1 Results and Discussion

–>

1.1 Analysis of Nearby Residues of Natural Porphyrins

The first part of the study aimed at providing statistics on the amino acid propensity to interact with hemes in natural proteins. We studied heme-b, heme-c, siroheme and verdoheme. Because we are not looking only at the iron environment, but instead at the environment of the entire microcycle, we did the analysis for any amino acid with potential contact with the heme. This was defined as any AA having at least one atom within the cutoff distances of 5 and 7 Angtroms (A).

Amino acid frequencies were obtained for distance cutoffs of 5A and 7A - these figures and data are shown in **FIXME ADD APPENDICES LATER** The trends in these data are very similar and therefore only the data pertaining to the 7A distance cutoff are discussed below.

1.2 AA Frequency

1.2.1 Heme-b

Amino Acid Frequencies in Binding Pocket

Figure 1.1 plots the frequency of each residue within 7A of heme-b. Immediately below is Figure 1.2, which plots the frequency of each residue within the entire monomer. **I'm thinking put tables in back they dont add much value here** Data tables are available in Appendix...

We gotta fuck ton of shit to discuss! Don't try to keep this short homie. You've put in WORK.

Beginning at the left of the figure 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/environment with the pyrole groups and methyl and vinyl groups.

Note hisitidine is highly evolutionarily conserved to coordinate Fe, and it may be affected by the molecules we've chosen based on what they used to coordinate - I don't want to examine that right now though so moving on Tyrosine, arginine, histidine appear next most frequently. The two propionate groups on heme are used to coordinate the heme in the binding pocket. These polar residues are therefore likely interact with the propionate groups, providing the polar interactions necessary, in addition to the nonpolar interactions above, to provide as hospitable of a binding environment as possible to coordinate the heme. In additon, the arginine and histidine groups are positively charged (at pH 7), **further facilitating interaction with the electronegative proprionate groups on the porphyrin ring of heme-b.** GOTTA FACT CHECK THIS HOMIE-YO.

Glycine is a small residue and cannot form significant interactions within its environment; however, its frequency, or lack thereof (compared to background frequency, discused later), suggests the binding pocket may not require as much flexibility or... spatial considerations as in the rest of the protein. This would logically follow from the need for conserved binding sites.

Serine glu asp* 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 polarpositively charged. The negative charge is unlikely of importance in interaction with heme-b, however these polar amino acids likely again interact with the propionate groups on heme; only, infrequently. What is most interesting is why lysine is in such low abundance relative to the other polar, positively charged residues, arginine and histidine. Perhaps lysine's fairly linear structure prevents it from fitting into the binding pocket; however, arginine is also somewhat linear and features prominently. The exact reason for why this could be is beyond the scope of this study.

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, capable of stronger interactions with the electronegative propionate groups.

met, trp Methionine and tryptophan appear very infrequently in the binding pocket. All nonpolar amino acids already mentioned do not possess a sulfur (thio something?) bond in their structure; perhaps it is less favorable for the sulfur atom to interact with the porphyrin ring than another carbon. 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 taking up less space.

Cystine appears most infrequently of all the amino acids in the binding pocket. This is quite surprising - cystine is highly evolutionarily conserved to coordinate the iron in the binding pocket. 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. 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. **also need to talk about this with his**

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; quite logical, given there are only two polar propionate groups on a large porphyrin ring that is otherwise nonpolar. 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 intercations and the population of nopolar residues in the binding pocket should be considered when examining the binding environment of heme.

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 this may affect the significance of the frequencies of the amino acids in the binding pocket is unclear - those amino acids are still employed and placed such as

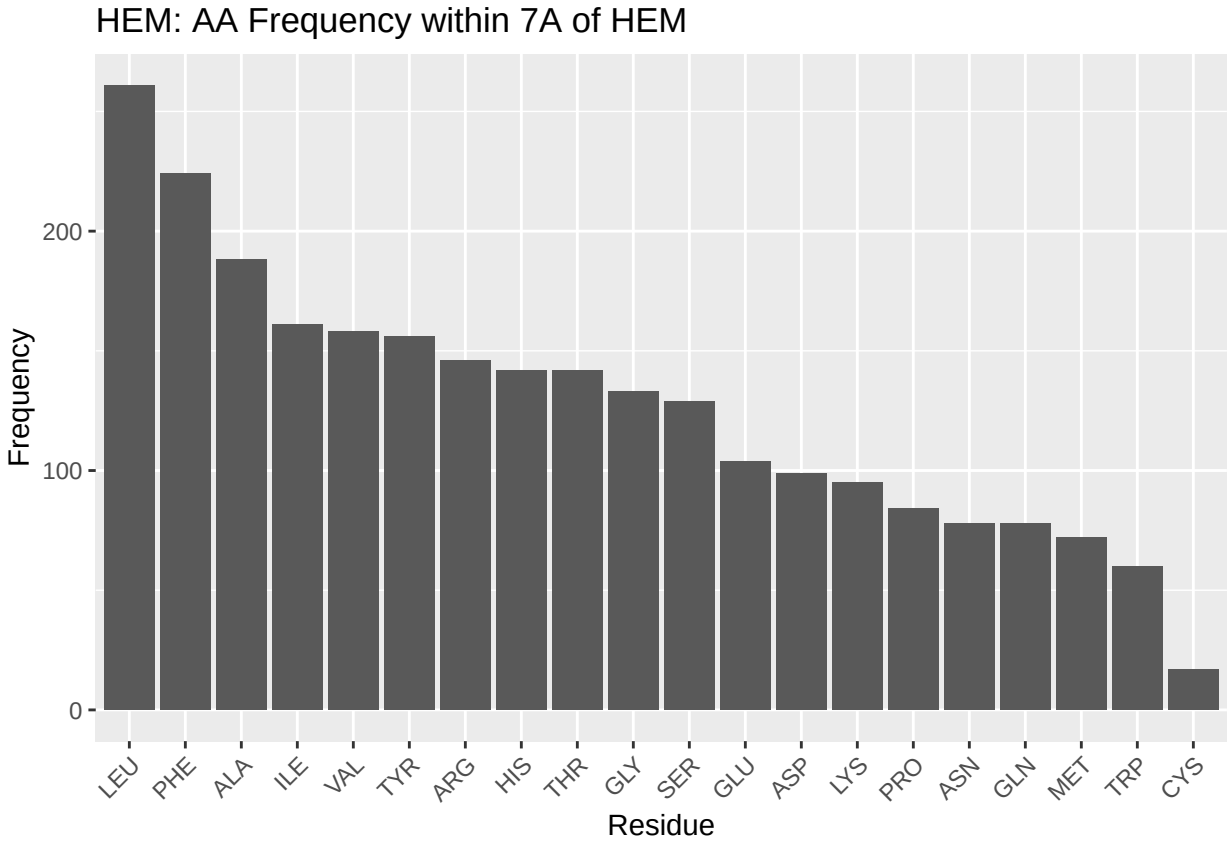

Figure 1.1: HEM: AA Frequency within 7A

to bind the heme, rather than being a random assortment of residues. However, a in depth examination of simlarities 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 **??** 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 pulation 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

