Structural Analysis of Hemoprotein Binding Sites

Patrick Finnerty

Hemoproteins are able to catalyze an extraordinarily broad range of reactions. This class of proteins includes cytochrome P450s and peroxidases, among many others. Hemoproteins by definition include heme to achieve their chemistry; but heme molecules themselves vary considerably, and different heme molecules are employed depending on the function of a particular hemoprotein. At present, there is a knowledge gap in the literature of the structural features required to bind heme in hemoprotein binding sites.

In this study, we investigated the binding environments of heme-b, heme-c, siroheme, and verdoheme. A high-throughput framework was developed to generate, process, and analyze data derived from the Protein Data Bank (PDB) files used in the study. Amino acid frequencies, volume, surface areas, and angular relations within the heme binding sites were all obtained and examined for a diverse set of hemoproteins. Overall, we find that nonpolar residues and their interactions with heme may be of greater importance to heme binding than previously thought, suggesting that polar interactions, while necessary, are insufficient to efficiently bind heme. We also report several values for binding pocket volumes, surface areas, and angular data that may be of utility in artificial metalloenzyme design. The framework developed for this study is highly adaptable and may facilitate similar investigations for other proteins and ligands.

Table of Contents

# Lay Summary

We investigated how heme, a molecule involved in many biological processes, binds to proteins. The study was conducted computationally. 3D structures of proteins were downloaded from a database, and run through software specialized for viewing molecules. This software was used to examine the region on the protein where heme binds (binding pocket). The software also predicted the volume and surface areas of the binding pocket. The data produced were analyzed using statistical software.

We found that a greater proportion of the binding pocket associates with the hydrophobic and uncharged parts of the heme molecules more frequently, and therefore likely more strongly, than previously thought. These results suggest that these hydrophobic, electrically inert parts of the pockets may be very important to consider when studying, or designing proteins containing heme. We also report the volume and surface area data that were predicted, which support the above conclusion and may also be of help when studying or designing these proteins.

# 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?**](#ref-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([Kleingardner and Bren 2015](#ref-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?**](#ref-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?**](#ref-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?**](#ref-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?**](#ref-Du2017); [Lalonde 2016](#ref-Lalonde2016)). The limiting factor preventing its deployment has been the struggle to increase enzymatic efficiency and therefore yield of processes employing the enzyme([**Girvan2016?**](#ref-Girvan2016); [**Li2020?**](#ref-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

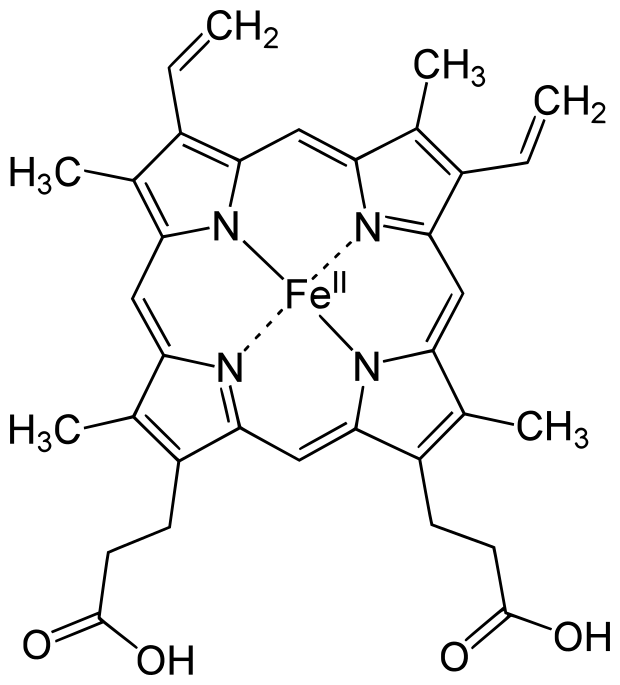


Figure 0.1: Heme-b (HEM)

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?**](#ref-Poulos2014); [Kleingardner and Bren 2015](#ref-Kleingardner2015)).

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([Barrows and Poulos 2005](#ref-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?**](#ref-Hayashi2009)).

### 0.1.2 Heme-c

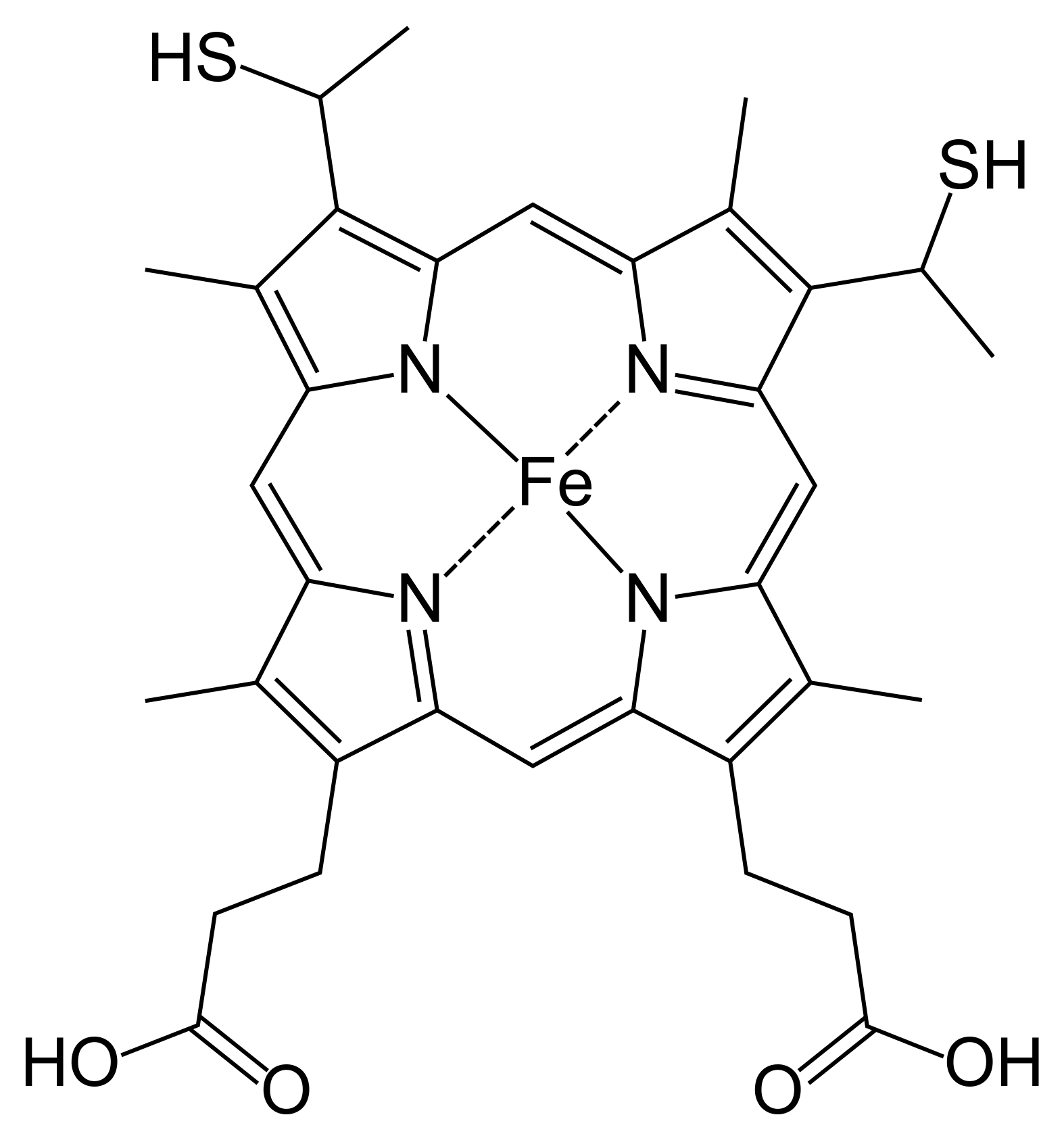


Figure 0.2: Heme-c (HEC)

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. ([Bowman and Bren 2008](#ref-Bowman2008); [Kleingardner and Bren 2015](#ref-Kleingardner2015))

