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 started with 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 in contact with the heme. This was done 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.

1.1.1 Heme-b

1.2 AA Frequency

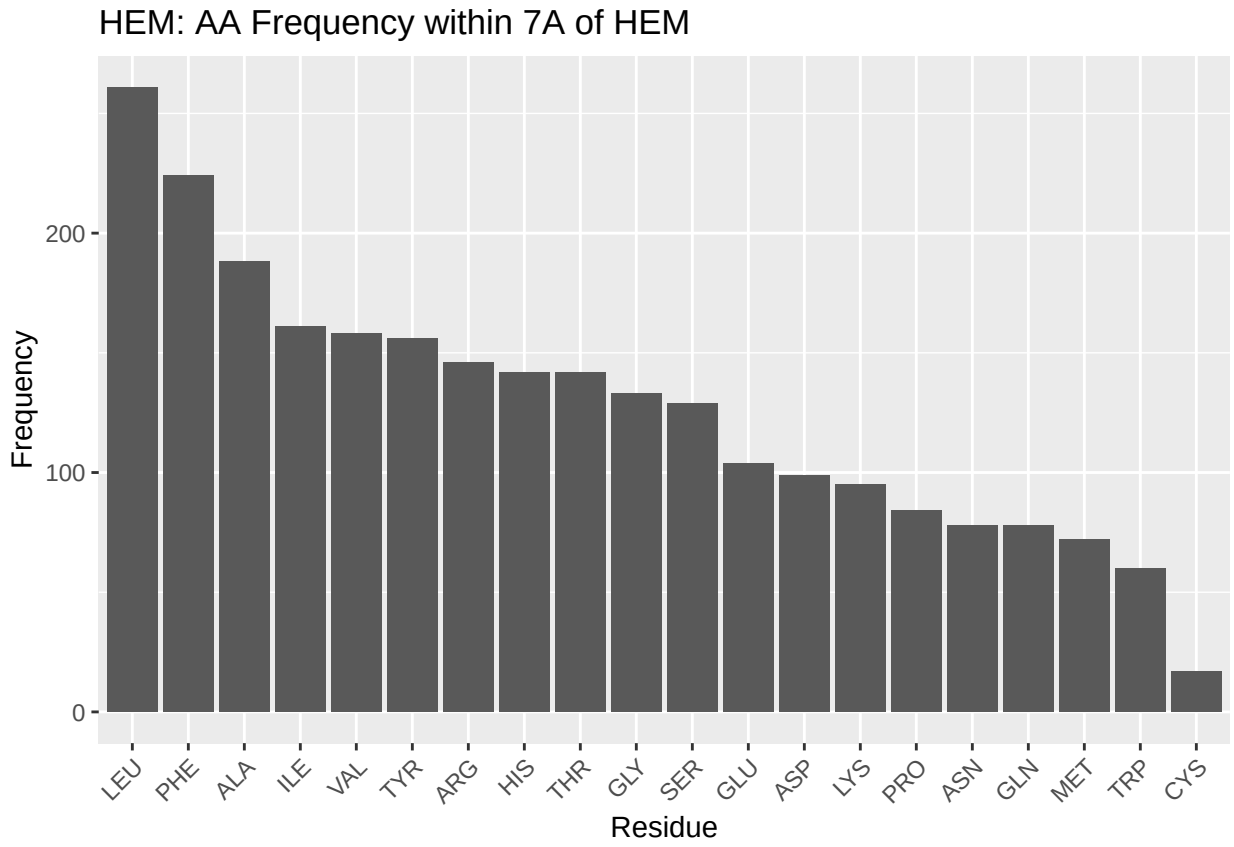

Figure 1.1: HEM: AA Frequency

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
ASP	99
LYS	95
PRO	84
ASN	78

Table 1.1: HEM AA Freq *(continued)*

Residue	Freq
GLN	78
MET	72
TRP	60
CYS	17

1.2.1 Heme-c

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

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

1.2.2 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

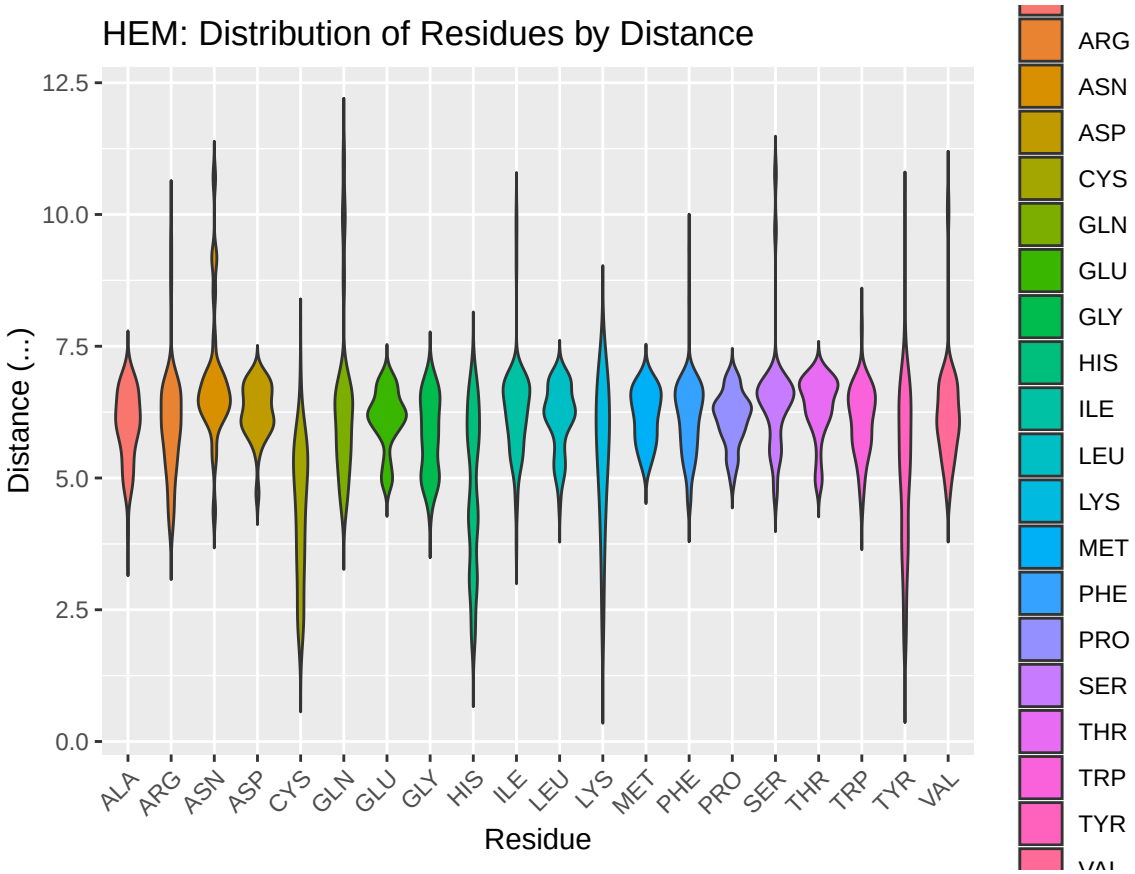

Figure 1.2: HEM: AA Frequency

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 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 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 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.

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** 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 **??**

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