RESEARCH ARTICLE
Bull Biotechnol (2023) 4(1):17-29
https://doi.org/10.51539/biotech.1262979 Journal homepage: www.dergipark.org.tr/biotech

Bulletin of Biotechnology

Functional and structural identification of iron-binding proteins on tomato (*Solanum lycopersicum* L.) proteome via *in silico* approaches

Yiğit Küçükçobanoğlu[1*], Lale Yıldız Aktaş[1]

[*1]*Department of Biology, Faculty of Science, Ege University, İzmir, Turkey*

*Corresponding author : y.kucukcobanoglu@gmail.com
Orcid No: https://orcid.org/0000-0002-9856-5506

Received : 10/03/2023
Accepted : 09/06/2023

Abstract: Iron-plant interactions have crucial roles in crop production growth and development. In this study, we have analyzed the whole proteome of tomato (*Solanum lycopersicum* L.) plants for iron-binding proteins. A total of 213 iron-binding protein candidates were identified in the study. Out of these 213 proteins, 45 were selected for modeling and validated with a high confidence level by using different computational analyses. Results showed that Glu, Cys, Asp, and His amino acid residues were indicators of iron-binding proteins. Besides, mechanistic insights of iron-binding proteins were analyzed by molecular dynamics simulations. Simulation results proved the conformational stabilization of proteins. Validated proteins were further analyzed for subcellular localization, clustered for molecular functions and biological processes. According to the results, iron-binding proteins were mostly located in the chloroplast. Also, these proteins are involved in different molecular and biological roles ranging from oxidation-reduction processes and electron transport chain to protein repair mechanisms. This report provides structural and functional properties of iron-binding proteins for tomato proteome. The study may assist in future research on plant physiology, protein engineering, or bioengineering.

Keywords: iron-binding proteins; tomato; structural bioinformatics; functional analysis; three-dimensional modeling

1 Introduction

Metal ions have a vital role in biological reactions, including photosynthesis, respiration, and water oxidation in living organisms (Lu et al. 2013). The importance of metal ions in these processes comes forward due to their ability to bind proteins. (Garcia et al. 2006). Among metals, iron (Fe) (an essential micronutrient for plants) has roles in several processes like photosynthesis, respiration, or DNA synthesis on plant growth and development (Briat et al. 1995). Different oxidation states of Fe exist in nature, like reduced ferrous (Fe^{+2}) or oxidized ferric (Fe^{+3}) (Pehkonen, 1995). Iron-binding proteins are essential cofactors for the plant cells (Andreini et al. 2017). Iron-binding proteins are a component of chloroplast proteins like Rieske proteins, which have a crucial role in photosynthesis (Allen, 2004; Briat et al. 2007). Furthermore, iron-binding proteins take charge of DNA metabolism or protein translation metabolic pathways (Braymer and Lill 2017).

Metalloproteins (proteins that contain metal ions) are important protein classes with their catalytic, regulatory, or structural roles via metal atoms (Shi and Chance, 2008). In the last decades, several analysis techniques have been used to investigate metalloproteins based on mass spectrometry like inductively coupled plasma-mass spectrometry (ICP-MS) (Pröfrock and Prange 2012), nuclear analytical techniques like high-resolution spatial speciation analysis by synchrotron radiation X-ray fluorescence (SR-XRF) (Pushie et al. 2014), or affinity techniques like immobilized metal affinity chromatography (IMAC) (Chang et al. 2017). However, these analytical techniques have limitations due to time-consuming processes for sample preparation and data acquisition. Additionally, these techniques are expensive for instrumentation and chemical supply. Given all these disadvantages, using new approaches such as computational biology has become inevitable to focus on these significant proteins.

Nowadays, developments based on bioinformatics make it feasible to analyze big data of the whole genome or proteome. Iron-binding proteins can be identified, modeled, or functionally analyzed using bioinformatics tools for plant proteomes. Iron-binding sites of proteins can be analyzed by using protein sequences or structures with online tools like MetalPredator (Valasatava et al. 2016), MetSite (Sodhi et al. 2004), MetalloPred (Naik et al. 2011), or IonCom (Hu et al. 2016). Besides, there are various bioinformatic tools available like MetalS^2 (Andreini et al. 2013), MetalS^3 (Valasatava et al.

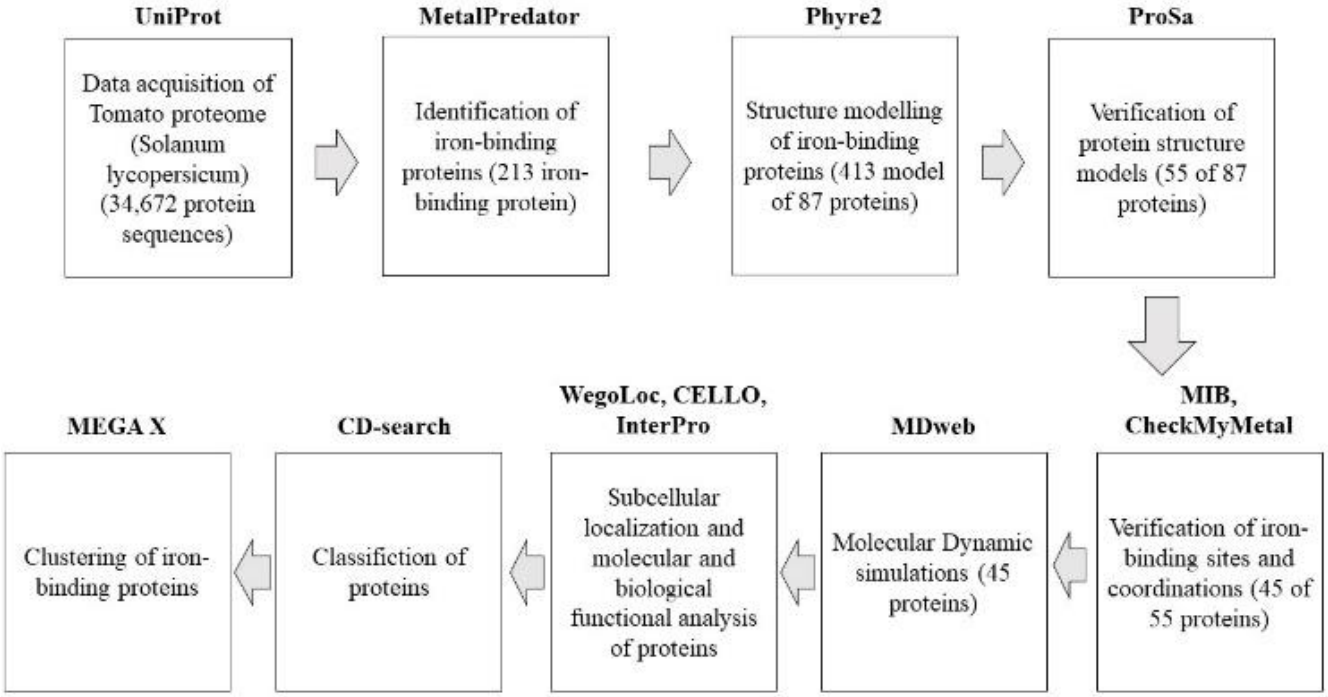

Fig. 1 Schematic demonstration of the analysis workflow for identifying and classifying iron-binding proteins from the tomato proteome

2014), MIonSite (Qiao and Xie 2019), MetalDetector V2.0 (Passerini et al. 2011), CheckMyMetal (CMM) (Zheng et al. 2017). The interaction of binding sites of metalloproteins with transition metals like iron, zinc, copper, manganese, and cadmium makes them possible to identify by using sequence or structural information of proteins via bioinformatic tools. In addition, the MetalPredator tool is a specified tool for finding iron-binding proteins using Hidden Markov Model profiles (HMM) and structural motif bindings of iron (Valasatava et al. 2016). *Xanthomonas translucens* pv. *Undulosa* proteome was analyzed for iron-binding proteins, and ~9.8% of the proteome had iron-binding protein motifs using *in silico* approaches. These proteins had different functions like transport or carbohydrate metabolism (Sharma et al. 2017). In another research, iron-binding proteins of wheat proteome were analyzed, and results showed that iron-binding proteins were involved in 21% of biological processes and 39% of molecular functions (Verma et al. 2017).

Tomato (*Solanum lycopersicum* L.) is included in the Solanaceae family, which contains most of the economic crops like potato or pepper. With over 10000 cultivars, tomato is cultivated in a 5 million ha area, and its production was estimated at 180 million tons worldwide in 2019 (FAO, 2020). Due to the importance of tomato crop production, understanding the interactions between iron metal ions and tomato proteome can benefit crop growth and development.

The literature review showed no research on iron-binding proteins on the whole tomato (*S. lycopersicum*) proteome. This study aimed at identifying, classifying, and 3D modeling the iron-binding proteins on tomato proteome using *in silico* analyses. A summarized scheme of the workflow is illustrated in Fig. 1. The results highlight the understanding of the functions of iron-binding proteins in tomato, which could be beneficial for new challenges for tomato crop growth and development. Also, these results can offer different approaches to iron-related nutrient management for crop production.