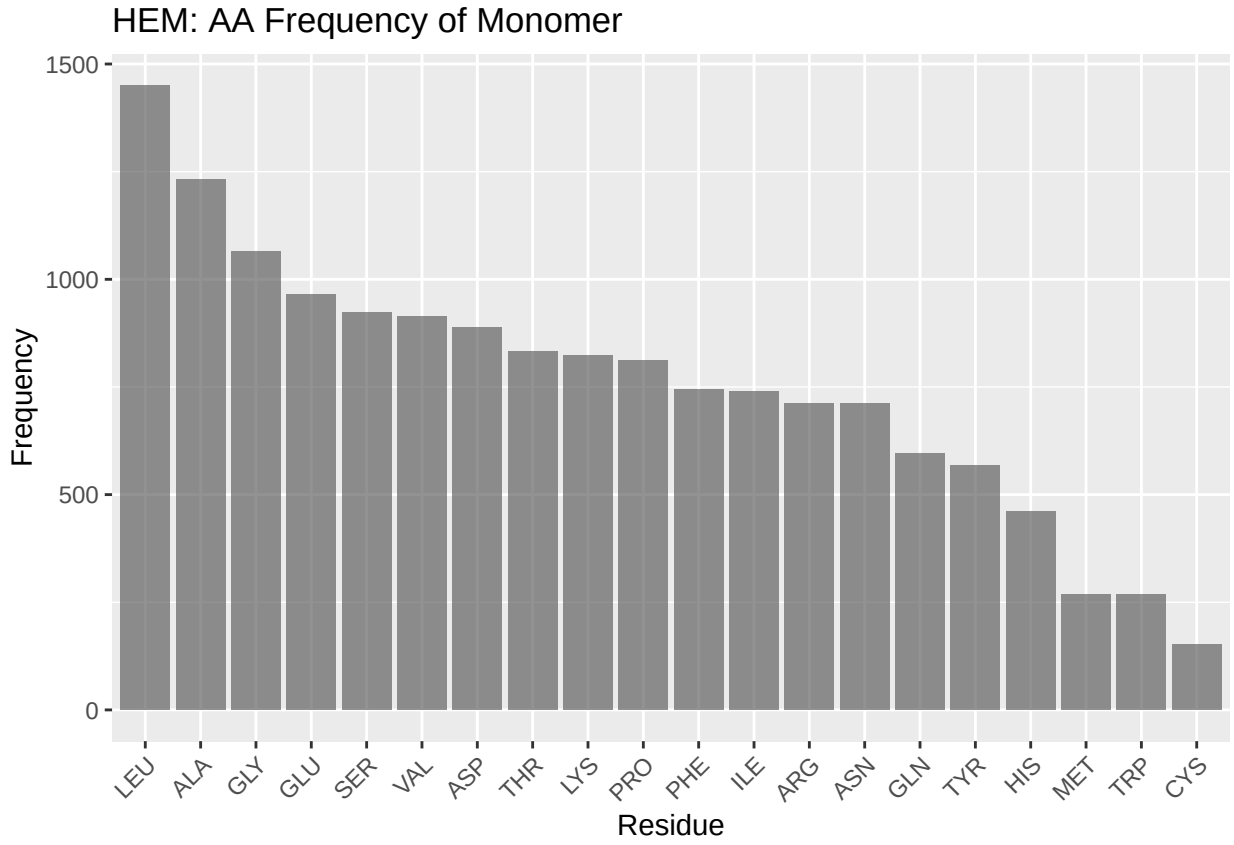

Figure 1.2: HEM: AA Frequency of Monomer

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.

Table 1.1: HEM AA Freq

Residue	Freq
LEU	261
PHE	224
ALA	188
ILE	161
VAL	158
TYR	156
ARG	146
HIS	142
THR	142
GLY	133
SER	129
GLU	104

Table 1.1: HEM AA Freq *(continued)*

Residue	Freq
ASP	99
LYS	95
PRO	84
ASN	78
GLN	78
MET	72
TRP	60
CYS	17

Distributon of Amino Acids over Distance

After an exhaustive exploration of the relative frequencies of amino acids in the binding pocket, the figure below likely appears fairly straightfoward. Figure **??** 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: Cys, His, Tyr. Most residues center their distribution at around 6A, although Lys seems more biased than the remaining residues to be a bit closer. Cystidine and histidine may be at least in part explained to be close due to their use as coordnating 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 lkely more notable. It cannot form coordination bonds with the heme iron, but tyrosine residues do interact with the propionate groups; and 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 extreme imortance of tyrosine to interact with propionate groups, or instead the need for tyrosine to be in close proximity in order to form such interactions, is beyond the scope of this study.

