INTRODUCTION

Proteins form the very basis of life. More than half of the dry weight of human cells is made up of proteins. They regulate a variety of activities in all known organisms, from replication of the genetic code to transporting oxygen, and are generally responsible for regulating the cellular machinery and consequently, the phenotype of an organism. Proteins accomplish their task by three-dimensional tertiary and quaternary interactions between various substrates such as ligands, DNA and RNA, and other proteins and form a molecular complex. The recognition process is mediated by a distinct region on the surface of the protein, which is referred to as the binding site, and which forms the trigger to set the protein into action (Bergner and Günther, 2004). More than 30% of all proteins in the cells exploit one or more metals to perform their specific functions (Gray, 2003), over 40% of all enzymes contain metals (Andreini et al., 2008). The amino acids that regularly act as metal ligands in proteins are thiolates of cysteines, imidazoles of histidines, carboxylates of glutamic and aspartic acids, and phenolates of tyrosines. Numerous cellular processes require metal ions as cofactors for enzymatic reactions or as structural components of proteins (McCall et al., 2000; Rees, D.C. 2002). Proteins capable of binding one or more metal ions or metal containing cofactors are metalloproteins and play an important role for energy metabolism, molecular metabolism, and signal transduction in biological systems. The elucidation of the structure and function of these metalloproteins is central to understanding the regulatory mechanisms associated with biological functioning.

The importance of metal ions in biological system is increasingly drawing attention as indicated with the recent emergence of terms such as metallome and metallomics. There are 13 metals which are essential for plants and animals (Bertini et al., 1994; Lippard and Berg, 1994). Four of these, sodium, potassium, magnesium and calcium, are present in large quantities and are known as bulk metals (Fenton, 1995). The remaining nine, which are present in small quantities, are vanadium, chromium, molybdenum, manganese, iron, cobalt, nickel, copper and zinc, and are known as the trace metals. Most of the trace metals are found
as natural constituents of proteins. Nature has taken advantage of the special properties of the metal ions and tuned them by protein encapsulation to perform a wide variety of specific functions associated with life processes (Bertini et al., 1994; Lippard and Berg, 1994).

Iron and copper are redox-active metals (i.e. can switch between oxidized and reduced forms: \( \text{Cu}^{2+}/\text{Cu}^{1+} \) and \( \text{Fe}^{3+}/\text{Fe}^{2+} \)) and often participate in electron transfer (Fenton, 1995) in respiration and photosynthesis processes, small redox-active metalloproteins facilitate electron-transfer reactions by alternately binding to specific integral membrane proteins that often contain several metal sites. Iron and copper are also involved in dioxygen (\( \text{O}_2 \)) storage and carriage via metalloproteins (e.g. hemoglobin, myoglobin and hemocyanin). Furthermore, iron storage as well as iron and copper transfer are facilitated in the cells by specific proteins. In contrast to iron and copper, zinc serves as a super acid center in several metalloenzymes, promoting hydrolysis or cleavage of a variety of chemical bonds. Representative proteins that use catalytic zinc ions are carboxypeptidases, carbonic anhydrase, and alcohol dehydrogenase. In addition, zinc ions often play structural roles in proteins (e.g. Superoxide dismutase and zinc-finger motifs). Most of the other trace metals have been identified as parts of metalloenzymes (Lippard and Berg, 1994; Gray, 2003). For example, manganese is found as a cofactor in mitochondrial SOD, inorganic phosphatase and photosynthesis system II. Nickel functions in enzymes such as urease and several hydrogenases. Both molybdenum and vanadium are found in nitrogenases, where they are present in larger clusters containing also iron and sulfur ions (Fenton, 1995).

Role of metal ions in drug targeted proteins such as copper, zinc and calcium are identified and evaluated in this work. The Alzheimer’s disease amyloid precursor protein (APP) is a major regulator of neuronal copper homeostasis (Barnham et al., 2003; Malakooti et al., 2014). APP has a copper binding domain located in the N-terminal cysteine-rich region that can strongly coordinate Cu (II) and reduce it to Cu (I). APP is a member of a multigene family, and the CuBD sequence is similar among the different APP family paralogs and orthologs, suggesting an overall conservation in its function or activity. It has been demonstrated that this domain can modulate copper homeostasis and production of A\( \beta \), a peptide that plays a central role in the progression of Alzheimer’s disease. The structure
of this domain and residues (His-147, His-151, Tyr-168, Met-170) involved in coordinating copper and the possible mechanism for copper reduction. The nature and orientation of these residues constitute a novel copper binding site. His-147 and His-151 were necessary for copper binding (Multhaup et al., 1996; Spoerri et al., 2012). The orientation of these residues in the three-dimensional structure indicates that, with very small side-chain movements, a tetrahedral metal binding site suitable for coordinating Cu(I) is formed (Fig. 1.1). Such a site is reminiscent of the blue copper proteins that bind copper with a tetrahedral arrangement of ligands consisting of two histidines, a methionine, and a cysteine residue (Adman, 1991). The coordination of Cu (II) to the tetrahedrally arranged His-147, His-151, Tyr-168, and Met-170 (Fig. 1.1) can explain the redox chemistry associated with Cu binding to APP. In general, four coordinate Cu (II) ions favor a square planar coordination sphere about the metal, whereas Cu (I) generally prefers a tetrahedral arrangement (Casella and Gullotti, 1993).