2 Materials and Method

2.1 Dataset acquisition of tomato (*S. lycopersicum*) proteome

Tomato (*S. lycopersicum*) proteome has 34672 proteins based on UniProt database as reference proteome (Proteome ID: UP000004994). The protein sequence dataset of the whole tomato proteome was downloaded in FASTA format from the UniProt database (https://www.uniprot.org/proteomes/UP000004994).

2.2 Identification of iron-binding proteins

The whole proteome sequence dataset of tomato was analyzed using an online MetalPredator program to identify iron-binding proteins (Valasatava et al. 2016). MetalPredator identifies proteins using Pfam domains and iron-binding motifs called Minimal Functional sites (MFSs) with e-value lower than 10^{-3}.

2.3 Modeling structures of iron-binding proteins

Three-dimensional (3D) structures of identified iron-binding proteins were modeled using online Protein Homology/Analogy Recognition Engine V2.0 (Phyre2) software. Phyre2 is a sequence-based modeling program for protein 3D structures that uses advanced remote homology (Kelley et al. 2015). Modeled 3D structures were selected according to the criteria: confidence ≥ 90%, query ≥ 50%, and identification ≥ 30%. Filtered structures quality was verified by the ProSA-web program using structure templates (Wiederstein and Sippl, 2007). ProSA-web program verifies models by z-scores depending on C^{α} potentials of protein structure coordinates.

2.4 Determination of iron-binding sites on verified protein models

Verified 3D models of iron-binding proteins were analyzed using Metal Ion-Binding site prediction and docking server (MIB) to identify Fe^{2+} and Fe^{3+} binding structure models (Lin et al. 2016). MIB analyses were based on the fragment transformation method using structure models, and the tool aligns the structure to metal ion binding residues. The results were analyzed depending on the alignment scores of iron-binding motifs.

CheckMyMetal (CMM) program was used for both validation and coordination geometries of iron-binding sites on protein structures (Zheng et al. 2017). The CMM program can evaluate structures for geometrical and other irregularities in iron-binding sites with different parameters such as bond valences or metal-binding sites. Three-dimensional protein structures were visualized using Mol* 3D viewer (Sehnal et al. 2021).

2.5 Molecular dynamics simulations of iron-binding proteins

Iron-binding structure models of iron-binding proteins were analyzed via MDweb server for molecular dynamics (MD) simulations (Hospital et al. 2012). MDweb server can perform MD simulations using different algorithms like AMBER, GROMACS, or NAMD. In our study, AMBER Full MD setup -ff99SB* (Hornak & Simmerling, including Best &Hummer psi modification)- toolkit was used for setup, solvation, and equilibration. In this toolkit, sodium ions have been used for neutralization, and octahedron box of TIP3P water molecules with a spacing distance of 15 Å was used for solvation. The energy minimization of the structure was completed in 500 steps of the conjugate gradient with a force constant of 50Kcal/mol. The equilibration step was started with heating solvent to 300k, then continued with reducing the force constant respectively 5.0 Kcal/mol, 2.5 Kcal/mol, 1 Kcal/mol, and finally completed by simulation without restrains. Simple box solvent molecular dynamics simulations were performed via a constant Number of particles, Pressure, and Temperature (NPT) with 2.0 fs time steps, 2.5 ps total time, 300k temperature, and 50 output frequency steps. For the results, the root mean square deviation (RMSD) was determined for protein backbone residues and ligands.

2.6 Subcellular localization, molecular functions, and biological process analysis of iron-binding proteins

Cellular and subcellular localization of iron-binding proteins were analyzed by using weighted gene ontology term-based subcellular localization prediction (WegoLog) (Chi and Nam, 2012) and subcellular localization predictor (CELLO V2.5) web tools (Yu et al. 2006). Both web tools are used for analyzing sequence similarity and gene ontology information to predict the subcellular localization of proteins based on a support vector machine (SVM) classifier. But tools have been using different datasets to analyze localization, such as BaCello dataset for WegoLoc and Park and Kneisha dataset for Cello.

Functional analysis of identified and verified iron-binding proteins was performed using online InterPro software, which classifies proteins into families and predicts domains and important sites (Blum et al. 2021). The program uses predictive models known as signatures and classifies proteins for subcellular localization, biological processes, and molecular functions.

2.7 Classification and clustering of iron-binding proteins

The identified 55 iron-binding proteins were classified for functional domains using the Conserved Domain Database (CDD) tool on the National Centre for Biological Institute database (Marchler-Bauer et al. 2010). CDD search tool classification analysis was performed for the criteria: CDD–58235 position-specific scoring matrix (PSSM) database with an expectation value (e-value) of 0.01.

Iron-binding protein sequences were clustered using the maximum parsimony method with 2000 bootstrap replications of the Subtree-Pruning-Regrafting (SPR) algorithm in the MEGA X program (Kumar et al. 2018). The phylogenetic tree was visualized using the EvolView v3 web tool (Subramanian et al. 2019).

3 Results

3.1 Identification and structure modeling of iron-binding proteins on tomato proteome

The tomato proteome dataset, which includes 34672 protein sequences, was retrieved from the UniProt database. The proteome dataset was analyzed to determine iron-binding proteins via MFSs using the MetalPredator program. Approximately 0.6% of whole tomato proteome, corresponding to 213 proteins, were identified as iron-binding proteins depending on Pfam patterns on minimal functional sites of proteins. Potential 213 iron-binding proteins have been 3D modeled using protein sequences on Phyre2 software. The 3D protein structure modeling results showed 413 candidate models for 87 protein sequences. The candidate protein models were filtered according to the criteria: confidence ≥ 90%, query ≥ 50%, and identification ≥ 30%. Filtered models were selected for structure quality analysis. Finally, filtered structure model quality was verified by the ProSA-web program depending on the z-scores of structure models (Supplementary Table S1). Results have shown that the ProSA-web program verified 55 protein structures.

3.2 Determination of iron-binding sites on verified protein models

The selected 55 iron-binding protein models were analyzed using the MIB program to identify both Fe^{2+} and Fe^{3+} binding motifs on 3D protein structures. Iron had two convertible oxidation states, Fe^{2+} and Fe^{3+}, involved in different biochemical pathways. MIB program analyses were based on protein structures and used to predict iron-binding residues. The results showed that 50 of 55 identified proteins were modeled and could bind both Fe^{2+} and Fe^{3+} (Supplementary Table S2). Glutamic acid (Glu), Cysteine (Cys), Aspartic acid (Asp), and Histidine (His) amino acids residues were involved in the iron bindings of proteins, and the distribution of iron-binding amino acid residues was shown in Fig. 2. According to the results, Fe^{2+} binding residues were determined as 37 Cys, 21 Glu, 19 Asp, 19 His, 5 Asn, and 4 Ala, while Fe^{3+} binding residues were 33 Cys, 32 Glu, 16 Asp, 11 His, 4 Gln, and 2 Ala amino acids. The binding residues of each iron-binding protein were addressed in Supplementary Table S2

Validation and coordination geometry analyses of iron-binding protein structures were performed by using CheckMyMetal (CMM) web tool. Results showed that 45 of 50 iron-binding proteins were verified at a high confidence level, 42 Fe^{2+} binding sites and 41 Fe^{3+} binding sites were determined with various geometries on iron-binding proteins. Coordination geometries of many iron-binding protein structures were classified as freely bound. These structures did not show any coordination geometry with iron ions because the poor coordination caused deviation from the ideal geometry. Besides, different coordination geometries like square planar, trigonal planar, trigonal bipyramidal, or tetrahedral by CMM analysis were observed. Fe^{2+} binding protein geometries indicated 20 poorly coordinated, 12 free coordination, 8 square planar coordination, and 2 tetrahedral coordination. Also, Fe^{3+} binding protein geometries revealed 17 free coordination, 16 poorly coordinated, 3 square planar coordination, 2 tetrahedral and trigonal bipyramidal coordination, and 1 trigonal planar coordination (Supplementary Table S3). Selected 3D structure images of iron-binding proteins were illustrated in Fig. 3.

