Journal of **Concepts in Structural Biology & Bioinformatics**

STRUCTURE & FUNCTION ARTICLES

Research article: Master's research based.

Structural and Functional Binding Motifs in Porphyrin Proteins: Insights into Ligands and Biological Function

MEGHARBI Mohamed El-Mahdi[1], BITAR Mohamed[1] and RACHEDI Abdelkrim [*,1]

[1] Laboratory of Biotoxicology, Pharmacognosy and biological valorisation of plants, Faculty of Sciences, Department of Biology, University of Saida - Dr Moulay Tahar, 20100 Saida, Algeria.

[*]**Correspondence:** RACHEDI Abdelkrim – Email: abdelkrim.rachedi@univ-saida.dz, bioinformatics@univ-saida.dz

Published: 03 August 2023

Abstract

Understanding the underlying principles governing the relationship between protein structure and function is crucial for advancing our knowledge of biological functions, including those related to proteins and macromolecules in health and disease. Knowledge aspects generated in this work, in relation to protein ligand binding environment, would contribute in better understanding of biology mechanisms and is instrumental in the endeavours for designing novel drugs and developing biotechnology solutions.

The research presented in this article examines a collection of 3D structures of porphyrin proteins, encompassing 51 full chains from 21 entries in the Protein Databank (PDB), originating from various organisms and species. The investigation identified, constructed and characterised 43 unique structural and functional binding motifs. These motifs are associated with essential biological functions such as oxygen transport, storage, light harvesting, and energy production.

The findings and analysis obtained in this work have been annotated in a database named Porphyrin Proteins Binding Structural Motifs (PPBSMs). This database has been made available online, through the web server of the University of Saida, to providing a valuable resource for scientists and researchers worldwide.

Availability: The PPBSMs database is available freely here: https://bioinformatics.univ-saida.dz/prjs/ppbsms/

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

Introduction

The biological function of proteins is intricately linked to their three-dimensional (3D) structure, which determines their biochemical activities. Within molecules, specific arrangements or patterns of atoms, functional groups, or molecular fragments, known as structural motifs, have significant implications for the physicochemical properties, biological activity, and pharmacokinetic behavior of therapeutic compounds.

Structural motifs are essential for protein function as they facilitate interactions with substrates, cofactors, and other ligands necessary for various biological processes. For instance, the helix-turn-helix motif (Aravind L, 2005) is a vital structural motif involved in DNA binding and gene transcription regulation, highlighting the complexity of identifying such motifs based solely on amino acid sequences.

The field of structural biology continually expands our knowledge of these motifs, and this project aims to explore their role in porphyrin containing proteins (Smith, K. M., & Ito, S., 2017, Wiley-VCH, 2011), such as Hemoglobin, Cytochromes, and other functionally important proteins. Porphyrin proteins are critical for essential processes in living organisms, including respiration and cellular metabolism. Dysregulation of these proteins can lead to diseases, including cancer. Therefore, characterizing the binding motifs associated with their functionality would greatly contribute to our understanding of their function and the identification of potential novel drugs.

Porphyrin proteins exhibit various structural motifs crucial for their specific functions. For example, hemoglobin and myoglobin, which are heme binding proteins (Trent JT, Hargrove MS, 2002) containing an Iron ion, possess a "globular" domain responsible for oxygen binding and transportation. Additionally, they have an "α-helical" domain that provides structural stability. Cytochromes (Ortiz de Montellano, P. R., 2005, Chiancone, E., & Ceci, P., 2010), on the other hand, feature a "cytochrome" motif involved in electron transfer reactions. Cobalamin or vitamin B12 (Scott, A. I., 1998), important for DNA synthesis and nerve function, relies on the porphyrin ring system that binds a Cobalt ion. Furthermore, porphyrin groups binding Magnesium ions are integral components of Chlorophyll A, B, D, and bacterial chlorophyll, essential parts of Photosystem I (Hill, R., & Bendall, D. S., 2014) and Photosystem II (Järvi, S., *et. al.*, 2015) behind the photosynthesis and phosphorylation vital processes.

This research project aims to investigate and discover structural motifs associated with a range of proteins from different species that bind various types of porphyrin groups, including Heme and its derivatives, Chlorophyll types A, B, D, and F, and Cobalamin (Megherbi M. El.-M., Bitar M., 2023). By unraveling the relationship between these motifs and protein functionality, valuable insights can be gained, leading to potential advancements in drug discovery and a deeper understanding of pathological alterations in these proteins.

Materials and Methods

In order to realise the structural study of this project, the binding structural motif may be found in the Porphyrin proteins, structural bioinformatics methods involving database creation, data annotation and programming were employed. The required steps that we followed to achieve the goals from this study are summarised in the following major steps:

I. Protein Structures Identification and Data Preparation:

Experimentally determined protein structures of more than 2000 complexes are publically available through the Protein Data Bank – PDB (Berman *et. al.*, 2000). Such wealth of structural data provides the ground for fundamental research seeking the understanding of biological function and in health and disease situations.

The PDB assigns unique identifiers, PDB IDs, which are used to access structural data for each and every entry in the PDB. The PDB ID consists of four characters, and it is assigned when a new structure is deposited in the database.

Data preparation for Porphyrin proteins in complex with ligands, for the purpose of ligand (i.e. Porphyrin) binding motifs calculation, involved 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. Porphyrin Protein Structures:

The PDB entries used in this project amount to 21 structures each with its own PDB id selected based on being determined by x-ray crystallography method and of high resolution. This makes it easy for researchers to access and compare structural data for porphyrin-containing proteins from different sources. The 21 structures encompassed 51 full protein chains most bound to a one porphyrin ligand or more. **Table 1**, below, represents the list of porphyrin-protein classes and their protein names, PDB id, Resolution among other useful information.

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

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	2nwp	ANAEROBIC COBALT CHELATASE (CBIK) FROM SALMONELLA TYPHIMURIU COMPLEX WITH METALATED TETRAPYRROLE		1.9	19.7	Co2+	SIR	COBALT SIROHYDROCHLORIN

Table 1: List of the porphyrin-containing protein classes and structures 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, Porphyrin ligand name, Metal ion type used by the porphyrin ligand.

II. Identification and Processing of Ligand Binding Data:

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 at the Department of Biology by Dr. Abdelkrim Rachedi, **Figure 2**, has been used to carry out calculation of the ligand motifs environment details within a cut-off distance of ≤ 4.0 Å.

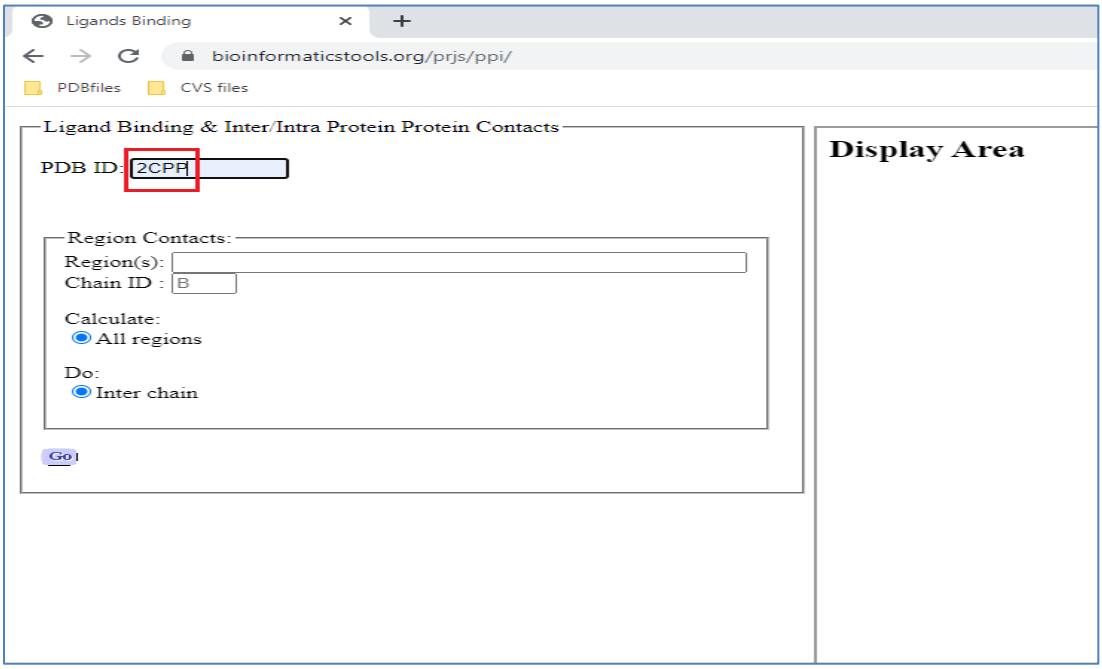


Figure 2. Capture of the interface of the **PPI** tool (https://bioinformaticstools.org/prjs/ppi/).

