People's Democratic Republic of Algeria

The Ministry of Higher Education and Scientific Research

Faculty of Science

University of Saida - Dr.Moulay Taher

Department: Biology

Memory

Presented for obtaining the Master's degree in: Biology

Specialty: Biochemistry

THEME:

Bioinformatics based Characterization of Porphyrin Protein Binding Structural Motifs and Construction of a relevant Database: Unravelling some Essentials of Protein Structure-Function Relationships (https://bioinformatics.univ-saida.dz/prjs/ppbsms/).

- **Prepared by :**

➢ Megherbi Mohamed El Mehdi
➢ Bitar Mohamed

- **In front of jury commission, composed by**

❖ President: Dr. Bellil Yahia	MCA	Univ. of Saida - Dr. Moulay Taher.
❖ Examiner: Dr. Benabbou Taha Ahmed	MCB	Univ. of Saida - Dr. Moulay Taher.
❖ Supervisor: Dr. Abdelkrim Rachedi	MCA	Univ. of Saida - Dr. Moulay Taher

2022/2023

Acknowledgement:

We thank our great God for all he gives us.

First and foremost, we would like to extend our heartfelt appreciation to our mentor, Dr. RACHEDI Abdelkrim, for his invaluable guidance and advice, which greatly contributed to the successful completion of this final assignment study.

We would also like to extend our profound respect to all the professors who played a significant role in our training and education. Special recognition is due to the members of the jury for their willingness to evaluate and assess this project.

Furthermore, we would like to extend our thanks to all our colleagues in Master 2 Biochemistry, whose support and collaboration have been instrumental throughout this journey.

Dedication

I extend my sincere dedication, recognition, and profound respect to my beloved parents, who epitomize the embodiment of love and tenderness. This humble work stands as a tribute to their unwavering support and guidance, which have been instrumental in shaping my academic journey. Their profound impact on my life deserves the utmost appreciation, through this dedication:

- **To my mother, may Allah have mercy on her soul.**
- **To my father, may Allah protect him.**
- **A special dedication for my grandmother.**
- **To my brother: Mohamed.**
- **To my beloved sister: Houda.**
- **To the angel: Wassim.**
- **To my partner: Bitar Mohamed.**
- **To all my professors.**

Not to mention my colleagues in master 2 Biochemistry 2023.

Mehdi

Dedication

I extend my sincere dedication, recognition, and profound respect to my beloved parents, who epitomize the embodiment of love and tenderness. This humble work stands as a tribute to their unwavering support and guidance, which have been instrumental in shaping my academic journey. Their profound impact on my life deserves the utmost appreciation.

Through this dedication.

- **To my Twin: Abderrahmane.**
- **To my sisters.**
- **To my partner: Megherbi Med El Mehdi.**
- **To all my professors.**

Not to mention my colleagues in master 2 Biochemistry 2023.

Mohamed

Abstract

Abstract

This project seeks to explore more the basis behind Structure-Function relationship in biological context of macromolecules; the proteins in the case of this study. Furthermore, the study draws attention to results that would touch upon distant evolutionary relations across species.

The understanding of the structure-function relationship in important for deeper studies of the subject of biological function of proteins, and macromolecules in general, in both health and disease situations.

One way to undertake the kind of study is to analyse the proteins structures involved in selected biological functions and examine their ligand binding environments.

In this project, a set of 3D-structures of porphyrin proteins representing 51 full chains from 21 Protein Databank (PDB) entries from different source organism or species have been studied in terms of their binding sites environment around and number porphyrin groups. Thus study implemented structural bioinformatics tools developed at the Department of Biology, Then University of Saida,

The study resulted in the identified and construction a set of **43 unique Structural and Functional Binding Motifs** what have been characterised and analysed their binding site residues content and properties. These motifs are associated with a number of vital biological functions including oxygen transport, storage, light harvesting and energy production and more.

The data and analysis pertaining to this project have been stored in a Flat-File database type named as **Porphyrin Proteins Binding Structural Motifs – PPBSMs**.

The database PPBSMs has been made available online for the world wide community of scientist and researchers though the Web-server of University of Saida via the following web-address:

https://bioinformatics.univ-saida.dz/prjs/ppbsms/

Keyword; Porphyrin proteins, Porphyrin ligands, Heme, Chlorophyll, B12, Structural Bioinformatics Structural & Functional Motifs, Ligand Binding Environment, Amino Acids, Residues, Databases.

Résumé

Résumé

Ce projet vise à explorer davantage les bases de la relation structure-fonction dans le contexte biologique des macromolécules, en particulier les protéines dans le cadre de cette étude. De plus, l'étude attire l'attention sur des résultats qui toucheraient aux relations évolutives éloignées entre les espèces.

La compréhension de la relation structure-fonction est importante pour des études approfondies sur le sujet de la fonction biologique des protéines et des macromolécules en général, tant dans les situations de santé que de maladie.

Une façon d'entreprendre ce type d'étude consiste à analyser les structures protéiques impliquées dans des fonctions biologiques sélectionnées et à examiner leurs environnements de liaison avec les ligands.

Dans ce projet, un ensemble de structures 3D de protéines porphyrines représentant **51 chaînes complètes issues de 21 entrées de la Protein Data Bank** (PDB) provenant de différentes sources d'organismes ou d'espèces a été étudié en termes d'environnement des sites de liaison et du nombre de groupes porphyrine. Ainsi, l'étude a mis en œuvre des outils de bioinformatique structurale développés au Département de biologie de l'Université de Saida.

L'étude a permis **d'identifier et de construire un ensemble de 43 motifs structuraux et fonctionnels de liaison uniques** qui ont été caractérisés et analysés pour leur contenu en résidus et leurs propriétés de site de liaison. Ces motifs sont associés à un certain nombre de fonctions biologiques essentielles, notamment le transport de l'oxygène, le stockage, la capture de la lumière et la production d'énergie, et d'autres encore.

Les données et les analyses relatives à ce projet ont été stockées dans une base de données de type Flat-File appelée **Porphyrin Proteins Binding Structural Motifs - PPBSMs**.

La base de données **PPBSMs** a été rendue disponible en ligne pour la communauté mondiale de scientifiques et de chercheurs via le serveur Web de l'Université de Saida à l'adresse suivante:

https://bioinformatics.univ-saida.dz/prjs/ppbsms/

Mots-clés: protéines porphyrines, ligands porphyrines, hème, chlorophylle, B12, bioinformatique structurale, motifs structuraux et fonctionnels, environnement de liaison avec les ligands, acides aminés, résidus, bases de données.

الملخص

الملخص:

يهدف هذا المشروع إلى استكشاف أسس العلاقة بين الهيكل (التركيب الفراغي) والوظيفة في السياق البيولوجي للجزيئات الضخمة؛ وتحديداً البروتينات في حالة هذه الدراسة. علاوة على ذلك، تتعرض نتائج الدراسة إلى استنتاج أولي يمس العلاقات التطورية بين الأنواع البعيدة.

فهم العلاقة بين الهيكل والوظيفة أمر مهم لدراسة أعمق حول موضوع وظيفة البروتينات والجزيئات الضخمة بشكل عام، في الحالات الصحية والمرضية على حد سواء.

واحدة من الطرق للقيام بهذا النوع من الدراسة هي تحليل التراكيب الفراغية للبروتينات المساهمة في وظائف بيولوجية محددة وفحص بيئات إرتباط المواقع الفعالة بالمواد أو الليجاندات المرتبطة ذات العلاقة.

في هذا المشروع، تمت دراسة مجموعة من الهياكل ثلاثية الأبعاد لبروتينات البورفيرين Porphyrin proteins **و قد ضمت 51 سلسلة كاملة موجودة في 21 مدخل في قاعدة بيانات البروتينات** (PDB) من مصادر تمثل أنواع حيوية مختلفة. وقد تم تنفيذ هذه الدراسة باستخدام أدوات المعلوماتية_الحيوية التركيبية Structural Bioinformatics التي تم تطويرها في قسم البيولوجيا بجامعة سعيدة.

أسفرت الدراسة **عن تحديد وبناء مجموعة من 43 نمط تركيبي ووظيفي متفرد**، وتم تحليل محتوى وخصائص مواقع الإرتبط. هذه الأنماط التركيبية تتعلق بعدد من الوظائف البيولوجية الحيوية الهامة بما في ذلك نقل الأكسجين وتخزينه وحصد الضوء وإنتاج الطاقة وغيرها.

تم تخزين البيانات والتحاليل المتعلقة بمشروع البحث في قاعدة بيانات من نوع Flat-Files و التي تم تسميها بـ Porphyrin Proteins Binding Structural Motifs - PPBSMs.

قاعدة البيانات PPBSMs متوفرة عبر الإنترنت لمجتمع العلماء والباحثين على مستوى العالم من خلال خادم الويب لجامعة سعيدة عبر العنوان التالي:

https://bioinformatics.univ-saida.dz/prjs/ppbsms/

الكلمات المفتاحية: بروتينات البورفيرين، ليجاندات البورفيرين، الهيم، الكلوروفيل، بي 12، البيوإنفورماتيك الهيكلي، الأنماط الهيكلية والوظيفية، بيئة ربط الليجاند، الأحماض الأمينية، البقايا، قواعد البيانات.

Table of contents

Table of Contents :

Table of contents

Table of contents

Table of contents

List of Abbreviations:

PDB: Protein Data Bank .

PPI: Protein-Protein Interaction.

SSFS: Sequences Structures Function Server.

wwPDB: World Wide Protein Data Bank .

NMR : Nuclear Magnetic Resonance.

NOS: Nitric oxide synthases.

HEM: Hemoglobin.

CLA: Chlorophyll A.

CHL: Chlorophyll B.

CL7: Chlorophyll D.

F6C: Chlorophyll L.

BCL: Bacteriochlorophyll.

HEC: Heme C.

SIR: Sirohydrochlorin.

PSI: Photosystem I.

PSII: Photosystem II.

URL: Uniform Resources Locator.

List of Figures

List of Figures:

List of Figures

List of Figures

List of Tables:

General Introduction

General Introduction

The biochemical function of a protein is generally dictated by the 3D structure of the polypeptide chain. Structural motifs refer to specific 3D arrangements or patterns of atoms, functional groups, or molecular fragments within a molecule. These motifs play a significant role in determining the physicochemical properties, biological activity, and pharmacokinetic behavior of a medicament.

The biochemical function of protein structure is governed by many factors including structure motifs which are specific regions in proteins that are associated, amongst a number of roles, such as binding substrates, cofactors and other ligands necessary for relevant biological functions. One example of importantly functional structural motifs is **helix-turnhelix** motif which can bind DNA and help in gene transcription regulation and is difficult to identify from the amino acid sequence alone (EMBL-EBI, Voet and Voet, 2004).

The list of structural motifs identification is growing as research in the field of structural biology continues. The study undertaken in this project, looks at whether structural motifs play role in Porphyrin binding proteins such as Hemoglobin, Cytochromes and many other functional proteins. Porphyrins proteins, and are essential for many processes in living organisms, including respiration and cellular metabolism which can be behind illness cases and even human life loses when gone wrong such as in cases of cancers. Characterising binding motifs that may be linked to the functionality of these protein would play major contribution into the understanding better their function and their pathological alterations which can help in novel-drug discoveries.

Porphyrin proteins are characterized by a number of structural motifs that are important for their function. For example, hemoglobin and myoglobin which are a subclass of the heme binding proteins (containing Iron ion) contain a "globular" domain that binds and carries oxygen, as well as an "alpha-helical" domain that helps to stabilize the protein's structure. Cytochromes contain a "cytochrome" motif that is important for electron transfer reactions. Cobalamin or vitamin B12 functional in DNA synthesis and nerve function relies on porphyrin ring system that binds Cobalt ion instead. Also, porphyrin groups that bind Magnesium ion are

essential part of Chlorophyll A and B and Bacteria chlorophyll and play roles in photosynthesis and phosphorylation.

This research study project is about an investigation and discovery of structural motifs in relation to a number of proteins, from different species, binding porphyrin groups of different types including Heme and derivatives, Chlorophyll types A, B, D and F and Cobalamin.

It should be noted that this project implements techniques that fall under the field of Structural Bioinformatics as this latter uses informatics science to study and analyze biological structures such as proteins structure motifs and their biological functions.

This project is distributed as follows:

First chapter: A literature review through generality on the protein and Structural Motifs. The chapter covers protein possible motifs, Structural motifs, Amino Acids, Porphyrins definition, Porphyrins classes and the PDB database (Protein Data Bank).

Second chapter: Describes the Materials and Methods used to study and analysis of the data pertaining to the themes of the project.

Third chapter: Results and Discussion, contains presentation of the results obtained from the structural data analysis followed by discussion of what the results may mean and indicate to.

This chapter is ended with a general conclusion around the benefits of the study and future orientations.

Chapter I

I .Generality on Proteins and Structural Motifs:

Proteins are linear chains of covalently connected molecules called amino acids. Their sequences are encoded in DNA segments called genes. Protein is a complex organic molecule that plays numerous critical roles in living organisms. They are composed of one or more chains of amino acids linked together by peptide bonds. Proteins are involved in various functions such as structural support, transport of molecules, catalyzing biochemical reactions, and regulation of gene expression, among others.(Alberts B, Johnson A, Lewis J, Raff, M., Roberts, K., & Walter, P.2002).

I.1.Protein Sequence Motifs:

Protein motifs refer to small, conserved amino acid sequences or patterns within a larger protein sequence that have specific functions. These motifs can be structural, functional, or both, and they play a crucial role in protein-protein interactions, protein localization, and enzymatic activity (Eddy SR .1998).

Some common examples of protein motifs include the the helix-turn-helix motif, zinc finger motif, and the leucine zipper motif. These motifs are often found in DNA-binding proteins involved in DNA transcription regulation through the interaction with specific sequences (Keskin O, Nussinov R. 2007).

I.1.1 Helix-turn-helix motif:

This motif consists of two alpha helices separated by a turn composed by few amino acids. The motif is commonly found in DNA-binding proteins, **Figure 1**. where the two alpha helices fit into the major groove of DNA to form a protein-DNA complex. Examples of proteins containing the helix-turn-helix motif include the Lac Repressor protein and the homeodomain proteins (Aravind L, Anantharaman V, Balaji S, Babu MM, Iyer LM .2005).

Figure 1. A typical winged Helix-turn-helix motif (PDB: 3JSO)

I.1.2.Zinc finger motif: This motif involves the coordination of a zinc ion by several cysteine and histidine residues, forming a finger-like structure. Zinc finger motifs, **Figure 2**. are found in a variety of DNA-binding proteins, RNA-binding proteins, and enzymes. Examples of proteins containing zinc finger motifs include the steroid hormone receptor and the transcription factor TFIIIA (Laity JH, Lee BM, Wright PE .2001).

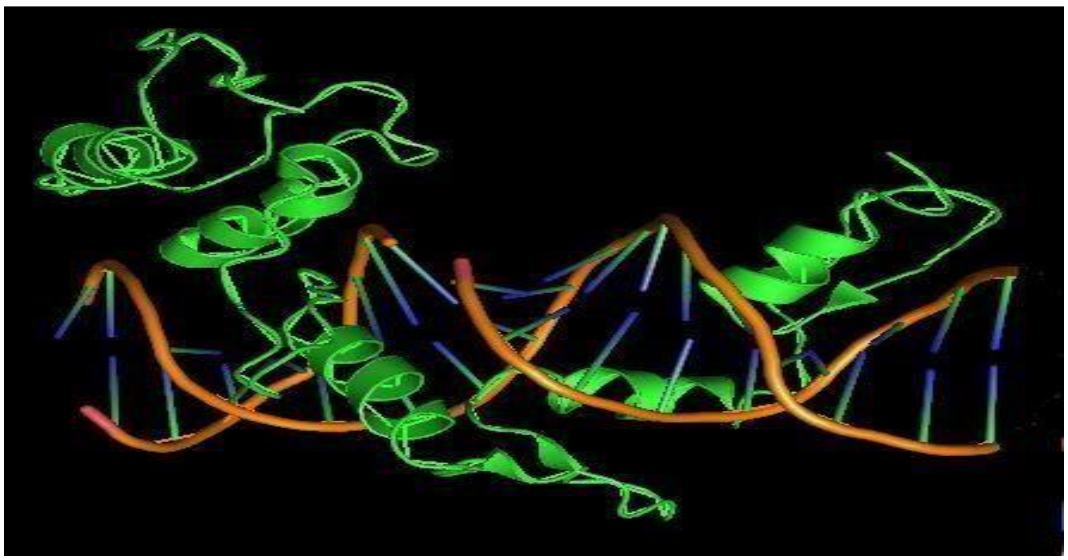


Figure 2. A typical Zinc finger motif

I.1.3.Coiled-coil motif: This motif is formed by two or more alpha helices winding around each other to form a supercoil. The coiled-coil motif, **Figure 3**.is commonly found in proteins involved in cell signaling and structural proteins such as myosin. Examples of proteins containing the coiled-coil motif include the transcription factor c-Jun and the protein keratin (Lupas AN, Gruber M 2005).

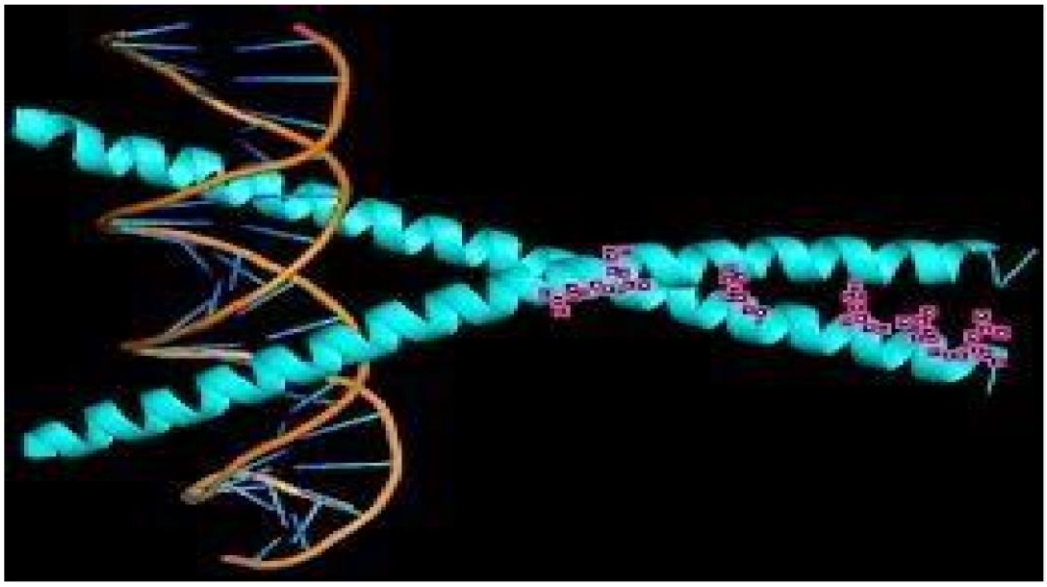


Figure 3. A typical Coiled-coil motif.

I.2. Protein Structural Motifs:

Structural motifs are recurring three-dimensional arrangements of amino acid residues or nucleotides that play important roles in the folding, stability, and function of proteins and nucleic acids.

There are three types of structural motifs:

I.2.1.α-Helices: α-helices are right-handed coils of amino acid residues stabilized by intramolecular hydrogen bonds, **Figure 4.** They are one of the most common structural motifs found usually associated with the secondary-structure level of proteins and are important in protein-protein interactions, DNA binding, and membrane-spanning domains (Richardson, J. S. 1981).

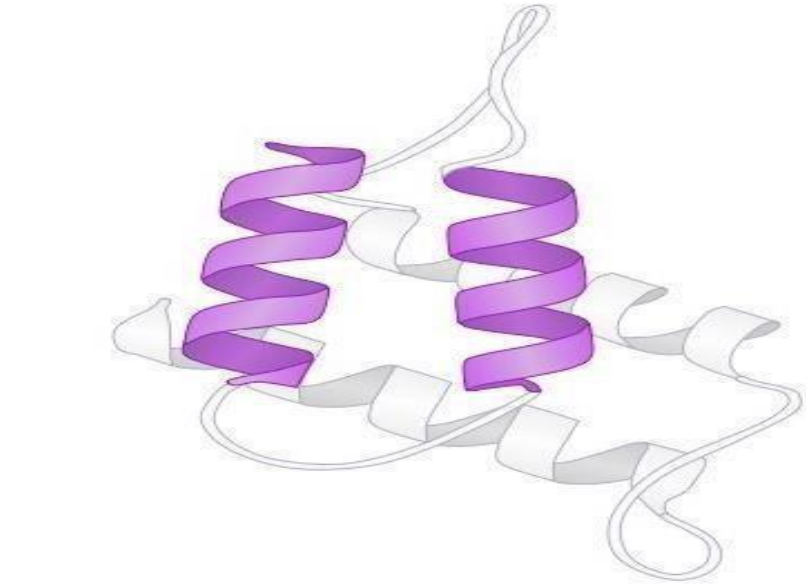


Figure 4. Secondary structure of alpha-helix

I.2.2.β-Sheets: β-sheets are planar arrangements of amino acid residues stabilized by hydrogen bonds between adjacent strands, **Figure 5.** They are important in protein-protein interactions, enzymatic catalysis, and as part of the core of many proteins (Janin, J., 1997).

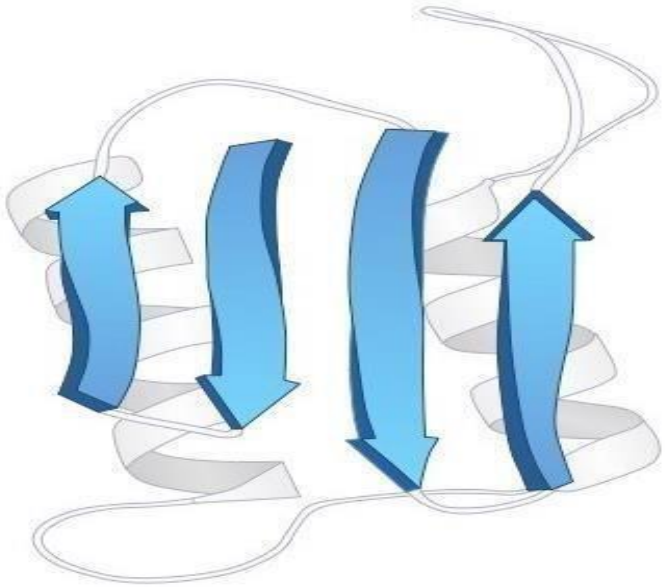


Figure 5. Secondary structure of protein Beta-sheets

I.2.3.Loops: Loops are flexible regions of a protein that connect the secondary structural elements (helices and sheets), **Figure 6**. They often form functional sites, including active sites for enzymatic catalysis and ligand-binding sites (Chothia, C. & Lesk, A. M., 1986).

