International Journal of Therapeutic Applications, Volume 10, 2013, 11-18

ELUCIDATION OF IRON BINDING PATTERNS THROUGH INSILICO APPROACHES IN HUMAN IRON BINDING PROTEINS

[1]Muthukumaran Sivashanmugam, [3]AnushaKuppuswamy, [2]UmashankarVetrivel, [1]K.N. Sulochana*, Umashankar Vetrivel*

[1*] Department Of Biochemistry and Cell Biology,Sankara Nethralaya, 18, College Road, Nungambakkam, Chennai – 600 006, India

[2] Centre for Bioinformatics, Vision Research Foundation,Sankara Nethralaya, 18, College Road, Nungambakkam, Chennai – 600 006, India

[3]Prathyusha Institute of Technology and Management; Thiruvallur, India- 602025.

ABSTRACT

Human iron binding proteins play a crucial role in various cellular activities. Cellular iron imbalance can lead to many disease conditions. A comprehensive understanding of iron binding patterns in proteins is necessary to develop proper therapeutic interventions. By this study, we have collated the present classification strategies adopted in classifying the iron binding proteins. Additionally, we have also identified five novel iron binding patterns inferred and validated through bioinformatics studies. The outcome of this study will greatly help in database development for human iron binding proteins and also aid in prediction of iron binding signatures in uncharacterized proteins. Moreover, the identified iron binding patterns will also pave way for the design of iron sequestering peptides which shall be utilized as modulators in diseases arising due to iron imbalance.

Key words : Iron, iron binding proteins, phylogenetic analysis, iron binding pattern, peptide therapy.

1. INTRODUCTION

Iron (Fe) is an important trace element which plays a major role in production of RBC (haemopoiesis), transport of oxygen [1], production of enzymes, amino acids, hormones and also in proper functioning of immune system [2]. Fe present in myoglobin helps in proper storage & diffusion of oxygen into muscle cells [1]. It also helps in many of the crucial functions like conversion of sugar to energy, electron transport chain, DNA synthesis and redox reactions [1, 2, 3, and 4]. Fe imbalance may lead to various neurodegenerative diseases like Alzheimer's disease, Parkinson's disease, epilepsy, and multiple sclerosis and very commonly anemia. Hence, iron homeostasis should be highly organized with proper uptake of iron from diet, controlled transport across biological membranes by iron chaperones, distribution of iron throughout the body and finally detoxification and also excretion from the body.

During bacterial infection, free iron is retained in the body by iron withholding trigger mechanism stimulated by the immune system. This prevents the utilization of free iron by the bacteria for its survival. Bacteria also produce an iron binding molecules called siderophores, which also aids in iron absorption. Individuals with increased iron content are more prone to bacterial infections. As a defensive mechanism to combat this condition, hepcidin is produced to induce anti inflammatory functions in humans [5].

Iron deficiency in human is also referred as sideropenia or hypoferremia. In this disorder, storage of iron is depleted and the level of hemoglobin in the blood is also reduced. Children and premenopausal women are prone to this disease. The main causes of iron deficiency are chronic bleeding, inadequate intake of substance

Corresponding author:
suloravi@gmail.com, vumashankar@gmail.com

that interferes iron absorption, malabsorption syndrome, blood donation and inflammation due to bacterial growth.Iron deficiency can be controlled withappropriate dietary supplements [6].In general, oral drugs that contain iron in form of ferrous sulfate isadministered to sideropenia patients [7].

The two main reasons of iron overloading or hemochromatosis are blood transfusion and mutationsin HFE gene (High Fe). Iron overloadingleads to conditions like diabetes mellitus, irregular heartbeat, arthritis, liver cancer, depression, hypothyroidism, hypogonadism, gall bladder disease and many neurodegenerative diseases [8].Iron chelators are used for treatment ofiron overload. Desferrioxamine, ferric chrome, paraphernalia are some of the iron chelators used for the treatment. But these chelators are poorly absorbed in the gastrointestinal tract and many of them also exhibit non-specificity to iron [8]. Moreover, these chelators also produce many side effects. Iron quenching through peptide chelators shall prove to be biologically safe therapy. Hence, in this study we propose a conserved sequence and structurebased *insilico* method which shall provide insights towards the design of potential Iron chelating sequences.