The PPI tool takes a PDB id, example 2CPP, which is the cytochrome P450 protein structure and upon clicking the **Go** button, Figure 2, a list of links to the binding environment of the ligands bound to the structure in question, Figure 3. The HEM group is the porphyrin ligand of the 2CPP structure. The details of the binding environment of the HEM group are generated through the link **'Environment'**, **Table 2**.

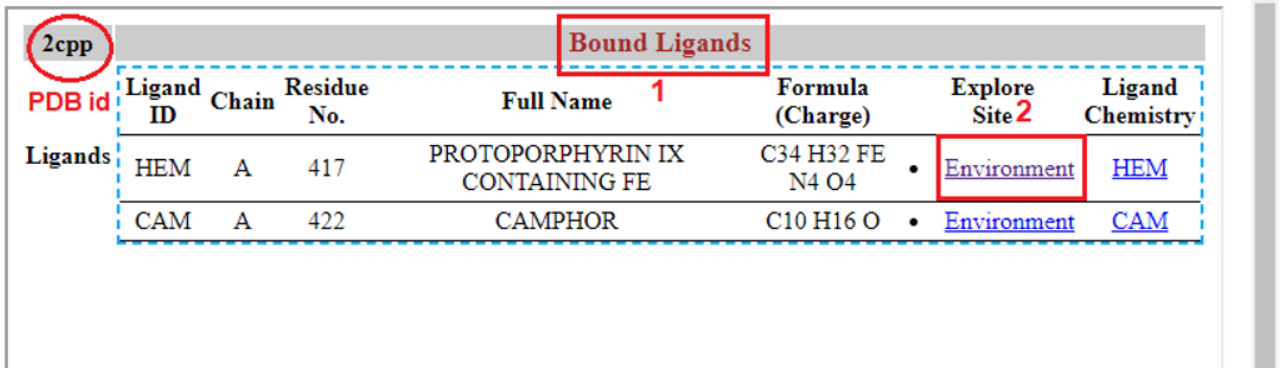


Figure 3. Capture of the following steps for the data extraction

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 organised in the following columns:

- **The columns under the title "Protein or NA Residues":** These columns show the atoms responsible for the binding of the enzyme amino-acids residues 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 responsible for the binding of the ligand HEM, its number and the ligand id.

- **The columns under the title "Bonds":** 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**.. etc

Using the PPI system, the binding environment details of all ligand associated with the 21 PDB porphyrin proteins were calculated, Figure 4. The data generated by the PPI, in CSV format, is downloaded and parsed using a PHP script used which in turn annotates the data and store it in a database.

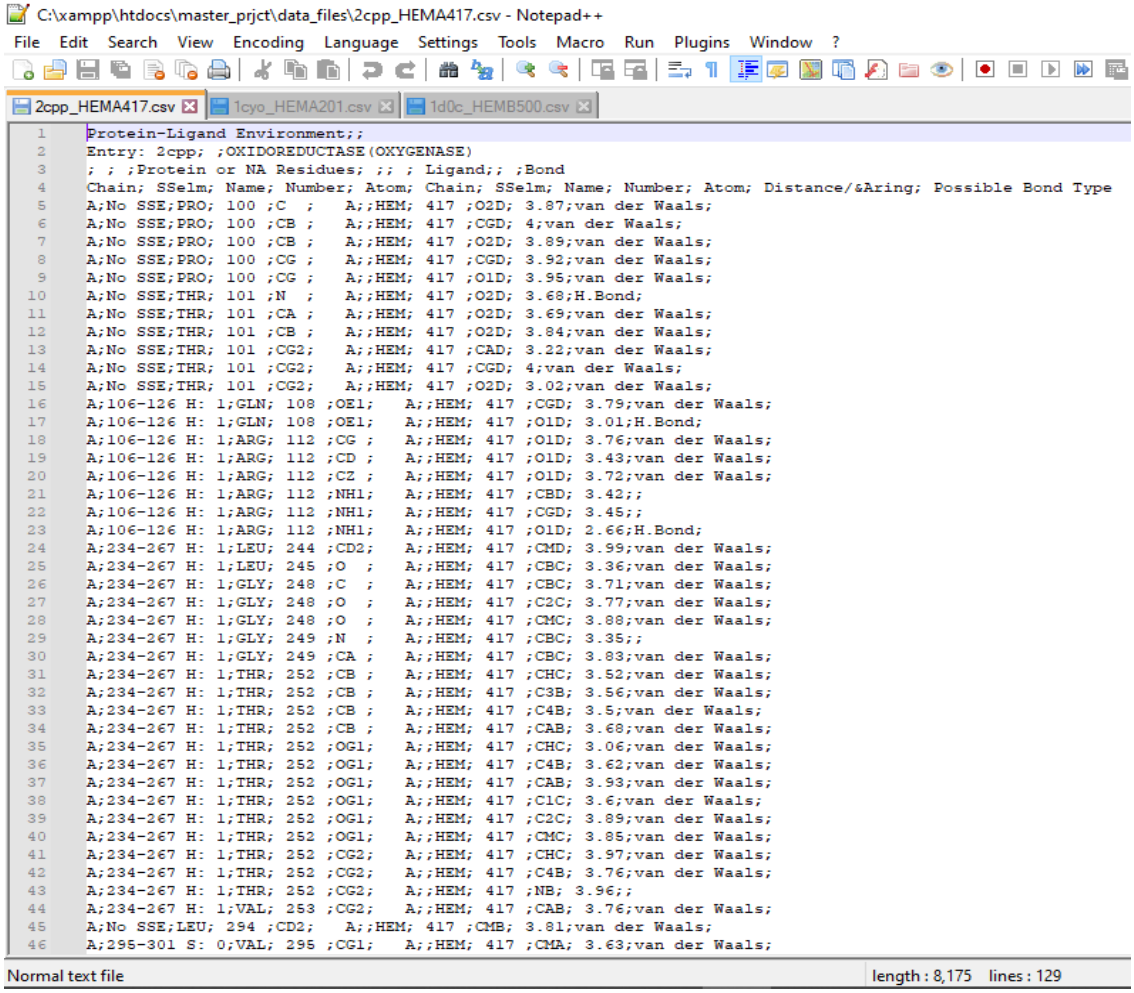


Figure 4. An example of a CSV file which contain the binding details of the protein residues with the HEM ligand and PDB id : 2CCP, chain A.

III. Construction of the Binding Motifs and Representation:

In the binding details presented earlier, the residues that interact with the ligand are associated with certain secondary structure elements represented in the column **SSelm**, in **Table 2**, which points at the association of the protein binding residues with specific regions that represent secondary structure elements. The annotation of the relevant data is illustrated by the example of 2CPP where the ligand HEM, chain A as explained in the following:

- **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 α-helix, designated by the symbol **H**.
- **The protein region labeled "295-301 S:0"** corresponds to the secondary structure β-strand, designated by the symbol **S**.

In this specific case of 2CPP, the pattern (motif) 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. Hence, this type of calculated data is further annotated, in this work, as Structural and Functional Binding Motifs shorten here as Binding Motifs.

These graphical representations, eg. **..HH.SS……H**, can be used to analyse the structure of these molecules and to better understand their function. Rasmol (Sayle R. A., Milner-White E. J., 1995), a molecular graphics tool, can display various types of secondary structure motifs, such as α-helices, β-strands, and loop regions, in a variety of graphical representations.

Three types of structural representation are adopted throughout the article:

α-helix representation (H): The α-helix is a common structural motif in proteins, consisting of a helical arrangement of amino acids. In Rasmol, α-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 α-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**.

