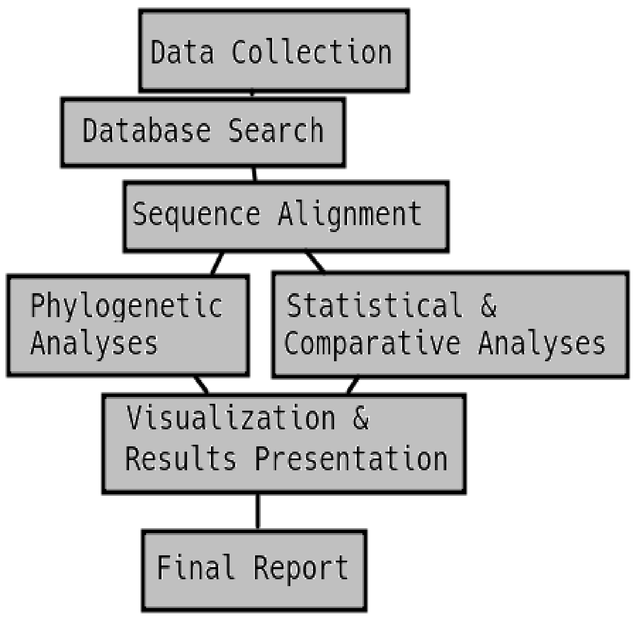
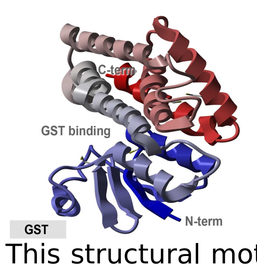


Figure 1. Draft of Pipeline for Evolutionary analysis pipeline of the GST, CCE, and CYP gene families in primates. Data Collection: research relevant material. Database search: search NCBI/ Ensembl for a variety of primate species. Allignment: Use SeaView or alternative MSA methods to prepare for phylogenetic analysis. Data Preparation: involving checking for duplicates, verifying correct allignment, removal of stop codons & flagging relevant genes. Phylogenetic comparative methods to assess selection pressures and evolutionary trends. Visualization: Phylogenetic trees, heatmaps, and statistical plots to visualize gene family evolution across primates.

GST (Glutathione S-transferase)



The complete GST gene family consists of 16 genes in six subfamilies: alpha (GSTA), mu (GSTM), omega (GSTO), pi (GSTP), theta (GSTT) and zeta (GSTZ) with evidence of functional polymorphic variation (Nebert and Vasiliou, 2004, Townsend and Tew, 2003). All GST isozymes share a similar topology, consisting of two domains. The N-terminal domain, which makes up about one-third of the protein, forms the G site and consists of four β-sheets flanked by three α-helices.

This structural motif is common to thioredoxin and other proteins that bind GSH or cysteine. It also contains a catalytically essential residue, such as tyrosine, serine, or cysteine, which directly interacts with the thiol group of GSH. The C-terminal domain is predominantly α-helical, and along with a loop from the N-terminal domain, it forms the H site. Variations in the H site amino acids determine the enzyme's substrate specificity. Structural differences in the C-terminal domain are observed in the Alpha, Theta, and Mu classes, with the Alpha and Theta classes featuring an additional C-terminal α-helix, while the Mu class has an extra loop (Townsend and Tew, 2003).

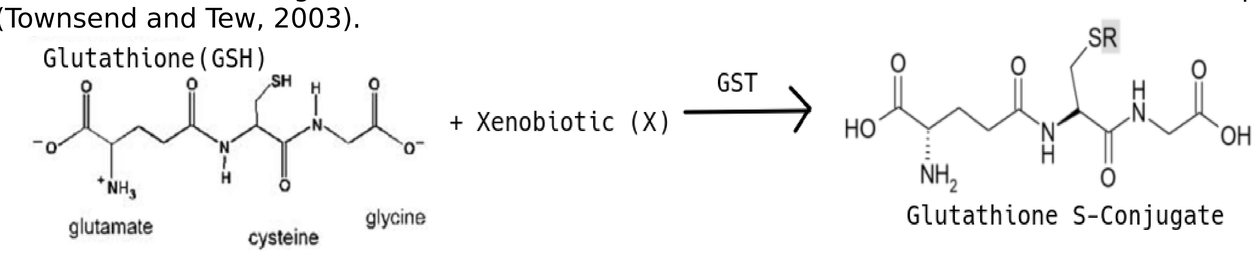
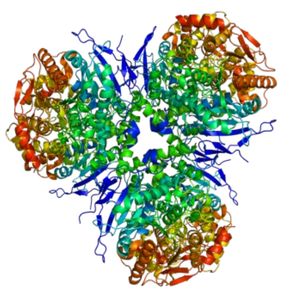