2. MATERIALS METHODS

The sequence data of different iron binding proteins from human were obtained from UniProtKB database (www.uniprot.org). Totally, 149 sequences were found as "iron binding proteins in *Homo sapiens*" as per UniProtKB test entry at the time of search. All these proteins collected were found to be involved in iron metabolism. However, only 107 proteins out of 149 were found to have direct iron binding function and were further used for the study. Finally, the conserved iron binding patterns were identified by adopting the protocol as implemented by Mathangi *et al.,* 2012 in case of copper binding proteins [9].

The 107 iron binding proteins were further classified based on their respective molecular functions as per gene ontology predictions, which fall into 4 major known categories namely, enzymes, electron transport system, metal ion binding region, transporter proteins and an unknown category (unknown function). Further, the major known categories were classified based on Enzyme type using EC number which resulted sorting in 6 enzyme class namely, oxidoreductases, transferases, hydrolases, lyases, isomerases and ligases. Further both the enzymatic and non-enzymatic proteins were classified into major categories based on cellular localization in human [10].

Multiple sequence alignment of all 107 iron binding proteins was performed and the dendrogram based on neighbor-joining method was generated using ClustalW [http://www.genome.jp/tools/clustalw]. Further, the sequences were classified based on the Clade formation. The sequences of each clade were separately subjected to multiple sequence alignment using MEGA and ClustalW software. Finally, the conserved regions specific to each clade were identified corresponding iron binding patterns were design and were subsequently validated for iron binding property by PRATTdatabase[11]and GOmotif database [12].The patterns were verified using GO motif database to get information on Gene Ontology (molecular function, biological process, cellular components)[12].

The 3D structure of iron binding proteins co-crystallized with iron were retrieved from Protein Data Bank to identify key interacting residues and also to explore the various chemical bonding patterns of iron binding proteins with different oxidation forms of iron at the time of search[13]. The Iron binding cavities of these proteins were analyzed using Ligplus software plugin [14] installed in PyMOL [15].

3. RESULTS

In this study we attempted to classify the iron binding proteins based on the documented protocol as implemented by our group in case of copper

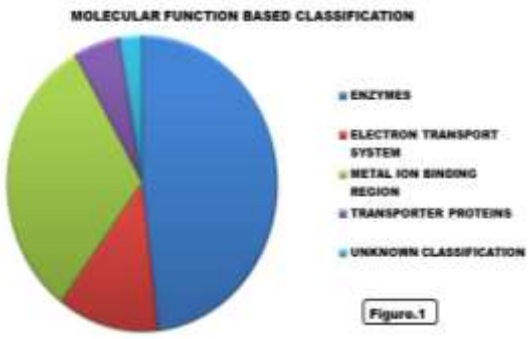

Figure.1. Distribution of different categories of iron binding proteins based on molecular function

binding proteins [9]. Totally 107 iron binding proteins were obtained from UniProtKB and were grouped into 5 major categories based on their molecular function. The 107 iron binding proteins are classified as enzymes, electron transport system, metal ion binding region, transporter proteins and proteins with unknown classification as shown in [**Fig.1]**.

3.1 ENZYME BASED CLASSIFICATION

The enzymes identified in this study were further sorted into six classes namely, oxidoreductases, transferases, hydrolases, lyases, isomerases andligases based on the Enzyme commission number [**Fig. 2**].

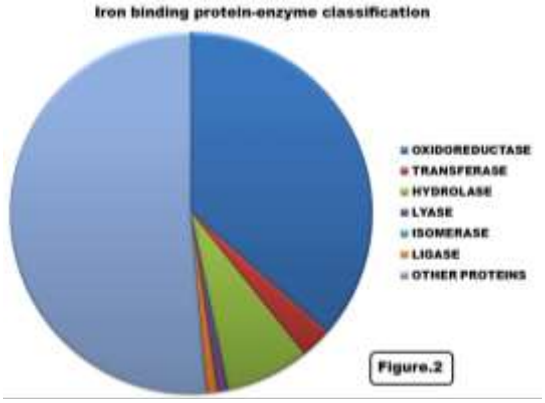

Figure. 2. Distribution of different iron binding proteins based on enzyme type.