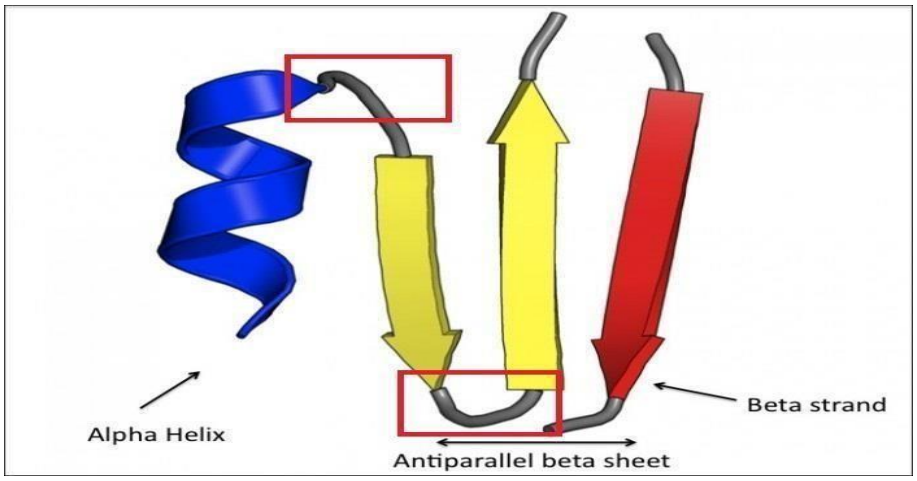


Figure 6. The loops regions between α-Helices and β-Sheets

3. Amino Acids :

Amino acids are the basic building blocks of proteins, and they serve as the nitrogenous backbones for compounds like neurotransmitters and hormones. In chemistry, an amino acid is an organic compound that contains both an amino (-NH2) and carboxylic acid (-COOH) functional group, hence the name amino acid, **Figure 7**.

Proteins are long chains or polymers of a specific type of amino acid known as an alphaamino acid. Alpha-amino acids are unique because the amino and carboxylic acid functional groups are separated by only one carbon atom, which is usually a chiral carbon. In this article, we will solely focus on the alpha-amino acids that make up proteins.(LaPelusa A, Kaushik R. StatPearls, 2022).

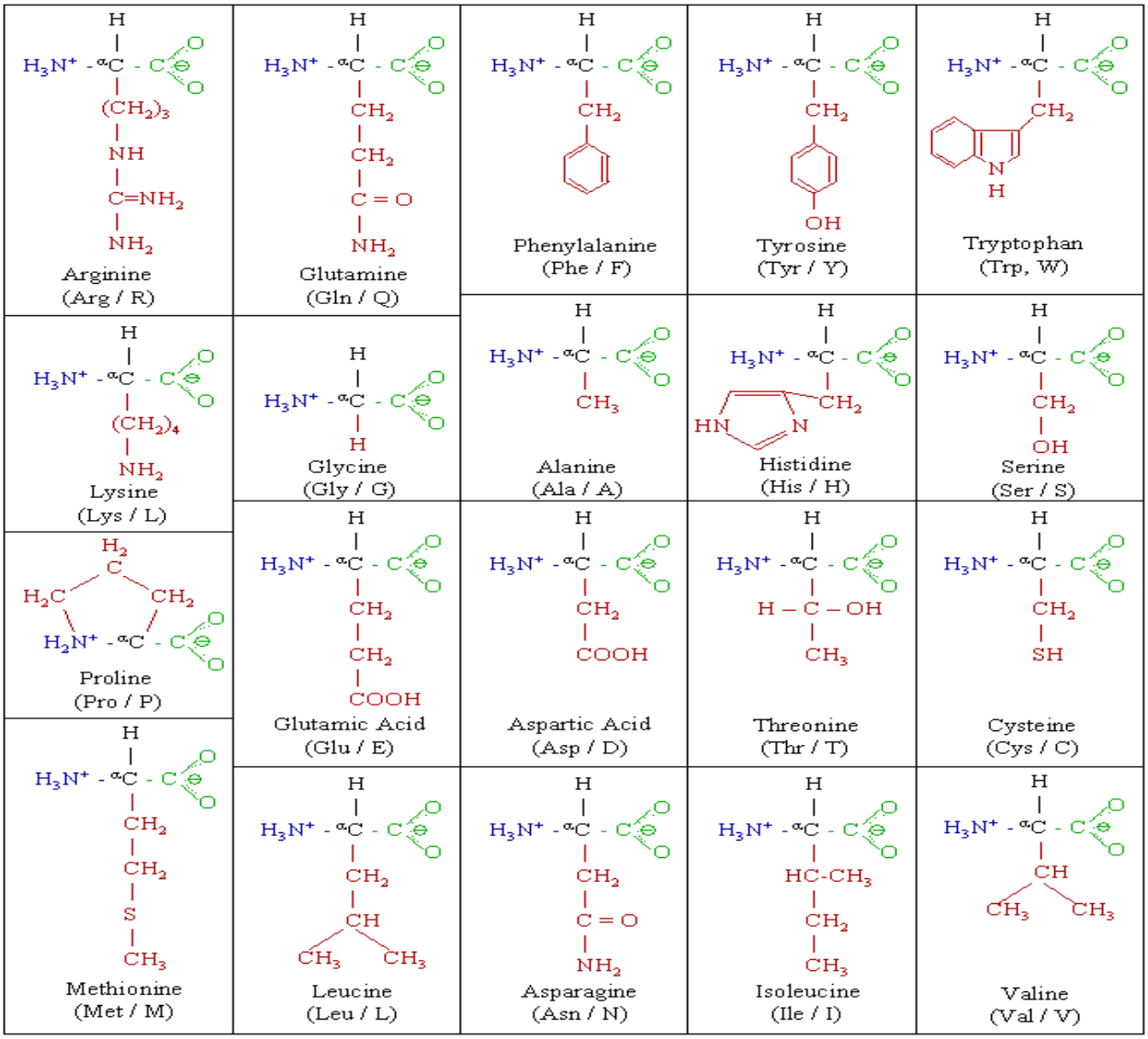


Figure 7. List of the 20 commonly occurring standard amino acids along with their single-letter and three-letter abbreviations. The main difference among amino acids lies in their side chains. The side chains canbe categorized into different groups based on their properties, such as hydrophobic,hydrophilic, acidic, basic, aromatic, etc.

1. The Porphyrins :

Porphyrins are a class of organic compounds characterized by a cyclic tetrapyrrole structure, **Figure 8**. These compounds consist of four pyrrole rings interconnected by methine bridges, forming a larger macrocyclic ring (Wiley-VCH. 2011). Porphyrins are notable for their ability to coordinate with metal ions, such as iron, magnesium, or zinc, at the center of the ring.

They play crucial roles in various biological processes, including oxygen transport (as in heme) and photosynthesis (as in chlorophyll).(Smith, K. M., & Ito, S. (Eds.). 2017).

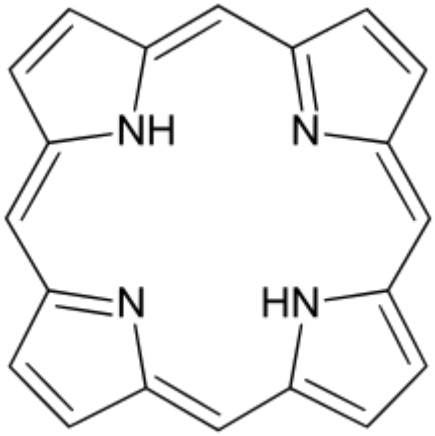


Figure 8. Porphine ring the nucleus of the porphyrin ring system

II. 2. The Porphyrins classes :

This protein family the porphyrins is divided into a several classes that is contain a proteins each one of theme bonded with a specific ligand : HEM, CLA, CHL, CL7, B12, HEC, SIR, F6C. which are they Related to a deferent biological functions.The deferent Porphyrins classes.

II. 2.1. Photosystem I (PSI) :

Photosystem I (PSI), also known as P700, is one of the key players in the process of oxygenic photosynthesis. This large membrane protein complex utilizes light energy to transfer electrons from the lumenal electron carriers plastocyanin or cytochrome c6 across the photosynthetic membrane to the stromal/cytosolic electron carriers ferredoxin or flavodoxin. The resulting proton gradient is used for adenosine triphosphate (ATP) production by the ATP synthase, while the electrons end up in carbon fixation. With a molecular weight of 1 million Da, trimeric cyanobacterial PSI is one of the largest membrane protein complexes with known structure. About one-third of its molecular weight comes from cofactors, mainly chlorophylls, and also carotenoids, phylloqinones, and iron–sulfur clusters.(I. Grotjohann, P. Fromme.2013).

This class contain the light harvesting protein.

II. 2.1.1. Light Harvesting Protein :

Light harvesting proteins are specialized membrane-bound proteins found in photosynthetic organisms, including plants, algae, and certain bacteria, **Figure 9**.These proteins play a crucial role in the absorption and transfer of light energy during photosynthesis.

Light harvesting proteins are responsible for capturing photons and transferring the absorbed energy to the reaction centers, where the conversion of light energy into chemical energy takes place. They contain pigment molecules, such as chlorophylls and carotenoids, which absorb light across a range of wavelengths. The energy captured by light harvesting proteins is utilized to drive the electron transfer processes that generate ATP and reduce NADP+ in photosynthetic organisms. (Blankenship, R. E. 2014).

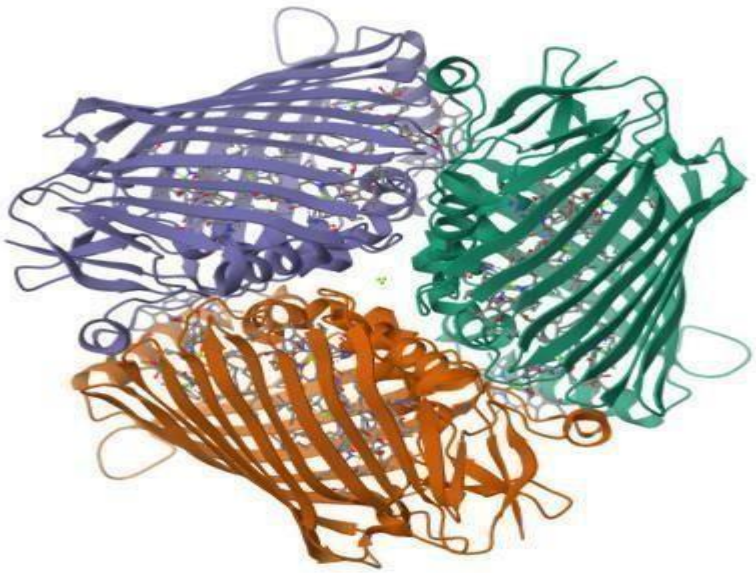


Figure 9.Structure of 3D model of Light harvesting protein.

From PDB : 7DKZ.

II. 2.2. Photosystem II (PSII) :

Photosystem II (PSII), also known as P680, is another pigment-protein complex involved in the initial stage of the light-dependent reactions of photosynthesis. It absorbs light with a maximum wavelength of around 680 nanometers. PSII captures photons and uses their energy to oxidize water molecules, releasing oxygen gas, protons (H+), and electrons. These highenergy electrons are then transferred through an electron transport chain to PSI, ultimately leading to the generation of ATP through chemiosmosis. PSII replenishes its electron supply by splitting water molecules, a process known as photolysis, and is crucial for sustaining the electron flow in the photosynthetic system.(Hill, R., & Bendall, D. S. 2014). This class contain the PSII with PSB27 complex:

II. 2.2.1. PSII with PSB27, PSB28, and PSB34 complex :

The association of PSII with Psb27, Psb28, and Psb34 is crucial for the proper functioning and adaptation of PSII to varying environmental conditions, ensuring efficient light capture, photoprotection, and optimal photosynthetic performance, **Figure 10**.

Psb27 is involved in the assembly and stabilization of PSII complexes, contributing to their proper formation and function. It assists in the incorporation of PSII subunits and cofactors into the mature complex, ensuring its structural integrity.

Psb28 is known to participate in the repair and reassembly of damaged PSII complexes, especially under stressful conditions such as high light or oxidative stress. It aids in the replacement of damaged subunits, contributing to the recovery of PSII functionality.

Psb34 is involved in the regulation of PSII activity, potentially acting as a photoprotective mechanism. It helps modulate the balance between light capture and photoprotection by regulating the flow of excitation energy within the PSII antenna system. (Järvi, S., Suorsa, M., & Aro, E. M. 2015).

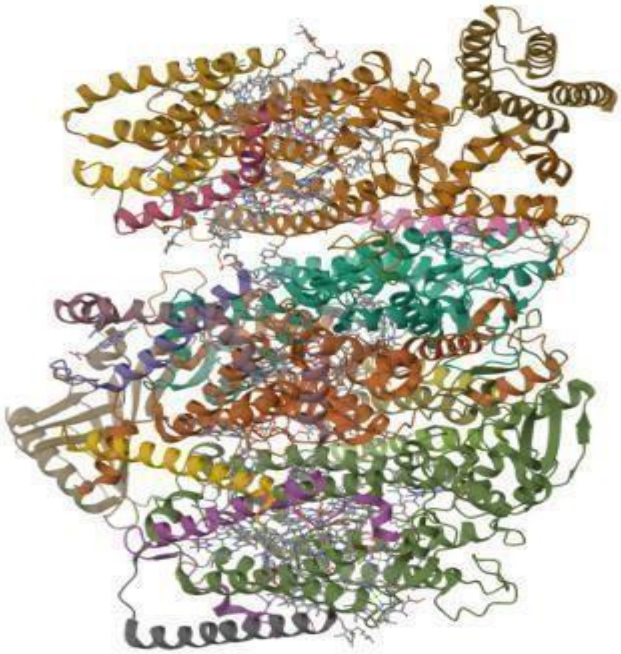


Figure 10.Structure of 3D model of The Photosystem II (PSII) core protein.

From PDB : 7NHP.

II. 2.2.2. Photosystem II CORE :

The Photosystem II (PSII) core refers to the central protein complex in the photosynthetic apparatus responsible for capturing light energy and driving the initial steps of photosynthesis. It consists of the core subunits that form the reaction center and the associated electron transfer components of PSII, **Figure 11**.

The PSII core is composed of a core reaction center, where light energy is absorbed and used to initiate electron transfer, and several peripheral antenna proteins that assist in capturing light and transferring the energy to the reaction center. The core reaction center includes key subunits such as D1, D2, CP43, and CP47, which house the

chlorophyll and other cofactors involved in the primary photochemical reactions.(Nelson, N., & Yocum, C. F. 2006).

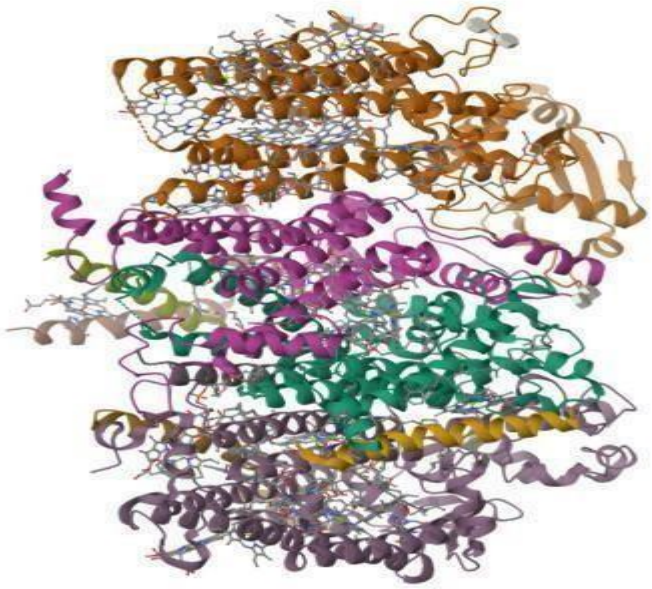


Figure 11.Structure of 3D model of The Photosystem II (PSII) core protein.

From PDB : 7SA3.

II. 2.3. oxygen transport :

The oxygen transport system refers to the biological mechanisms responsible for carrying oxygen from the lungs or respiratory organs to the body's tissues and cells.

It involves the circulation of oxygenated blood through the cardiovascular system, facilitated by the binding of oxygen to hemoglobin molecules in red blood cells.(Lehninger, A. L., Nelson, D. L., & Cox, M. M. 2008).

This class contain the hemoglobin protein :

II. 2.3.1. Hemoglobin :

Hemoglobin is a globular protein found in red blood cells that is responsible for transporting oxygen from the lungs to the body's tissues, **Figure 12**.Hemoglobin is composed of four subunits, each containing a heme group that binds to oxygen. The protein undergoes a conformational change upon binding to oxygen, allowing it to release the oxygen when it reaches its destination.(Perutz, M. F. 1989).

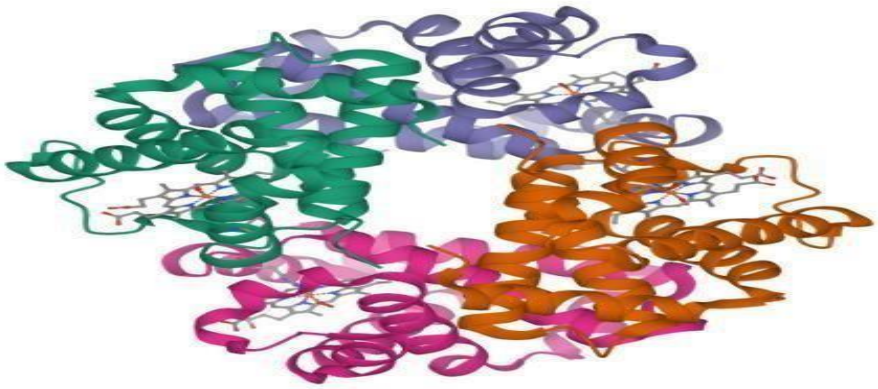


Figure 12.Structure of Oxy T State Haemoglobin-Oxygen bound at all four haems.

From PDB: 1GZX.

II. 2.4. oxygen storage :

Oxygen storage refers to the processes by which oxygen is stored within the body for later use when the demand for oxygen exceeds the immediate supply. It involves the binding of oxygen to various molecules and structures within tissues, such as myoglobin in muscle cells or hemocyanin in certain invertebrates, allowing for the storage and release of oxygen as needed. (Lehninger, A. L., Nelson, D. L., & Cox, M. M. 2008).

This class contain the myoglobin and the cytoglobin proteins :

II. 2.4.1.Myoglobin:

Myoglobin is a globular protein found in muscle cells that is responsible for storing and releasing oxygen as needed during muscle contraction. Myoglobin is composed of a single subunit that contains a heme group, **Figure 13**.allowing it to bind to oxygen. (Kendrew, J. C., & Dickerson, R. E. 1958).

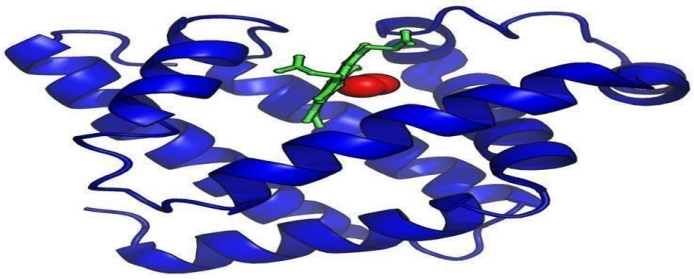


Figure 13.Cartoon representation of oxymyoglobin isolated from the sperm whale.

From PDB: 1MBO.

II. 2.4.2. Cytoglobin :

Cytoglobin (Cygb) is a globin protein that is expressed in various tissues,including the brain, heart, and lungs. It is thought to play a role in oxygen transport and storage, as well as in cytoprotection against oxidative stress, **Figure 14**.Cygb has been implicated in a wide range of physiological and pathological processes, including cancer, cardiovascular disease, and neurodegeneration.(Trent JT, Hargrove MS 2002).

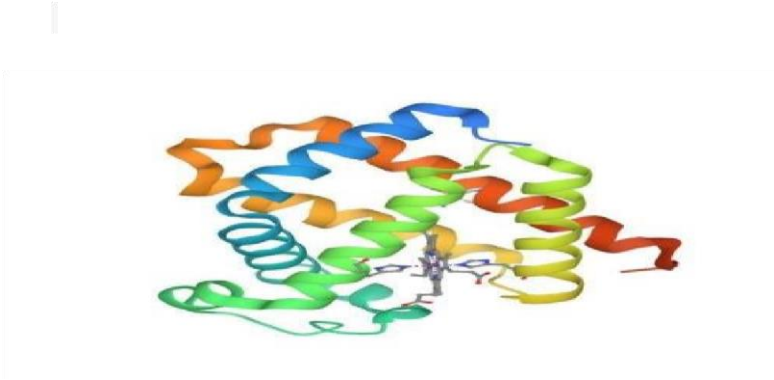


Figure 14.structure of 3D model of cytoglobin protein

From PDB: 2DC3 .

II. 2.5. Electron transport :

Electron transport refers to the sequential transfer of electrons through a series of proteins and molecules in a membrane-bound system. It plays a vital role incellular respiration and photosynthesis, facilitating the production of energy-rich molecules such as ATP. During electron transport, electrons are passed from oneelectron carrier to another, creating a flow of electrons that drives the synthesis of ATP and other metabolic processes.(Berg, J. M., Tymoczko, J. L., & Gatto, G. J. 2015).

This class contain the cytochrome P450 and b5:

II. 2.5.1. Cytochrome P450 :

Cytochrome P450 is a family of enzymes involved in the metabolism of a wide variety of compounds, including drugs and toxins, **Figure 15**. These enzymes contain a heme group that allows them to oxidize the compounds they metabolize.(Ortiz de Montellano, P. R. 2015).

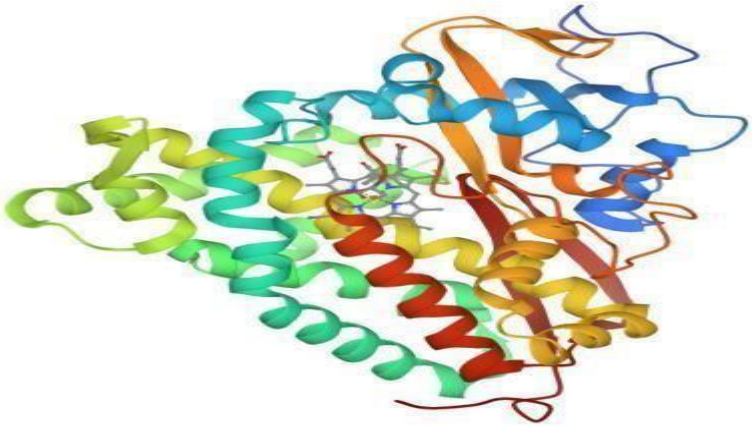


Figure 15.structure of Cytochrome P450 with heme

From PDB: 2CPP.