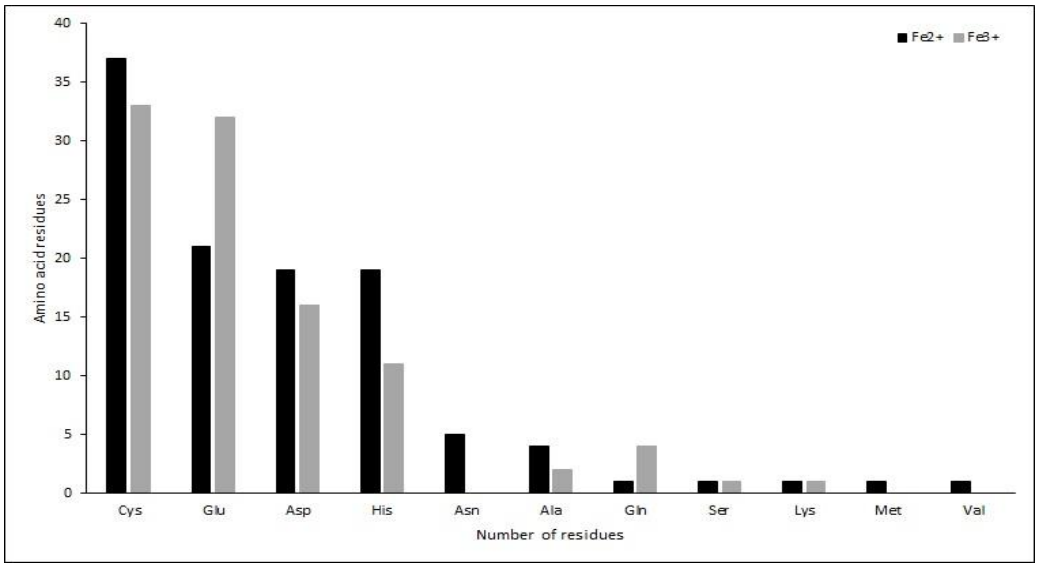

Fig. 2 Distribution of iron-binding amino acid residues

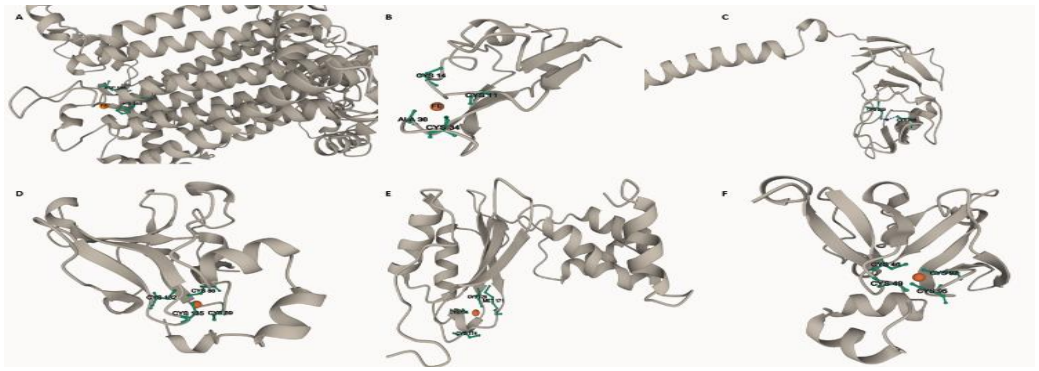

Fig. 3 Visualization of iron-binding proteins in three dimensional structures and coordination geometries; A)sp|Q2MIA0|PSAA_SOLLC protein Fe^{2+} square planar coordination geometry, B)sp|Q2MI49|PSAC_SOLLC protein Fe^{3+} square planar coordination geometry, C)tr|A0A3Q7JLM2|A0A3Q7JLM2_SOLLC protein Fe^{3+} trigonal planar coordination geometry, D)tr|A0A3Q7ERR3|A0A3Q7ERR3_SOLLC protein Fe^{3+} trigonal bipyramidal coordination geometry, E)tr|A0A3Q7I5P3|A0A3Q7I5P3_SOLLC protein Fe^{2+} tetrahedral coordination geometry, F)tr|A0A3Q7EI02|A0A3Q7EI02_SOLLC protein Fe^{3+} tetrahedral coordination geometry

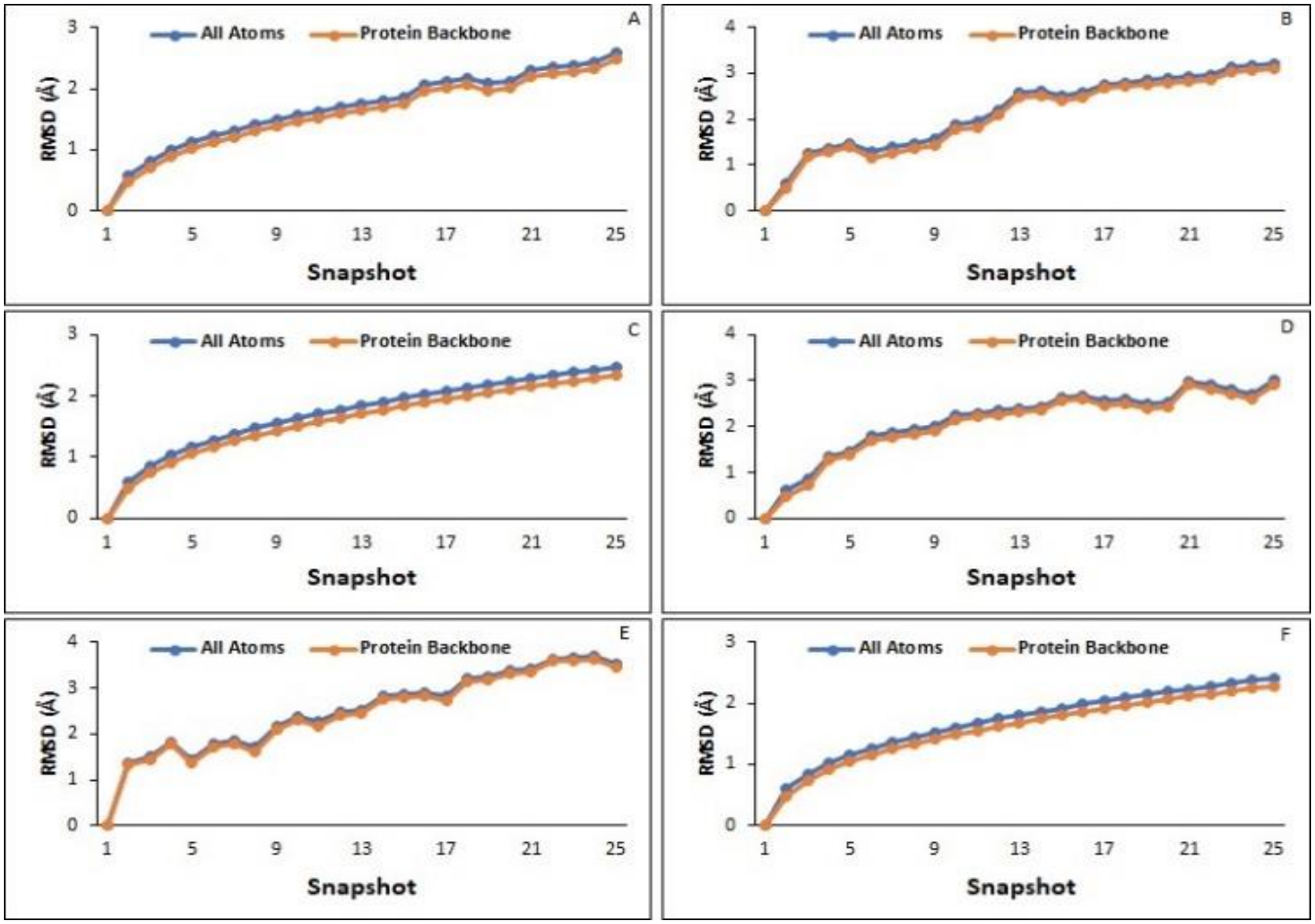

Fig. 4 RMSD results of iron-binding proteins A)sp|Q2MIA0|PSAA_SOLLC Fe^{2+}, B)sp|Q2MI49|PSAC_SOLLC Fe^{3+}, C)tr|A0A3Q7JLM2|A0A3Q7JLM2_SOLLC Fe^{3+}, D)tr|A0A3Q7ERR3|A0A3Q7ERR3_SOLLC Fe^{3+}, E)tr|A0A3Q7I5P3|A0A3Q7I5P3_SOLLC Fe^{2+}, F)tr|A0A3Q7EI02|A0A3Q7EI02_SOLLC Fe^{3+}

3.3 Molecular dynamics simulations of iron-binding proteins

After the iron-binding site verification at 45 iron-binding proteins, docked structures for both Fe^{2+} and Fe^{3+} via MIB have been selected for MD simulations. MD simulations were performed to investigate the mechanistic insight of iron-binding proteins. Structural changes of proteins were determined using RMSD analysis for 25 snapshots. Results showed that the time evolution of average RMSD values on all proteins varied between 1,626 Å - 2,994 Å for Fe^{2+} and 1,626 Å - 2,140 Å for Fe^{3+} (Supplementary Table S4). The time-course changes of RMSD distribution plots for selected proteins compared to protein backbones were shown in Fig. 4. According to the MD simulation results, iron-protein binding structures achieved stable conformations in the simulations.

3.4 Subcellular localization, molecular functions, and biological process analysis of iron-binding proteins

The remaining 45 proteins with high confidence levels were analyzed to determine subcellular localization, molecular function, and biological process. Subcellular localization analyses were completed using WegoLoc and CELLO v2.5 web tools. Cellular components, molecular functions, and biological processes were analyzed using the InterPro web tool. Subcellular localization of iron-binding proteins analyzed via WegoLoc showed that proteins were localized in the chloroplast (33%), cytoplasm (31%), mitochondrion (29%), and nucleus (7%) (Fig. 5a). CELLO v2.5 analysis results showed that subcellular localizations of the iron-binding proteins were chloroplastic (33%), nuclear (24%), mitochondrial (20%), cytoplasmic (9%) extracellular (7%), and plasma membranal (7%) (Fig. 5b). Total 16 protein was found as consistent localizations for both WegoLoc and CELLO v2.5 programs (Supplementary Table S5).