3.2 PROTEIN LOCALIZATION BASED CLASSIFICATION

Identification of sub-cellular localization of iron binding proteins will be helpful in understanding the key features like nature of solvent in which the protein is located, overall charge of the protein, metabolic pathways involved etc. Iron binding proteins in this study were found to be localized in mitochondria, cytoplasm, nucleus and endoplasmic reticulum as shown in **[Fig.3]**.

3. 3 CONSTRUCTION OF PHYLOGENETIC TREE

Phylogenetic tree was constructed using the 107 iron binding proteins and the similar proteins were grouped together in a single internode. Default parameters for Unweighted Pair Group Mean Averagemethod in ClustalWwas used for the

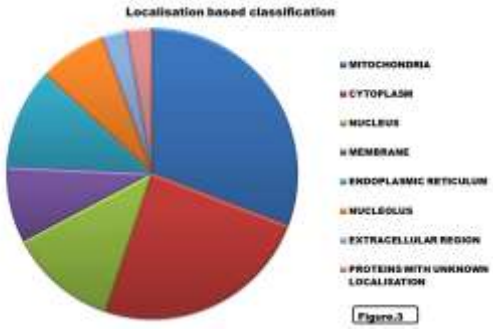

Figure. 3. Distribution of various localization of selected iron binding proteins.

construction of phylogenetic tree. The different iron binding groups among the 107 proteins were sub-classified based on the clade formation.

3.4 IDENTIFICATION OF IRON BINDING MOTIF

Based on the sub-classified clades with proper topology,6 different iron binding motifs were identified. This was done by multiple sequence

TABLE. 1. PATTERN DETAILS OF IRON BINDING PROTEINS

S.No	Pattern identified	Number of sequence obtained from GOmotif database	Number of sequence obtained from PRATT database	Classification
1	**PFSXGX RXCXG**	29	29	Cyto-chrome protein
2	**DGX(13)GP**	174	187	Iron sulfur cluster protein
3	**DXHX(26)D**	478	524	Heme protein
4	**HXDX(7)Y**	320	400	Iron sulfur cluster protein
5	**WXXAXRCX G**	4	6	Others
6	**DXHX(64)D**	494	538	Others

alignment of proteins among the individual cluster followed by identification of conserved residues in the alignment pertaining to iron binding activity.

3.5 PATTERNDESIGN & VALIDATION

The identified motifs werefurther analyzed for regular expressions and consensus patterns were obtained and tabulated [**Table.1**]. Further, the designed patterns were searched against different pattern database using GOMotif and PRATT. This identified all the proteins which were used to derive these patterns and is suggestive of the predictive accuracy. Earlier studies revealed that the heme binding motif for bacteria, plant and mammalian cytochrome proteinsshare a common pattern which was also observed synonymously in this work [16]. Moreover, we also identified few novel sequence patterns which shall possess iron binding activity. Each pattern identified was named based on the respective phylogeny lineage it represents.

3.7 STRUCTURAL STUDY OF HUMAN IRON BINDING PROTEIN

Iron exist in different protonation states in human body like Fe(III), Fe(II). It also coexists in combination with other atoms, molecules or compounds (i.e.) it can combine with porphyrin rings to form heme and also it can combine with sulfur to form iron sulfur cluster. In this study, 220 iron containingatomic coordinate structure files were retrieved from Protein Data Bank by text mining using “iron” as keyword at the time of search. The retrieved structures were further classified into 4 major types as heme binding proteins[Heme], iron sulfur cluster proteins[FeS], ferrous binding proteins[Fe(II)] and ferric binding proteins[Fe(III)] based on protonation states and iron coexistence [**Fig 4**]. Iron forms interactwith other molecules through hydrogen bonds, coordinate covalent bonds and hydrophobic bonds. In general, **Cys, His, Asp, Tyr, Ser, Gln** and **Arg** are the amino acid residues

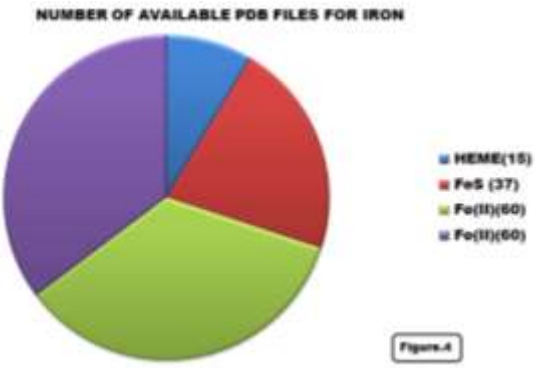

Figure.4. Pdb files available for different iron protonation state.