II. 2.5.2. Cytochrome b5 :

Cytochrome b5 is a small protein that is involved in electron transfer reactions in a variety of metabolic pathways, **Figure 16**.It contains a heme group that allows it to transfer electrons between enzymes. (Chiancone, E., & Ceci, P.2010).

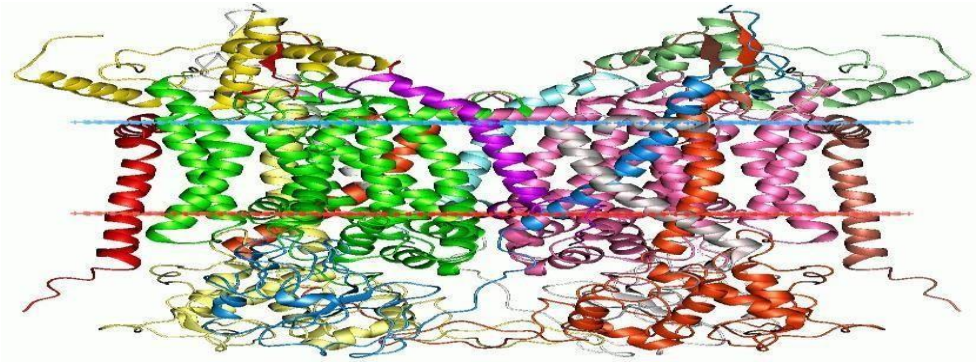


Figure 16.structure of Cytochrome b5 with heme
From PDB : 1CYO.

II. 2.5.3. Cytochrome C :

Cytochrome *c* is a heme protein that is present in and can easily be isolated from mitochondria of all eukaryotic organisms, **Figure 17**.The amino acid sequence of the protein moiety was among the first sequences that could be elucidated. This was the starting point for comparative studies about sequence variations found in cytochrome *c* from a wide range of species. A phylogenetic tree constructed on the basis of this information was found to be biologically

significant and became exemplary for subsequent studies on molecular evolution. The function of cytochrome *c* in the respiratory chain as an electron carrier is well established. More recently, an additional role of cytochrome *c* was discovered: its release from mitochondria into the cytosol triggers apoptosis – the programmed cell death.
(H. Tuppy, G. Kreil, 2013).

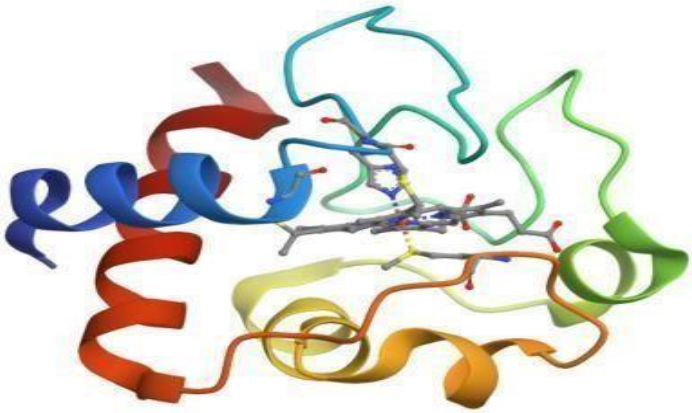


Figure 17.structure of Cytochrome C with heme

From PDB: 1HRC

II. 2.6. Peroxidases:

Peroxidases are a family of enzymes that catalyze the oxidation of various substrates using hydrogen peroxide.They contain a heme group that serves as a cofactor in the oxidation reaction.(Dunford, H. B. 1999). this class contain the horseradish peroxidase protein :

II. 2.6.1. Horseradish Peroxidase :

Horseradish peroxidase (HRP) is an enzyme found in the roots of horseradish plants (Armoracia rusticana), **Figure 18**.It belongs to the peroxidase family of enzymes and is widely used in various biochemical and molecular biology applications. HRP catalyzes the oxidation of a wide range of substrates by utilizing hydrogen peroxide (H2O2) as a cosubstrate. It plays a crucial role in many laboratory techniques, including enzyme-linked immunosorbent assays (ELISA), immunohistochemistry, and Western blotting, where it is used as a reporter enzyme to detect the presence of specific target molecules.(Welch, G. R. 1980).

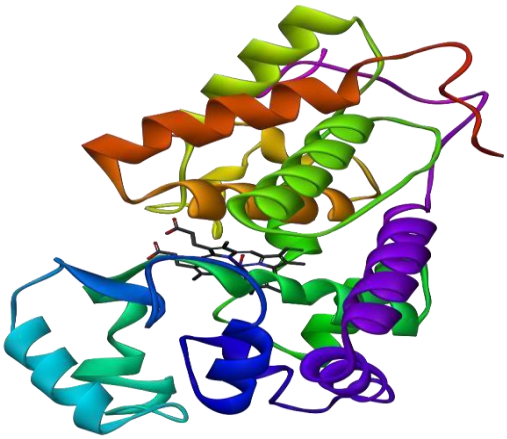


Figure 18.structure of 3D model of the horseradish peroxidase protein

From PDB: 1ARV.

II. 2.6.2. Lignin peroxidase :

Lignin peroxidase is an enzyme belonging to the class of peroxidases that is involved in the degradation of lignin, a complex polymer found in plant cell walls, **Figure 19**.It plays a crucial role in the breakdown of lignin by catalyzing the oxidative cleavage of lignin's aromatic subunits. Lignin peroxidase is produced by certain fungi and bacteria, and its activity contributes to the recycling of lignin and the breakdown of plant biomass.(Hofrichter, M., Kellner, H., Pecyna, M. J., Ullrich, R., & Scheibner, K. 2020).

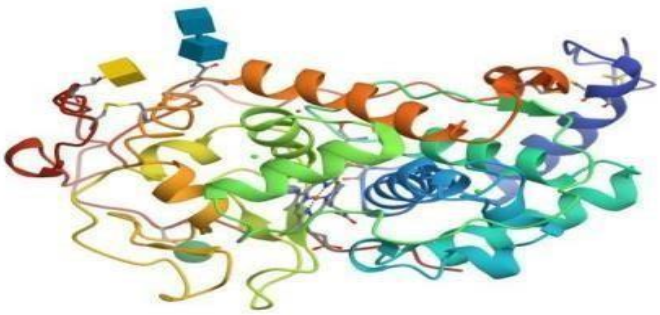


Figure 19.structure of 3D model of the Lignin peroxidase protein

From PDB: 1LLP.

II. 2.6.3. Chloroperoxidases:

Chloroperoxidases (CPO) are a family of enzymes that catalyze the oxidation of halide ions (e.g., chloride) by hydrogen peroxide (H2O2) to produce hypohalous acids, which have antimicrobial activity, **Figure 20**. CPOs are found in various organisms, including fungi and bacteria, and are involved in a wide range of physiological processes, including defense against pathogens and environmental stress. (Hofrichter M, Ullrich R, Pecyna MJ, Liers C, Lundell T 2010)

2010)

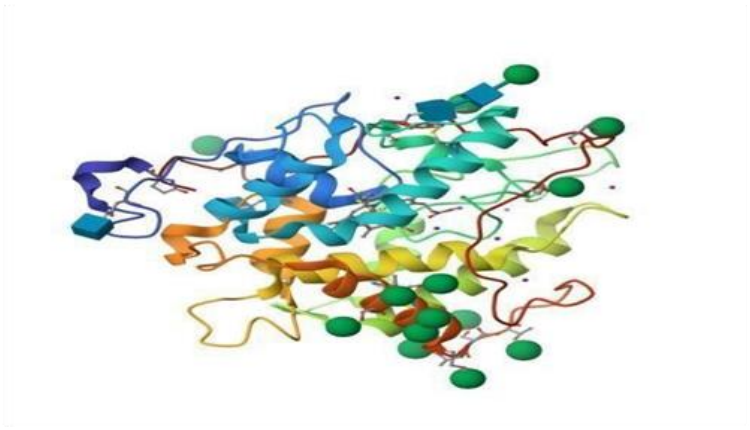


Figure 20.structure of 3D model of chloroperoxidases protein.

From PDB: 2CIW.

II. 2.7.Catalase :

Catalases are enzymes that catalyze the decomposition of hydrogen peroxide into water and oxygen. Catalases are found in almost all aerobic organisms and play a crucial role in protecting cells from oxidative damage caused by reactive oxygen species.

The reaction catalyzed by catalase is as follows: 2 H2O2 → 2 H2O + O2.
(Scandalios JG 2005).

This class contain the Human Erythrocyte catalase :

II. 2.7.1. Human Erythrocyte catalase :

Human erythrocyte catalase is an enzyme found in red blood cells that plays a vital role in protecting the body from oxidative damage, **Figure 21**.Catalase is responsible for the breakdown of hydrogen peroxide (H2O2) into water and molecular oxygen. By efficiently decomposing H2O2, catalase prevents the accumulation of this reactive oxygen species, which can cause cellular damage and oxidative stress. Human erythrocyte catalase serves as an essential antioxidant defense mechanism, contributing to the maintenance of cellular homeostasis and overall health. (Vanderver, A. L., Wolff, J., & Milstien, S. 2021).

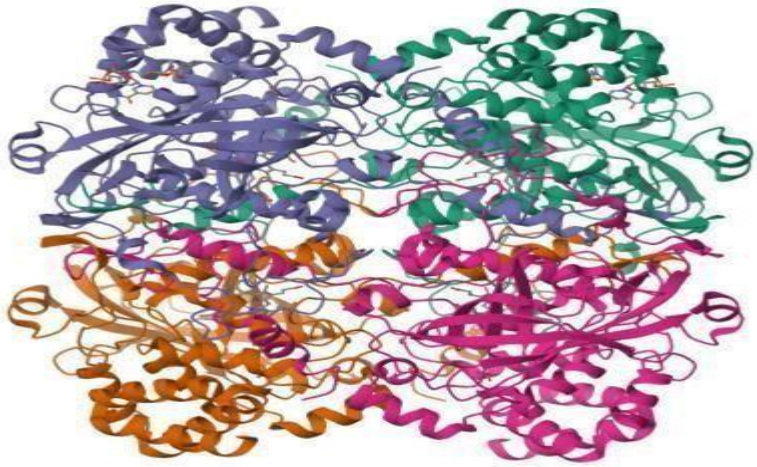


Figure 21.structure of 3D model of Human erythrocyte catalase protein

From PDB: 1DGF.

II. 2.8.Oxidoreductase :

Oxidoreductases are a class of enzymes that catalyze oxidation-reduction reactions by transferring electrons between substrates. These enzymes facilitate the transfer of electrons from a donor molecule (oxidation) to an acceptor molecule (reduction). They play a crucial role in various biological processes, including cellular respiration, photosynthesis, and metabolism. Oxidoreductases are involved in the transfer of electrons during redox reactions, contributing to the conversion of energy and the synthesis of important biomolecules.(Nelson, D. L., Cox, M. M., & Lehninger, A. L.2008).

This class contain the.

II. 2.8.1. Nitric Oxide Synthases:

Nitric oxide synthases (NOS) are a family of enzymes that catalyze the production of nitric oxide (NO) from the amino acid L-arginine, **Figure 22**.Nitric oxide is a signaling molecule involved in a wide range of physiological processes, including vasodilation, neurotransmission, and immune response. There are three isoforms of NOS: neuronal NOS (nNOS), endothelial NOS (eNOS), and inducible NOS (iNOS). Each isoform is expressed in different cell types and has distinct functions. (Alderton WK, Cooper CE , Knowles RG 2001) .

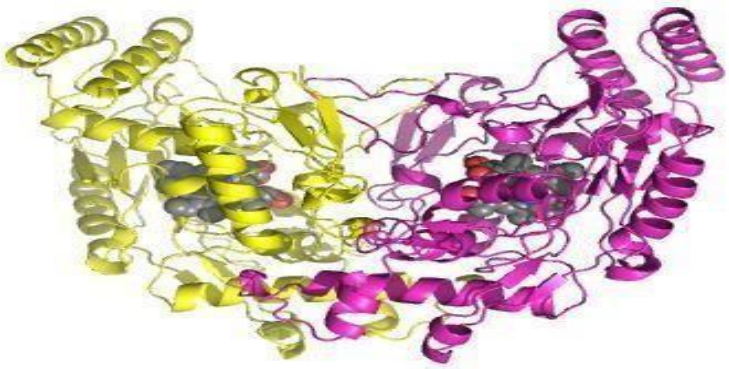


Figure 22.structure of 3D model of the Nitric oxide synthases protein

From PDB: 1D0C.

II. 2.8.2. Heme Oxygenases :

Heme oxygenases are enzymes that catalyze the breakdown of heme to produce biliverdin, carbon monoxide, and iron, **Figure 23**. This reaction is important in the recycling of heme and the regulation of cellular heme levels.(Tenhunen, R., Marver, H. S., & Schmid, R.1969).

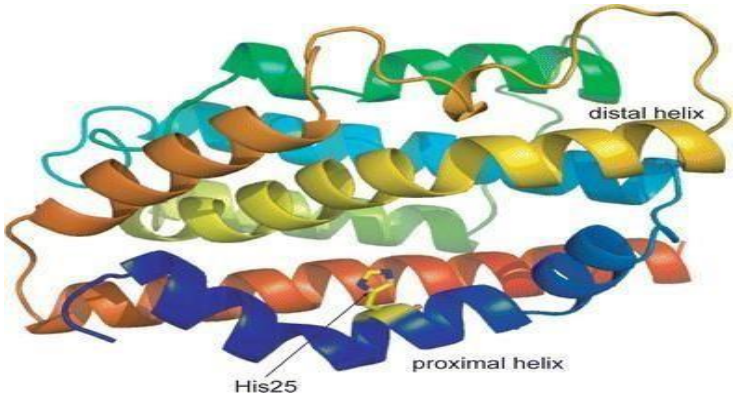


Figure 23.structure of 3D model of Heme Oxygenases protein

From PDB: 1IW0.

II. 2.8.3. Flavihemoglobin:

Flavihemoglobin is a type of bacterial hemoglobin that contains a flavin cofactor and is involved in the regulation of gene expression in response to changes in oxygen levels, **Figure 24**. It was first identified in the bacterium Alcaligenes eutrophus and has since been found in other bacteria. Flavihemoglobin is unique in that it contains both a heme group, which binds oxygen, and a flavin adenine dinucleotide (FAD) cofactor, which allows it to sense changes in oxygen levels and regulate the expression of genes involved in metabolism, respiration, and stress response.(Fernández - Justel, D. , Zbilut, J. P., & Roca, F. J.2018) .

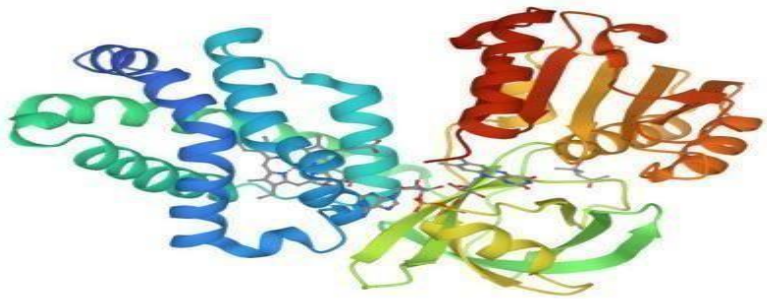


Figure 24.structure of 3D model of the flavihemoglobin From PDB: 6O0A.

II. 2.8.4. Chloroperoxidases:

Chloroperoxidases (CPO) are a family of enzymes that catalyze the oxidation of halide ions (e.g., chloride) by hydrogen peroxide (H2O2) to produce hypohalous acids, which have antimicrobial activity, **Figure 25**. CPOs are found in various organisms, including fungi and bacteria, and are involved in a wide range of physiological processes, including defense against pathogens and environmental stress.(Hofrichter M, Ullrich R, Pecyna MJ, Liers C, Lundell T 2010).

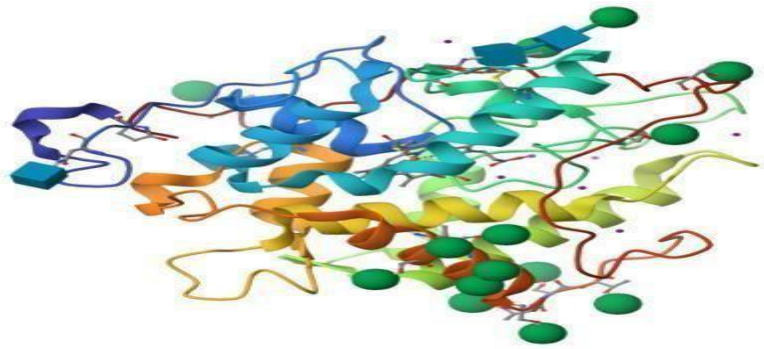


Figure 25.structure of 3D model of chloroperoxidases protein

From PDB: 2CIW.

II. 2.9.Oxygene binding :

Oxygen binding refers to the process by which oxygen molecules (O2) attach to specific molecules or proteins, enabling the transport and delivery of oxygen throughout the body. This binding typically involves the association of oxygen with proteins such as hemoglobin in red blood cells or myoglobin in muscle cells. (Lehninger et al., 2008).

This class contain the Neuroglobin :

II. 2.9.1. Neuroglobin :

Neuroglobin is a protein that is predominantly expressed in neurons and plays a role in the transport and storage of oxygen in the brain, **Figure 26**. It is a member of the globin family of proteins, which also includes hemoglobin and myoglobin. Neuroglobin is characterized by its ability to bind oxygen and other small molecules such as carbon monoxide and nitric oxide, and its expression is upregulated in response to hypoxia (low oxygen levels).(Khan, A. A., & Mao, X. O.2018).

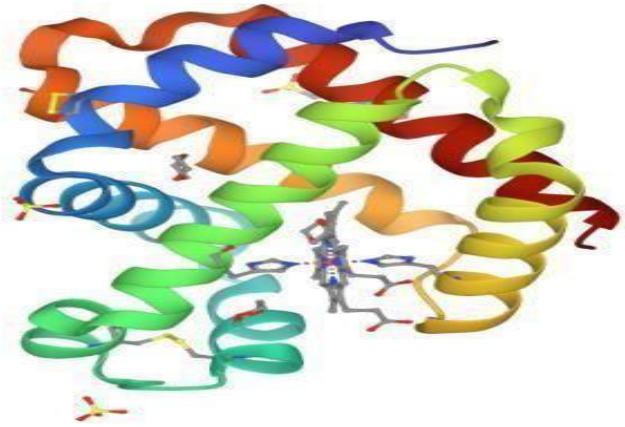


Figure 26.structure of the 3D model neuroglobin with heme.

From PDB: 7VQG.

II. 2.10.Isomerase :

An isomerase is a class of enzymes that catalyze the conversion of one isomer into another by facilitating the rearrangement of atoms within a molecule. Isomerases play a vital role in various biological processes, including metabolism, signal transduction, and biosynthesis. These enzymes enable the interconversion of isomeric forms of molecules without changing their overall molecular formula.(Nelson, D. L., Cox, M. M., & Lehninger, A. L.2008).

II. 2.10.1. Methylmalonyl-CoA mutase :

Methylmalonyl-CoA mutase is an enzyme that plays a crucial role in the metabolism of certain fatty acids, amino acids, and cholesterol, **Figure 27**.It catalyzes the isomerization of methylmalonyl-CoA to succinyl-CoA, a key step in the degradation of branched-chain amino acids and odd-chain fatty acids. This enzymatic activity is essential for maintaining the balance of metabolites and energy production within cells. Deficiencies or mutations in methylmalonylCoA mutase can lead to methylmalonic acidemia, a metabolic disorder characterized by the accumulation of methylmalonic acid and associated health complications.(Peters, H. L., Sweetman, L., & Nyhan, W. L. (2021).

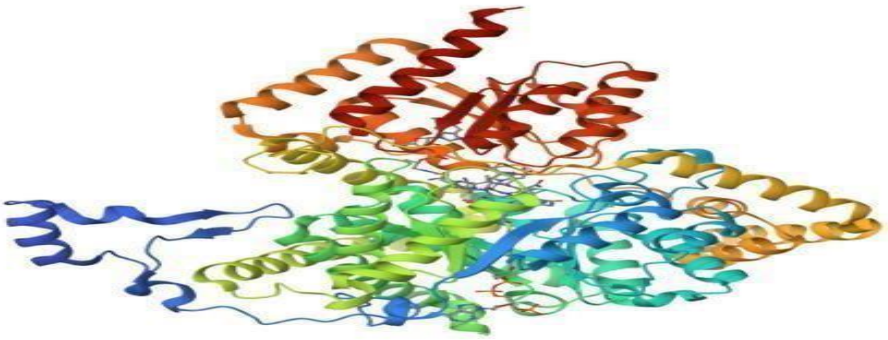


Figure 27.structure of the 3D model Methylmalonyl-CoA mutase

From PDB: 2XIQ.

II. 2.10.2. Beta-methylaspartate-glutamate mutase :

is a enzyme Glutamate mutase converts (*S*)-glutamate to (*S*,*S*)-3-methylaspartate, whereas 2methyleneglutarate mutase isomerizes 2-methyleneglutarate to (*R*)-3-methylitaconate, **Figure 28**.Both enzymes occur in amino acid fermenting clostridia , require coenzyme B12 (adenosylcobalamin) and catalyze a reversible, radical carbon skeleton rearrangement. The enzymes are assayed using the consecutive enzymes in their fermentation pathways: methylaspartase and methylitaconate isomerase, which produce the higher absorbing mesaconate and 2,3-dimethylmaleate, respectively. Although the quaternary structures of glutamate mutase and 2-methyleneglutarate mutase are different, the binding mode of coenzyme B_{12} and their catalytic mechanism appear to be very similar.(Wolfgang Buckel, ... Oskar Zelder, in Methods in Enzymology, 2022)

Figure 28.structure of the 3D model Beta-methylaspartate-glutamate mutase

From PDB: 6H9E.

II. 2.11.Lyase :

A lyase is a class of enzymes that catalyze the cleavage or formation of chemical bonds within a molecule, resulting in the formation of two separate products. Lyases facilitate this process without the addition or removal of water, making them distinct from hydrolases. These enzymes play a critical role in various biological pathways, such as metabolism and DNA repair, by breaking down or forming specific chemical bonds to generate new compounds.(Nelson, D. L., Cox, M. M., & Lehninger, A. L.2008).

This class contain the Sirohydrochlorin cobaltochelatase protein :

II. 2.11.1. Sirohydrochlorin cobaltochelatase :

Sirohydrochlorin cobaltochelatase is an enzyme that plays a crucial role in the biosynthesis of cobalamin (vitamin B12) in certain bacteria and archaea, **Figure 29**. It catalyzes the insertion of a cobalt ion into the sirohydrochlorin macrocycle, converting it into adenosylcobinamideGDP, an intermediate in the cobalamin synthesis pathway. This enzyme is responsible for the final step in the cobalamin biosynthesis pathway and is essential for the production of functional vitamin B12, which is involved in various important biological processes, including DNA synthesis, methionine synthesis, and methyl group transfer reactions. (Scott, A. I.1998).