Molecular and functional analyses of iron-binding proteins were performed using InterPro web tool. Results showed that analyzed iron-binding proteins were involved in 51 biological processes, 128 molecular functions, and 15 cellular components. iron-binding proteins were checked for their implication in biological processes showing that a large part was related to oxidation-reduction processes and electron transport chain (14%), and tricarboxylic acid cycle (8%). Other functions (6%) were determined as protein repair, oxidative stress, rRNA processing, rRNA methylation, tRNA methylation, and photosynthesis. Also, a small part of the proteins belonged to another 11 processes with different rates like DNA repair (4%) or tRNA modification (2%) (Fig. 6).

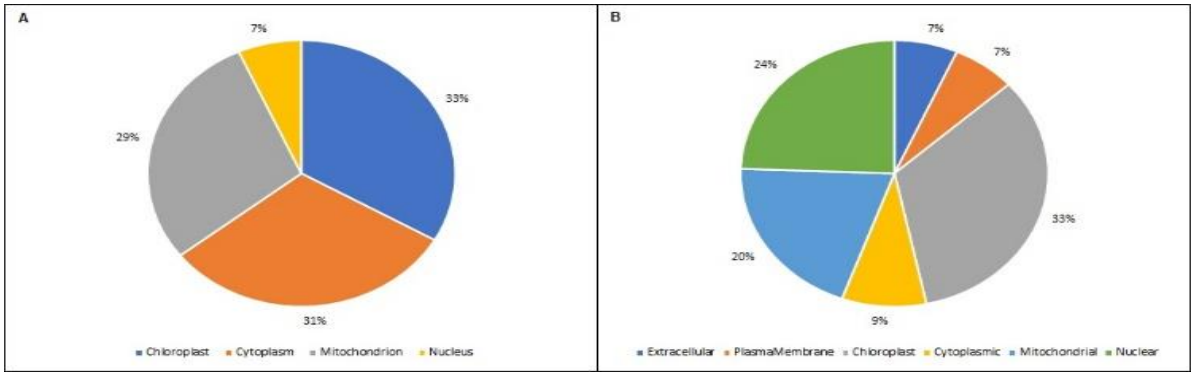

Fig. 5 Subcellular localization of iron-binding proteins; A) WegoLoc analysis, B) CELLO v2.5 analysis

Molecular functions of iron-binding proteins were mostly classified as 20% iron-sulfur cluster binding, 13% electron transfer activity, and 11% for 2 iron-2 sulfur (2Fe-2S) cluster binding. The other molecular functions were 9% catalytic activity, 7% oxidoreductase activity, and 6% iron 4 sulfur cluster binding (Fig. 7).

In addition, the analyzed iron-binding proteins were joined into cellular components. Most of the proteins were part of membranes (20%). Other proteins were determined as components at photosystems I, thylakoid, integral components of membranes, and nucleus (13%). A small part of the proteins were components of the thylakoid membrane, chloroplast, mitochondrion, and molybdopterin synthase complex (7%) (Fig. 8).

3.5 Classification and clustering of iron-binding proteins

Selected high confidence level 45 iron-binding proteins were classified for functional domains using the CDD search tool. These functionally classified proteins were clustered using the maximum parsimony method with 2000 bootstrap on the MEGA X program. iron-binding proteins were classified into 28 superfamily categories, which were assembled into 15 categories depending on their biological or molecular functions. These categories were oxidation-reduction, protein repair, posttranslational modifications, gene regulation, metal-ion binding, DNA binding, rRNA processing, ATP metabolism, translation, tRNA modification, DNA repair, electron transfer, photosynthesis mechanisms, and reductase and transferase activities (Fig. 9).

The roles of iron-binding proteins were demonstrated through ferredoxin in electron transfer, succinate dehydrogenase in oxidation-reduction processes, Rieske in photosynthesis, and thioredoxin in gene regulation (Supplementary Table 1). All the identified iron-binding proteins were associated with important roles in different metabolic pathways of tomato plants.

