People's Democratic Republic of Algeria

The Ministry of Higher Education and scientific research

FACULTY OF SCIENCES

University Dr. Moulay Taher

Department : Biology

Memory

Presented for obtaining the Master's degree in: Biology

Specialty: Biochemistry and Cell Physiology

THEME:

Research on Possible Structural & Functional Motifs in Amino Acids Degradation and Urea Cycles

URL: http://bioinformaticstools.org/prjs/uadfms

Prepared by:

- Raimes Mira
- Bessaih Aicha

In front of the jury commission, composed by

President:	Mr. Berrezoug Halimi	Univ. Dr. Moulay Taher- Saida.
Examiner:	Dr. Yahia Nasrellah	Univ. Dr. Moulay Taher.-Saida
Supervisor:	Dr. Abedlkrim Rachedi	Univ. Dr. Moulay Taher- Saida

2015/2016


Acknowledgment

We thank our great God for all he gives us .

Firstly a big thank you to our mentor Dr. RACHEDI Abdelkrim for his guidance and valuable advice that helped us a lot in achieving this final assignment study.

We thank all who helped us during the work on this project.

To all the professors who contributed to our training and education, we would like to express our deep respect which is extended especially to the members of the jury for agreeing to examine this project.

Our thanks also go to all our colleagues in master 2 Biochemistry and Cellular Physiology and exclusively to Ould Ahmed Salem Abdellahi and Mebarka .


Dedication

I dedicate this modest work with recognition and respect to my dear parents the symbols of love and tenderness.

To my sister: Afafe Salma .

To my brothers: Abdellatif, Abelssabor, Abdelghafor , Abdelbari

To all my family.

To all my professors.

To all my friends .

To my partner: Raimes Mira.

Not to mention my Colleagues in master 2 Biochemistry and Cellular Physiology 2016.

Aicha

Dedication

I dedicate this modest work in recognition and respect to my

dear parents symbols of love and tenderness .

To my sister: Karima.

To my brothers: Kada, Elhadj, Mohamed.

To all my family.

To all my professors.

To all my friends.

To my partner: Bessaih Aicha .

Not to mention my Colleagues in master 2 Biochemistry

and Cellular Physiology 2016.

Mira

Table of Contents

Table of Contents:

Table of Contents

Table of Contents

List of Abreviations :

ARG: Arginase.

ARG: Arginine.

AS1 : Argininosuccinate.

ASAT: Aspartate Aminotransferase .

ASL : Argininosuccinate Lyase .

ASP: Aspartic Acid.

ASS: Argininosuccinate Synthetase .

CATH: **C**lass ,**A**rchitecture ,**T**opology and **H**omologuos.

CIR: Citrulline.

CPS I: Carbamoyl Phosphate Synthetase I.

GDH: Glutamate Dehydrogenase.

GLU: Glutamic Acid.

HAR: N-Omega-Hydroxy-L-Arginine .

LGB : Ligand Binding Tool.

NET: Tetraethyl Ammonium Ion .

NMR: Nuclear Magnetic Resonance.

NNH: Nor-N-Omega-Hydroxy-l-Arginine.

OTC: Ornithine Transcarbamoylase.

PAO: N-(phosphonoacetyl)-L-Ornithine.

PDB: Protein Data Bank .

PL P : Pyridoxal-5'-Phosphate.

SCOP: Structural Classification of Protein .

SSFS : Sequences Structures Function Server.

URL : Uniform Resources Locator.

List of figures

List of figures

List of figures

List of tables

List of tables :

List of tables

List of tables

Indexes

.

Summary:

This project has been undertaken in order to carry out a study in the fundamentals behind the Structure and Function relationship of proteins.

The understanding of such a relationship is important in the discovery of the biological function of proteins in both of normal and pathogenic situations. One way to undertake this kind of study is to analyze the enzymes involved in metabolic pathways and examine their ligands binding environment.

A set of enzymes involved in the Amino Acids degradation and Urea cycles has been selected for the study presented in this project.

This study necessitated the use of structural data that represent three-dimensional structures of the enzymes in complex with their ligands (and/or analogues). The structural data related to the enzymes involved in the metabolic cycles can be extracted from the international database known as the Protein Data Bank or the PDB.

16 protein/enzyme structures extracted from the PDB have been analyzed in this study using the techniques of structural bioinformatics. This has led to the discovery of a set of secondary structure configurations named in this study as **Structural & Functional Motifs** which are deemed to be important in the process of binding the ligands by the enzymes and hence in the biological function of the enzymes.

This study has identified, defined and characterized the **Structural & Functional Motifs** associated with the enzymes understudy and their ligands. The ligand binding details and graphical representation has been stored in a Flat-Files database.

To share the data and results with the scientific community at the local and international levels, the Flat-Files database has been uploaded into an online database and made available on the Internet through the following web address:

http://bioinformaticstools.org/prjs/uadfms

Key words: Proteins, Enzymes, Ligands, Structural & Functional Motifs, PDB, Amino Acids, Urea, Structure, Function, Databases, Structural Bioinformatics.

Résumé :

Ce projet a été entrepris dans le but de réaliser une étude sur les fondamentaux derrière la relation entre la structure et la fonction des protéines.

La compréhension d'une telle relation est importante dans la découverte de la fonction biologique des protéines dans les deux situations normales et pathogènes. Une façon de réaliser l'étude de la fonction biologique des enzymes impliquées dans les voies métaboliques est d'examiner l'environnement de liaison des ligands de ces enzymes.

Un ensemble d'enzymes impliquées dans la dégradation et Urée cycles Acides aminés a été sélectionnée pour l'étude présentée dans ce projet.

Cette étude a nécessité l'utilisation des données de structure qui représentent les structures tridimensionnelles des enzymes dans un complexe avec leurs ligands (et / ou analogues). Les données structurelles liées aux enzymes impliquées dans les cycles métaboliques peuvent être extraites de la base de données international connu sous la Protein Data Bank ou PDB.

Les structures tridimensionnelles de 16 protéines / enzymes extraites de la PDB ont été analysés dans cette étude en utilisant les techniques de la Bioinformatique Structurelle. L'étude a conduit à la découverte d'un ensemble de ce qu'on appelle ici comme **Motifs Structurales & Fonctionnels** qui sont considérés d'être importants dans la liaison des ligands par les enzymes et donc importants dans leur fonction.

Ce travail a identifié, défini et caractérisés ces **Motifs Structurales & Fonctionnels** associés aux enzymes de cette étude et leurs ligands. Les calcules et les détails de liaison des ligands et leurs représentation graphique ont été stocké dans une base de données de type Flat-Files.

Pour partager les données et les résultats avec la communauté scientifique aux niveaux local et international, une base de données en ligne a été crée et mis à disposition sur l'Internet à l'adresse web suivante:

http://bioinformaticstools.org/prjs/uadfms

Les mots clés : Protéine, Enzyme, Ligand, PDB, Acides Aminés, Motifs Structurales & Fonctionnels, Urée, Fonction, Base de données, Bioinformatique Structurelle.

الملخص:

تم تنفيذ هذا المشروع من أجل إجراء دراسة في الأسس وراء العلاقة بين بنية و وظيفة البروتينات.

فهم هذه العلاقة مهم في اكتشاف الوظيفة البيولوجية للبروتينات في كل من الحالات العادية والمرضية.
أحد الطريق التي يمكن انتهاجها لدراسة هذا الموضوع لفهم الوظيفة البيولوجية هو دراسة الأسس الهيكلية لكيفية ارتباط المواد المتفاعلة أو الليجندات مع الإنزيمات المشاركة في المسارات الأيضية.

ولهذا فقد تم اختيار دراسة مجموعة الإنزيمات المسؤولة عن هدم الأحماض الأمينية وعن دورة اليوريا وتقديمها في هذا المشروع.

هذه الدراسة استدعت استخدام التراكيب ثلاثية الأبعاد للإنزيمات وهي في حالة إرتباط مع ليجنداتها الطبيعية أو نظائرها.

يمكن الحصول على البيانات ثلاثية الأبعاد للإنزيمات مرتبطو بليجنداتها من قاعدة البيانات الدولية للتراكيب الفراغية والمعروفة بـ Protein Data Bank أو الـ PDB.

لقد تم تحميل التراكيب الفراغية لـ 16 بروتين من قاعدة البيانات أعلاه والتي تم تحليلها باستخدام تقنيات المعلوماتية الحيوية الهيكلية الأمر الذي سمح باكتشاف بنيات هيكلية سميت في هذه الدراسة بـ **الوحدات الهيكلية الوظيفية** والتي يمكن أن تكون أساسية في عملية إرتباط الليجندات بالإنزيم والتي بالتالي تكون مهمة في محاولة فهم العلاقة بين التركيب الفراغي للإنزيمات ووظيفها.

أهم مساهمة قام بها هذا المشروع هو تعريف و تحديد ثم وصف هذه **الوحدات الهيكلية الوظيفية** المكتشفة في الإنزيمات تحت الدراسة. تفاصيل إرتباط الليجندات والتمثيل البياني لها تم تخزيها في قاعدة بيانات من نوع **الملفات المبسوطة** أوFlat-Files database

بغرض تبادل ومشاركة البيانات والنتائج مع المجتمع العلمي على المستويات المحلية والدولية، تم تحميل قاعدة بيانات الملفات المبسوطة في شكل نسخة متاحة على شبكة الإنترنت من خلال العنوان التالي:

http://bioinformaticstools.org/prjs/uadfms

الكلمات المفتا حية:

PDB، الأحماض الأمينية، الوحدات الهيكلية الوظيفية، بروتينات، إنزيمات، التركيب الفراغي، لجيندات، الوظيفة البيولوجية، المعلوماتية الحيوية الهيكلية، قواعد البيانات.

General Introduction:

The biochemical function of a protein is generally dictated by the 3D structure of the polypeptide chain.

The research in this work is focused on examining the relationship between structure and function of enzymes which catalyzes specific reactions in the amino acids degradation and urea cycles.

Enzymes in complex with ligands related to the said metabolic pathways have been structurally studied using the structures available from the Protein Data Bank – PDB.

Analysis performed in this project included calculating the binding environment details between the enzymes and their ligands has led to the identification, definition and characterization of a set of structural elements composed of secondary structures and loop regions which hold the residues responsible for the binding of the ligands and thus the carry out of enzymes function. Based on such assumptions, these identified elements which are referred to, in this thesis, as the Structural and Functional Motifs since they seem to reoccur based on the types of enzymes and the function s responsible for.

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 and nucleic acids.

This project is distributed as follows:

First chapter; which is a literature review thorough the concepts of metabolism, enzymes, proteins, amino acid degradation, urea cycle, the PDB, classification of protein structures.

Second chapter describes the Materials and Methods used to study and analyze the data pertaining to theme of the project.

In the 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**.**

I. Generality on the protein:

Proteins are linear chains of covalently connected molecules called amino acids. Their sequences are encoded in DNA segments called genes.

Proteins participate in almost all activities that take place within an organism and perform a huge variety of functions. Some of them are **enzymes** that catalyze biochemical reactions, and metabolism. Others have structural or mechanical functions, such as the proteins of the cytoskeleton, which form a system of scaffolds to maintain a cell's shape. Proteins are also important in processes of the so-called immune response, in cell adhesion, cell signaling, and in the cell cycle. [1].

I.1.Protein

I. 1.1.Amino Acids

In general terms, an amino acid is a molecule containing both amine and carboxyl functional groups.

In biochemistry, what really go under the name of amino acids are only the 20 standard natural amino acids (see figure 1.). With the exception of Proline, these all adhere to the same template, including an α-carbon to which the amine and the carboxyl groups and a variable side-chain are bonded. What drives the folding process and thus leads to the final three-dimensional structure of the protein are the different physicochemical properties of the side-chains. The amino acids in a protein are linked by peptide bonds formed in a dehydration reaction. For this reason, proteins are often also called peptides (or polypeptides if they are particularly long). [1].