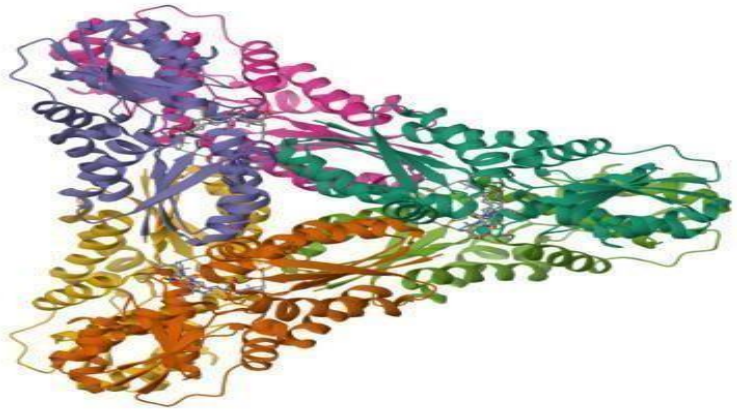


Figure 29.structure of the 3D model Sirohydrochlorin cobaltochelatase

From PDB: 2XWP.

III. The protein data bank(PDB) :

The Protein Data Bank (PDB) is a digital repository that provides a collection of threedimensional structures of biological macromolecules, including proteins, nucleic acids, and complex assemblies.The PDB is maintained by the Worldwide Protein Data Bank (wwPDB) organization and is a critical resource for researchers in structural biology, biochemistry, and molecular biology, **Figure 30**. The PDB stores information about the atomic coordinates of each molecule, as well as other experimental data, such as diffraction data, nuclear magnetic resonance data, and electron microscopy data. These structures are determined by experimental techniques such as X-ray crystallography, NMR spectroscopy, and cryoelectron microscopy., the websites of this member organisations (PDBe, PDBj, [1] and RCSB[1]).

The first link returned, which is : **http://www.rcsb.org/pdb/home**

.

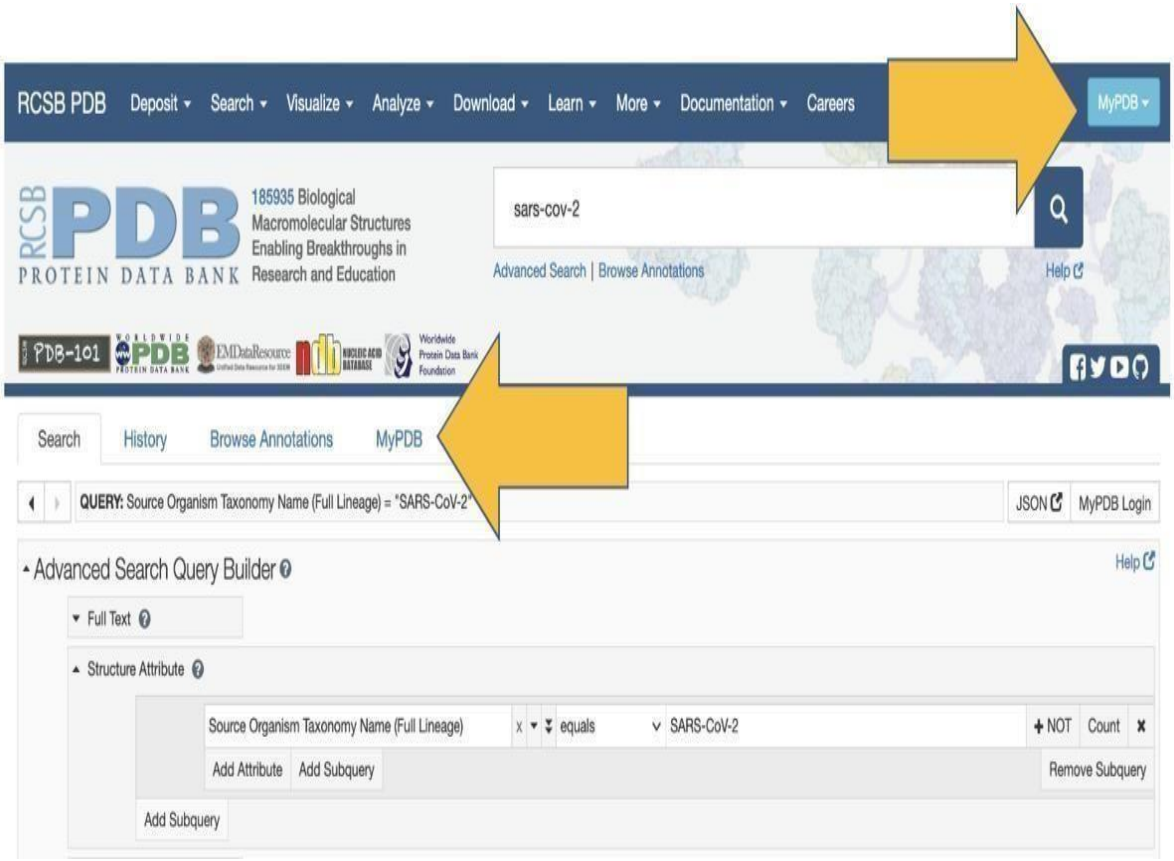


Figure 30.PDB – Protein Data Bank main page.

III.1.File format :

The Protein Data Bank (PDB) file format is a standard file format for storing the atomic coordinates and other information of biomolecular structures. PDB files use a plain text format and are identified by the file extension ".pdb". Each PDB file contains information about a single macromolecule, including the type of macromolecule, the sequence of amino acids or nucleotides, and the three-dimensional coordinates of each atom in the molecule. An example of a PDB file can be accessed through its unique identifier, such as for the protein myoglobin (PDB ID: 1mbn), the URL is: **https://www.rcsb.org/structure/1MBN.**(Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., ... & Bourne, P. E. 2000).

III.2.Resolution :

Resolution is a measure of the quality of the three-dimensional structure of a macromolecule, typically determined by X-ray crystallography. It is defined as the smallest distance between two atoms that can be reliably distinguished in the electron density map, and is usually reported in units of angstroms (Å).(Karplus, P. A., & Diederichs, K. 2012).

III. 3.Refinement factor (R-factor) :

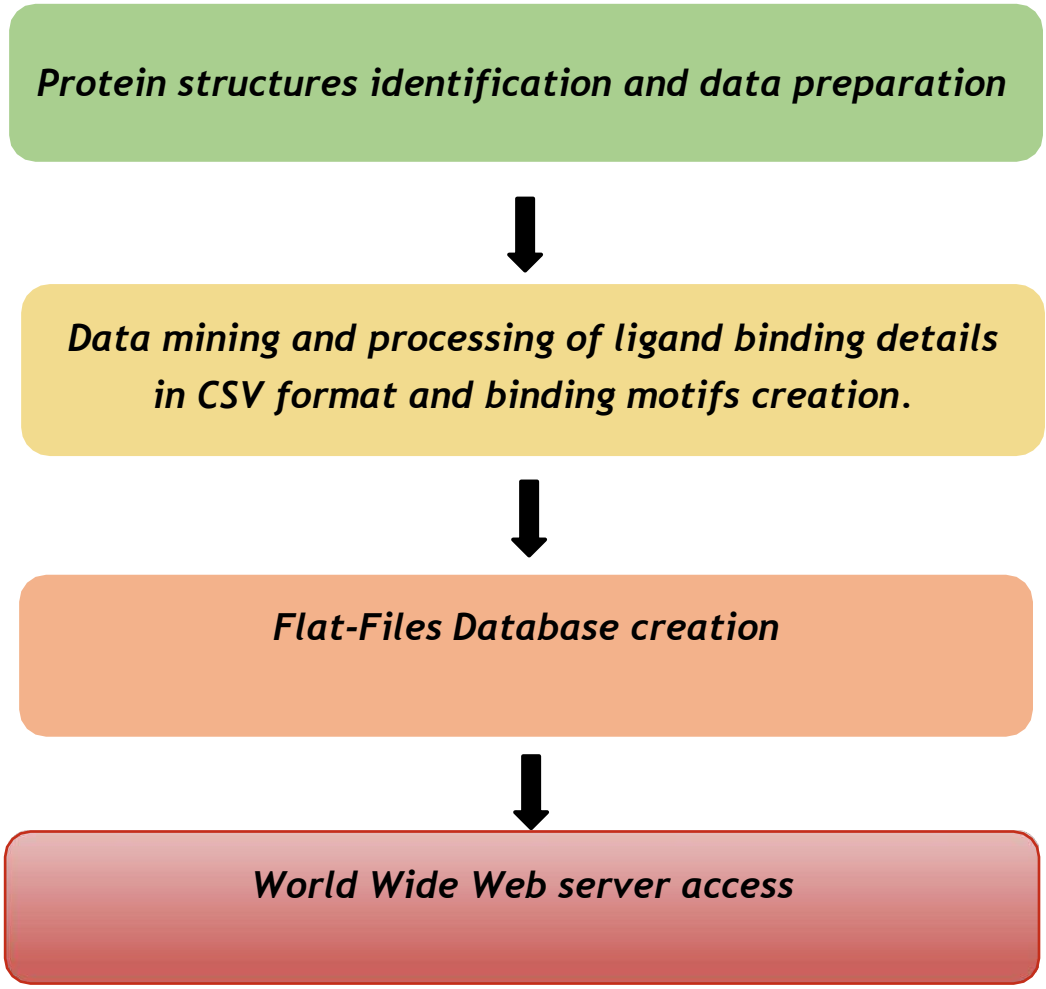
The refinement factor is a measure of the quality of the structural information contained in a protein structure determined using X-ray crystallography. It provides a measure of how well the observed diffraction data agrees with the calculated structure, and is typically used as an indicator of the quality of the final refined structure.(Rachedi A, 2013).

Chapter II

Introduction:

In order to realize the structural study of this project, the binding structural motif may be found in the Porphyrin proteins, structural bioinformatics methods involving database creation and programming were employed .

The required steps that we followed to achieve the goals from this study are explained in the following:



I. Protein structures identification and Data Preparation:

As we explained in the previous chapter, the PDB is the database which provides Structural data for the protein, we can also use the PDB to study the function of Proteins.

The PDB assigns a unique identifier to each entry, called a PDB ID, which is used to reference the structure in scientific literature and other databases. The PDB ID consists of four characters, and it is assigned when a new structure is deposited in the database.

Data preparation for Porphyrin motifs structures may involve extracting relevant information such as the protein name, PDB ID, amino acid sequence, and Porphyrin coordination geometry. This data can be extracted from the PDB or wwPDB using their online search tools or through programmatic access using web APIs.

I.1. Protein structures (PDB entries):

The PDB entries used in this project amount to 12 structures each with its own PDB id, we choose the PDB id with a high resolution.

In addition to providing a database of structures, the PDB also plays a key role in the identification and entry of Porphyrin-containing proteins. Proteincrystallographers who determine the structure of a protein containing Porphyrin are required to deposit their data in the PDB, which then assigns a unique identifier, or PDB id, to the entry.

All of the studied structures have been found to have determined by the X-ray Crystallography, This makes it easy for researchers to access and compare structural data for porphyrin-containing proteins from different sources. The Table 1., below, represents the list of porphyrin proteins class and their protein name, PDB id. Resolution and R-factor which reflect the quality of the structure under study as indicated in the table.

I.1.1. PDB entries list of the porphyrin protein family:

Class	Enzyme Class/Gene Name	Prt.name	Source organism	PDB ID	Title of PDB	Method	esolution (angstro	R-Value	Metal Ions	Ligands	Lig_Name
Photosystem I (PSI)	fmoA	LIGHT HARVESTING PROTEIN	Chlorobaculum tepidum	3BSD	EVOLUTION OF PHOTOSYSTEM I - FROM SYMMETRY THROUGH PSEUDO-SYMMETRY TO ASYMMETRY.	X-ray diffraction	2.3	17.8	Mg2+	BCL	BACTERIO-CHLOROPHYLL A
Photosystem I (PSI)	1.97.1.12	LIGHT HARVESTING PROTEIN	Pisum sativum	7DKZ	Human Erythrocyte STRUCTURE OF PLANT PHOTOSYSTEM I-LIGHT HARVESTING C SUPERCOMPLEX AT 2.4 ANGSTROM RESOLUTION.	X-ray diffraction	2.39	19.1	Mg2+	CLA/CHL	CHLOROPHYLL A / B
Photosystem II (PSII)	1.10.3.9	PSII WITH PSB27; PSB28; AND PSB34	Thermosynechococcus elongatus bp-1	7NHP	STRUCTURE OF PSII-I (PSII WITH PSB27; PSB28; PSB34)	ELECTRON MICROSCOPY	2.72	N/A	Mg2+	CLA	CHLOROPHYLL A
Photosystem II (PSII)	1.10.3.9	PHOTOSYSTEM II CORE	Synechococcus sp. pcc 7335	7SA3	STRUCTURE OF A MONOMERIC PHOTOSYTEM II CORE COMPLEX FROM A CYANOBACTERIUM ACCLIMATED TO FAR-RED LIGHT	ELECTRON MICROSCOPY	2.25	N/A	Mg2+	CLA/CL7/F6C	CHLOROPHYLL A / D / F
oxgen transport	HBA1; HBA2	Hemoglobin	Homo sapiens	1GZX	Oxy T State Haemoglobin - Oxygen bound at all four haems	X-ray diffraction	2.1	19.5	Fe2+	HEM	PROTOPORPHYRIN IX CONTAINING FE
oxgen storage	MB	Myoglobin	Physeter catodon	1MBO	Stucture and refinement of oxymyglobin	X-ray diffraction	1.6	24.3	Fe2+	HEM	PROTOPORPHYRIN IX CONTAINING FE
oxgen storage	CYGB	Cytoglobin	Homo sapiens	2DC3	CRYSTAL STRUCTURE OF HUMAN CYTOGLOBIN AT 1.68 ANGSTROMS RESOLUTION	X-ray diffraction	1.68	14.2	Fe2+	HEM	PROTOPORPHYRIN IX CONTAINING FE
Electron transport	CYCS	Cytochrome C	Equus caballus	1HRC	HIGH-RESOLUTION THREE-DIMENSIONAL STRUCTURE OF HORSE HEART CYTOCHROME C		1.9	17.9	Fe2+	HEC	HEME C
Electron transport	1.14.15.1	Cytochrome P450-cam / Camphor 5-monooxygenase	Pseudomonas putida	2CPP	Crystal structure of cytochrome P450-cam with camphor bound	X-ray diffraction	1.63	19	Fe3+	HEM	PROTOPORPHYRIN IX CONTAINING FE

Electron transport	1.14.15.1	Cytochrome P450-cam / Camphor 5-monooxygenase	Pseudomonas putida	2CPP	Crystal structure of cytochrome P450-cam with camphor bound	X-ray diffraction	1.63	19	Fe3+	HEM	PROTOPORPHYRIN IX CONTAINING FE
Electron transport	CYB5A	Cytochrome B5	Bos taurus	1CYO	BOVINE CYTOCHROME B(5)	X-ray diffraction	1.5	16	Fe2+	HEM	PROTOPORPHYRIN IX CONTAINING FE
Peroxidase	1.11.1.7	Horseradish peroxidase C1A	Arthromyces ramosus	1ARV	CRYSTAL STRUCTURES OF CYANIDE-AND TRIIODIDE-BOUNDFORMS OF ARTHROMYCES RAMOSUS PEROXIDASE AT DIFFERENT PH VALUES	X-ray diffraction	1.6	0.178	Fe2+	HEM	PROTOPORPHYRIN IX CONTAINING FE
Peroxidase	1.11.1.14	Lignin peroxidase	Phanerochaete chrysosporium	1LLP	THE CRYSTAL STRUCTURE OF LIGNIN PEROXIDASE AT 1.70 RESOLUTION REVEALS A HYDROXY GROUP ON THE CBETA OF TRYPTOPHAN 171: A NOVEL RADICAL SITE FORMED	X-ray diffraction	1.7	0.162	Fe3+	HEM	PROTOPORPHYRIN IX CONTAINING FE
Peroxidase	1.11.1.10	Chloroperoxidase	Caldariomyces fumago	2CIW	Chloroperoxidase iodide complex	X-ray diffraction	1.15	15	Fe2+	HEM	PROTOPORPHYRIN IX CONTAINING FE
Catalase	1.11.1.6	Human Erythrocyte catalase	Homo sapiens	1DGF	ACTIVE AND INHIBITED HUMAN CATALASE STRUCTURES: LIGAND AND NADPH BINDING AND CATALYTIC MECHANISM.	X-ray diffraction	1.5	17.2	Fe2+	HEM	PROTOPORPHYRIN IX CONTAINING FE
oxidoreductase	1.14.13.39	Nitric oxide synthas	Bos taurus	1D0C	BOVINE ENDOTHELIAL NITRIC OXIDE SYNTHASE HEME DOMAIN COMPLEXEDWITH 3-BROMO-7-NITROINDAZOLE	X-ray diffraction	1.65	25.9	Fe2+	HEM	PROTOPORPHYRIN IX CONTAINING FE
oxidoreductase	1.14.14.18	Heme oxygenase	Corynebacterium diphtheria	1IW0	Crystal structure of a heme Oxygenase (hmuo) from Corynebacterium diphtheriae Complexed with heme in The ferric state	X-ray diffraction	1.4	16.8	Fe2+	HEM	PROTOPORPHYRIN IX CONTAINING FE
oxidoreductase	1.14.12.17	Flavihemoglobin	Malassezia yamatoensis	6O0A	fCRYSTAL STRUCTURE OF FLAVOHEMOGLOBIN FROM MALASSEZIA YAMATOE BOUND FAD AND HEMEDETERMINED BY IRON SAD PHASING	X-ray diffraction	1.7	16.9	Fe2+	HEM	PROTOPORPHYRIN IX CONTAINING FE

oxidoreductase	1.14.14.18	Heme oxygenase	Corynebacterium diphtheria	1IW0	Crystal structure of a heme Oxygenase (hmuo) from Corynebacterium diphtheriae Complexed with heme in The ferric state	X-ray diffraction	1.4	16.8	Fe2+	HEM	PROTOPORPHYRIN IX CONTAINING FE
oxidoreductase	1.14.12.17	Flavihemoglobin	Malassezia yamatoensis	6O0A	fCRYSTAL STRUCTURE OF FLAVOHEMOGLOBIN FROM MALASSEZIA YAMATOE BOUND FAD AND HEMEDETERMINED BY IRON SAD PHASING	X-ray diffraction	1.7	16.9	Fe2+	HEM	PROTOPORPHYRIN IX CONTAINING FE
Oxygene binding	NGB (https://www.ncbi.nlm.nih.gov/gene/58157)	Neuroglobin	Homo sapiens	7VQG	THE X-RAY STRUCTURE OF HUMAN NEUROGLOBIN A15C MUTANT	X-ray diffraction	1.35	14.5	Fe2+	HEM	PROTOPORPHYRIN IX CONTAINING FE
ISOMERASE	5.4.99.2	Methylmalonyl-CoA mutase; mitochondrial	Homo sapiens	2XIQ	CRYSTAL STRUCTURE OF HUMAN METHYLMALONYL-COA MUTASE IN COMPL ADENOSYLCOBALAMIN AND MALONYL-COA	X-ray diffraction	1.95	16.3	Co3+	B12	COBALAMIN
ISOMERASE	5.4.99.1	Beta-methylaspartate-glutamate mutaseSTRUCTURE	Clostridium cochlearium	6h9e	STRUCTURE OF GLUTAMATE MUTASE RECONSTITUTED WITH HOMO-COENZY	X-ray diffraction	1.82	13.8	Co3+	B12	COBALAMIN
LYASE	4.99.1.3	Sirohydrochlorin Cobaltochelatase	Salmonella enterica	2awp	ANAEROBIC COBALT CHELATASE (CBIK) FROM SALMONELLA TYPHIMURIU COMPLEX WITH METALATED TETRAPYRROLE		1.9	19.7	Co2+	SIR	COBALT SIROHYDROCHLORIN

Table 1. List of the porphyrine classes structure used in the study accompanied with the title and PDB entry, ligands names, resolution and R-factor which reflect the quality of the quality of the structures .

II. Extraction and process data in a CSV format and data Mining :

In order to study the porphyrin structural motifs and their relationship with their function in the 21 selected porphyrin proteins, the bioinformatics tool Protein-Protein Interaction (PPI) was used to extract the binding details between the different types of porphyrin rings existing in the selected protein.

The PPI is a customized version of the **SSFS** (Sequence Structure and Function Server), which is a bioinformatics tool developed here at the Department of Biology by Dr. Abdelkrim Rachedi (Golovin *et. al.*, 2005), **Figure 31**. has been used to carry of calculation binding motifs, Ligand motifs environment details, the url address to access and use the tool :

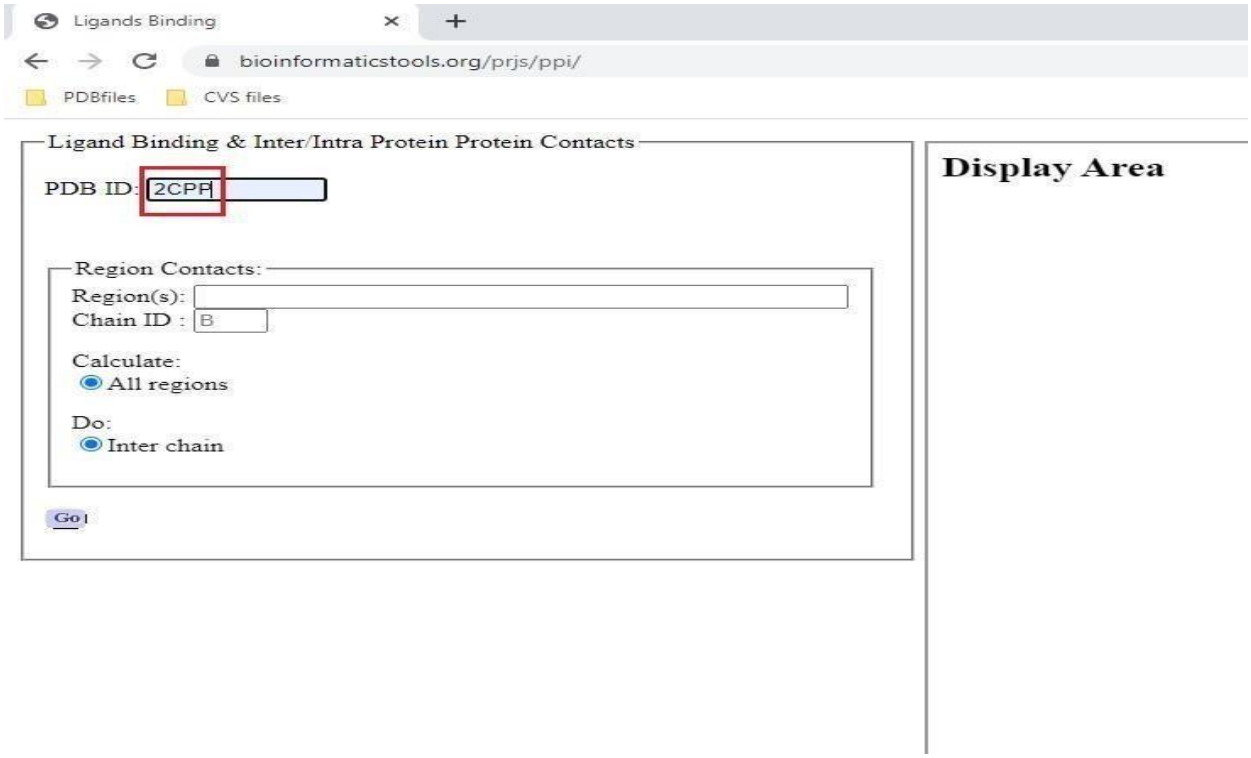


Figure 31.capture of the interface of the **PPI** tool.