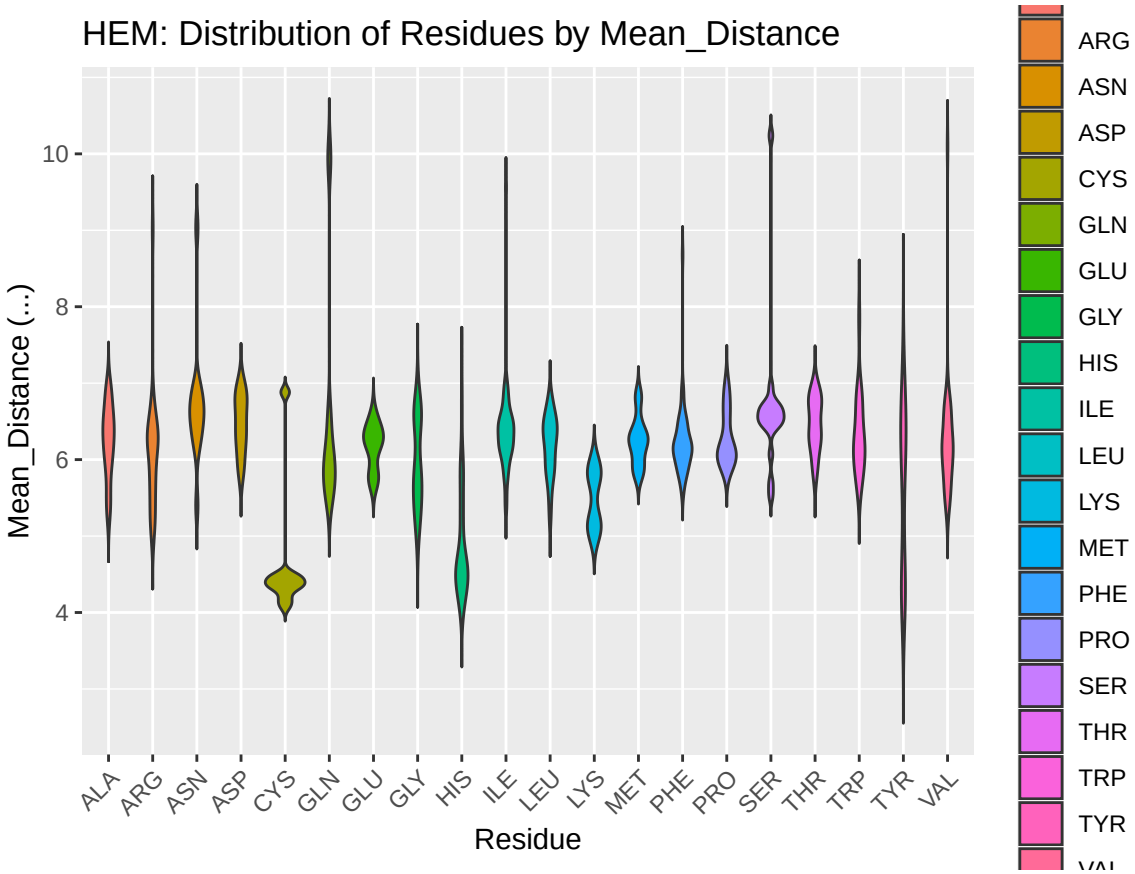

Figure 1.3: HEM: AA Distances

1.2.2 Heme-c

Leucine and alanine again are highly frequent for HEC, followed by quite similar trends and therfore 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 covalently binds to CYS, and this may explain that frequency: but as for glycine, perhaps the covalent binding of CYS is sufficient for other interactions to be of lower priority, and the flexibility and shape of the binding pocket to be prioritized, therefore leading to more glycines being included in order to shape the pocket favorably. **I feel like this is grasping at straws without much support, add more qualifiers or just leave it at not discussing much**

Generally, heme-c mirrors heme-b in a strong showing of alanine and leucine, followed by a divergence in the frequency of amino acids and therefore a struggle to

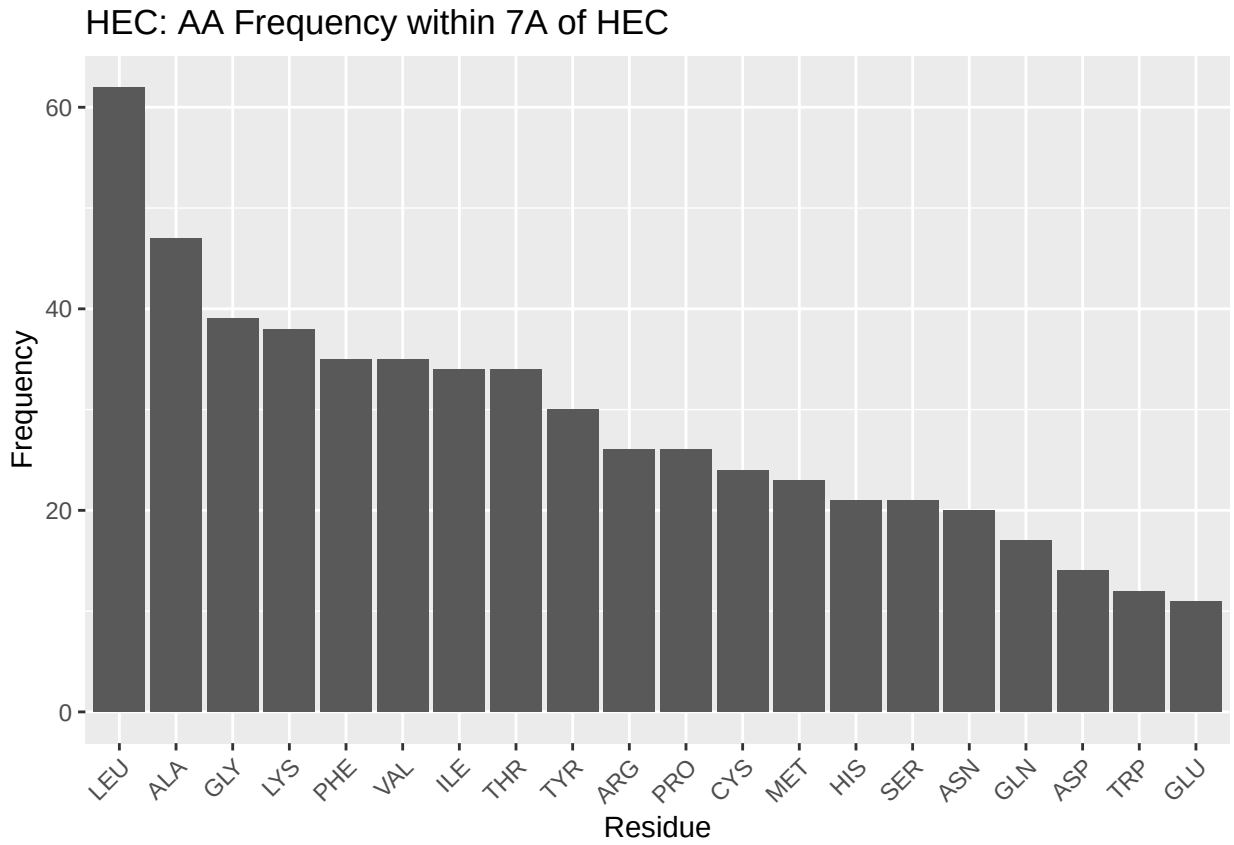

Figure 1.4: HEC: AA Frequency

form any meaningful discussion when it comes to comparing the binding pocket frequencies against background frequencies.

Table 1.2: HEC AA Freq

Residue	Freq
LEU	62
ALA	47
GLY	39
LYS	38
PHE	35
VAL	35
ILE	34
THR	34
TYR	30
ARG	26
PRO	26
CYS	24
MET	23
HIS	21

Table 1.2: HEC AA Freq *(continued)*

Residue	Freq
SER	21
ASN	20
GLN	17
ASP	14
TRP	12
GLU	11

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 covalently bonds to HEC's iron atom, providing further justification for its proximity. However, for HEC, Lys and Met also are very proximal. The methionine residues... for thioester bonds?? **HELP** Lysine is polar and positively charged, and therefore in the case of HEC being covalently bonded, and therefore reliably, specifically, positioned, the environment may be such that a lengthy, polar, charged residue may also be positioned well enough to be consistently nearby and forming interactions with the propionate groups. **review this one home dawg**

1.2.3 Verdoheme

Table 1.3: VERDOHEME AA Freq

Residue	Freq
LEU	16
ALA	13
TYR	13
ARG	11
GLY	11
PHE	11
GLU	10
SER	10
VAL	9
LYS	8

Table 1.3: VERDOHEME AA Freq *(continued)*

Residue	Freq
ASN	7
HIS	7
MET	7
THR	7
GLN	6
ILE	6
ASP	4

1.2.4 Siroheme

Table 1.4: SRM AA Freq

Residue	Freq
ARG	83
GLN	51
CYS	43
LYS	42
THR	40
ASN	39
GLY	37
ALA	35
PHE	31
VAL	31
ASP	30
LEU	20
SER	20
MET	18
ILE	17
PRO	17
HIS	15
TRP	10
TYR	6
GLU	2

1.2.5 Heme-c

Interacts with the propionates, the polar groups within the porphyrin - 1.