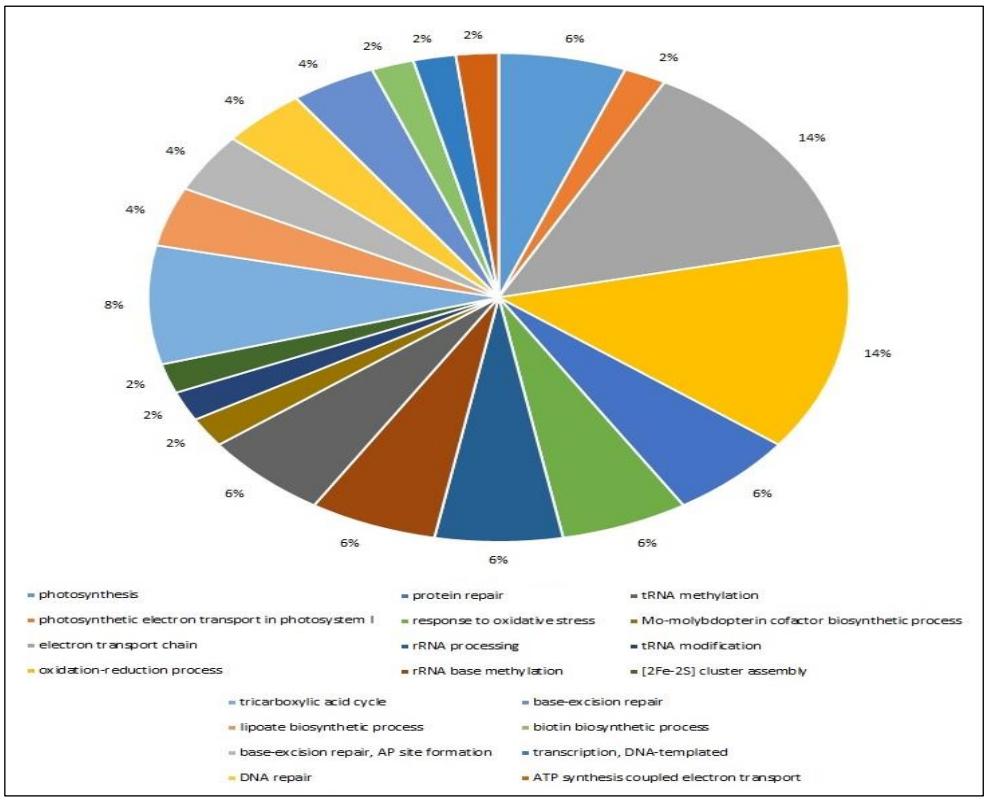

Fig. 6 Biological processes involving iron-binding proteins

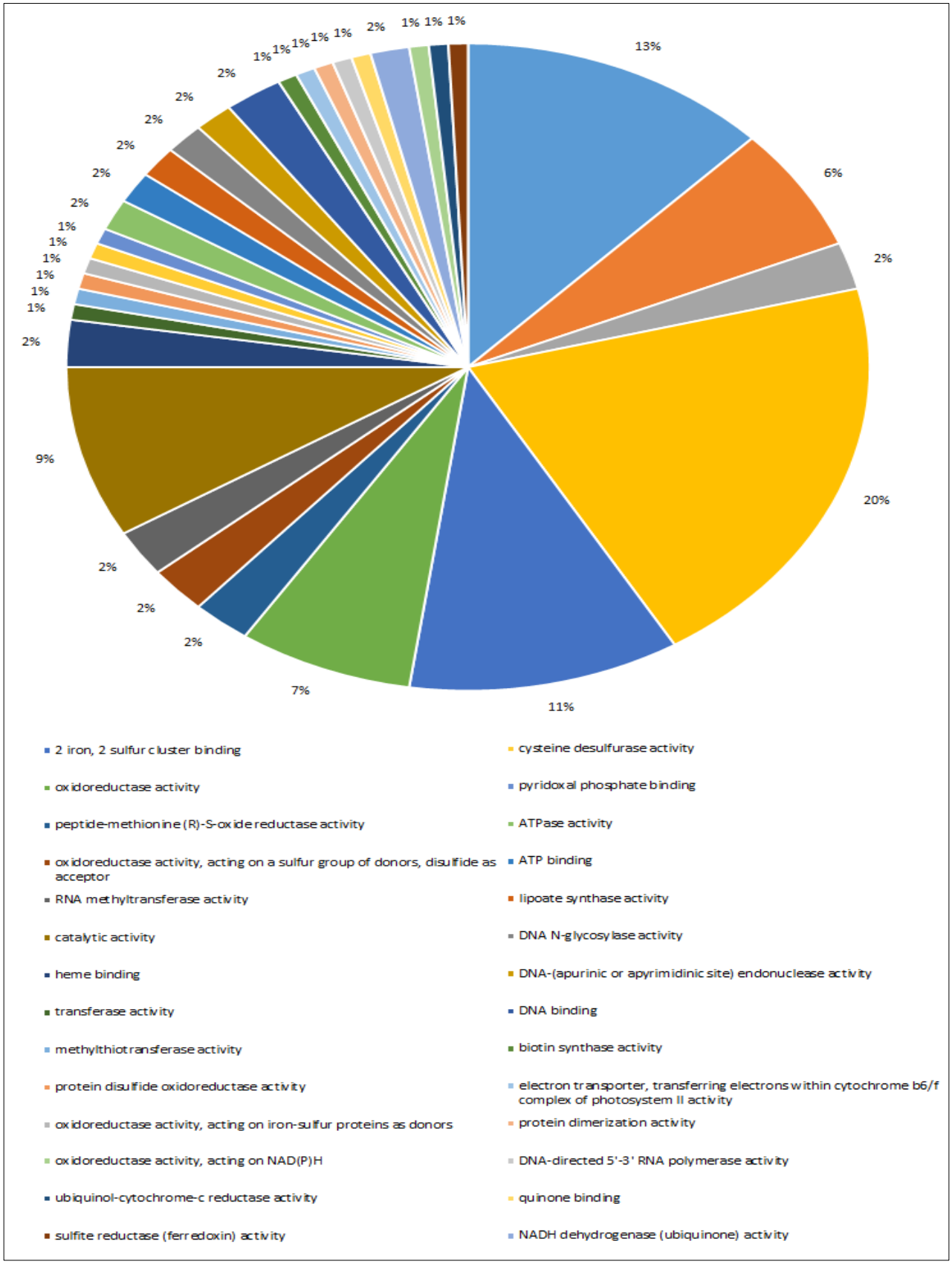

Fig. 7 Molecular functions involving iron-binding proteins

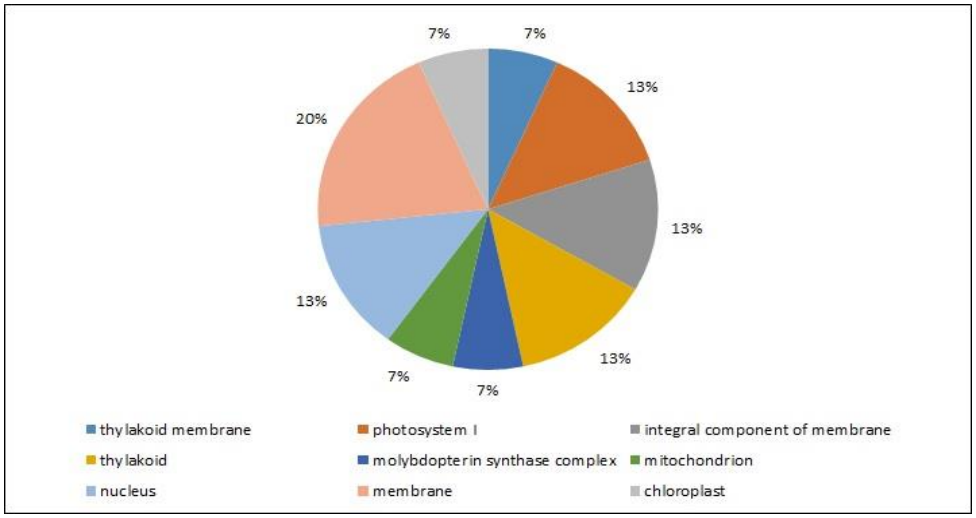

Fig. 8 Cellular components involving iron-binding proteins

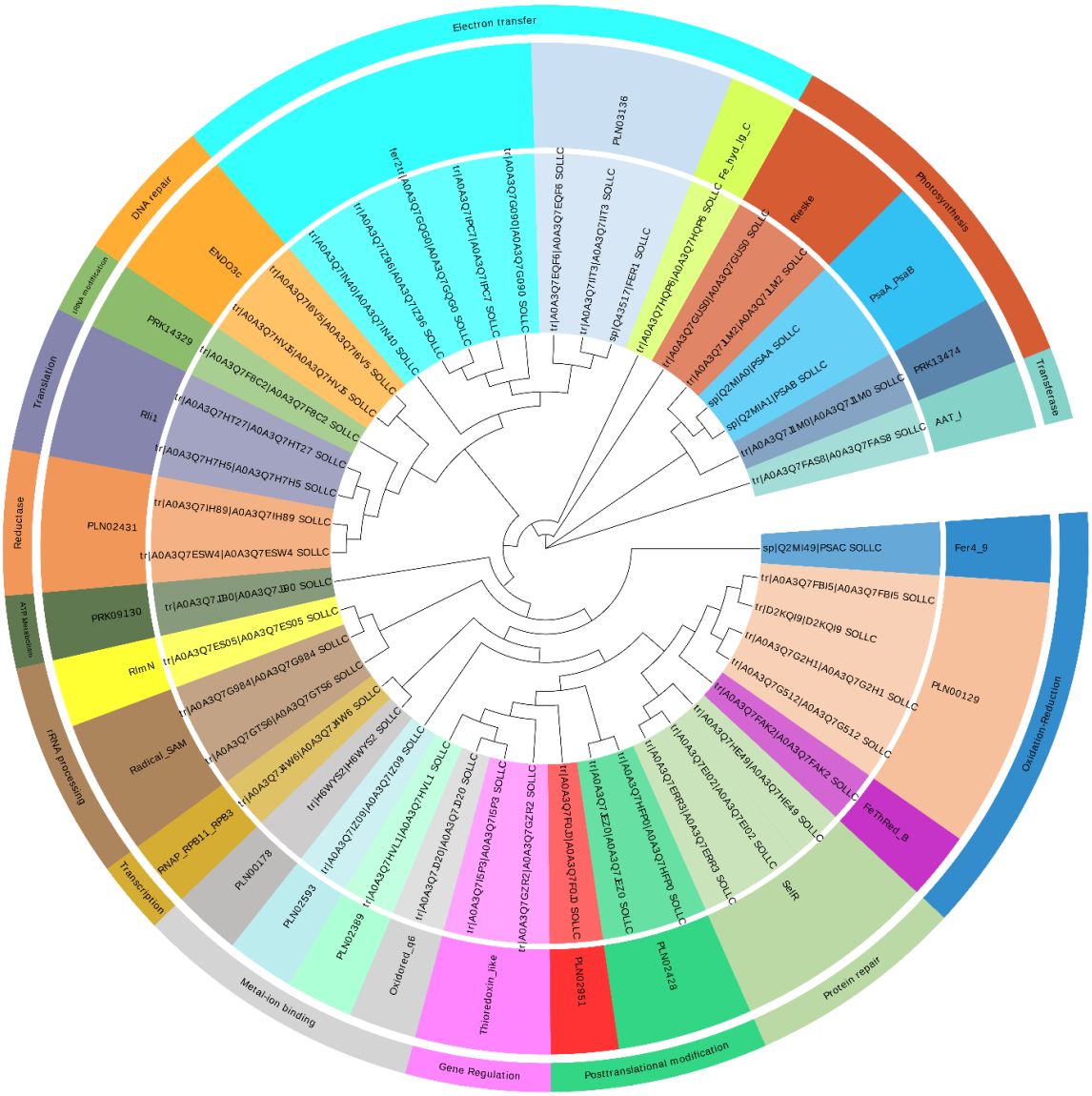

Fig. 9 Functional classification of clustered iron-binding proteins

4 Discussion

Tomato is an economic crop with many cultivars worldwide. Iron is one of the essential micronutrients, and it has a crucial role in plant growth and development. In this study, iron-binding proteins of tomato proteome have been identified, classified, and analyzed for biological functions. In the first step of the study, the MetalPredator program was used to find iron-binding proteins on tomato proteome sequences. MetalPredator is a customized program for iron-binding proteins and uses the HMM model profiles for Pfam domains and structural motifs defined as MFSs via MetalPDB (Andreini et al. 2012; Valasatava et al. 2016; Putignano et al. 2018). MetalPredator has been used for various bacterial proteomes like *Paenibacillus polymyxa* for iron-binding biosynthesis (Li et al. 2021) or *Spironucleus salmonicida* for stress response genes (Stairs et al. 2019). Also, wheat proteome has been analyzed for iron-binding proteins using the MetalloPred tool, and approximately 1.1% of wheat proteome has been determined as iron-binding proteins (Verma et al. 2017). Our results showed that 0.6% of tomato proteome contained iron-binding proteins. There is no report in the literature addressing iron-binding proteins in tomato plants, especially the determination of iron-binding proteins using MetalPredator. This study demonstrates, for the first time, a complete analysis of all iron-binding proteins of the tomato proteome.

