Several significant databases aim to organize the protein universe at a high level, such as Pfam relying on sequence information, and both SCOP and CATH utilizing protein structural data. These databases categorize proteins into families or superfamilies based on measures of either sequence or structural similarity. While these databases are essential for outlining broad structural relationship, they often present conflicting classifications, lacking information on evolutionary relationships among individual superfamily components[21]. SCOP superfamilies contain protein families that are assumed to be evolutionary related based on sequence and structural similarity and functional commonalities. Ludin et.al [21] investigated how ferritin-like proteins are classified across Pfam, SCOP, and CATH. Notably, this superfamily encompasses a diverse range of proteins, including iron-storing ferritins, methane monooxygenases, the small subunit of RNR R2, rubrerythrins, bacterioferritins, Dps (DNA binding protein from starved cells that protects against oxidative DNA damage), and Dps-like proteins. As discussed by Ludin et.al at the superfamily level, the classification of the “ferritine-like” superfamily appears consistent across these databases but does differ in the amount of information provided regarding the relationships and functions of superfamily constituents. So although the classification in all three databases is hierarchical, they do not encompass all level of functional and evolutionary information. The low sequence similarities across this superfamily make it feasible to construct sequence-based phylogenies only for specific subsets. Consequently, addressing this challenge requires efforts to integrate structural information with sequence-based phylogenies Lundin et al. [21] and Malik et al. [22] delved into the evolutionary relationships of this superfamily by creating a phylogenetic network. They employed the distance-based NeighborNet network method, utilizing distances calculated through structure-based alignment methods. To reconstruct the previously published structural phylogeny of the ferritin-like superfamily, we utilized the same protein structures within this superfamily as Lundin et al. and Malik et al. The dataset specifically focuses on the SCOP superfamily, Ferritin-like (a.25.1) encompassing two manually curated protein families: Ferritin (a.25.1.1) and RiboNucleotide Reductase-like [RNR] (a.25.1.2). The “Ferritin” family contains ferritins, bacterioferritins, and Dodecameric ferritin homolog (Dps) proteins and the “RiboNucleotide Reductase-like” family contains the activating subunit of class I ribonucleotide reductase (RNR R2), BMM, and Fads. Following this, we computed the Structure Profile Energetics (SPE) for each protein and determined the distances between SPEs. The reconstruction of the phylogenetic tree was achieved using the neighbor-joining method[23].

Our findings indicate that energetic phylogenies of the ferritin-like superfamily reveal meaningful relationships among superfamily members, aligning with known evolutionary connections and functional roles. A key conclusion consistent from the previous structural phylogeny of this superfamily (Lundin et al. 2012) was that all three classification systems have just one Ferritin family, which we reproduce at a high level. Our results that is consistent with the previous results suggest that this family could be further split into four subgroups, separating out “Dps and related”, “Rubrerythrin”, and “Bacterioferritins” from the remainder of the Ferritins.