β-strand representation (S): B-strands are another common structural motif in proteins, consisting of a sheet-like arrangement of amino acids. In Rasmol, β-strands can be represented as arrows or ribbons. The arrow representation shows the strand as an arrow-like structure, while the ribbon representation shows the strand as a twisted ribbon. To display a β-strand 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 α-helices and β-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.**

III.1. Representation of Binding Details in Global View:

Below is a Rasmol script, **Figure 5**, that produces a graphical representation of the motif **"..HH.SS.....H"** as shown in **Figure 6**.

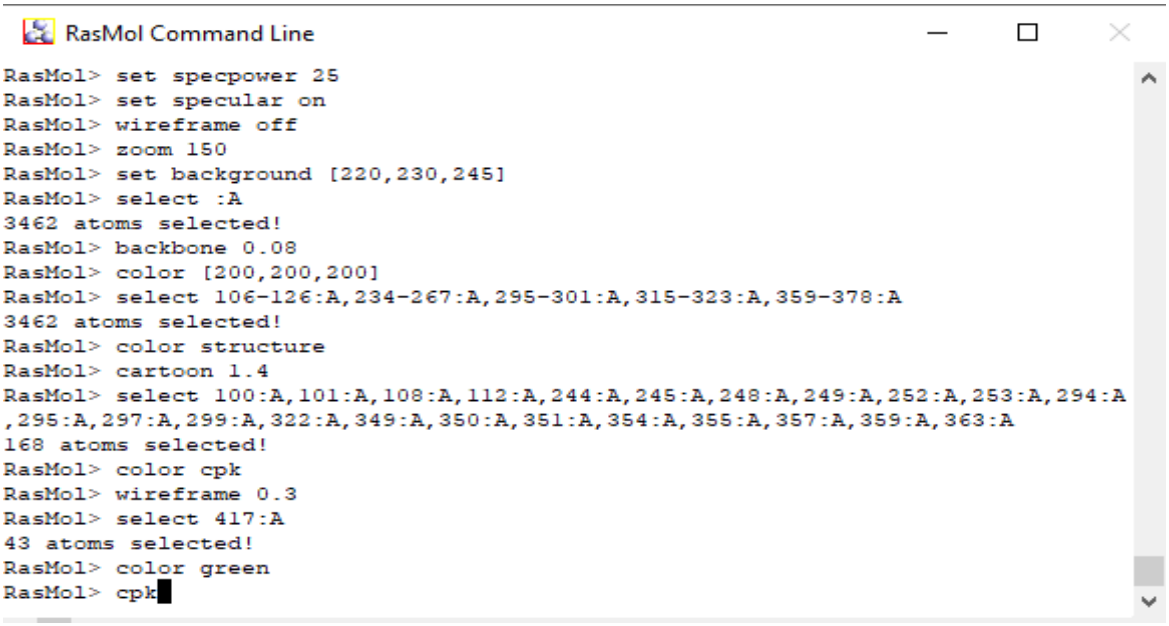


Figure 5. Capture of the RasMol script to create the general motif representation.

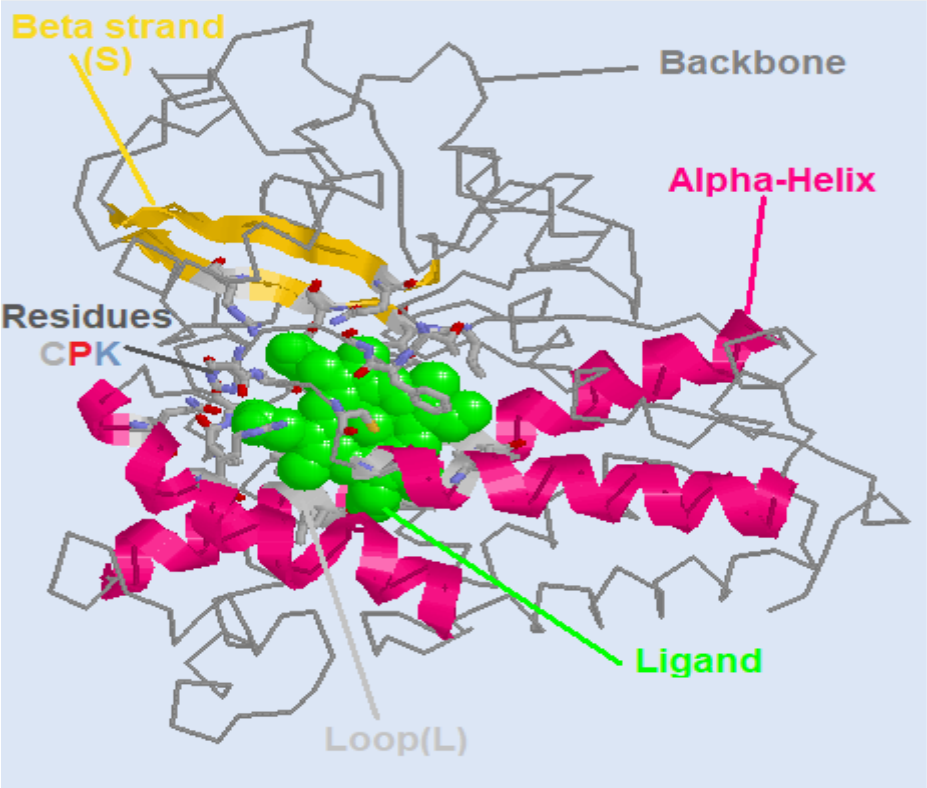


Figure 6. Capture of the global RasMol representation of the binding motifs. α–helices (red) and β–strands (yellow) with the interacting residues (CPK colours) and the porphyrin ligand HEM (green). Backbone of the whole protein cytochrome P450 (PDB id : 2CPP chain A) that show a gobal perspective.

III.2. Motif-Ligand only representation:

Below is a Rasmol script, **Figure 7**, that generates a graphical representation of the **"..HH.SS.....H"** motif with the ligand HEM visible with the binding residues, **Figure 8**.

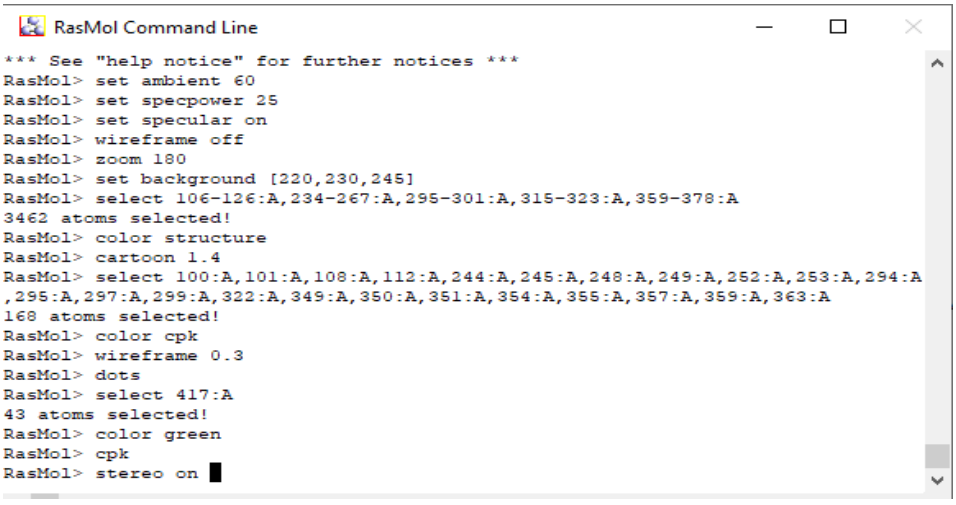


Figure 7. Capture of the RasMol script to create the motif representation.

