I'm gonna need like 10 Irish Car Bombs, or margaritas when this is all over
UAB

Pat the Great and Powerful

29 June 2021

Abstract

Metalloproteins compose approximately 40 percent (look up how to do percents in latex) of all known proteins, and use some metallic group to accomplish their chemistry. One such metallic group is heme. Heme is a member of the porphyrin family, which are able to catalyze a broad range of reactions. Heme in particular catalyzes many different reactions and is present in many proteins. However, the underlying structural requirements to host heme in a protein are not well studied.

In this study, all heme or heme-c containing proteins as of xx were downloaded and processed in order to determine underlying structural characteristics these proteins may have in common. Parameters that were examined include: xx. Overall, we found: xx. These results may have implications for protein engineering; or if I fucked up this illustrates the difficulty of the field and demonstrate the wide range of acceptable environments of heme; it may therefore be more appropriate to take a more hands-on approach until perhaps other computational methods evolve to better examine structure-function relationships.

See? Not so bad of a worst-case scenario. Just, an unusual sentiment to see in modern science.

Lay Summary

Proteins are biological molecules that perform essential functions in our bodies and in all living things. They are responsible for everything from transporting oxygen in our blood to photosynthesis in plants. They can be extracted from living things and cultured in laboratories and factories. They can then be used as drugs, or be used to perform functions in industrial processes.

Many proteins require additional molecules to perform their function, and these molecules must be bound, docked to the molecule like a boat in a harbor. The molecules can be bound to the protein inside a specialized space, or pocket, within or on the surface of the protein.

One class of proteins is hemoproteins - proteins that use the molecule heme to perform their function. The heme is critical to the proper function of these proteins. Heme enables many specific chemical reactions to be performed, or assisted by the protein. In hemoglobin, in blood, the heme molecule allows the overall protein of hemoglobin to carry and transport oxygen around the body.

But the specifics of the pocket that binds heme in these hemoproteins is not well understood. What are the conditions inside the pocket? Is it a very specific size or can it vary? Is there anything in common among the pockets of many different hemoproteins? These are the questions this research hoped to answer; or at the very least, provide some data and lay the groundwork for others to continue the research later.

This investigation was not conducted in a lab. Rather, using various software packages and the 3D structures of hemoproteins published in a database, various properties of these hemoproteins were calculated. The data produced were analyzed with various statistical methods to extract potentially useful information.

Overall, the following was found: Grad school sucks but bravas are delicious.

Acknowledgments

In case anyone reads this in the future, some context may be appreciated: I attended and completed this Master's during the COVID-19 global pandemic from September 2020 to September 2021.

Thanks professors

Thanks lab

Thanks UAB

Thanks Spain, and Catalonia, allowing me in and then also having public health measures unlike Donny's America

Thanks classmates

Thanks fam, friends

Thanks to the media and the creators of media that facilitated the survival of my sanity through the pandemic.

Finally, I'd like to quote a well-known artist from California. He was referencing his own work, but I wholly identify with his appreciation for the subject of his esteem:

"Last but not least, I wanna thank me. I wanna thank me for believing in me. I wanna thank me for doing all this hard work. I wanna thank me for having no days off. I wanna thank me for, for never quitting. I wanna thank me for always being a giver, and trying to give more than I receive. I wanna thank me for trying to do more right than wrong. I wanna thank me for just being me at all times."

– Calvin Cordozar Broadus Jr.

Contents

List of Figures

List of Tables

Introduction

Proteins may catalyze reactions, and many require ligands to enable their chemistry. A significant portion of all proteins, approximately 40%, require a metallic group as a ligand in order to function correctly - these proteins are known as metalloproteins.

One of these metallic groups is heme. Heme is a member of the porphyrin family, a group of molecules capable of catalyzing a broad range of reactions. Heme can catalyze many different reactions and is present in many proteins. However, the underlying structural requirements to host heme in a protein are not well understood. [MAY ADD CITATIONS]

There have only been a handful of studies dedicated to understanding the structure-chemical relationship between heme and the proteins that use heme for their chemistry (these proteins are known as hemoproteins).

In the most significant previous work, approximately 125 hemoproteins were studied[2]. Although pdbs were thoroughly examined and the datasets were culled, the sample size of this study is very small compared to the amount of hemoproteins available in the pdb a decade later (10,000 HEM-containing proteins and xx). The dataset is also limited in that there is a somewhat homogenous group of proteins examined (?). The characteristics examined were limited to: xx.

It is hypothesized that the following characteristics all have an impact on the binding of heme and function of the hemoprotein: XXXXXXXXX.

In this study, some of these characteristics were examined. They include: XX. The remainder are thus far not feasible to calculate.

All of these characteristics have implications in the field of protein engineering or basic research into hemoproteins. Examples of the uses of these results include [SUPER BLOOD STUDY] and [OTHER PROTEIN ENGINEERING STUFF]. Not sure how much we can reference those other papers besides doing that besides in the conclusion.

Notable results from some of the prior studies include: xx and xx. These characteristics are also examined in this dataset, while some are not due to different study approaches.

Methods

Datasets

**Remember to alter how this header's size appears... later. Several datasets were constructed to examine the full ... (this sentence belongs in the intro, I think) The primary dataset of heme-containing proteins (HEM) was composed by finding approximately 30 (specify exact number) of types of proteins that are present in the PDB. This included heme oxygenase, myoglobin, cytochrome P450, among many others.

Datasets: HEM, HEC(?), SRM, VER/VEA.

Four datasets were constructed for this study.

The primary dataset of heme-containing proteins (HEM) was constructed by searching for 30 different classes of proteins; this enabled a dataset of diverse proteins to account for any structural deltas to achieve different chemistry. Additional samples of each class of protein were then added to the dataset, bringing the total to XX. The PDBs were restricted to the following criteria to ensure quality: XX. A full list of proteins and their source organism used in the study is available in table XX.

The heme-c dataset followed the same criteria (XX). Similar proteins were searched for as HEM, depending on availability in the PDB/possibility with chemistry. This dataset was anticipated to be fairly similar to HEM, and so only contains XX samples. The full table is available in table XX.

The siroheme dataset (SRM) contains fewer samples than HEM or HEC due to the limited structures available. A search for 'SRM' as of 26 July 2021 produced 52 structures. Not all of these structures contain siroheme. A full-text search for "siroheme" produces many more results, but very few are complex with siroheme. SAH appears commonly used to complex the relevant siroheme proteins, however examination of whether this guarantees acceptable results/estimates of SA/V etc. is outside the scope of this study. No quality criteria were employed, but all proteins are within: XX.

The verdoheme dataset (VER/VEA) is very limited. A search for "ver-

doheme" as of 26 July 2021 produces 12 results. From these results only 4 usable proteins are available. All PDBs fall within XX criteria.

Some PDBs in HEM-dataset contain PDBs where there is a double-molecule representation of heme. If this has not been taken care of ** FIX ME!!!* then write something here.

The scripts used in this study were modified depending whether HEM/HEC/SRM/VER/VEA were being processed, and depending on the distance from the ligand of interest being examined (i.e. 5-7A). This is discussed in further detail below MAYBE (FIXME!).

Confirming data quality and details/PDB detail table

All PDBs used in the study were scanned/text-parsed with a python script. This script grabbed a bunch of relevant qualities, like molecule purpose, source organism, resolution (XX FIX), and PDB code to confirm. The data produced are in table X.

Preprocessing/before chimera/monomers

Many of the PDBS downloaded are multimeric structures. These were all processed into monomeric structures, by selecting a single chain (chain A) and eliminating all others chains in each PDB. This makes examination easier and more representative when data are aggregated; multimers with more pockets would otherwise skew the data and be the majority represented in the results.

All scripts below were paused while running for visual examination; in rare cases the process of conversion to monomers resulted in errors processing, especially for volume measurements. This is discussed below if this issue was not corrected (FIXME!! XX).

Examination in Chimera/acquiring results

Multiple scripts were written and applied in Chimera. These scripts are divided up based on what results are produced.

In the scripts the choice of 5A or 7A as a distance from the ligand is arbitrarily chosen. This is a distance that generally accounts for all residues

able to interact with the ligand. The results are presented in both 5A and 7A sets to account for the variability introduced by these cutoffs.

Volume

Volume of the binding pockets for each ligand are calculated using surfnet. Atoms within 5-7A of the ligand are selected and the pocket volume they form with the ligand is calculated. The surfnet algorithm works by... making triangles along the molecular surface, I guess, until the distance cutoff.

Volume of the ligands is also calculated, using a different method; the above method with surfnet is not possible to employ for single molecules outside the pocket. First the ligands were isolated from their pdb. The molecular solvent/surface area was calculated (Accesible and excluded). The surface area and the volume of the resulting... blob, is given by Chimera. This is not the same method as employed by surfnet to acquire volume and represents a limitation in the study (FIXME! IDK IF THIS SHOULD GET DISCUSSED HERE)

Images of this operation are available in XX.

Surface Area

Surface area of the pockets is calculated using a similar method as above. The atoms within 5-7A of the ligand are selected. The 'surf' operation is applied. This creates a surface... IDK how this algorithm works (FIXME!). Both accessible and excluded solvent area are outputted.

Surface area of the ligands is calculated as noted above in Volume section.

Images of this operation are available in XX.

Distances

Distances (NOT THE ANGLEDIST stuff) were calculated by selecting all atoms within 5-7A of Fe in the ligands. Each atom's distance to the Fe was calculated by using the distance operation in Chimera (confirm this is precisely what we do), which simply draws a line between the atom and the Fe atom. (FIXME! IN DISCUSSION, DISCUSS WHY METAL COORDINATION IS GARBAGE)

The residue each atom is in is reported. Therefore, each

Angles Residues - Heme Plane

Angles Fe-CA-CB

Amino Acid Frequency in Pockets

title

Importing to R and stastical analysis

FIgures wer also made in R.

- Download from PDB using the script they've provided at RCSB for many, many files
- Use UCSF Chimera to determine:
  - Volume
  - SA
  - Nearby AA
- R to process raw data and produce tables
- Whatever other software we use to achieve the other results. E.g. E, or availability to solvent etc. likely will stick w Chimera I suspect. Or somehow implement the Python script to open both chimera for the first part of what we've done or for something else later. The script we've written is a python script, not a chimera script. We're initializing it with chimera and excluding the necessary code... to initialize chimera and specify chimera to receive the commands

Equipment

Results

Attempt to reference **??** figure: **??** References

Discussion

Limitations of this brilliant work: 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 Visual examination itself to OK the parameters/algorithms can introduce bias Precision of algorithms needs to be evaluated** or at least IDK how PRECISE they are

Conclusion

this is just as master's and in basic research don't feel the need to replicate what took some god forsaken, sad, overworked, impoverished PhD students years + with help of their PIs and with generous word fluff to hide fuck ups. 0 -¿ thesis in approx. 3-4 months during global catastrophe is nifty

Bibliography

[1] Chris E Cooper et al. "Engineering tyrosine residues into hemoglobin enhances heme reduction, decreases oxidative stress and increases vascular retention of a hemoglobin based blood substitute". In: *Free Radical Biology and Medicine* 134 (June 2019), pp. 106–118. ISSN: 18734596. DOI: 10.1016/j.freeradbiomed.2018.12.030.

[2] Ting Li, Herbert L Bonkovsky, and Jun Tao Guo. "Structural analysis of heme proteins: Implications for design and prediction". In: *BMC Structural Biology* 11 (2011). ISSN: 14726807. DOI: 10.1186/1472-6807-11-13.