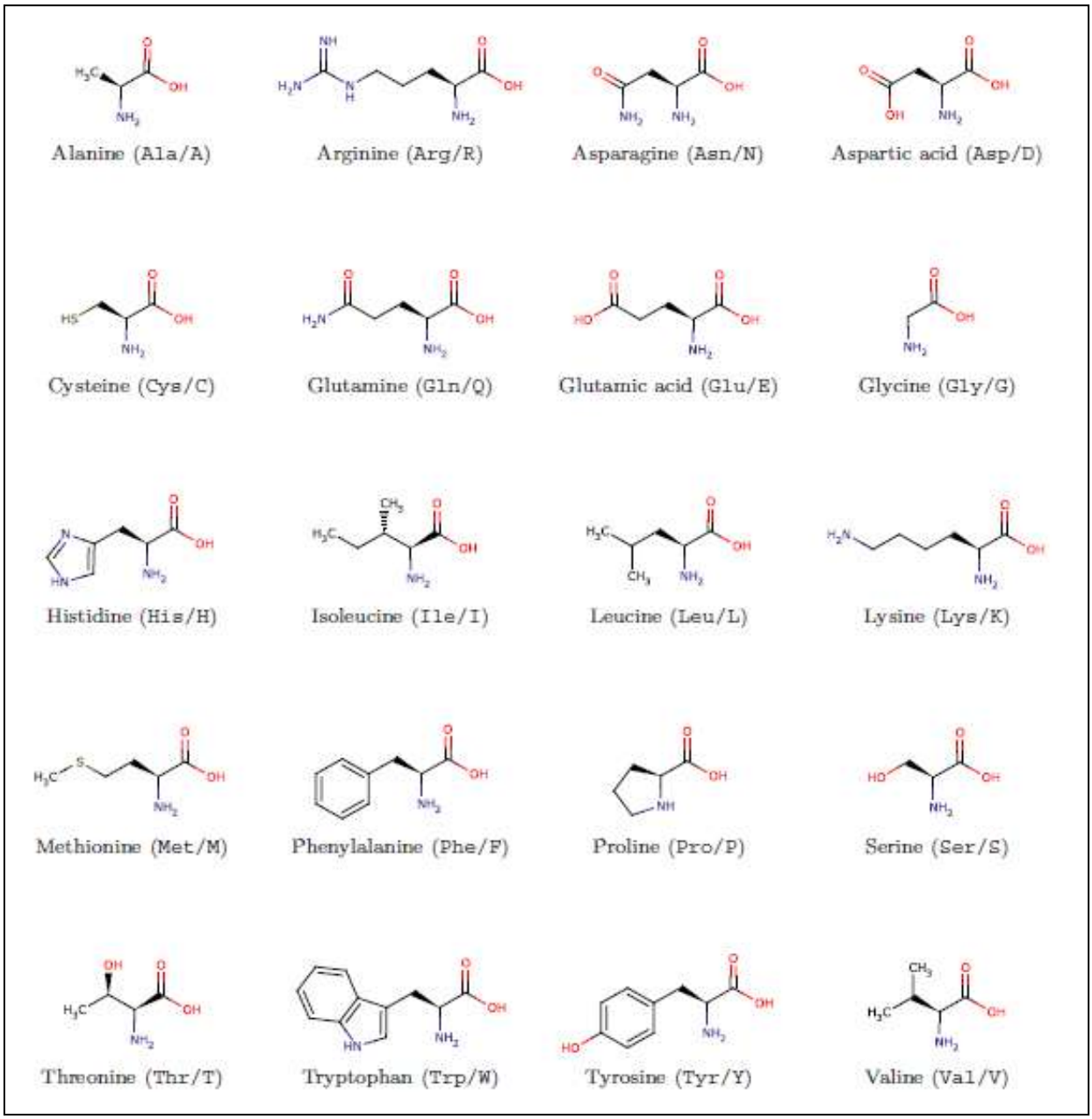


Figure n° 1: The 20 standard natural amino acids in their skeletal representation.

In parentheses are respectively their three-letter and one-letter codes. As can be seen, proline deviates from the scheme the other amino acids adhere to, in that its N-end nitrogen is involved in an unusual ring with the side-chain. This, incidentally, makes proline technically an imino acid rather than an amino acid.

I. 1.2. Protein Structure:

.

Proteins have several different levels of organization. They become highly organized and efficient biological machines through many types of ionic and molecular interactions within the protein itself.[2]

Most proteins fold into unique three-dimensional structures, which appear to be determined by their primary structure, that is, by the sequence of amino acids actually composing them. Assembled together in the native three-dimensional protein structure, the amino acids enlisted in the primary structure organize themselves in regularly recurrent local structural motifs mostly stabilized by means of hydrogen bonds. The most common examples of such structural motifs are alpha-helices and beta-strands The local arrangements of a polypeptide chain are collectively called secondary structure, while the way in which the polypeptide chain (eventually locally organized in secondary structure domains) finally folds in the three-dimensional space is called tertiary structure. The latter is generally stabilized by non-local interactions, most commonly by the formation of a hydrophobic core, but also through hydrogen bonds, disulphide bonds and salt bridges. Finally, in many cases, two or more polypeptide chains, called in this context protein subunits, can form larger complexes, which then constitute what is commonly regarded as the protein's quaternary structure. [1]

I. 1.2.1. Primary structure:

The primary structure is the sequence of amino acids constituting the polypeptide chain:

$R_1R_2 \ldots R_n$. [3]

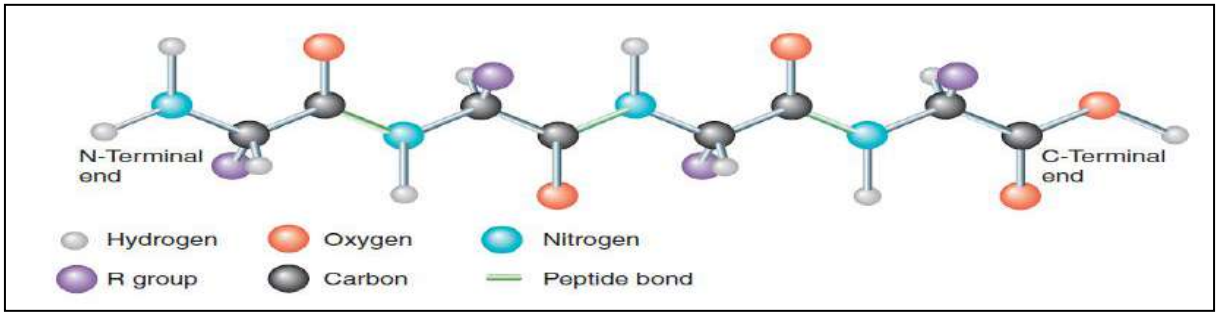


Figure n° 2: The primary structure of a protein.

I. 1.2.2. Secondary structure :

The secondary structure represents the local conformation of the polypeptide chain. Three main types of secondary structures are found: α-helices, β-sheets and loops.

I. 1.2.2.1.α-helix :

α-helix is stabilized with hydrogen bonds between the C=O group in the main chain of residue i and the N−H group in the main chain of residue i+4. In such a regular structure, all residues are involved in hydrogen bonds. Generally, there are two other kinds of bonding though they are much less frequent. The 3.10 -helices and π-helices are characterized by hydrogen bonds between residues i and i+ 3, and between residues I and i+5, respectively.

An α-helix is geometrically considered as a chain of periodic tours which correspond to a 5.4°A translation along the helix axis. Each tour contains, on average, 3.6 amino acids, thus the amino acids are translated 1.5°A along the axis. The structure of an α -helix is illustrated in Figure n°3. [3]

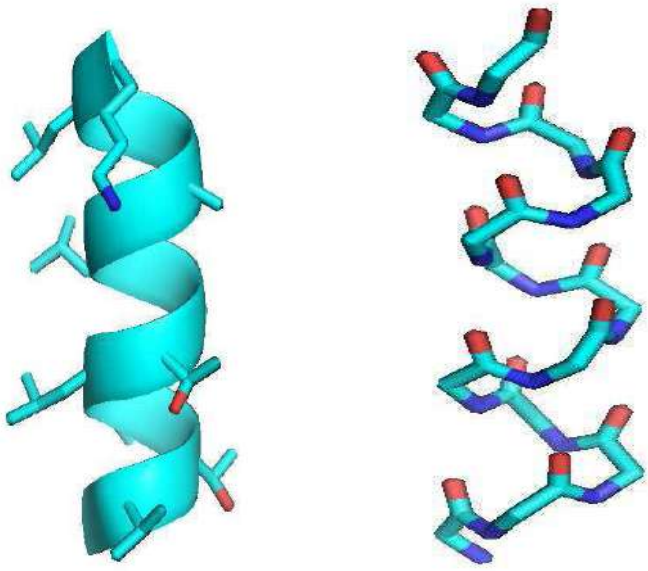
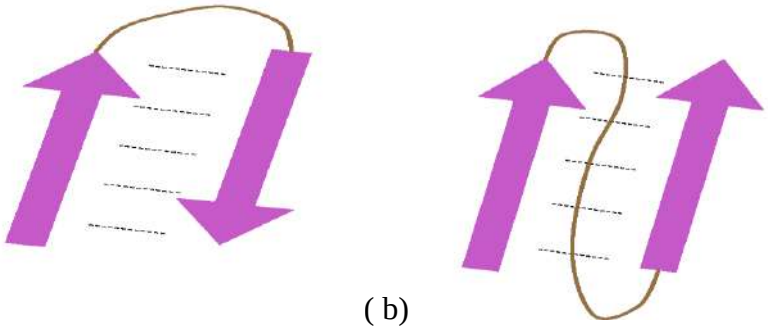


Figure n°3: Structure of an α-helix.

I. 1.2.2.2. β-sheet :

A β-sheet is composed of β-strand subunits. A β-strand can be considered as degenerated helix with 2 amino acids per tour. Each strand interacts with its neighbors through hydrogen bonds between the C=O and N−H groups in the main chains. As in helices, all residues in a regular β-sheet are involved in hydrogen bonds. This bonding associates the β-strands to each other, making the β-sheet stable. [3]

β-sheets are separated into two types regarding whether the constitutive β-strands are: parallel ; antiparallel, which is determined by the direction of the pairing β-strands (see Figure n°4). The β-sheet structure generated by antiparallel pairing is found more frequently than the one with parallel pairing, as the former is naturally more stable thanks to a better arrangement of residues. [3]



(a) (b)

Figure n° 4: Antiparallel pairing (a) and parallel pairing (b) of β-strands.

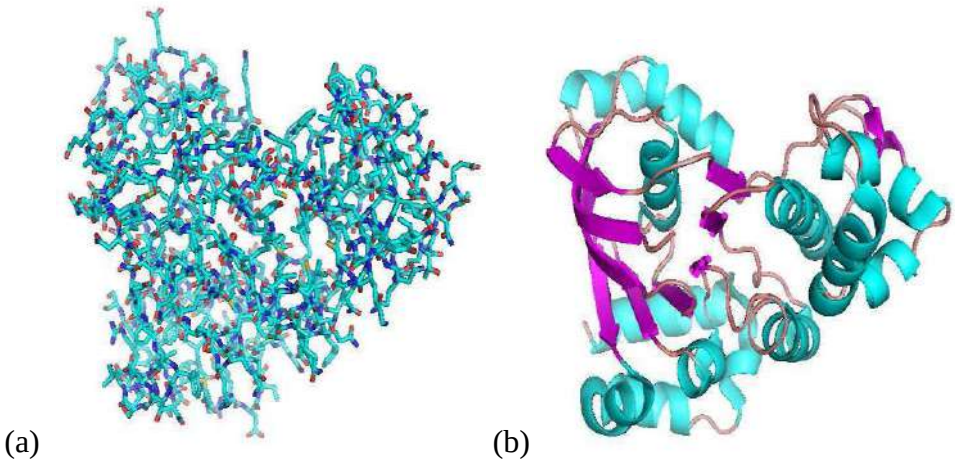
I. 1.2.2.3. Loop:

Loop, also known as random coil, is the all other category of secondary structure. In general, loops are not structured in the way that α-helices or β-sheets are; they are the portion of the protein that resembles “cooked spaghetti”. They can be flexible or rigid, and usually serve as connectors between α-helices and β-strands.