Metal ions bind to proteins as ligands by donating an electron pair. The side chains of cysteine (Johnson et al. 2005), histidine (Ciofi-Baffoni et al. 2018), glutamic acid (Chaud et al. 2002), and aspartic acid (Caetano-Silva et al. 2015) residues interact with both Fe^{2+} and Fe^{3+} ions. Iron-binding residues interaction sites are analyzed by setting an iron-ion binding template of at least 3.5 Å between the metal ion and two residues (Lin et al. 2016). In this study, MIB analysis showed Fe^{2+} and Fe^{3+} ion interactions on 50 iron-binding proteins. Both Fe^{2+} and Fe^{3+} ion binding were related to the canonical metal-binding sites. These sites are located at the interface of two domains connected by a single long α-helix (Vigouroux et al. 2020). Analyzed proteins had an interaction with mostly cysteine (Cys), aspartic acid (Asp), histidine (His), and glutamic acid (Glu) residues. Besides, a fewer number of the analyzed proteins interacted with glutamine (Gln), tyrosine (Try), asparagine (Asn), alanine (Ala), methionine (Met), valine (Val), and lysine (Lys) residues. Respectively, free γ- and δ- carboxyl groups of aspartic and glutamic acid residues could bind iron ions (Storcksdieck et al. 2007). Iron ions could bind with histidine residues due to the imidazole side chain (Nemirovskiy and Gross, 1996). Cysteine residues could bind iron ions due to sulfur groups (Giles et al. 2003). Additionally, 89% of coordinating atoms bind to side chains of Asp, His, Glu, and Cys residues (called canonical amino acids) on iron-ion binding sites (Sánchez-Aparicio et al. 2021). Another study reported that the major iron-ion binding protein residues of sea cucumber as His, Glu, Cys, and Asp amino acids (Sun et al. 2017). According to the literature, our results are coherent with previous reports for metal-ion binding sites on proteins.

Iron has two interconvertible oxidation states: Fe^{2+} and Fe^{3+}. Consequently, Fe ions can be implicated in vital oxidation-reduction reactions in biological systems. Those reactions are a part of significant processes like respiration, photosynthesis, or DNA synthesis. Depending on the oxidation states of iron, it can be in different coordination environments (Sánchez et al. 2017). In our study, most iron-binding proteins showed free or poor coordination. Iron ions were primarily coordinated by O or N atoms in the first coordination sphere, but if lone pair of electrons occupies more space than a bonding pair, deviation from ideal geometry may occur (Gillespie, 1992). Reports showing that various protein complexes without the ideal designated geometries are consistent with our results (McLaughlin et al. 2012). Our coordination geometries analysis for Fe^{2+} binding proteins displayed 18% as square planar and 4% as tetrahedral geometries. On the other hand, Fe^{3+} binding proteins showed 6% square planar and 4% for trigonal bipyramidal and tetrahedral coordination geometries. In both cases, square planar geometry was supported by negatively charged amino acids like Glu or His to prevent higher coordination numbers (Pascualini et al. 2015). The square planar geometry can also arise from the arrangement of four binding pyridine units on the bis-porphyrin heme protein model (Chen et al. 1999). Square planar geometry corresponds to a high-field and low-spin electronic arrangement like in pheophytin, an electron carrier of chlorophyll (Bechaieb et al. 2018). Tetrahedral coordination geometry on iron-binding complexes has been proceeded by inorganic sulfides or protein-based ligands like cysteine residues (Pandelia et al. 2015). Also, tetrahedral coordination in iron-binding proteins is regulated by coordinating four cysteine sulfur atoms like in Rubredoxin (Todorovic and Teixeira, 2018). Trigonal bipyramidal coordination geometry can be promoted by coordinating N and three O atoms like on YtgA protein (Luo et al. 2019). Various coordination geometries resulted from differences in electronegativity between iron and amino acid residues.

MD simulations were performed to understand the mechanistic insight of iron-binding proteins via structural conformations. Iron ions were docked into targeted residues iron-binding proteins using the MIB tool. MD simulations showed that protein structures have stable conformations according to the average RMSD values (between 1,626 Å - 2,994 Å for Fe^{2+} and 1,626 Å - 2,140 Å for Fe^{3+}). Evaluating the accuracy of the dynamic binding process is complicated due to multiple iron coordinating geometries (Hu and Shelver, 2003). In this study, Q2MI49|PSAC_SOLLC protein showed square planar geometry for Fe^{3+} via Cys11, Cys14, Cys34, and Ala36. The RMSD values for these residues were respectively changing (0.06 Å, 0.04 Å, 0.03 Å, 0.07 Å) when compared with the protein backbone. Furthermore, A0A3Q7EI02_SOLLC protein showed tetrahedral coordination geometry via Cys46, Cys49, Cys92, and Cys95, with RMSD values of 0,01 Å, 0,22 Å, 0,03 Å, 0,01 Å when compared with the protein backbone (Data not shown). Although lower RMSD values have proved conformational stabilization of analyzed iron-binding proteins, differently charged areas generated by the side chains of amino acid residues could be the reason for the issue (Bernacchioni et al. 2016).

Both subcellular localization analysis results in WegoLoc and Cello v2.5 showed that iron-binding proteins were mostly localized in the chloroplast. Total 16 protein were detected in

similar localization. These are 9 proteins for chloroplast, 4 proteins for mitochondrion, 2 proteins for nucleus and 1 protein for cytoplasm (Supplementary Table S5). However, there were differences between these analyses in other localizations because WegoLoc and Cello use different datasets. WegoLoc tool is based on the BaCello dataset for plants with 491 proteins (Pierleoni et al. 2006). Meanwhile, the Cello v2.5 tool is based on Park and Kneisha dataset with 7589 eukaryotic proteins (Yu et al. 2006). Regarding datasets, results from the Cello v2.5 tool could be considered more accurate than WegoLoc for plants.

This study classified 45 high confidence level iron-binding proteins from different protein families according to subcellular localization, molecular functions, and biological processes. These proteins have essential roles in electron transfer, catalytic activity, metal ion binding, protein repair, and oxidation reduction processes. Iron-binding proteins plays important roles as both carrier proteins and metalloproteins in plant metabolism including from photosynthesis to protein repair. Carrier iron-binding proteins take charge for moving ions and molecules across the membranes. Also, metalloproteins are contains metal ions as cofactors. In this study, determined iron-binding proteins commonly act as carrier proteins for electron transfer and oxidation reduction processes. On the other hand, other iron-binding proteins act as metalloproteins for metal-ion binding or Fe-S clusters. The biological and molecular functions of these high confidence level iron-binding proteins were presented (Fig. 9, Supplementary Table 1), and the descriptions were derived from InterPro (Blum et al. 2021) and NCBI-CDD (Marchler-Bauer et al. 2010) databases.

Functional domains of the identified proteins were analyzed by detecting homologs using the CDD search tool and protein super families clustered via the sequences using the maximum parsimony method. This method is a character-based phylogenetic tree construction method (Kannan and Wheeler, 2012) and offers better performance than the maximum likelihood method due to the high heterogeneity of sequence datasets (Kolaczkowski and Thornton, 2004).

5 Conclusion

In this study, using computational methods, iron-binding proteins of tomato proteome were identified via structural formations at high confidence level. A total of 42 iron-binding proteins were identified via 3D structures, metal ion binding geometries, and structural conformations (MD analysis) on tomato proteome. Different tools were used to analyze the identified iron-binding proteins subcellular localization, molecular function, and biological process roles. Two different subcellular localization tools proved that iron-binding proteins are primarily localized in the chloroplast. These proteins are commonly involved in oxidation-reduction biological processes. Molecular functions of iron-binding come forward on iron-sulfur cluster binding and electron transfer activity. Finally, iron-binding proteins were classified as superfamilies and clustered for homologous proteins. Clustered proteins revealed the importance of iron-binding proteins on electron transfer and oxidation-reduction processes. Results showed that iron-binding proteins play important roles in the growth and development of tomato plants through biochemical and physiological functions. These results provide a base for iron-binding proteome analyses using computational plant sciences methods. Considering these proteomic analyses will be more critical in the future for research areas like protein engineering or bioengineering.

Acknowledgements

This study is funded by Ege University Scientific Research Projects (FDK-2019-21028), Council of Higher Education Scholarship 100/2000, and TUBITAK 2211c scholarship.

Authors' contributions:

Yiğit Küçükçobanoğlu and Lale Yıldız Aktaş contributed to the study conception, design, data collection, writing - orginal draft preparation, writing-, review and editing.

Conflict of interest disclosure:

The authors declare that they have no conflict of interest.

References

Allen JF (2004) Cytochrome b6f: structure for signalling and vectorial metabolism. Trends Plant Sci 9:130–137.

Andreini C, Cavallaro G, Lorenzini S, Rosato A (2012) MetalPDB: a database of metal sites in biological macromolecular structures. Nucleic Acids Res 41:D312–D319.

Andreini C, Cavallaro G, Rosato A, Valasatava Y (2013) MetalS2: a tool for the structural alignment of minimal functional sites in metal-binding proteins and nucleic acids. J Chem Inf Model 53:3064–3075.

Andreini C, Rosato A, Banci L (2017) The relationship between environmental dioxygen and iron-sulfur proteins explored at the genome level. PLoS One 12:e0171279.

Bechaieb R, Lakhdar ZB, Gérard H (2018) DFT and TD-DFT studies of Mg-substitution in chlorophyll by Cr (II), Fe (II) and Ni (II). Chemistry Africa 1:79–86.

Bernacchioni C, Pozzi C, di Pisa F, Mangani S, Turano P (2016) Ferroxidase Activity in Eukaryotic Ferritin is Controlled by Accessory-Iron-Binding Sites in the Catalytic Cavity. Chemistry–A European Journal 22:16213–16219.