II.1 .Binding Details Data Generation and Downloading in CSV files:

For extracting and processing the binding environment data details, 21 3d-structures from the PDB pertaining to the porphyrin proteins, see Table 1, processed using their PDB IDs and binding data have been generated in CSV format. The steps involved in the process are outlined as shown in **Figure 32**.

First step we insert the PDB ID in the interface of the site click **Go** in this case for a example insert the PDB ID of the **cytochrome P450** "2CPP", After that we click on ligand binding that lead us to a new page we gonna choose and click on explore environment of the protein and click on it and download the **CSV** file

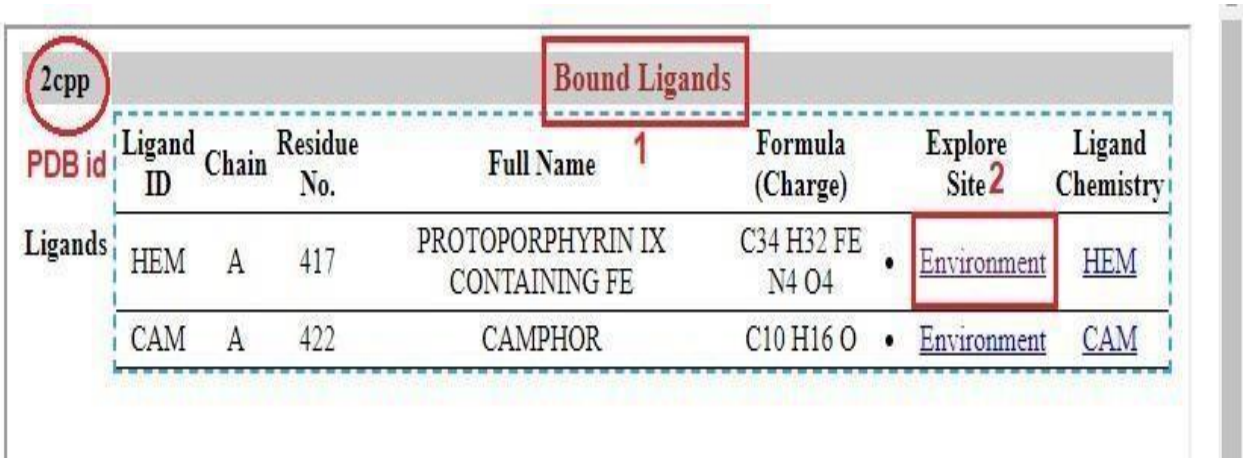


Figure 32.capture of the following steps for the data extraction

Entry: 2cpp	OXIDOREDUCTASE(OXYGENASE)

Protein-Ligand Environment

Protein or NA Residues						Ligand					Bonds	
Chain	SSelm	Name	Number	Atom		Chain	Name	Number	Atom		Distance/ Å	Possible Bond Type
A	No SSE	THR	101	CG2		A	HEM	417	CAD		3.22	van der Waals
A	No SSE	THR	101	CG2		A	HEM	417	O2D		3.02	van der Waals
A	106-126 H: 1	GLN	108	OE1		A	HEM	417	O1D		3.01	H.Bond
A	106-126 H: 1	ARG	112	CD		A	HEM	417	O1D		3.43	van der Waals
A	106-126 H: 1	ARG	112	NH1		A	HEM	417	CBD		3.42	
A	106-126 H: 1	ARG	112	NH1		A	HEM	417	CGD		3.45	
A	106-126 H: 1	ARG	112	NH1		A	HEM	417	O1D		2.66	H.Bond
A	234-267 H: 1	LEU	245	O		A	HEM	417	CBC		3.36	van der Waals
A	234-267 H: 1	GLY	249	N		A	HEM	417	CBC		3.35	
A	234-267 H: 1	THR	252	CB		A	HEM	417	C4B		3.5	van der Waals
A	234-267 H: 1	THR	252	OG1		A	HEM	417	CHC		3.06	van der Waals
A	295-301 S: 0	ASP	297	CG		A	HEM	417	O2A		3.4	van der Waals
A	295-301 S: 0	ASP	297	OD1		A	HEM	417	O2A		3.2	H.Bond
A	295-301 S: 0	ASP	297	OD2		A	HEM	417	O2A		2.74	H.Bond

A	295-301 S: 0	ARG	299	NH1		A	HEM	417	O2A		3.07	H.Bond
A	295-301 S: 0	ARG	299	NH2		A	HEM	417	CGA		3.45	
A	295-301 S: 0	ARG	299	NH2		A	HEM	417	O1A		2.82	H.Bond
A	No SSE	THR	349	O		A	HEM	417	CMB		3.3	van der Waals
A	No SSE	PHE	350	CE1		A	HEM	417	C3B		3.48	van der Waals
A	No SSE	HIS	355	O		A	HEM	417	CAA		3.44	van der Waals
A	No SSE	HIS	355	CB		A	HEM	417	O1A		3.37	van der Waals
A	No SSE	HIS	355	ND1		A	HEM	417	CGD		3.36	
A	No SSE	HIS	355	ND1		A	HEM	417	O2D		2.58	H.Bond
A	No SSE	CYS	357	CB		A	HEM	417	NA		3.31	
A	No SSE	CYS	357	CB		A	HEM	417	FE		3.2	Metalic Bond
A	No SSE	CYS	357	SG		A	HEM	417	NA		3.39	
A	No SSE	CYS	357	SG		A	HEM	417	NB		3.27	
A	No SSE	CYS	357	SG		A	HEM	417	NC		3.13	
A	No SSE	CYS	357	SG		A	HEM	417	ND		3.41	

Table 2. The binding environment details of the HEM bound cytochrome P450cam with camphor (PDB id : 2cpp)

The ligand binding details shown in the above table is organized in the following columns :

- **The columns under the title "Protein residues":** These columns show the atoms of the enzyme residues (AA) that bind with the ligand. The residues are also denoted in terms of what secondary elements (α-helix, β-sheet or loop) they may belong to.
- **The columns under the title "Ligand":** These columns show the atoms of the ligand HEM, its number and the ligand id.

- **The columns under the title "Ligand":** These columns show the distance between atoms (**Å: Angstroms**) and the possible bonds which can for example be a **Hydrogen** or **Van der Waals bonds**.

Using the PPI system, the binding environment details of all ligand associated with the 21 PDB porphyrin proteins have been calculated, Parse the PDB file using a scripting language like **php** to extract the relevant data.

In this case, we want to extract the coordinates for the heme group. Once the data has been extracted, write it to a CSV file. The CSV file should have the following columns: x-coordinate, y-coordinate, z-coordinate, and element symbol.this data have been collected after stored onto a system of organized files **.**

II.2. Data mining for the binding details :

We using the notepad ++ and with the aid of 2 scripts language wich are the php and the html for the creating of the protein database after the extraction of the data from folders named data_files wich contain the csv files of the pdb ids and the table of the porphyrin prots clasess the folders exist in the same general folder master_project with a deferent path name.

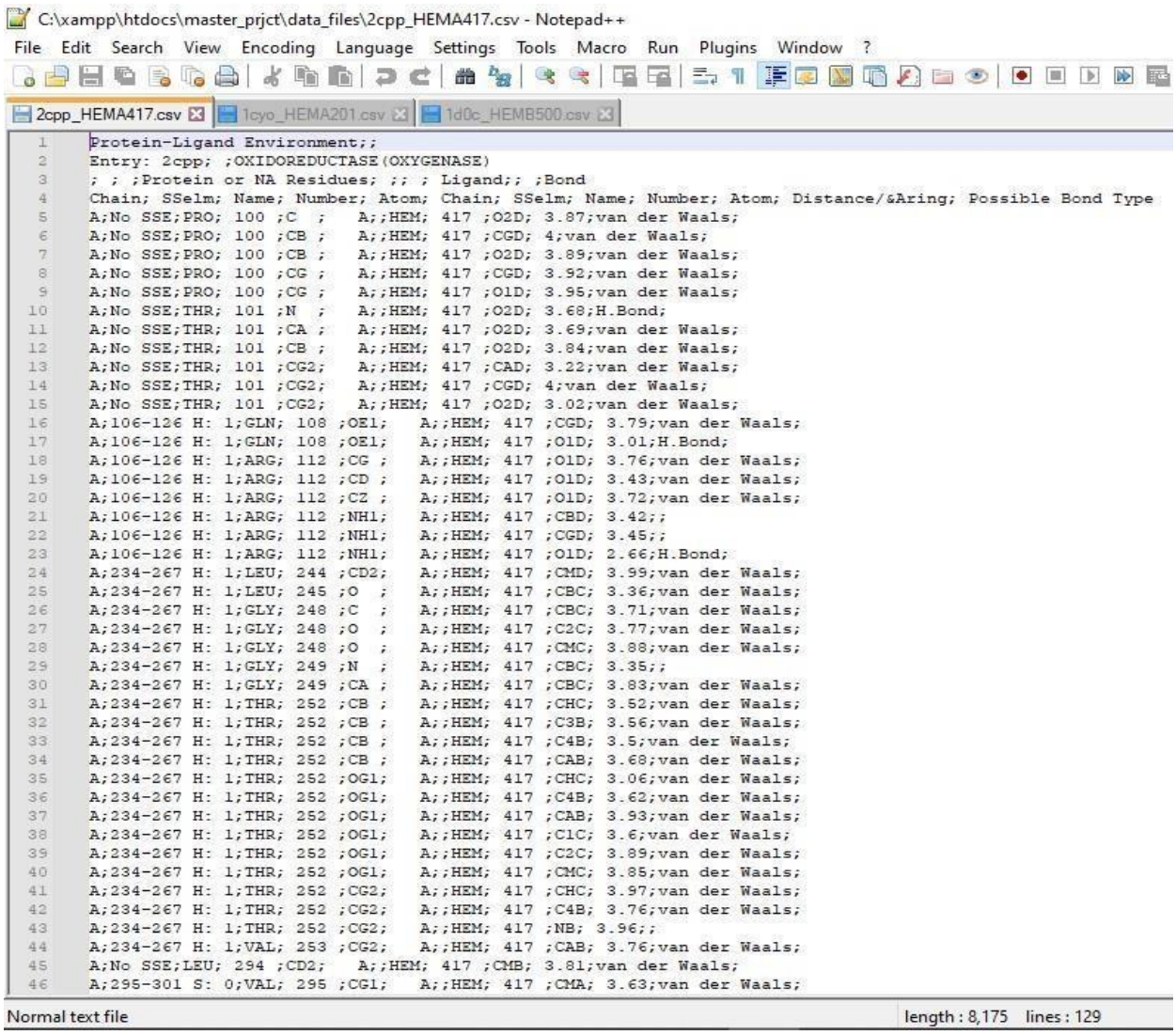


Figure 33.Notepad ++ example of a csv file, named 2cpp_HEMA417.csv Which contain the binding details of the protein residues with the HEM ligand and PDB id : 2CCP, chain A.

II.3. Binding Motifs Construction and Representation:

In the binding details presented earlier, the residues that interact with the ligand are associated with certain secondary structure elements. Table 2, which is located above, shows the association of the protein binding residues with specific regions that represent secondary structure elements. The annotation for this is illustrated by the following example, which pertains to the ligand HEM and its association with the PDB ID 2CPP, chain A:

- **The protein region labeled "no SSE"** indicates the absence of secondary structure, implying that the binding residues belong to a loop region and are denoted by the symbol (**.**).
- **The protein region labeled "106-126H:1, 234-267H:1"** corresponds to the secondary structure alpha-helix, designated by the symbol H.
- **The protein region labeled "295-301 S:0"** corresponds to the secondary structure beta-strand, designated by the symbol S.so in our case the pattern representing the binding site of the ligand HEM found in the table above is : **..HH.SS……H**

These patterns have been observed for all ligand binding sites and appear to be associated with specific functions that recur consistently (see chapter 3). They can be further annotated as structural and functional motifs, also known as binding motifs.

II.4. Graphical Representation of the Motifs:

In order to do a graphical representation of the motifs we needed to use a tool named Rasmol molecular graphics program (Sayle A.Roger & Milner-White E.J., 1995) which is a molecular visualization program that is used to generate graphical representations of macromolecules such as proteins, DNA, and RNA.

For the graphical representation we using the PPI for the download of the 4 scenes before the edition and modification of this scene. Notepad++ has also been used for the modification of the scenes before doing the capture of the images after that we stored the graphics representation which are 4 scenes in a folder named images this folder are in the same path of the General folder of the master project

For the downloading of this scenes we used **PPI** after the insertion of the PDB id and clicked on the download as shown in the **Figure 34**.

A	359-378 H: 1	ALA	363	CB	A		HEM	417	CAB	3.88	van der Waals
A	Water	HOH	511	O	A		HEM	417	CAA	3.71	van der Waals
A	Water	HOH	511	O	A		HEM	417	CGA	3.61	van der Waals
A	Water	HOH	511	O	A		HEM	417	O1A	2.66	H.Bond
A	Water	HOH	536	O	A		HEM	417	CMB	3.94	van der Waals
A	Water	HOH	566	O	A		HEM	417	CMC	3.51	van der Waals
A	Water	HOH	652	O	A		HEM	417	CGA	3.7	van der Waals
A	Water	HOH	652	O	A		HEM	417	O2A	3.09	H.Bond

Download CSV: File

Download: Ras Scene 1 Ras Scene 2 Ras Scene 3 Ras Scene 4 Download: Run file

Figure 34.capture of the following steps for the downloading of the rasmol scenes.

These graphical representations can be used to analyze the structure of these molecules and to better understand their function.Rasmol can display various types of secondary structure motifs, such as alpha-helices, beta-strands, and loop regions, in a variety of graphical representations.

Once you have generated a graphical representation of the binding motif in Rasmol, you can save it as an image file, such as a PNG or JPEG. You can then view and share the image as needed, There is three types of representation :

Alpha-helix representation(H): The alpha-helix is a common structural motif in proteins, consisting of a helical arrangement of amino acids. In Rasmol, alpha-helices can be represented as cylinders or ribbons. The cylinder representation shows the helix as a tube-like structure, while the ribbon representation shows the helix as a twisted ribbon. To display an alpha-helix in Rasmol, the user can select the helix region and use the "cartoon" command to choose between the cylinder or ribbon representation. For example, the command "cartoon cylinder" will display the helix as a cylinder show as **Red ribbons**.

Beta-strand representation(S): Beta-strands are another common structural motif in proteins, consisting of a sheet-like arrangement of amino acids. In Rasmol, beta-strands can be represented as arrows or ribbons. The arrowrepresentation shows the strand as an arrow-like structure, while the ribbon representation shows the strand as a twisted ribbon. To display a betastrand in Rasmol, the user can select the strand region and use the "cartoon" command to choose between the arrow or ribbon representation. For example, the command "cartoon ribbon" will display the strand as a ribbon show as **Yellow ribbons**.

Loop region representation (.): Loop regions are non-repetitive sections of proteinstructure that connect alpha-helices and beta-strands. In Rasmol, loop regions can be represented as coils or lines. The coil representation shows the loop as a coiled structure, while the line representation shows the loop as a straight line. To display a loop region in Rasmol, the user can select the loop region and use the "cartoon" command to choose between the coil or line representation. For example, the command "cartoon coil" will display the loop as a coil show as a **Light Grey strips.**

There are three types of Rasmol images were produced for each ligand binding data:

- **Binding details general view representation, Figure 36-b**.
- **Motif+Ligand+Binding Residues(dots form) representation, Figure 37-b**.
- **Motif+ Ligand+ Binding Residues(cpk form) representation, Figure 38-b**.
- **Stereo graphic representation, Figure 39-b**.

To generate the visual representation of the patterns in the binding sites, the Rasmol program employs a scripting language. This scripting language instructs the program on how to depict the molecular data in a graphical format. The resulting graphical representation can be observed in the figure. Essentially, the script serves as a set of commands that specifies how the program should generate the images, including the size, shape, color, and position of the different molecular components.

These commands may include instructions to highlight specific atoms, bonds, or residues that are of interest, or to manipulate the perspective and lighting conditions to optimize the visual clarity of the final image.

For the rasmol graphical representation we using the PPI for the download of the 4 rasmol scene before the edition and modification of this scene using notepad ++ .

II.4.1. binding details general view representation :

Blow is a Rasmol script, see **Figure n°35-a,** that produces a graphical representation of the **"..HH.SS……H"** motif without displaying the ligand "HEM", see **Figure n°35-b.**

RasMol Command Line

```
RasMol> set specpower 25
RasMol> set specular on
RasMol> wireframe off
RasMol> zoom 150
RasMol> set background [220,230,245]
RasMol> select :A
3462 atoms selected!
RasMol> backbone 0.08
RasMol> color [200,200,200]
RasMol> select 106-126:A,234-267:A,295-301:A,315-323:A,359-378:A
3462 atoms selected!
RasMol> color structure
RasMol> cartoon 1.4
RasMol> select 100:A,101:A,108:A,112:A,244:A,245:A,248:A,249:A,252:A,253:A,294:A
,295:A,297:A,299:A,322:A,349:A,350:A,351:A,354:A,355:A,357:A,359:A,363:A
168 atoms selected!
RasMol> color cpk
RasMol> wireframe 0.3
RasMol> select 417:A
43 atoms selected!
RasMol> color green
RasMol> cpk
```

Figure 35-a.Capture of the RasMol script to create the general motif representation shown in **Figure 35-b**.

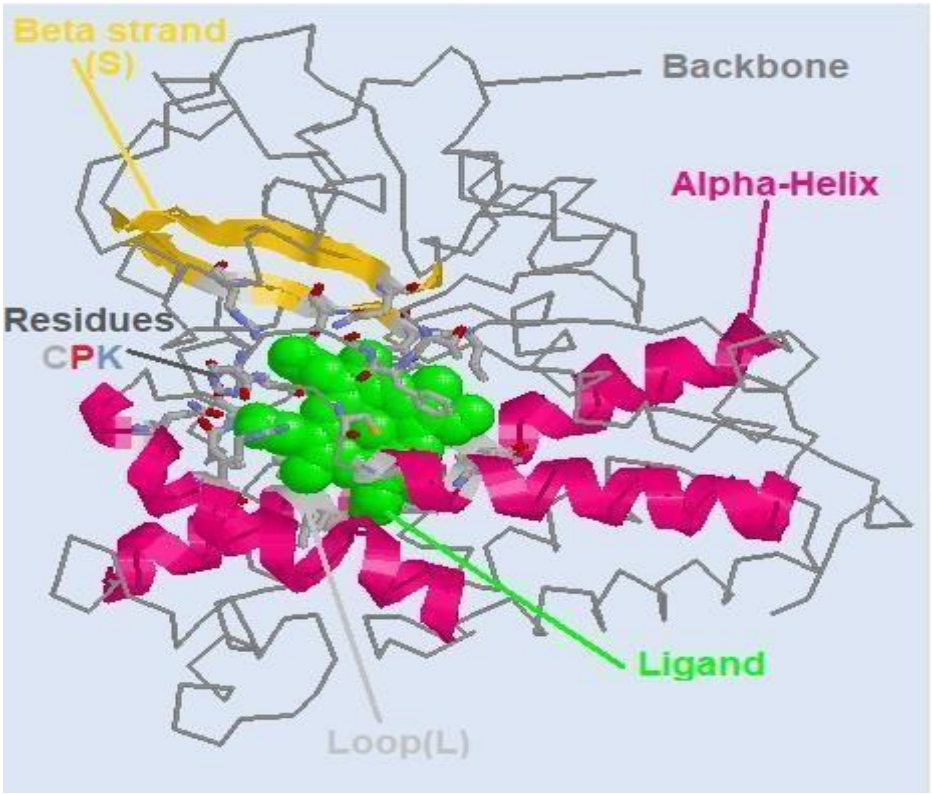


Figure 35-b.capture of the RasMol representation of the bidning motifs with displaying the binding residue and ligand HEM case of cytochrome P450 (PDB id : 2CPP chain A).

II.4.2. Motif+ Ligand+ Binding Residues(dots form) representation:

Below is a Rasmol script, **Figure 36-a**.that generates a graphical representation of the **"..HH.SS……H"** motif with the ligand "HEM" visible but without showing the binding Residues, **Figure 36-b**.

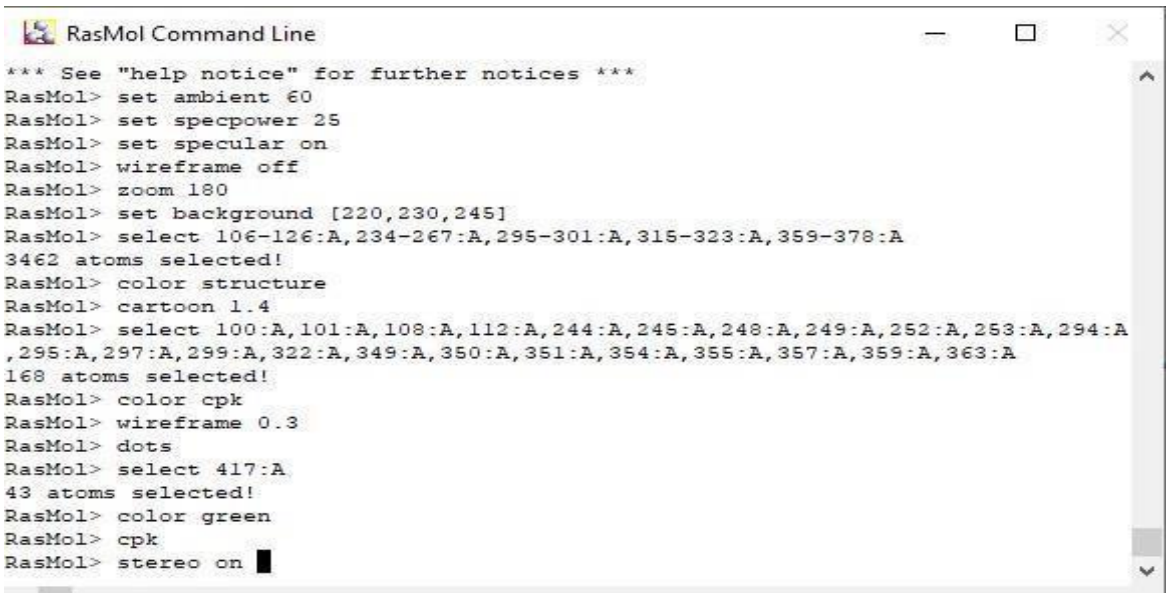


Figure 36-a.Capture of the RasMol script to create the motif representation shown in **Figure 36-b**.