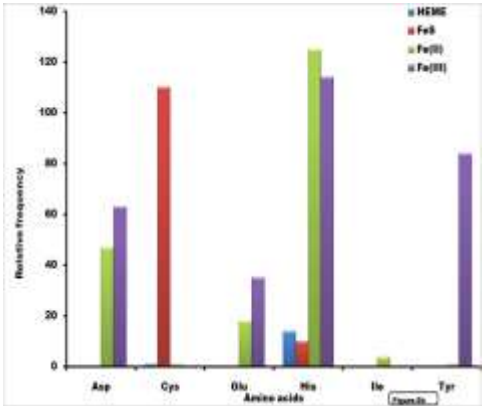

Figure. 5a. Residues forming hydrogen bond with different forms of iron.

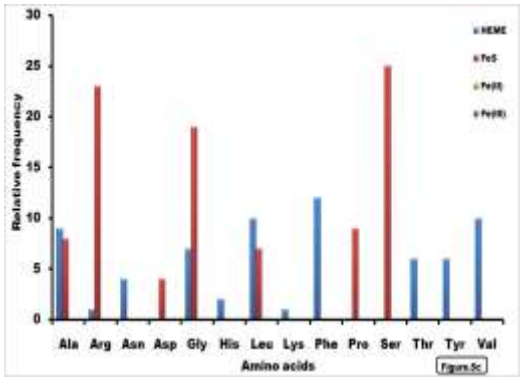

Figure.5b.Residues forming coordinate covalent bond with iron.

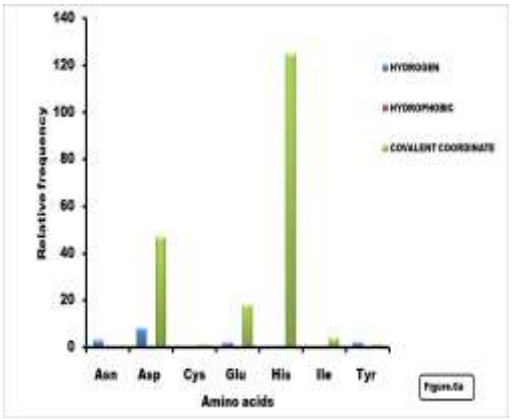

Figure. 5c. Residues forming hydrophobic interaction with iron.

that plays a vital role in interaction of iron with its interacting proteins [17]. Hence, the structures used in this study were analyzed for the bonding nature, protonation states of iron and the data is graphically represented in **Fig 5a, 5b & 5c**. Similarly, the key iron interacting residues of each protonation state were analyzed and represented in **Fig 6a, 6b, 6c & 6d**. In table 2 and table 3, the set of key interacting residues present in the binding pocket of Fe (III) and Fe (II) binding proteins were illustrated.

