Introduction

Proteins that contain metal ion cofactor(s) are known as metalloproteins, and they compose approximately 40% of all proteins. The metal ion can be free, or bound within a ligand. One such ligand is heme, a complex of iron and a porphyrin ring (**should I go deeper about porphyrins?**). Heme is employed by many metalloproteins to catalyze a broad range of reactions.

Hemoglobin and myoglboin are likely familiar proteins, and use heme to store and transport oxygen. Other examples of enzymes that employ heme are peroxidases, catalases (a type of peroxidase), nitric oxide synthases, heme oxygenases, and cytochrome p450s. Peroxidases and catalases catalyze oxidation-reduction reactions using a histidine-bound heme, with catalases in particular specializing in the decomposition of hydrogen peroxide. Nitric oxide synthases catalyze the reaction of L-arginine into nitric oxide, which is vitally important to cellular signaling (**Poulos2014**). Free heme molecules may be released upon degradation of hemoproteins (especially hemoglobin), however, heme is prooxidative and therefore toxic to cells and must be cleared. Heme oxygenases assist in the degradation of heme, and are regarded as potential therapeutics, due to anti-inflammatory effects[**Araujo2012**].

The enzymes with arguably the most potential applications, cytochrome P450s function as powerful monooxygenases. They participate in many reactions: capable oxidizing a wide range of substrates, serving to oxidize carbohydrates, steroids, fatty acids; catalyzing hormone degradation and synthesis; and degrading the majority of drugs(**Poulos2014**). Due to their extraordinary utility and range of reactions, cytochrome p450s are of great interest in the protein engineering field. Cytochrome P450s have the potential to be used in industrial biocatalysis, e.g. in

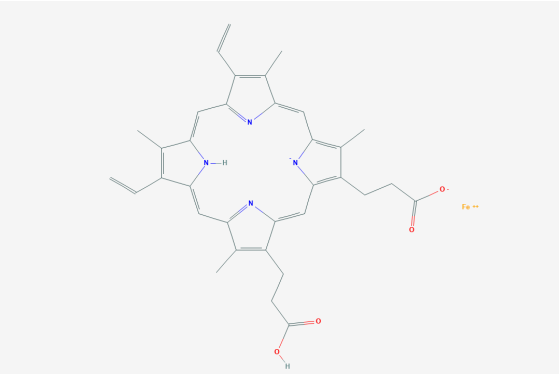

Figure 1: Heme-b (HEM)

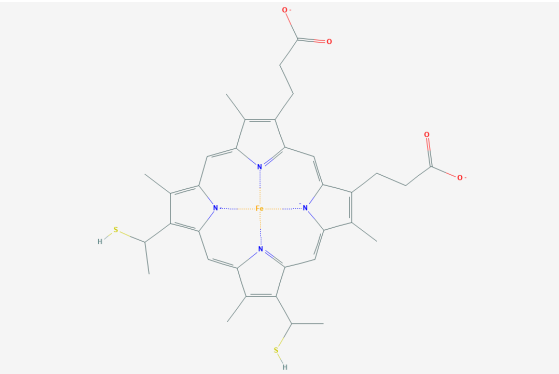

Figure 2: Heme-c (HEC)

pharmaceutical production, bioremediation of environmental pollutants[**Du2017**, **Lalonde2016**]. The limiting factor preventing its deployment has been the struggle to increase enzymatic efficiency and therefore yield of processes employing the enzyme[**Girvan2016**, **Li2020**].

Echoing the many applications of its host enzymes, the heme molecule itself has its own variety. There are several "hemes", structurally and chemically different, that are used to achieve different chemical reactions: in this study, we examined heme-b, heme-c, siroheme, and verdoheme. Their structures are shown in Figures 1-5.

The most common "heme", heme-b, has been discussed above. It binds principally via noncovalent interactions with its binding pocket. Heme-c is derived from heme-b; it binds, with few exceptions, covalently to cysteine residues in the binding pocket, forming thioester bonds between the residues and heme-c vinyl groups. Its function is much more specific than heme-b, mostly serving as an elctron carrier. The reason for this is not abundantly clear, but several studies suggest that

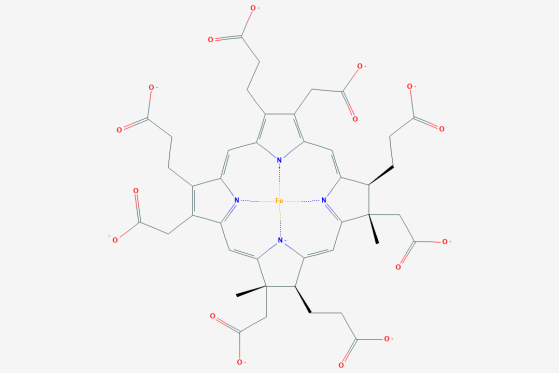

Figure 3: Siroheme (SRM)