Sometimes, the term “secondary structure” is used to refer to the portions of the protein for which the secondary structure is structured: α-helix and β-sheet. In particular, the term secondary structural elements refers to the non-loop regions of the protein. [4]

I. 1.2.3.Tertiary structure:

The tertiary structure is the tridimensional conformation of the polypeptide chain, i.e. the relative coordinates of all atoms constituting the protein. This level of structure is essentially stabilized by hydrophobic interaction. There is a considerable difference on the precision of description between secondary and tertiary structures. Hence, the super secondary structure appears as an intermediary description level. This describes the secondary structure as well as its interactions. [3] Figure n°5.



(a) (b)

Figure n° 5: Tertiary structure (a) and super-secondary structure (b)

I. 1.2.3. Quaternary structure:

When the protein is a multi-subunit complex, i.e. a composition of several polypeptides chains, the quaternary structure describes the arrangement of these chains (stoichiometry, Interaction interface, symmetry,. . .). Figure n °6 presents the quaternary structure of human hemoglobin, which is a heterotetramer (α2β2) composed of two heterodimers (αβ). [3]

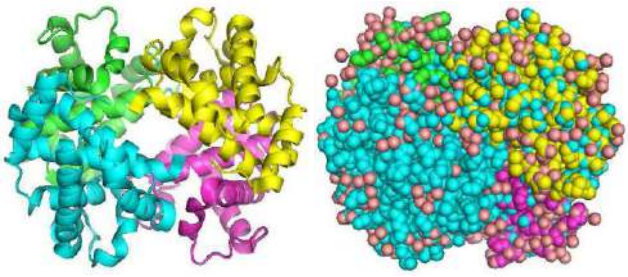


Figure n° 6: Quaternary structure of human hemoglobin (PDB: 1MKO).

I. 1.3.Protein Motifs :

The term **motif** is used in two different ways in structural biology. The first refers to a particular amino-acid sequence that is characteristic of a specific biochemical function. An example is the so-called zinc finger motif, which is found in a widely varying family of DNA-binding proteins (Figure n°7). [5]

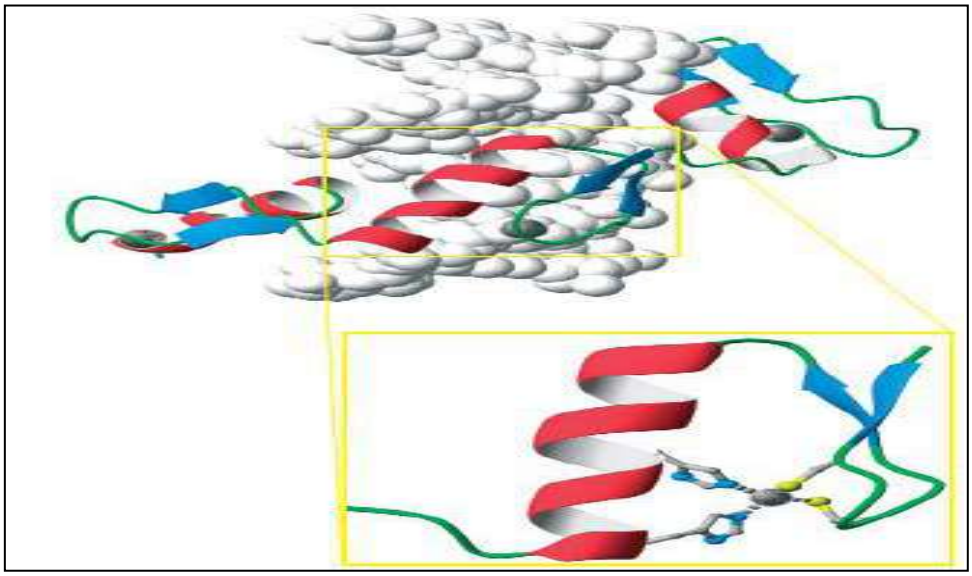


Figure n° 7: Zinc Finger Motif**.** A fragment derived from a mouse gene regulatory protein is shown, with three zinc fingers bound spirally in the major groove of a DNA molecule. The inset shows the coordination of a zinc atom by characteristically spaced cysteine and histidine residues in a single zinc finger motif.

The second, equally common, use of the term motif refers to a set of contiguous secondary structure elements that either have a particular functional significance or define a portion of an independently folded domain. Along with the functional sequence motifs , the former are known generally as **functional motifs**. An example is the **helix-turn-helix** motif found in many DNA-binding proteins (Figure n°8). [5]

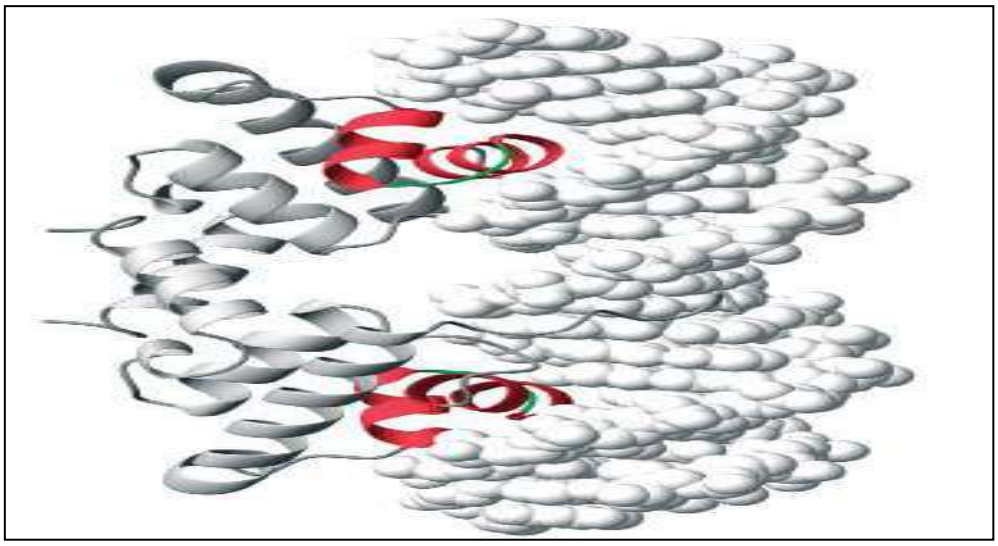


Figure n° 8: Helix-turn-helix. The DNA-binding domain of the bacterial gene regulatory protein lambda repressor , with the two helix-turn-helix motifs shown in color. The two helices closest to the DNA are the reading or recognition helices, which bind in the major groove and recognize specific gene regulatory sequences

This simple **structural motif** will not exist as a stably folded domain if expressed separately from the rest of its protein context, but when it can be detected in a protein that is already thought to bind nucleic acids, it is a likely candidate for the recognition element. Examples of structural motifs that represent a large part of a stably folded domain include the four-helix bundle (Figure n°9), a set of four mutually anti-parallel alpha helices that is found in many hormones as well as other types of proteins; the Rossmann fold , an alpha/beta twist arrangement that usually binds NAD cofactors; and the **Greek-key motif**, an all-beta-sheet arrangement found in many different proteins and which topologically resembles the design found on ancient vases. As these examples indicate, these structural motifs sometimes are suggestive of function, but more often are not: the only case here with clear functional implications is the Rossmann fold.). [5]

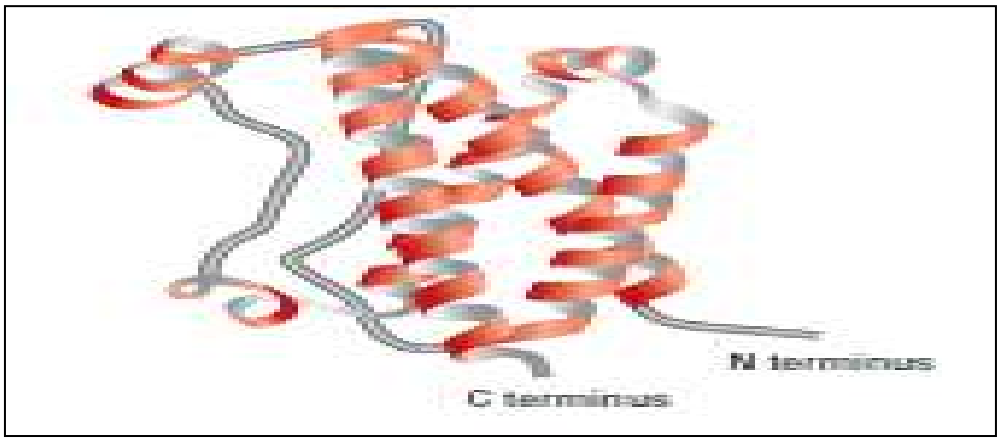


Figure n° 9: Four-helix bundle motif . The four-helix bundle motif can comprise an entire protein domain, and occurs in proteins with many different biochemical functions. Shown here is human growth hormone, a signaling molecule.

I. 1.4.Enzymes:

Enzymes are proteins functioning as catalysts that speed up reactions by lowering the activation energy. A simple and succinct definition of an enzyme is that it is a biological catalyst that accelerates a chemical reaction without altering its equilibrium. During the reactions the enzymes themselves undergo transient changes. In the overall process, enzymes do not undergo any net change. The enzyme catalysts regulate the structure and function of cells and organisms. [6]

I. Nature of Enzymes :

Enzymes are proteins. However, without the presence of non-protein component called cofactor, many enzyme proteins lack catalytic activity. When this is the case, the inactive protein component of an enzyme is termed the Apoenzyme, and the active enzyme, including cofactor, the holoenzyme. The cofactor may be an organic molecule, when it is known as a coenzyme or it may be a metal ion. Some enzymes bind cofactors more tightly than others. When a cofactor is bound tightly (that it is difficult to remove without damaging the enzyme) it is sometimes called a prosthetic group**.** [7]

I. 1.4.2. Classification of Enzymes :

By common convention, an enzyme's name consists of a description of what it does, with the word ending in"-ase". The International Union of Biochemistry and Molecular Biology has developed a nomenclature for enzymes, the enzyme commission (EC) numbers. The Enzyme Commission divided enzymes into six main classes, on the basis of total reaction catalyzed. Each enzyme is described by a sequence of four numbers, preceded by EC.The first number broadly classifies the enzyme based on its reaction mechanism. [7]

- **Oxidoreductases:** catalyze oxidation/reduction reactions.
- **Transferases:** transfer a functional group (e.g. a methyl of phosphate group).
- **Hydrolases:** catalyze the hydrolysis of various bonds.
- **Lyases:** cleave various bonds by means other than hydrolysis and oxidation.
- **Isomerases:** catalyze isomerization changes within a single molecule.
- **Ligases:** join two molecules with covalent bonds.

I. 1.4.3. Active Site:

Enzymatic catalysis relies on the action of amino acid side chains arrayed in the active center. Enzymes bind the substrate into a region of the **active site** in an intermediate conformation. [6]

I. 1.5. Methods of determining protein structure:

Three-dimensional protein structures are determined by two methods: x-ray crystallography and NMR spectroscopy. Protein structures can be determined to a _ne detail, describing the relative position of every single atom within the protein. [4]

I. 1.5.1. Method of X-ray crystallography :

X-ray crystallography is the predominant method of protein structure determination. The process begins by growing crystals of a purified protein sample. Once the crystals have grown sufficiently large, X-ray beams are applied to the crystal,and the structure is determined by studying the diffraction pattern. While this process might sound simple in a brief summary, it is not. [4].

The procedure for obtaining the protein structure using X-ray crystallography can be divided into three steps: Obtaining the crystal, Recording diffraction outputs, Processing diffraction patterns. [8] See figure n°10.