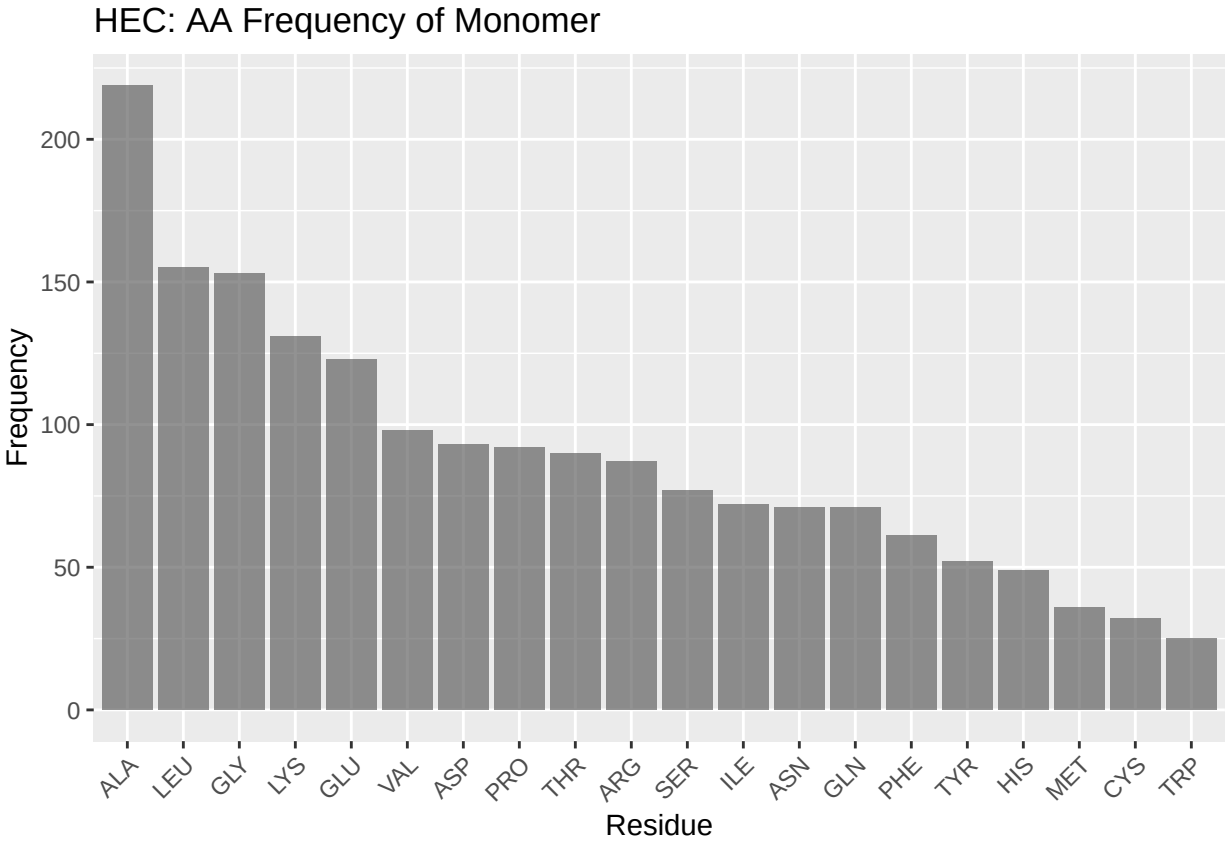

Figure 1.5: HEC: AA Frequency of Monomer

Thus, we confirm with (**Li2011**) that the binding site for heme-b, and now additionally heme-c, is enriched with nonpolar amino acids. This strongly suggests that the nonpolar amino acids are crucial to the binding of heme, likely due to the favorable interaction between the nonpolar amino acids and the porphyrin ring of heme-b and heme-c. The degree to which the percentage of nonpolar acids in the binding pocket may affect the binding of heme cannot be determined in this analysis. But the agreement with the residues closest to heme (and therefore likely to coordinate) is discussed below. ### Siroheme

The binding pocket for siroheme is highly enriched with arginine, in addition to other charged residues such as lysine and cysteine. Given the function of siroheme is to transfer multiple electrons per reaction, the enrichment of the binding environment with charged residues is perhaps not surprising. Addtionally, for siroheme in particular, the geometry, the placement of these residues in order to