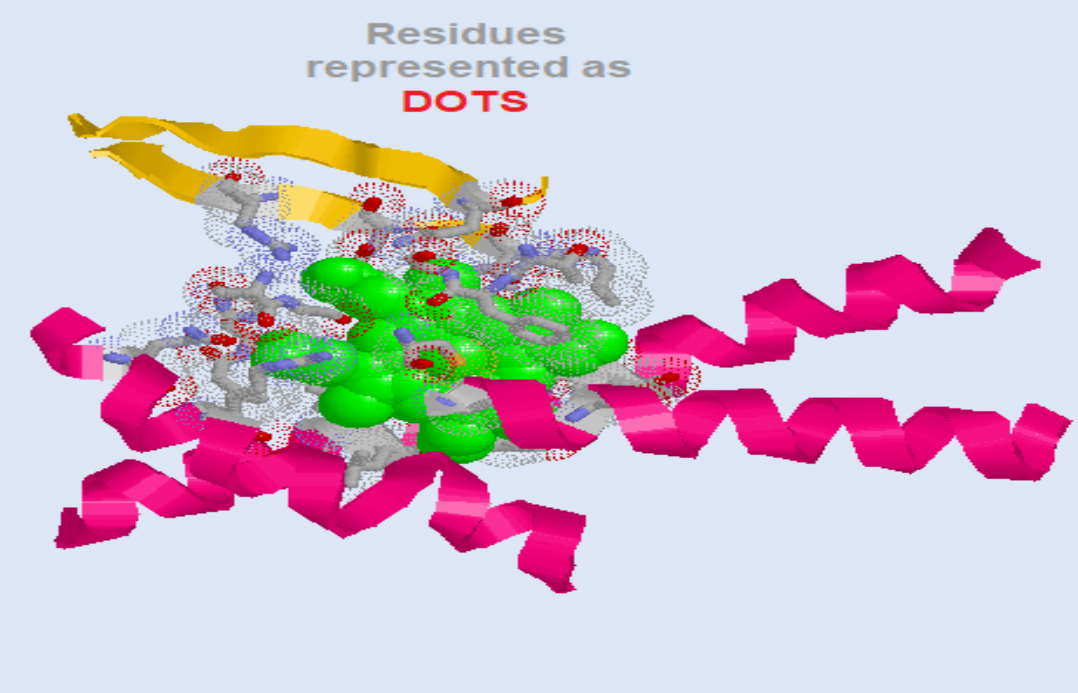


Figure 8. Capture of RasMol representation of the binding motifs where the residues represented as dots (around van-der-waals radius) for the case of cytochrome P450 (PDB id : 2CPP chain A).

Patterns representation of the binding motifs and Rasmol graphical sceense, seen above, have been done for all of the 21 structures including the relavant 51 chains of the porphyrin proteins studied in this project.

IV. Annotation of Binding Motifs and Creation of the PPBSMs Database:

The data generated in this research work including the constracted binding motifs, details of the interactions with the different types of porphyrin ligands, graphics representations have be annotated in a Flat-Files database named **P**orphyrin **P**roteins **B**inding **S**tructural **M**otif**s** (**PPBSMs**). For illustration purposes, only chain A of 6 PDB ids , are depicted for the rasmol scenes, **Figure 9**.

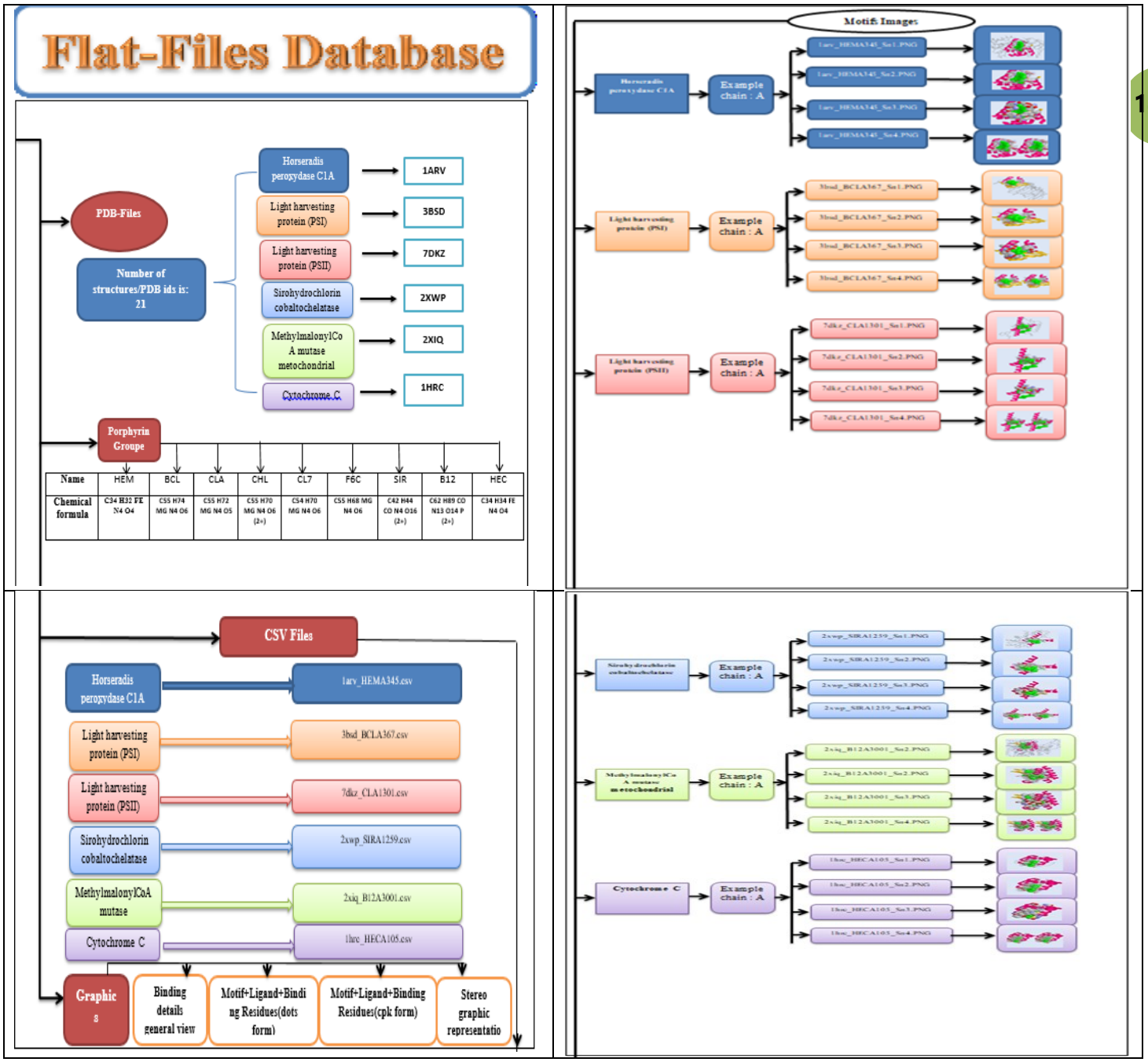


Figure 9. The database schema representing the architecture of the created Flat-File database; the **left side** show the arrangement and classification of the files containing the csv-files which contain the binding environment details while the **right side** of the figure show the arrangement of the 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. PPBSMs Database online access:

The **PPBSMs** has been mounted on the "Bioinformatics" service based on the University of Saida server, and PHP programming scripts have been developed to facilitate the tasks of the database querying. This is done to enable sharing of the data and results with the scientific community both locally and internationally.

The database is made freely accessible from the following web address: **https://bioinformatics.univ-saida.dz/prjs/ppbsms/ - Figure 12**, refer to next section.

Results

I.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 β-strands are represented with **S** and loop-regions are represented with '**.**', see **Figure 10** (B1). In addition, residues involved in the binding are presented with single letter code of amino-acids, see **Figure 10** (B2).