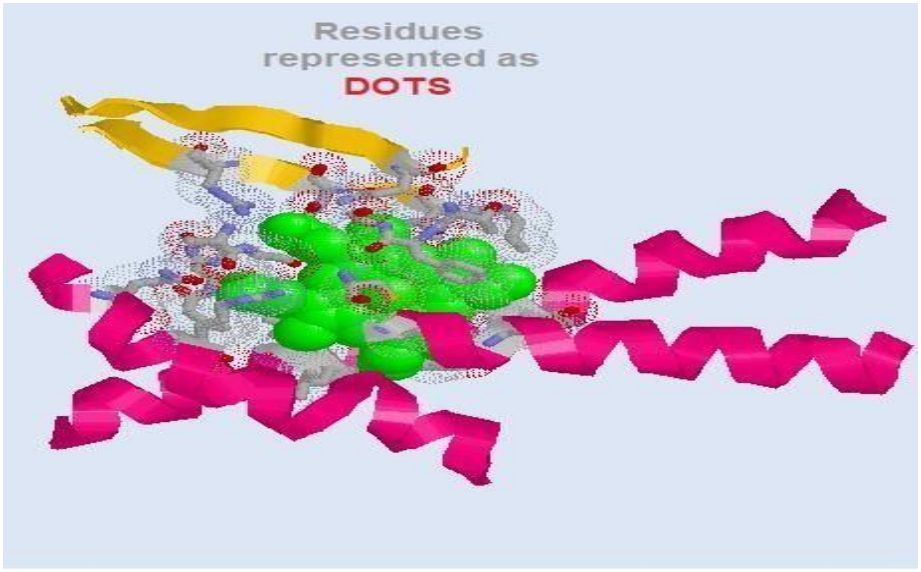


Figure 36-b.Capture of RasMol representation of the binding motifs where the residues represented as dots in the case of cytochrome P450 (PDB id : 2CPP chain A).

II.4.3. Motif+ Ligand+ Binding Residues(cpk form) representation:

Below is a Rasmol script, **Figure 37-a**.that generates a graphical representation of the **"..HH.SS……H"** motif, including both the ligand "HEM" and the binding residues, **Figure 37b**.

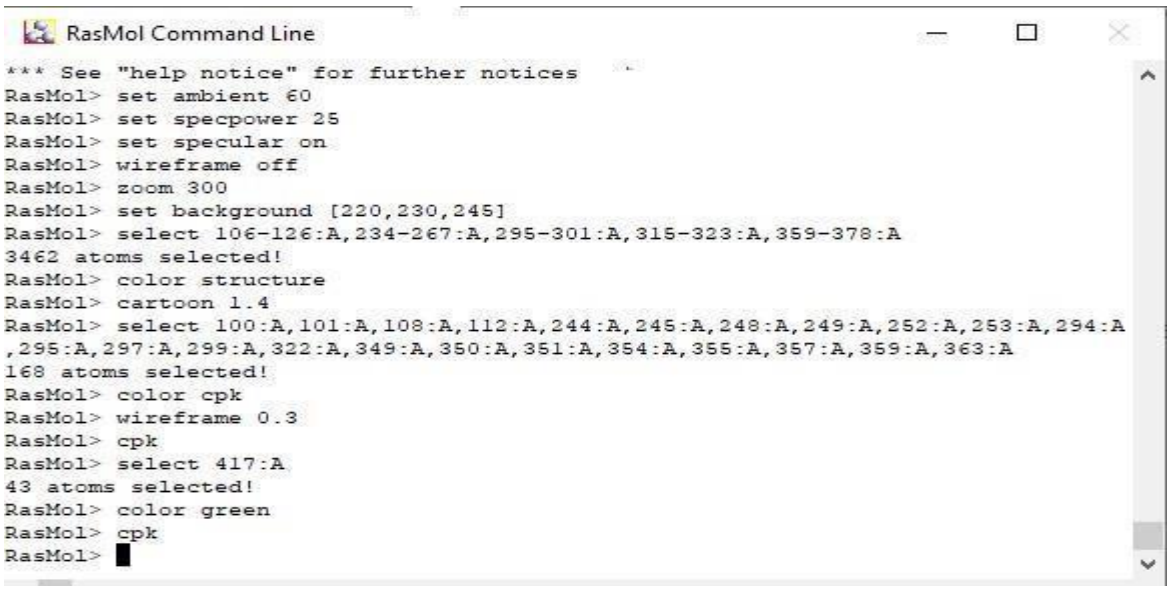


Figure 37-a.capture of the RasMol script to create the motif representation show in **Figure 37-b**.

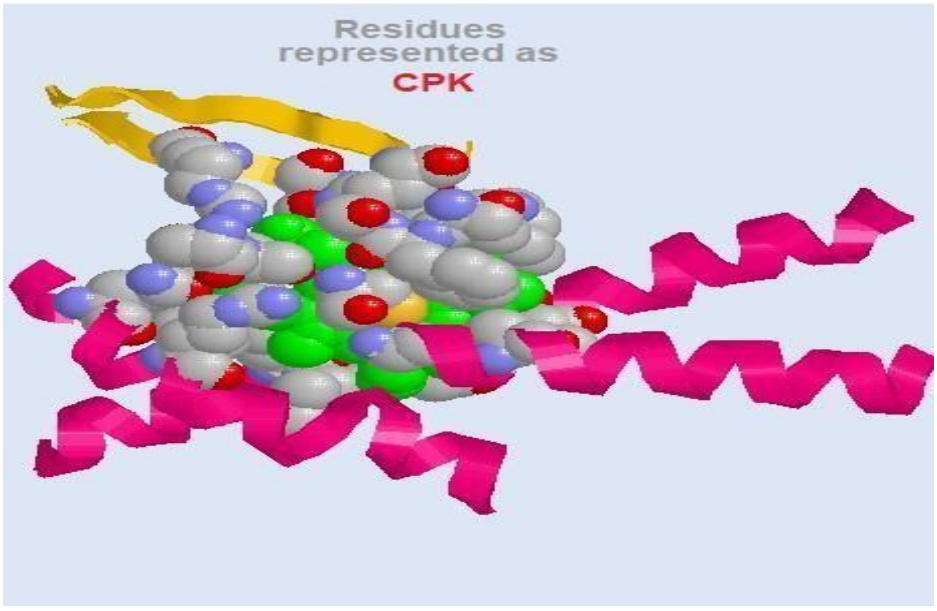


Figure 37-b.Capture of RasMol representation of the binding motifs where the residues represented as cpk in the case of cytochrome P450 (PDB id : 2CPP chain A).

II.4.4. Stereo graphic representation :

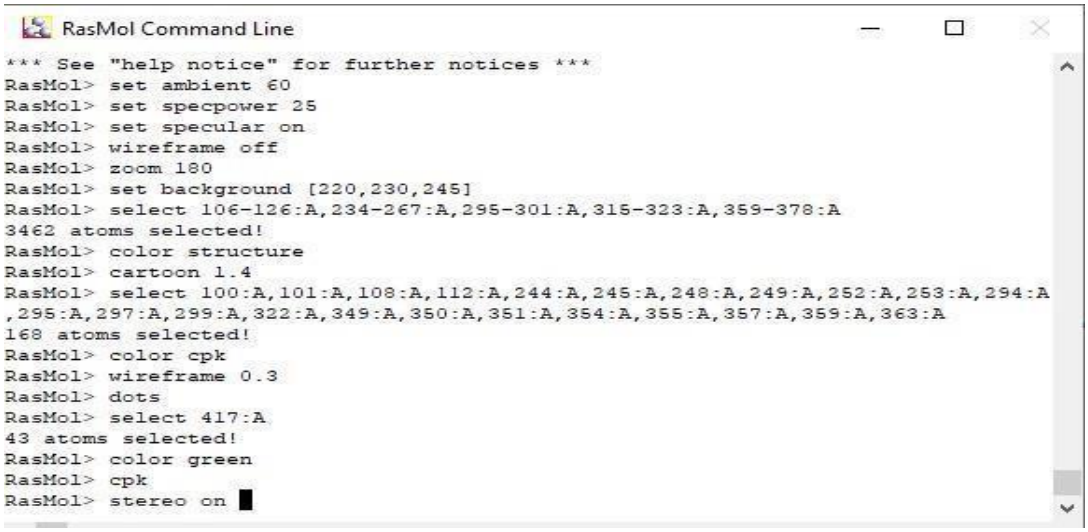


Figure 38-a. Screen-capture of the RasMol script to create the motif representation show in **Figure 39-b**.

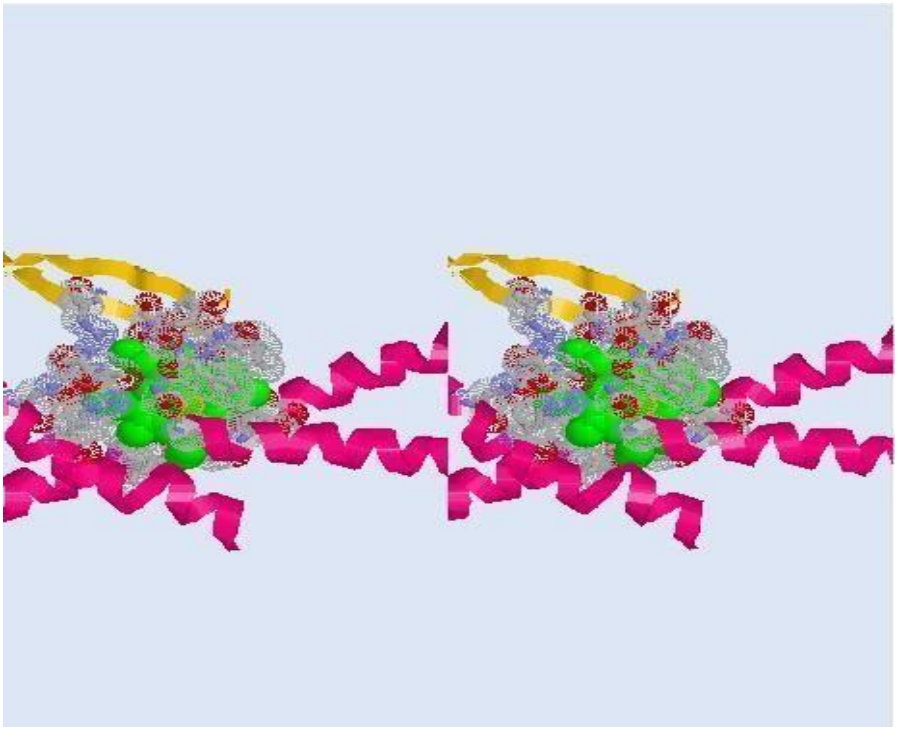


Figure 38-b. Capture of RasMol representation of the binding motifs in stereo displaying

NOTE: it should be noted here that the three types of the graphical representations show above for the case of the ligand **HEM** in the hemeprotein cytochrome P450 (PDB id: 2CCP, chain A) wich

this graphical representations are done for all the ligand binding Motifs in all the porphyrin proteins studied in this project,(**Index**)

III. Flat-Files Database creation:

For the creation of the Flat File database, **Figure 39**.we using three folders wich are the data_files, PDB files, graphics. The 1st folder data_files contain the all the csv_files of the 21 selected porphyrin proteins, the 2nd folder contain the all the PDB id of the proteins, the 3rd folder contain the the rasmol seans fot the creation of the motifs graphics

For the illustrion we choosing only 6 PDB id, for the rasmol scenes for the rasmol seans we choose chain A for the rasmol seans

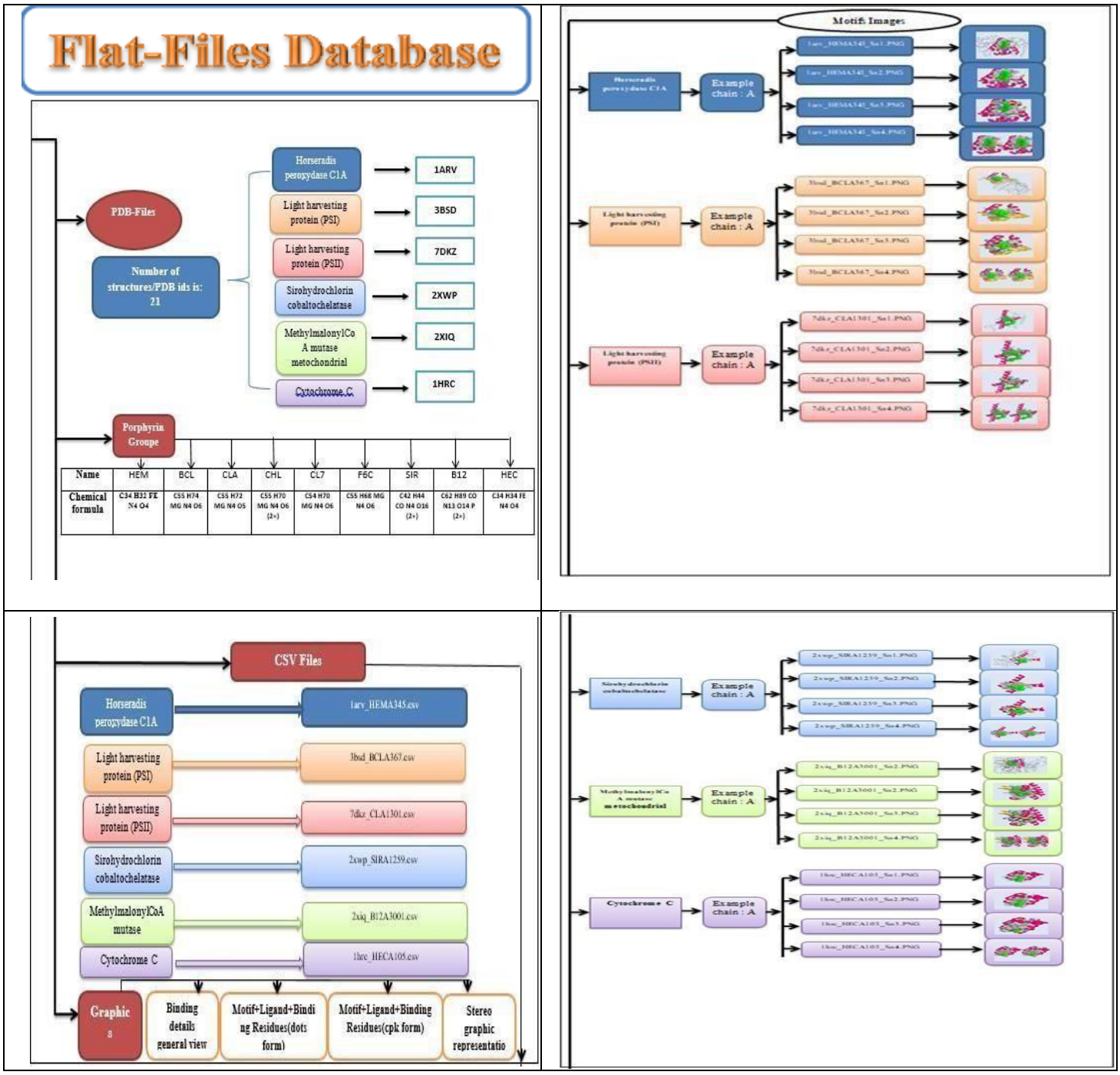


Figure 39.The database schema representing the architecture of the created Flat-File database; the **left side** show the arrangement and classification of the files containing the csv-files which contain the binding environment details while the **right side** of the figure show the arrangement of the graphics files containing the motifs in Rasmol scenes for the illustrations we choose 6 proteins such protein contained one of porphyrin group (ligand), for the chlorophyll family we mention only the CLA(chlorophyll A).

IV. World Wide Web Database :

The supervisor of this project has mounted the database on the "Bioinformatics" server of the University of Saida, and developed programming scripts to facilitate searching the database. This was done to enable sharing of the data and results with the scientific community both locally and internationally.

The database has been given the name, Porphyrin Proteins Binding Structural Motifs (PPBSMS) and is accessible online from the following web address:

https://bioinformatics.univ-saida.dz/prjs/ppbsms/

Chapter III

I. Introduction

As shown in Chapter II, this project has resulted in the creation of a readily accessible online database that provides details for a constructed set of novel structural motifs, referred to as Porphyrin Proteins Binding Structural Motifs. These motifs binds the different types of porphyrin ligands found in a selected set of biologically important protein and enzyme that fall under the category of Porphyrin Proteins. The biology relevance and importance of such proteins are explained in Chapter I.

This chapter presents and discusses the importance of the understanding how porphyrin proteins bind their ligands which would shed more light on the relation between structure and biological function in this category of proteins and by extension in all types of proteins,

The database is named the **Porphyrin Proteins Binding Structural Motifs – PPBSMs** and is accessible via the web-site: https://bioinformatics.univ-saida.dz/prjs/ppbsms/ and reside on the web-server of the University of Saida.

II. Results outlines:

II.1. Porphyrin Binding Structural Motifs:

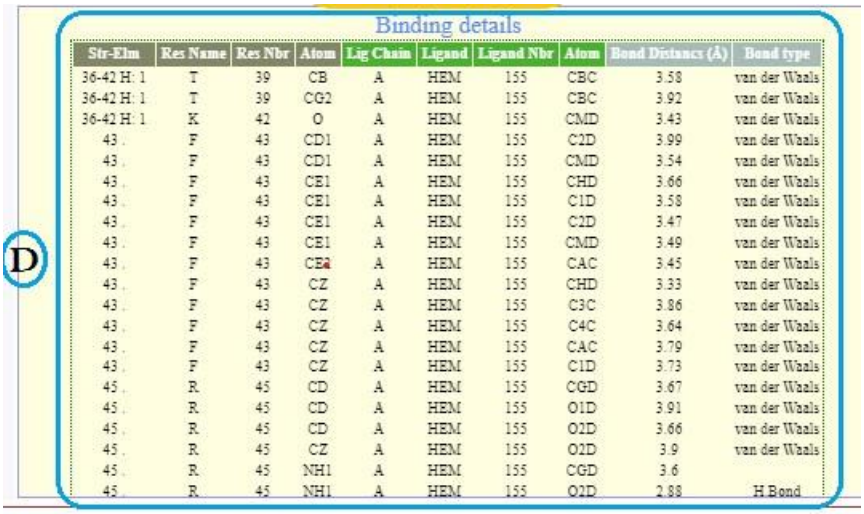
The binding motifs are the 3D-structurally arranged secondary structure elements which contribute with residue that interact with the porphyrin ligand in each of the porphyrin proteins selected in this study.

The motifs are represented in string form where α-helices are represented with **H** and βstrands are represented with **S** and loop-regions are represented with „**.**“, see Figure 41 (B1). In addition, residues involved in the binding are presented with single letter code of aminoacids, see Figure 41 (B2).

Figure 40. Structural binding linear prsentation **B1**: **H** denotes α-helix and „**.**" denotes loop region. **B2**: Amino acids shown with approximation of their belonging to the secondary structure elements.

II.2. Binding details:

The binding details summarise the bond lengths and types of each and every residue atoms involved in the binding of the porphyrin ligand. It also show the ranges of the secondary structure and loop regions, see Figure 42.



Binding details

Str-Elm	Res Name	Res Nbr	Atom	Lig Chain	Ligand	Ligand Nbr	Atom	Bond Distance (Å)	Bond type
36-42 H: 1	T	39	CB	A	HEM	155	CBC	3.58	van der Waals
36-42 H: 1	T	39	CG2	A	HEM	155	CBC	3.92	van der Waals
36-42 H: 1	K	42	O	A	HEM	155	CMD	3.43	van der Waals
43 .	F	43	CD1	A	HEM	155	C2D	3.99	van der Waals
43 .	F	43	CD1	A	HEM	155	CMD	3.54	van der Waals
43 .	F	43	CE1	A	HEM	155	CHD	3.66	van der Waals
43 .	F	43	CE1	A	HEM	155	C1D	3.58	van der Waals
43 .	F	43	CE1	A	HEM	155	C2D	3.47	van der Waals
43 .	F	43	CE1	A	HEM	155	CMD	3.49	van der Waals
43 .	F	43	CEA	A	HEM	155	CAC	3.45	van der Waals
43 .	F	43	CZ	A	HEM	155	CHD	3.33	van der Waals
43 .	F	43	CZ	A	HEM	155	C3C	3.86	van der Waals
43 .	F	43	CZ	A	HEM	155	C4C	3.64	van der Waals
43 .	F	43	CZ	A	HEM	155	CAC	3.79	van der Waals
43 .	F	43	CZ	A	HEM	155	C1D	3.73	van der Waals
45 .	R	45	CD	A	HEM	155	CGD	3.67	van der Waals
45 .	R	45	CD	A	HEM	155	O1D	3.91	van der Waals
45 .	R	45	CD	A	HEM	155	O2D	3.66	van der Waals
45 .	R	45	CZ	A	HEM	155	O2D	3.9	van der Waals
45 .	R	45	NH1	A	HEM	155	CGD	3.6	
45 .	R	45	NH1	A	HEM	155	O2D	2.88	H Bond

Figure 41. Partial binding details of the secondary structure and loop elements. Estimated atomic bonding distances and types are shown.

Note that both Figures 41 and 42 can be understood better in the context of the global Figure 44 show further below.

III. Presentation of results:

III.1. Online Access and Database Querying :

The online version of the database "**PPBSMs**" can be uploaded by invoking the URL address mentioned above, see Figure 43.