### 0.1.3 Siroheme

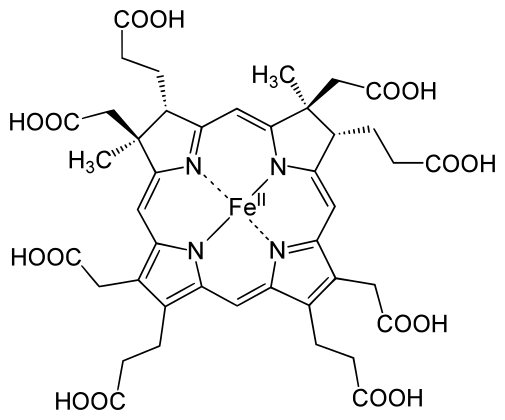
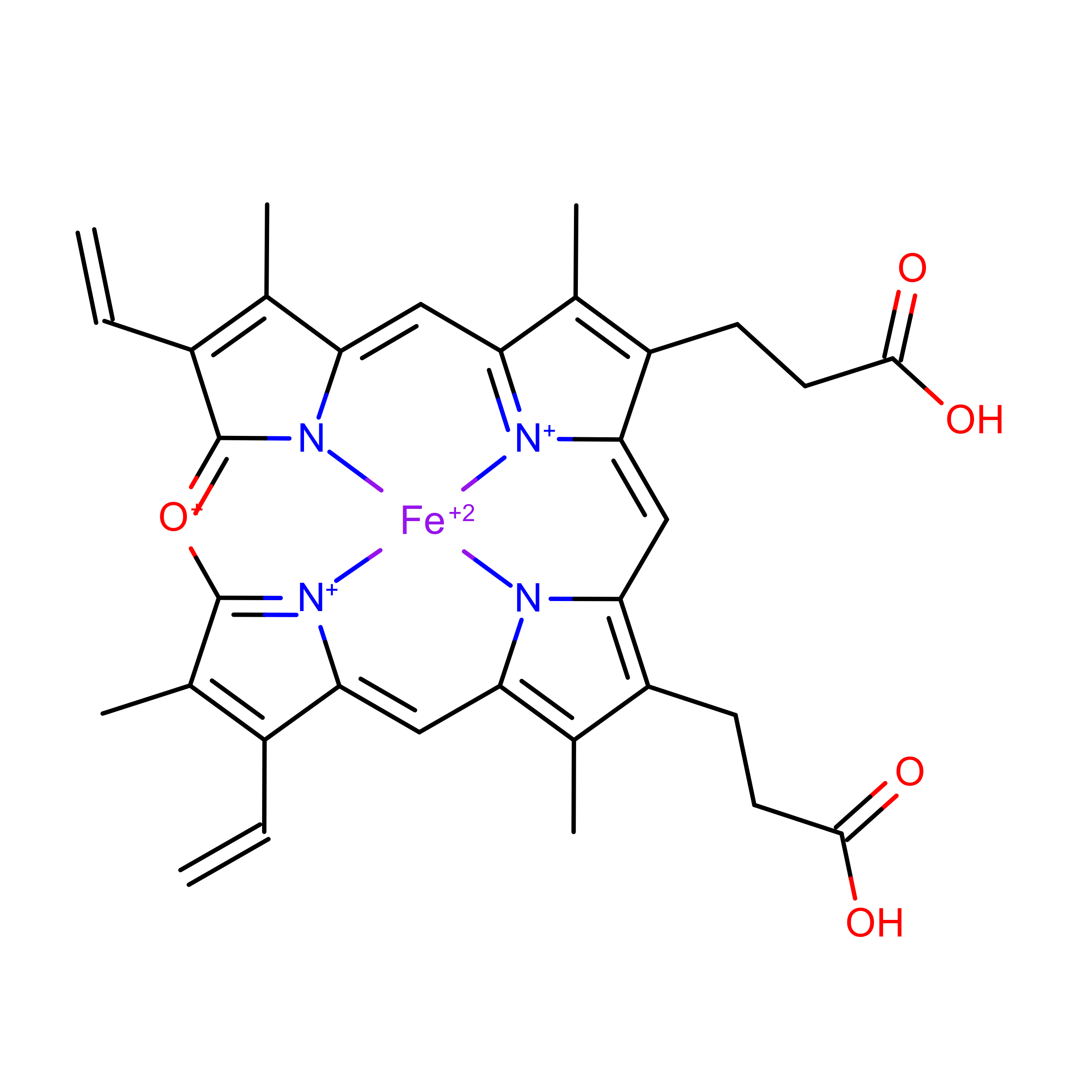
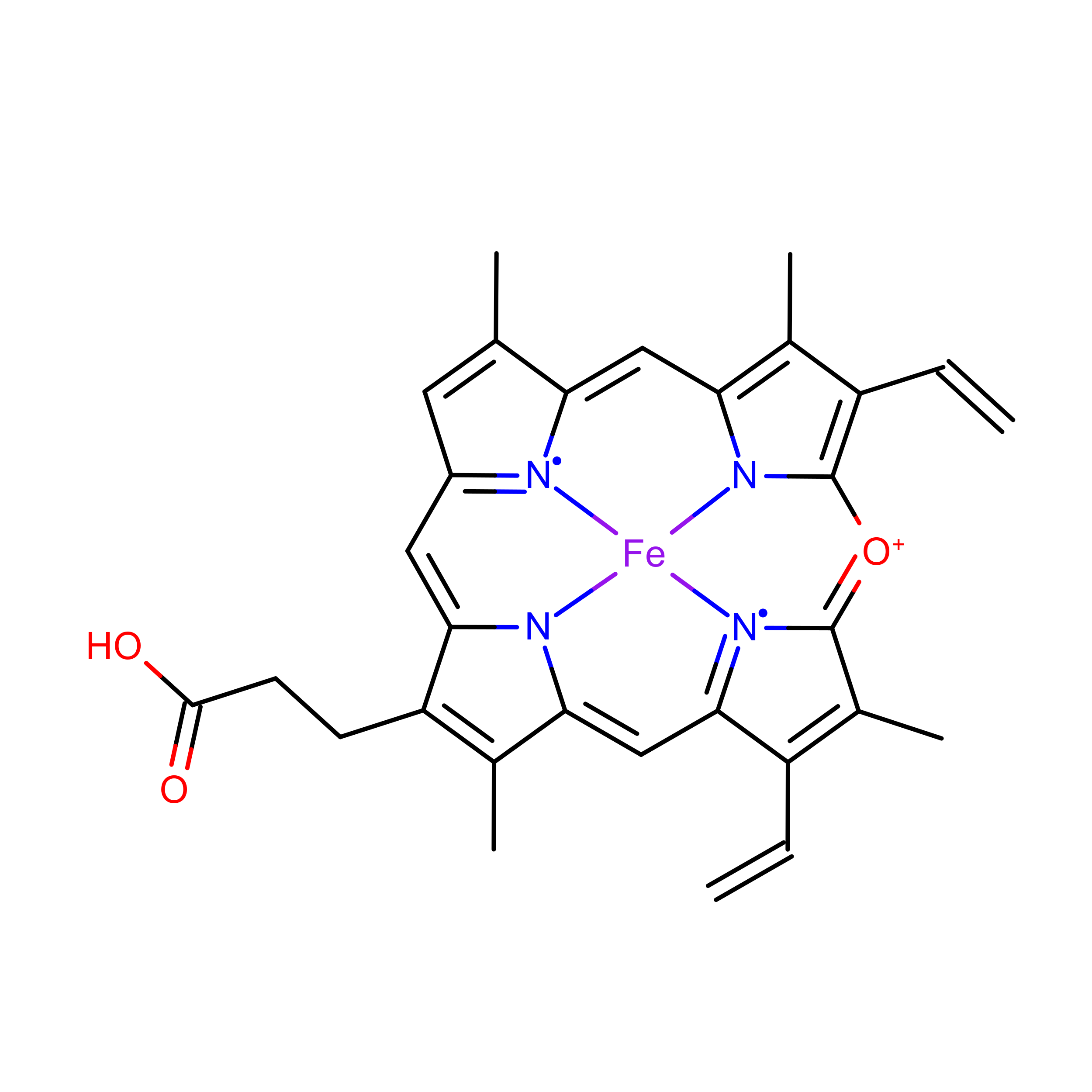


Figure 0.3: Siroheme (SRM)

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?**](#ref-Tripathy2010)). The reason for the use of siroheme in this reaction 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.([Brânzanic, Ryde, and Silaghi-Dumitrescu 2019](#ref-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([Lai et al. 2010](#ref-Lai2010); [Sato et al. 2007](#ref-Sato2007)). While a product of prior reactions wthin heme oxygenase, verdoheme appears to be oriented and bound differently ([**Lad2004?**](#ref-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 ([Ishida et al. 2004](#ref-Ishida2004)).

A fairly recent study conducted a structural analysis of 125 hemoprotein chains([**Li2011?**](#ref-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.

# 1 Methods

All scripts (as well as raw data, results, and this document) are available on GitHub([Finnerty 2021](#ref-Finnerty_hemebinding_2021)).

## 1.1 Datasets

A list of PDBs was assembled that represented either a representative sample of a variety of proteins, with a resolution better than 3A, (HEM and HEC) or, all proteins containing these ligands were downloaded from the PDB (in the case of SRM, VER, VEA). Not all downloaded PDBs were appropriate for this study (e.g. contained superimposed structures) and therefore the amount of PDBs was culled. The datasets are current as of 16 August 2021.

The size of the datasets actually used in the study were as follows: HEM (n=58), HEC(n=13), SRM (n=9), VER (n=2) and VEA (n=2), which are merged for a combined n=4 for VERDOHEME.

The name of all proteins used in the study and their source organism are provided tables within Appendix 5.1.

## 1.2 Pre-processing

Many of the PDBs downloaded were multimeric structures. The number of subunits per protein would skew results and overrepresent especially large multimeric proteins. Therefore, to only allow for one heme binding site per PDB, all downloaded PDBs were converted to monomeric structures. This was achieved by saving a single chain (chain A) of each PDB and eliminating all other chains. The single chain was then saved as a PDB and used in all subsequent scripts. Part of the script is reproduced below:

from chimera import runCommand as rc  
# select chain A, a single unit  
rc("sel :.a")  
# select everything else  
rc("sel invert sel")  
# delete everything else besides that chain A  
rc("del sel")  
# now save the monomer:  
rc(("write format pdb 0 "+unexpandedResultPath+activeLigand+"/%s")%  
 (fn + ".mono.pdb"))

## 1.3 Processing Monomers

UCSF-Chimera was used to generate all data in this study. Multiple Python scripts were employed to achieve a high-throughput process where all monomeric PDBs could be processed in the same session.

Chimera was used to predict the following qualities: Volume of the ligand binding pocket, accessible and excluded surface area of the ligand, and accessible and excluded surface area of the binding pocket. These calculations require a population of atoms to be selected for the calculation.

Atoms were selected within a distance cutoff, to be considered as potentially interacting with the ligand or forming the binding pocket. Distance cutoffs from the ligand of 5A and 7A were chosen; for the predicted qualities, the algorithms were run twice to get values at 5A and 7A. For the distance and angle calculations, only the 7A distance cutoff was used, as the cutoff does not factor into any calculations and may be set during analysis.

As these cutoffs are selected arbitrarily, data from the 5A and 7A runs are overlaid in the figures reported in Appendix 4. Data tables are also provided in Appendix 5.

### 1.3.1 Amino Acid Frequency

Amino acids within the bounds of the lower and upper distance cutoff were selected and recorded. These were then counted for frequency per residue.

### 1.3.2 Volume Calculations

Volume of the binding pocket was predicted via Surfnet ([**Laskowski1995?**](#ref-Laskowski1995)), and run with default parameters of Grid Interval = 1.0 and Distance Cutoff = 10.0 (the latter option does not relate to the distance cutoff from the ligand). Surfnet is the molecular volume calculation tool implemented within UCSF Chimera. The script used selects the residues around heme to consider as the bounds of the pocket, but effectively ignores heme’s presence as its calculates the volume, as if the pocket were empty:

from chimera import runCommand as rc  
# Select the atoms within 7A of heme.   
#Then, of that selection, keep everything but heme.  
rc("sel :"+activeLigand+" za < "+angstromDistance)  
# this is the syntax that accomplishes our desired selection  
rc("sel sel &~:"+activeLigand)   
interface\_surfnet("sel","sel")  
rc("sop split #") # acquire the individual pockets that have been generated  
rc("measure volume #") # measures volume of individual pockets  
# in R we keep only the largest volume

Surfnet, at least in this investigation, was prone to generating very small volumes. During analysis these were removed and only the largest volume generated is recorded, since the largest volume generated and identified is most likely the binding pocket. Two figures below demonstrate a run where one good pocket is produced, and one where a few very small “bubbles” are generated:

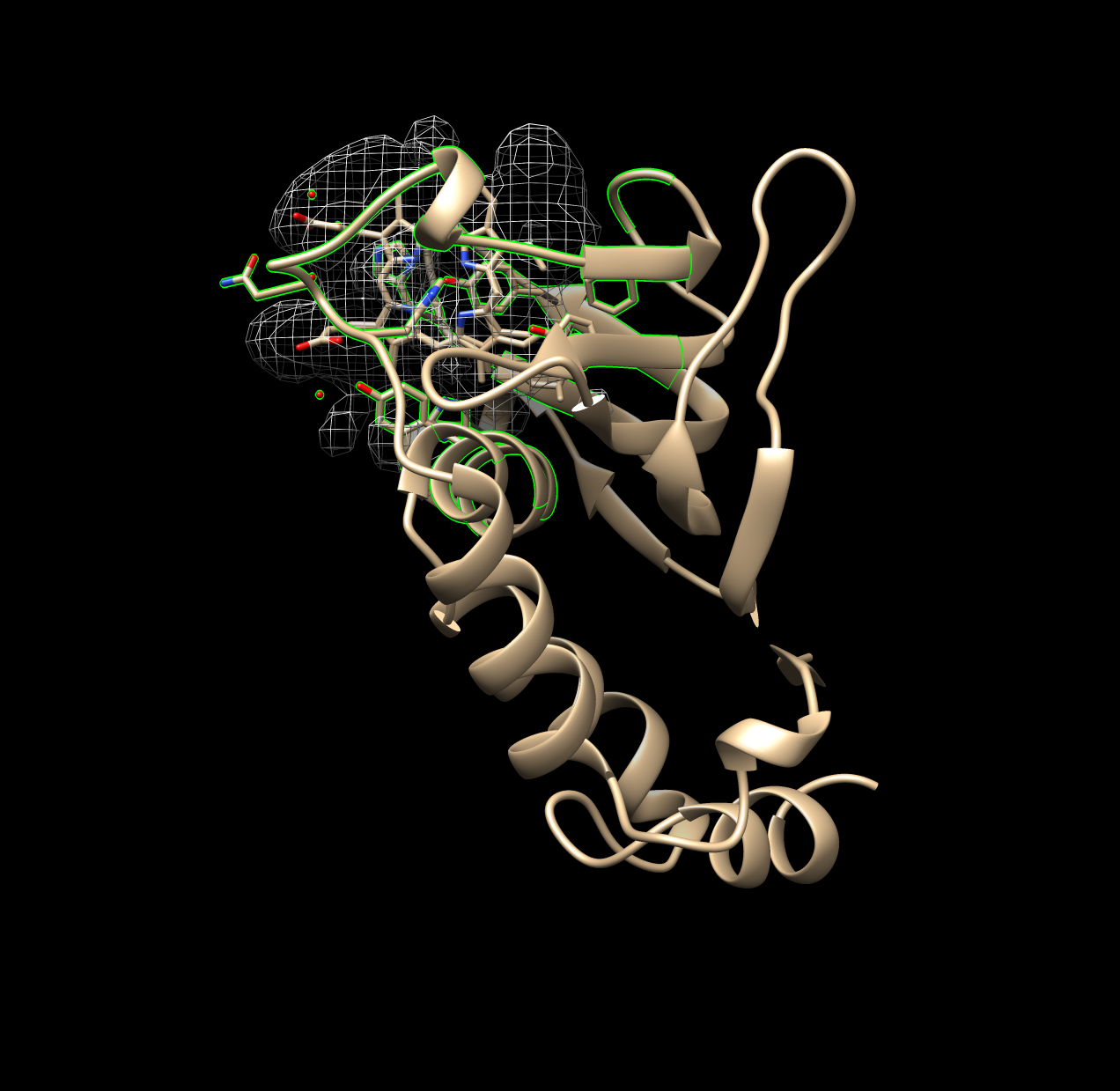


Figure 1.1: Good Example of Surfnet Run (1B2V)