Figure 2. Illustration of how GSTs function in detoxification. GSTs are a family of phase II detoxification enzymes that catalyze the conjugation of glutathione (GSH), a tripeptide composed of glutamate, cysteine, and glycine, to various xenobiotics (foreign compounds) and endogenous substrates. GST attaches the thiol group (-SH) from glutathione to the xenobiotic resulting in a conjugate: Glutathione S-conjugate that is usually more hydrophilic (water-soluble) than the original xenobiotic, facilitating its excretion. GSTs work by lowering the pKa of the sulfhydryl solution to around 6.5 when GSH binds to the active site. This allows GSH to exist as the thiolate (GS^{2-}) anion at neutral pH. Catalysis by GST occurs through the enzyme's ability to facilitate GS^{2-} formation and bind hydrophobic electrophilic compounds at separate G- and H-sites. For certain substrates, such as benzyl and phenethyl isothiocyanates and alkyl dihalides, GSTs can catalyze both forward and reverse reactions, potentially increasing toxicity rather than detoxifying. The active enzyme exists as a dimer of two subunits, further supporting its essential role in cellular detoxification (Nebert and Vasiliou, 2004)

CCE (Carboxylesterase)



CCE enzymes hydrolyze compounds containing ester, amide, or thioester linkages, playing roles in prodrug bioconversion, lipid mobilization, and protein trafficking. The family consists of six isoforms (CES1 to CES6), with CES1 and CES2 being key for drug metabolism. CES1 has therapeutic potential for treating organophosphate poisoning and acute cocaine overdose, and its inhibitors may treat hypertriglyceridemia, obesity, and atherosclerosis. CES1 and CES2 have undergone extensive gene duplication in species with omnivorous and herbivorous diets, while CES3, CES4, and CES5 remain with one gene per species, reflecting their more specialized roles in xenobiotic metabolism (Di, 2019; Williams et al., 2010).

CES enzymes are membrane-bound, primarily localized in the endoplasmic reticulum (ER), and feature an N-terminal signaling peptide for ER targeting and a C-terminal retention sequence. CES1 has three binding sites: active, side door, and Z-side, with a large, flexible active site that accommodates diverse substrates (Williams et al., 2010).

The hydrolysis mechanism involves a catalytic triad of glutamate, histidine, and serine, with the serine alcohol attacking the ester substrate's carbonyl carbon. This forms a tetrahedral intermediate, stabilized by the oxyanion hole, leading to product release. This NADPH-independent process does not require cofactors and can also catalyze transesterification, crucial for cholesterol homeostasis (Di, 2019).

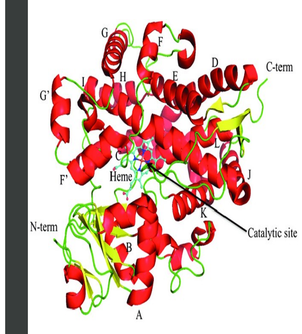
A Concern

GSTs are upregulated in response to oxidative stress and are often overexpressed in tumors, where they neutralize The drug by conjugating them with glutathione, reducing its efficacy. Genetic variations within human GST isozymes may influence both cancer susceptibility and treatment outcomes. Drugs that are not substrates for GST-catalyzed thioether bond formation, still select for resistance suggesting dual functionality where in addition to their role in drug metabolism, they can regulate mitogen-activated protein kinases (MAPKs), e.g, GSTM1 regulates the heat shock-sensing pathway by inhibiting ASK1, a kinase involved in apoptosis. Under stress, dissociation of the GSTM1:ASK1 complex activates ASK1, but forced GSTM1 expression can block this activation, reducing apoptosis. Similarly, GSTP1 regulates JNK, a protein involved in apoptosis. Under oxidative stress, dissociation of the GSTP1:JNK complex activates JNK, inducing apoptosis. Inhibition of GSTP1 enhances JNK activity, promoting cell death. Thus, GST overexpression can limit drug efficacy by influencing apoptosis pathways in addition to drug metabolism (Townsend and Tew, 2003).