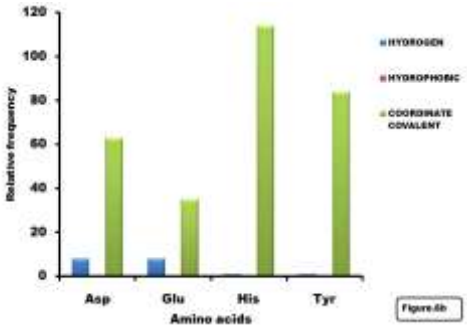

Figure. 6a. Interaction of iron binding residues with Fe (II) binding proteins

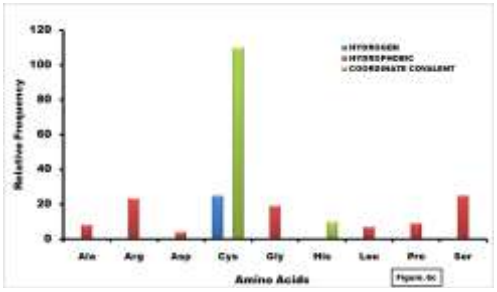

Figure. 6b. Interaction of iron binding residueswith Fe (III) binding proteins

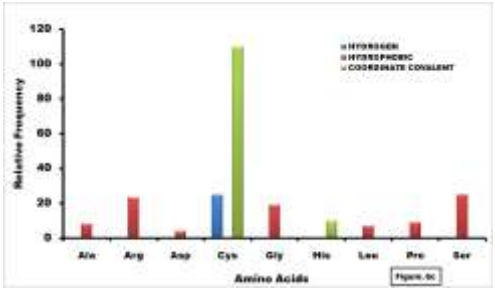

Figure. 6c. Interaction of iron binding residues with Fe-S binding proteins

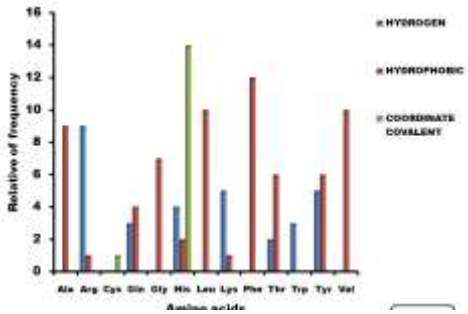

Figure. 6d. Interaction of iron binding residues with heme binding proteins

4. DISCUSSION

By this study, we have attempted to classify iron binding proteins based on the molecular function which inferred that majority of these proteins to be enzymes (49%). Furthermore, only few proteins fall under the category of iron transporters (6%) and proteins involved in electron transport system (11%). Few other proteins were classified as proteins of unknown function as per gene ontology classification. In the Enzyme category, 52 iron binding enzymes were classified further classified into 6 classes using their respective EC number. Most of these enzymes were Oxidoreductases which catalyze redox reactions. None of the iron binding enzymes fall under the category of Isomerases. Many of the iron binding enzymes were found to be localized in mitochondria and cytoplasm, as per the sub-cellular localization prediction. However, few of these enzymes were also predicted to span in membrane, nucleolus and extracellular region.

Phylogenetic analysis of 107 iron binding proteins revealed the formation 32 clades in the dendrogram. 6 clades with proper topology were used for generating sequence patterns. Finally, six significant patterns were designed: PFSXGXRXCXG for cytochrome proteins, DGX(13)GP for iron sulfur cluster proteins, DXHX(26)D for Serine/threonine-protein phosphatase, HXDX(7)Y for lysine specific demethylase family, WXXAXRCXG for nitric oxide synthase proteins and DXHX(64)D for Egl homologs. These designed sequence patterns were validated using the GOmotif and PRATT search tools. Further, the search results of PRATT and GOmotif were analyzed for the occurrence of the sequences which were utilized to design the respective patterns. Moreover, other novel proteins retrieved through

this search were analyzed for similar iron binding activity which yielded promising results.

The structural analysis of iron binding proteins involved a detailed study of the hydrogen bonding pattern, covalent coordinate bond formation and hydrophobic bond in the iron binding cavity. A hydrogen bond is a type of attractive (dipole-dipole) interaction between an electronegative atom and a hydrogenatombonded to another electronegative atom. Hydrogen bond in iron binding protein helps in the interaction of dioxygen to iron. Generally, hydrogen bondsare stronger than Vander Waals forces, but weaker than covalent bonds or ionic bonds.As per this study, **Cys, Asp, Gln** and **Arg**in iron binding cavities were found to contribute more towards hydrogen bonding with iron. **Cys**was found to be major contributor for hydrogen bonding in case of iron binding cavities of iron sulfur cluster proteins. **Arg** was found to be the major hydrogen bond contributors in heme shown in [**Fig.5a]**.

The coordinate covalent interactions are essential for the iron-protein interactions. In general, **His** in all of its protonation statescontributes towards coordinate covalent interactions with iron.However, in case of iron sulfur cluster proteins **Cys** residues are the only contributor of coordinate covalent bond with iron. Moreover, Ferrous and Ferric states of iron were found to form coordinated covalent bonds with proteins through **Asp** and**Glu** as shown in [**Fig.5b]**.

Hydrophobic bonds in iron binding proteins play an important role in stabilizing the interactions between iron and the protein. In this study, hydrophobic interactions were mainly observed in heme binding and iron sulfur cluster proteins. **Phe, Val, Leu, Ala** and **Gly**residues of heme binding proteins form hydrophobic bonds with heme.In case of Fe-S cluster, **Ser, Arg, Gly** and **Pro** were found to confer hydrophobic interaction [**Fig.5c]**.