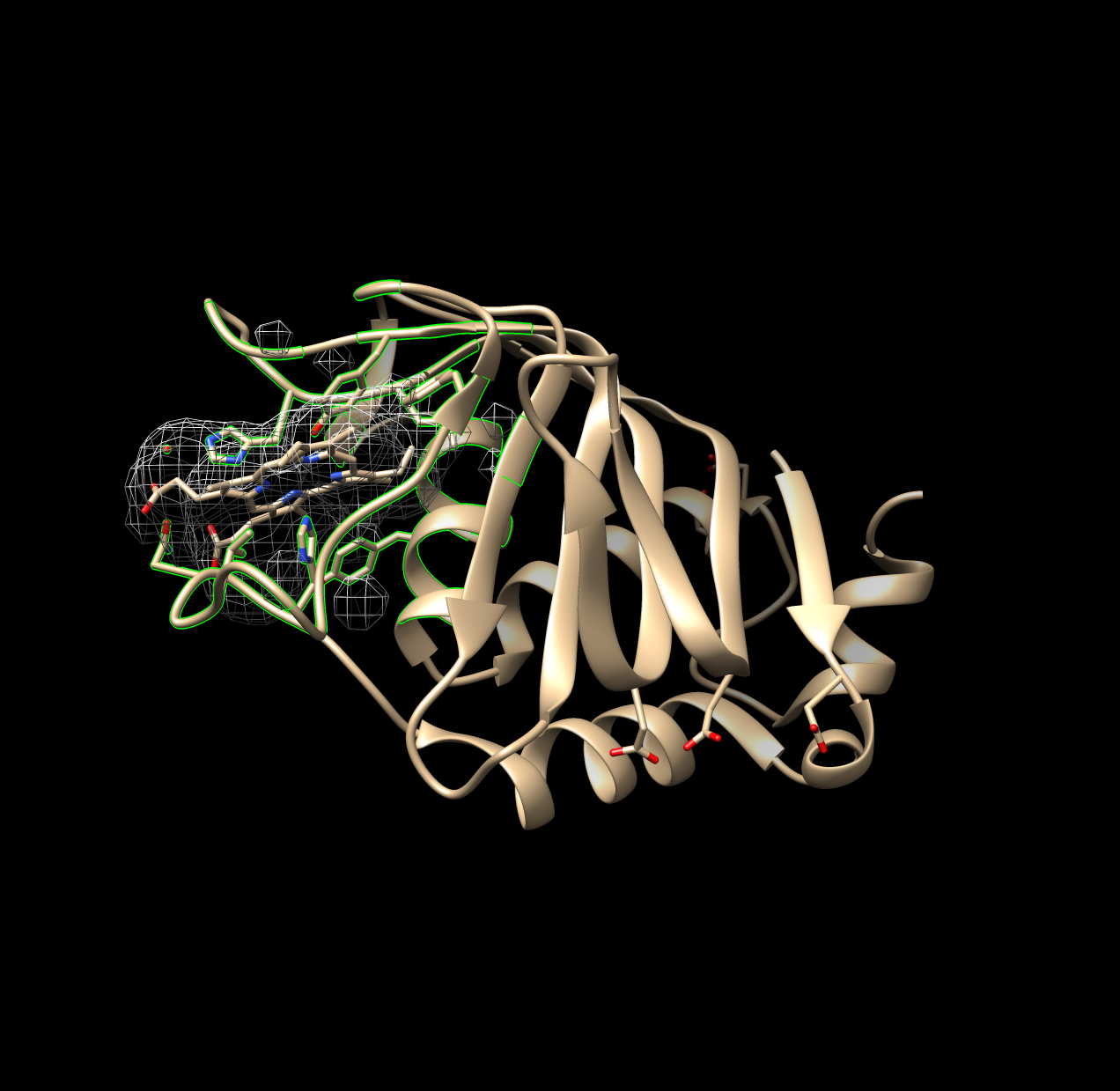


Figure 1.2: Non-Ideal Example of Surfnet Run (1DKH)

### 1.3.3 Surface Area Calculations

Solvent excluded and solvent accessible surface areas of both the ligand and the binding pocket were calculated using Chimera’s “surf” algorithm, which itself is an implementation of a program called MSMS ([**Sanner1996?**](#ref-Sanner1996)).

These two measures are similar but not the same. Solvent accessible surface area represents the surface area of the protein that a solvent molecule (i.e. water) may interact with. It is calculated by rolling a sphere on the Van der Waals surface of the protein, and the *center of the sphere* is recorded as the bounds of the accessible surface area. Solvent excluded surface area is calculated the same way, rolling a sphere on the Van der Waals surface of the protein, but instead the *point of contact of the sphere against the Van der Waals surface* is recorded as the excluded surface area. The solvent excluded surface area may therefore be considered the bounds of the protein itself, versus the solvent accessible surface area, which can be considered the bounds at which a solvent may interact with the protein([**Sanner1996?**](#ref-Sanner1996)).

### 1.3.4 Distance Calculations

Distances of amino acids from the ligand could not be calculated accurately nor precisely in a direct way. Instead, distances for each atom composing a residue were calculated. This was achieved using a built-in function of chimera; the syntax is not straightforward, but part of the script is shown below. The distances of all atoms within a residue were averaged, and this value was taken as the mean distance of the entire residue and used in subsequent steps.

from chimera import runCommand as rc  
#select and define the Fe atom  
rc("sel :HEM@Fe")   
# index to acquire the one atom selected  
fe = chimera.selection.currentAtoms()[0]   
# select all atoms within angstromDistance of Fe (also de-selects Fe)  
rc("sel sel za < "+settings.angstromDistance)   
# define this selection of atoms within distance as a list  
nearbyAtoms = chimera.selection.currentAtoms()  
# parse and print the distances (and coordinates) of these atoms   
for i in nearbyAtoms:  
 print "Atom being analyzed...", i, "... Distance to Fe...",  
 #prints distance between atom i and the Fe atom  
 i.coord().distance(fe.coord())

The data produced in this step therefore include the mean distance of each amino acid. Distances are traceable per residue and atoms in each residue; this data was used to construct the distributions of amino acids over distance, and the angular data below are cross-referenced with this list of distances.

### 1.3.5 Planar Angle Calculations

Individual residues and the ligand were defined as axes. The angle between each residue’s axis and the axis of the ligand were calculated. Each axis functions essentially as a separate plane. This employed the “define axis,” and “angle” functions of Chimera; the Axes/Planes/Centroids Structural Analysis function of Chimera via GUI.

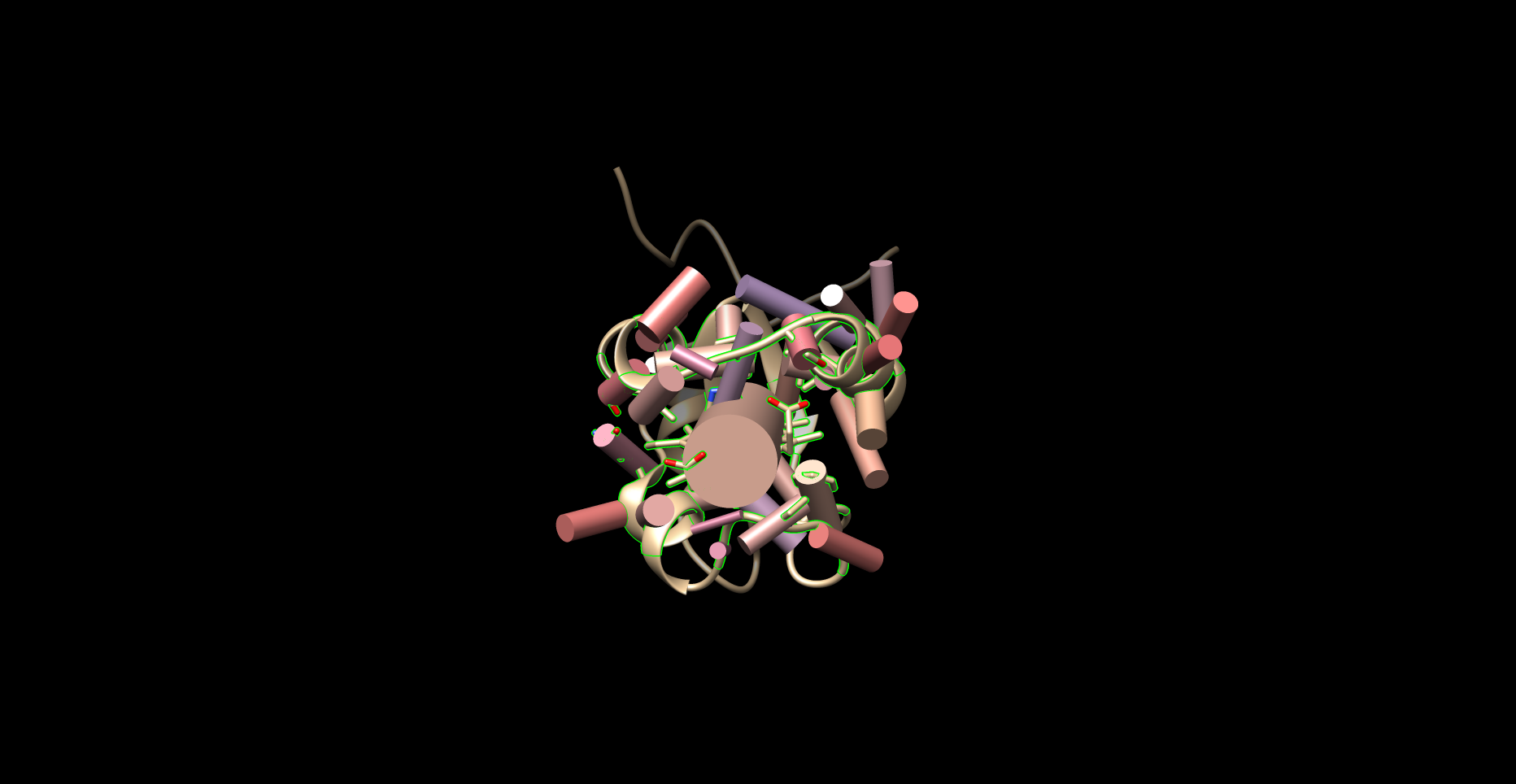


Figure 1.3: Example of Planar Angles Calculation (1B5M)

### 1.3.6 CA-CB-Fe Calculations

Residues within the distance cutoff were examined one by one. The angle of between each residue’s carbon alpha (CA) and carbon beta (CB) and the Fe of the ligand was calculated, using the “angle” function of Chimera. The ligand nor the Fe atom were compared with themselves.