![Fig. 1.1: The metal binding site. (a) View of the putative metal binding site consisting of residues His-147, His-151, Tyr-168, and Met-170. (b) A model of Cu (I) coordinated in a tetrahedral configuration to His-147, His-151, Tyr-168, and Met-170.](image)

Blue-copper proteins (also called cupredoxins) are an important class of metalloproteins that carry out electron-transfer reactions (Lippard and Berg, 1994). The name comes from their intense blue absorption around 630 nm, which is a result of unique copper coordination. The best characterized blue-copper protein with respect to the role of metals in folding is *Pseudomonas aeruginosa* azurin. This is a small (128 residues) protein that is
believed to facilitate electron transfer in denitrification/respiration chains (Adman, 1991). In recent years, it was proposed that the physiological function of azurin in *P. aeruginosa* involves electron transfer directly related to the cellular response to oxidative stress (Vijgenboom et al. 1997). Notably, in 2002 *P. aeruginosa* azurin was demonstrated to interact with the tumor-suppressor gene product p53 and act as an anti-cancer agent in cell-culture studies (Yamada et al. 2002). Thus, not only is *P. aeruginosa* azurin an excellent model, it is also a putative cancer-drug candidate.

Azurin has one α-helix and eight β-strands that fold into a β-barrel structure arranged in a double-wound Greek-key topology (Adman, 1991; Nar et al. 1992b) (Fig. 2). In *P. aeruginosa* azurin, the redox-active copper (Cu$^{1+}$/Cu$^{2+}$) is coordinated by two histidine imidazoles (His46 and His117), one cysteine thiolate (Cys112), and two weaker axial ligands, sulfur of methionine (Met121) and carbonyl of glycine (Gly45) in a trigonal bipyramidal geometry. It has been suggested that the polypeptide fold defines the exact geometry of the metal site, leading to the unusual Cu$^{2+}$ coordination in azurin as well as in other blue-copper proteins (Wittung-Stafshede et al. 1998a). In vitro, *P. aeruginosa* azurin can bind many different metals in the active site (e.g. zinc). Crystal structures of apo- and holo-azurin (apo=without cofactor; holo=with cofactor) have shown that the overall three-dimensional structure is identical with and without a metal (copper or zinc) cofactor (Nar et al. 1991, 1992a, b). Since the zinc form of azurin is always a by-product upon azurin over-expression in *Escherichia coli*, and both copper and zinc may be present in the cells where the azurin polypeptide is produced.
Matrix metalloproteinases (MMPs, matrixins) are a family of secreted and membrane-bound zinc-dependent endopeptidases that have the combined capacity to degrade all the components of the extracellular matrix. The catalytic domain contains the zinc-binding motif XEXXHXXGXXH, where three histidine residues coordinate a zinc ion. Additionally, the catalytic domain also contains a conserved methionine residue, forming a “Met-turn,” which contributes to protect the catalytic zinc (Ganea et al. 2007; Nagase et al. 2006). MMP enzymes are strongly involved in a kaleidoscope of normal, pathological, physiological, and biological processes such as embryogenesis, normal tissue remodeling, wound healing, and angiogenesis, and in diseases such as atheroma, arthritis, cancer, and tissue ulceration (Sekhon, 2010).

Owing to the advances in crystallography, time-resolved spectroscopic methods and protein engineering since the 1970s, we now have high-resolution structures of many metalloproteins and know a great deal about their mechanisms of action (Gray, 2003). Metal-binding proteins have been identified by such experimental approaches as absorbance spectroscopy, gel electrophoresis, metal-affinity columns and shift assay, chromatography, mass spectroscopy, NMR, and combined spectroscopic studies. However, some of these methods generally require a purified or semi-purified target of interest, do not facilitate identification of unknown targets form complex protein mixtures, or require multi-step
processes and very specialized equipment, which limit their application ranges. Therefore, there is a need to explore other methods including computational approaches for facilitating the identification of metal-binding proteins to complement these experimental methods. Computational protein structure analysis is one of the cornerstones of bioinformatics. It seems strange how little attention and bioinformatics community has paid to metalloproteins and other complex proteins.

Protein farnesyltransferase (FTase) is a zinc metalloenzyme that catalyzes the addition of a farnesyl isoprenoid to a conserved cysteine in peptide or protein substrates (Huang et al., 1997). Zinc ion plays an important catalytic role in FTase, most likely by activation of the cysteine thiol of the protein substrate for nucleophilic attack on the isoprenoid (Huang et al, 1997). It has been reported that mammalian FTase is a heterodimer of α and β subunits and the β subunit contains a zinc at the CAAX motif binding site, coordinated by Asp297β, Cys299β, His362β and a water molecule (Park et al., 1997; Dunten et al., 1998; Long et al., 1998; Long et al., 2002).