Appendices

Figures

Figure A.1: HEM AA Frequency 7A

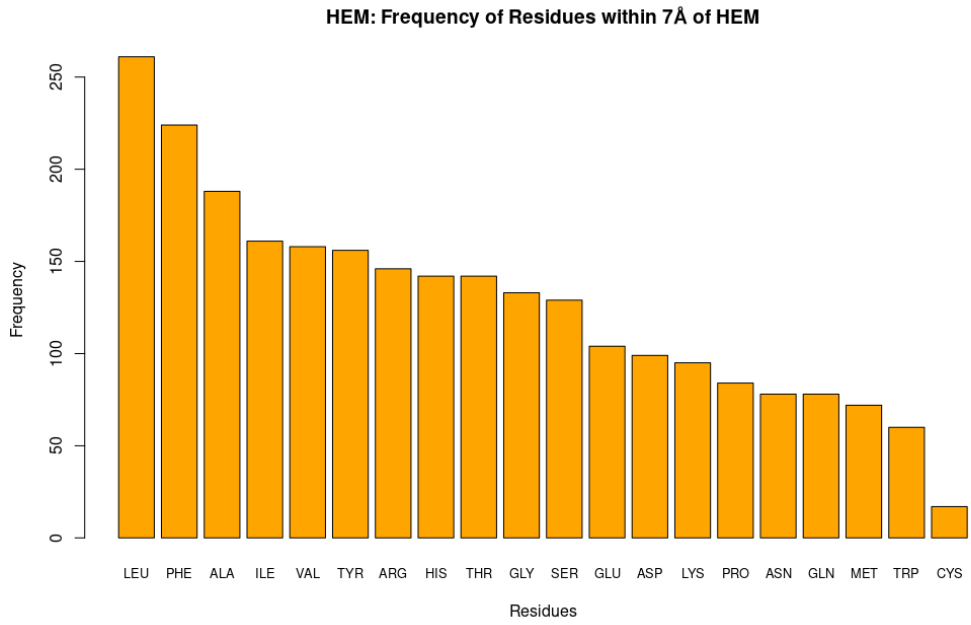


Figure A.2: HEC AA Frequency 7A

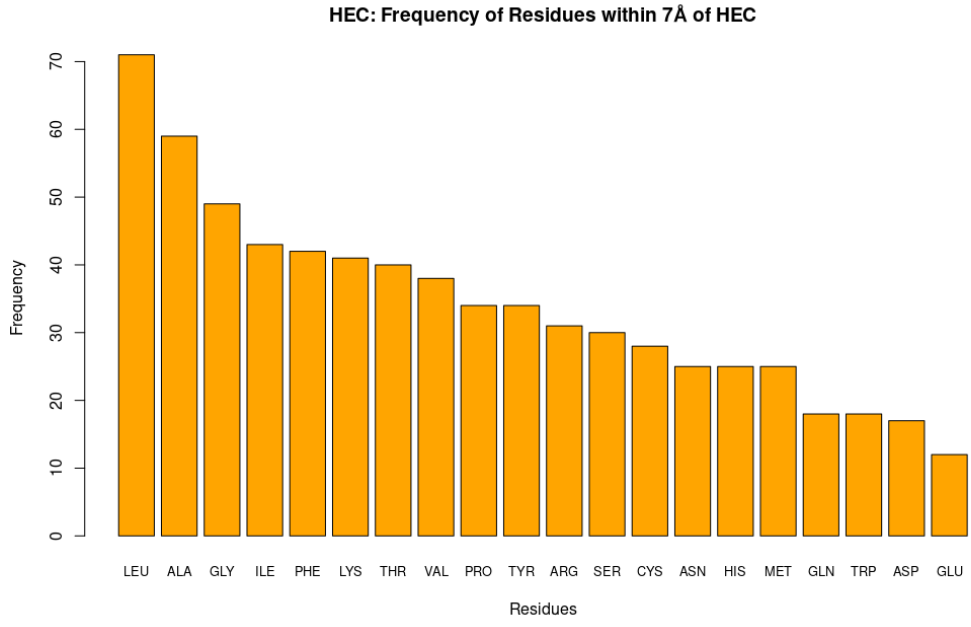


Figure A.3: SRM AA Frequency 7A

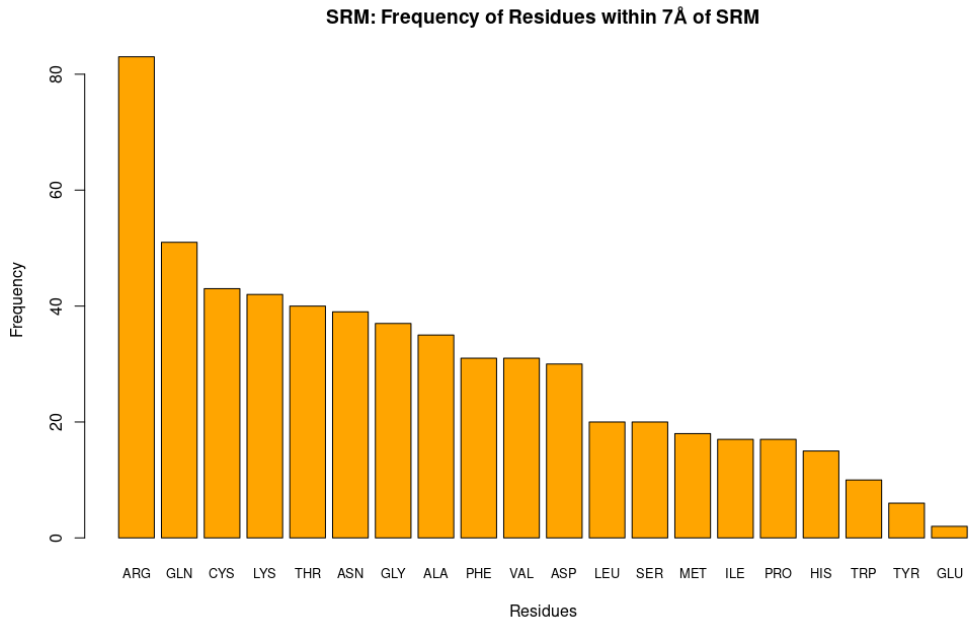


Figure A.4: VERDOHEME AA Frequency 7A

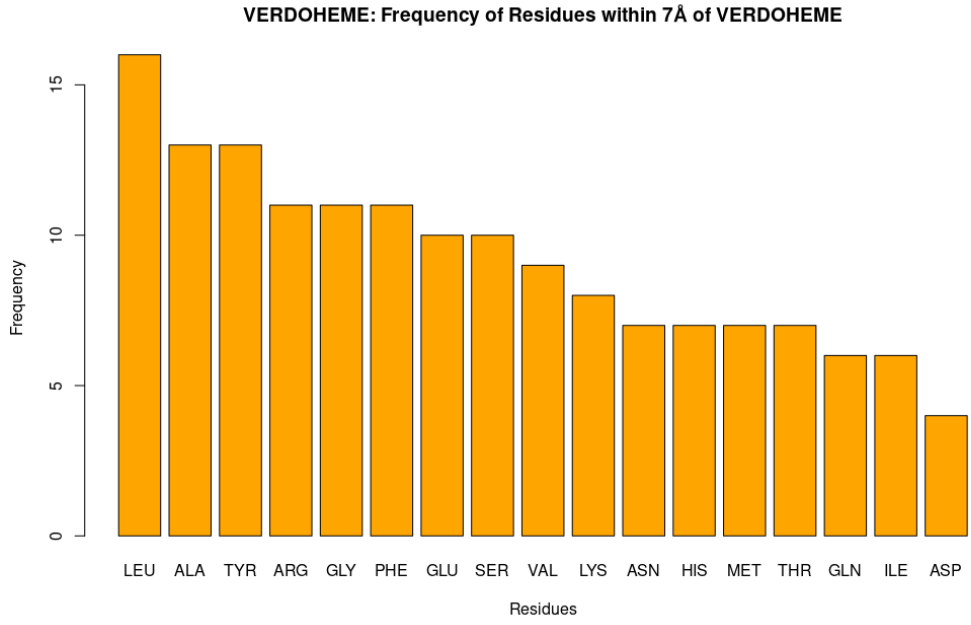


Figure A.5: HEM CACBFe Data

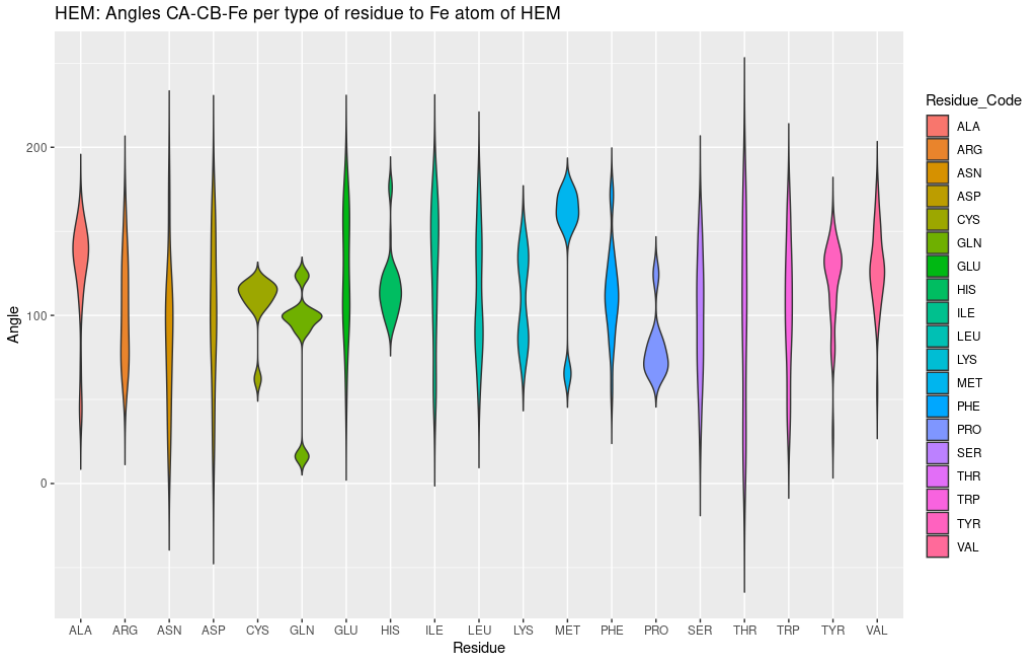


Figure A.6: HEC CACBFe Data

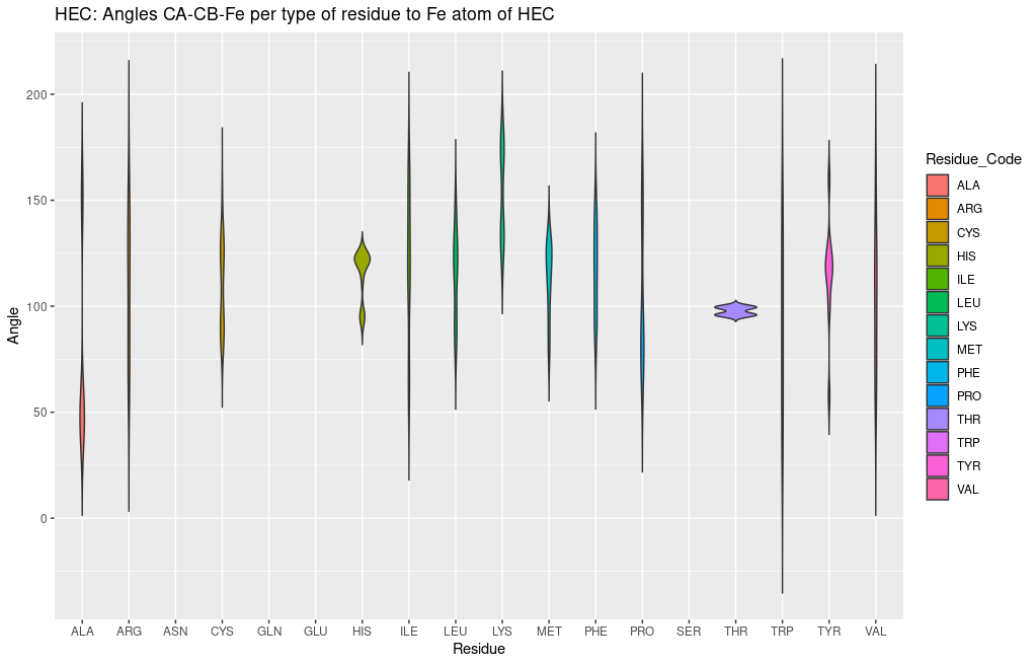


Figure A.7: SRM CACBFe Data

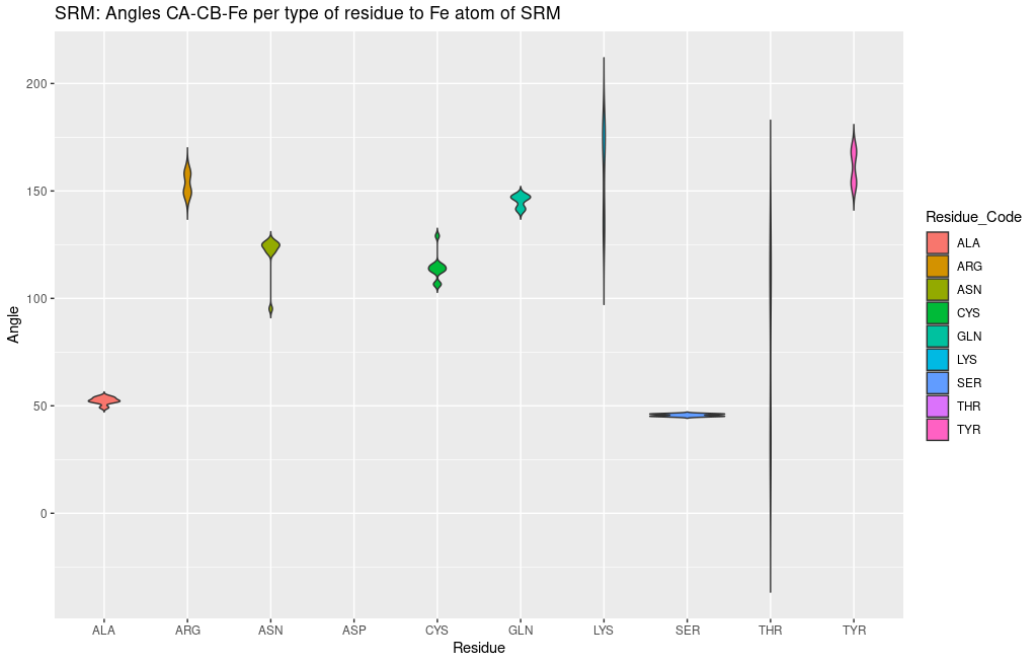


Figure A.8: VERDOHEME CACBFe Data

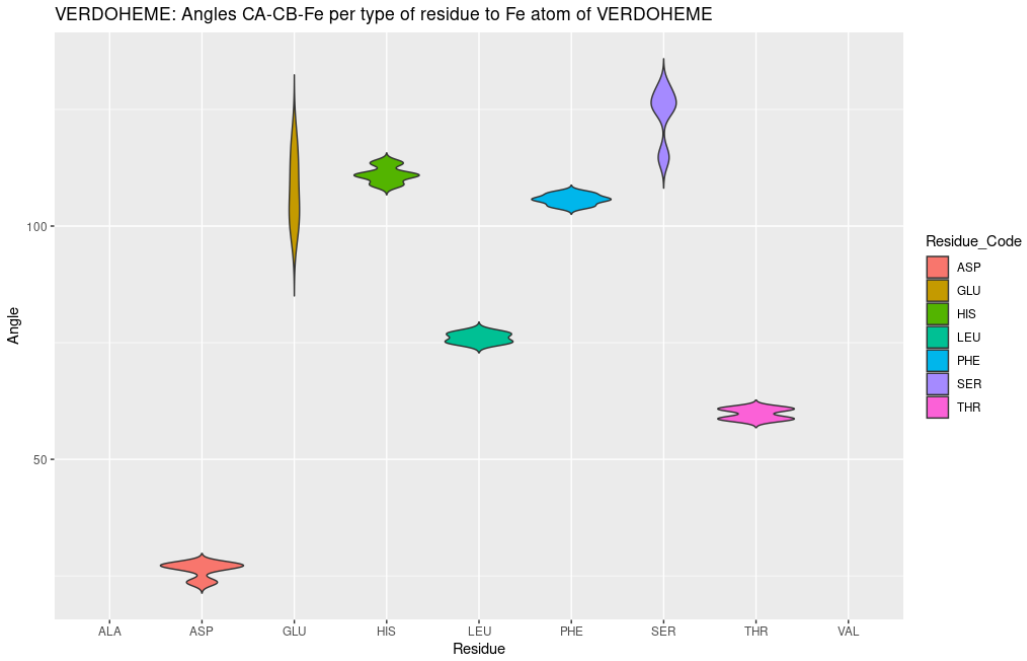


Figure A.9: HEM Closest Residue Data

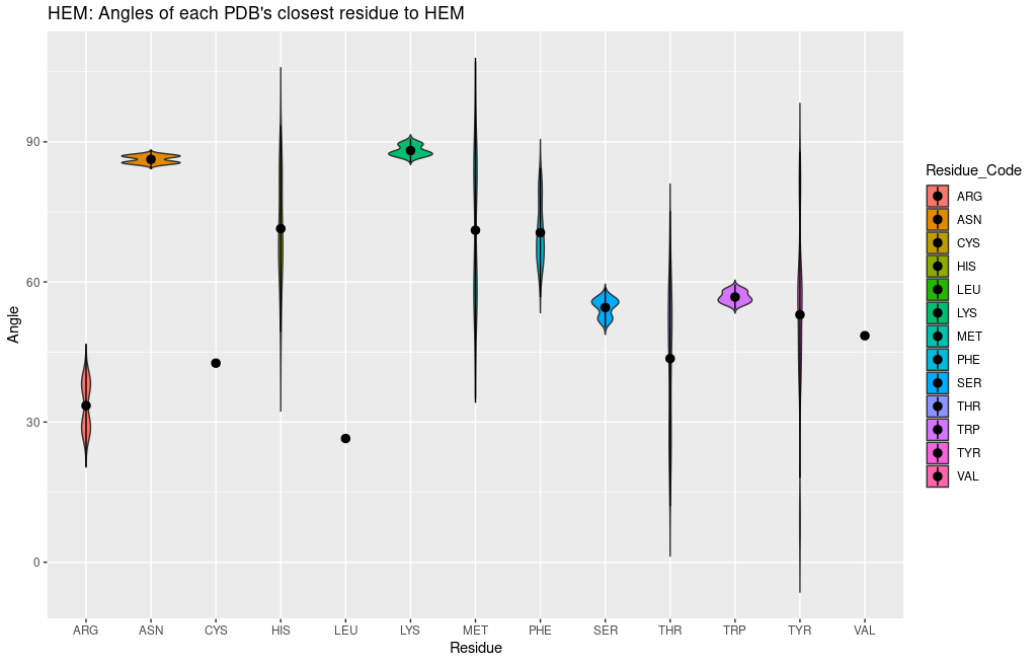


Figure A.10: HEC Closest Residue Data

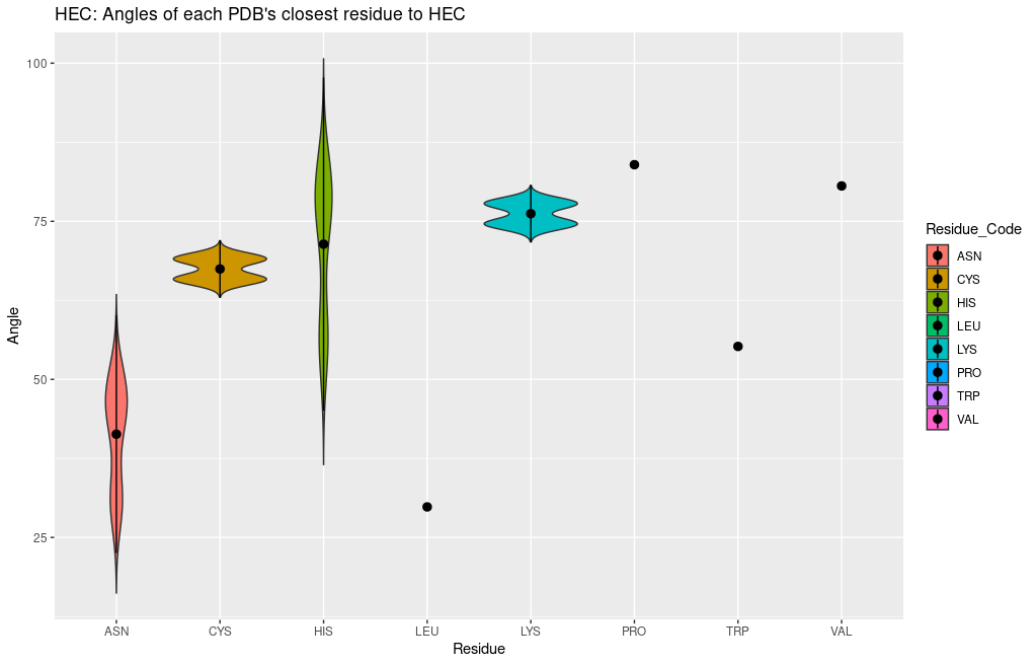


Figure A.11: SRM Closest Residue Data

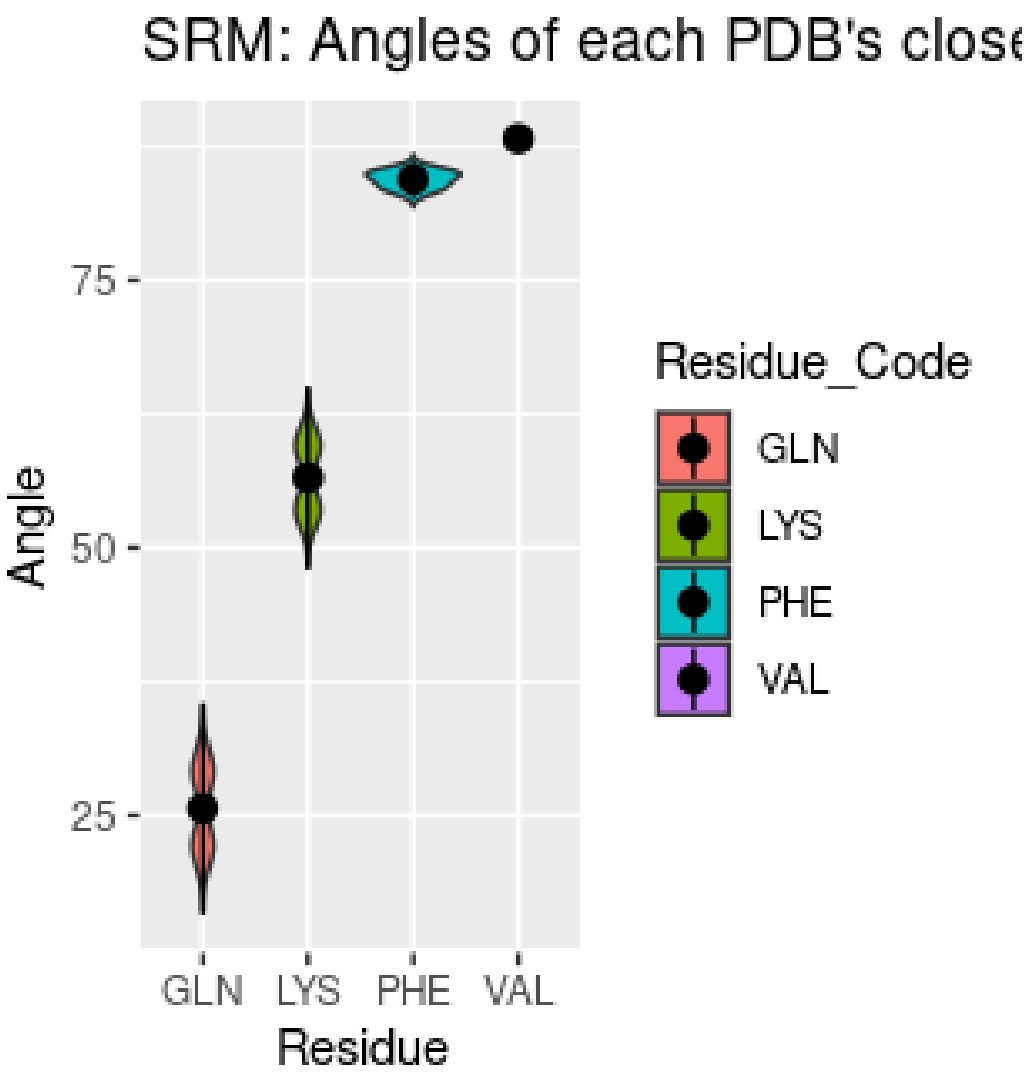


Figure A.12: VERDOHEME Closest Residue Data

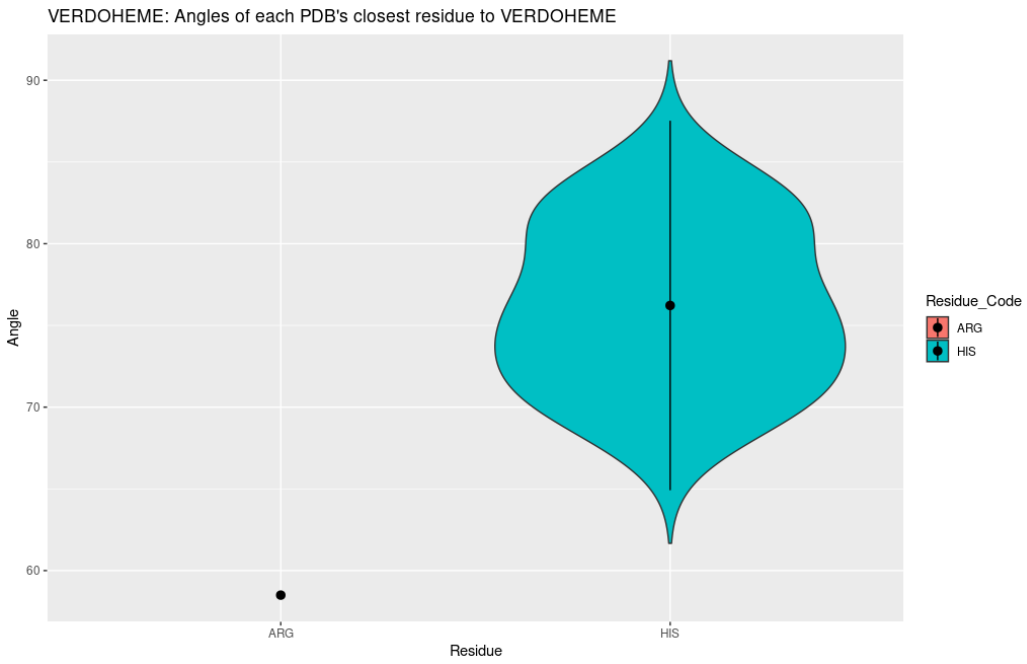


Figure A.13: HEM Coordinating Residue Data

Figure A.14: HEC Coordinating Residue Data

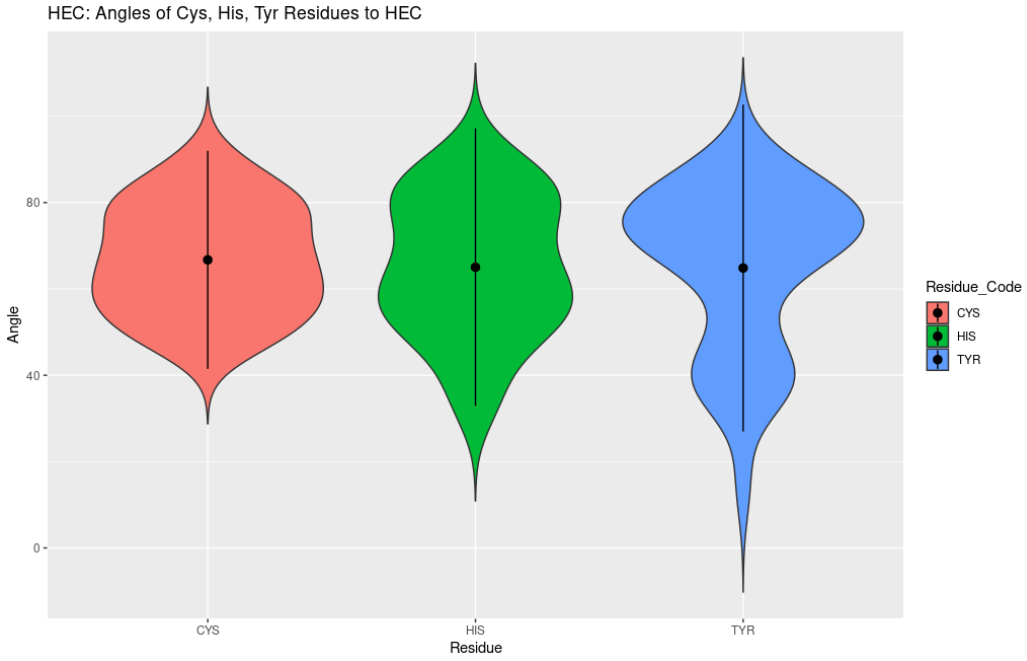


Figure A.15: SRM Coordinating Residue Data

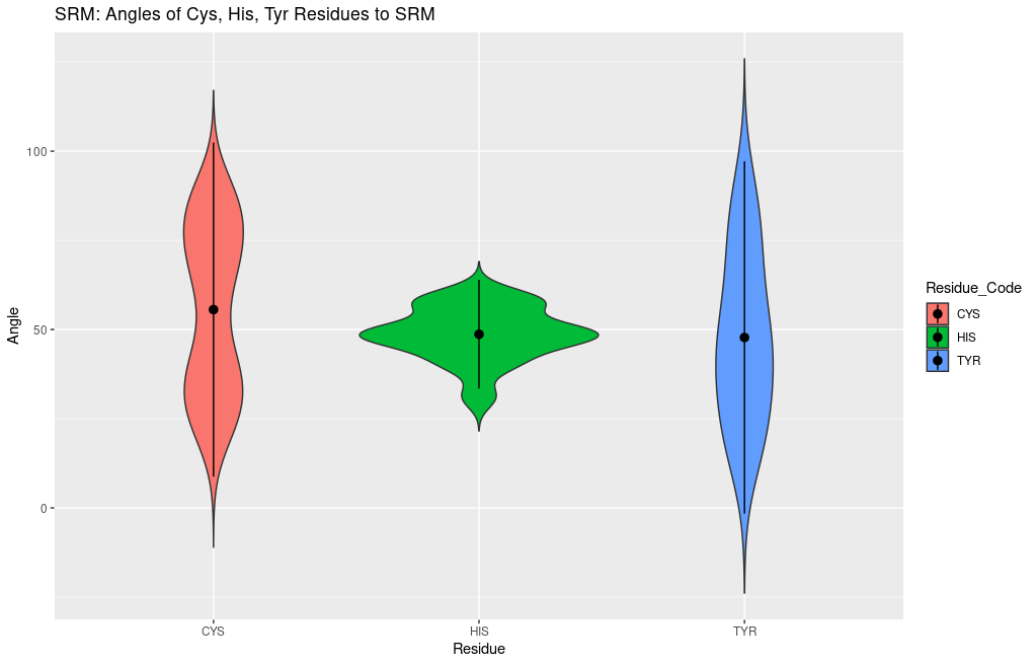


Figure A.16: VERDOHEME Coordinating Residue Data

Figure A.17: HEM Ligand Accessible Surface Area

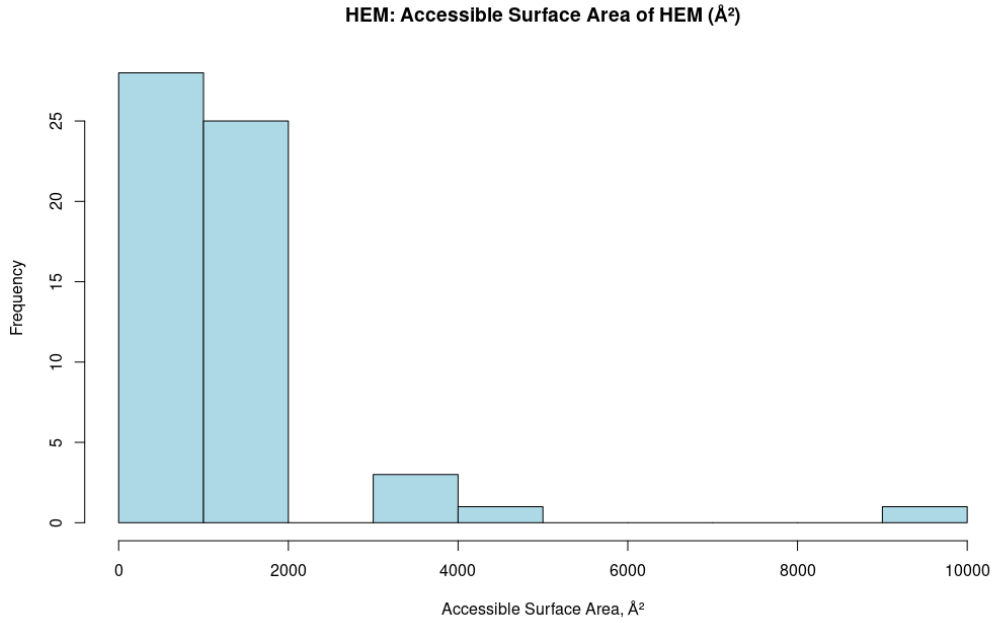


Figure A.18: HEC Ligand Accessible Surface Area

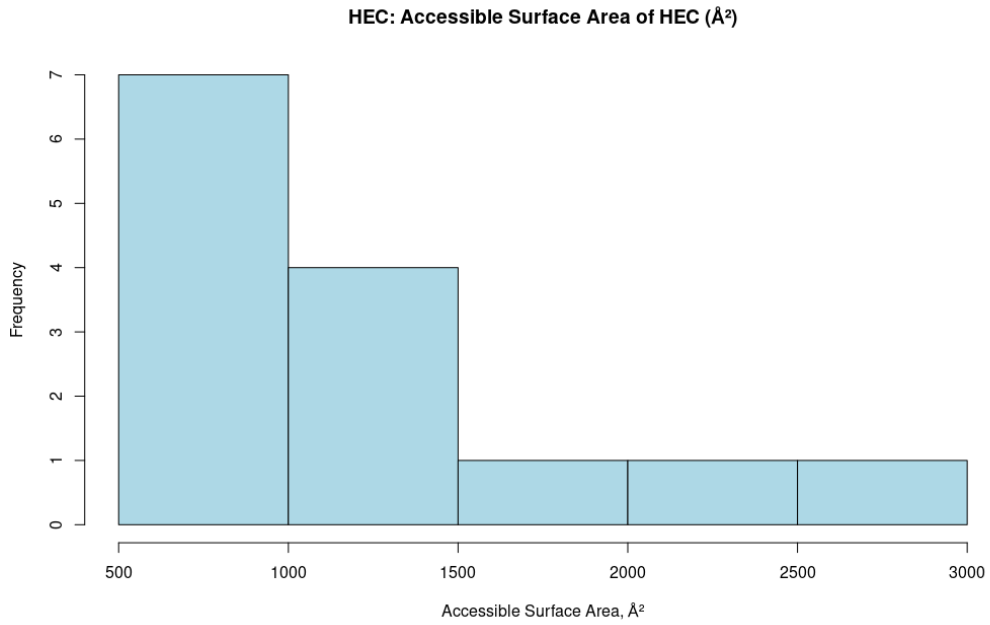


Figure A.19: SRM Ligand Accessible Surface Area

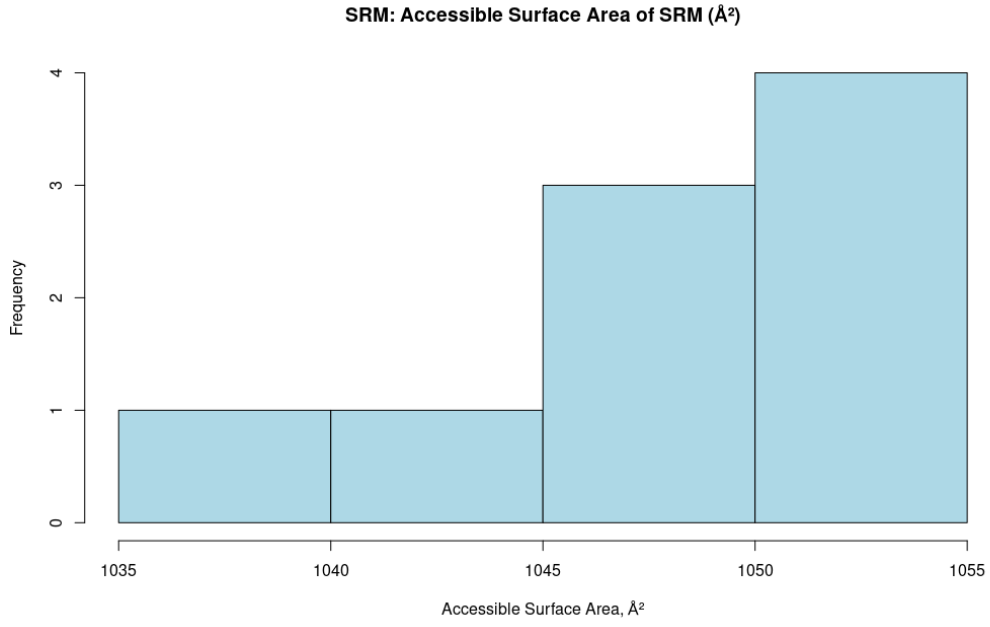


Figure A.20: VERDOHEME Ligand Accessible Surface Area

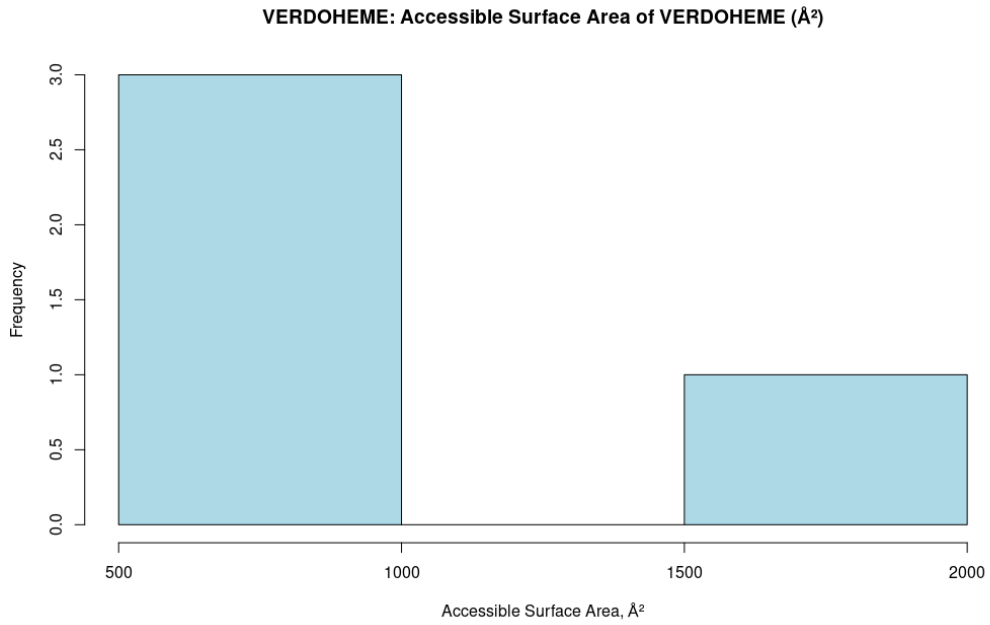


Figure A.21: HEM Ligand Excluded Surface Area

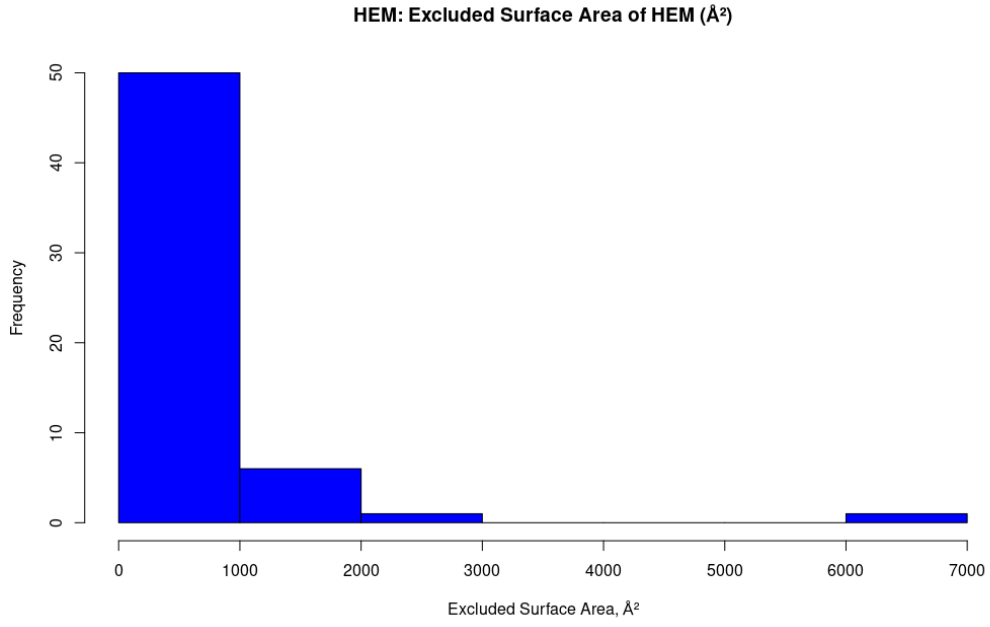


Figure A.22: HEC Ligand Excluded Surface Area

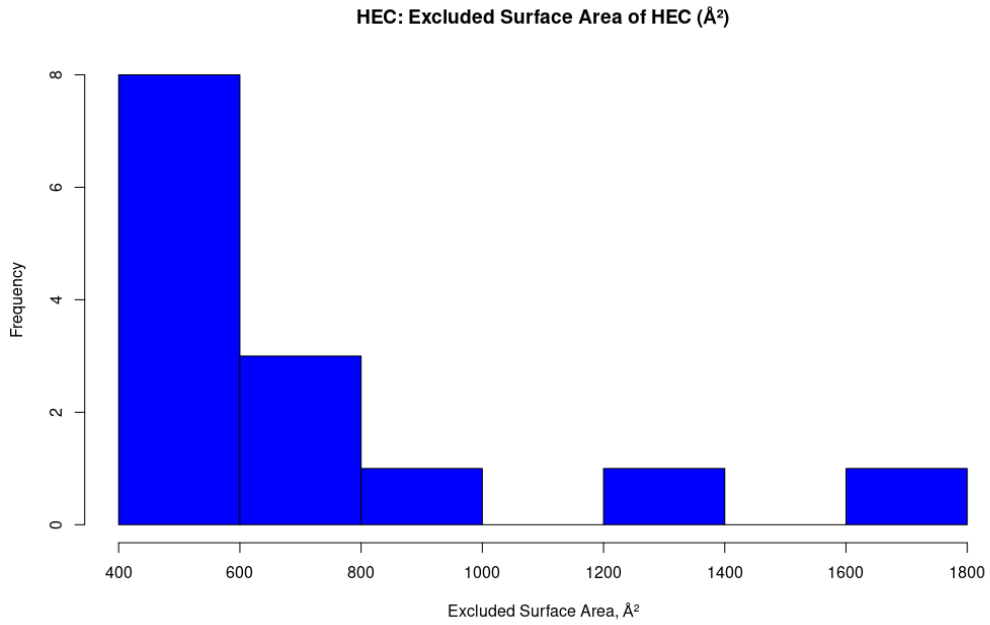


Figure A.23: SRM Ligand Excluded Surface Area

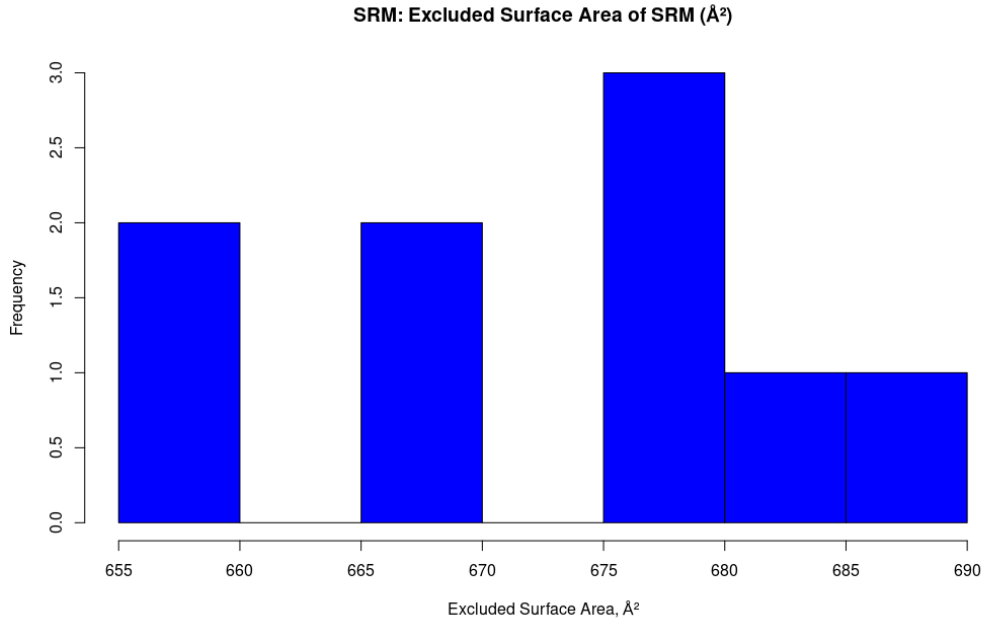


Figure A.24: VERDOHEME Ligand Excluded Surface Area