## 1.4 Import to R

The data produced by Chimera and the Python scripts were stored as .txt files. These files were imported to R and processed from .txt files into organized data formats. All plots and tables were constructed using R and imported directly to this document using Rmarkdown.

# 2 Results and Discussion

## 2.1 Analysis of Residues Nearby Each Heme Molecule

We began the study by acquiring data to elucidate and quantify the propensity of amino acids to interact with heme (HEM, HEC, SRM, VEA/VER) in its binding environment. This study focused on potential interactions with the entire heme molecule, including the porphyrin ring and attached groups; therefore, any amino acids with potential interactons with the heme iron, porphyrin ring, or groups on the porphyrin ring (e.g. vinyl, propionate groups), were included in the data gathered for this section. A potentially interacting amino acid was therefore defined as any amino acid with at least one atom within the distance cutoffs (5 and 7 Angstroms (A)) from the heme *molecule*.

Amino acid frequencies were obtained for residues within the distance cutoffs of 5A and 7A - these data are overlaid in several figures in Appendix 4.1. The trends in these data are very similar and therefore only the data pertaining to the 7A distance cutoff are discussed below.

### 2.1.1 Heme-b

#### 2.1.1.1 Amino Acid Frequencies in Binding Pocket

Figure 2.1 plots the frequency of each residue within 7A of heme-b.

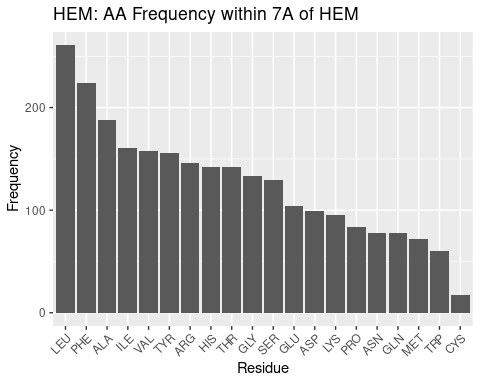


Figure 2.1: HEM: AA Frequency within 7A

Beginning at the left of Figure 2.1 and moving right, large, nonpolar amino acids appear most frequently within 7A: LEU and PHE; ILE appears less frequently than these two amino acids but nonetheless is in high frequency. Small, nonpolar amino acids ALA and VAL also appear very frequently. As the majority of the heme-b molecule is made up of the nonpolar porphyrin ring, these amino acids are therefore likely in such high frequency to provide the nonpolar interactions with the pyrole groups and methyl and vinyl groups.

Tyrosine, arginine, histidine appear next most frequently. The two propionate groups on heme make polar interactions with salt bridges formed between arginine groups within the binding environment([Barrows and Poulos 2005](#ref-Barrows2005)). Therefore, the tyrosine and histidine likely form polar interactions with the portion of the propionate groups not interacting with the arginine salt bridges. This, in addition to the nonpolar interactions above, likely provides as hospitable of a binding environment as possible to coordinate the heme. It should be noted histidine is one of the residues that coordinates the iron atom, and this may therefore increase its frequency in the binding pocket.

Glycine is a small residue and cannot form significant interactions within its environment; however, its frequency, or lack thereof (compared to background frequency, discussed below), suggests the binding pocket may not require as much flexibility or spatial considerations as in the rest of the protein.

Next appear serine, glutamate (glutamic acid) and aspartate (aspartic acid) and lysine. These are polar residues, and glutamate and aspartate are negatively charged; lysine is polar too, but positively charged (at pH 7). The negative charge on glutamate and aspartate may explain why they are fairly infrequent: although polar, the negative charge may induce a repulsion effect with the propionate groups. Or, it may be that other amino acids are preferable, as is likely the case for lysine. Lysine is polar and positively charged; but arginine residues forms the salt bridges necessary for propionate to make polar interactions. And histidine coordinates the heme iron. Therefore the infrequency of lysine may be less due to a problem with lysine and more due to a preference for other polar amino acids.

Proline is a small nonpolar amino acid in low frequency; the trend for heme-b, at least, appears to be to favor large nonpolar amino acids in the binding pocket. This may suggest that a large amount of nonpolar interactions, per residue, is favored in the binding pocket, perhaps because of the limited space available to position residues to interact with heme.

Asparagine and glutamine are both medium-sized polar amino acids; given the trends already discussed it is surprising these are not in greater abundance. But as with proline, it may simply be a matter of maximizing the benefit of the interactions that may be formed with the heme; while asparagine and glutamine are polar, amino acids like arginine and histidine are both polar and positively charged (and arginine forms salt bridges), capable of stronger interactions with the electronegative propionate groups.

Methionine and tryptophan appear very infrequently in the binding pocket. Tryptophan is very surprising to find as second-to-least frequent. It is a large nonpolar amino acid - but perhaps its single, potential hydrogen bond, although weak, is enough to prefer completely nonpolar residues. Or, with its size, it is preferable to have more numerous, smaller nonpolar residues that can favorably interact with the porphyrin while reducing steric hindrance of other residues in the environment. The reason for methionine’s low frequency is not clear, perhaps for similar reasons as with proline, where more intensely nonpolar residues are preferred, rather than less nonpolar residues being unfavorable.

Cystine appears most infrequently of all the amino acids in the binding pocket. This is quite surprising - cystine is the other highly conserved residue that may coordinate the heme iron. Perhaps the sample of PDBs used in this study mostly use histidine to coordinate the iron - but this would only account for one residue in the binding pocket per pdb. Therefore these results suggest that while cystidine may be well suited to coordinate the iron in heme, it is poorly suited to form any nonpolar interactions with the porphyrin ring, leaving the task up to other, more suitably, intensely nonpolar amino acids.

Moving away from discussing individual amino acid populations, what is especially notable of the data for heme-b is that nonpolar residues appear in much greater frequency than polar residues. Nonpolar interactions with heme are therefore more numerous than polar interactions. Their multiplicity may also suggest that they are potentially of greater importance than previously thought. At the very least, these results suggest that polar interactions and coordination of the iron atom, while necessary for heme binding, are insufficient, and that nonpolar interactions and the population of nonpolar residues in the binding pocket should be considered when examining the binding environment of heme.

#### 2.1.1.2 Comparison with Background Amino Acid Frequencies

While the frequencies of amino acids in the binding pocket have been discussed, it may also be of interest to compare against the background amino acid frequency, the general frequency of amino acids within the entire monomer. The degree to which any results may affect the significance of the frequencies of the amino acids in the binding pocket is unclear - those amino acids are interacting with the heme. However, an in depth examination of similarities and differences may reveal that some amino acids may simply be extremely highly conserved by chance and by virtue of their numerous population, rather than some chemical benefit.

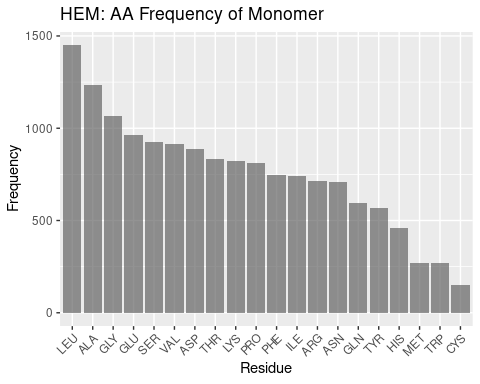


Figure 2.2: HEM: AA Frequency of Monomer

Figure 2.2 documents the frequencies of amino acids overall within the monomer.

Leucine and alanine, as in the binding pocket frequencies, are highly frequent in the overall monomer. This may suggest their prevalence in the binding pocket may simply be due to a high population of leucine and alanine in hemoproteins.

However, after these two amino acids the tendencies in frequency for the binding pocket and the monomer at large diverge. Glycine is in high frequency - likely due to more complex geometry e.g. helices outside the binding pocket. In interest of brevity, the remaining frequencies are summed up thus: the same trends that appear to exist in the binding pocket do not appear to exist in the monomer at large. While the order of frequencies in conserved binding pockets can be rationalized, justifying the overall frequencies in monomers invites significant speculation.

#### 2.1.1.3 Distributon of Amino Acids by Distance

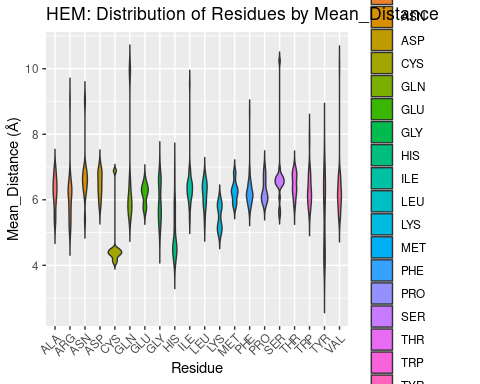


Figure 2.3: HEM: Residue Distribution by Distance

After an exhaustive exploration of the relative frequencies of amino acids in the binding pocket, Figure 2.3 below is fairly straightfoward. It plots the distribution of amino acids in the binding pocket against their distance from the iron of the heme.

We find that only a few residues come in close contact (<4A) of the heme: cysteine, histidine, and tyrosine. Most residues center their distribution at around 6A, although lysine seems more biased than the remaining residues to be a bit closer. Cysteine and histidine may be at least in part explained to be close due to their use as coordinating residues; histidine, being in greater frequency, may also be this close due to favorable interactions with the porphyrin ring.

The proximity of tyrosine however, is more notable. It cannot form coordination bonds with the heme iron, but tyrosine residues do interact with the propionate groups. Tyrosine is also required for redox reactions, and part of the population of tyrosine residues may therefore be in close proximity to heme to facilitate electron transfer in various enzymes ([**Poulos2014?**](#ref-Poulos2014)). These results suggest that of all potentially interacting polar/positively charged residues, tyrosine is the most likely at least to be in close proximity to the heme molecule. Whether this illustrates an importance of tyrosine to interact with propionate groups, or instead the need for tyrosine to be in close proximity in order to form such interactions, or simply demonstrates involvement in oxidation/reduction reactions, is beyond the scope of this study.

### 2.1.2 Heme-c

#### 2.1.2.1 Amino Acid Frequencies in Binding Pocket

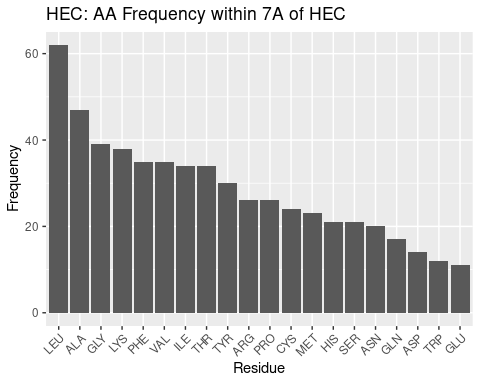


Figure 2.4: HEC: AA Frequency within 7A

Leucine and alanine again are highly frequent for HEC, followed by quite similar trends, and therefore HEC will not be as thoroughly discussed as HEM. The most notable differences may be that GLY and CYS are in far higher frequency than in heme. Heme-c almost always forms covalent bonds with cysteine residues, and this may explain that frequency. But as for the high frequency of glycine, the reason for its abundance is unclear, although it seems it may have an important role in heme-c pockets.