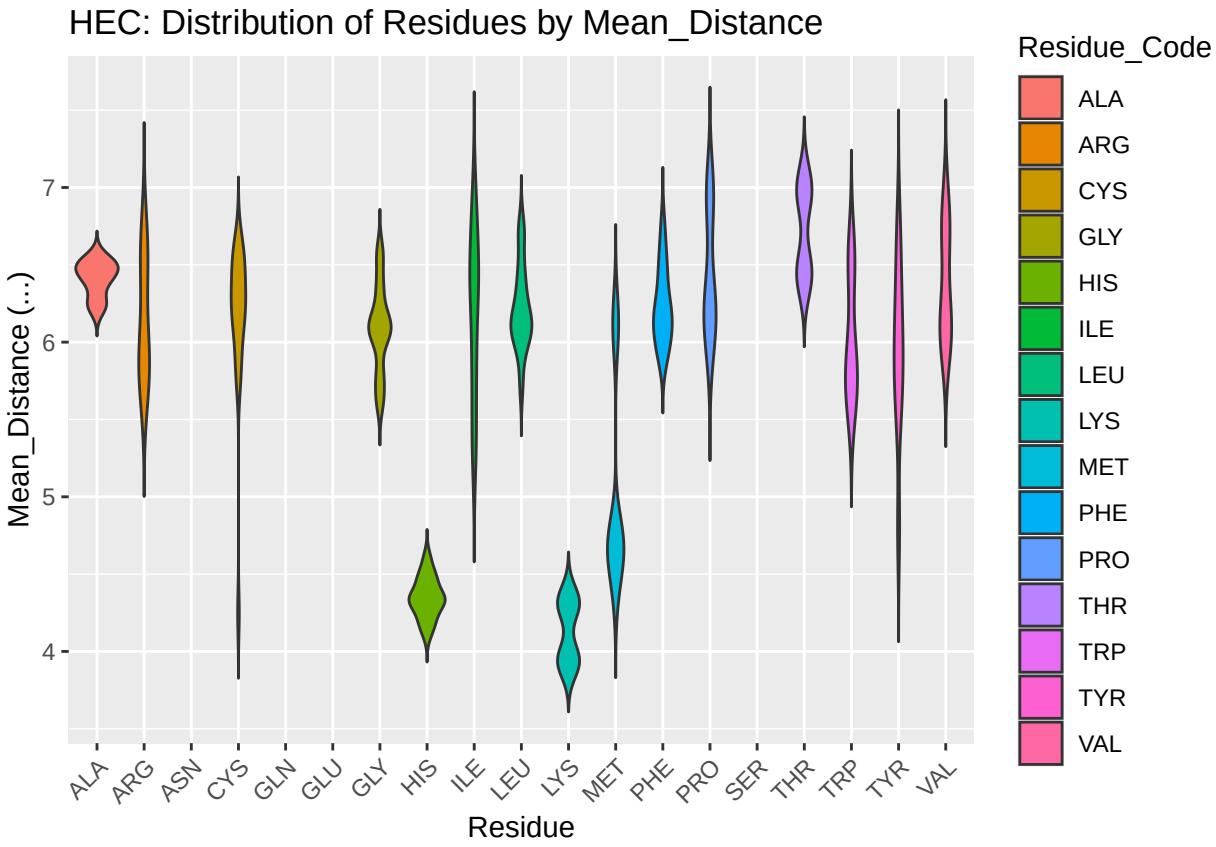

Figure 1.6: HEC: AA Distances

coordinate the movement of electrons may be of interest; however this is beyond the scope of this study.

1.2.6 Verdoheme

Verdoheme is an intermediate formed within the binding pocket of heme oxygenase during the degradation of heme; this is the same environment that binds heme at the beginning of the reaction. Verdoheme is oriented differently in the pocket, however overall the amino acid composition of the pocket should be quite similar if not the same. But this is not reflected in our results. Rather, relative to the amino acid... profile (?) of heme-b and heme-c, arginine is in greater abundance in the binding pocket. Leucine, alanine and phenylalanine are still amongst the most frequent amino acids, again providing nonpolar interactions.

These results may suggest that verdoheme is reoriented, and potentially spatially relocated, to interact with more arginine residues than would be possible if the

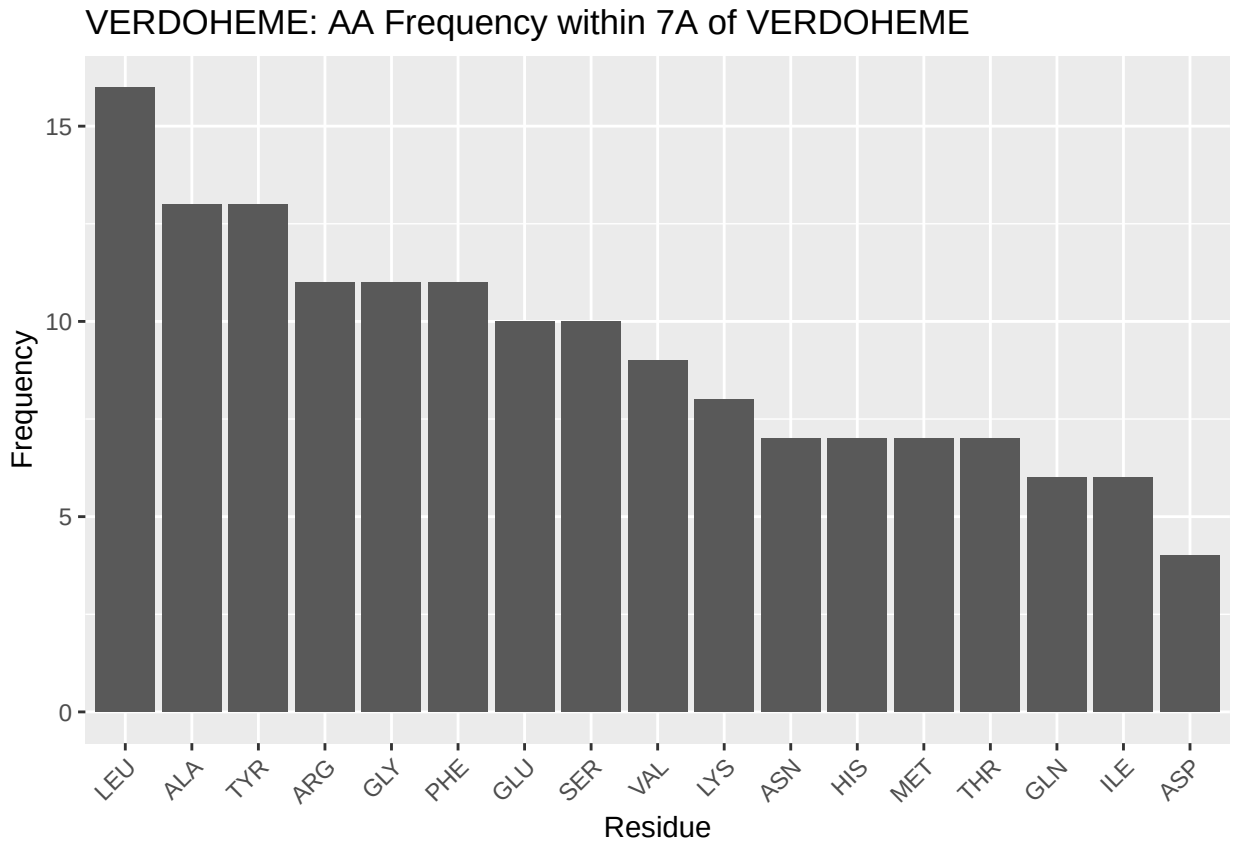

Figure 1.7: VERDOHEME: AA Frequency

original heme orientation were maintained; the amount of space verdoheme would have to move within the binding pocket is discussed below. However, it is also possible, likely, this is an artefact of the very small sample size of verdoheme PDBs used in this study (n=4). Upon the availability of additional verdoheme-containing PDBs in the future this result may be reexamined.

1.3 Volume Discussion

Figures can be found in Appendix **??**..