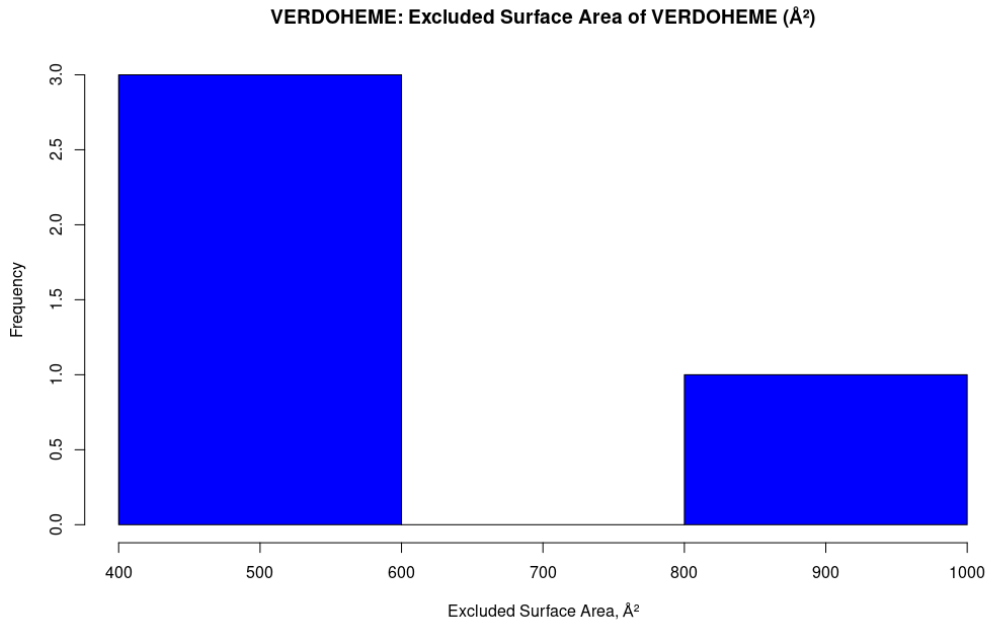


Figure A.25: HEM Planar Angles

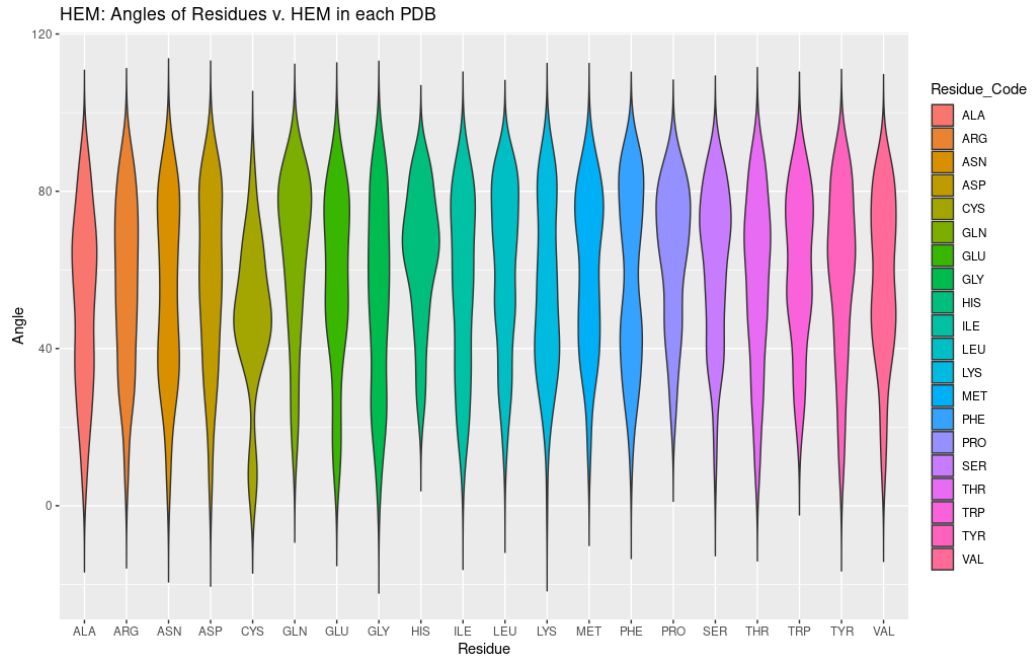


Figure A.26: HEC Planar Angles

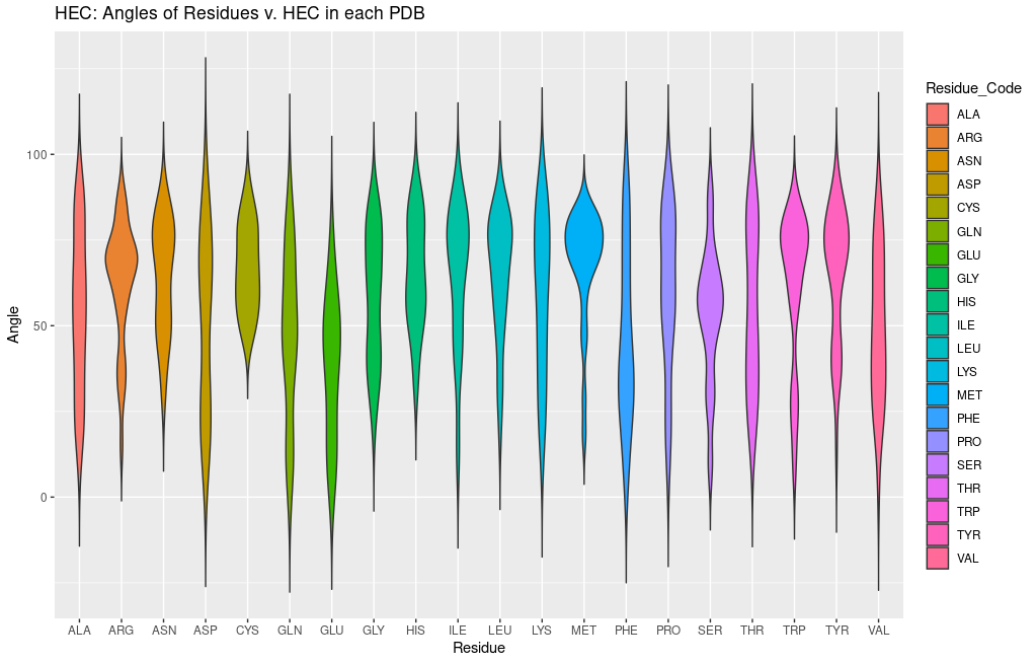


Figure A.27: SRM Planar Angles

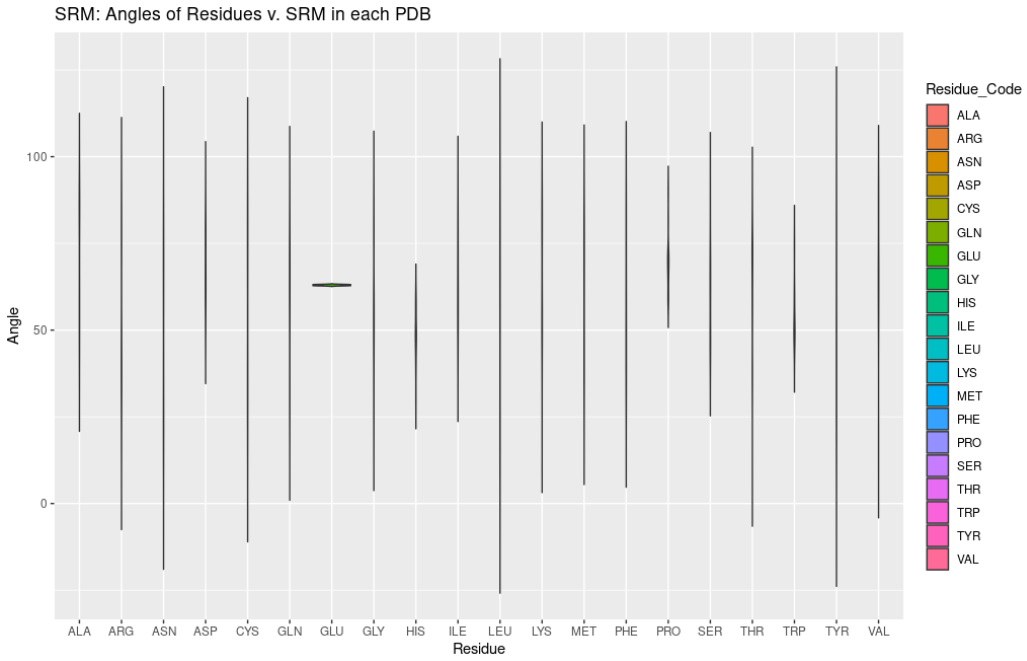


Figure A.28: VERDOHEME Planar Angles

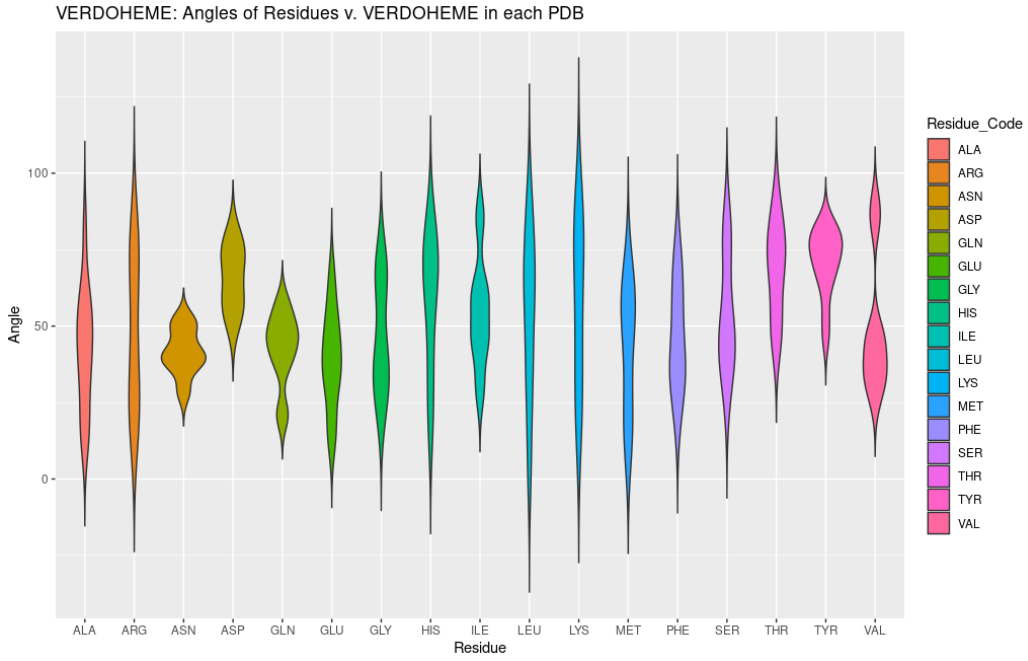


Figure A.29: HEM Pocket Accessible Surface Area

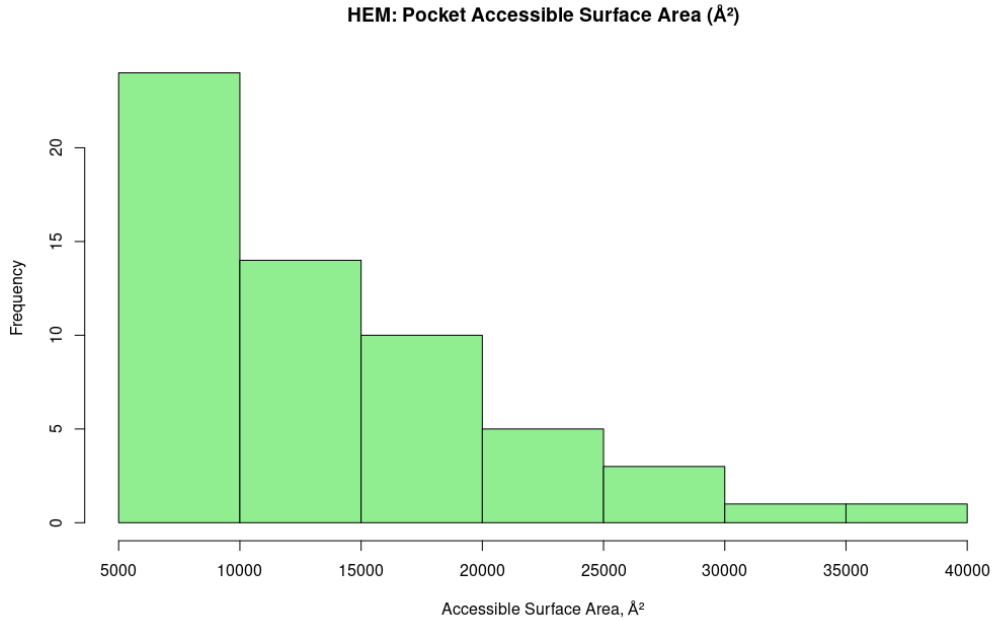


Figure A.30: HEC Pocket Accessible Surface Area

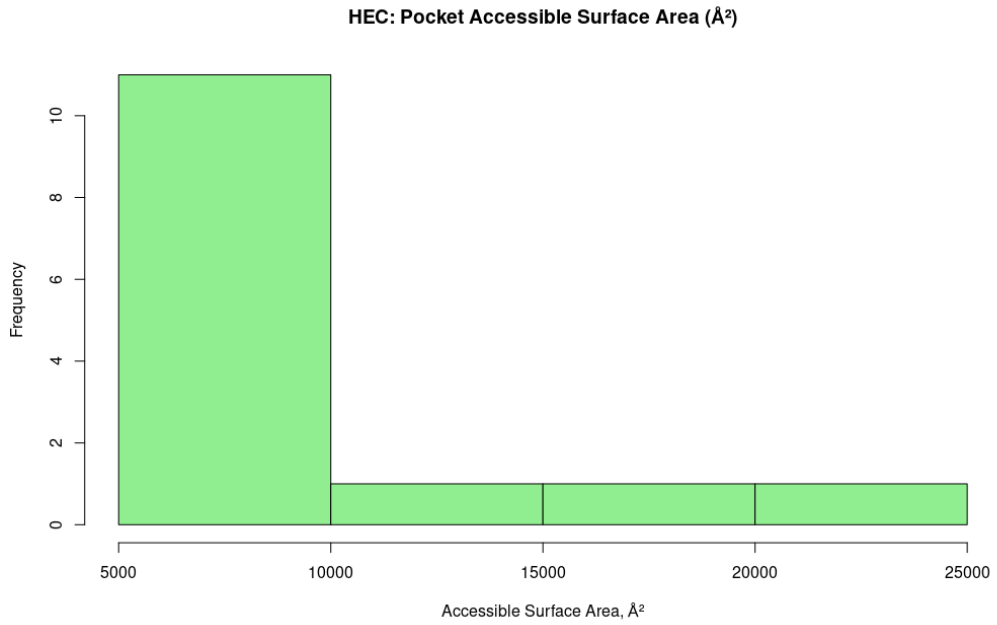


Figure A.31: SRM Pocket Accessible Surface Area

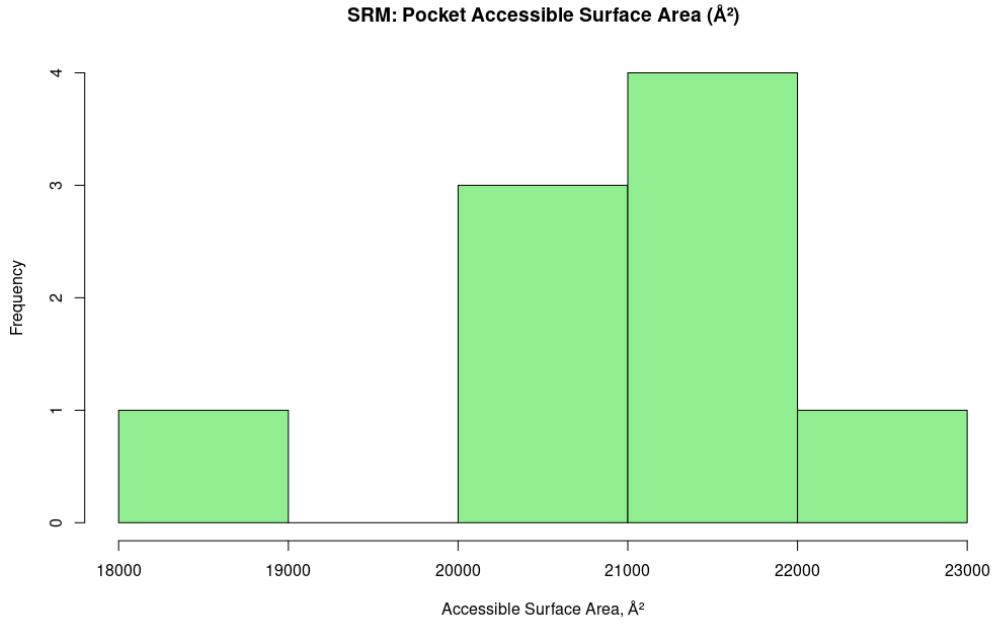


Figure A.32: VERDOHEME Pocket Accessible Surface Area

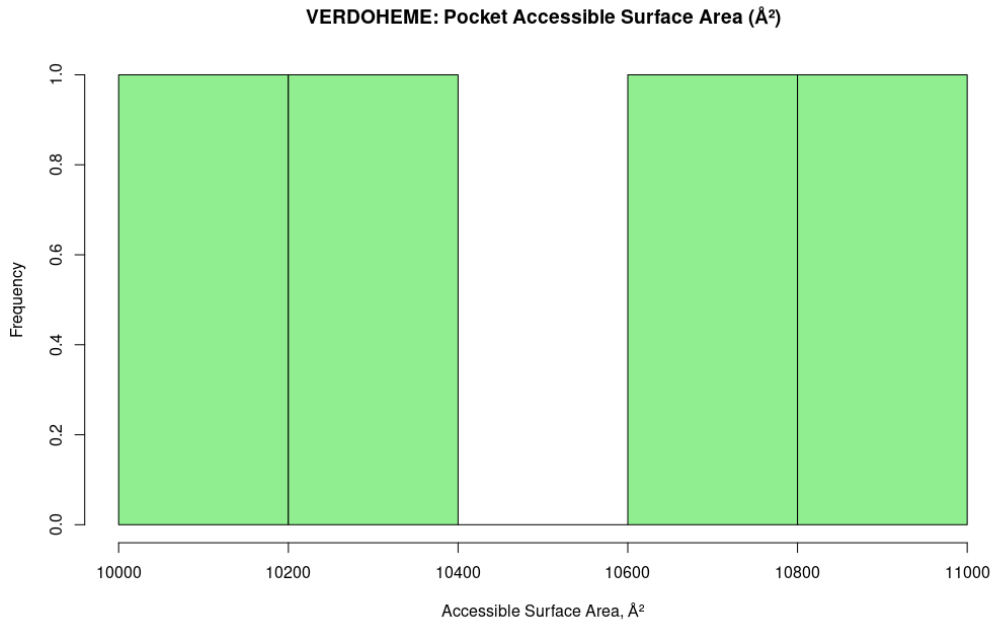


Figure A.33: HEM Pocket Excluded Surface Area

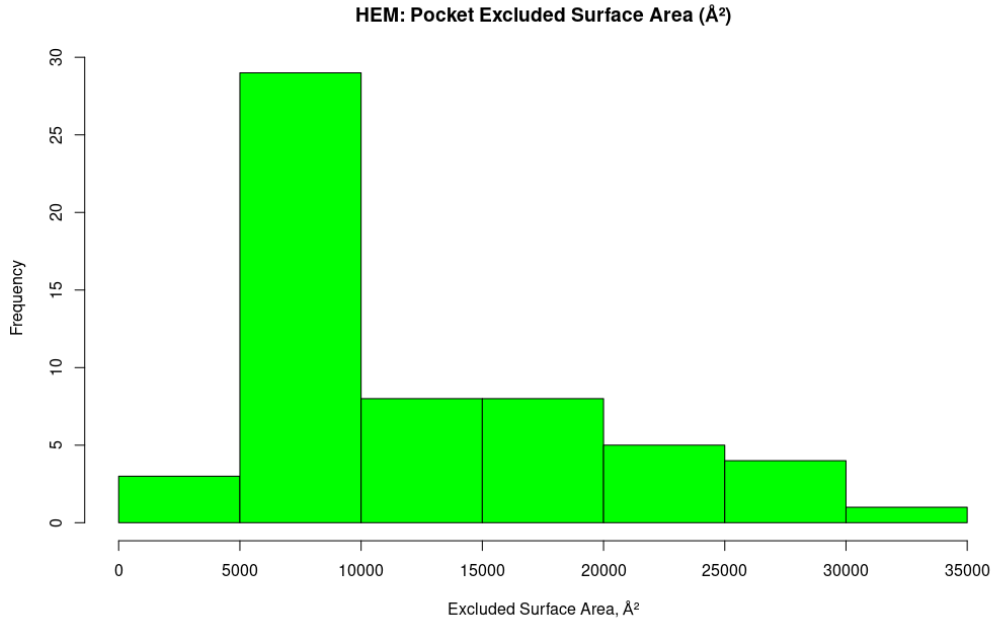


Figure A.34: HEC Pocket Excluded Surface Area

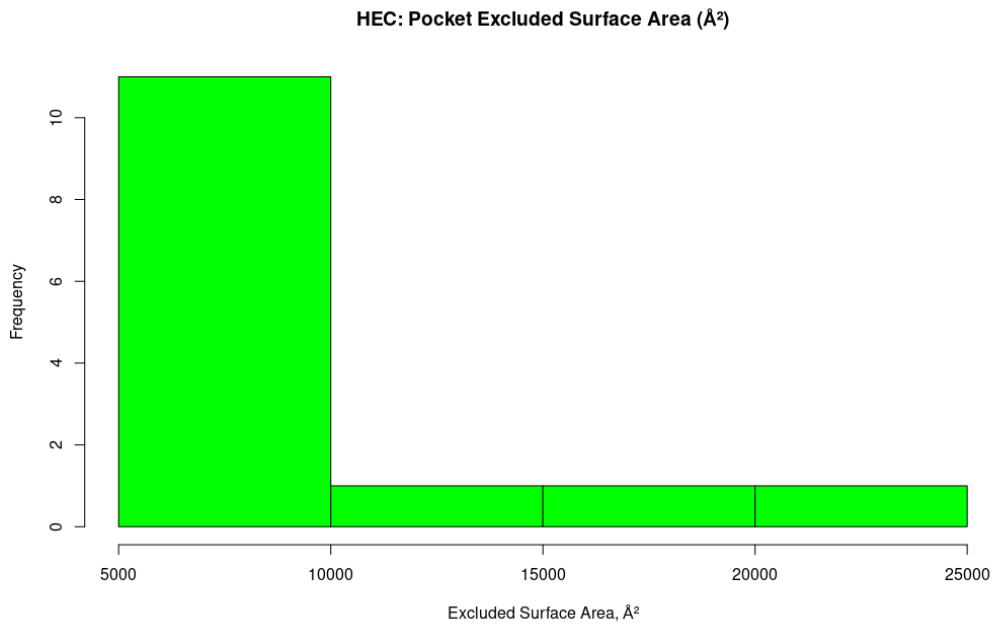


Figure A.35: SRM Pocket Excluded Surface Area

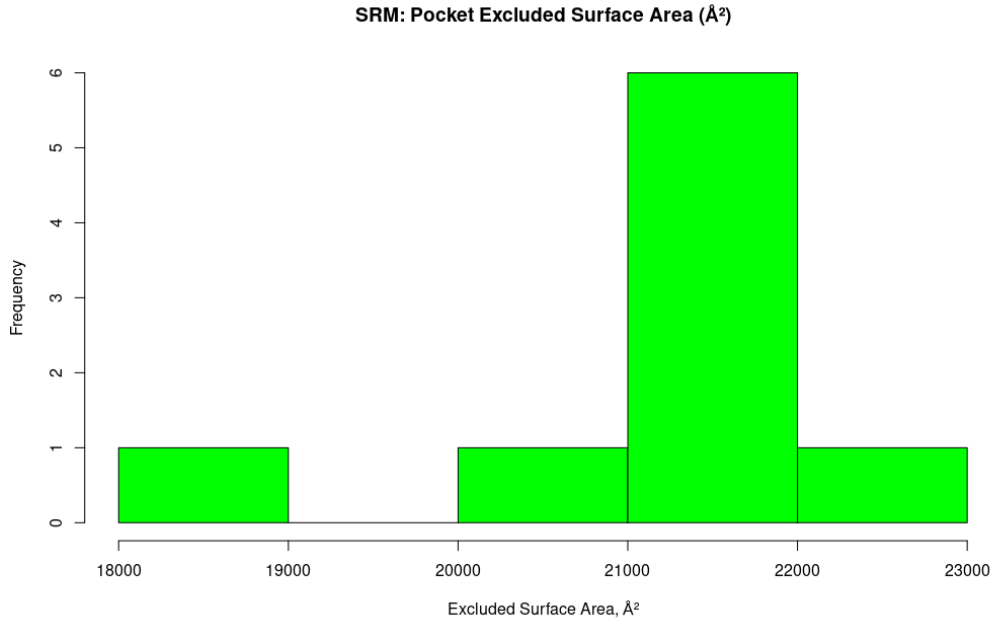


Figure A.36: VERODHEME Pocket Excluded Surface Area

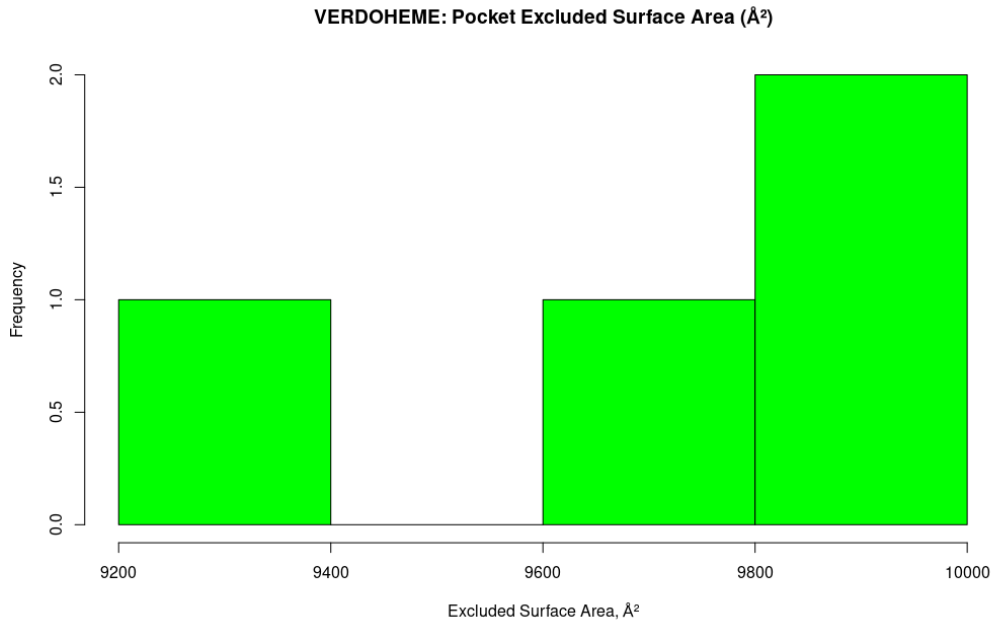


Figure A.37: HEM Pocket Volume

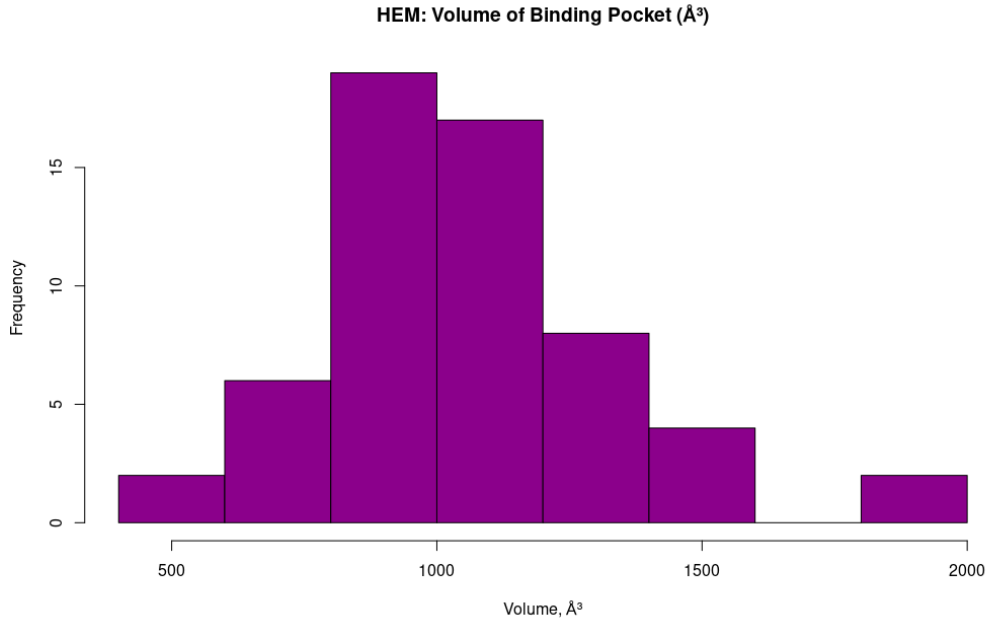


Figure A.38: HEC Pocket Volume

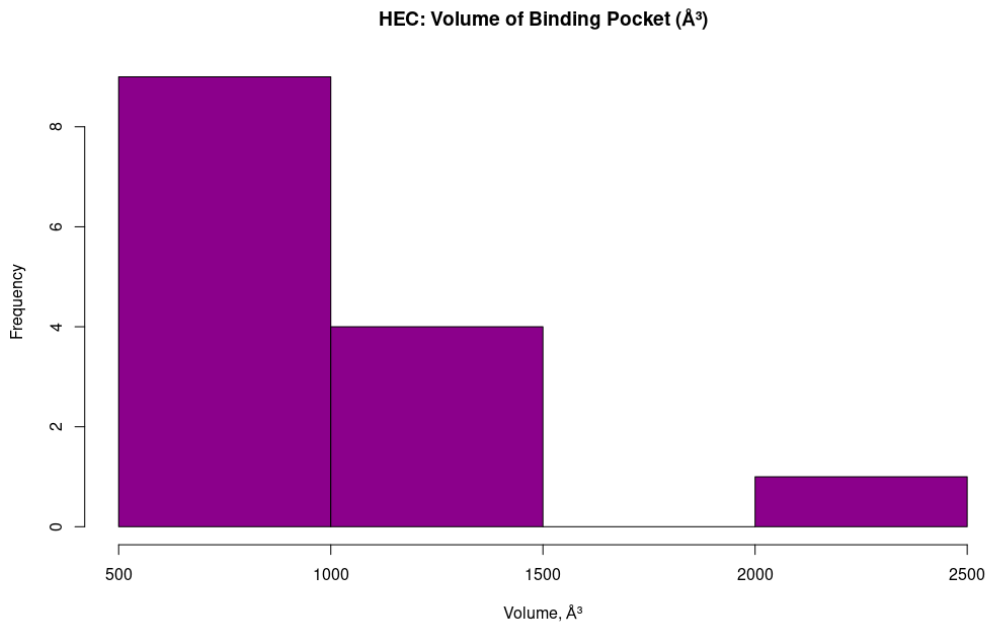


Figure A.39: SRM Pocket Volume

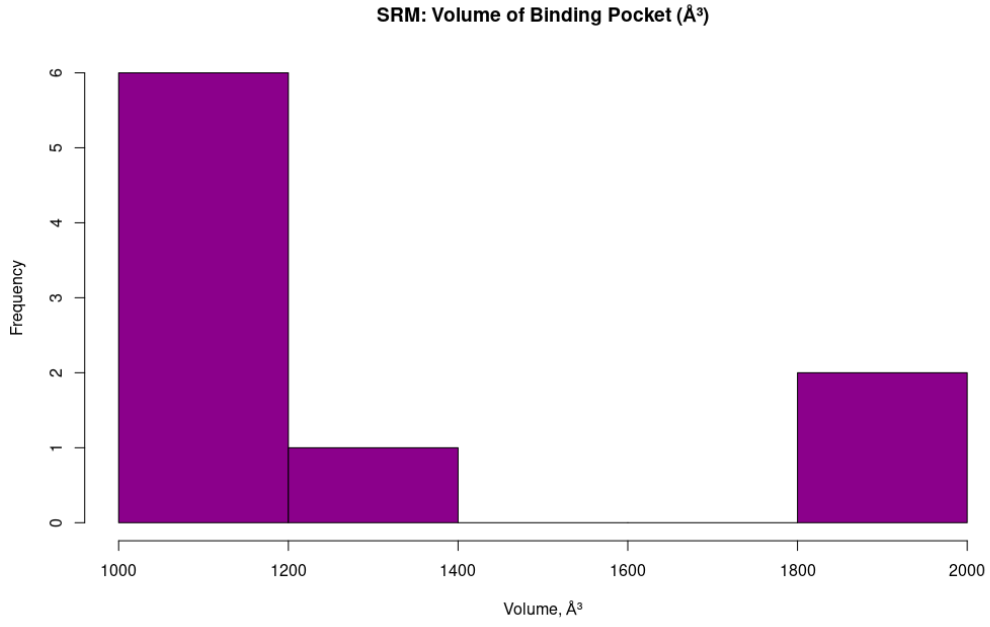
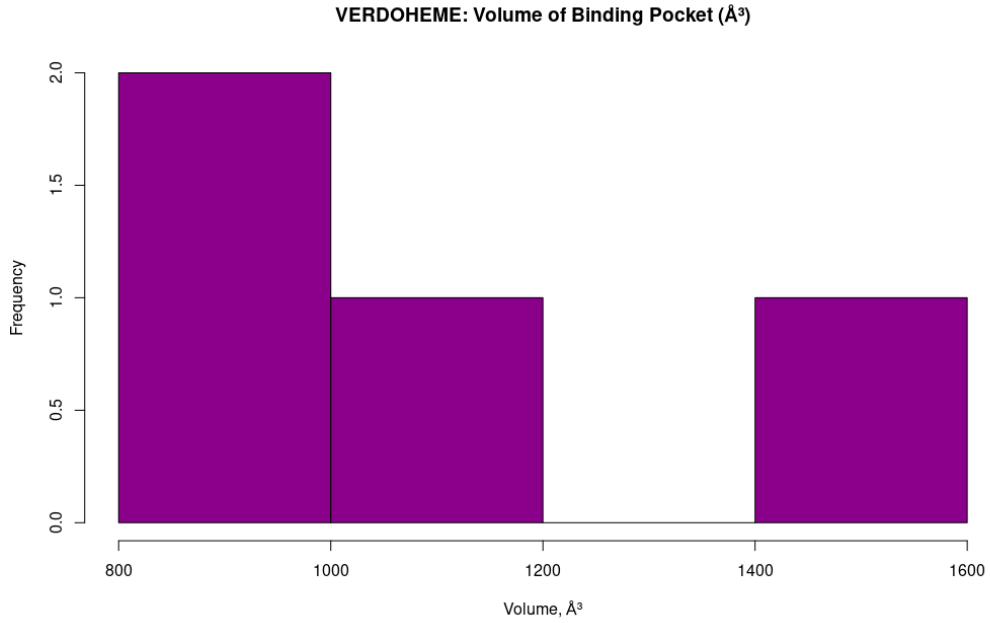


Figure A.40: VERDOHEME Pocket Volume



Tables

B.1 AA Frequency

B.2 CACBFe Data

B.3 Coordinating Residue Data

Table B.9: perhaps

PDB_ID	Residue_Number	Distance	Angle	Residue_Code
1B2V	128	7.698	62.901	HIS
1B2V	133	7.967	34.190	HIS
1B2V	137	6.042	27.438	TYR
1B2V	32	2.694	51.415	HIS
1B2V	55	3.537	15.312	TYR
1B2V	75	1.611	39.160	TYR
1B2V	83	4.010	56.778	HIS
1B2V	86	10.019	89.073	TYR
1B5M	30	6.817	64.022	TYR
1B5M	39	2.727	87.693	HIS
1B5M	63	1.901	71.272	HIS
1B5M	74	8.245	36.329	TYR
1DK0	128	7.430	68.576	HIS
1DK0	133	7.840	40.101	HIS
1DK0	137	5.957	32.086	TYR
1DK0	32	2.549	48.497	HIS