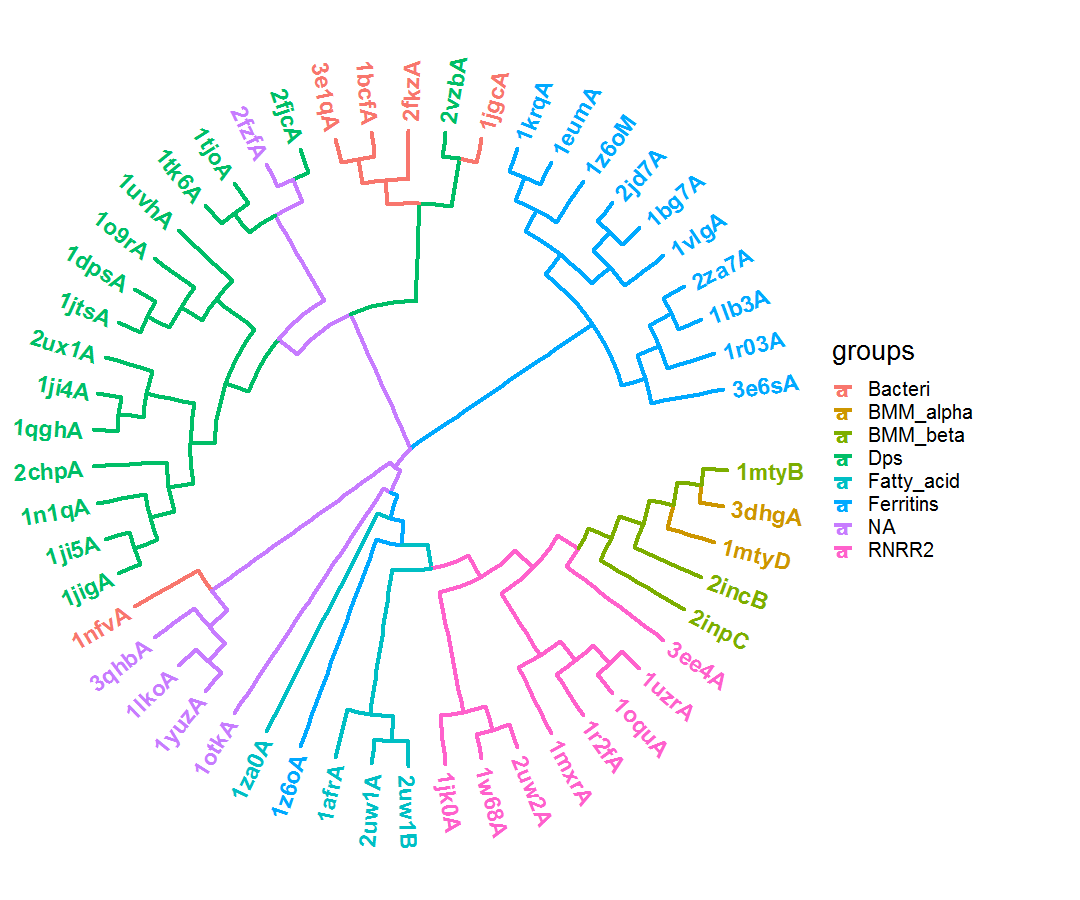
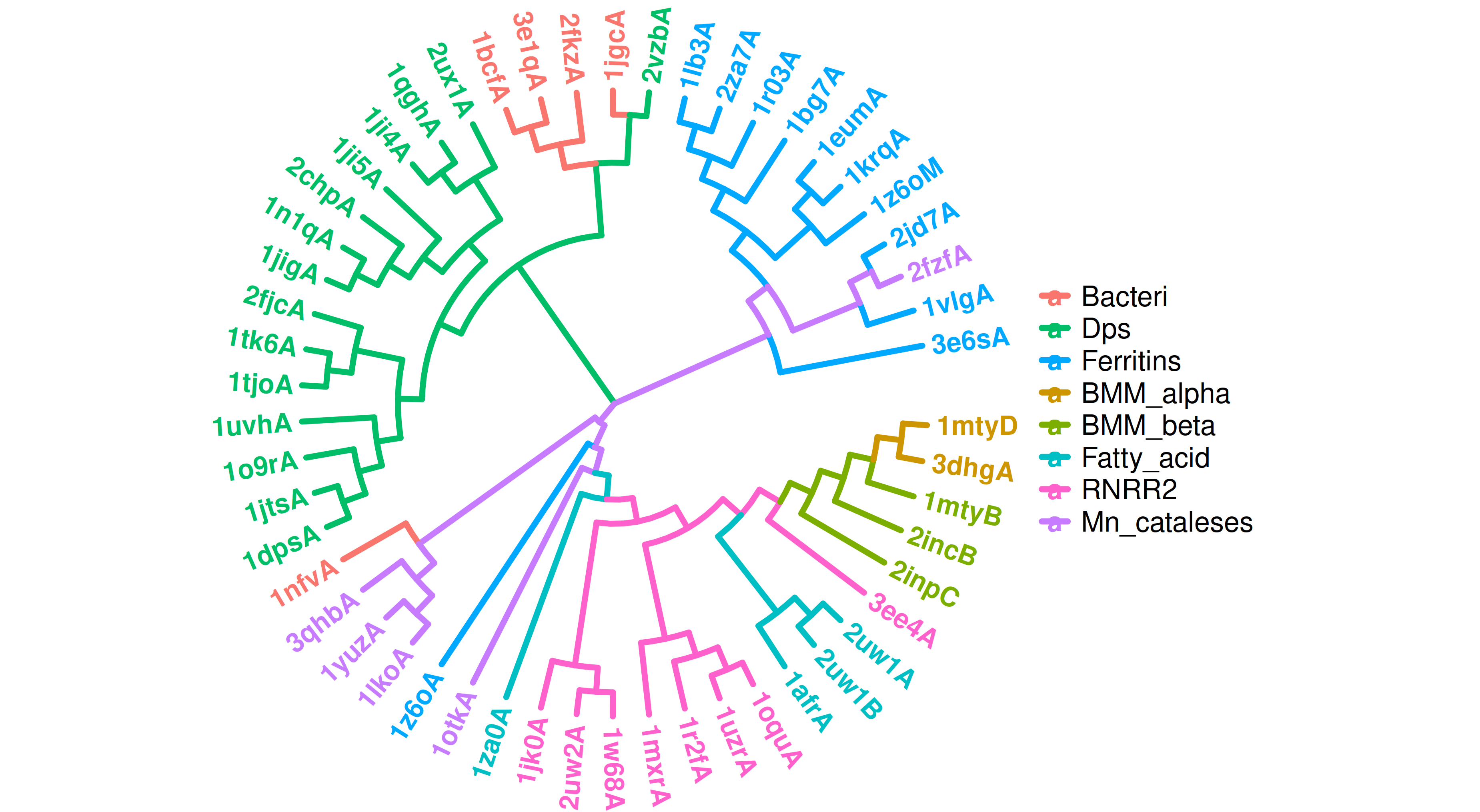
Additionally, although SCOP and CATH have a single overarching RNR-like family, these proteins are classified into three distinct families by Pfam, Phenol\_Hydrox (PF02332), Ribonuc\_red\_sm (PF00268), and Fatty acid desaturase (PF03405). We find consistently high support for this more detailed sequence-based classification, as well as the further separation of the BMMs into BMMa and BMMb.