#### 2.1.2.2 Comparison with Background Amino Acid Frequencies

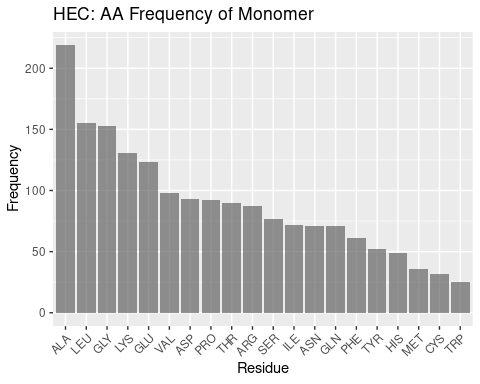


Figure 2.5: HEC: AA Frequency of Monomer

Generally, the heme-c monomer AA frequency profile appears similar to the heme-b monomer, with a high frequency of alanine and leucine, followed by a divergence in the frequency of amino acids and therefore a struggle to form any meaningful discussion when it comes to comparing the binding pocket frequencies against background frequencies.

#### 2.1.2.3 Distributon of Amino Acids by Distance

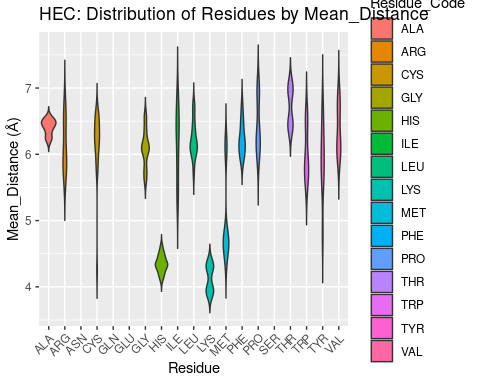


Figure 2.6: HEC: Residue Distribution by Distance

The distribution of amino acids over distance from the heme iron for HEC is similar to HEM, with some exceptions. Cys, His, Tyr again are amongst the closest residues to HEC, likely for the same reasons of very strong polar interactions or coordination. Additionally, cysteine forms covalent, thioether bonds with heme-c, providing further justification for its proximity. However, for heme-c, lysine and methionine also are very proximal. The methionine residues are nonpolar, small, neutral; lysine is polar and positively charged; neither of these residues are favored to be included in the heme-b binding environment despite very similar structures. The reason for their inclusion so close to the binding pocket is therefore unclear, but based on their distribution, and lysine being even more close proximity than heme, the results suggest these two residues may have important roles.

### 2.1.3 Verdoheme

#### 2.1.3.1 Amino Acid Frequencies in Binding Pocket

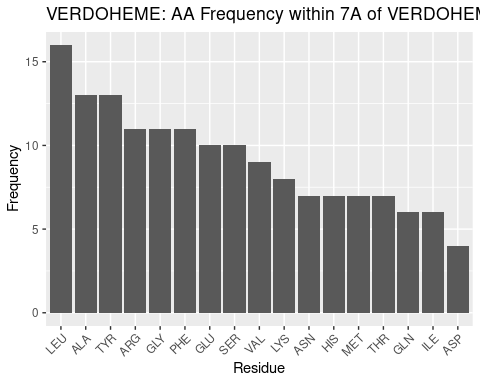


Figure 2.7: VERDOHEME: AA Frequency within 7A

The frequqency of amino acids in the binding pocket for verdoheme is dissimilar from heme-b and heme-c above. This is fairly surprising, given that verdoheme is an intermediate in the binding pocket for heme-b within heme oxygenases.

Leucine and alanine are again most frequent, but after these, results diverge. Tyrosine and arginine are next most frequent - surprising, given that this is still the same pocket that bound heme-b. The data for heme-b indicate more frequent nonpolar residues before tyrosine. Chemically, it may be that as heme-b is oxidized, there is greater need for polar interactions; this would help to explain the high frequency of polar residues, but does not explain the shift in amino acid frequencies within what would be expected to be a similar binding pocket - all verdoheme PDBs in this in study were sourced from heme oxgenase proteins. Some heme oxygenases are included for heme-b, but they are amongst a diverse set of proteins. Therefore, the heme oxygenase environment may simply be host to more polar residues than normal for hemoproteins. This also agrees with tyrosine’s inclusion in redox reactions, and it may be favored to be present in heme oxygenase. These results indicate more about the heme oxygenase environment than the verdoheme binding environment - however, these two are inseparable, given verdoheme only appears in heme oxygenase. A dedicated investigation to the heme binding environment for heme oxygenase, may therefore be warranted in future study.

Glycine is the next most frequent - it is in lower frequency, relatively, for heme-b. As with other heme molecules, it is not clear as to what the role of glycine is in binding verdoheme.

#### 2.1.3.2 Comparison with Background Amino Acid Frequencies

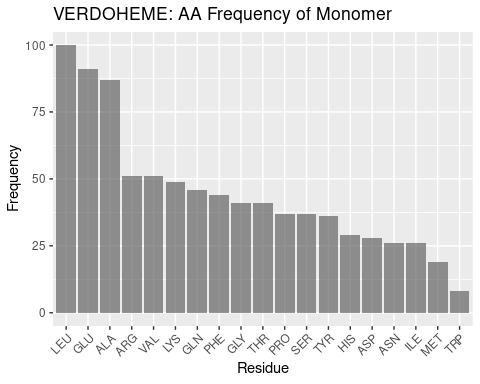


Figure 2.8: VERDOHEME: AA Frequency of Monomer

Besides the frequencies of leucine and alanine, which have been found for heme-b and heme-c above to be highly frequent in hemoproteins at large, the frequency profiles for the verdoheme binding environment and monomers is shown to be quite dissimilar, supporting the results for the binding environment as unique, not simply due to background frequency.

#### 2.1.3.3 Distributon of Amino Acids by Distance

The low sample size for verdoheme leads here to a poor figure with few residues plotted. This is likely attributable to an insufficient amount of distances and residues to cross-reference against each other, an operation that successfully occurs for all similar graphs. Regardless, the data that are plotted will be discussed.

The highly conserved histidine for hemoproteins is exclusively within 5A for verdoheme. This result again suggests that at least some of the data for verdoheme may be highly biased because of the small sample size - heme-b data included a greater range for histidine. Or, perhaps for heme oxygenases heme is solely present to coordinate the iron atom, leading to all His residues being nearby verdoheme. The close proximity of glycine to verdoheme is also unexpected and unable to be explained without further study. The remainder of the residues plotted appear to follow the trends seen in for the other types of heme, with distance values centered around 6A and unremarkable distributions.

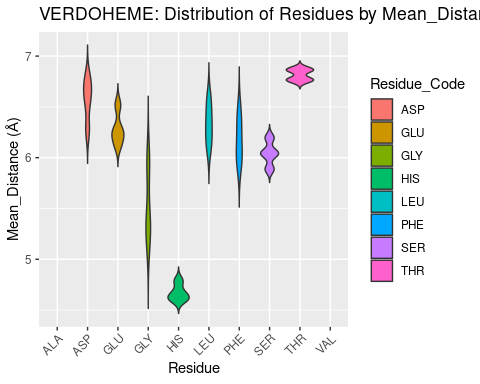


Figure 2.9: VERDOHEME: Residue Distribution by Distance

### 2.1.4 Siroheme

#### 2.1.4.1 Amino Acid Frequencies in Binding Pocket

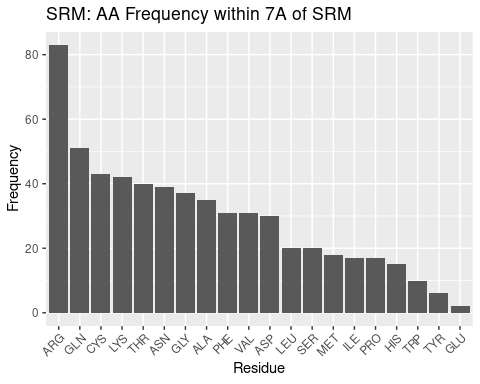


Figure 2.10: SRM: AA Frequency within 7A

Siroheme, with a structure highly dissimilar to the other heme molecules examined, should be expected to have a different amino acid frequency profile – and indeed we confirm this in our results.

Nonpolar residues are not the most abundant in the siroheme binding pocket. In fact, disproportionately frequent to the rest of the residues in the binding pocket is arginine. Siroheme is saturated with carboxyl and propionate groups; the entire porphyrin ring surrounded by polar, electronegative groups. And therefore a polar, positively charged amino acid such as arginine is reasonable to expect in the binding pocket – what is striking, however is the extreme preference for arginine; such a profile does not exist for the other types of heme. This can be at least explained; siroheme contains propionate groups that likely still form polar interactions with arginine salt bridges, and the carboxyl groups may also form polar interactions with arginine, therefore highly favoring arginine’s presence in the binding pocket.

Arginine is followed by other polar amino acids: glutamine, cystine, lysine, threonine, and asparagine; a more homogenous trend than seen for the other heme molecules, in that the trend is not interrupted by residues of opposing polarity as for other types of heme. Though these results could be expected, they demonstrate the extent to which siroheme’s binding pocket is dominated by polar residues. The preference for arginine out of all polar amino acids may be attributed to its positive charge, and ability to form salt bridges that interact with the propionate groups; lysine also has a positive charge and is polar, but does not form salt bridge interactions with the propionate groups. Cysteine is used to coordinate the iron of siroheme, and while this did not significantly affect the frequency for other heme molecules, it is still possible this increases the value for cysteine for siroheme.

After this group of polar amino acids, glycine is the next most frequent. Glycine has been situated at about a median frequency for other heme molecules, so perhaps its frequency here, slightly above the median, is of note. Again, for glycine in particular, the reason for its particular frequency cannot be determined from this data, but it appears to have some role.

Finally we come to several nonpolar amino acids: alanine, phenylalanine, and valine. These amino acids define roughly the median of the frequency data. With all the polar groups on siroheme, it might be expected that only polar interactions would be desirable. However, the not minuscule frequency of these residues suggests nonpolar interactions still occur in the binding pocket; the porphyrin ring remains, as well as some methyl groups.

After these nonpolar residues the remaining frequencies do not follow a clear trend but regardless are discussed. After aspartate the remaining frequencies are considerably lower. This may be an artefact of a small sample size, or may suggest the remaining residues form, if any, far less favorable interactions with the heme.

Aspartate appears next most frequently; it is a polar, negatively charged amino acid (at pH 7). Siroheme is saturated with other electronegative groups; perhaps there is some repulsion between these groups and aspartate – this could explain why, despite being a polar residue, arginine does not appear very frequently in the binding pocket.

Leucine is the first of the residues of diminished frequency. It is nonpolar. It, and, skipping a frequency, methionine, isoleucine, and proline, appear less frequently, and therefore are likely disfavored from forming the relatively few nonpolar interactions that do occur. Why is not clear - other small, nonpolar residues, and other lengthy nonpolar residues appear in the pocket in greater frequency.

Serine appears just less frequently than leucine, and in this context may likely be considered a polar residue that is not as strongly polar or positively charged and therefore less preferred to include in the binding pocket to form polar interactions with siroheme as other residues.

Histidine appears quite infrequently. As with siroheme, other, more strongly polar and perhaps less bulky residues are likely preferred.

Tryptophan is the least frequent nonpolar residue. The presence of a weak hydrogen bond and its size may preclude its inclusion in the binding pocket in lieu of more uniformly nonpolar residues that take up less space and can better form interactions with the few nonpolar areas of siroheme.

Tyrosine and glutamate are the least frequent polar residues. This is in stark opposition to the other heme molecules - tyrosine seemed to be favored for other heme molecules to form interactions with the propionate groups. Glutamate is also extremely infrequent, even in spite of its similarity to aspartate. Both are electronegative at pH 7 - glutamate’s extra carbon may provide sufficient steric hindrance to render it less favored. In either case, the infrequency of these residues and the tendencies of other, more intensely polar or nonpolar amino acids to be more populous, suggests tyrosine and glutamate, in the siroheme binding environment, do not interact strongly enough to be favored over other polar residues.