Type A influenza virus is becoming a worldwide pandemic threat due to its virulence and transmissibility in people. The viral surface protein neuraminidase, along with hemagglutinin, classifies influenza subtypes and fulfills an important role in viral propagation by cleaving a terminal sialic acid from host cell surfaces; thus, neuraminidase (N1) is a key drug target for controlling flu infection (Lawrenz et al., 2010). The highly pathogenic influenza strain H1N1 is currently treated with inhibitors of the viral surface protein neuraminidase (N1). Structures of N1 indicate a conserved, high affinity calcium binding site located near the active site has been identified through crystallography (Russell et al., 2006; Varghese et al., 1983; Xu et al., 2008; Collins et al., 2008) and Proton Induced X-ray Emission (PIXE) experiments (Taylor et al., 1993) though the calcium density was not resolved or discussed for some holo structures of group 1 neuraminidase (Russel et al., 2006). The specific role of this calcium ion is unknown, but experiments have supported its importance for wildtype enzyme activity (Chong et al., 1991) and enzyme thermostability (Burmeister et al., 1994). Commonly performed fluorometric neuraminidase activity assays include calcium salts in the activity buffer and follow the protocol developed originally in Potier, et al.( Potier et al., 1979) where
increased calcium ion concentration was found to augment activity. Additionally, a recent crystal structure of calcium-deficient group 2 neuraminidase shows destabilization of the key active site residue R292 near the calcium binding site (Smith et al., 2006).

Approximately one third of all of structurally-determined proteins are metal-bound, and large percentages of metals present in human body are bound to proteins. This emphasizes the crucial role of metal ions in stabilizing protein structure (Jernigan et al., 1994). Most of metalloproteins play a key role in biological process, and therefore have been considered to be a promising target for drug discovery. The metal ion in the active site can participate in the enzymatic reaction as a receptor for the lone pair electrons, which has an effect of weakening the chemical bonds in the substrate. Due to such an essential role in enzymatic reaction, a chemical group that can bind to the central metal ion is necessary in the molecular structure of a small-molecule inhibitor that can regulate the activity of metalloproteins. This has made it much more difficult to design an inhibitor of a metalloprotein than that of a protein without a metal cofactor. Despite such a difficulty, a number of chemical groups that chelate and inactivate a metal ion at the active site of metalloproteins have been reported including carboxylate, phosphate, hydroxamate, -keto acid, and diol moieties. Although some databases and web servers for metalloproteins have already been reported, all the biochemical information has been limited to the identification of metal binding residues (Hsin et al., 2008), coordination geometry around the metal ion (Castagnetto et al., 2002; Andreini et al., 2009a), and catalytic mechanism (Andreini et al., 2009b).

In this work, we focus on conservation analysis of residues coordinating with some of the metals commonly found in the Protein Data Bank (PDB), namely: Ca, Cu, and Zn. Proteins coordinating with some of these metals have recently been analyzed based on the composition and geometry of the metal-binding site. All metal-binding proteins used in this study are collected from a comprehensive search of Protein Data Bank. A total of 33295 metal-binding protein sequences were obtained. Most of these proteins can be classified into one of the 10 metal-binding classes, in calcium-binding, cobalt-binding, copper-binding, iron-binding, magnesium-binding, manganese-binding, nickel-binding, potassium-binding, sodium-binding...
and zinc-binding class respectively. Some proteins were found to belong to more than one class. Therefore, we constructed MetalloDB to provide information of metal ions insight into a metalloprotein. Despite the critical role of metalloproteins in fundamental processes in vivo, little is known about the biological mechanisms of metalloprotein folding and assembly (Bartnikas and Gitlin, 2001). Here we have reported role of metal in catalytic site of drug target proteins such as Matrix metalloproteinase, Farnesyltransferase, Neuraminidase and Thioredoxin and conclude that (i) The presence of a metal often stabilizes the native protein. (ii) Coordination of a metal prior to polypeptide folding can dramatically accelerate formation of functional metalloproteins. (iii) On the other hand, some polypeptides must form well-defined metal sites before metal binding can occur.

We emphasize that conformational changes of metallopeptides play an important role in a range of human diseases. For example, Cu²⁺ ion has been shown to induce aggregation of amyloid-forming peptides (Villanueva et al. 2004). Further understanding of metallopeptide and metalloprotein folding may also aid in curing of diseases related to metal metabolism, such as Menkes syndrome and Wilson’s disease, which both involve erroneous cofactorprotein interactions.
OBJECTIVES

- Identification of Metal binding protein complexes from Protein Data Bank (PDB) and other protein web resources.
- Classification and characterization of metaloproteins on the basis of their binding pockets.
- Identification and classification of disease target on the basis of motifs present at possible binding regions.
- *In silico* virtual screening of metal based legends and targets on the basis of results obtained from the objective 2nd and 3rd.