CCEs overexpression can lead to the breakdown of chemotherapy agents or reduced activation of prodrugs, thus lowering their therapeutic effectiveness (Di, 2019). Their ability to convert prodrugs into active forms or deactivate drugs complicates cancer treatment and drug response. Notably, CES1 has shown promise as a target for improving drug pharmacokinetics and efficacy.

CYP enzymes play a significant role in both activating and deactivating drugs. Certain CYPs, like CYP2B6, & CYP3A4, can activate prodrugs, such as cyclophosphamide, enhancing their pharmacological activity. However, CYPs can also decrease the effectiveness of chemotherapeutics by altering their pharmacokinetics. This can occur through decreased bioavailability and increased elimination rates, especially in individuals with high CYP expression or due to drug-drug interactions that induce CYP enzymes. Additionally, intratumoral CYP expression can lead to localized drug resistance by deactivating drugs directly within tumors. Drugs like docetaxel, paclitaxel, and vincristine are particularly susceptible to this intratumoral metabolic resistance, as they are metabolized by CYP enzymes which significantly reduce their pharmacological activity (Ingelman-Sundberg and Lauschke, 2020).

CYP (Cytochrome P450)



The CYP enzymes are a large diverse family of heme-containing enzymes that play a role in the metabolism of a wide range of substances, Their name is due to how they are bound to membranes within a cell (cyto) & contain a heme pigment (chrome and P) that absorbs light at a wavelength of 450 nm when exposed to carbon monoxide. CYPs are designated into families/subfamilies, based on % amino acid sequence identity. they are assigned to a family designated by an Arabic numeral when they share >=40% identity, & assigned to a subfamily by a letter when they share >= 55% (Nebert, Wikvall and Miller, 2013)

There are >50 CYP450 enzymes, but the CYP1A2, CYP2C9, CYP2C19, CYP2D6, CYP3A4, and CYP3A5 enzymes metabolize 90% of drugs & are predominantly expressed in the liver. In mammals there are 18 families encoding 57 genes in humans. These families notibly contain more genes than the other 15- corresponds to how mammals such as humans are more exposured to chemicals & drugs. It is also notable that genes from CYP1-4 are inducable by many stimuli however the other gene families are never really induced- suggesting a conserved role in critical life functions & thus, are more associated with disease (Nebert, Wikvall and Miller, 2013).. Key subfamilies include CYP1, CYP2, CYP3, CYP4, CYP5, CYP6, CYP7, CYP8, CYP11, CYP17, CYP19, CYP21, CYP24, CYP27, CYP39, CYP46, and CYP51.

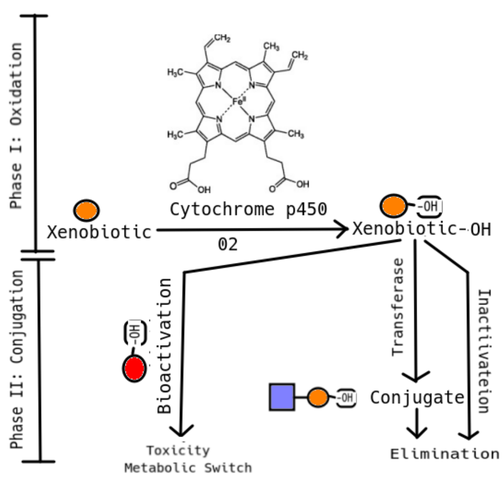


Figure 3. CYP450 enzymes catalyze oxidation reactions on xenobiotics and endogenous compounds, adding an oxygen atom to form a hydroxyl group (-OH), increasing the compound's polarity. The heme group facilitates this transfer. In Phase II, transferase increases the polarity, making the compound more hydrophilic for easier excretion via urine or bile.

Conclusions

Studying the evolution of these three gene families provides insight into how they have adapted over time to interact with a wide range of xenobiotics, how variations in these genes have influenced their functional roles in different species, how variants may impact responses to treatment and how they have contributed to the development of drug resistance mechanisms. further exploration of these gene families could lead to more personalized and targeted therapeutic strategies, allowing for more effective drug development and individualized treatment plans based on specific gene variants