1DK0	55	3.779	19.455	TYR
1DK0	75	1.716	40.042	TYR

Table B.9: perhaps *(continued)*

PDB_ID	Residue_Number	Distance	Angle	Residue_Code
1DK0	83	3.979	62.320	HIS
1DKH	128	9.160	41.849	HIS
1DKH	133	7.963	28.255	HIS
1DKH	137	6.532	26.390	TYR
1DKH	32	2.969	50.187	HIS
1DKH	55	3.321	11.809	TYR
1DKH	75	2.303	45.976	TYR
1DKH	83	3.350	43.522	HIS
1DKH	86	9.941	74.151	TYR
1ICC	30	7.787	63.247	TYR
1ICC	39	2.907	78.752	HIS
1ICC	63	2.338	57.814	HIS
1ICC	74	9.353	38.522	TYR
1IPH	128	4.997	33.997	HIS
1IPH	212	8.210	67.518	HIS
1IPH	275	5.348	76.198	HIS
1IPH	392	7.521	56.309	HIS
1IPH	415	2.177	62.200	TYR
1N45	114	10.393	54.942	TYR
1N45	134	6.110	73.753	TYR
1N45	137	7.892	81.032	TYR
1N45	25	2.302	69.116	HIS
1P3T	112	6.840	67.730	TYR
1P3T	113	8.134	41.741	CYS
1P3T	184	3.832	79.193	TYR
1P3T	23	2.782	67.542	HIS
1P3T	53	9.402	65.300	HIS
1QHU	175	9.923	88.691	TYR
1QHU	176	5.915	62.683	TYR
1QHU	197	3.965	54.403	TYR
1QHU	204	5.533	47.589	TYR
1QHU	213	2.769	79.430	HIS
1QHU	222	6.023	77.401	HIS
1QHU	224	8.727	63.901	HIS

Table B.9: perhaps *(continued)*

PDB_ID	Residue_Number	Distance	Angle	Residue_Code
1QHU	253	9.228	87.634	HIS
1QHU	265	2.093	83.910	HIS
1QHU	271	4.447	65.497	HIS
1QHU	56	7.104	61.749	HIS
1QJS	175	9.211	88.749	TYR
1QJS	176	5.217	60.789	TYR
1QJS	197	3.368	56.340	TYR
1QJS	204	5.912	48.525	TYR
1QJS	213	3.425	82.802	HIS
1QJS	217	7.939	50.557	HIS
1QJS	225	8.182	81.298	HIS
1QJS	254	8.866	87.803	HIS
1QJS	266	2.146	82.026	HIS
1QJS	272	4.375	61.689	HIS
1QJS	56	6.112	65.715	HIS
1SI8	197	4.659	73.198	HIS
1SI8	337	2.133	58.339	TYR
1SI8	341	6.477	59.843	HIS
1SI8	54	4.940	26.890	HIS
1SY2	104	9.925	88.447	TYR
1SY2	105	8.182	68.754	TYR
1SY2	120	8.228	63.661	HIS
1SY2	124	7.544	61.698	HIS
1SY2	134	5.245	80.180	TYR
1SY2	17	9.754	62.084	TYR
1SY2	28	6.477	43.207	TYR
1SY2	40	5.181	30.456	TYR
1SY2	41	6.450	67.587	CYS
1SY2	58	7.252	86.657	TYR
1SY2	59	2.120	85.351	HIS
1SY2	60	8.111	80.724	TYR
1SY2	69	8.550	56.645	TYR
1U9U	30	5.740	62.876	TYR