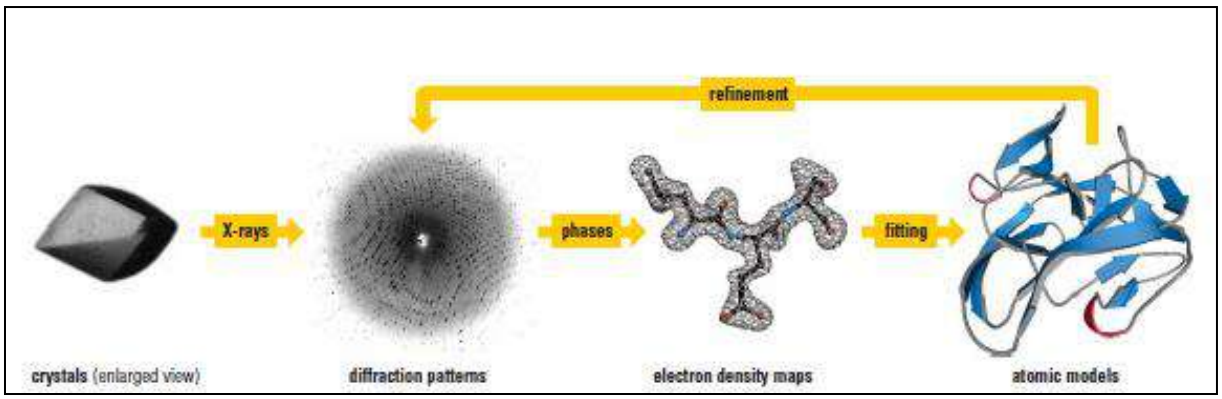


Figure n°10: Structure determination by X-ray crystallography.

I. 1.5.2 .Method of Nuclear Magnetic Resonance (NMR) spectroscopy:

Nuclear Magnetic Resonance (NMR) spectroscopy does not require protein crystals, but "merely" a highly concentrated and purified sample of the protein in question, at a slightly lowered pH. The protein is then put in a strong magnetic field, and subjected to radio frequency (RF) pulses. This puts the nuclei of certain atoms of the protein in an excited state, and as they return to equilibrium, they emit RF radiation. Structural information can then be inferred from the frequencies and intensities of the emitted radiation and from coupling between the frequencies of individual nuclei. Like crystallography, determining the structure from the observed data is a complex modeling process itself, and the technique is not viable on all proteins. Certain proteins are not stable in concentrated solutions at lowered pH . [4]

In general, NMR spectroscopy is not viable on larger proteins due to technical limitations.see figure n°11.

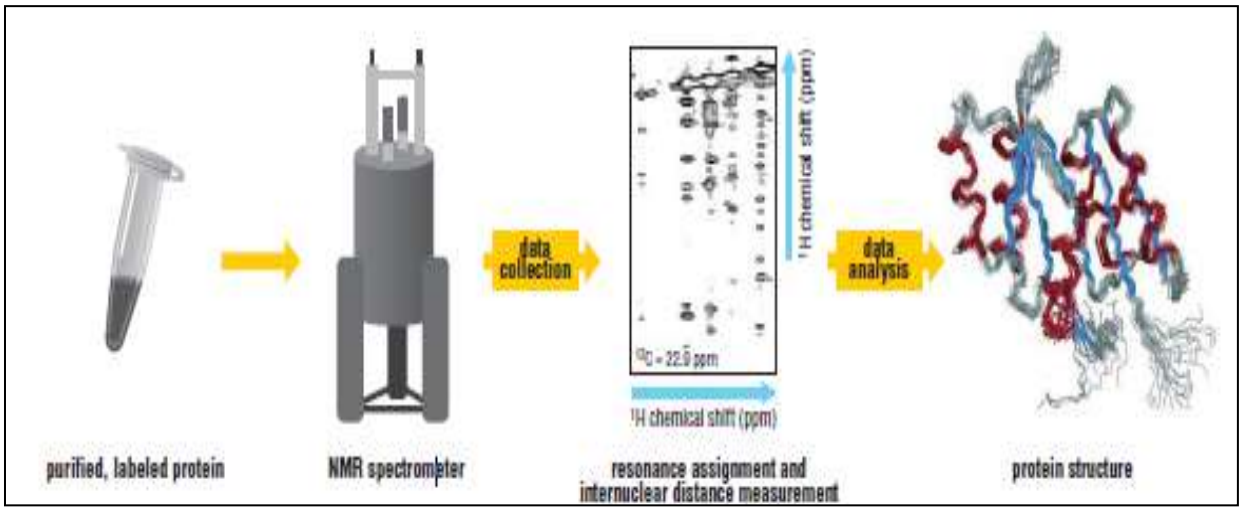


Figure n° 11: Structure determination by NMR.

II. Metabolism:

One of the fundamental properties of all living organisms is the process of metabolism, where by organic compounds are synthesized (anabolism) and broken down (catabolism).

The metabolism (derived from the Greek 'metabolismos', meaning 'change') of a whole cell is an extremely complex system, but it can be broken down into subsystems and pathways, which are comprised of individual reactions that change one compound into another.[9]

II. 1. Catabolism of amino acids:

The carbon skeletons of amino acids are important energy sources in some dietary situations. Use of these carbon skeletons requires proper disposal of ammonia (NH_3), a toxic by-product of amino acid catabolism.[10]

II. 1.1.Transaminations :

The nitrogen component of amino acids, the α - amino groups, must be removed before the carbons can be used in other metabolic pathways.There are several ways that this can be achieved. The first step in the catabolism of most amino acids is the transfer of their α -amino group to α -ketoglutarate where the products are α- ketoacids and glutamate. This transfer of amino groups from one carbon skeleton to another is

catalyzed by a family of transaminases which are also called as aminotransferases. Most of the amino acids undergo these reaction except Lysine and Threonine [11] ·figure n °12.

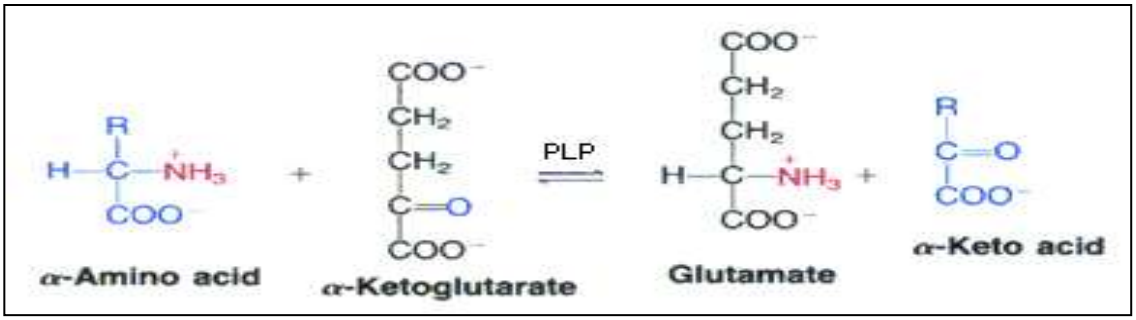


Figure n° 12: Transamination of amino acids.

II. 1.1.1.Aspartate Aminotransferase (ASAT):

Aspartate aminotransferase (AST) catalyses the transfer of an amino group form aspartate to α-ketoglutarate to form oxaloacetate and glutamate. This transamination reaction is reversible. This enzyme can also catalyses the transfer of an amino group from glutamate to oxaloacetate to form aspartate and α-ketoglutarate. [12].

- **Reaction:** L-aspartate + 2-oxoglutarate = oxaloacetate + L-glutamate.

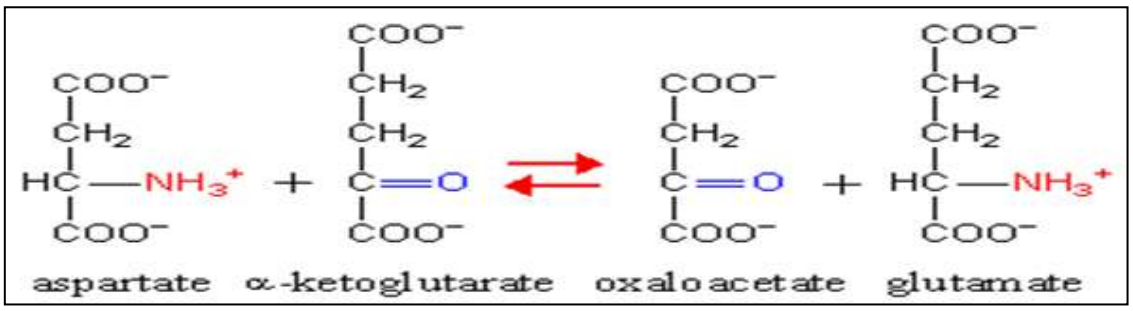


Figure n° 13: Aspartate transaminase reaction.

Aminotransferases utilize a coenzyme - **pyridoxal phosphate** - which is derived from Vitamin B6. The functional part of pyridoxal phosphate is an aldehyde functional group attached to a pyridine ring. Catalysis involves a Schiff base intermediate. [11]

This is a typical pyridoxal dependent enzyme. The crystal structure of this enzyme has been solved. The lysine that forms the Schiff base with the aldehyde of pyridoxal phosphate is Lys-268. Adjacent to the pyridoxal cofactor is the binding site for aspartate/oxaloacete. When aspartate binds to the active site of the enzyme the α-amino group displaces Lys-268 to form the external aldimine . The next step in the enzyme catalysed pathway is the abstraction of the proton from the α-carbon to generate the quinonoid intermediate. The general base that abstracts this proton is the same Lys-268. The protoned Lys-268 then transfers this proton to the aldehyde carbon to generate the ketamine intermediate. [12]

II. 1.2. Oxidative Deamination :

In contrast to transaminase reactions, oxidative deamination yields an α-keto acid with release of the amino group as free ammonia. **Glutamate dehydrogenase** in **liver** is the most important enzyme involved. Glutamate is the only amino acid that is rapidly deaminated – remember α-ketoglutarate collects amino groups on glutamate. Glutamate dehydrogenase then produces ammonia, regenerating α-ketoglutarate. .[10]

The direction of the glutamate dehydrogenase reaction depends on levels of substrates, including the ratio of oxidized/reduced coenzymes. After a protein-containing meal, the reaction proceeds in the direction of amino acid degradation and ammonia production, but the reverse reaction can also be used to synthesize glutamate .Glutamate dehydrogenase is unusual in that it can use either NAD^+ or $NADP^+$ it usually uses NAD^+ for oxidative deamination and NADPH**:** for reductive amination, but doesn't have to.[10]

❖ **Reaction:** L-glutamate + H_2O + $NADP^+$ = 2-oxoglutarate + NH_3 + NADPH + H^+.

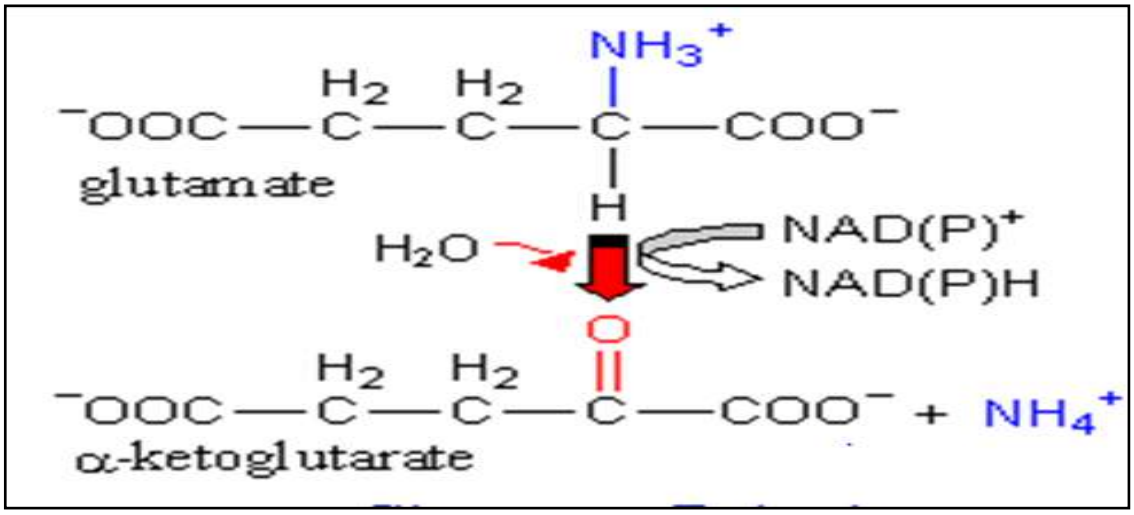


Figure n° 14: Reversible reaction catalyzed by GDH.

II. 2.The Urea Cycle:

Excess amino Nitrogen from amino acids is removed as ammonia, which is toxic to the human body. Some ammonia is excreted in urine, but nearly 90% of it is utilized by the liver to form urea, which is highly soluble and is passed in to circulation for being excreted by the kidneys.