Besides HEM, much of the volume data for the different ligands did not closely agree between different distance cutoffs. The degree to which the method used to determine the volume of the binding pocket, Surfnet as implemented in UCSF Chimera, is affected by the distance cutoff is therefore demonstrated to be quite great. Most of the data did however fall within reasonable bounds, regardless

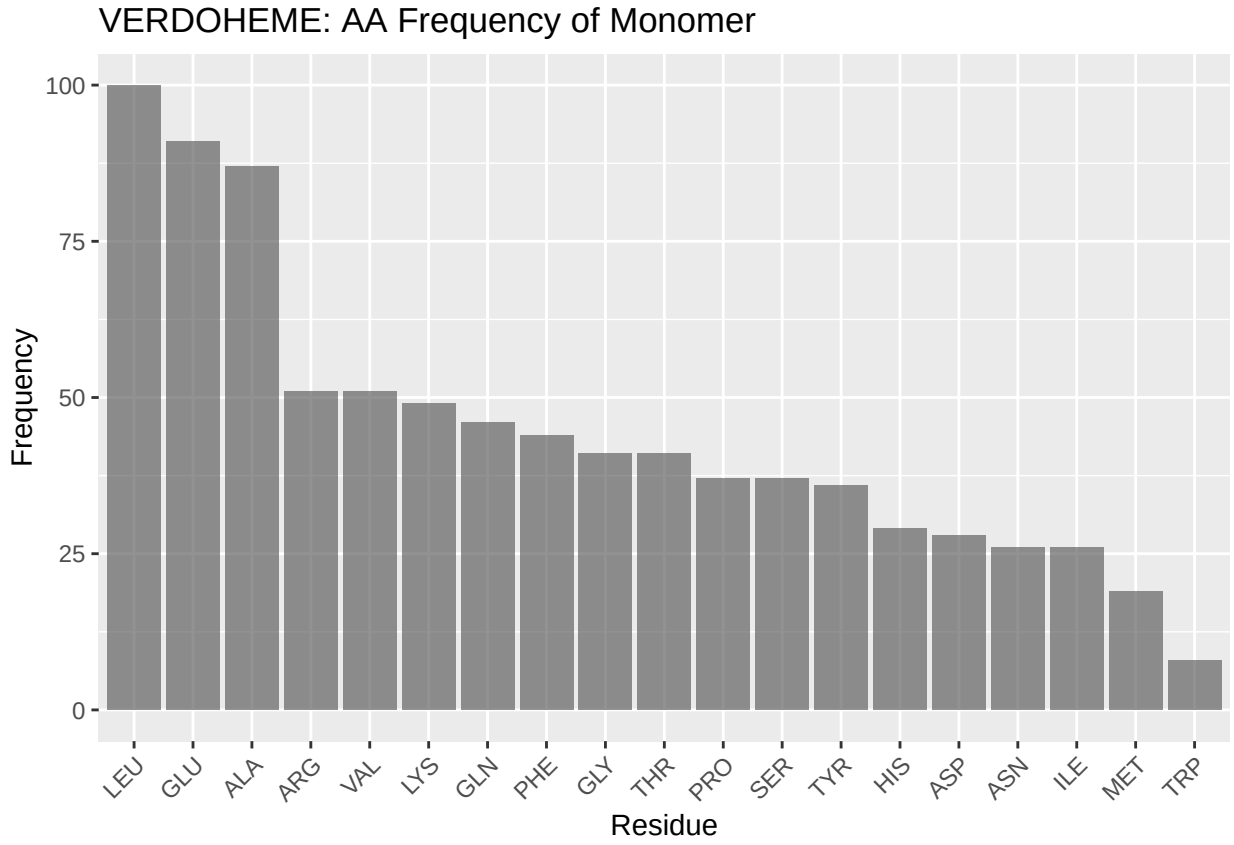

Figure 1.8: VERDOHEME: AA Frequency of Monomer

of distance cutoffs.

Data for heme-b, heme-c and siroheme somewhat agree within a tight range of volumes, with heme-b in particular showing the best agreement of data at approximately 800-1200A^3. However, the *range* of the volume data, for all ligands, is considerable, approximately 1000A^3.

There are then at two conflicting conclusions from these data.

The the volumes of the binding pocket, as determined by calculation, are shown to not require uniform size. This suggests that the binding pocket need not be a perfect fit for each protein's respective ligand. As far as implications for protein engineering, these results suggest that de novo design of a hemoprotein may not require significant work as far as sizing of a binding pocket.

Alternatively, the results demonstrate the difficulty of calculating molecular volumes, and how error-prone at least this method may be. Comparison with

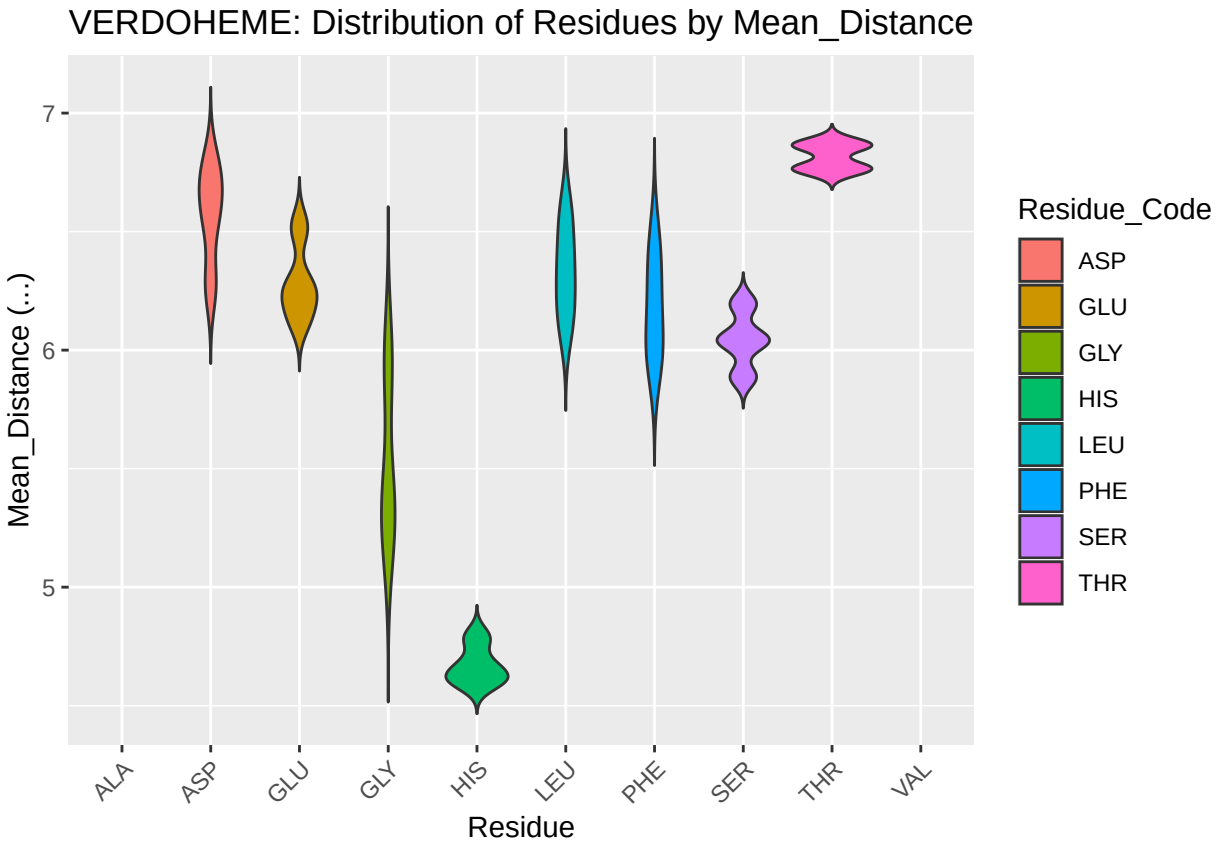

Figure 1.9: VERDOHEME: AA Distances

other methods of calculating molecular volumes, or with experimental data, would be warranted in future study.

also need to highlight differences in excluded v. accessible ## Surface Area of Ligands Figures can be found in Appendix **??** for ligands' solvent excluded surface area, and in Appendix **??** for ligands' solvent accessible surface area.