There are several proteins that lie outside of the major groupings in our networks, all of which are classified by CATH as Ferritins, and most of which are also classified by SCOP as ferritins. For example, in our network 1otkA is closer to the RNRs rather than the ferritins. Pfam classifies this protein into PaaA\_PaaC, with 1otkA the only member of PaaA\_PaaC. Another protein is 3ee4A which Pfam classifies as Ribonuc\_red\_sm, possibly because of its sequence similarity to RNR R2 proteins (30). In our network this structure clearly occupies an outgroup position relative to the RNR R2 structures. This is functionally consistent with its ligand-binding pocket, which indicates that it is a substrate oxidizing enzyme, and its lack of competence as an RNR R2 (30, 31).

In the Fad group, there is a distinct cluster of plant Fads (2uw1A-B, 1afrA), whereas the Mycobacterium tuberculosis protein (1za0A) appears more distantly related. As discussed by Lundin et al. [21] unfortunately, this is the only solved structure of a bacterial Fad. It is also one of a paralogous pair and not the one considered functional. The structure of the functional Fad has not yet been possible to solve (28). With such a skewed data set, it is difficult to judge how well our structure-based network identifies evolutionary relationships within the Fad group.

Using sequence-based phylogenies (22, 27), BMMs have been suggested to have evolved by duplication and divergence, leading to distinct di-iron binding catalytic (α) and non-metal binding (β) subunits. Although substrate specificity is generally considered low in BMMs (27), we can identify one distinct subgroup both in the α and β clans containing only proteins annotated as soluble methane monooxygenases (1mhyD B, 1mtyD B, and 1xvbA C for α and β components, respectively), and another subgroup consisting of toluene monooxygenases (2incA B and 3dhgA B). This indicates that our structural analysis is able to recover relatively recent evolutionary relationships in addition to the more distant examples described above. Indeed, within both clans, we see that the phenol hydroxylase subunits (2inpC A) also appear distinct within this part of the network.

Our qualitative analysis provides compelling support for the separation of two SCOP families within this superfamily, namely ferritin (a.25.1.1) and RNR-like (a.25.1.2), as well as their subdivision into distinct groups within our phylogenetic tree. These groupings closely align with prior studies conducted by Lundin et al. [21] and Malik et al.[22]. Figure 8 visually illustrates that the overarching classification delineated in SCOP, CATH, and Pfam (http://pfam.xfam.org/) is faithfully recapitulated through our energy-based phylogenetic analysis.

****

*Figure 9: Energy-based phylogenetic of the ferritin-like superfamily. The two large SCOP families, ferritins (a.25.1.1; Bacteri, Ferritins, Dps and NA) and ribonucleotide reductase-like (a.25.1.2; BMM\_alpha, BMM\_beta, Fatty\_acid and RNRR2) separated in the tree with smaller groups.*