Daily excretion of urea amounts to about 30g with a protein intake of nearly 100g in the food. It is less with lower protein intake. The urea-cycle starts in the mitochondrial matrix of hepatocytes and few of the steps occur in the cytosol: the cycle spans two cellular compartments. The first amino group to enter the cycle is derived from ammonia inside the mitochondria. Some ammonia also arrives at the liver via the portal vein from the intestine, when it is produced by bacterial oxidation of amino acids.[11]

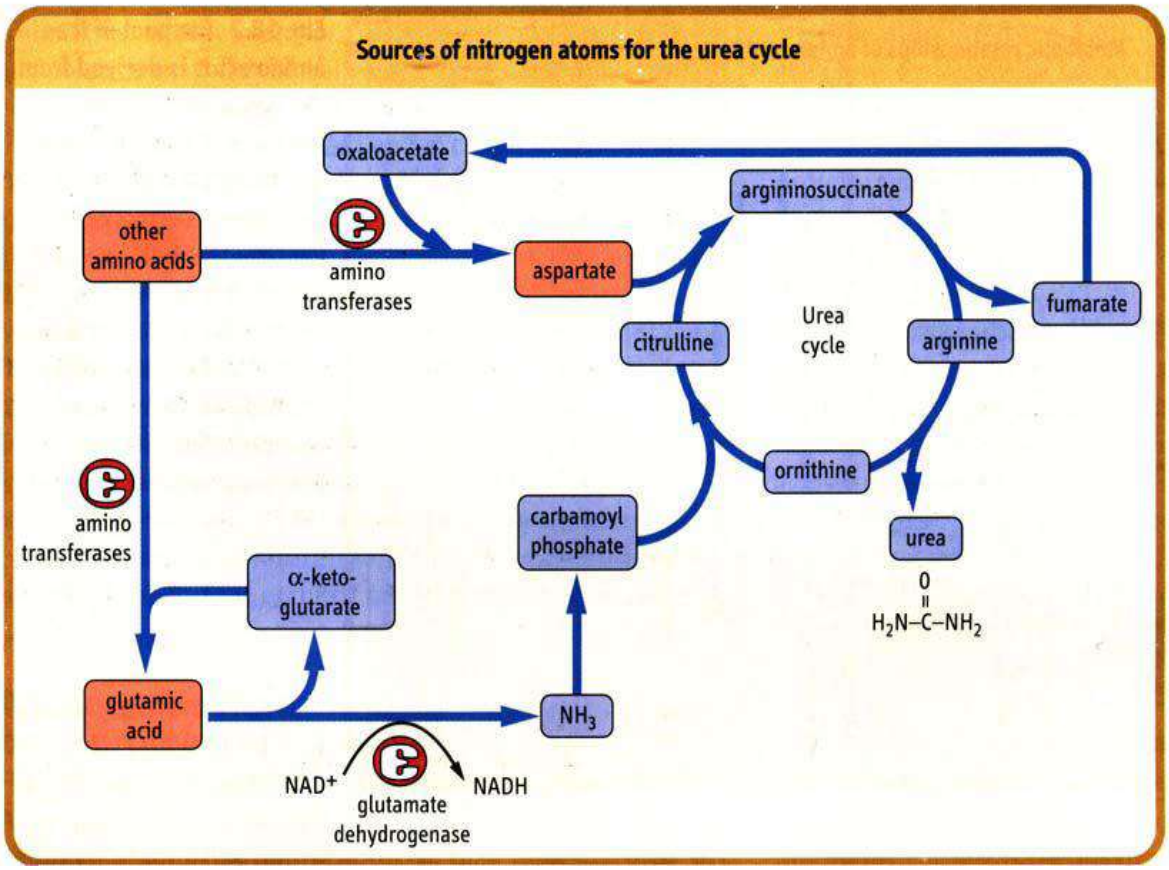


Figure n° 15: The Urea Cycle.

II. 2.1. The reactions :

Step 1: formation of carbamoyl phosphate from ammonia, bicarbonate and ATP.

CO2 from bicarbonate and NH4 from the two sources mentioned above combine together in the liver mitochondria to form carbamoyl phosphate in presence of ATP and Mg2+ by the enzyme Carbamoyl phosphate synthetase I (CPSI).

❖ **Reaction:**

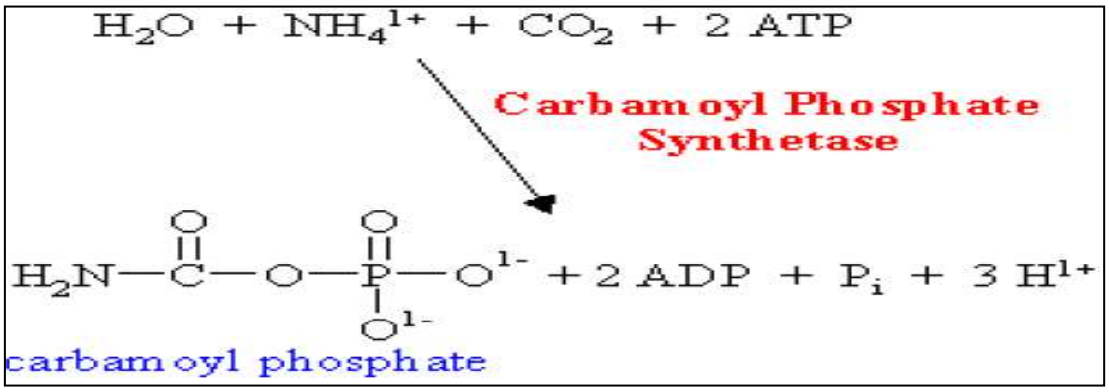


Figure n° 16: Formation of carbamoyl phosphate.(step 1)

Step 2: Formation of citrulline from ornithine and carbamoyl phosphate.

Carbamoyl phosphate reacts with ornithine transferring the carbamoyl moiety to produce citrulline: by the enzyme i.e. ornithine transcarbomylase (OTC).

❖ **Reaction:**

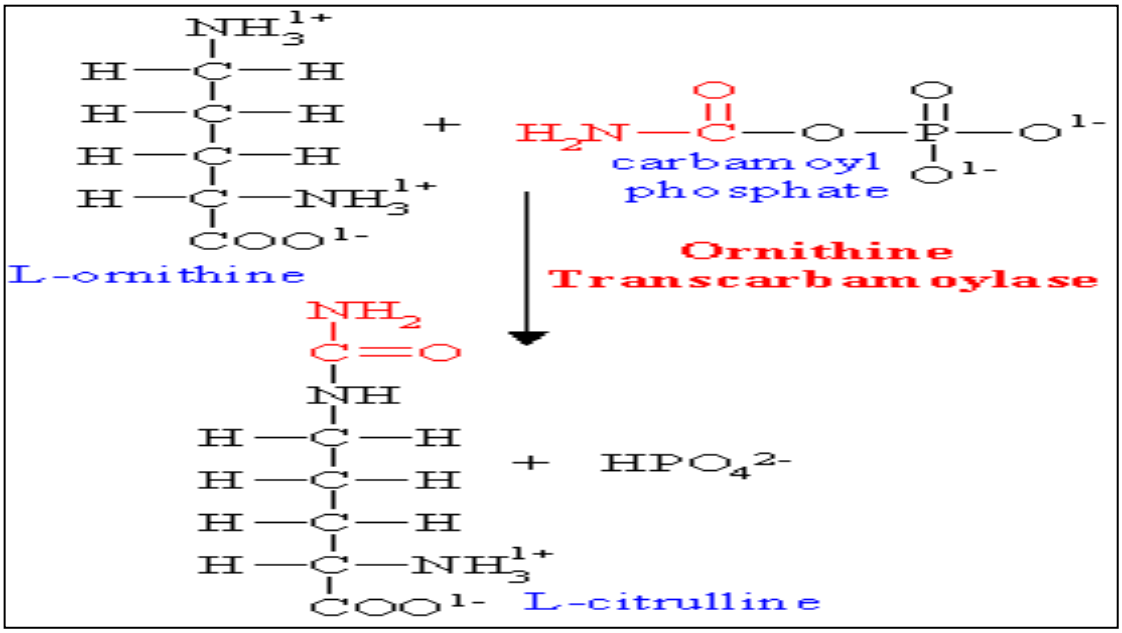


Figure n° 17: Formation of citrulline.

Step 3: Formation of arginosuccinate from citrulline and aspartate.

Argininosuccinic acid is formed by the reaction of Aspartic acid and citrulline: the NH_2 group of the former is linked to – CO group of the latter. The enzyme reqused is argininosuccinic acid synthase. (ASS).

❖ **Reaction**: ATP + L-citrulline + L-aspartate = AMP + diphosphate + N(omega)- (L-arginino)succinate

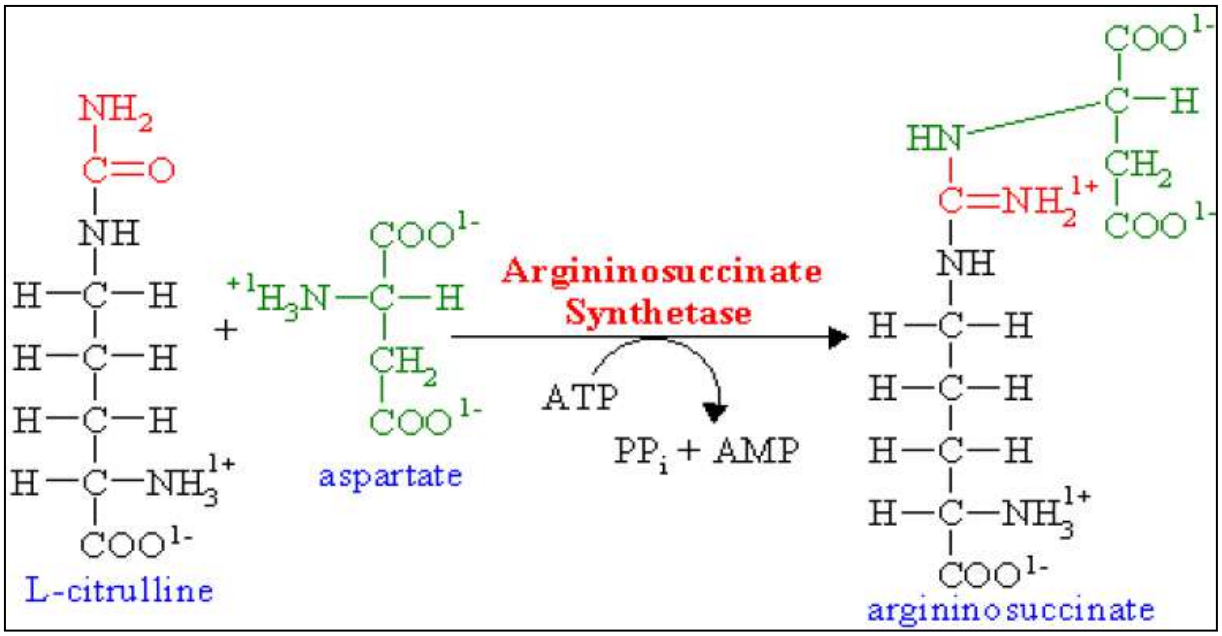


Figure n° 18: Formation of arginosuccinate (Step 3).

Step 4: **Formation of arginine**

Argininosuccinic acid is cleaved to form Arginine and fumerate by the enzyme Arginiosuccinate lyase (ASL). Fumerate goes to the pool of TCA-cycle.

❖ **Reaction :**

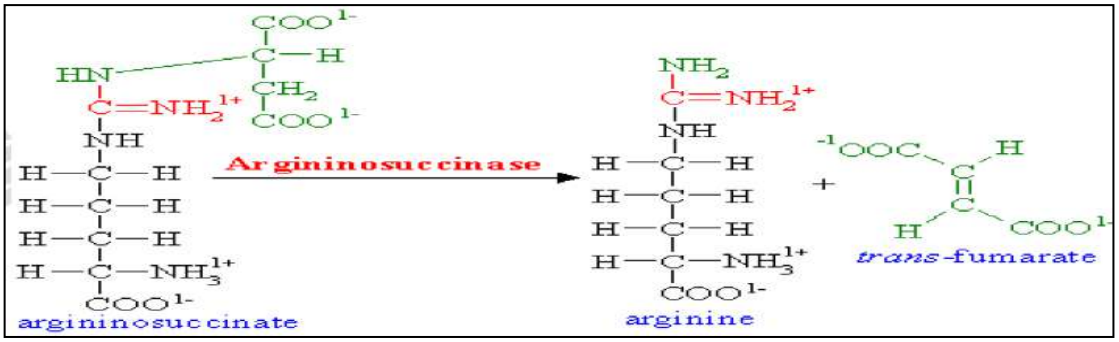


Figure n° 19: Formation of arginine (step 4).