1U9U	39	2.971	80.451	HIS

Table B.9: perhaps *(continued)*

PDB_ID	Residue_Number	Distance	Angle	Residue_Code
1U9U	58	4.883	76.301	TYR
1U9U	63	2.015	66.393	HIS
1VGI	114	10.625	53.936	TYR
1VGI	134	6.540	71.010	TYR
1VGI	137	7.705	77.807	TYR
1VGI	25	2.587	72.142	HIS
1VGI	58	10.081	74.439	TYR
1ZVI	415	3.064	46.871	CYS
1ZVI	588	8.761	51.436	TYR
1ZVI	706	4.912	89.229	TYR
2BHJ	194	2.661	52.816	CYS
2BHJ	361	12.114	68.203	CYS
2BHJ	367	8.780	48.571	TYR
2BHJ	483	9.006	20.946	TYR
2BHJ	484	10.510	86.316	TYR
2BHJ	485	4.831	83.349	TYR
2CJ0	105	4.893	56.765	HIS
2CJ0	107	6.057	78.344	HIS
2CJ0	29	2.636	47.217	CYS
2CN4	128	7.392	64.449	HIS
2CN4	133	7.798	35.290	HIS
2CN4	137	6.148	28.073	TYR
2CN4	55	3.374	16.581	TYR
2CN4	75	1.571	45.523	TYR
2CN4	83	3.639	61.039	HIS
2CPO	105	5.051	56.257	HIS
2CPO	107	6.010	77.723	HIS
2CPO	29	2.568	49.291	CYS
2E2Y	103	5.279	62.538	TYR
2E2Y	93	2.884	86.534	HIS
2E2Y	97	3.306	68.715	HIS
2FC2	124	8.199	67.859	HIS
2FC2	235	8.971	51.247	TYR
2FC2	351	8.654	20.451	TYR

Table B.9: perhaps *(continued)*

PDB_ID	Residue_Number	Distance	Angle	Residue_Code
2FC2	353	5.379	80.273	TYR
2FC2	62	2.737	54.005	CYS
2IIZ	185	2.384	59.725	TYR
2IIZ	187	6.234	26.392	HIS
2IIZ	224	2.426	61.464	HIS
2IIZ	266	11.868	34.256	HIS
2IIZ	272	11.141	87.103	HIS
2IPS	109	5.096	73.103	HIS
2IPS	312	8.997	60.684	TYR
2IPS	351	3.260	28.391	HIS
2IPS	444	5.609	58.367	HIS
2J0P	196	2.433	75.104	HIS
2J0P	318	0.747	79.570	TYR
2J0P	85	7.030	89.141	HIS
2J0P	89	9.449	71.784	HIS
2J18	105	4.950	57.689	HIS
2J18	107	6.040	78.324	HIS
2J18	29	2.562	47.527	CYS
2O6P	120	7.278	85.869	TYR
2O6P	132	2.414	56.191	TYR
2O6P	134	7.422	61.077	HIS
2O6P	135	8.862	51.119	HIS
2O6P	136	3.338	86.464	TYR
2O6P	39	9.838	81.179	TYR
2O6P	41	8.850	82.547	TYR
2O6P	52	2.644	77.760	TYR
2O6P	74	6.398	86.342	HIS
2Q6N	369	5.361	52.983	HIS
2Q6N	436	1.918	40.993	CYS
2R7A	201	6.780	70.643	TYR
2R7A	67	2.239	73.259	TYR
2SPL	103	5.381	59.825	TYR
2SPL	64	4.293	73.719	HIS