This measure, and also the solvent accessible surface area of the ligands, is effectively of the surface area of each ligand itself - distance cutoffs and binding environments are not expected to affect the results. Significant variability in the data may therefore suggest problems in calculation.

Both accessible and excluded surface area data are fairly well centered around some value for all ligands, except for verdoheme. For the latter this may be attributed to the small sample size. But for other ligands, and also for verdoheme, the outliers in the data may suggest errors in calculation and potential inclusion of parts of the binding pocket itself in calculation.

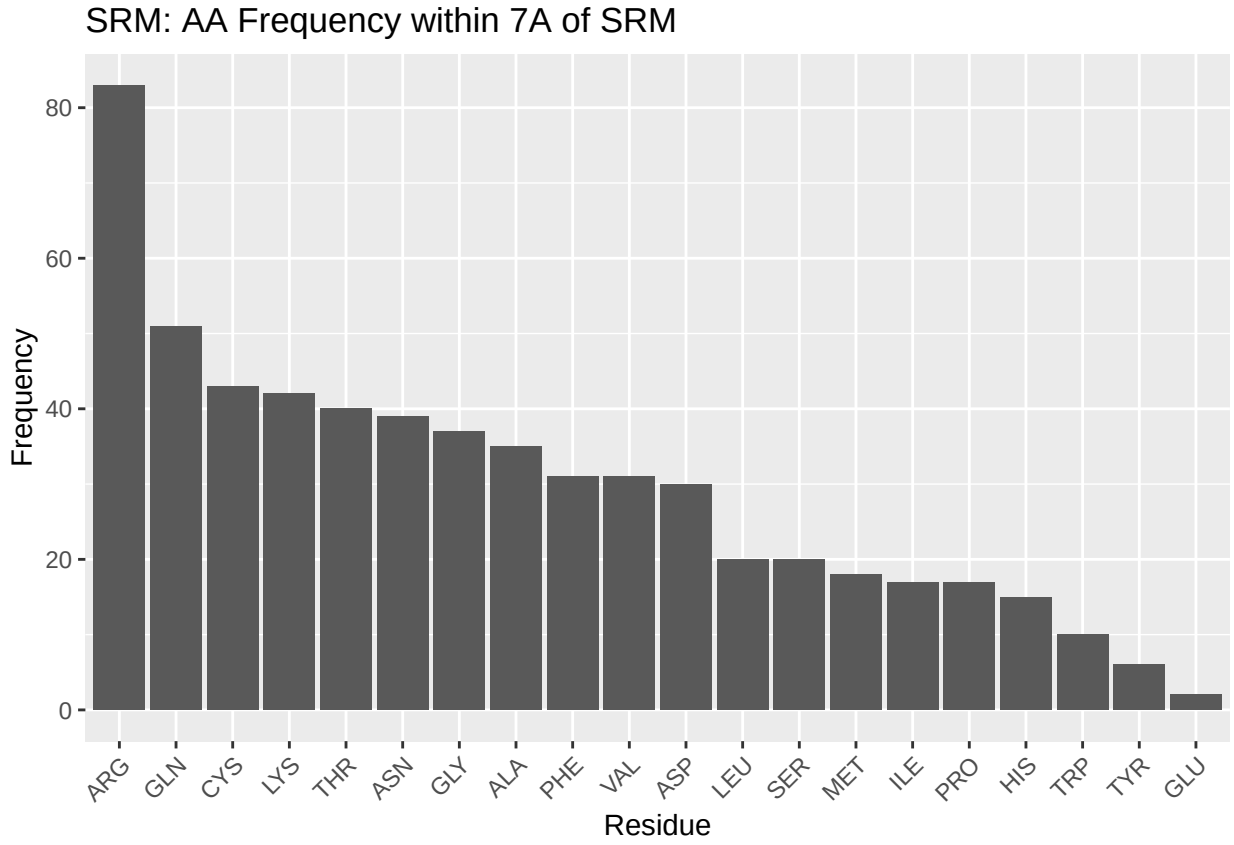

Figure 1.10: SRM: AA Frequency

1.4 Pocket Solvent Excluded Surface Area

We'll do same treatment for pocket surface areas, skipping ahead to angles/distances

Figures can be found in **??**

Figures can be found in Appendix **??**.

1.5 All Planar Angles

Figures can be found in **??**

These data, for all ligands, except potentially for heme-c, largely serve to compare as noise for the next section. The planar angles of all residues, falling within the upper distance cutoff of 7A, are plotted.