Fe(III) is the bi-cation form of iron which is commonly called as ferric form of iron. Normally the ferric form of iron is reduced to ferrous form [Fe(II)] by enzymatic action of ferric reductase which occurs in enterocytes of intestine. **Asp, Glu, Tyr** and **His** are the fourresidues of iron binding proteinswere found to be involved in binding with Fe(III). Among the above mentioned residues, **His** is found to found to form Covalent coordinated bonds in majority of the Fe(III) binding proteins compared **Tyr, Asp** and **Glu**. Moreover, no hydrophobic interactions were observed [**Fig.6a]** and also the set of key interacting residues in the binding pocket of the Fe(III) binding proteins were illustrated in[**Table.2]**

TABLE. 2. SET OF KEY RESIDUES PRESENT IN THE CAVITY OF Fe(III) BINDING PROTEIN

S.No	KEY INTERACTING RESIDUES IN THE CAVITY	NO OF PDB FILES AVAILABLE
1	Tyr, Tyr, His, Asp	35
2	Glu, His, His, Asp	15
3	Glu, Glu, Glu, His, Asp	4
4	Glu, His, His	4
5	Asp, His, His	5
6	Others	20

Fe(II) is the ferrous form of iron which is observed many of iron binding class of human proteins. This protonated state of iron is the most widely absorbed form of dietary iron. Fe(II) interacts with proteins through coordinated covalent bonds and hydrogen bonding with **His, Asp, Glu, Ile, Tyr** and**Asn** residues **[Fig.6b]**. However, no hydrophobic interactions were observed as same as in case of Fe(III) binding proteins. The set of key Fe(III) interacting residues in the binding pocket of the Fe(III) binding proteins were tabulated in the **[Table.3].**

TABLE. 3. SET OF KEY RESIDUES PRESENT IN THE CAVITY OF Fe(II) BINDING PROTEINS

S.No	key interacting residues in the cavity	NO OF PDB FILES AVAILABLE
1	His, His, Asp	27
2	His, His, Glu	3
3	His, His, Asp, Glu	6
4	His, His, Asp, Asp	7
5	others	13

Ironsulfur cluster proteins exist in different forms like Fe-S, Fe_2S, Fe_2S_2 and Fe_4S. The key iron and sulphur interacting residues observed in iron sulfur clusters are **Gly, Pro, Ala, Leu, Arg, Cys, Ser, His, Asp.**Among these residues, only **Cys** and **His** were found to be involved coordinated covalent bond with iron, whilst, other residues were to interact exclusively with sulphur atom **[Fig.6c]**.

Heme is an iron binding compound which interacts to a centered Fe atom through its 4 porphyrin rings. **Gly, Ala, Val, Leu, Arg, Lys, Try, Gln, Cys, Thr, His, Phe** and **Trp**residueswere found to be mainlyinvolved heme-protein interactions in case of heme binding proteins. Among all the residues mentioned above, **His**and **Cys**form direct coordinated covalent bond with Fe atom thereby, stabilizing and accommodating the Heme molecule. Whereas, **Gly, Ala, Val, Leu** and **Phe** found to contribute more towards hydrophobic interactions with Heme molecule. Moreover, **Arg, Lys, Try, Gln, Trp** and **Thr** were found be more involved in the formation of hydrogen and hydrophobic bonds with heme as shown in **[Fig.6d]**.

CONCLUSION

Iron is one of the important trace elements in the human body and it becomes essential to understand its role various cellular processes. This outcome of this study shall aid in creation ofa knowledge base on the sequence patterns and molecular interactions observed in binding proteins. The sequence and structural patterns designed in this study will be significantly help in identifying unreported novel iron binding proteins.Moreover, the sequence and structural insights explained in this study shall pave way for designing of peptides which can be utilized for therapeutic interventions in case of disease caused due to iron imbalance.

ACKNOWLEDGMENT

Support: We sincerely thank Sankara Nethralaya for granting access to use CLC Main Workbench.

REFERENCE

1. Beard JL. Iron biology in immune function, muscle metabolism and neuronal functioning. J Nutr. 2001 Feb; 131(2S-2):568S-579S; discussion 580S. Review. PubMed PMID: 11160590.