Step 5: Hydrolysis of arginine to form ornithine and urea .

Arginine gets cleared off to urea and ornithine by the cytosolic enzyme arginase.(ARGS)

Ornithine is thus re-generated and can be transported in to the mitochondrion to initiate another round of the urea - cycle.

- **Reaction:**

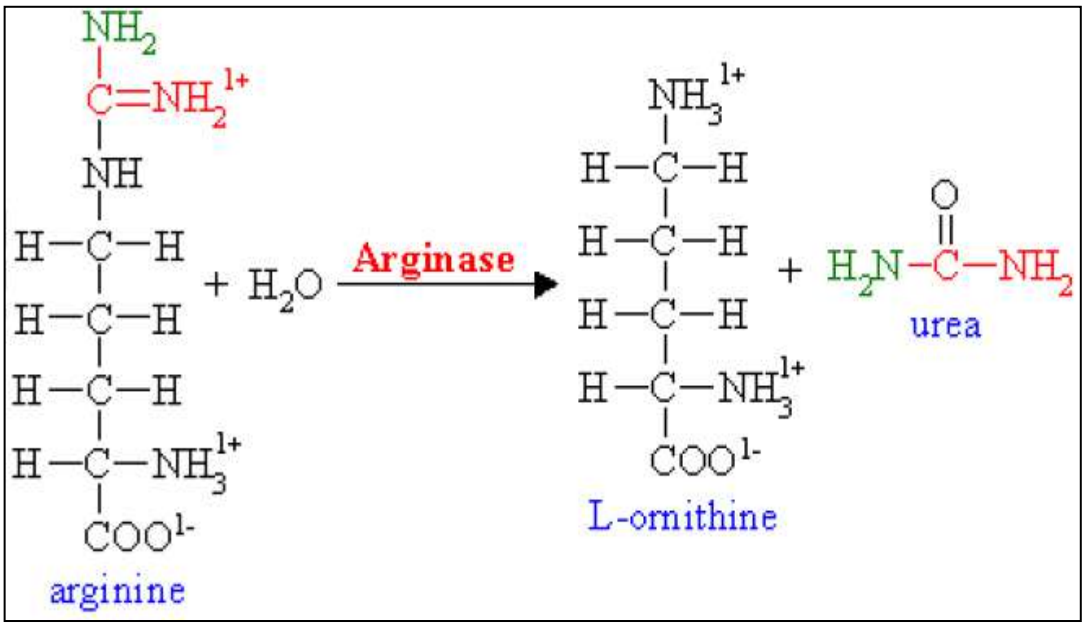


Figure n° 20: Formation of ornithine and urea. (step 5)

II. 2.2 Disorders of Urea Cycle Function

Disruption of the urea cycle in mammals may result from impaired portal blood flow, severe liver disease, lack of urea cycle enzymes, or lack of urea cycle substrates. Although relatively little is known about urea cycle function in companion animals, data from humans and laboratory animals may be of some benefit in predicting diseases that may occur in companion animal species.[13]

III. The Protein Data Bank (PDB) :

The **Protein Data Bank** (**PDB**) is a crystallographic database for the three-dimensional structural data of large biological molecules, such as proteins and nucleic acids. The data, typically obtained by X-ray crystallography, NMR spectroscopy, or, increasingly, cryo-electron microscopy, and submitted by biologists and biochemists from around the world, are freely accessible on the Internet via the websites of its member organisations (PDBe, PDBj,[1] and RCSB[]). The PDB is overseen by an organization called the World wide Protein Data Bank, ww PDB.[14]. the first link returned, which is: http://www.rcsb.org/pdb/home

Figure n° 21: PDB – Protein Data Bank.

III. 1.File format :

The file format initially used by the PDB was called the PDB file format.

This original format was restricted by the width of computer punch cards to 80 characters per line. Around 1996, the "macromolecular Crystallographic Information file" format, mmCIF, which is an extension of the CIF format started to be phased in. mmCIF is now the master format for the PDB archive. An XML version of this format, called PDBML, was described in 2005. The structure files can be downloaded in any of these three formats. In fact, individual files are easily downloaded into graphics packages using web addresses:

- For PDB format files, use, e.g., http://www.pdb.org/pdb/files/4hhb.pdb.gz or http://pdbe.org/download/4hhb
- For PDBML (XML) files, use, e.g., http://www.pdb.org/pdb/files/4hhb.xml.gz or http://pdbe.org/pdbml/4hhb

The "4hhb" is the PDB identifier. Each structure published in PDB receives a four-character alphanumeric identifier, its PDB ID. (This cannot be used as an identifier for biomolecules, because often several structures for the same molecule in different environments or conformations are contained in PDB with different PDB IDs.).[14]

III. 2.Resolution :

Resolution is a measure of the quality of the data that has been collected on the crystal containing the protein or nucleic acid.[14]

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

The refinement factor is a statistical value, in percentage, which reflects the quality of protein and nucleic acid structures and is the main value of the level of errors may be associated with the final structures.[15]

IV. Structural Classification of Proteins :

IV. 1. The scop database:

Scop aims to provide a detailed and comprehensive description of the structural and evolutionary relationships between all proteins whose structure is known, including all entries in Protein Data Bank (PDB). It is available as a set of tightly linked hypertext documents which make the large database comprehensible and accessible. In addition, the hypertext pages offer a panoply of representations of proteins, including links to PDB entries, sequences, references, images and interactive display systems. World Wide Web URL **http://scop.mrc-lmb.cam.ac.uk/scop/** is the entry point to the database (MRCsite).[16]

IV. 2 .CATH:

Cath is a novel hierarchical classification of protein domain structures, which clusters proteins at four major levels, class(C), architecture(A), topology(T) and homologous super family (H). Class, derived from secondary structure content, is assigned for more than 90% of protein structures automatically. Architecture, which describes the gross orientation of secondary structures, independent of connectivities, is currently assigned manually. The topology level clusters structures according to their toplogical connections and numbers of secondary structures. The homologous super families cluster proteins with highly similar structures and functions. The assignments of structures to toplogy families and homologous super families are made by sequence and structure comparisons. CATH, can be reach on the Web at this URL: **http://www.cathdb.info**.[16]

IV. 3. Classification Levels :

At the top level, each domain belongs to one out of four classes. The class is determined according to the secondary structure composition. The classes are:

- **Mainly alpha** Domains with mainly alpha helixes and few beta strands.
- **Mainly beta** Domains with mainly beta strands and few alpha helixes
- **Alpha and beta** Domains with both beta strands and alpha helixes
- **Other** Domains with few secondary structure elements or irregular structures (i.e., those domains that do not belong to any of the previous classes).[17]

Introduction:

In order to realize the structural study of the ligand binding environment found in the enzyme complexes selected to study in this project, methods involving informatics, databases and programming were employed:

The steps followed in to achieve the goals of this study are summarized in the following:

1. **Protein structures identification and Data Preparation.**
2. **Binding details Calculations and Data Mining.**
3. **Data Storing and Flat-Files Database creation.**
4. **World Wide Web Database.**

1.Protein structures identification and Data Preparation:

As it has been explained in the previous chapter, the PDB is the database which provides structural data for proteins and nucleic acids. Every structure in the Protein Data Bank (PDB) is stored as an entry which is given an identification code or PDB ID.

The structures of the enzymes including the binding ligands involved in the metabolic pathway understudy have been identified as explained below.

1.1.Protein structures (PDB entries) :

The PDB entries used in this project amount to 16 structures each with its own PDB id. All of the studied structures have been found to have determined by the X-ray crystallography method and are list below depending on the degradation pathway cycle they belong to. Same presentation is done for the ligands found bound in the enzyme complexes.

The table below represents the list of protein structures in complex with ligands. Resolution and R-factor which reflect the quality of the structures under study are also shown in the tables.

1.1.1. List of PDB entries:

ENZYME	CLASS	PDB ID	TITLE OF PDB	Methode	RESOLUTION (angstrom)	R-Value (%)
Glutamate dehydrogenase	**oxidoreductase**	3ETD	Glutamate dehydrogenase complexed with bithionol	x-ray diffraction	2.5	0.239
		3MVO	Bovine glutamate dehydrogenase complexed with eu3+	x-ray diffraction	3.23	0.261
		3MVQ	Bovine glutamate dehydrogenase complexed with zinc	x-ray diffraction	2.94	0.223
Aspartate aminotransferase	**Transferase**	3II0	Crystal structure of human Glutamate oxaloacetate transaminase 1 (GOT1)	x-ray diffraction	2.05	0.180

Table n° 1: List of protein structures used in the study with the title of PDB entry, resolution and R-factor which reflect the quality of the structure (catabolism of amino acid).

ENZYME	CLASS	PDB ID	TITLE OF PDB
Carbamoyl Phosphate Synthetase	Ligase	1JDB	Carbamoyl phosphate synthetase from escherichia coli.
		1T36	Crystal structure of E. coli carbamoyl phosphate synthetase small subunit mutant c248d complexed with uridine 5'-monophosphate.
	Amidotransferase	1A9X	Carbamoyl phosphate synthetase: caught in the act of glutamine hydrolysis.
	Ligase	1KEE	Inactivation of the amidotransferase activity of carbamoyl phosphate synthetase by the antibiotic acivicin.
ornithine transcarbamoylase	Transferase	1OTH	Crystal structure of human ornithine transcarbamoylase complexed with n-phosphonacetyl-l-ornithine.
argininosuccinate synthase	ligase	2NZ2	Crystal structure of human argininosuccinate synthase in complex with aspartate and citrulline.

	Ligase	**1J1Z**	Crystal Structure of Thermus thermophilus HB8 Argininosuccinate Synthetase in complex with substrate
Argininosuccinate lyase	**Lyase**	**1K7W**	Crystal Structure of S283A Duck Delta 2 Crystallin Mutant
	Lyase	**1TJW**	Crystal Structure of T161D Duck Delta 2 Crystallin Mutant with bound argininosuccinate
Arginase	**Hydrolase**	**3KV2**	high resolution structure of human arginase i in complex with the strong inhibitor n(omega)-hydroxy-nor-l-arginine (nor-noha)
		3LP7	crystal structure of human arginase i in complex with inhibitor n(omega)-hydroxy-l-arginine (noha).
		3CEV	arginase from bacillus caldevelox, complexed with l-arginine

Table n° 2: List of protein structures used in the study with the title of PDB entry, Resolution and R-factor which reflects the quality of the structures (urea cycle).

1.1.2. List of ligands:

The tables bellow represents all of the ligands used in this project that are bound with the enzymes of the Amino Acids degradation (**Table n°3**) and those bound with the Urea cycle enzymes (**Table n°4**):

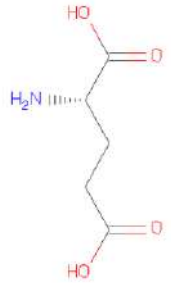
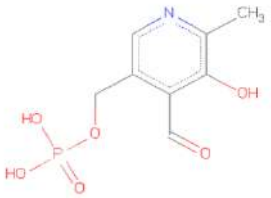
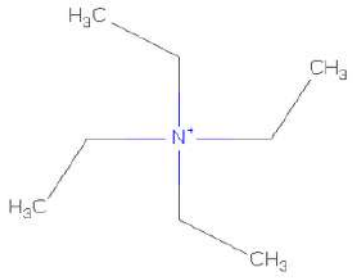
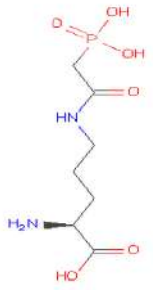
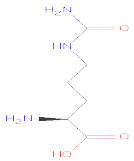
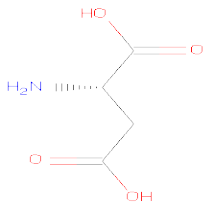
Ligand ID	Ligand name	LIGAND FORMULA	Ligand Chemistry	PDB COD
GLU	GLUTAMIC ACID	C5 H9 N O4		3ETD 3MVO 3MVQ
PLP	PYRIDOXAL-5'-PHOSPHATE	C8 H10 N O6 P		3II0

Table n°3: List of the Ligands in complex with Amino Acids Degradation enzymes.