In the notable exception of heme-c (HEC), Figure **FIXME: insert figure name**

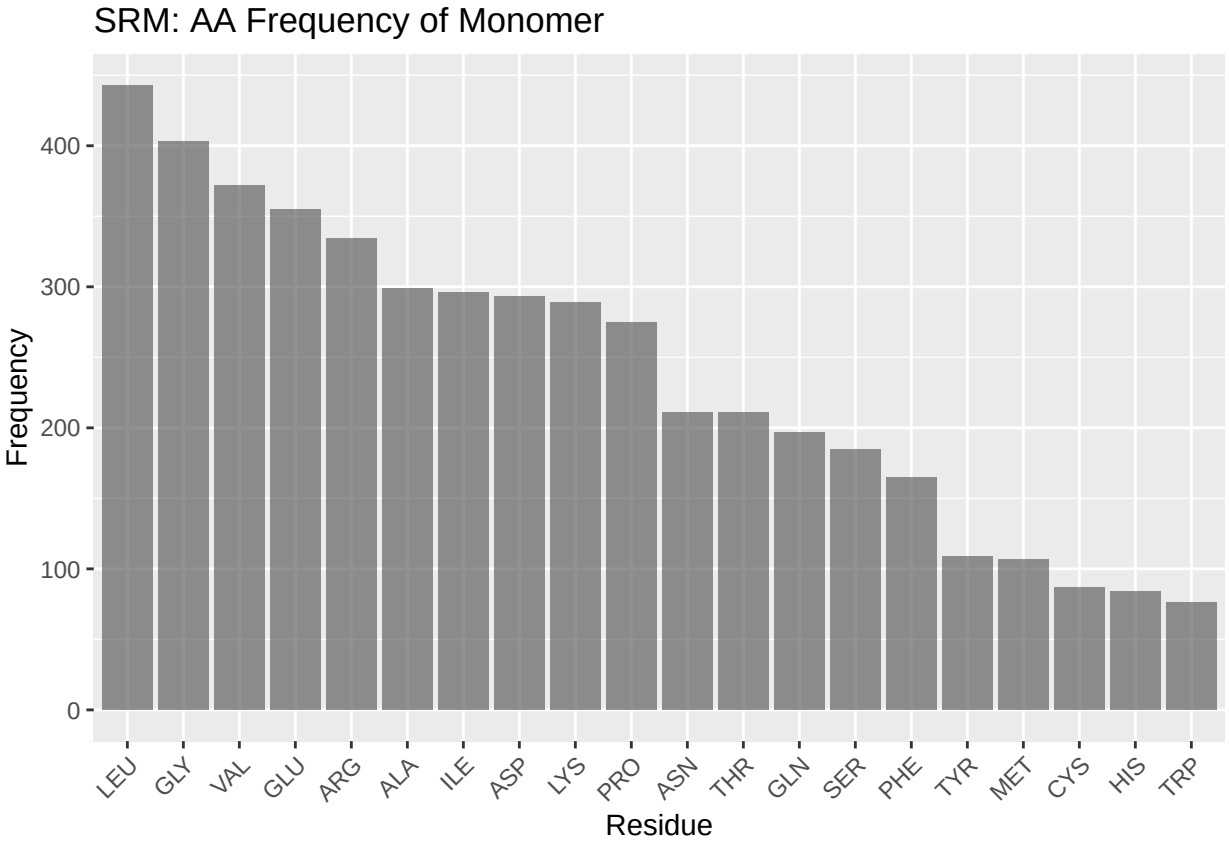

Figure 1.11: SRM: AA Frequency of Monomer

for this ligand seems to suggest that GLU, MET and LYS have fairly specific planar angles with the ligand. However, heme-c is not coordinated to achieve its chemistry but rather covalently bound, to histidine. Lysine is effectively the median of amino acid frequency for heme-c, methionine is even less frequent and glutamine is the least frequent amino acid. For the latter two amino acids their tight range of planar angles is therefore likely an artefact of a small sample size of amino acids. However, for lysine the tight range of angles may be significant; this is dicussed further below.

1.6 Planar Angles of Closest Residues

Figures can be found in **??**

1.7 All CA-CB-Fe Angles

Figures can be found in **??**

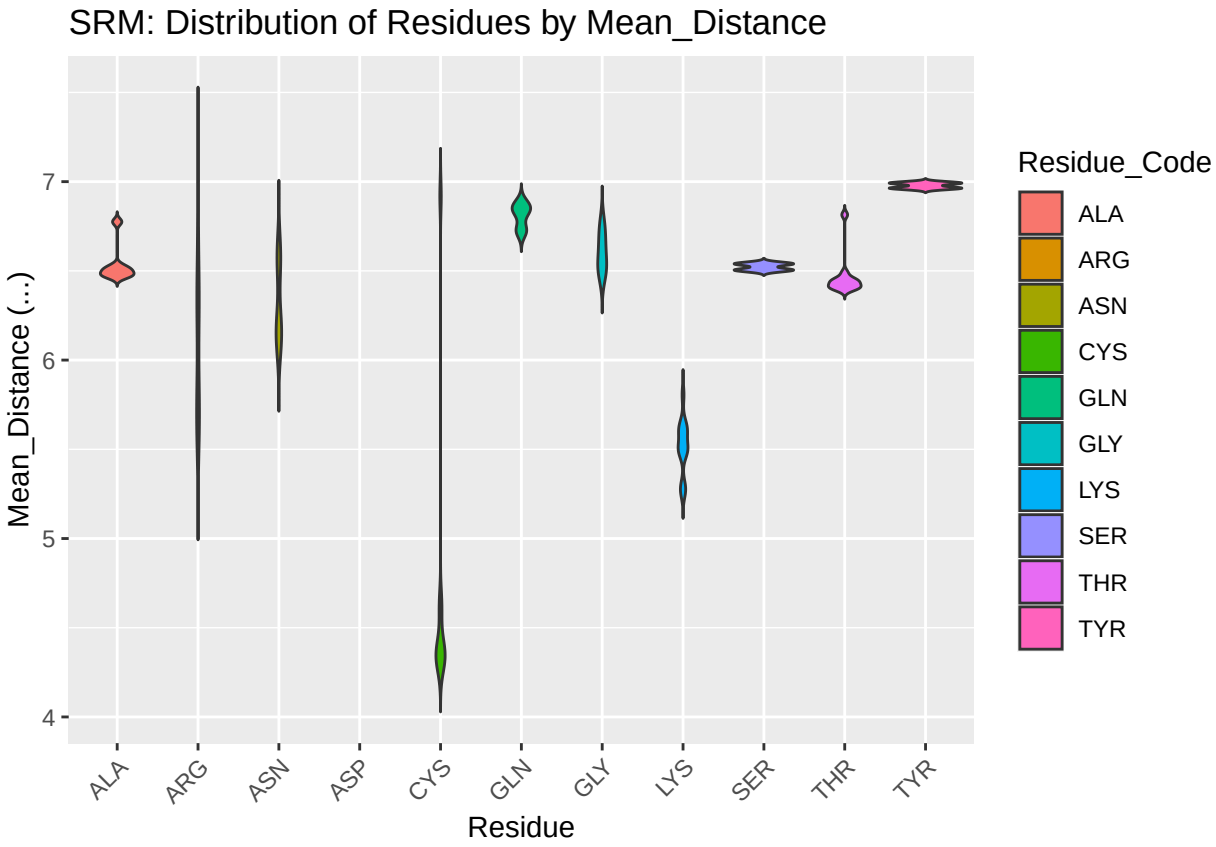

Figure 1.12: SRM: AA Distances

1.8 CA-CB-Fe Angles of Closest Residues

Figures can be found in **??**

Most notably, HIS is the only residue that appears near verdoheme.

1.9 Limitations of the Study

Limited sample size

Limited experimental data to reference to verify

NO experimental data in this study to verify, all theoretical

Only one software package/few algorithms used to calculate all these properties. Others were evaluated but none are compared w.

Algorithms may introduce bias based on how they work e.g. all the bubbles

Arbitrary selection of parameters; some based on rule of thumb or visual evaluation but all or almost all arbitrary

Unknown if the qualities measured are truly the most critical for the heme binding. Some papers suggest other properties may also be important but cannot be calculated, at least right now, e.g. ionic bonding strength etc.

Visual examination itself to OK the parameters/algorithms can introduce bias