#### 2.1.4.2 Comparison with Background Amino Acid Frequencies

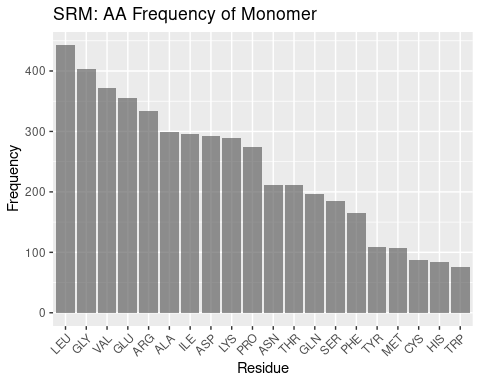


Figure 2.11: SRM: AA Frequency of Monomer

Compared to the other heme molecules, siroheme’s binding pocket amino acid frequencies are even more different than the background frequencies. Arginine is far and away the most frequent amino acid in the binding pocket; leucine is the most populous amino acid in the monomer overall, seeming to follow a trend amongst the hemoproteins examined so far. Again, discussing the remainder of the frequencies of the monomer would be conjecture, but it is worthwhile to note that the pocket frequencies appear unique against the background.

#### 2.1.4.3 Distributon of Amino Acids by Distance

Residues appear less uniformly distributed over distance for siroheme binding pockets when compared against the distribution for other heme molecules. Cysteine is the only residue that comes within 5A of siroheme; it is used to coordinate the iron in siroheme, so this result is expected. The lack of other residues being within 5A, differing from other heme molecules, suggests the many carboxyl and propionate groups on siroheme prevent, or preclude the need for closer interaction except for coordinating residues.

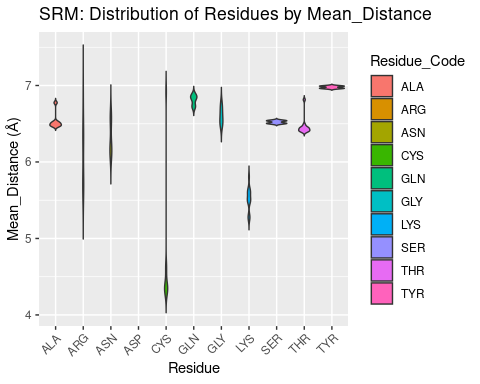


Figure 2.12: SRM: Residue Distribution by Distance

## 2.2 Volume of Heme Binding Pockets

Figures are shown below.

Volume results were rather spread out, with close agreement only found for heme-b. In general, volume for all heme molecules regardless of distance cutoff centered at approximately 1200 A³. This result may be useful in protein engineering efforts, especially for selection or design of binding pockets.

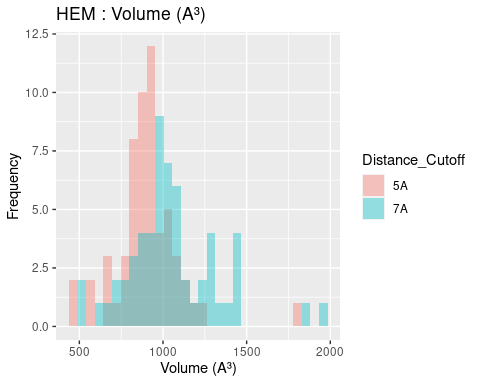


Figure 2.13: HEM: Volume of Binding Pocket

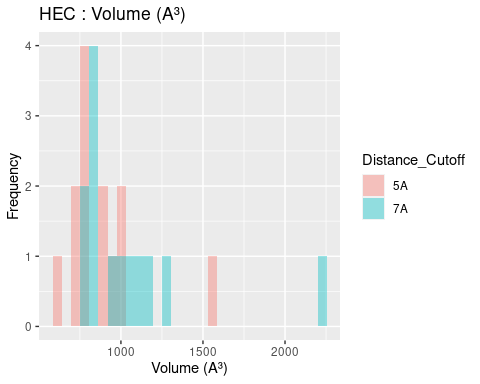


Figure 2.14: HEC: Volume of Binding Pocket

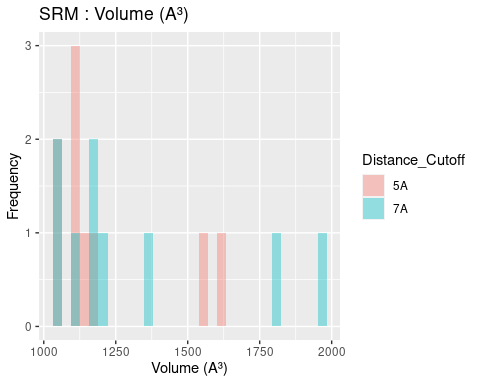


Figure 2.15: SRM: Volume of Binding Pocket

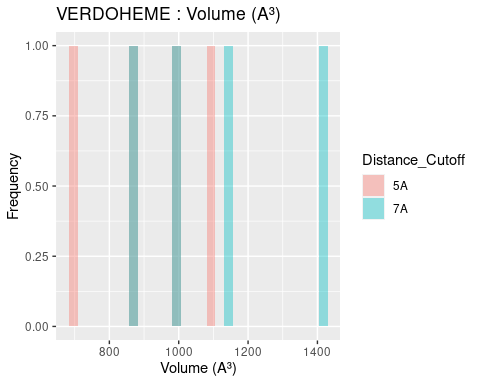


Figure 2.16: VERDOHEME: Volume of Binding Pocket

## 2.3 Surface Areas of Heme Molecules and Their Binding Pockets

### 2.3.1 Surface Area of Heme Molecules

Both solvent accessible and solvent excluded surface areas were calculated for heme molecules and binding pockets. The differences between these two measures were discussed in Section 1.3.3. The results are extremely similar for solvent accessible and solvent excluded surface areas; and therefore only solvent accessible surface area, a measure more practically interpreted into chemical phenomena, is discussed below. Figures and data for solvent excluded surface areas are available in Appendix 4.2, 4.3, and 5.4. Solvent accessible surface area figures are shown below; data tables are available in Appendix 5.4.

The solvent accessible surface area for all heme *molecules* themselves centers around values of 1000 A². This result is reasonable, given the similarity in size and structure of all heme molecules, in spite of the attached groups. Figures are shown below; full data tables are available in Appendix 5.4. The extreme outliers are likely artefacts of the method used to calculate surface area and potential conflicts with the method used to convert multimeric proteins to monomers.

## 2.4 Ligand Solvent Accessible Surface Area

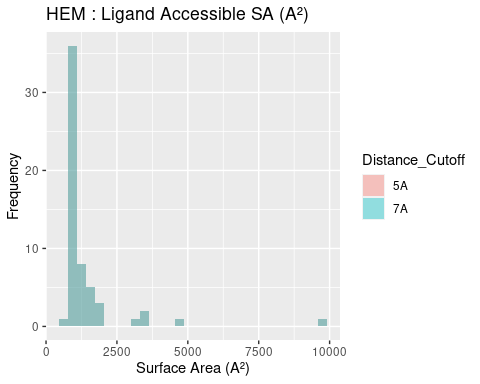


Figure 2.17: HEM: Ligand Accessible Surface Area

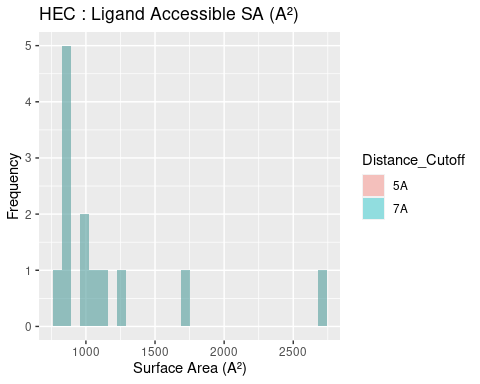


Figure 2.18: HEC: Ligand Accessible Surface Area

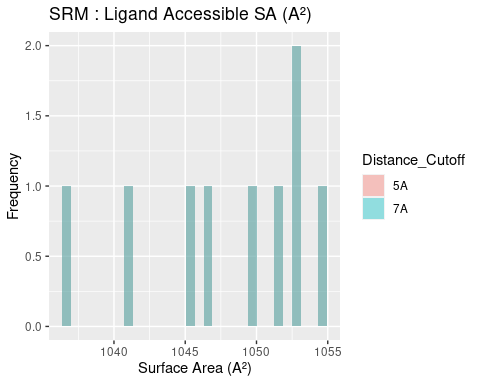


Figure 2.19: SRM: Ligand Accessible Surface Area

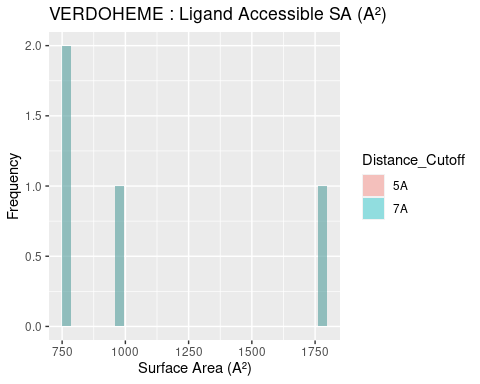


Figure 2.20: VERDOHEME: Ligand Accessible Surface Area

### 2.4.1 Surface Area of Binding Pockets

## 2.5 Pocket Solvent Accessible Surface Area

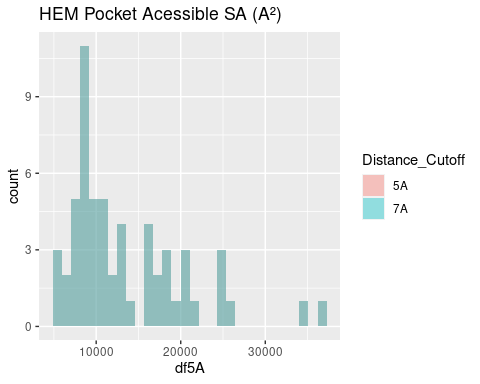


Figure 2.21: HEM: Pocket Accessible Surface Area

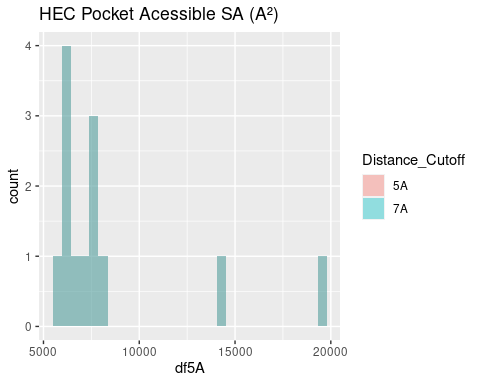


Figure 2.22: HEC: Pocket Accessible Surface Area

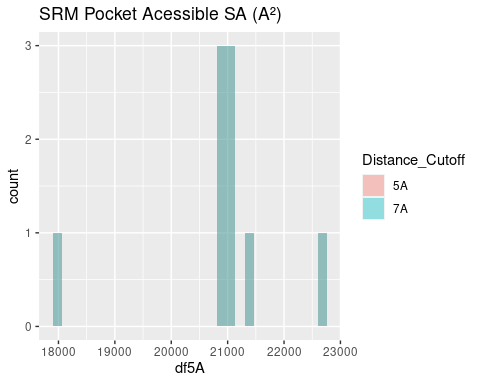


Figure 2.23: SRM: Pocket Accessible Surface Area

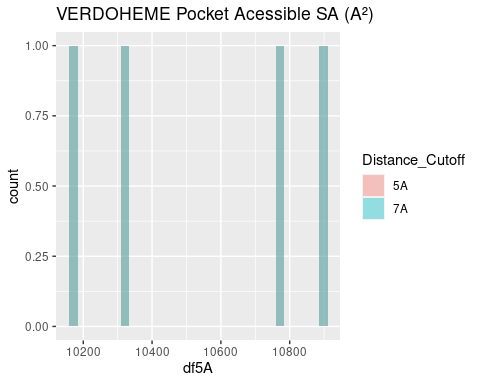


Figure 2.24: VERDOHEME: Pocket Accessible Surface Area

The surface area of binding pockets is more varied than the heme surface areas.