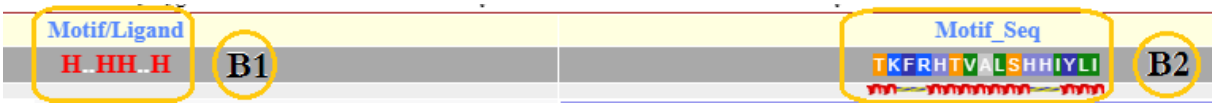
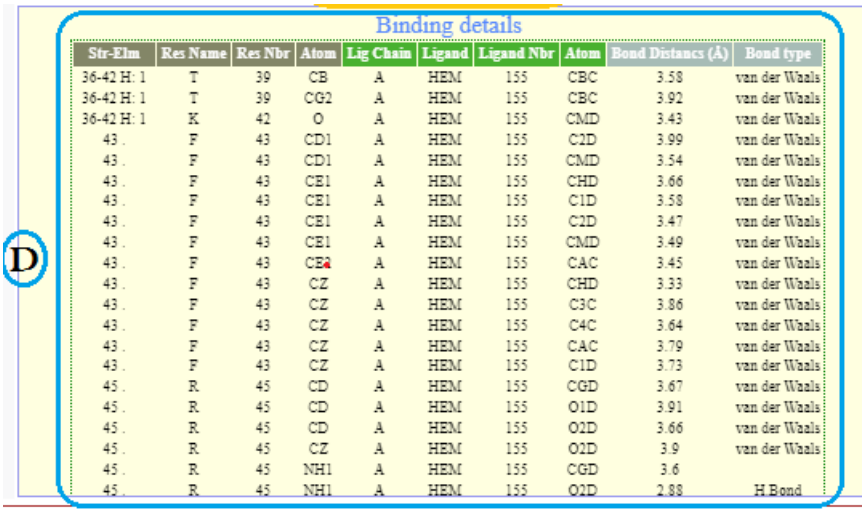


Figure 10. Structural binding linear prsentation **B1**: **H** denotes α-helix and '**.**' Denotes loop region. **B2**: Amino acids shown with approximation of their belonging to the secondary structure elements.

I.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 11**.



Binding details

Str-Elm	Res Name	Res Nbr	Atom	Lig Chain	Ligand	Ligand Nbr	Atom	Bond Distancs (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	CE2	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 11. Partial binding details of the secondary structure and loop elements. Estimated atomic bonding distances and types are shown.

Note. Both Figures 10 and 11 can be understood better in the context of the global figure, **Figure 13**, shown further below.

II. Presentation of results:

II.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.

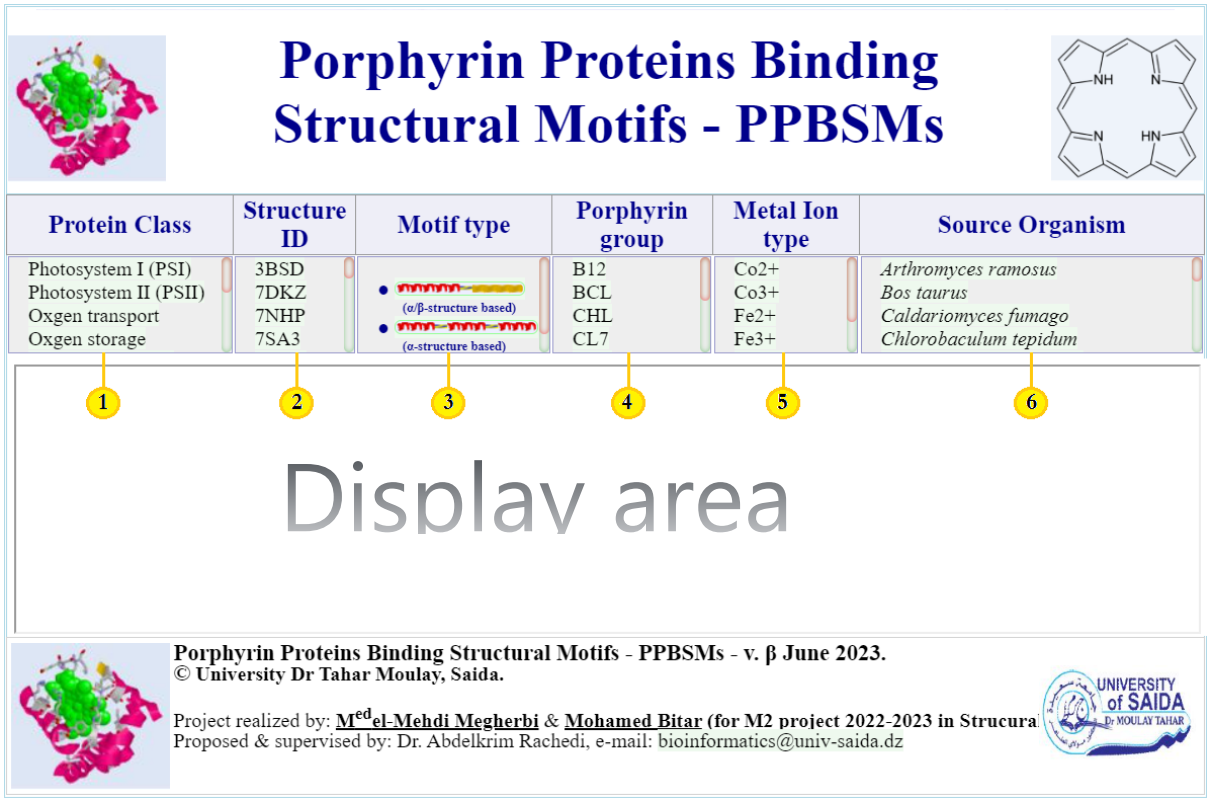


Figure 12. The main web interface of **PPBSMs** database as screen-captured from the web address, see next sections for explanations on the highlighted areas.

II.2 Database Methods of Querying and Results Display:

As shown above in **Figure 12**, 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. This list allows for querying by:

1. This list allows for querying data by **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 as shown in **Figure 13**.

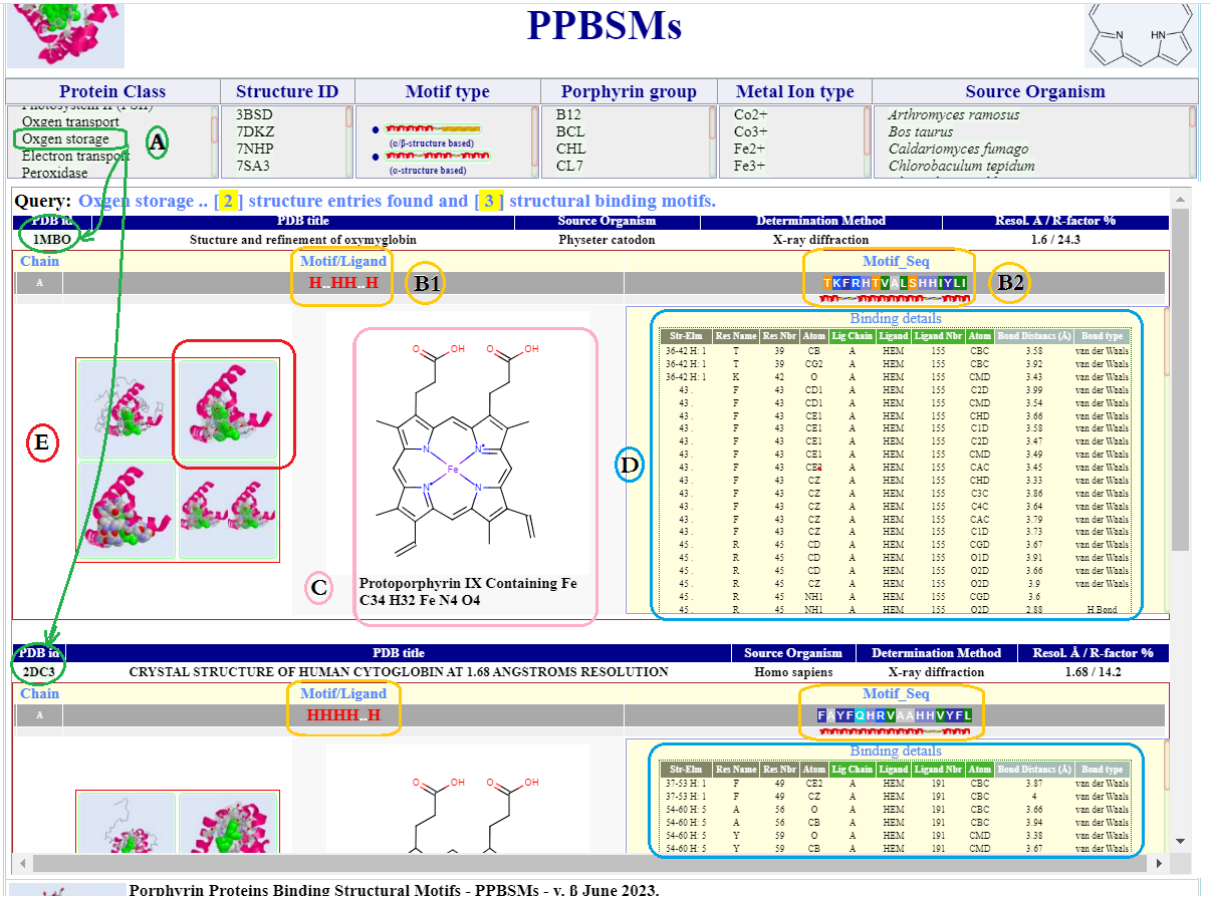


Figure 13. A screen-shot shows the six methods of searching **PPBSMs** database. The red highlighted areas represent the different types of results:

III. 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 unique binding motifs 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 "**II. Porphyrins Binding Tendency for Structural Motif type**"

II. 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 in **Table 3**, can be used to draw a number of preliminary characterisation of the motifs and extract a list of their properties:

II.1. Motifs Classification:

II.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).