2. Andrews NC. Disorders of iron metabolism. N Engl J Med. 1999 Dec 23;341(26):1986-95. Review. Erratum in: N Engl J Med 2000 Feb 3;342(5):364.PubMed PMID: 10607817.

3. Chitambar CR, Narasimhan J. Targeting iron-dependent DNA synthesis with gallium and transferrin-gallium. Pathobiology. 1991; 59(1):3-10. PubMed PMID:1645976.

4. Dugan TR, Bill E, MacLeod KC, Christian GJ, Cowley RE, Brennessel WW, Ye S, Neese F, Holland PL. Reversible C-C bond formation between redox-active pyridine ligands in iron complexes. J Am Chem Soc. 2012 Dec 19;134(50):20352-64. doi:10.1021/ja305679m. Epub 2012 Dec 11. PubMed PMID: 23181620.

5. Lammi-Keef CJ, CouchSC, Philipson EH, ed. Dietary diversification and modification of iron. Handbook of Nutrition and Pregnancy. Nutrition & Health. Totowa, New Jersey: Humana Press. 2008 pp. 350–351. doi:10.1007/978-1-59745-112-3. ISBN 978-1-59745-112-3.

6. http://www.emedicinehealth.com/anemia/article_em.htm#anemia_overview

7. Hoppe M, Hulthén L, Hallberg L. The relative bioavailability in humans of elemental iron powders for use in food fortification. Eur J Nutr. 2006 Feb; 45(1):37-44. PubMed PMID: 15864409.

8. Kalinowski DS, Richardson DR. The evolution of iron chelators for the treatment of iron overload disease and cancer. Pharmacol Rev. 2005 Dec; 57(4):547-83. Review. PubMed PMID: 16382108

9. Mathangi Jayaram, Sudarshana N, Vidya Natarajan, Umashankar , Sulochana K N. Elucidation of structure and function in families of copper binding proteins in Homo sapiens(2012).Int J Pharm Bio Sci. 647 - 662.

10. Lodish H, Berk A, Zipursky SL, et al. Molecular Cell Biology. 4th edition. New York: W. H. Freeman; 2000.

11. Jonassen I, Collins JF, Higgins DG. Finding flexible patterns in unaligned protein sequences.

Protein Sci. 1995 Aug;4(8):1587-95. PubMed PMID: 8520485.

12. Bristow F, He R, Van Domselaar G. GOmotif: A web server for investigating the biological role of protein sequence motifs. BMC Bioinformatics. 2011 Sep 26;12:379. doi: 10.1186/1471-2105-12-379. PubMed PMID: 21943350.

13. Gilliland GL, Berman HM, Battistuz T, Bhat TN, Bluhm WF, Bourne PE, Burkhardt K, Feng Z, Iype L, Jain S, Fagan P, Marvin J, Padilla D, RavichandranV,Schneider B, Thanki N, Weissig H, Westbrook JD, Zardecki C. The Protein Data Bank. ActaCrystallogr D BiolCrystallogr. 2002 Jun;58(Pt 6 No 1):899-907. Epub 2002 May 29. PubMed PMID: 12037327.

14. Wallace AC, Laskowski RA, Thornton JM. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. Protein Eng. 1995 Feb;8(2):127-134. PubMed PMID: 7630882.

15. Hagelueken G, Ward R, Naismith JH, Schiemann O. MtsslWizard: In Silico Spin-Labeling and Generation of Distance Distributions in PyMOL. ApplMagn Reson.2012 Apr; 42(3):377-391. Epub 2012 Feb 3. PubMed PMID: 22448103.

16. Li T, Bonkovsky HL, Guo JT. Structural analysis of heme proteins: implications for design and prediction. BMC Struct Biol. 2011 Mar 3;11:13. doi:10.1186/1472-6807-11-13. PubMed PMID: 21371326.

17. Eckenroth BE, Steere AN, ChasteenND, Everse SJ, Mason AB. How the binding of human transferrin primes the transferrin receptor potentiating iron release at endosomalpH. Proc NatlAcadSci U S A. 2011 Aug 9; 108(32):13089-94. doi:10.1073/pnas.1105786108. Epub 2011 Jul 25. PubMed PMID: 21788477.