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 complexed with another molecule. One such molecule is heme, a complex of iron and a porphyrin ring. Heme is employed by many metalloproteins to catalyze a broad range of reactions[**Poulos2014**].

There are many types of heme; but all of them are a coordination complex composed of an iron atom coordinated and bound to a modified porphyrin ring. Porphyins are composed of four pyrrole subunits (pentagonal structures of four carbons and a nitrogen) that are bound together via methine (i.e. carbon) bridges. Porphyrins are considered macrocycles, molecules with large ring structures. The ring within porphyrins is heterocyclic, and considered aromatic, and therefore acts as a large resonant structure capable of transferring electrons. Coupled with an iron atom to enable reduction-oxidation reactions and the macrocycle of the porphyrin ring, the overall heme complex is therefore highly suited for reactions involving electron transfer[**Kleingardner2015**].

Proteins containing heme are known as hemoproteins - hemoglobin and myoglobin are well-known examples, using heme to store and transport oxygen. Other examples of hemoproteins 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, including 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 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**].

Thus far, only heme-b and its respective hemoproteins have been discussed (although heme-b is the most abundant and most employed type of heme). Other types of heme are structurally and chemically different, and are used to achieve different chemical reactions. In this study, we examined heme-b, heme-c, verdoheme, and siroheme. Their structures are shown in Figures 1-5, and each heme is individually discussed below.

0.1 Types of Heme

0.1.1 Heme-b

The most common heme is heme-b. It is employed by the vast majority of hemoproteins. It is composed of an iron and porphyrin ring complex with attached vinyl and ethyl groups, and with the addition of two propionate groups. The iron atom is usually coordinated to a histidine or cysteine, depending on the enzyme[**Poulos2014**, **Kleingardner2015**].
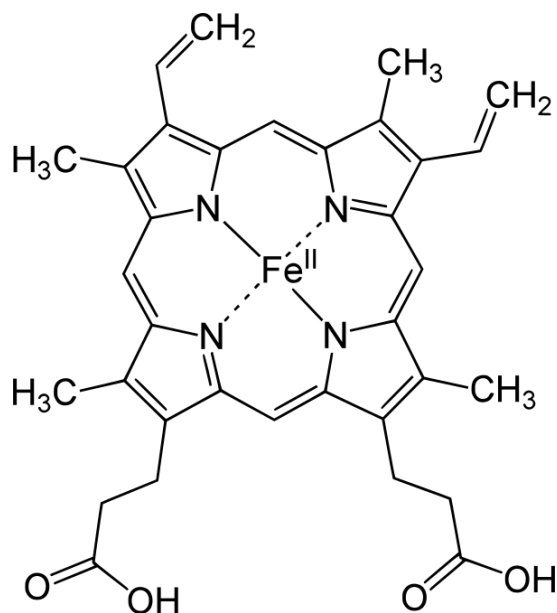
Figure 1: Heme-b (HEM)

The two propionate groups stabilize the heme in the pocket by forming polar interactions with salt bridges formed by arginine residues in the binding environment[**Barrows2005**]. This behavior is the same for heme-b and likely verdoheme. It has also been suggested that the propionate groups may also serve to exclude solvent from the binding environment, potentially acting to expel and repel water molecules [**Hayashi2009**].

0.1.2   Heme-c

Heme-c is derived from heme-b. It is fairly similar to heme, with two notable differences: the iron atom binds, with few exceptions, covalently to cysteine residues in the binding pocket; and its two vinyl groups form thioether bonds with amino acids in the protein binding pocket. Its function is much more specific than heme-b, mostly serving as an electron carrier rather than catalyzing a plethora of reactions. The reason for this is not abundantly clear, but several studies suggest that because of its covalent bonding, heme-c has an electronic potential that is can be far lower and in general more broad, and more specifiable, than heme-b. [**Bowman2008**, **Kleingardner2015**]
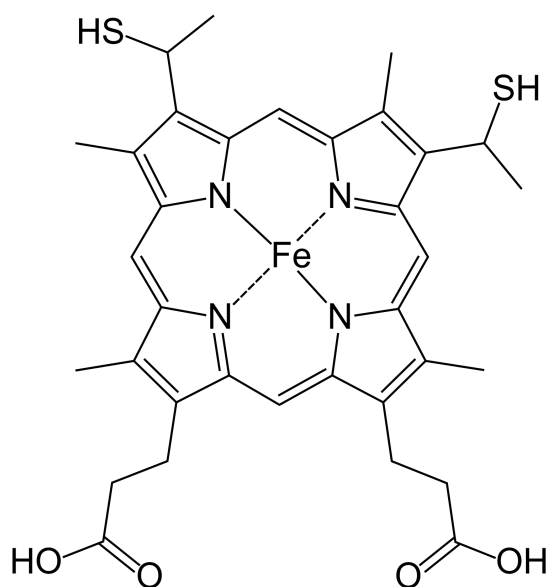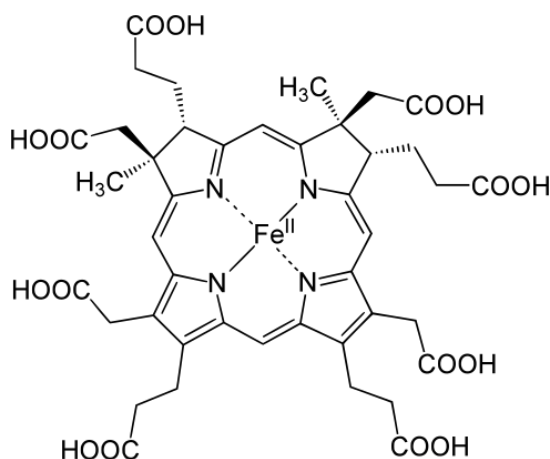
Figure 2: Heme-c (HEC)