Bertsova Y v, Kostyrko VA, Baykov AA, Bogachev A v (2014) Localization-controlled specificity of FAD: threonine flavin transferases in Klebsiella pneumoniae and its implications for the mechanism of Na+-translocating NADH: quinone oxidoreductase. Biochimica et Biophysica Acta (BBA)-Bioenergetics 1837:1122–1129.

Blum M, Chang H-Y, Chuguransky S, Grego T, Kandasaamy S, Mitchell A, Nuka G, Paysan-Lafosse T, Qureshi M, Raj S (2021) The InterPro protein families and domains database: 20 years on. Nucleic Acids Res 49:D344–D354.

Borukhov S, Nudler E (2008) RNA polymerase: the vehicle of transcription. Trends Microbiol 16:126–134.

Braymer JJ, Lill R (2017) Iron–sulfur cluster biogenesis and trafficking in mitochondria. Journal of Biological Chemistry 292:12754–12763.

Briat J-F, Curie C, Gaymard F (2007) Iron utilization and metabolism in plants. Curr Opin Plant Biol 10:276–282.

Briat J-F, Fobis-Loisy I, Grignon N, Lobréaux S, Pascal N, Savino G, Thoiron S, von Wirén N, van Wuytswinkel O (1995) Cellular and molecular aspects of iron metabolism in plants. Biol Cell 84:69–81.

Caetano-Silva ME, Bertoldo-Pacheco MT, Paes-Leme AF, Netto FM (2015) Iron-binding peptides from whey protein hydrolysates: Evaluation, isolation and sequencing by LC–MS/MS. Food Research International 71:132–139.

Chang Y-Y, Li H, Sun H (2017) Immobilized metal affinity chromatography (IMAC) for metalloproteomics and phosphoproteomics In: Inorganic and Organometallic Transition Metal Complexes with Biological Molecules and Living Cells , pp 329–353. Elsevier.

Chaud M v, Izumi C, Nahaal Z, Shuhama T, Bianchi M de LP, Freitas O de (2002) Iron derivatives from casein hydrolysates as a potential source in the treatment of iron deficiency. J Agric Food Chem 50:871–877.

Chen LX, Lee PL, Gosztola D, Svec WA, Montano PA, Wasielewski MR (1999) Time-resolved X-ray absorption determination of structural changes following photoinduced electron transfer within Bis-porphyrin Heme protein models. J Phys Chem B 103:3270–3274.

Chi S-M, Nam D (2012) WegoLoc: accurate prediction of protein subcellular localization using weighted Gene Ontology terms. Bioinformatics 28:1028–1030.

Ciofi-Baffoni S, Nasta V, Banci L (2018) Protein networks in the maturation of human iron–sulfur proteins. Metallomics 10:49–72.

DeFraia CT, Wang Y, Yao J, Mou Z (2013) Elongator subunit 3 positively regulates plant immunity through its histone acetyltransferase and radical S-adenosylmethionine domains. BMC Plant Biol 13:1–13.

Dong J, Lai R, Nielsen K, Fekete CA, Qiu H, Hinnebusch AG (2004) The Essential ATP-binding Cassette Protein RLI1 Functions in Translation by Promoting Preinitiation Complex Assembly*♦. Journal of Biological Chemistry 279:42157–42168.

Dutta S, Teresinski HJ, Smith MD (2014) A split-ubiquitin yeast two-hybrid screen to examine the substrate specificity of atToc159 and atToc132, two Arabidopsis chloroplast preprotein import receptors. PLoS One 9:e95026.

FAO F (2020) FAOSTAT statistical database. Rome: Food and Agriculture Organisation of the United Nations.

Freedman RB, Hirst TR, Tuite MF (1994) Protein disulphide isomerase: building bridges in protein folding. Trends Biochem Sci 19:331–336.

Garcia JS, de Magalhães CS, Arruda MAZ (2006) Trends in metal-binding and metalloprotein analysis. Talanta 69:1–15.

Giles NM, Watts AB, Giles GI, Fry FH, Littlechild JA, Jacob C (2003) Metal and redox modulation of cysteine protein function. Chem Biol 10:677–693.

Gillespie RJ (1992) The VSEPR model revisited. Chem Soc Rev 21:59–69.

Gueguen V, Macherel D, Jaquinod M, Douce R, Bourguignon J (2000) Fatty acid and lipoic acid biosynthesis in higher plant mitochondria. Journal of Biological Chemistry 275:5016–5025.

Hase T, Schürmann P, Knaff DB (2006) The interaction of ferredoxin with ferredoxin-dependent enzymes In: Photosystem I , pp 477–498. Springer.

Hospital A, Andrio P, Fenollosa C, Cicin-Sain D, Orozco M, Gelpí JL (2012) MDWeb and MDMoby: an integrated web-based platform for molecular dynamics simulations. Bioinformatics 28:1278–1279.

Houston NL, Fan C, Xiang Q-Y, Schulze J-M, Jung R, Boston RS (2005) Phylogenetic analyses identify 10 classes of the protein disulfide isomerase family in plants, including single-domain protein disulfide isomerase-related proteins. Plant Physiol 137:762–778.

Hu X, Dong Q, Yang J, Zhang Y (2016) Recognizing metal and acid radical ion-binding sites by integrating ab initio modeling with template-based transferals. Bioinformatics 32:3260–3269.

Hu X, Shelver WH (2003) Docking studies of matrix metalloproteinase inhibitors: zinc parameter optimization to improve the binding free energy prediction. J Mol Graph Model 22:115–126.

Jahns P, Graf M, Munekage Y, Shikanai T (2002) Single point mutation in the Rieske iron–sulfur subunit of cytochrome b6/f leads to an altered pH dependence of plastoquinol oxidation in Arabidopsis. FEBS Lett 519:99–102.

Jin S, Hu Y, Fu H, Jiang S, Xiong Y, Qiao H, Zhang W, Gong Y, Wu Y (2021) Identification and characterization of the succinate dehydrogenase complex iron sulfur subunit B gene in the Oriental River Prawn, Macrobrachium nipponense. Front Genet 12.

Johnson DC, Dean DR, Smith AD, Johnson MK (2005) Structure, function, and formation of biological iron-sulfur clusters. Annu Rev Biochem 74:247.

Kannan L, Wheeler WC (2012) Maximum parsimony on phylogenetic networks. Algorithms for Molecular Biology 7:1–10.

Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE (2015) The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc 10:845–858.

Kolaczkowski B, Thornton JW (2004) Performance of maximum parsimony and likelihood phylogenetics when evolution is heterogeneous. Nature 431:980–984.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 35:1547.

Kurisu G, Kusunoki M, Katoh E, Yamazaki T, Teshima K, Onda Y, Kimata-Ariga Y, Hase T (2001) Structure of the electron transfer complex between ferredoxin and ferredoxin-NADP+ reductase. Nat Struct Biol 8:117–121.

Li Q, Li Y, Li X, Chen S (2021) Identification of Genes Involved in Fe–S Cluster Biosynthesis of Nitrogenase in Paenibacillus polymyxa WLY78. Int J Mol Sci 22:3771.

Lin Y-F, Cheng C-W, Shih C-S, Hwang J-K, Yu C-S, Lu C-H (2016) MIB: metal ion-binding site prediction and docking server. J Chem Inf Model 56:2287–2291.

Litomska A, Ishida K, Dunbar KL, Boettger M, Coyne S, Hertweck C (2018) Enzymatic thioamide formation in a bacterial antimetabolite pathway. Angewandte Chemie International Edition 57:11574–11578.

Lu Y, Chakraborty S, Miner KD, Wilson TD, Mukherjee A, Yu Y, Liu J, Marshall NM (2013) Metalloprotein design In: Comprehensive Inorganic Chemistry II (Second Edition): From Elements to Applications , pp 565–593. Elsevier Ltd.

Luo Z, Neville SL, Campbell R, Morey JR, Menon S, Thomas M, Eijkelkamp BA, Ween MP, Huston WM, Kobe B (2019) Structure and metal binding properties of Chlamydia trachomatis YtgA. J Bacteriol 202:e00580–19.

Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR (2010) CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Res 39:D225–D229.

McLaughlin MP, Retegan M, Bill E, Payne TM, Shafaat HS, Pena S, Sudhamsu J, Ensign AA, Crane BR, Neese F (2012) Azurin as a protein scaffold for a low-coordinate nonheme iron site with a small-molecule binding pocket. J Am Chem Soc 134:19746–19757.

Naik PK, Ranjan P, Kesari P, Jain S (2011) MetalloPred: A tool for hierarchical prediction of metal ion binding proteins using cluster of neural networks and sequence derived features. Journal of Biophysical Chemistry 02:112–123.

Nakayama M, Akashi T, Hase T (2000) Plant sulfite reductase: molecular structure, catalytic function and interaction with ferredoxin. J Inorg Biochem 82:27–32.

Nemirovskiy O v, Gross ML (1996) Complexes of iron (II) with cysteine-containing peptides in the gas phase. J Am Soc Mass Spectrom 7:977–980.