2SPL	93	2.890	88.954	HIS

Table B.9: perhaps *(continued)*

PDB_ID	Residue_Number	Distance	Angle	Residue_Code
2SPL	97	3.454	67.846	HIS
2VEB	11	10.694	45.226	TYR
2VEB	112	5.772	72.895	TYR
2VEB	120	2.602	79.839	HIS
2VEB	121	8.762	74.095	HIS
2VEB	139	9.078	84.230	TYR
2VEB	141	4.252	67.226	TYR
2VEB	147	7.919	35.550	TYR
2VEB	61	7.823	23.496	TYR
2VEB	69	9.700	45.126	HIS
2VEB	72	10.947	86.614	TYR
2VEB	73	8.395	36.659	TYR
2VEB	85	7.353	68.252	TYR
2VEB	9	7.229	61.290	TYR
3HX9	25	10.506	40.629	HIS
3HX9	28	12.242	77.576	HIS
3HX9	56	10.073	84.565	HIS
3HX9	75	4.713	50.709	HIS
3HX9	78	7.572	67.872	HIS
3MVF	104	9.681	89.023	TYR
3MVF	105	8.198	68.065	TYR
3MVF	124	7.204	65.593	HIS
3MVF	134	5.389	81.636	TYR
3MVF	28	6.166	45.316	TYR
3MVF	40	7.066	4.606	TYR
3MVF	41	6.200	62.235	CYS
3MVF	58	7.239	83.767	TYR
3MVF	59	2.119	87.977	HIS
3MVF	69	8.541	55.944	TYR
3QZN	158	9.826	89.747	HIS
3QZN	168	7.173	70.767	HIS
3QZN	170	3.267	72.518	TYR
3QZN	83	2.699	67.495	HIS
3QZN	87	3.369	84.821	TYR

Table B.9: perhaps *(continued)*

PDB_ID	Residue_Number	Distance	Angle	Residue_Code
3QZZ	112	6.443	67.761	TYR
3QZZ	120	2.589	74.693	HIS
3QZZ	121	8.549	75.618	HIS
3QZZ	139	9.018	88.169	TYR
3QZZ	141	4.386	72.731	TYR
3QZZ	147	7.440	38.893	TYR
3QZZ	69	10.498	44.008	HIS
3QZZ	72	11.678	89.789	TYR
3QZZ	73	8.935	40.394	TYR
3QZZ	85	7.238	67.945	TYR
3QZZ	9	7.361	61.739	TYR
3SIK	136	2.207	52.942	TYR
3SIK	140	2.707	63.829	TYR
3SIK	58	4.929	80.739	TYR
3TGC	104	9.882	87.772	TYR
3TGC	105	8.089	69.194	TYR
3TGC	120	8.242	64.898	HIS
3TGC	124	7.654	60.860	HIS
3TGC	134	5.409	80.162	TYR
3TGC	17	9.646	62.418	TYR
3TGC	28	6.645	42.557	TYR
3TGC	40	5.341	29.632	TYR
3TGC	41	6.277	64.234	CYS
3TGC	58	7.103	88.065	TYR
3TGC	59	2.123	87.207	HIS
3TGC	69	8.528	63.416	TYR
3VP5	108	7.320	84.173	TYR
3VP5	149	2.237	49.264	HIS
3VP5	169	7.299	73.135	TYR
3VP5	72	2.856	45.918	HIS
3VP5	91	2.713	32.406	TYR
3ZJS	11	11.512	38.869	TYR
3ZJS	112	6.683	68.690	TYR
3ZJS	120	2.355	73.923	HIS

