

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[Li2011]. 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. **Li2011**

so this is what I really want (**Li2011**)

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

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.