II.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).

II.2. Motifs Binding Tendency:

II.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.

II.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 .

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

III. 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.

IV. 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.

V. Graphical Representation of the Binding Motifs:

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

The graphical presentation, **Figure 14**, 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** (refer to Table 3).

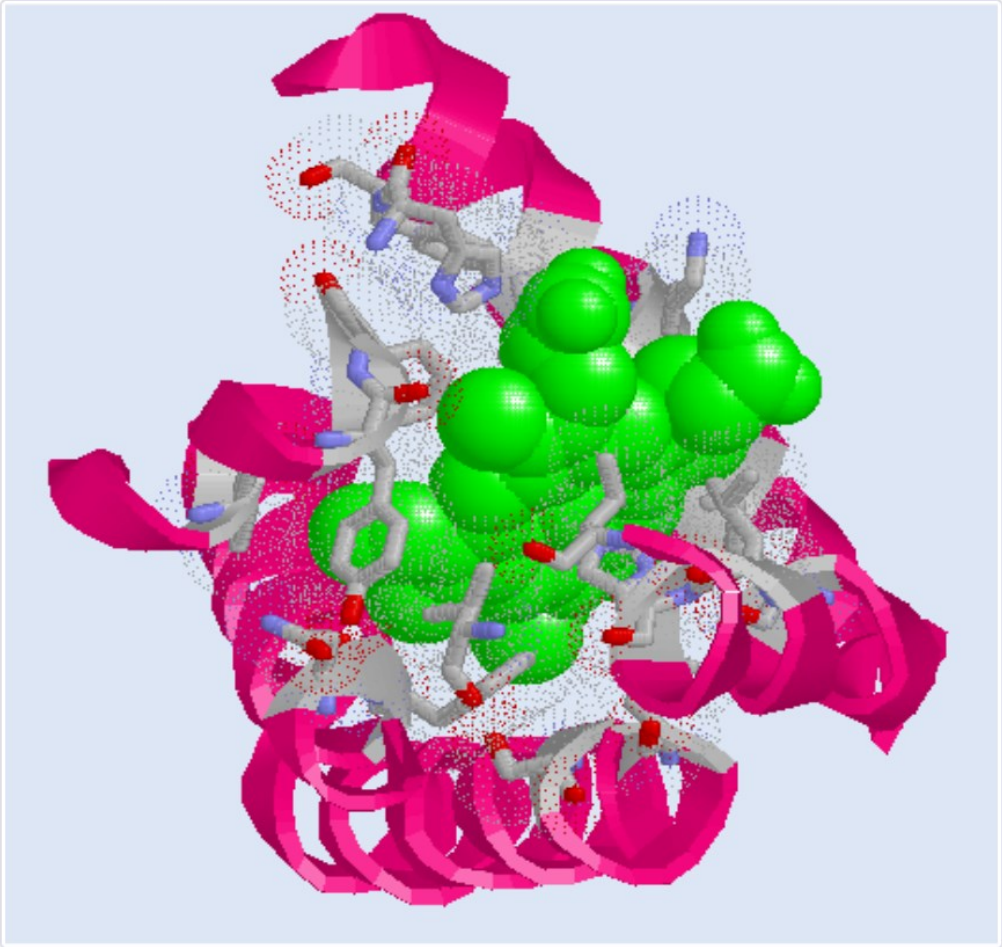


Figure 14. The Structural Motif: [**HHHHH..HH**] – binding the HEM group (green). The residues making the actual binding are LFFHKVAFLLHLVNFLL (PDB: 1GZX).

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

The graphical presentation, **Figure 15**, 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** (refer to Table 3).

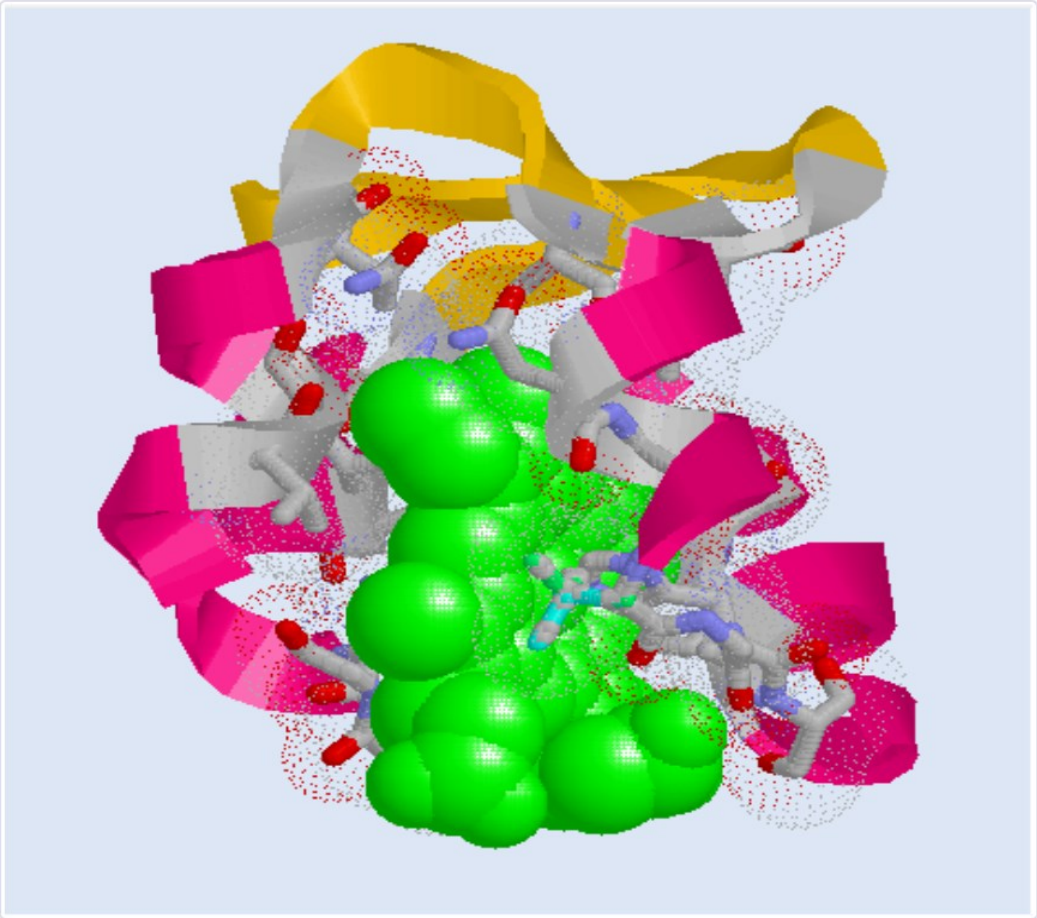


Figure 15. Structural Motif: [SSH…HSH….H] – binding the HEM group (green).The residues making the actual binding are LLFHPGVLQANFVGHSALS (PDB: 1CYO).

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

The graphical presentation, **Figure 16**, displays the BCL porphyrin group (Bacteriochlorophyll A), in green van-der-waals representation, binding an α/β-structure motif. Below is an example of the binding motif **SSSHH.S** (refer to Table 3).