Table B.9: perhaps *(continued)*

PDB_ID	Residue_Number	Distance	Angle	Residue_Code
3ZJS	121	8.548	74.154	HIS
3ZJS	139	9.181	86.879	TYR
3ZJS	141	4.273	73.885	TYR
3ZJS	147	6.821	34.613	TYR
3ZJS	61	7.100	42.808	TYR
3ZJS	69	11.236	20.941	HIS
3ZJS	72	11.509	89.819	TYR
3ZJS	73	8.882	36.925	TYR
3ZJS	85	6.847	70.295	TYR
3ZJS	9	7.722	56.668	TYR
4B8N	39	6.321	69.096	TYR
4B8N	48	2.856	87.524	HIS
4B8N	71	1.755	70.933	HIS
4B8N	79	8.264	10.051	TYR
4CDP	193	2.409	74.031	HIS
4CDP	315	0.578	83.078	TYR
4CDP	83	6.916	82.362	HIS
4CDP	87	9.545	73.878	HIS
4I3Q	307	5.653	50.994	TYR
4I3Q	442	2.140	34.781	CYS
4JET	135	8.219	65.297	HIS
4JET	140	8.332	30.141	HIS
4JET	55	2.434	11.357	TYR
4JET	75	1.844	47.089	TYR
4JET	81	3.516	54.183	HIS
4MF9	209	2.485	63.203	HIS
4MF9	210	9.455	50.161	HIS
4MF9	212	10.680	70.794	HIS
4MYP	209	6.410	71.980	TYR
4MYP	280	2.119	56.836	TYR
4MYP	289	5.495	20.187	TYR
4NL5	25	11.173	26.069	HIS
4NL5	28	11.403	84.281	HIS
4NL5	75	2.734	46.347	HIS

Table B.9: perhaps *(continued)*

PDB_ID	Residue_Number	Distance	Angle	Residue_Code
4NL5	78	7.980	36.780	HIS
4UZV	106	2.473	79.507	HIS
4UZV	109	6.542	40.788	TYR
4UZV	143	10.030	21.014	TYR
4UZV	146	7.929	59.304	TYR
4UZV	149	6.704	31.972	TYR
4UZV	93	5.892	59.726	TYR
4XZD	135	7.425	73.244	HIS
4XZD	140	7.989	42.711	HIS
4XZD	55	3.139	12.231	TYR
4XZD	75	1.999	46.839	TYR
4XZD	81	3.940	67.684	HIS
4XZD	84	10.535	68.861	HIS
4Y1Q	135	8.162	68.871	HIS
4Y1Q	140	8.287	35.206	HIS
4Y1Q	55	2.854	8.751	TYR
4Y1Q	81	3.830	61.474	HIS
4Y1Q	84	10.097	68.905	HIS
5CN5	103	5.250	60.339	TYR
5CN5	36	11.713	23.540	HIS
5CN5	64	3.729	84.840	HIS
5CN5	93	2.970	82.799	HIS
5CN5	97	3.227	71.762	HIS
5GJ3	140	3.091	58.494	TYR
5GJ3	239	3.919	62.993	TYR
5GJ3	332	10.593	66.350	TYR
5KZL	103	10.371	47.478	TYR
5KZL	123	7.226	70.521	TYR
5KZL	126	8.028	69.074	TYR
5KZL	15	2.756	59.949	HIS
5O1L	153	9.769	76.368	TYR
5O1L	154	10.556	60.343	TYR
5O1L	198	2.803	66.467	HIS
5O1L	232	9.116	44.567	HIS

Table B.9: perhaps *(continued)*

PDB_ID	Residue_Number	Distance	Angle	Residue_Code
5O1M	153	9.343	70.903	TYR
5O1M	154	9.770	65.166	TYR
5O1M	198	2.839	64.463	HIS
5O1M	232	8.715	48.734	HIS
5VEU	307	6.114	62.140	TYR
5VEU	441	2.030	42.614	CYS
6A2J	123	10.183	45.992	HIS
6A2J	171	9.787	69.068	TYR
6A2J	173	11.472	28.364	TYR
6A2J	177	8.156	51.730	TYR
6A2J	181	6.110	82.785	TYR
6A2J	184	5.553	82.865	HIS
6A2J	191	12.388	81.985	CYS
6A2J	216	2.542	63.468	HIS
6A2J	278	2.230	63.931	HIS
7C74	109	5.222	70.733	HIS
7C74	312	8.496	58.176	TYR
7C74	351	3.380	25.953	HIS
7C74	441	9.298	6.308	CYS
7C74	444	5.535	54.836	HIS
7DMR	109	5.094	62.306	HIS
7DMR	312	9.451	63.699	TYR
7DMR	351	3.061	31.126	HIS
7DMR	441	9.664	9.491	CYS
7DMR	444	5.076	55.605	HIS

Table B.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
GLN	78
MET	72
TRP	60
CYS	17

Table B.2: HEC AA Freq

Residue	Freq
LEU	71
ALA	59
GLY	49
ILE	43
PHE	42
LYS	41
THR	40
VAL	38
PRO	34
TYR	34
ARG	31
SER	30
CYS	28
ASN	25
HIS	25
MET	25
GLN	18
TRP	18
ASP	17
GLU	12

Table B.3: 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

Table B.4: 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
ASN	7
HIS	7
MET	7
THR	7
GLN	6
ILE	6
ASP	4

Table B.5: HEM CACBFe Data

PDB_ID	Residue_Number	Residue_Code	Angle
1B2V	137	TYR	107.0750
1B2V	140	MET	173.7920
1B2V	32	HIS	116.3150
1B2V	37	VAL	150.5390
1B2V	41	ASN	79.4068
1B2V	42	SER	82.8367
1B2V	75	TYR	132.4540
1B2V	77	LEU	57.1497
1B2V	83	HIS	102.9160
1B2V	84	THR	18.8827
1B5M	35	PHE	126.8820
1B5M	39	HIS	101.8130
1B5M	40	PRO	84.9302
1B5M	45	VAL	132.2220
1B5M	46	LEU	104.5310
1B5M	58	PHE	85.0021
1B5M	61	VAL	142.4900
1B5M	63	HIS	125.8380
1B5M	67	ALA	143.9450
1DK0	137	TYR	107.9930
1DK0	140	MET	173.4760
1DK0	32	HIS	116.4470
1DK0	33	THR	13.7171
1DK0	37	VAL	154.2260
1DK0	41	ASN	80.6960
1DK0	42	SER	80.4760
1DK0	75	TYR	131.4420
1DK0	77	LEU	58.1793
1DK0	83	HIS	102.7520
1DK0	84	THR	19.3165
1DKH	137	TYR	103.9420
1DKH	140	MET	172.2070
1DKH	32	HIS	121.3750
1DKH	37	VAL	149.8520
1DKH	42	SER	32.8371
1DKH	75	TYR	125.4210
1DKH	77	LEU	66.1552
1DKH	83	HIS	122.9600
1DKH	84	THR	31.3703
1ICC	35	PHE	121.2740
1ICC	39	HIS	101.5070
1ICC	40	PRO	84.5709
1ICC	45	VAL	128.6010

Table B.6: HEC CACBFe Data

PDB_ID	Residue_Number	Residue_Code	Angle
1BBH	121	CYS	88.6062
1BBH	124	CYS	118.4660
1BBH	125	HIS	95.2502
1BBH	129	ARG	148.1750
1BBH	16	TYR	126.0380
1BBH	17	GLU	46.8470
1BBH	19	MET	132.1620
1BBH	58	TYR	118.4030
1S56	126	VAL	116.6120
1S56	33	TYR	98.2768
1S56	46	PHE	100.7840
1S56	54	LEU	117.0640
1S56	58	GLN	114.9080
1S56	77	MET	79.9304
1S56	80	VAL	122.1110
1S56	81	HIS	112.6780
1S56	86	ILE	163.7880
1S56	94	VAL	156.6730
1W2L	18	CYS	83.0319
1W2L	21	CYS	129.4480
1W2L	22	HIS	122.1140
1W2L	32	PRO	80.5165
1W2L	34	PHE	94.2433
1W2L	60	SER	107.3410
1W2L	61	ILE	64.6202
1W2L	75	VAL	68.5700
1W2L	76	MET	95.5351
1W2L	77	PRO	84.7339
1W2L	80	TYR	159.9880
2BC5	10	LEU	145.5220
2BC5	101	CYS	122.7380
2BC5	102	HIS	96.2948
2BC5	106	ARG	119.2950
2BC5	3	LEU	93.4646
2BC5	65	PHE	87.4034
2BC5	7	MET	112.0730
2BC5	98	CYS	83.1994
2BC5	99	ASN	26.5703
2BH5	100	LYS	174.4600
2BH5	102	PHE	125.9060
2BH5	15	CYS	93.4388
2BH5	18	CYS	129.9250
2BH5	19	HIS	122.4230

Table B.7: SRM CACBFe Data

PDB_ID	Residue_Number	Residue_Code	Angle
1ZJ8	129	ASP	96.5485
1ZJ8	134	GLN	147.3840
1ZJ8	166	ARG	157.1260
1ZJ8	207	LYS	172.2200
1ZJ8	209	LYS	132.2160
1ZJ8	465	ASN	126.9150
1ZJ8	466	SER	46.1914
1ZJ8	467	CYS	106.8380
1ZJ8	468	ALA	54.3434
1ZJ8	69	TYR	168.2380
1ZJ8	97	ARG	148.8370
2AKJ	109	ARG	148.4620
2AKJ	142	THR	112.5850
2AKJ	179	ARG	150.8160
2AKJ	224	LYS	179.3020
2AKJ	484	ASN	125.4620
2AKJ	485	SER	45.1203
2AKJ	486	CYS	106.3630
2AOP	116	ASN	95.1407
2AOP	121	GLN	146.9480
2AOP	153	ARG	144.7120
2AOP	215	LYS	157.3800
2AOP	217	LYS	135.7480
2AOP	481	ASN	121.7600
2AOP	483	CYS	115.6650
2AOP	83	ARG	162.1930
3B0G	109	ARG	157.7590
3B0G	142	THR	114.5110
3B0G	179	ARG	150.2730
3B0G	224	LYS	175.7930
3B0G	483	ASN	124.8060
3B0G	484	THR	31.8530
3B0G	485	CYS	114.2180
3B0G	486	ALA	52.3271
3VKP	109	ARG	159.0060
3VKP	142	THR	114.3200
3VKP	179	ARG	149.5410
3VKP	224	LYS	175.8260
3VKP	483	ASN	125.9350
3VKP	484	THR	32.2678
3VKP	485	CYS	113.1560
3VKP	486	ALA	52.9419
3VLX	109	ARG	157.8390

Table B.8: VERDOHEME CACBFe Data

PDB_ID	Residue_Number	Residue_Code	Angle
2ZVU	135	THR	58.6713
2ZVU	138	LEU	76.7687
2ZVU	140	ASP	26.6732
2ZVU	142	SER	126.2110
2ZVU	207	PHE	105.4400
2ZVU	25	HIS	110.7510
2ZVU	28	ALA	120.0970
2ZVU	29	GLU	117.2590
3MOO	131	VAL	60.1702
3MOO	134	LEU	77.1733
3MOO	136	ASP	23.6316
3MOO	138	SER	125.3120
3MOO	20	HIS	111.0890
3MOO	201	PHE	104.2170
3MOO	24	GLU	110.6430
1TWN	135	THR	60.8151
1TWN	138	LEU	75.1317
1TWN	140	ASP	27.4847
1TWN	142	SER	114.7150
1TWN	207	PHE	105.9700
1TWN	25	HIS	113.5160
1TWN	29	GLU	100.1730
1TWR	138	LEU	75.0669
1TWR	140	ASP	27.4184
1TWR	142	SER	129.2760
1TWR	207	PHE	107.0750
1TWR	25	HIS	108.8640
1TWR	29	GLU	103.1100

Table B.10: HEC Coordinating Residue Data

	PDB_ID	Residue_Number	Distance	Angle	Residue_Code
6	1BBH	121	2.890	69.070	CYS
9	1BBH	124	5.495	73.170	CYS
10	1BBH	125	3.167	89.456	HIS
11	1BBH	128	8.052	42.662	TYR
17	1BBH	16	3.180	83.790	TYR
25	1BBH	58	5.857	74.986	TYR
52	1S56	33	5.860	73.693	TYR
74	1S56	72	6.816	73.544	TYR
81	1S56	81	2.639	80.865	HIS
86	1S56	90	5.880	69.009	HIS
93	1W2L	13	7.978	56.895	TYR
96	1W2L	18	3.700	79.901	CYS
99	1W2L	21	4.996	50.740	CYS
100	1W2L	22	2.258	62.051	HIS
114	1W2L	56	2.981	82.425	TYR
127	1W2L	72	6.533	88.308	TYR
133	1W2L	80	1.717	80.939	TYR
140	2BC5	101	5.489	89.234	CYS
141	2BC5	102	2.941	82.850	HIS
142	2BC5	105	7.621	44.145	TYR
162	2BC5	63	10.176	35.590	HIS
176	2BC5	98	2.814	62.529	CYS
190	2BH5	15	3.173	80.908	CYS
193	2BH5	18	5.226	56.447	CYS
194	2BH5	19	2.432	56.825	HIS
209	2BH5	55	3.172	79.925	TYR
217	2BH5	79	3.392	66.731	TYR
231	3EAH	150	2.557	47.992	CYS
250	3EAH	323	8.876	55.586	TYR
263	3EAH	441	4.705	85.347	TYR
267	3X15	12	2.836	75.877	CYS
270	3X15	15	5.324	60.268	CYS
271	3X15	16	2.626	56.339	HIS
279	3X15	34	6.546	79.080	TYR
300	4B2N	191	4.436	51.391	CYS
303	4B2N	194	6.387	46.261	CYS
304	4B2N	195	2.206	82.691	HIS
348	4B2N	367	8.067	15.008	TYR
358	4B2N	390	4.550	58.866	CYS
361	4B2N	393	5.897	49.659	CYS
362	4B2N	394	2.083	86.758	HIS
377	4B2N	501	6.883	69.986	TYR
386	4B2N	515	7.504	74.964	TYR

Table B.11: SRM Coordinating Residue Data

	PDB_ID	Residue_Number	Distance	Angle	Residue_Code
10	1ZJ8	135	6.040	66.721	TYR
11	1ZJ8	136	3.878	31.155	HIS
13	1ZJ8	161	7.230	55.869	CYS
41	1ZJ8	417	6.941	26.616	CYS
45	1ZJ8	423	6.211	42.879	CYS
49	1ZJ8	463	9.342	73.879	CYS
53	1ZJ8	467	3.281	87.220	CYS
58	1ZJ8	69	6.634	17.492	TYR
91	2AKJ	238	13.702	48.274	HIS
105	2AKJ	308	7.951	86.710	CYS
113	2AKJ	441	6.801	23.979	CYS
117	2AKJ	447	5.921	36.892	CYS
122	2AKJ	482	9.088	72.772	CYS
126	2AKJ	486	3.299	86.391	CYS
132	2AKJ	97	9.584	59.448	HIS
144	2AOP	123	3.090	39.920	HIS
166	2AOP	309	7.118	45.159	TYR
173	2AOP	434	6.661	23.660	CYS
179	2AOP	440	6.319	36.167	CYS
183	2AOP	479	9.195	74.728	CYS
187	2AOP	483	3.190	85.931	CYS
192	2AOP	84	10.533	38.954	CYS
217	3B0G	238	11.585	50.847	HIS
230	3B0G	308	9.447	72.387	CYS
237	3B0G	440	7.055	40.691	CYS
241	3B0G	446	5.032	24.524	CYS
247	3B0G	481	9.209	57.400	CYS
251	3B0G	485	3.287	73.017	CYS
257	3B0G	97	7.720	46.114	HIS
283	3VKP	238	12.998	48.519	HIS
296	3VKP	308	8.440	84.519	CYS
303	3VKP	440	6.837	28.134	CYS
307	3VKP	446	5.941	35.235	CYS
313	3VKP	481	9.358	66.162	CYS
317	3VKP	485	3.289	84.887	CYS
323	3VKP	97	9.436	56.728	HIS
348	3VLX	238	13.170	48.585	HIS
367	3VLX	440	6.821	27.250	CYS
371	3VLX	446	6.030	34.574	CYS
377	3VLX	481	9.390	67.736	CYS
381	3VLX	485	3.276	85.502	CYS
388	3VLX	97	9.567	58.570	HIS
413	3VLY	238	12.955	49.000	HIS

Table B.12: VERODHEME Coordinating Residue Data

	PDB_ID	Residue_Number	Distance	Angle	Residue_Code
1	2ZVU	114	9.736	45.880	TYR
3	2ZVU	134	7.026	63.216	TYR
6	2ZVU	137	8.229	70.326	TYR
31	2ZVU	25	2.497	70.790	HIS
40	2ZVU	58	9.916	66.224	TYR
42	3MOO	109	10.966	53.677	TYR
44	3MOO	129	12.192	56.279	HIS
46	3MOO	130	6.634	70.461	TYR
49	3MOO	133	7.960	73.999	TYR
66	3MOO	20	2.692	65.389	HIS
70	3MOO	205	8.691	18.767	HIS
74	3MOO	25	7.594	36.606	HIS
11	1TWN	114	11.189	76.003	TYR
32	1TWN	134	3.339	83.511	TYR
61	1TWN	137	7.144	80.837	TYR
33	1TWN	25	2.517	82.070	HIS
41	1TWR	114	11.476	76.363	TYR
421	1TWR	134	3.210	81.340	TYR
45	1TWR	137	7.391	78.764	TYR
701	1TWR	25	2.464	75.802	HIS