Figure 3: Siroheme (SRM)

0.1.3 Siroheme

Siroheme is even more limited in its applications, but highly specialized for its role. It is still an iron atom-porphyrin coordination complex, but it contains far more carboxyl and propionate groups than the other types of heme, making it highly polar. It is used exclusively in sulfite and nitrite reductases, which catalyze the reduction of the sulfates and nitrates plants uptake from the ground, providing the sources of nitrogen and sulfur used to produce nitrogen and sulfur-containing amino acids[**Tripathy2010**]. The reason for the use of siroheme in this reaction
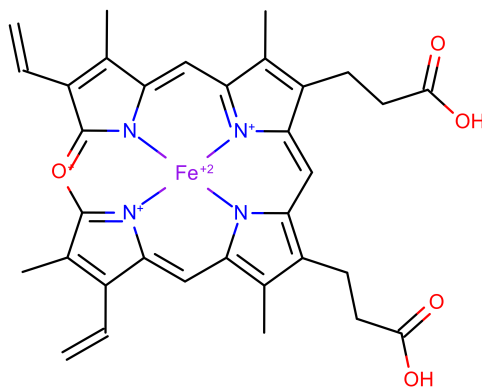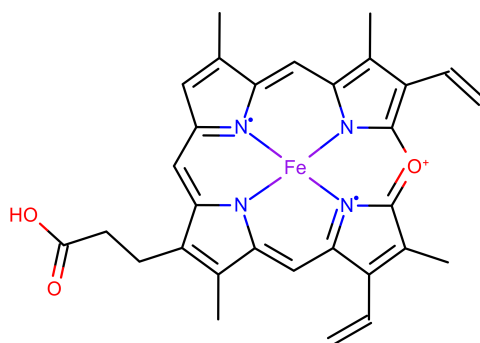
Figure 4: Verdoheme (VEA)

Figure 5: Verdoheme (VER)

rather than heme-b is not completely understood. But one study suggests that the bridge that siroheme forms between its catalyic iron atom, and the protein matrix environment (which also necessarily involves another cofactor, a cluster of cubane for electron transfer and provision) is more efficient at channeling electrons than the bridge that could be formed by heme.[**Branzanic2019**]

0.1.4 Verdoheme

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**]. The two structures used in the study, VEA and VER, are verdoheme at different stages of degradation, either partially oxidized or containing one less propionate group.

In summary, heme molecules can have very different structures and functions; they enable and catalyze an extraordinary amount, and extraordinarily diverse set of chemical reactions. They are important, not only as a study of how one class of molecule can be involved in a broad swath of reactions, but because hemoproteins have the potential to be of great value in biocatalysis, bioremediation, and pharmaceutical applications.

There is a significant barrier to the employment of hemoproteins in these areas, however: improving their efficiency far beyond what is found in nature. This is the field of artifical metalloproteins, or metalloenzymes: engineering metalloenzymes to improve them; increasing efficiency, stability, or even to introduce new reactions to heme's repetoire.

There are multiple methods employed to design these 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. For example, 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 environments for both heme-b and heme-c were analyzed, and relative frequencies per amino acid were reported. Cysteine, histidine, phenylalanine, methionine, and tyrosine were found to be the most abundant residues within the binding environments of both heme-b and heme-c.

The aforementioned study was published in 2011 – since then the PDB has been populated with far more hemoproteins. The focus of the study was on conformational differences induced by heme-binding, rather than the binding environment, although the relative frequencies of amino acids were reported. Interactions of the more abundant residues with heme-b or heme-c, including interactions with the porphyrin ring, were briefly discussed and this discussion will not be reproduced here.

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

The properties extracted and predicted of the heme molecules' binding environments were: the amino acid frequencies; the distances of the amino acids from the heme iron; the volume of the binding pocket; and the surface areas of both the hemes and the binding pocket. These data can be expected to be of use, or at least of interest, to efforts in artifical metalloenzyme design.

Additionally, angular data for the residues within the binding environment were obtained. These data were produced more for exploratory purposes and are not discussed extensively in this study. Specifically, planar angles and the angle between residues' alpha-carbon, beta-carbon, and heme iron (CA-CB-Fe) were obtained.

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.