Heme-b and verdoheme, being highly similar molecules, with the same propionate groups, and one the derivative of the other, have quite similar surface areas, centering around 10,000-11,000 A². This is useful as a baseline to discuss the surface area of the binding pockets of the other two heme molecules below.

The surface area of the binding pocket of heme-c is considerably lower than that of heme-b and verdoheme. Its values center around 7500 A². Heme-c is bound covalently to the hemoprotein, forming thioether bonds with cysteine residues at two sites; this result suggests that the covalent bonds may exclude these sites from interacting with water molecules. Further study would be required to confirm this phenomenon.

The surface area of siroheme’s binding pocket is far greater than that for other heme molecules: values center around 21000 A². Siroheme’s extra groups on the porphyrin ring do not appear to affect its own surface area, per above. However, it is effectively a very polar molecule and appropriately the binding pocket is highly saturated with very polar amino acids, as seen in the amino acid frequency analysis. The binding pocket is therefore completely different from the other heme molecules, and these populous, polar amino acids favorably interact with aqueous solvent, negating the need to bury any hydrophobic residues and reduce surface area.

## 2.6 Angular Data

As briefly mentioned in the introduction, angular data was generated but will not be discussed extensively. Figures may be found in Appendix 4.4 and 4.5 data tables may be found in Appendix 5.5 and 5.6. Amongst the results are tight distributions of planar angles and CA-CB-Fe angles for some residues; but much of the data demonstrates a broad range of angles that may be formed. The data may be useful for protein engineering and residue placement, but cannot be productively discussed and are therefore relegated to the appendices.

## 2.7 Limitations of the Study

A high throughput framework was built to conduct this study. However, guaranteeing the quality of PDBs to enable the scripts to function properly proved challenging, and the sample size is small, although diverse. This problem only exists for heme-b and heme-c – for siroheme and verdoheme, all structures in the PDB capable of being used, were used. Heme-b and heme-c would only require more trial and error, or pre-processing, to be input to the framework that has been built.

Although many hypotheses have been suggested in the discussion to explain the data, limited experimental data exists to confirm them. Future work may include wet lab experiments to confirm these hypotheses, such as mutating several hemoproteins to contain higher or lower percentages of nonpolar residues in the binding pocket, and observing how the binding of heme is affected.

Some of these data could also be analyzed more thoroughly, for example eliminating the coordinating amino acids from the amino acid frequency data. This was not possible here due to how the framework is constructed: coordinating residues are not identified, nor is a definition proposed to identify coordinating residues. Manual input of known coordinating residues would be necessary to be certain that they could be eliminated from the final dataset analyzed, but this was beyond the scope of this study.

UCSF-Chimera was used to generate all data used in this study; many algorithms have remained unchanged for some time (surface area calculations are sourced from MSMS (1996) and volume calculations from Surfnet (1995)). It would be well to compare with any new algorithms that are developed to calculate surface area or volume, or with any experimental data that may be used confirm these numbers.

The reason being for this desired orthogonality is that the algorithms themselves may certainly introduce bias based off how they work. Surfnet generates 3D-contour surfaces to identify cavities; in practice, many small “bubbles” or insignificant cavities were generated in the study, and are filtered out during analysis – the parameters chosen can also significantly influence the behavior of the algorithm; in this study, the default parameters appeared to generate the most reasonable binding pocket. But this assessment is based off subjective visual observation by the author, and therefore introduces further bias. One may expect applying the same algorithm with the same parameters to many PDBs may at least introduce the same bias to all samples, but the algorithm may distort some PDBs more than others depending on the shape and size of their binding pocket.

# 3 Conclusion

A knowledge gap in the binding environment for heme exists in the present literature. A high-throughput framework employing UCSF Chimera was constructed to process diverse sets of hemoproteins and output information about their binding pockets: amino acid frequencies and distances from heme, volume, surface area, angles. Data was gathered and predicted from representative and varied datasets for heme-b, heme-c, verdoheme, and siroheme, and their respective hemoproteins. R was used to analyze all data.

The results of this study suggest that binding pockets for hemoproteins have some requirements for binding that may have been overlooked to date. The data and their trends observed in this study demonstrate several phenomena.

First, the heme binding environments for heme-b, heme-c, and verdoheme contain high populations of nonpolar amino acids, suggesting nonpolar interactions may be of greater importance than previously thought to providing the necessary interactions to bind heme. The binding environment for siroheme, by contrast, is shown to be extremely enriched with polar amino acids, which is not very surprising; but this binding environment also still contains many nonpolar amino acids, reinforcing the idea that the polar interactions for all heme molecules, while necessary, may be insufficient for heme binding.

Second, most of the volume data for the binding pockets of all heme molecules centers around a value of 1200 A³. Surface areas of heme-b and verdoheme binding pockets are similar, approximately 10000 A², the surface area for heme-c is less, approximately 7500 A², and for siroheme is approximately 21000 A². These values may be useful in the design of artifical metalloenzymes.

Additionally, the seeming conservation of the volume size but the variety in pocket surface areas demonstrates that while the heme molecules may be of similar size and, besides attached groups, similar structure, the attached groups will significantly affect what interactions occur in the binding pockets, and therefore the shape and exposure to solvent in the binding pockets. Siroheme is strongly polar and its binding pocket has a large surface area and is therefore highly solvent exposed, as compared to heme-b which has more nonpolar groups that must be buried and therefore requiring a smaller surface area.

Finally, angular data were generated; but the phenomena observed, such as some residues having tight ranges of angles in relation to heme or the heme iron, cannot be interpreted as useful results, except perhaps for some protein engineering efforts that may have interest in the range or distribution of possible angles for a specific residue.

These results may be useful for the rational design of hemoproteins, with the importance of nonpolar interactions in particular likely of great interest. The framework constructed for this study can be applied to any list of PDBs and their respective ligands, thereby facilitating similar research for other proteins.