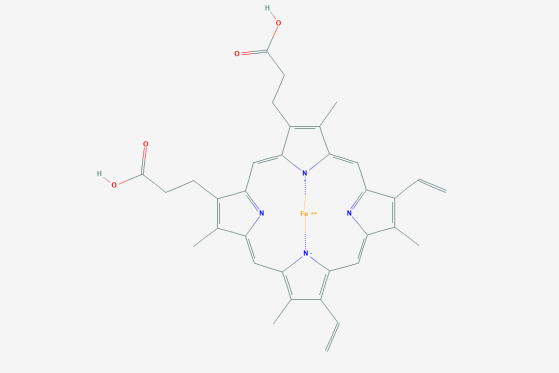

Figure 4: Verdoheme, VEA

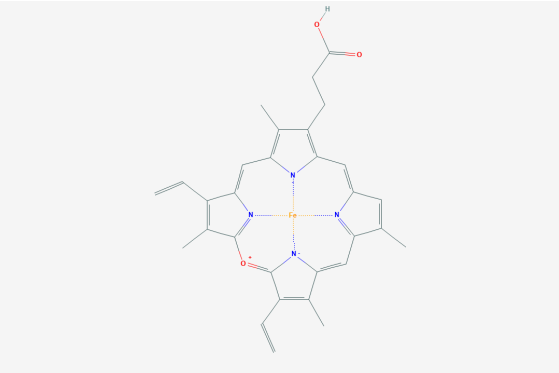

Figure 5: Verdoheme, VER

because of its covalent bonding, heme-c has an electronic potential that is broader and more specifiable than heme-b. [**Bowman2008**, **Kleingardner2015**]

Siroheme is even more limted in its applications but arguably of greater signifiance in nature. It is used exclusively in sulfite and nitrite reductases, which catalyze the reduction of the sulfates and nitrates plants uptake from the ground, and ultimately use to produce nitrogen and sulfur-containing amino acids(**Tripathy2010**). The reason for the use of siroheme in this reaction rather than heme is not completely understood, but one study suggests that, along with other cofactors involved in the reaction, siroheme forms a bridge to the catalyic iron atom in the porphyrin ring that is more efficient at channeling electrons than the bridge that could be formed by heme.[**Branzanic2019**] FIXME DEFINITELY CUT THIS DOWN IN WORDS.

Lastly, verdoheme is an intermediate product in the degradation of heme-b by heme oxygenase. When heme oxygenase degrades heme-b, biliverdin, carbon monoxide, and iron are produced; verdoheme is the precusor to bilverdin[**Lai2010**, **Sato2007**]. While a product of prior reactions wthin heme oxygenase, verdoheme appears to be oriented and bound differently [**Lad2004**].

In summary, heme molecules are used in a many reactions and several enzymes have the potential to be of great value either in biocatalysis or pharmaceutical applications. There are multiple methods employed to design molecules, but rational design in particular (basically, the mutation of certain residues based on an understanding of the structure-function relationships) is at least partially hampered by an incomplete understanding of the binding environment for heme. The importance of the binding environment was noted in a study seeking to design *de novo* heme-c based enzymes, and found the binding environment likely to be of importance in modulating redox potential [**Ishida2004**].

A fairly recent study conducted a structural analysis of 125 hemoprotein chains(**Li2011**). The study suggested hemoproteins undergo small conformational changes during binding; and that apo-form (ligand-containing) proteins may therefore be suitable for bioinformatics-based prediction and protein design. Additionally,

the heme binding environment of the protein dataset used in the study was found to be rich in aromatic and nonpolar residues.

The aforementioned study was published in 2011 - since then the PDB has been populated with far more hemoproteins. The 125 protein chains used in the study were sourced from a small dataset of proteins (FIXME: Double check but their Table 2 is like 20 proteins), compared to the amount of hemoproteins now available in the PDB. The focus of the study was on conformational differences induced by heme-binding.

need a better transition here

In this study, we focus on elucidating the binding environment for multiple heme molecules: heme-b (HEM), heme-c (HEC), siroheme (SER), and verdoheme (VEA/VER). UCSF Chimera was used to both extract and predict properties of a diverse set of hemoproteins. A robust and high-throughput framework was constructed to process the datasets for each heme molecule, requiring only inputs of which ligand is to be examined per dataset.

The properties extracted and predicted of the heme molecules were (I need to figure out how to do a bullet list in this format): +Volume +Solvent accessible and excluded surface areas of ligands and binding pockets +Amino acid frequencies within binding pockets +Planar angles of amino acids in binding pocket v. the ligand +CA-CB-Fe angles of amino acids in binding pocket v. the ligand +Distances of amino acids from the Fe atom of the ligand

These results may be of use in rational design of hemoproteins in future studies, or at least improve the understanding of the heme binding environment. **probably put something more... salesman-like here later**