Ligand ID	Ligand name	Ligand Formula	Ligand Chemistry	PDB IDs
NET	Tetraethylammonium Ion	C8 H20 N (1+)		1JDB 1T36 1A9X 1KEE
PAO	N-(PHOSPHONOACETYL) -L-ORNITHINE	C7 H15 N2 O6 P		1OTH
CIR	CITRULLINE	$C_6 H_{13} N_3 O_3$		2NZ2 1J1Z
ASP	ASPARTIC ACID	$C_4 H_7 N O_4$		2NZ2 1J1Z

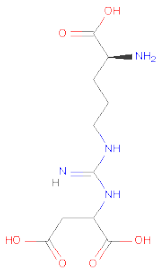
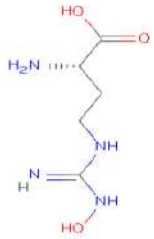
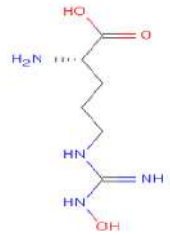
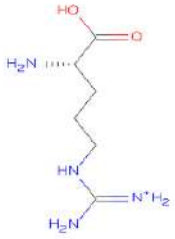
AS1	ARGININOSUCCINATE	C_{10} H_{18} N_4 O_6		1K7W 1TJW
NNH	NOR-N-OMEGA-HYDROXY-L-ARGININE	C5 H12 N4 O3		3KV2
HAR	N-OMEGA-HYDROXY-L-ARGININE	C6 H14 N4 O3		3LP7
ARG	ARGININE	C6 H15 N4 O2 (1+)		3CEV

Table n° 4: List of the Ligands in complex with urea cycle enzymes.

It should be noted that due to the difficulty of producing 3D-structures of enzymes bound to their natural substrates, analogue of these are used instead; Hence some of the ligands above are analogues to the natural substrates; for example the ligands NET, PAO, HAR and NHH are analogues to the natural substrates of this project. The PLP is a cofactor ligand. These analogues are, in addition, used by the structures' producers to study the various aspects of the enzymes binding sites and reaction dynamics.

2. Binding Details Calculations and Data Mining:

In order to study the structure-function relationship in the selected enzymes, calculation of the ligands binding environment is necessary.

The Ligand Binding tool**; lgb**, which is a light version of the bioinformatics tool "Sequence Structure and Function Server" SSFS [18], [19], has been used to carry out calculations of the ligands binding environment details through the url address: **http://bioinformaticstools.org/prjs/lgb/**. This has been done for all of the 16 PDB entries together with their bound ligand.

The table below, Table 05, is an example that represents the output of the **Lgb** for the enzyme Argininosuccinate synthethas and its ligand the Aspartate (id: ASP) as found in the PDB entry 1J1Z. The **Lgb** is used as shown in Figure n°22.

Figure n°22: Capture the interface of the site **Lgb**.

Entry: 1j1z	LIGASE									
Protein-Ligand Environment										
Protein Residues					Ligand				Bonds	
Chain	Sselm	Name	Number	Atom	Chain	Name	Number	Atom	Distance/Å	Possible Bond Type
A	No SSE	ALA	115	CA	A	ASP	530	O	3.19	van der Waals
A	No SSE	ALA	115	C	A	ASP	530	O	3.41	van der Waals
A	No SSE	ALA	115	CB	A	ASP	530	O	3.61	van der Waals
A	No SSE	THR	116	N	A	ASP	530	C	3.52	H.Bond
A	No SSE	THR	116	N	A	ASP	530	O	2.75	H.Bond
A	No SSE	THR	116	N	A	ASP	530	OXT	3.54	H.Bond
A	No SSE	THR	116	CA	A	ASP	530	O	3.86	van der Waals
A	No SSE	THR	116	CA	A	ASP	530	OXT	3.89	van der Waals
A	No SSE	THR	116	CB	A	ASP	530	O	3.96	van der Waals
A	No SSE	THR	116	CB	A	ASP	530	OXT	3.34	van der Waals
A	No SSE	THR	116	OG1	A	ASP	530	C	3.49	van der Waals
A	No SSE	THR	116	OG1	A	ASP	530	O	3.82	H.Bond
A	No SSE	THR	116	OG1	A	ASP	530	OXT	2.48	H.Bond
A	No SSE	THR	116	CG2	A	ASP	530	C	3.67	van der Waals
A	No SSE	THR	116	CG2	A	ASP	530	O	3.57	van der Waals
A	No SSE	THR	116	CG2	A	ASP	530	OXT	3.32	van der Waals
A	No SSE	GLY	119	N	A	ASP	530	OXT	3.82	H.Bond
A	No SSE	GLY	119	CA	A	ASP	530	CG	3.4	van der Waals
A	No SSE	GLY	119	CA	A	ASP	530	OD1	3.51	van der Waals
A	No SSE	GLY	119	CA	A	ASP	530	OD2	3.32	van der Waals
A	No SSE	GLY	119	CA	A	ASP	530	OXT	3.75	van der Waals
A	No SSE	GLY	119	C	A	ASP	530	CG	3.75	van der Waals
A	No SSE	GLY	119	C	A	ASP	530	OD1	3.57	van der Waals
A	No SSE	GLY	119	C	A	ASP	530	OD2	3.48	van der Waals
A	120-133 H: 1	ASN	120	N	A	ASP	530	CG	3.24	H.Bond
A	120-133 H: 1	ASN	120	N	A	ASP	530	OD1	3.14	H.Bond
A	120-133 H: 1	ASN	120	N	A	ASP	530	OD2	2.71	H.Bond
A	120-133 H: 1	ASN	120	CA	A	ASP	530	OD1	3.97	van der Waals
A	120-133 H: 1	ASN	120	CA	A	ASP	530	OD2	3.68	van der Waals
A	120-133 H: 1	ASN	120	C	A	ASP	530	OD1	3.77	van der Waals
A	120-133 H: 1	ASN	120	CB	A	ASP	530	OD2	3.74	van der Waals
A	120-133 H: 1	ASN	120	ND2	A	ASP	530	OD2	3.73	H.Bond
A	120-133 H: 1	ASP	121	N	A	ASP	530	CG	3.6	H.Bond
A	120-133 H: 1	ASP	121	N	A	ASP	530	OD1	2.72	H.Bond
A	120-133 H: 1	ASP	121	N	A	ASP	530	OD2	3.71	H.Bond

A	120-133 H: 1	ASP	121	CA		A	ASP	530	OD1	3.41	van der Waals
A	120-133 H: 1	ASP	121	CB		A	ASP	530	OD1	2.99	van der Waals
A	120-133 H: 1	ASP	121	CG		A	ASP	530	N	3.69	H.Bond
A	120-133 H: 1	ASP	121	CG		A	ASP	530	CG	3.84	van der Waals
A	120-133 H: 1	ASP	121	CG		A	ASP	530	OD1	3.11	van der Waals
A	120-133 H: 1	ASP	121	OD1		A	ASP	530	OD1	3.97	H.Bond
A	120-133 H: 1	ASP	121	OD2		A	ASP	530	N	2.62	H.Bond
A	120-133 H: 1	ASP	121	OD2		A	ASP	530	CA	3.8	van der Waals
A	120-133 H: 1	ASP	121	OD2		A	ASP	530	CB	3.93	van der Waals
A	120-133 H: 1	ASP	121	OD2		A	ASP	530	CG	3.52	van der Waals
A	120-133 H: 1	ASP	121	OD2		A	ASP	530	OD1	3.03	H.Bond
A	181-184 S: -1	GLU	184	OE1		A	ASP	530	CB	3.77	van der Waals

Table n° 5: The binding environment details of the ASP bound the enzyme ASS (PDB id: 1J1Z) as calculated by the **Lgb** system.

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 structure elements (helix, b-sheet or loop) they may belong to.
- **The columns under the title "Ligand":** These columns show the atoms of the ligand, its number and the ligand id.
- **The columns under the title "bond":** These columns represent the distance between atoms **(Å: Angstroms)** and the possible bonds which can for example be **Hydrogen** bonds or **van der Waals** bonds**..** etc.

Using the **lgb** system, the binding environment details of all the ligands associated with the 16 PDB structures have been calculated, collected and stored into a system of organized files (see below the section: **Flat-Files Database**).

2.1. Binding Motifs Constructions and Representation:

As seen above in the binding details, residues in contact with the ligand belong to some secondary structure elements.

The table n° 5, see above, report the protein binding residues association with precise region that may represent secondary structure elements. This annotation has been used to create **patterns** that would describe the ligand binding sites in an abstract manner as is shown in the following example (which is for the ligand **ASP** associated with the PDB id: 1J1Z, chain A):

- **The protein region "No SSE"** represents the lack of secondary structure which means that the binding residues belong to a loop region and is given the symbol **L**.
- **The protein region "120-133 H:1"** represents the secondary structure α-Helix and is symbolized as **H**.
- **The protein region "181-184 S: -1"** represents the secondary structure β-strand which is given the symbol **S**.

The pattern representing the binding site of the ligand ASP found in the table above is represented as follows: **LHS**.

These **patterns** which have been created for all of the ligands binding sites appear to be associated with types of functions and reappear accordingly, see Chapter III, and thus they can be better denoted or annotated as being **Structural & Functional Motifs**, also referred to below as simply **Binding Motifs**.

2.2. Graphical Representation of the Binding Motifs:

Graphical representations of the motifs in the ligand binding sites have been generated by the Rasmol molecular graphics program [20] where the helices (H) are shown as Red ribbons, β-strands (S) as Yellow ribbons and Loop regions (L) as Light Grey strips.

For reasons of clarifying the ligand binding sites, three types of Rasmol images were produced for each ligand binding case:

- ✓ **Motif-only**, see Figure 23-b.
- ✓ **Motif + Ligand** , see Figure 24-b.
- ✓ **Motif + Ligand + binding Residues**, see Figure 25-b.