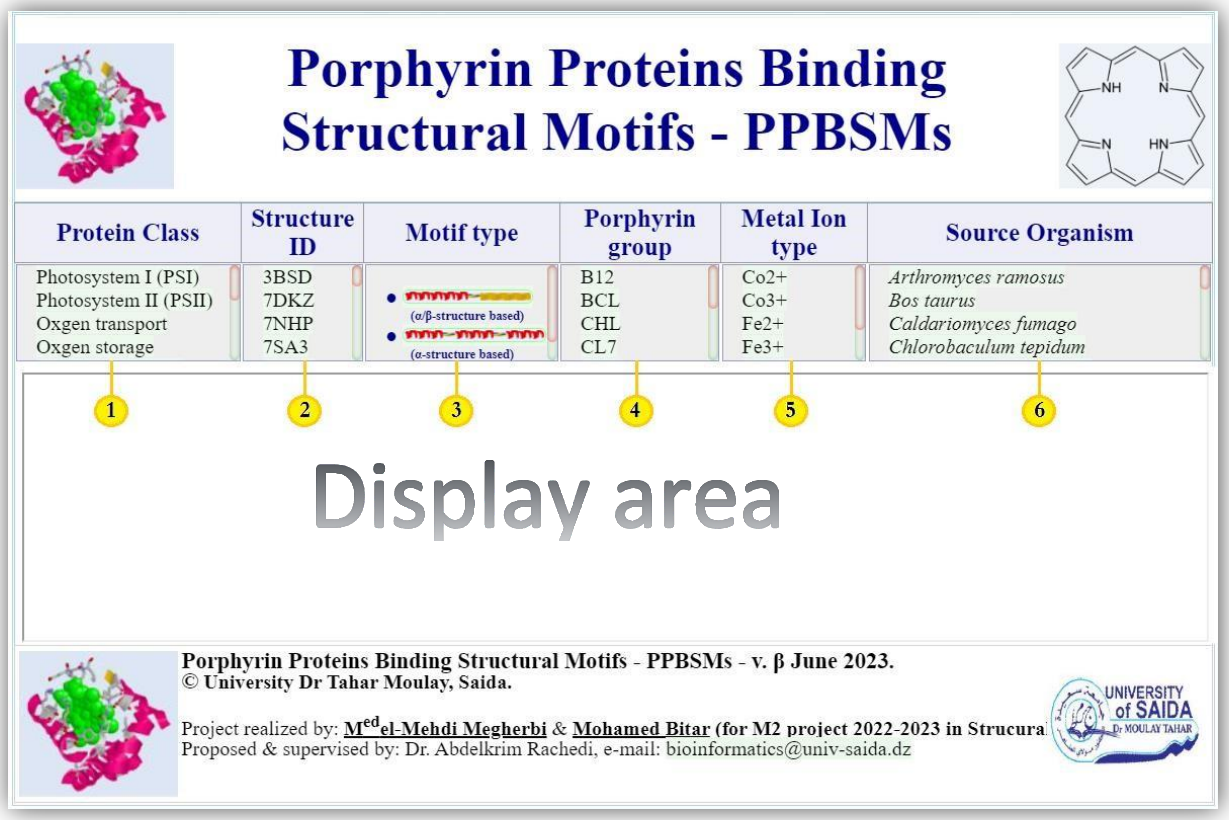


Figure 42. The main web interface of **PPBSMs** database as screen-captured from the web address, see next sections for explanations on the highlighted areas.

III.2. Database Methods of Querying and Results Display:

As shown above in Figure 43, the interface of **PPBSMs** allows for six (6) methods of searching the database content. For clarity, these methods of querying are yellow coloured and highlighted.

1. This list allows for querying by clicking on **Proteins Class**
2. This clickable list of **Structure ID** entries allows querying by PDB entry.
3. This allows for querying by **Motif type**; α-structure or α/β-structure based motifs.
4. This allows for querying by **Porphyrin group** or ligand id.
5. This clickable list of **Metal Ion type** allows querying by Ion type associated with the porphyrin group.
6. This clickable list of **Source Organism** allows querying by organism name from which the porphyrin protein is isolated.

The "**Display Area**" is the space region where querying results are displayed

III.2.1. Querying by Proteins classes:

This method of query allows selecting the protein class by clicking on the hyperlinks in **area** 1 **.** The search produces an output page of results that show binding details of the related porphyrin group (ligand) with the porphyrin protein which may be associated with more than one PDB entry, as shown in **Figure 43**.

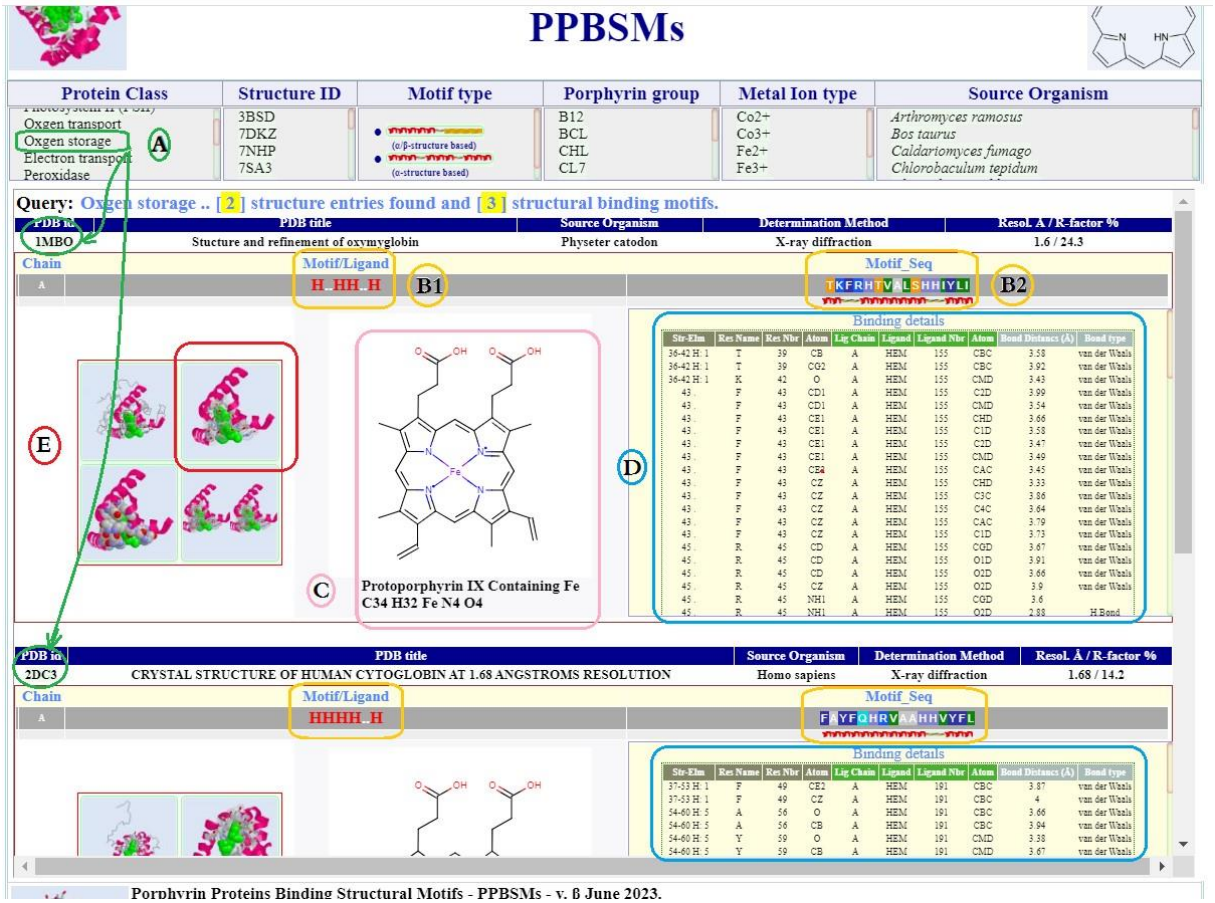


Figure 43. A screen-shot shows the six methods of searching **PPBSMs** database. The red highlighted areas represent the different types of results:

III.2.2. Querying by Structure ID:

The ligand binding details per structure ID can be searched by selecting the desired PDB entry and clicking on it from the hyperlinked structure ID lists, **area** 2 as in **Figure 44**. The results page would display the binding details associated with all querying methods.

III.2.3. Querying by Motif type:

The two major motif types, α-structure and α/β-structure based, can be explored by selecting one of the options from the list 3 as shown in **Figure 44**.

III.2.4. Querying by Porphyrin group:

This method of query is allowed selecting the ligand id to be explored by clicking on the hyperlinked porphyrin group list, **area** 4 as in **Figure 43.**

III.2.5. Querying by Metal Ion type:

This method of query allows the exploration of the motifs using the porphyrins metal ion types. This is done clicking on the hyperlinked metal ion list, **area** 5 as in **Figure 43.**

III.2.6. Querying by Source Organism:

This method of query is allowed selecting the organism name to be explored by clicking on the hyperlinked source organism list, **area** 6 as in **Figure 43.**

IV. Binding Motifs and Properties:

The total of 9 ligands belong to the Porphyrin group studied in this project bound to 51 protein chains associated with the 21 PDB entries. This resulted in the total number of 51 motif instances of which to 43 binding motifs are unique as shown below in Table 3.

Porphyrin group	Full Name	Number of unique motifs	Motifs
HEM	Protoporphyrin IX Containing Fe	19	**α-based:** HH...HH , .HHH , H..HH..HH , H..HH..H , HHHHH..HH , HHHH..H , HHHHHH , HHHHH , ..HHH.HH , HHHH.H, **α/β-based:** SSH...HSH....H , H...HH.S...SHH , ..HH.SS.....H , H....SSS.HH.....H , H....HHH.S....SH , H....H.S.SHHS. , H.....S..SHHHS. , ...SSSHH.....HH , ...SSS.HH.....HH ,
HEC	HEME C	1	**α-based:** .H.....H.H.....H,
BCL	Bacteriochlorophyll A	1	**α/β-based:** SSSHH.S
CLA	Chlorophyll A	15	**α-based:** .....HHH , ...HH , ...HH , HH , ..H.. , HHHHHHHHHHH , HHHHH , ..HHHH , H.HH , HH.HHHH , ...HHH , H.HH.. , HH.HHHHH , HHHHHHHHH , HHHHH..HH
CHL	Chlorophyll B	4	**α-based:** .H..HH , H...H. , ...H.. , ...H.
CL7	Chlorophyll D	1	**α-based:** HHH.HH
F6C	Chlorophyll F	2	**α-based:** H.HHH, HH..H

SIR	Cobalt Sirohydrochlorin	1	**α-based:** ..H...H....H...H
B12	Cobalamin	4	**α/β-based:** S.H.H..H.....H...HS...S.S... , S..H.H..H.....H...HS...S.S.... , ...HS...S.....HSH.....H , ...HS...S.....HSH.....H

Table 3. The porphyrin ligands and binding motifs categories. The data shows clear tendency of most of the porphyrin ligands to bind motifs of specific structural type. Coloured motifs are examples of similar motifs found in the function different porphyrin proteins. Refer for more discussion in the next section "**V. Porphyrins Binding Tendency for Structural Motif type**"

V. Porphyrins Binding Tendency for Structural Motif type:

Although the results of this study project are based on a small set of porphyrin proteins, the sample of motifs obtained, and summarised n Table 3, can be used to draw a number of preliminary characterisation of the motifs and extract a list of their properties:

V.1. Motifs Classification:

V.1.1. α-structure based motifs:

This type of motifs are exclusively made of α-helices and distributed residues from loops regions. These are seen with a group of motifs interacting with HEM, all the motifs binding HEC (HEME C), all Chlorophyll groups (A, B, D, F) and Sirohydrochlorin (SIR).

V.1.2. α/β-structure based motifs:

These are composed of both α-helices and β-strands (parts of β-sheets) in addition to residues from loop regions. These are found in a second group that bind with HEM, all the motifs that interact with Bacteriochlorophyll A (BCL) and Cobalamin (B12).

V.2. Motifs Binding Tendency:

V.2.1. Porphyrins tendency to bind both α-structure and α/β-structure based motifs. This is seen exclusively with the HEM group, see Figures 45 and 46.

V.2.2. Porphyrins tendency to bind only α-structure based motifs. This is seen with the HEC (HEME C) group, Chlorophyll groups (A, B, D, F) and Sirohydrochlorin (SIR), see Figures 48 and 49 .

V.2.3. Porphyrins tendency to bind only α/β-structure based motifs. This is seen with cases of Bacteriochlorophyll A (BCL) and Cobalamin (B12).

VI. Motifs Structure and Evolutionary Relationship:

As show in Table 3, the coloured motifs, show structural conservation of the binding motifs across the different species and biological function. This discovery supports evolutionary relationship between the different functional porphyrin groups notably seen in the motifs binding the HEME and Chlorophyll groups though both of which does different biological function.

VII. Motifs Structural Arrangement and Function relationship:

The groups of motifs reported above and in Table 3 notably those that bind Cobalamin (B12) and Sirohydrochlorin show quite a spread out types of motifs. The secondary structure element, α-helices and β-strands constructing these motifs are quite far from each other in sequence and are separated with large loop regions, however, they group together in close proximity in 3D-space, , see Figures 49 and 50.

Such a structural arrangement of distant regions enable the binding of the porphyrin groups thereby insuring their biological function. These cases enforce more the concept of structure-function relations ship.

VIII. Graphical Representation of Binding Motifs:

This section provides some examples of porphyrin groups in complex with their motifs.The graphics has been generated using the Rasmol software (Sayle A.Roger & Milner-White E.J., 1995).

VIII.1. HEM group binding motif all α-structure based:

The graphical presentation, Figure 45., displays the HEM porphyrin group (Protoporphyrin IX Containing Fe), in green van-der-waals representation, binding an all α-structure motif. Below is an example of the binding motif **HHHHH..HH** (Table 3).

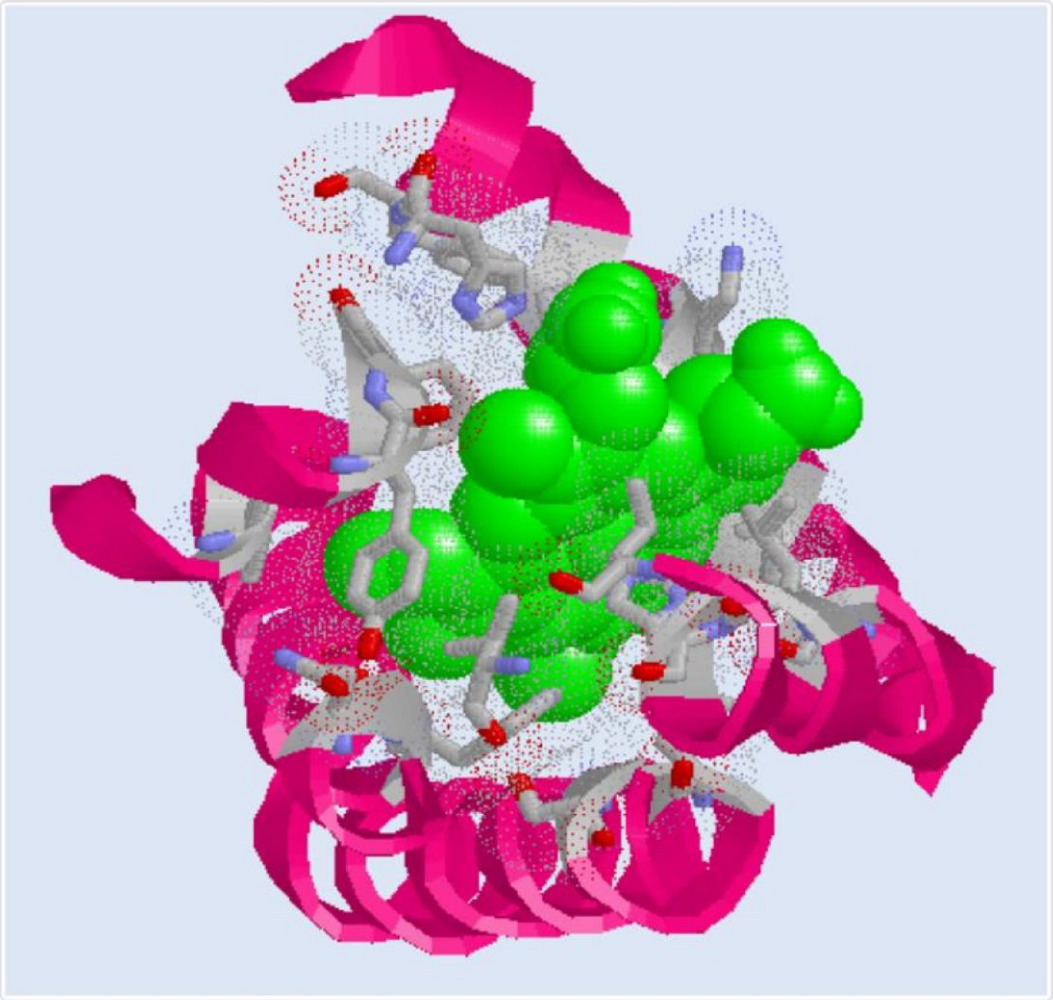


Figure 44. The Structural Motif: [HHHHH..HH] – binding the HEM group (green).The residues making the actual binding are LFFHKVAFLLHLVNFLL (PDB: 1GZX).

VIII.2. HEM group binding motif α/β-structure based:

The graphical presentation, Figure 46., displays the HEM porphyrin group (Protoporphyrin IX Containing Fe), in green van-der-waals representation, binding an α/β-structure motif. Below is an example of the binding motif **SSH...HSH....H** (Table 3).

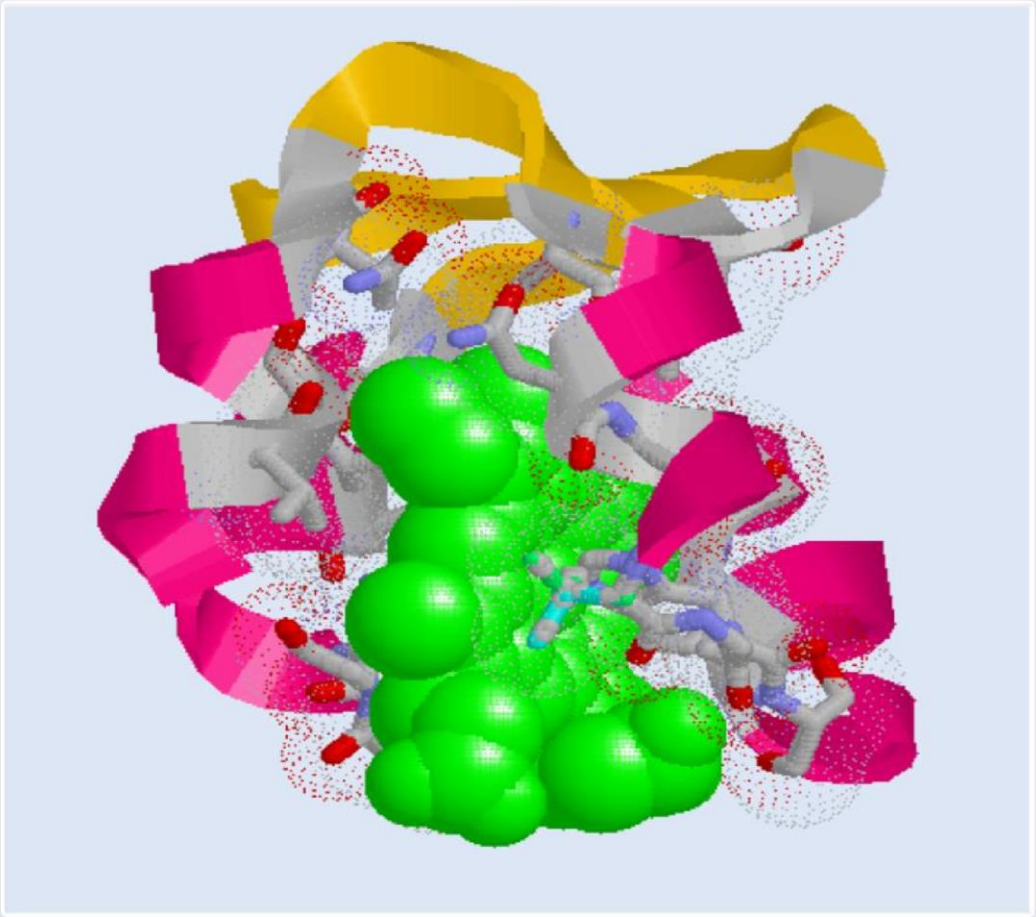


Figure 45. Structural Motif: [SSH...HSH....H] – binding the HEM group (green).The residues making the actual binding are LLFHPGVLQANFVGHSALS (PDB: 1CYO).

VIII.3. BCL group binding motif α/β-structure based:

The graphical presentation, Figure 47, displays the BCL porphyrin group (Bacteriochlorophyll A), in green van-der-waals representation, binding an α/βstructure motif. Below is an example of the binding motif **SSSHH.S** (Table 3).

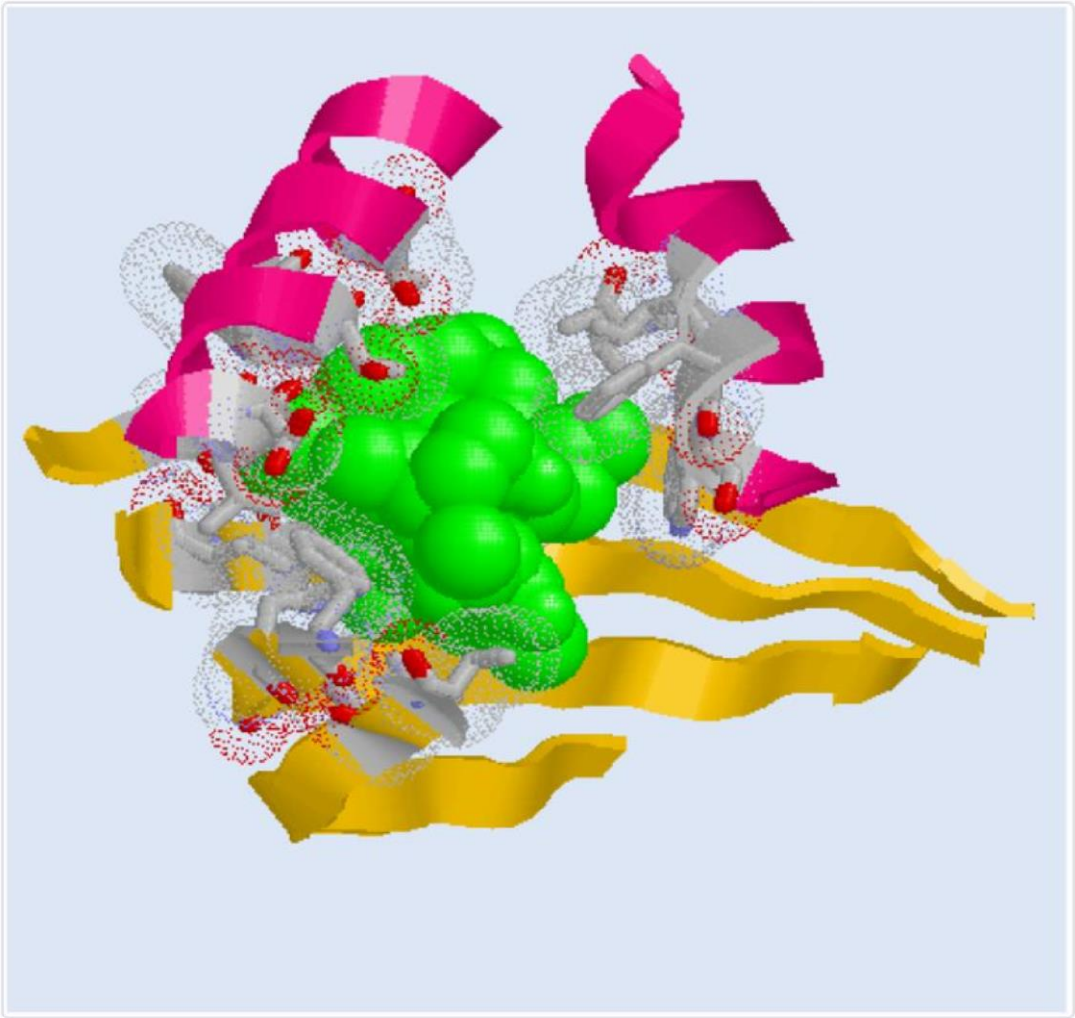


Figure 46. The Structural Motif: [SSSHH.S] – binding the BCL group (green).The residues making the actual binding are VVFHFMVDLTWTIFWIGSW (PDB: 3BSD).