# (APPENDIX) Appendix

# 4 Figures

## 4.1 AA Frequency

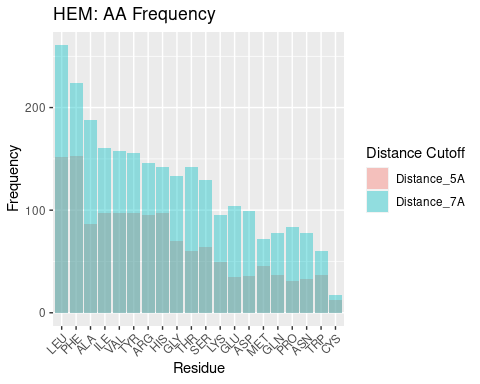


Figure 4.1: HEM: AA Frequency

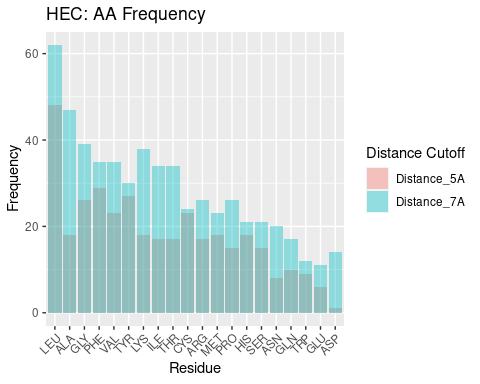


Figure 4.2: HEC: AA Frequency

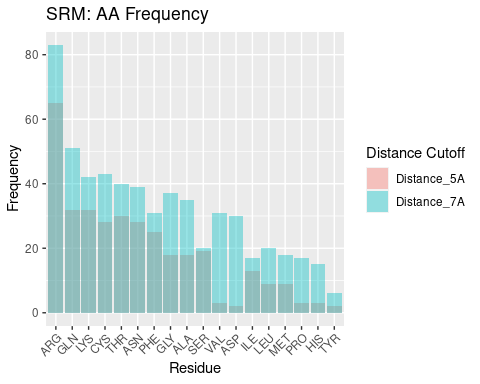


Figure 4.3: SRM: AA Frequency

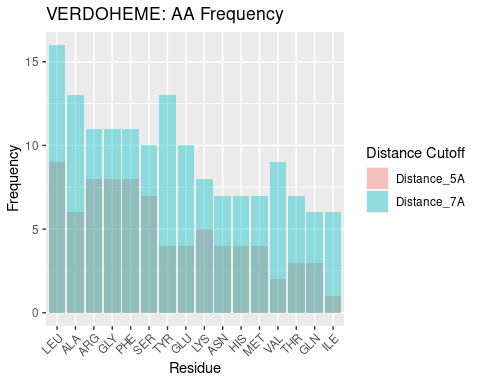


Figure 4.4: VERDOHEME: AA Frequency

## 4.2 Ligand Excluded Surface Area

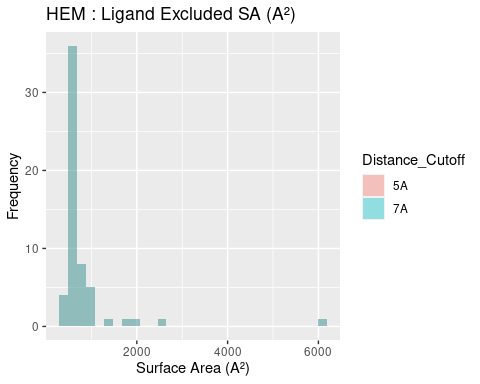


Figure 4.5: HEM: Ligand Excluded Suface Area

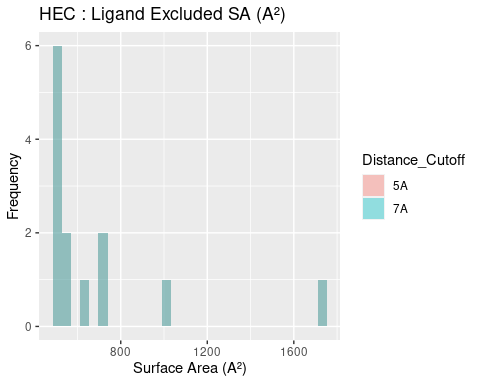


Figure 4.6: HEC: Ligand Excluded Suface Area

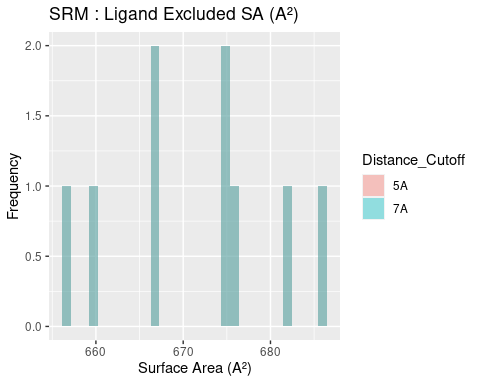


Figure 4.7: SRM: Ligand Excluded Suface Area

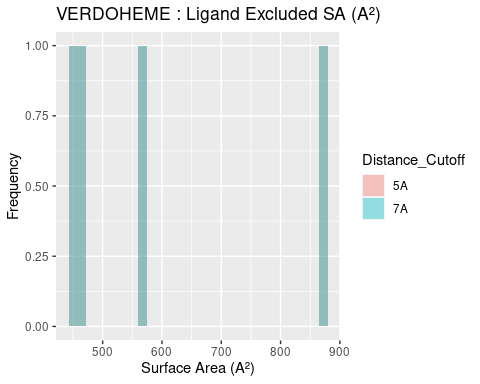


Figure 4.8: VERDOHEME: Ligand Excluded Suface Area

## 4.3 Pocket Excluded Surface Area

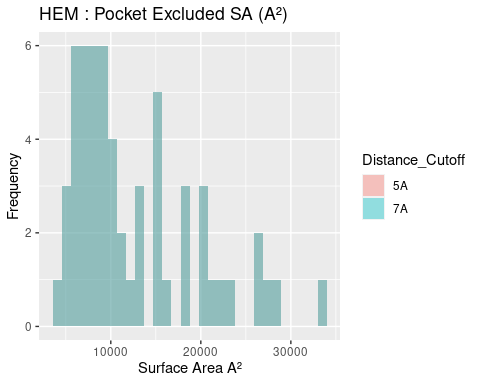


Figure 4.9: HEM: Pocket Excluded Surface Area

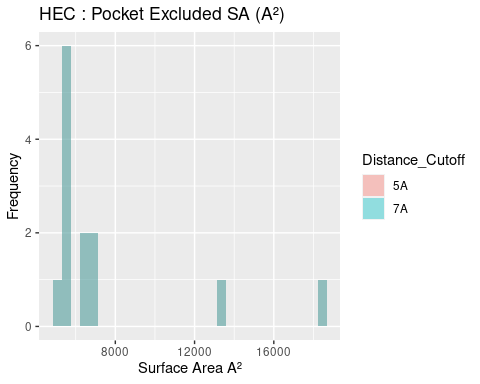


Figure 4.10: HEC: Pocket Excluded Surface Area

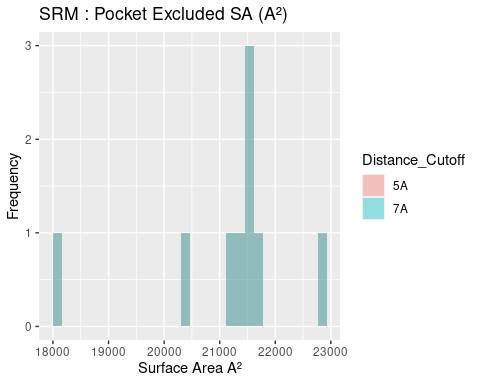


Figure 4.11: SRM: Pocket Excluded Surface Area

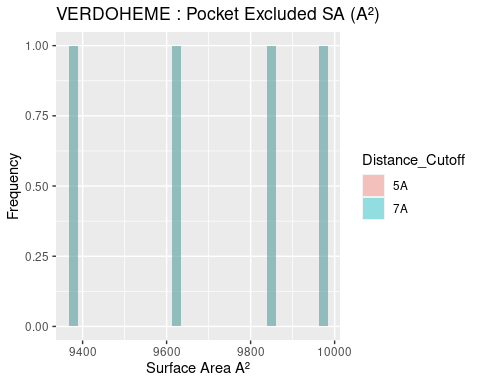


Figure 4.12: VERDOHEME: Pocket Excluded Surface Area

## 4.4 All Planar Angles

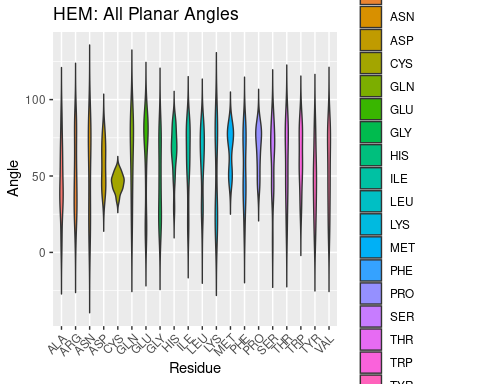


Figure 4.13: HEM: All Planar Angles

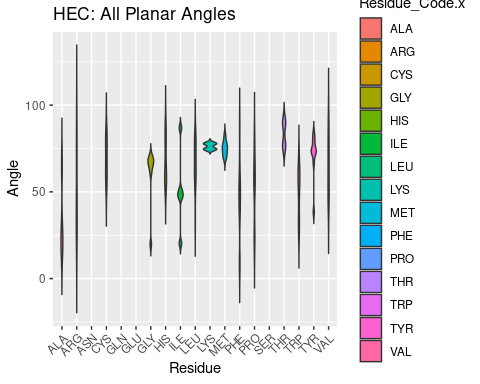


Figure 4.14: HEC: All Planar Angles

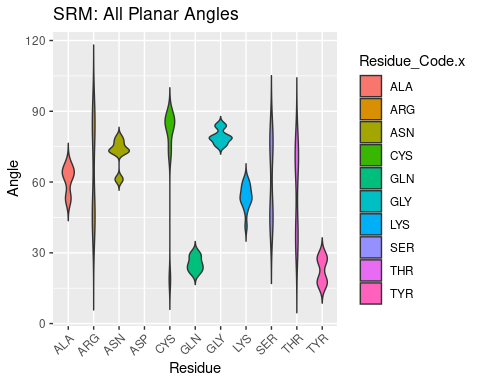


Figure 4.15: SRM: All Planar Angles

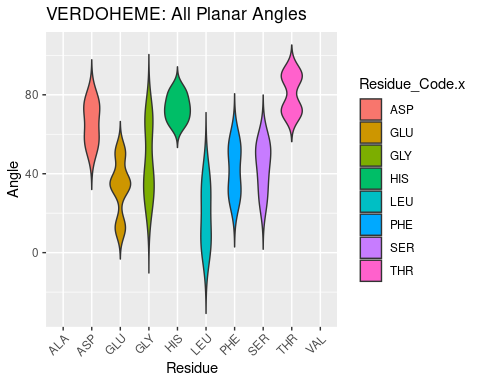


Figure 4.16: VERDOHEME: All Planar Angles

## 4.5 All CA-CB-Fe Angles

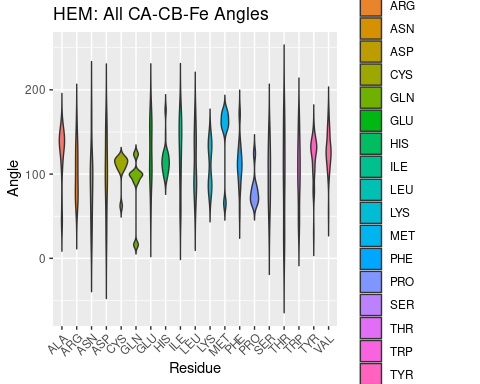


Figure 4.17: HEM: All CA-CB-Fe Angles

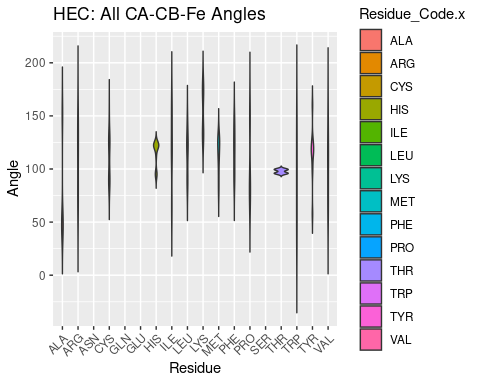


Figure 4.18: HEC: All CA-CB-Fe Angles

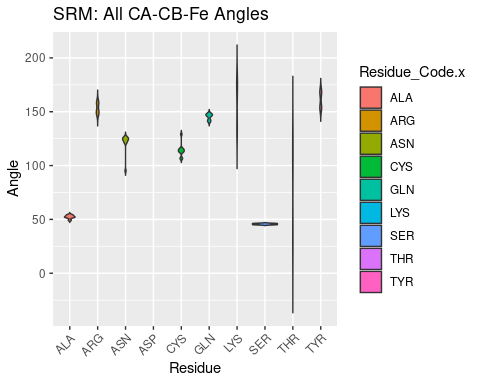


Figure 4.19: SRM: All CA-CB-Fe Angles

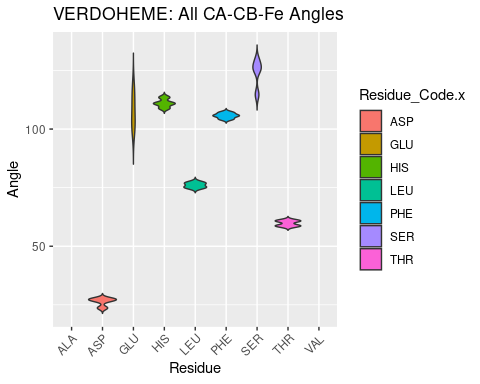


Figure 4.20: VERDOHEME: All CA-CB-Fe Angles

# 5 Tables

## 5.1 Molecule Names and Source Organisms

## 5.2 Amino Acid Frequencies at 5A Distance Cutoff

## 5.3 Distances

### 5.3.1 All Distances from Heme Fe to Atoms of Residues in Binding Pocket

### 5.3.2 Mean Distances of Each Residue in Binding Pocket

## 5.4 Volume and Surface Areas

### 5.4.1 Tables of Volume and Surface Areas, Distance Cutoff 7A

### 5.4.2 Tables of Volume and Surface Areas, Distance Cutoff 5A

## 5.5 All Planar Angles

## 5.6 All CA-CB-Fe Angles

# References

Barrows, Tiffany P., and Thomas L. Poulos. 2005. “Role of electrostatics and salt bridges in stabilizing the compound I radical in ascorbate peroxidase.” *Biochemistry* 44 (43): 14062–68. <https://doi.org/10.1021/bi0507128>.

Bowman, Sarah E. J., and Kara L. Bren. 2008. “The chemistry and biochemistry of heme c: Functional bases for covalent attachment.” *Natural Product Reports* 25 (6): 1118–30. <https://doi.org/10.1039/b717196j>.

Brânzanic, Adrian M. V., Ulf Ryde, and Radu Silaghi-Dumitrescu. 2019. “Why does sulfite reductase employ siroheme?” *Chemical Communications* 55 (93): 14047–49. <https://doi.org/10.1039/c9cc05271b>.

Finnerty, Patrick. 2021. “heme-binding.” <https://github.com/patimus-prime/heme-binding.git>.

Ishida, Manabu, Naoshi Dohmae, Yoshitsugu Shiro, Tadatake Oku, Tetsutaro Iizuka, and Yasuhiro Isogai. 2004. “Design and synthesis of de Novo cytochromes c.” *Biochemistry* 43 (30): 9823–33. <https://doi.org/10.1021/bi049546e>.

Kleingardner, Jesse G., and Kara L. Bren. 2015. “Biological Significance and Applications of Heme c Proteins and Peptides.” *Accounts of Chemical Research* 48 (7): 1845–52. <https://doi.org/10.1021/acs.accounts.5b00106>.

Lai, Wenzhen, Hui Chen, Toshitaka Matsui, Kohei Omori, Masaki Unno, Masao Ikeda-Saito, and Sason Shaik. 2010. “Enzymatic ring-opening mechanism of verdoheme by the heme oxygenase: A combined X-ray crystallography and QM/MM study.” *Journal of the American Chemical Society* 132 (37): 12960–70. <https://doi.org/10.1021/ja104674q>.

Lalonde, Jim. 2016. “Highly engineered biocatalysts for efficient small molecule pharmaceutical synthesis.” *Current Opinion in Biotechnology* 42: 152–58. <https://doi.org/10.1016/j.copbio.2016.04.023>.

Sato, Hideaki, Yuichiro Higashimoto, Hiroshi Sakamoto, Masakazu Sugishima, Kenichi Takahashi, Graham Palmer, and Masato Noguchi. 2007. “Electrochemical reduction of ferrous -verdoheme in complex with heme oxygenase-1.” *Journal of Inorganic Biochemistry* 101 (10): 1394–99. <https://doi.org/10.1016/j.jinorgbio.2007.05.016>.