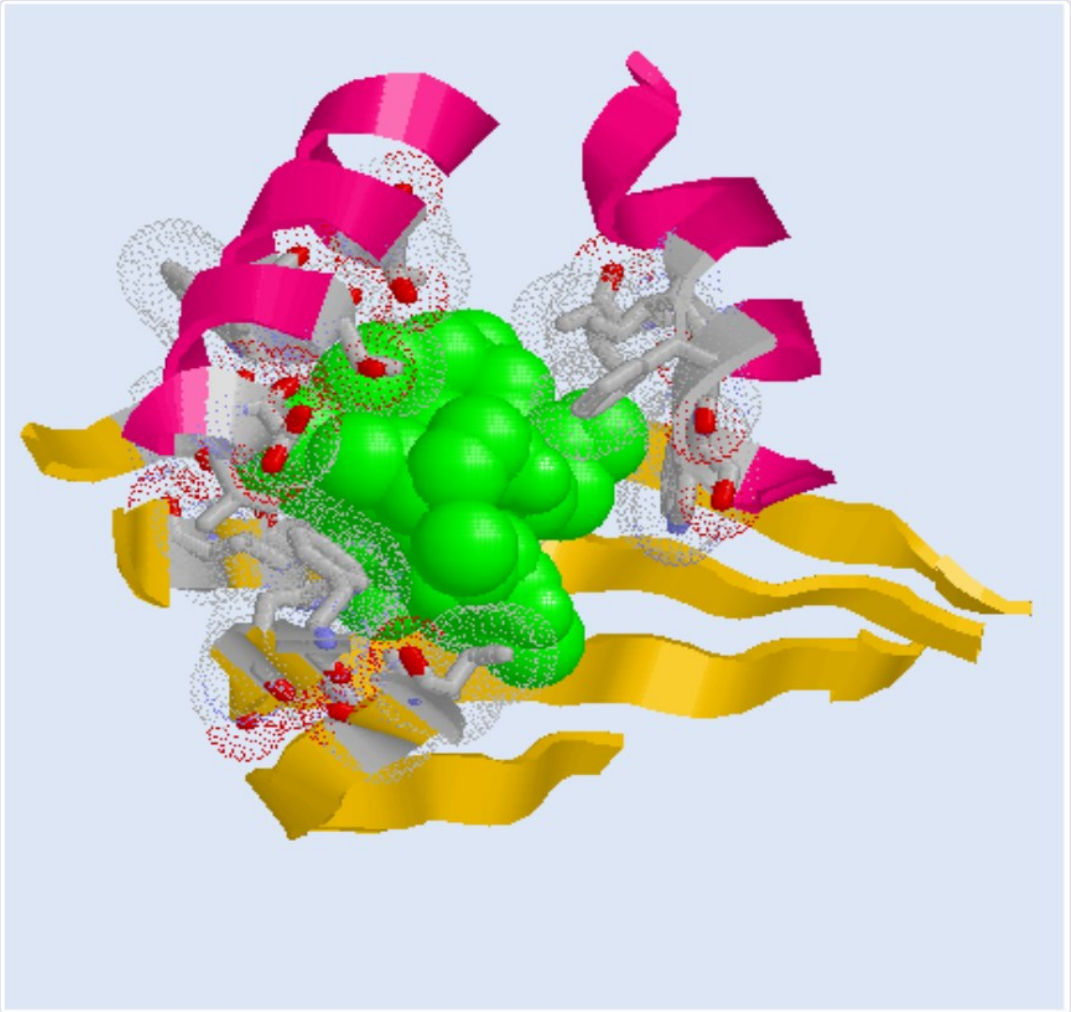


Figure 16. The Structural Motif: [SSSHH.S] – binding the BCL group (green).The residues making the actual binding are VVFHFMVDLTWTIFWIGSW (PDB: 3BSD).

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

The graphical presentation, **Figure 17**, 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** (refer to Table 3).

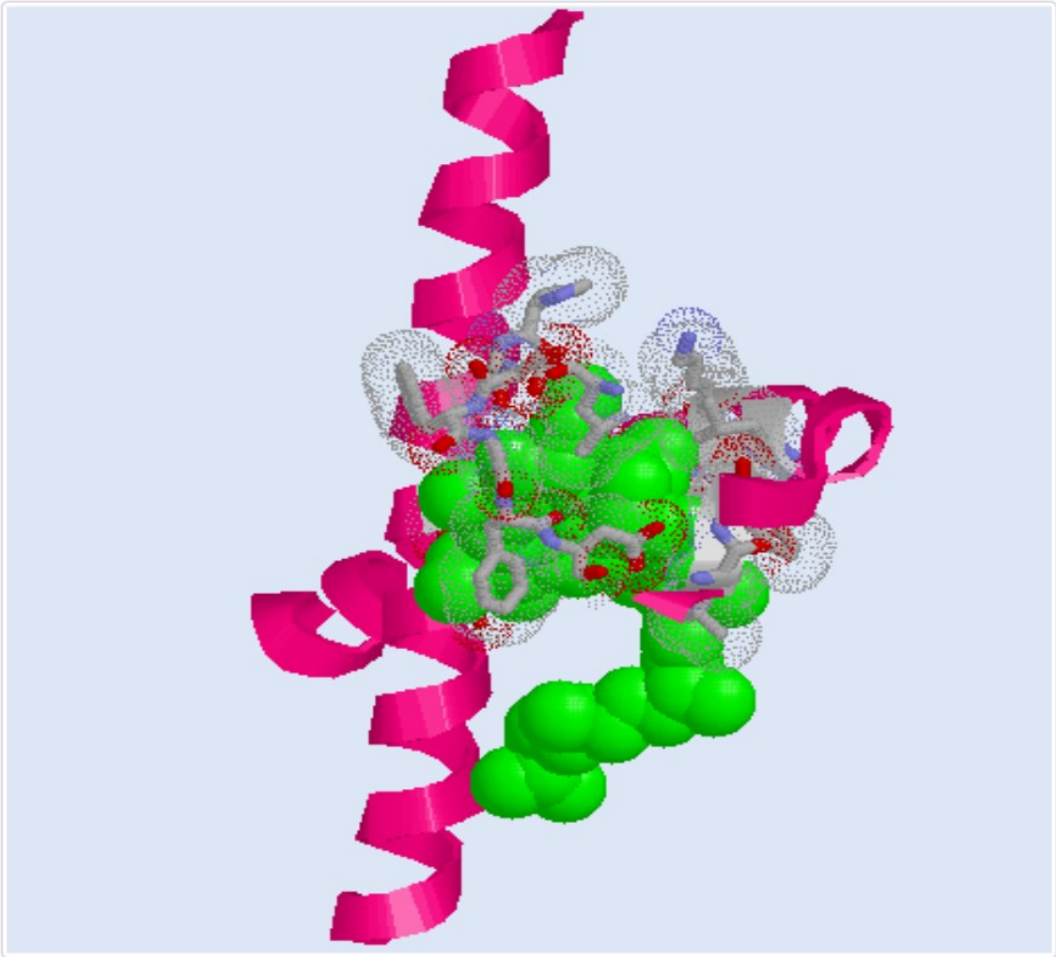


Figure 17. The Structural Motif: [.....**HHH**] – binding the BCL group (green).The residues making the actual binding are L P G D F G F D L G L F K S E H R L (PDB: 7DKZ)

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

The graphical presentation, **Figure 18**, 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** (refer to Table 3).