Pandelia M-E, Lanz ND, Booker SJ, Krebs C (2015) Mössbauer spectroscopy of Fe/S proteins. Biochimica et Biophysica Acta (BBA)-Molecular Cell Research 1853:1395–1405.

Pascualini ME, di Russo N v, Thuijs AE, Ozarowski A, Stoian SA, Abboud KA, Christou G, Veige AS (2015) A high-spin square-planar Fe (II) complex stabilized by a trianionic pincer-type ligand and conclusive evidence for retention of geometry and spin state in solution. Chem Sci 6:608–612.

Passerini A, Lippi M, Frasconi P (2011) MetalDetector v2. 0: predicting the geometry of metal binding sites from protein sequence. Nucleic Acids Res 39:W288–W292.

Pehkonen S (1995) Determination of the oxidation states of iron in natural waters. A review. Analyst 120:2655–2663.

Picciocchi A, Douce R, Alban C (2003) The Plant Biotin Synthase Reaction: IDENTIFICATION AND CHARACTERIZATION OF ESSENTIAL MITOCHONDRIAL ACCESSORY PROTEIN COMPONENTS *. Journal of Biological Chemistry 278:24966–24975.

Pierleoni A, Martelli PL, Fariselli P, Casadio R (2006) BaCelLo: a balanced subcellular localization predictor. Bioinformatics 22:e408–e416.

Pierrel F, Hernandez HL, Johnson MK, Fontecave M, Atta M (2003) MiaB Protein from Thermotoga maritima: CHARACTERIZATION OF AN EXTREMELY THERMOPHILIC tRNA-METHYLTHIOTRANSFERASE *. Journal of Biological Chemistry 278:29515–29524.

Pröfrock D, Prange A (2012) Inductively Coupled Plasma-Mass Spectrometry (ICP-MS) for Quantitative Analysis in Environmental and Life Sciences: A Review of Challenges, Solutions, and Trends. Appl Spectrosc 66:843–868.

Pushie MJ, Pickering IJ, Korbas M, Hackett MJ, George GN (2014) Elemental and Chemically Specific X-ray Fluorescence Imaging of Biological Systems. Chem Rev 114:8499–8541.

Putignano V, Rosato A, Banci L, Andreini C (2018) MetalPDB in 2018: a database of metal sites in biological macromolecular structures. Nucleic Acids Res 46:D459–D464.

Qiao L, Xie D (2019) MIonSite: Ligand-specific prediction of metal ion-binding sites via enhanced AdaBoost algorithm with protein sequence information. Anal Biochem 566:75–88.

Ramadan AM, Alnufaei AA, Khan TK, Ali HM, Eissa HF, Hassan SM (2021) The first report of RNA U to C or G editing in the mitochondrial NADH dehydrogenase subunit 5 (Nad5) transcript of wild barley. Mol Biol Rep 48:6057–6064.

Ravet K, Touraine B, Boucherez J, Briat J-F, Gaymard F, Cellier F (2009) Ferritins control interaction between iron homeostasis and oxidative stress in Arabidopsis. The Plant Journal 57:400–412.

Roldán-Arjona T, García-Ortiz M-V, Ruiz-Rubio M, Ariza RR (2000) cDNA cloning, expression and functional characterization of an Arabidopsis thaliana homologue of the Escherichia coli DNA repair enzyme endonuclease III. Plant Mol Biol 44:43–52.

Sánchez M, Sabio L, Gálvez N, Capdevila M, Dominguez-Vera JM (2017) Iron chemistry at the service of life. IUBMB Life 69:382–388.

Sánchez-Aparicio J-E, Tiessler-Sala L, Velasco-Carneros L, Roldán-Martín L, Sciortino G, Maréchal J-D (2021) BioMetAll: Identifying Metal-Binding Sites in Proteins from Backbone Preorganization. J Chem Inf Model 61:311–323.

Sehnal D, Bittrich S, Deshpande M, Svobodová R, Berka K, Bazgier V, Velankar S, Burley SK, Koča J, Rose AS (2021) Mol* Viewer: modern web app for 3D visualization and analysis of large biomolecular structures. Nucleic Acids Res 49:W431–W437.

Sello MM (2016) Comparative genomics of cytochrome P450 monooxygenases in newly explored pathogenic Oomycetes.

Sharma A, Sharma D, Verma SK (2017) Proteome wide identification of iron binding proteins of Xanthomonas translucens pv. undulosa: focus on secretory virulent proteins. BioMetals 30:127–141.

Shematorova EK, Slovokhotov IY, Shmakov VN, Khaliluev MR, Shpakovski DG, Klykov VN, Babak OG, Spivak SG, Konstantinov YM, Shpakovski G v (2019) Novel Interactions of Adrenodoxin-Related [2Fe-2S] Plant Ferredoxins MFDX1 and MFDX2 Indicate Their Involvement in a Wide Spectrum of Functions in Plant Mitochondria. Proceedings of the Latvian Academy of Sciences Section B Natural, Exact, and Applied Sciences 73:478–486.

Shi W, Chance MR (2008) Metallomics and metalloproteomics. Cellular and Molecular Life Sciences 65:3040–3048.

Sodhi JS, Bryson K, McGuffin LJ, Ward JJ, Wernisch L, Jones DT (2004) Predicting Metal-binding Site Residues in Low-resolution Structural Models. J Mol Biol 342:307–320.

Stairs CW, Kokla A, Ástvaldsson Á, Jerlström-Hultqvist J, Svärd S, Ettema TJG (2019) Oxygen induces the expression of invasion and stress response genes in the anaerobic salmon parasite Spironucleus salmonicida. BMC Biol 17:19.

Storcksdieck S, Bonsmann G, Hurrell RF (2007) Iron-Binding Properties, Amino Acid Composition, and Structure of Muscle Tissue Peptides from in vitro Digestion of Different Meat Sources. J Food Sci 72:S019–S029.

Subramanian B, Gao S, Lercher MJ, Hu S, Chen W-H (2019) Evolview v3: a webserver for visualization, annotation, and management of phylogenetic trees. Nucleic Acids Res 47:W270–W275.

Sun N, Cui P, Jin Z, Wu H, Wang Y, Lin S (2017) Contributions of molecular size, charge distribution, and specific amino acids to the iron-binding capacity of sea cucumber (Stichopus japonicus) ovum hydrolysates. Food Chem 230:627–636.

Todorovic S, Teixeira M (2018) Resonance Raman spectroscopy of Fe–S proteins and their redox properties. JBIC Journal of Biological Inorganic Chemistry 23:647–661.

Valasatava Y, Rosato A, Banci L, Andreini C (2016) MetalPredator: a web server to predict iron–sulfur cluster binding proteomes. Bioinformatics 32:2850–2852.

Valasatava Y, Rosato A, Cavallaro G, Andreini C (2014) MetalS3, a database-mining tool for the identification of structurally similar metal sites. JBIC Journal of Biological Inorganic Chemistry 19:937–945.

Verma SK, Sharma A, Sandhu P, Choudhary N, Sharma S, Acharya V, Akhter Y (2017) Proteome scale identification, classification and structural analysis of iron-binding proteins in bread wheat. J Inorg Biochem 170:63–74.

Vigouroux A, Aumont-Nicaise M, Boussac A, Marty L, lo Bello L, Legrand P, Brillet K, Schalk IJ, Moréra S (2020) A unique ferrous iron binding mode is associated with large conformational changes for the transport protein FpvC of Pseudomonas aeruginosa. FEBS J 287:295–309.

Walters EM, Johnson MK (2004) Ferredoxin:thioredoxin Reductase: Disulfide Reduction Catalyzed via Novel Site-specific [4Fe–4S] Cluster Chemistry. Photosynth Res 79:249–264.

Wang S, Chen W, Yang C, Yao J, Xiao W, Xin Y, Qiu J, Hu W, Yao H, Ying W, Fu Y, Tong J, Chen Z, Ruan S, Ma H (2016) Comparative proteomic analysis reveals alterations in development and photosynthesis-related proteins in diploid and triploid rice. BMC Plant Biol 16:199.

Wiederstein M, Sippl MJ (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res 35:W407–W410.

Yamori W, Kondo E, Sugiura D, Terashima I, Suzuki Y, Makino A (2016) Enhanced leaf photosynthesis as a target to increase grain yield: insights from transgenic rice lines with variable Rieske

FeS protein content in the cytochrome b6/f complex. Plant Cell Environ 39:80–87.

Yan F, Fujimori DG (2011) RNA methylation by Radical SAM enzymes RlmN and Cfr proceeds via methylene transfer and hydride shift. Proceedings of the National Academy of Sciences 108:3930–3934.

Yu C-S, Chen Y-C, Lu C-H, Hwang J-K (2006) Prediction of protein subcellular localization. Proteins: Structure, Function, and Bioinformatics 64:643–651.

Zheng H, Cooper DR, Porebski PJ, Shabalin IG, Handing KB, Minor W (2017) CheckMyMetal: a macromolecular metal-binding validation tool. Acta Crystallogr D Struct Biol 73:223–233.

Zhong Z, Ji Y, Fu Y, Liu B, Zhu Q (2015) Molecular characterization and expression analysis of the duck viperin gene. Gene 570:100–107.