VIII.4. CLA group binding motif all α-structure based:

The graphical presentation, Figure 48, displays the CLA porphyrin group (Chlorophyll A), in green van-der-waals representation, binding an all α-structure motif. Below is an example of the binding motif **.....HHH** (Table 3).

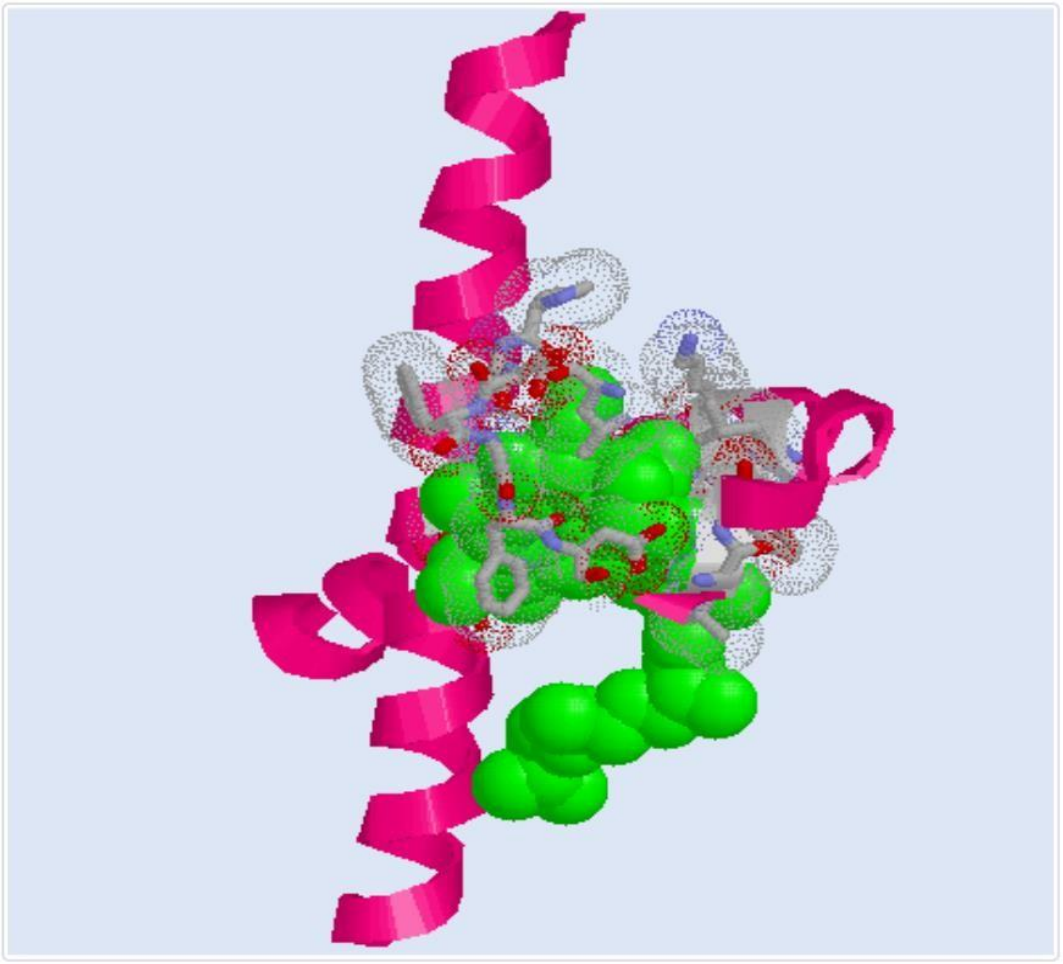
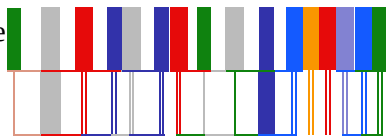


Figure 47. The Structural Motif: [.....HHH] – binding the BCL group (green).The residues making the actual binding are  (PDB: 7DKZ)

VIII.5. SIR group binding motif all α-structure based (spread out form):

The graphical presentation, Figure 49, displays the SIR porphyrin group (Cobalt Sirohydrochlorin), in green van-der-waals representation, binding an all α-structure motif. This type of binding motif takes a spread out from since the secondary structure elements are well spaced by regions of residues in loop structures. Below is an example of the binding motif **..H...H....H...H** (Table 3).

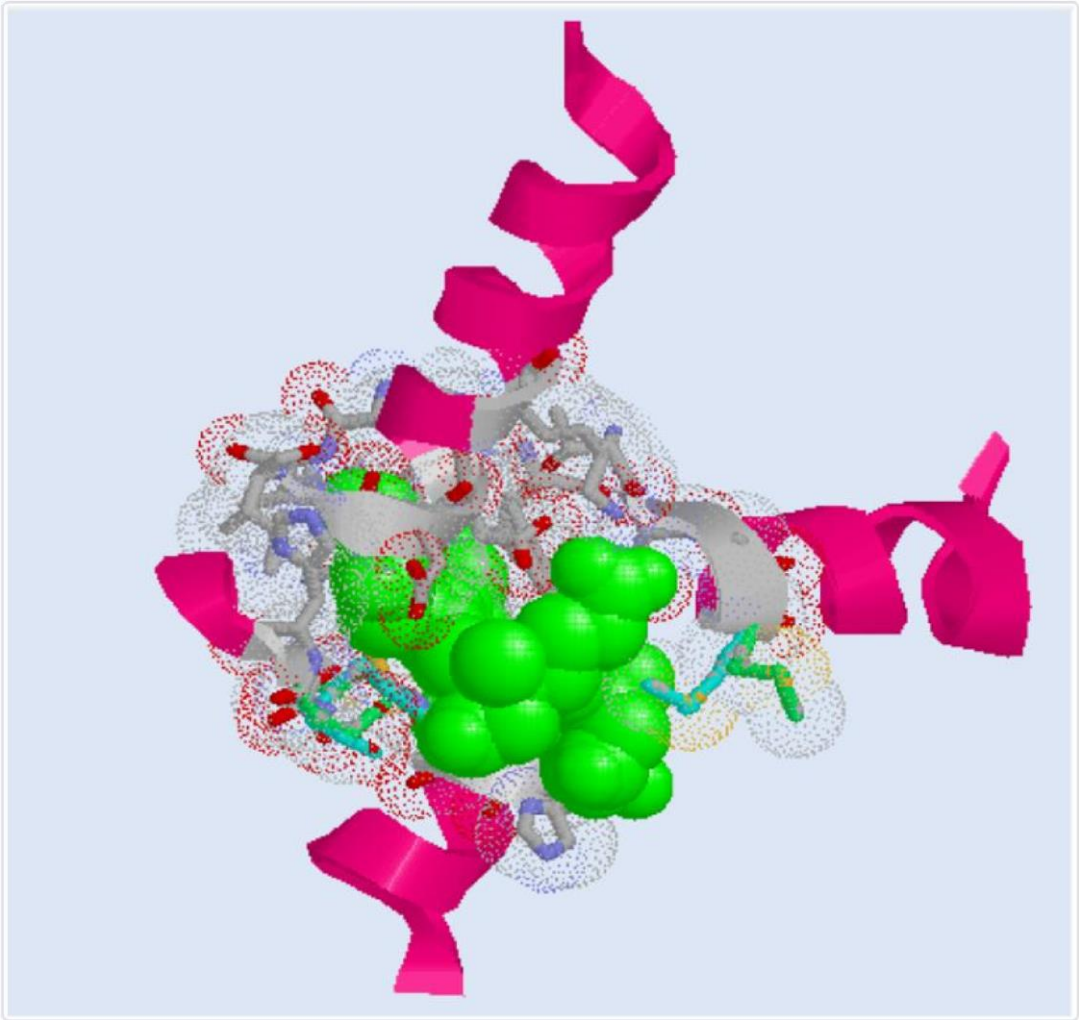


Figure 48. The Structural Motif: [..H...H....H...H] – binding the SIR group (green).The residues making the actual binding are FTSGMHIIGDEKHGASHMLVHA (PDB: 2xwp)

VIII.6. B12 (Cobalamin) group binding motif α/β-structure based (spread out form):

The graphical presentation, Figure 50, displays the B12 porphyrin group (Cobalamin), in green van-der-waals representation, binding an all α-structure motif. In this case too, the type of binding motif takes a spread out from since the secondary structure elements are well spaced by regions of residues in loop structures. Below is an example of the binding motif **S.H.H..H.....H...HS...S.S...** (Table 3).

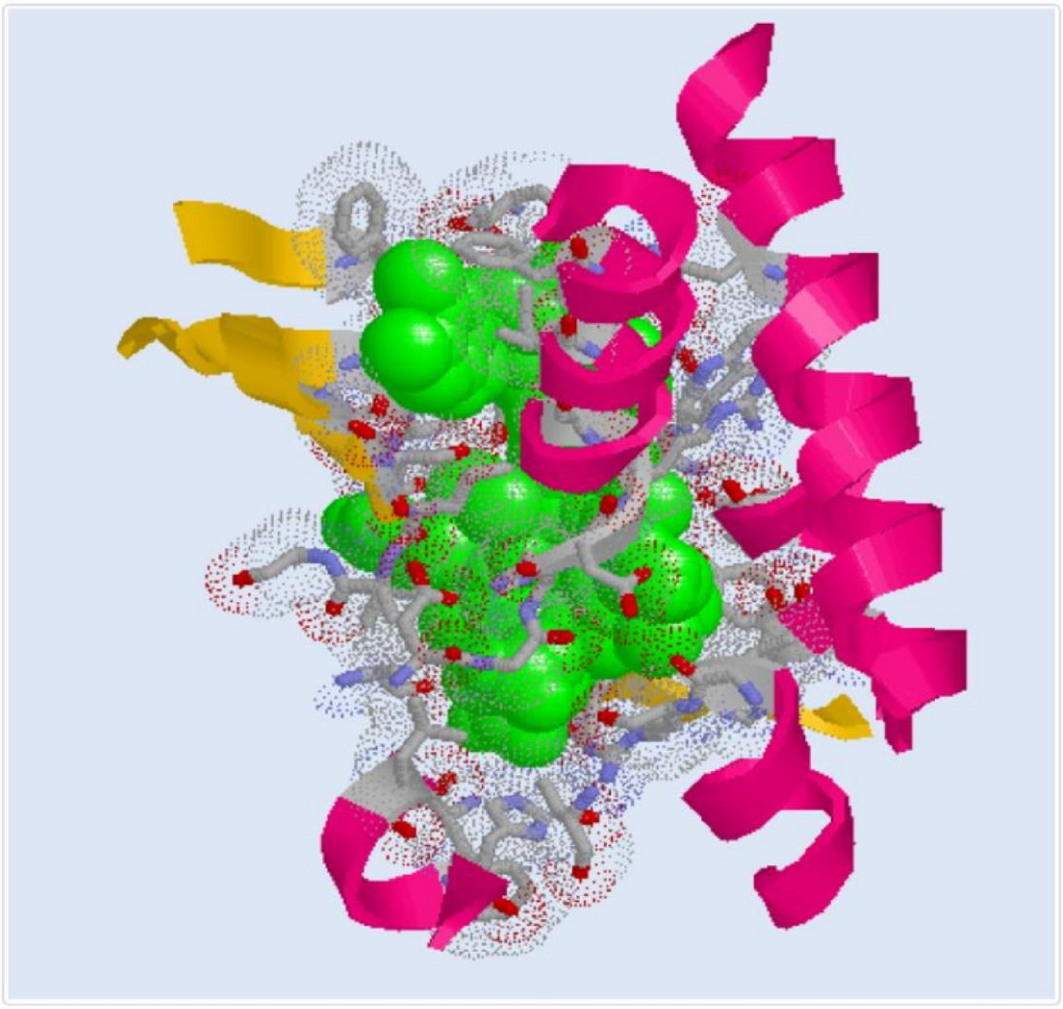


Figure 49. The Structural Motif: [S.H.H..H.....H...HS...S.S...] – binding the SIR group (green). The residues making the actual binding are YFLHAVRTYHEGWLEAGLQDGHDRGIFGSLAAGGVFGPT (PDB: 2XIQ)

General Conclusion

General conclusion

This project falls under the theme of Structural Bioinformatics and seeks to explore more the basis behind Structure-Function relationship in biological context of macromolecules; the proteins in the case of this study. Furthermore, the study draws attention to results that would touch upon distant evolutionary relations across species.

As revealed in the various analysis and deductions made in the Results and Discussions (Chapter III), this project has identified, defined and characterised a set of binding structural and functional motifs associated with a set of biologically important ligands/cofactors known collectively as Porphyrin groups. These ligands are relevant to vital biological function including oxygen transport, storage, light harvesting and energy production and more.

The project also identified the residues (amino acids) that are directly involved in the Porphyrin groups within the set of proteins selected in the study.

The protein structural elements (α-helices and β-strands) and loop regions that compose the structural binding motifs are considered, by this study, as providing important physical support on which the actual functional elements, i.e. the residues, are mounted, with their individual and collective physical and chemical properties, to carry out the specific biological function of the porphyrin proteins.

The discovery of structural similarity between the porphyrin binding motifs across distant species and functions is indicative of evolutionary relation between these types of proteins that use similar chemical groups such a the porphyrin planar cycles to achieve different biological functions. This would open further venues of research and discovery in this field of study.

The definition of the ligand binding sites, i.e. the binding structural motifs, and construction of a database accessible online and that provide such important data and analysis to researchers in the field would be very useful in deeper analysis of the protein function in health and pathological cases, in studies related to phylogenetic analysis, 3D-structure predictions and rational drug design.

However, such conclusions would better confirmed and further explored using larger data sets of Porphyrin proteins and the ligand types used by them. This is to be planned in future studies.

References

References

- Alberts B, Johnson A, Lewis J, Raff, M., Roberts, K., & Walter, P. (2002). Molecular Biology of the Cell. Garland Science.0-8153-3218-1

- Alderton WK, Cooper CE, Knowles RG (2001), "Nitric oxide synthases: structure, function and inhibition". The Biochemical Journal. 357 (Pt 3): 593–615.

- Aravind L, Anantharaman V, Balaji S, Babu MM, Iyer LM (Jul 2005),The many faces of the helix-turn-helix domain: transcription regulation and beyond. FEMS Microbiol Rev. 29(2):231-62. doi: 10.1016/j.femsre.2004.11.005.

- Berg, J.M., Tymoczko, J.L., Gatto Jr., G.J. and Stryer, L. (2015) Biochemistry. 8th Edition, W.H. Freeman & Company, New York, NY.

- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., ... & Bourne, P. E. (2000), The Protein Data Bank. Nucleic acids research, 28(1), 235-242.

- Blankenship,R.E.(2014),Molecular mechanisms of photosynthesis, 978-1-118-85568-0.

- Chothia, C., & Lesk, A. M. (1986),The relation between the divergence of sequence and structure in proteins. The EMBO journal, 5(4), 823-826.

- Chiancone, E., & Ceci, P. (2010),"The multifaceted capacity of cytochrome b5 in plants". Plant Signaling & Behavior, 5(1), 27-31.

- Eddy SR .(1998), Profile hidden Markov models. Bioinformatics, Volume 14, Issue 9, 1998, Pages 755–763.

- Fernández-Justel, D., Zbilut, J. P., & Roca, F. J. (2018),Flavohemoglobins: a colorful subfamily of hemoglobins with versatile biological functions. Journal of hematology & oncology, 11(1), 125. doi: 10.1186/s13045-018-0664-4.

References

- Golovin A., Dimitropoulos D., Oldfield T., Rachedi A. and Henrick "MSDsite., 2005, A Database Search and Retrieval System for the Analysis and Viewing of Bound Ligands and Active Sites. PROTEINS: Structure, Function, and Bioinformatics 58(1): 190-9.

- H. Tuppy, G. Kreil, (2013). in Encyclopedia of Biological Chemistry (Second Edition)

- Hofrichter, M., Kellner, H., Pecyna, M. J., Ullrich, R., & Scheibner, K. (2020),Fungal unspecific peroxygenases: heme-thiolate proteins that combine peroxidase and cytochrome P450 properties. In Handbook of Porphyrin Science (pp. 1-51). World Scientific Publishing.

- Hofrichter M, Ullrich R, Pecyna MJ, Liers C, Lundell T (2010), "New and classic families of secreted fungal heme peroxidases". Applied Microbiology and Biotechnology. 87 (3): 871–897.

- I. Grotjohann, P. Fromme (2013), Encyclopedia of Biological Chemistry (Second Edition). Pages 503-507.

- Janin, J. (1997),The third medium of life: Macromolecules. Oxford University Press.

- Järvi, S., Suorsa, M., & Aro, E. M. (2015),Photosystem II repair in plant chloroplasts—Regulation, assisting proteins and shared components with photosystem II biogenesis. Biochimica et Biophysica Acta (BBA) - Bioenergetics, 1847(9), 900-909).

- John Wiley & Sons; Dunford, H. B. (1999), "Heme peroxidases", 0-471-24244-6.

- Karplus, P. A., & Diederichs, K. (2012), Linking crystallographic model and data quality. Science, 336(6084), 1030-1033.

- Keskin O, Nussinov R. mars (2007),Similar Binding Sites and Different Partners: Implications to Shared Proteins in Cellular Pathways. The PubMed journal “structure”, 15(3):341-54.doi: 10.1016/j.str.2007.01.007.

References

- Khan, A. A., & Mao, X. O. (2018),Neuroglobin and Alzheimer's disease. Biochemicaland biophysical research communications, 497(2),733-737.doi: 10.1016/j.bbrc.2018.02.081.

- Laity JH, Lee BM, Wright PE(Feb 2001), Zinc finger proteins: new insights into structural and functional diversity. Curr Opin Struct Biol.;11(1):39-46. doi: 10.1016/s0959-440x(00)00181-9.

- LaPelusa A, Kaushik R. StatPearls [Internet](Nov 14; 2022) StatPearls Publishing; Treasure Island (FL). Physiology, Proteins. [PubMed]).

  Lehninger et al, Nelson, D. L., & Cox, M. M (2008),Lehninger principles of biochemistry. W.H. Freeman and Company.

- Lupas AN, Gruber M (2005),The structure of α-helical coiled coils. Advances in protein chemistry, 70: 37-78.

- Nelson, N., & Yocum, C. F. (2006),Structure and function of photosystems I and II. Annual Review of Plant Biology, 57, 521-565.

- Ortiz, de Montellano P. R. (2015) Cytochrome P450: Structure, Mechanism, and Biochemistry. 4th ed. Springer; New York.

- Peters, H. L., Sweetman, L., & Nyhan, W. L. (2021),Organic acidemias due to defects in the degradation of propionate and other organic acids. In Physician's Guide to the Diagnosis, Treatment, and Follow-Up of Inherited Metabolic Diseases (pp. 211-249). Springer.

- Richardson, J. S. (1981),The anatomy and taxonomy of protein structure. Advances in protein chemistry, 34, 167-339.

References

- Roger A. Sayle and E. James Milner-White. (1995), RasMol: Biomolecular graphics for all, *Trends in Biochemical Sciences* **20**(Sept):374-376.

- Scandalios JG (2005),"Oxidative stress: molecular perception and transduction of signals triggering antioxidant gene defenses". Brazilian Journal of Medical and Biological Research. 38 (7): 995–1014.

- Scott, A. I. (1998), Enzymatic synthesis of vitamin B12. Chemical reviews, 98(2), 491516. doi: 10.1021/cr960426r.

- Smith, K. M., & Ito, S. (Eds.). (2017).Porphyrins and related macrocycles: synthesis, spectroscopy, and potential applications. Royal Society of Chemistry.

- Tenhunen, R., Marver, H. S., & Schmid, R. (1969), "The enzymatic conversion of heme to bilirubin by microsomal heme oxygenase". Proceedings of the National Academy of Sciences, 64(3), 748-755.

- Trent JT, Hargrove MS (2002),"A ubiquitously expressedhuman hexacoordinate hemoglobin". The Journal of Biological Chemistry. 277 (20): 19538–19545.

- Vanderver, A. L., Wolff, J., & Milstien, S. (2021),Catalase: Biochemistry, molecular biology, and therapeutic implications. In Physiology and Pathology of Immunology (pp. 377-394). Academic Press.

- Via and Tramontano. (2011), Protein Structural Motifs: Identification, Annotation and Use in Function Prediction Sequence and Genome Analysis: Methods and Application II; ISBN: 9780980733051 .

- Welch, G. R. (1980),Horseradish peroxidase: A versatile tool in biology. Analytical Biochemistry, 106(2), 325-339. doi:10.1016/0003-2697(80)90118-5.

References

- Wiley-VCH 2011, "Porphyrin". Encyclopedia of Inorganic and Bioinorganic Chemistry DOI:10.1002/9781119951438.eibd0638.

- Nelson, D. L., Cox, M. M., & Lehninger, A. L. (2008), Lehninger principles of biochemistry; W.H. Freeman.

- Wolfgang Buckel, ... Oskar Zelder(2022), in Methods in Enzymology. DOI: 10.1016/bs.mie.2021.12.011 .

- Eddy SR .(1998)."Bioinformatics." Volume 14, Issue 9, pp. 755-763 . https://doi.org/10.1093/bioinformatics/14.9.755 .

- Hill, R., & Bendall, D. S. (2014),Photosystem I: Structure and function. Essays in biochemistry, 56, 41–61. https://doi.org/10.1042/bse0560041).

- PDB protein data bank: https://www.rcsb.org/pdb/home.

- https://www.rcsb.org/structure/1MBN.

- Sequence, Structure, Funtion Server (SSFS) https://www.bioinformaticstools.org/ssfs.

- PPI main page: https://www.bioinformaticstools.org/prjs/ppi.

- PPBSMS databaser https://bioinformatics.univ-saida.dz/prjs/ppbsms/.

EMBL-EBI Biomacromolecular structures. https://www.ebi.ac.uk/training/online/courses/biomacromolecularstructures/proteins/structural-motifs/.