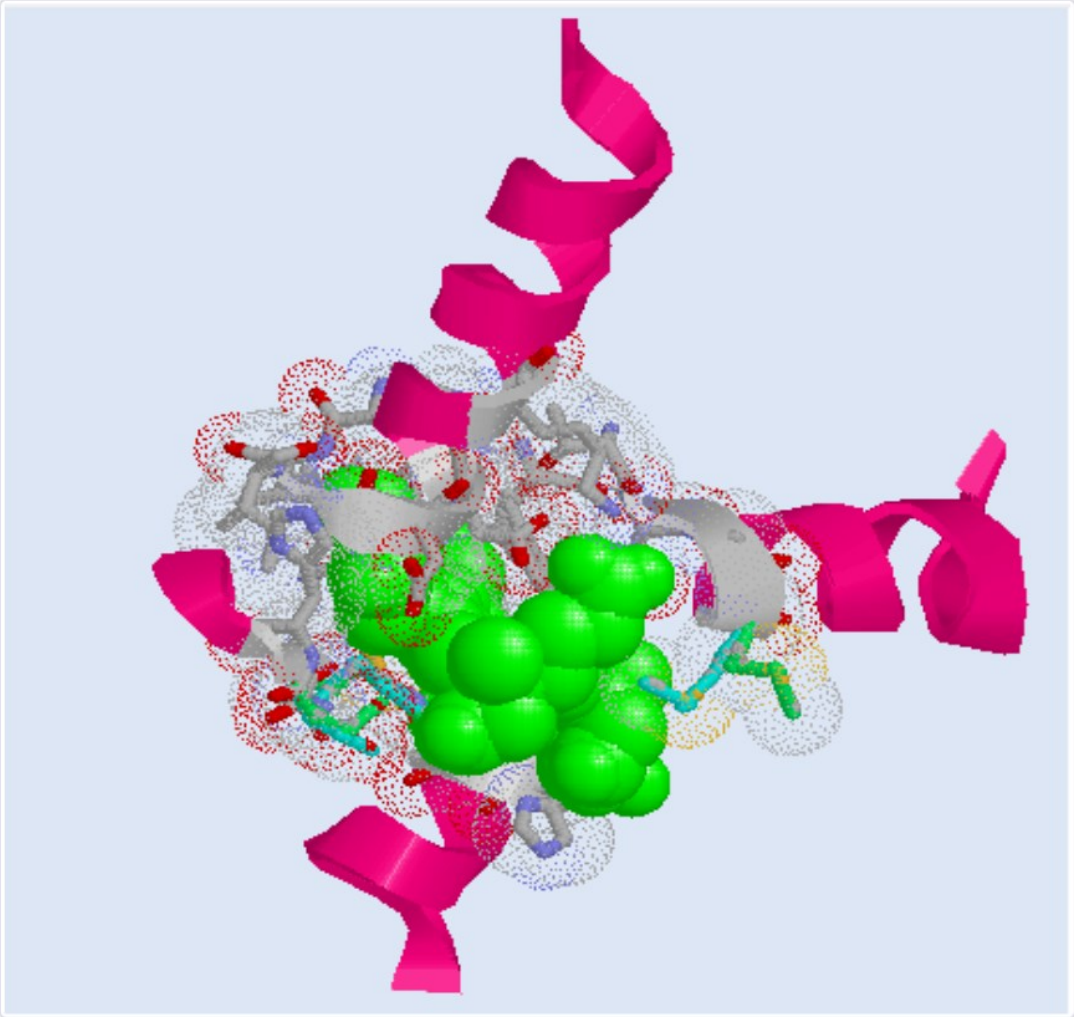


Figure 18. The Structural Motif: [..**H**...**H**....**H**...**H**] – binding the SIR group (green). The residues making the actual binding are FTSGMHIIGDEKHGASHMLVHA (PDB: 2xwp)

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

The graphical presentation, Figure 19, 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...** (refer to Table 3).

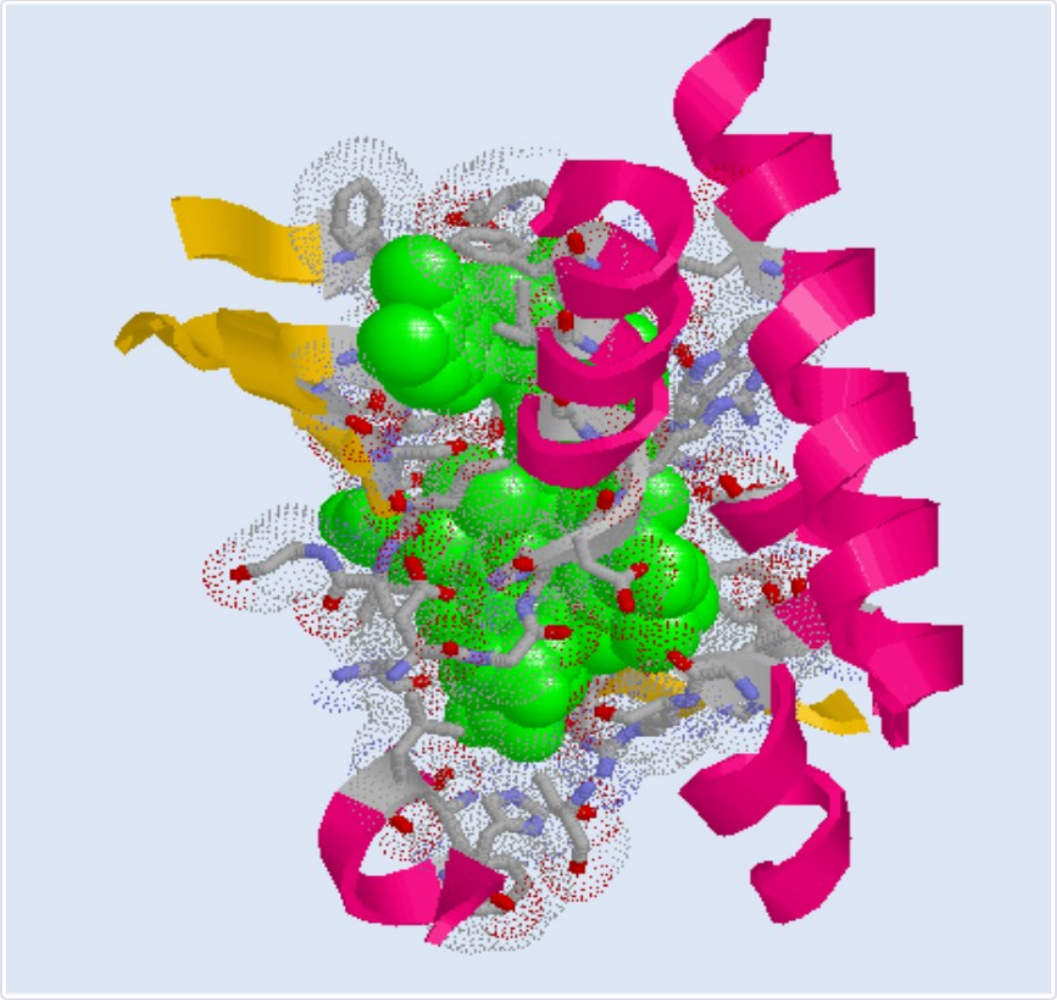


Figure 50. The Structural Motif: [S.H.H..H.....H...HS...S.S...] – binding the SIR group (green). The residues making the actual binding are:
YFLHAVRTYHEGWLEAGLQDGHDRGIFGSLAAGGVFGPT (PDB: 2XIQ)

Conclusion

This investigation research 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.

Future directions

The conclusions made above would be better confirmed and further explored using larger data sets of Porphyrin proteins and the ligand types used by them. Moreover, future work would include studies of the nature of the sequence of the motifs' amino-acids and explore the effect of their consecutive arrangement in the binding of the porphyrin ligands.

References

- 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.
- 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.
- Chiancone, E., & Ceci, P., 2010, "The multifaceted capacity of cytochrome b5 in plants". Plant Signaling & Behavior, 5(1), 27-31.
- Hill, R., & Bendall, D. S., 2014, Photosystem I: Structure and function. Essays in biochemistry, 56, 41–61. https://doi.org/10.1042/bse0560041).
- 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).
- Megherbi M. El Mehdi, Bitar M., 2023, 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/). Master Thesis, Library of the University of Saida.
- 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., 2005, "Cytochrome P450: structure, mechanism, and biochemistry". Springer Science & Business Me.
- Smith, K. M., & Ito, S. (Eds.), 2017, Porphyrins and related macrocycles: synthesis, spectroscopy, and potential applications. Royal Society of Chemistry.
- Scott, A. I., 1998, Enzymatic synthesis of vitamin B12. Chemical reviews, 98(2), 491-516. doi: 10.1021/cr960426r.
- Sayle, R. A., Milner-White E. J., 1995, RASMOL: biomolecular graphics for all, Trends Biochem Sci. Sep;20(9):374. doi: 10.1016/s0968-0004(00)89080-5.

- Trent JT, Hargrove MS, 2002, "A ubiquitously expressed human hexacoordinate hemoglobin". The Journal of Biological Chemistry. 277 (20): 19538–19545.
- Wiley-VCH, 2011, "Porphyrin". Encyclopedia of Inorganic and Bioinorganic Chemistry.