Table B.13: HEM Molecule and Organism Data

PDB_ID	Molecule_Name	Source_Organism
1B2V	PROTEIN (HEME-BINDING PROTEIN A);	SERRATIA MARCESCENS;
1B5M	CYTOCHROME B5;	RATTUS NORVEGICUS;
1DK0	HEME-BINDING PROTEIN A;	SERRATIA MARCESCENS;
1DKH	HEME-BINDING PROTEIN A;	SERRATIA MARCESCENS;
1ICC	CYTOCHROME B5 OUTER MITOCHONDRIAL MEMBRANE	RATTUS NORVEGICUS;
1IPH	CATALASE HPII;	ESCHERICHIA COLI;
1N45	HEME OXYGENASE 1;	HOMO SAPIENS;
1P3T	HEME OXYGENASE 1;	NEISSERIA MENINGITIDIS;
1QHU	PROTEIN (HEMOPEXIN);	ORYCTOLAGUS CUNICULUS;
1QJS	HEMOPEXIN;	ORYCTOLAGUS CUNICULUS;
1SI8	CATALASE;	ENTEROCOCCUS FAECALIS;
1SY2	NITROPHORIN 4;	RHODNIUS PROLIXUS;
1U9U	CYTOCHROME B5;	BOS TAURUS;
1VGI	HEME OXYGENASE 1;	RATTUS NORVEGICUS;
1ZVI	NITRIC-OXIDE SYNTHASE, BRAIN;	RATTUS NORVEGICUS;
2BHJ	NITRIC OXIDE SYNTHASE;	MUS MUSCULUS;
2CJ0	CHLOROPEROXIDASE;	CALDARIOMYCES FUMAGO;
2CN4	HEMOPHORE HASA;	SERRATIA MARCESCENS;
2CPO	CHLOROPEROXIDASE;	LEPTOXYPHIUM FUMAGO;
2E2Y	MYOGLOBIN;	PHYSETER CATODON;
2FC2	NITRIC OXIDE SYNTHASE;	BACILLUS SUBTILIS;
2IIZ	MELANIN BIOSYNTHESIS PROTEIN TYRA, PUTATIVE;	SHEWANELLA ONEIDENSIS;
2IPS	LACTOPEROXIDASE;	BOS TAURUS;
2J0P	HEMIN TRANSPORT PROTEIN HEMS;	YERSINIA ENTEROCOLITICA;
2J18	CHLOROPEROXIDASE;	CALDARIOMYCES FUMAGO;
2O6P	IRON-REGULATED SURFACE DETERMINANT PROTEIN C;	STAPHYLOCOCCUS AUREUS SUBSP. AUREUS;
2Q6N	CYTOCHROME P450 2B4;	ORYCTOLAGUS CUNICULUS;
2R7A	BACTERIAL HEME BINDING PROTEIN;	SHIGELLA DYSENTERIAE;
2SPL	MYOGLOBIN;	PHYSETER CATODON;
2VEB	PROTOGLOBIN;	METHANOSARCINA ACETIVORANS;
3HX9	PROTEIN RV3592;	MYCOBACTERIUM TUBERCULOSIS;
3MVF	NITROPHORIN-4;	RHODNIUS PROLIXUS;
3QZN	IRON-REGULATED SURFACE DETERMINANT PROTEIN A;	STAPHYLOCOCCUS AUREUS SUBSP. AUREUS;
3QZZ	METHANOSARCINA ACETIVORANS PROTOGLOBIN;	METHANOSARCINA ACETIVORANS;
3SIK	CONSERVED DOMAIN PROTEIN;	BACILLUS ANTHRACIS;
3TGC	NITROPHORIN-4;	RHODNIUS PROLIXUS;
3VP5	TRANSCRIPTIONAL REGULATOR;	LACTOCOCCUS LACTIS;
3ZJS	PROTOGLOBIN;	METHANOSARCINA ACETIVORANS;
4B8N	CYTOCHROME B5-HOST ORIGIN;	OSTREOCOCCUS TAURI VIRUS 2;
4CAT	CATALASE;	PENICILLIUM JANTHINELLUM;
4CDP	PUTATIVE HEME/HEMOGLOBIN TRANSPORT PROTEIN;	ESCHERICHIA COLI;
4I3Q	CYTOCHROME P450 3A4;	HOMO SAPIENS;
4JET	HEMOPHORE HASA;	YERSINIA PESTIS;
4MF9	HEMIN DEGRADING FACTOR;	PSEUDOMONAS AERUGINOSA;
4MYP	IRON-REGULATED SURFACE DETERMINANT PROTEIN A;	LISTERIA MONOCYTOGENES;
4NL5	HEME-DEGRADING MONOOXYGENASE HMOB;	MYCOBACTERIUM TUBERCULOSIS;
4UZV	HEMOGLOBIN;	THERMOBIFIDA FUSCA TM51;
4XZD	EXTRACELLULAR HEME ACQUISITION HEMOPHORE HASA;	YERSINIA PSEUDOTUBERCULOSIS IP 32953;
4Y1Q	EXTRACELLULAR HEME ACQUISITION HEMOPHORE HASA;	YERSINIA PSEUDOTUBERCULOSIS IP 32953;
5CN5	MYOGLOBIN;	EQUUS CABALLUS;
5GJ3	PERIPLASMIC BINDING PROTEIN;	ROSEIFLEXUS SP. RS-1;
5KZL	HEME OXYGENASE;	LEPTOSPIRA INTERROGANS;
5O1L	RUBBER OXYGENASE;	STREPTOMYCES SP. (STRAIN K30);
5O1M	RUBBER OXYGENASE;	STREPTOMYCES SP. (STRAIN K30);
5VEU	CYTOCHROME P450 3A5;	HOMO SAPIENS;
6A2J	HEME A SYNTHASE;	BACILLUS SUBTILIS (STRAIN 168);
7C74	LACTOPEROXIDASE;	BOS MUTUS;
7DMR	LACTOPEROXIDASE;	BOS MUTUS;

Table B.14: HEC Molecule and Organism Data

PDB_ID	Molecule_Name	Source_Organism
1BBH	CYTOCHROME C';	ALLOCHROMATIUM VINOSUM;
1S56	HEMOGLOBIN-LIKE PROTEIN HBN;	MYCOBACTERIUM TUBERCULOSIS;
1W2L	CYTOCHROME OXIDASE SUBUNIT II;	RHODOTHERMUS MARINUS;
2BC5	SOLUBLE CYTOCHROME B562;	ESCHERICHIA COLI;
2BH5	CYTOCHROME C-550;	PARACOCCUS VERSUTUS;
3EAH	NITRIC OXIDE SYNTHASE, ENDOTHELIAL;	HOMO SAPIENS;
3X15	CYTOCHROME C552;	AQUIFEX AEOLICUS VF5;
4B2N	70 KDA PROTEIN;	XANTHOMONAS SP.;
5KPF	CYTOCHROME C ISO-1;	SACCHAROMYCES CEREVISIAE;
5LFT	CYTOCHROME C ISO-1;	SACCHAROMYCES CEREVISIAE;
5T8W	CYC1P;	SACCHAROMYCES CEREVISIAE;
6VDQ	3-METHYL-L-TYROSINE PEROXYGENASE;	STREPTOMYCES LAVENDULAE;
6WZA	SOLUBLE CYTOCHROME B562;	ESCHERICHIA COLI;
6XNK	CYTOCHROME C;	HOMO SAPIENS;

Table B.15: SRM Molecule and Organism Data

PDB_ID	Molecule_Name	Source_Organism
1ZJ8	PROBABLE FERREDOXIN-DEPENDENT NITRITE REDUCTASE NIRA;	MYCOBACTERIUM TUBERCULOSIS;
2AKJ	FERREDOXIN–NITRITE REDUCTASE, CHLOROPLAST;	SPINACIA OLERACEA;
2AOP	SULFITE REDUCTASE HEMOPROTEIN;	ESCHERICHIA COLI;
3B0G	NITRITE REDUCTASE;	NICOTIANA TABACUM;
3VKP	NITRITE REDUCTASE;	NICOTIANA TABACUM;
3VLX	NITRITE REDUCTASE;	NICOTIANA TABACUM;
3VLY	NITRITE REDUCTASE;	NICOTIANA TABACUM;
3VLZ	NITRITE REDUCTASE;	NICOTIANA TABACUM;
5H8V	SULFITE REDUCTASE [FERREDOXIN], CHLOROPLASTIC;	ZEA MAYS;

Table B.16: VERDOHEME Molecule and Organism Data

PDB_ID	Molecule_Name	Source_Organism
2ZVU	HEME OXYGENASE 1;	RATTUS NORVEGICUS;
3MOO	HEME OXYGENASE;	CORYNEBACTERIUM DIPHTHERIAE;
1TWN	HEME OXYGENASE 1;	HOMO SAPIENS;
1TWR	HEME OXYGENASE 1;	HOMO SAPIENS;

B.4 Molecule and Organism Data

B.5 Merged Data

B.6 Minimum Distance Data

B.7 Planar Angles

Table B.17: HEM Merged Data

PDB_ID	Volume_Data	HEM_Excluded_SA	HEM_Accessible_SA	Pocket_Excluded_SA	Pocket_Accessible_SA
1B2V	893.60	502.042	820.988	7276.09	8232.60
1B5M	672.79	490.050	800.780	4695.01	5512.20
1DK0	966.72	505.258	837.157	7237.94	8217.58
1DKH	1010.70	509.042	828.131	7402.34	8175.94
1ICC	1000.40	499.585	811.357	5079.72	6028.23
1IPH	1345.60	501.603	814.652	33983.80	34094.40
1N45	978.98	560.384	983.238	9944.50	10779.30
1P3T	987.05	509.939	829.611	9530.67	10410.80
1QHU	1389.20	573.686	1002.160	18503.10	18257.20
1QJS	1102.30	573.266	1000.380	18588.40	18584.10
1SI8	965.57	646.643	1184.070	23711.20	25120.40
1SY2	918.34	501.850	817.749	8960.76	9610.23
1U9U	738.55	496.132	813.773	4675.76	5632.32
1VGI	870.44	577.234	1002.530	9615.29	10248.20
1ZVI	1435.90	701.091	1129.540	19918.60	20968.20
2BHJ	1438.30	836.576	1290.530	20102.30	20762.60
2CJ0	809.62	2653.180	4835.280	12749.60	12892.20
2CN4	526.88	576.760	961.348	9617.23	11917.70
2CPO	886.17	1846.490	3329.540	13081.60	12995.60
2E2Y	994.92	811.270	1607.370	7531.94	8240.75
2FC2	1091.40	1011.190	1669.900	18383.50	18552.10
2IIZ	1015.60	731.342	1393.160	13651.70	14031.40
2IPS	1242.40	618.252	1075.560	27760.50	25814.10
2J0P	1281.80	1030.510	1873.810	15192.90	15871.10
2J18	841.67	1962.990	3556.340	12675.10	12779.00
2O6P	788.05	499.017	822.121	6234.84	7200.43
2Q6N	1030.10	644.365	1040.080	20051.10	19747.50
2R7A	1284.50	507.098	845.182	11255.10	12389.00
2SPL	1055.70	589.706	1029.660	7588.36	8105.94
2VEB	886.06	762.309	1454.750	9840.72	10401.80
3HX9	1844.50	785.442	1168.200	5819.08	7189.03
3MVF	1271.40	576.502	1009.950	8559.24	9573.08
3QZN	726.52	664.858	1221.330	6133.24	7179.49
3QZZ	977.30	496.950	825.255	8523.59	9708.28
3SIK	492.15	498.621	823.565	6495.38	7739.06
3TGC	969.87	524.380	853.710	8712.77	9181.94
3VP5	1094.60	602.790	1050.820	9801.82	10810.80
3ZJS	788.74	528.419	860.137	9568.10	10130.40
4B8N	841.27	569.302	990.216	4560.39	5458.66
4CAT	1933.90	484.341	778.502	28372.40	36788.30
4CDP	1053.70	1425.050	3141.090	14733.50	15887.40
4I3Q	1220.50	510.623	845.108	21946.50	21093.70
4JET	1010.80	495.992	818.131	7887.81	8695.85
4MF9	1286.50	488.695	790.732	15669.80	16791.30
4MYP	610.72	963.019	1834.680	6285.40	7351.53
4NL5	1088.70	576.669	1003.400	5715.52	6894.72
4UZV	1184.10	526.584	844.058	7378.28	8322.74
4XZD	932.14	498.788	816.032	8028.32	8752.50
4Y1Q	952.23	494.939	806.960	7905.84	8785.04
5CN5	1070.30	663.162	1223.640	7629.45	8117.34
5GJ3	1108.20	756.603	1131.670	11394.00	12591.80
5KZL	914.22	483.760	805.567	9662.03	10431.00
5O1L	1438.70	801.519	1447.270	15538.20	16876.00
5O1M	1431.30	493.850	799.331	16096.90	15912.50
5VEU	964.76	993.578	1502.660	20900.80	20425.90
6A2J	1015.90	6183.450	9902.920	14870.30	15888.00
7C74	1155.10	497.527	820.381	26111.40	25094.20
7DMR	1083.40	1049.750	1916.950	26004.00	24563.80

Table B.18: HEC Merged Data

PDB_ID	Volume_Data	HEC_Excluded_SA	HEC_Accessible_SA	Pocket_Excluded_SA	Pocket_Accessible_SA
1BBH	969.51	514.130	829.817	6441.44	7514.06
1S56	1103.60	643.733	1075.840	6711.26	7477.96
1W2L	756.08	702.711	1240.680	5042.58	5485.50
2BC5	1166.20	569.905	997.324	5489.91	6306.02
2BH5	814.15	508.637	844.494	6359.51	6975.70
3EAH	1280.90	993.430	1697.130	18413.40	19313.80
3X15	823.59	496.328	802.584	5722.90	7493.62
4B2N	934.13	1255.940	2122.260	21914.30	22455.80
5KPF	778.79	568.036	1007.680	5485.51	6155.84
5LFT	809.40	1720.870	2719.000	5539.47	6315.96
5T8W	858.74	511.519	848.952	5755.48	6458.40
6VDQ	977.52	510.534	846.299	13399.60	14076.40
6WZA	1040.10	713.997	1095.240	5529.40	6385.75
6XNK	2214.40	499.687	835.610	6737.92	8143.17

Table B.19: SRM Merged Data

PDB_ID	Volume_Data	SRM_Excluded_SA	SRM_Accessible_SA	Pocket_Excluded_SA	Pocket_Accessible_SA
1ZJ8	1960.2	656.508	1036.43	20388.7	21432.8
2AKJ	1810.2	659.667	1041.00	21673.6	20933.7
2AOP	1040.5	682.170	1045.18	18119.8	18016.0
3B0G	1189.9	666.995	1054.40	21496.8	21033.9
3VKP	1178.0	675.050	1049.85	21279.3	20964.9
3VLX	1164.8	667.013	1052.76	21470.0	21037.0
3VLY	1061.8	675.293	1046.41	21476.6	21022.1
3VLZ	1123.2	676.360	1051.40	21433.5	20901.8
5H8V	1360.8	685.850	1052.56	22885.9	22713.3

Table B.20: VERDOHEME Merged Data

PDB_ID	Volume_Data	VERDOHEME_EXCLUDED_SA	VERDOHEME_ACCESSIBLE_SA	POCKET_EXCLUDED_SA	POCKET_ACCESSIBLE_SA
2ZVU	984.51	560.791	969.143	9633.81	10317.3
3MOO	864.48	870.228	1772.07	9371.88	10170.3
1TWN	1145	448.81	759.632	9966.97	10896.8
1TWR	1426	469.982	783.313	9854.01	10775.6

Table B.21: HEM Planar Angles

PDB_ID	Residue_Number	Distance	Angle	Residue_Code
1B2V	75	1.611	39.160	TYR
1B5M	63	1.901	71.272	HIS
1DK0	75	1.716	40.042	TYR
1DKH	34	1.542	85.521	ASN
1ICC	63	2.338	57.814	HIS
1IPH	415	2.177	62.200	TYR
1N45	25	2.302	69.116	HIS
1P3T	23	2.782	67.542	HIS
1QHU	265	2.093	83.910	HIS
1QJS	266	2.146	82.026	HIS
1SI8	337	2.133	58.339	TYR
1SY2	125	1.264	87.748	LYS
1U9U	63	2.015	66.393	HIS
1VGI	34	2.186	58.008	MET
1ZVI	409	1.815	56.622	TRP
2BHJ	188	2.142	55.507	TRP
2CJ0	108	2.241	56.216	SER
2CN4	75	1.571	45.523	TYR
2CPO	108	2.167	52.106	SER
2E2Y	93	2.884	86.534	HIS
2FC2	56	1.708	58.198	TRP
2IIZ	185	2.384	59.725	TYR
2IPS	101	3.075	84.120	MET
2J0P	318	0.747	79.570	TYR
2J18	108	2.222	55.263	SER
2O6P	132	2.414	56.191	TYR
2Q6N	303	1.595	56.344	THR
2R7A	257	2.179	26.488	LEU
2SPL	93	2.890	88.954	HIS
2VEB	145	2.314	71.125	PHE
3HX9	81	2.600	86.961	ASN
3MVF	125	1.169	89.550	LYS
3QZN	83	2.699	67.495	HIS
3QZZ	145	2.258	67.209	PHE
3SIK	136	2.207	52.942	TYR
3TGC	125	1.484	87.095	LYS
3VP5	149	2.237	49.264	HIS