In order to create the images for the graphical representation of the motifs in the binding sites, Rasmol program uses a script language that tells it what to and how to represent the molecular data in the graphical mode as seen in the figures **Figure 23-a, Figure 24-a and Figure 25-a.**

2.2.1. Motif-only presentation:

The following is an example of a Rasmol script, see **Figure n°23-a**, which generates the Rasmol graphics representation for the motif **LHS** without showing the ligand **ASP** nor the binding residues**, see**
Figure n°23-b:

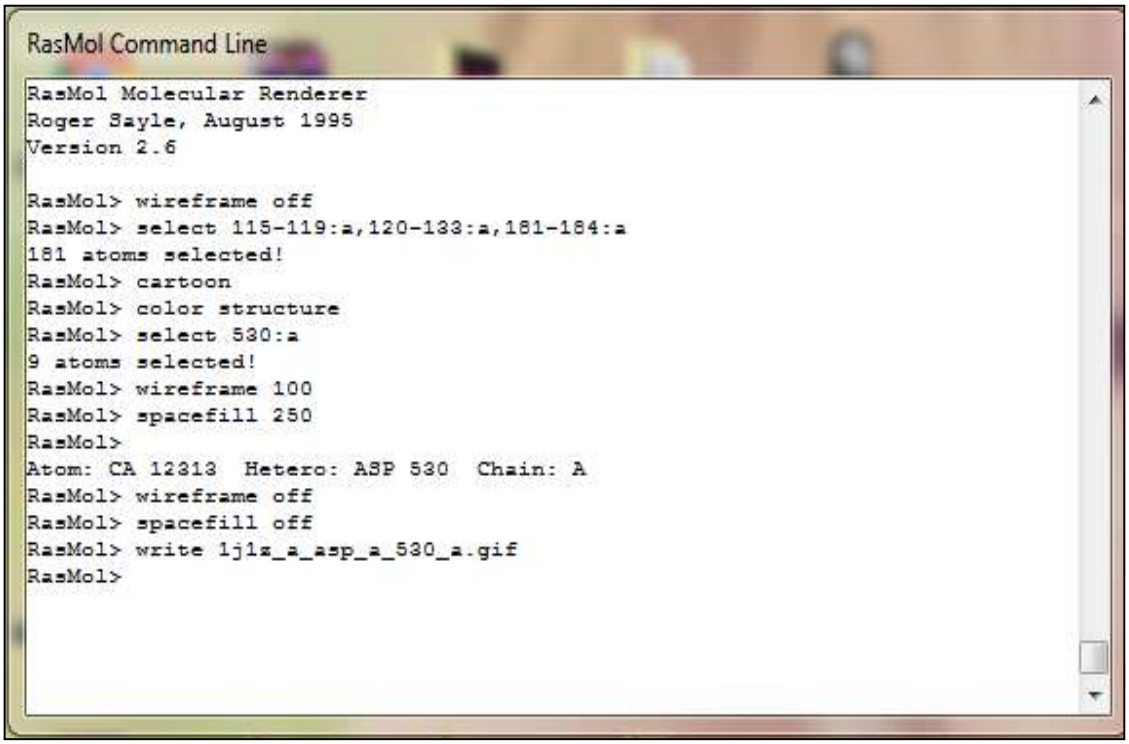


Figure n° 23-a: capture of the RasMol script to create the motif representation shown in **Figure 23-b.**

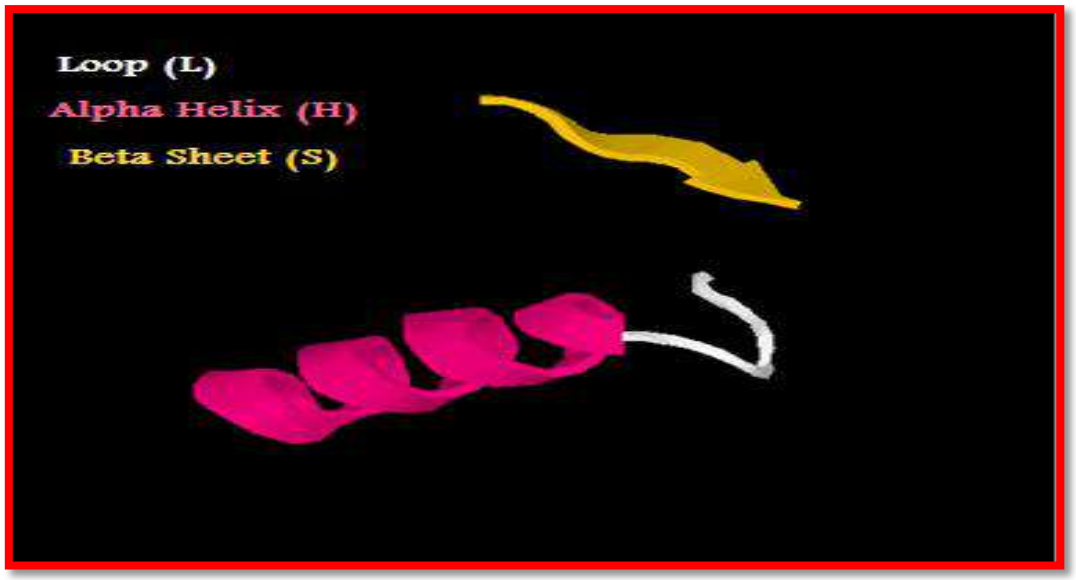


Figure n°23-b: Capture of RasMol representation of the binding motifs where the ligand ASP and binding residue are not shown in the case of ASS (PDB id: 1J1Z, chain A)

2.2.2. Motif + Ligand representation:

The following is an example of a Rasmol script, see **Figure n°24-a**, which generates the Rasmol graphics representation for the motif **LHS** where the ligand ASP is shown but not the binding residues, see **Figure n°24-b:**

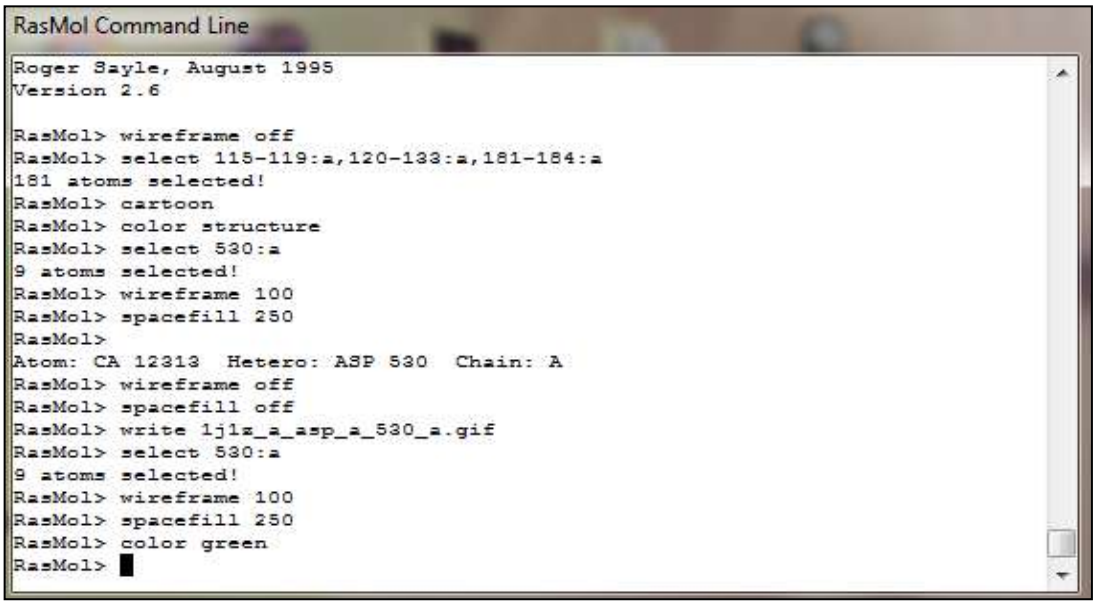


Figure n°24-a: Capture of the RasMol script to create the motif representation shown in **Figure n°24-b.**

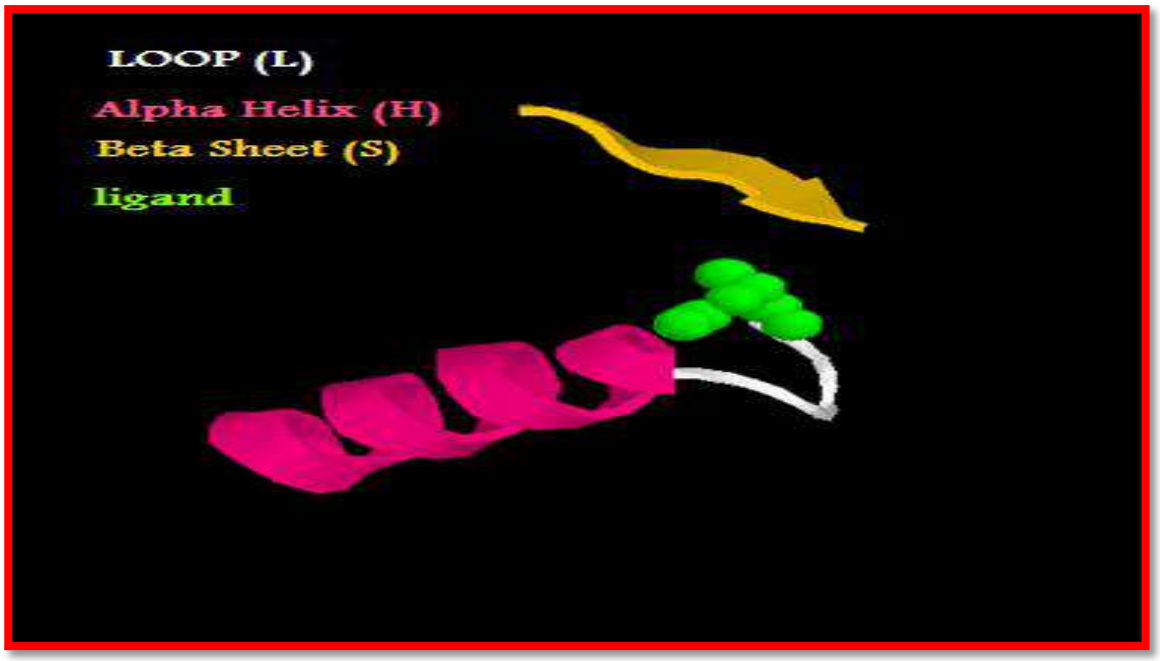


Figure n°24-b: Capture of RasMol representation of the binding motifs where the binding residues are not shown but the ligand ASP is shown in the case of ASS (PDB id: 1J1Z, chain A)

2.2.3. Motif + Ligand+ Binding Residues representation:

The following is an example of a Rasmol script, see **Figure n°25-a**, which generates the Rasmol graphics representation for the motif **LHS** with the ligand **ASP** and the binding residues**, see Figure n°25-b**:

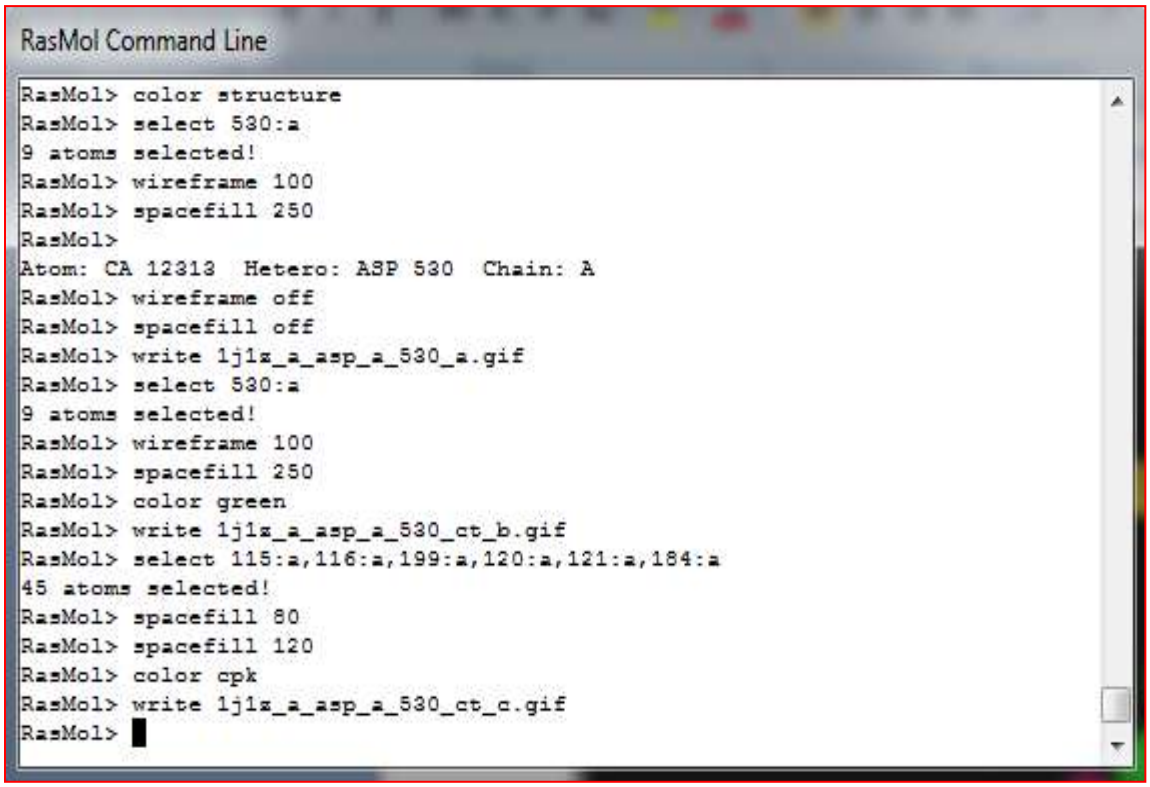


Figure n° 25–a: Capture of the RasMol script to create the motif representation shown in **Figure n° 25-b.**

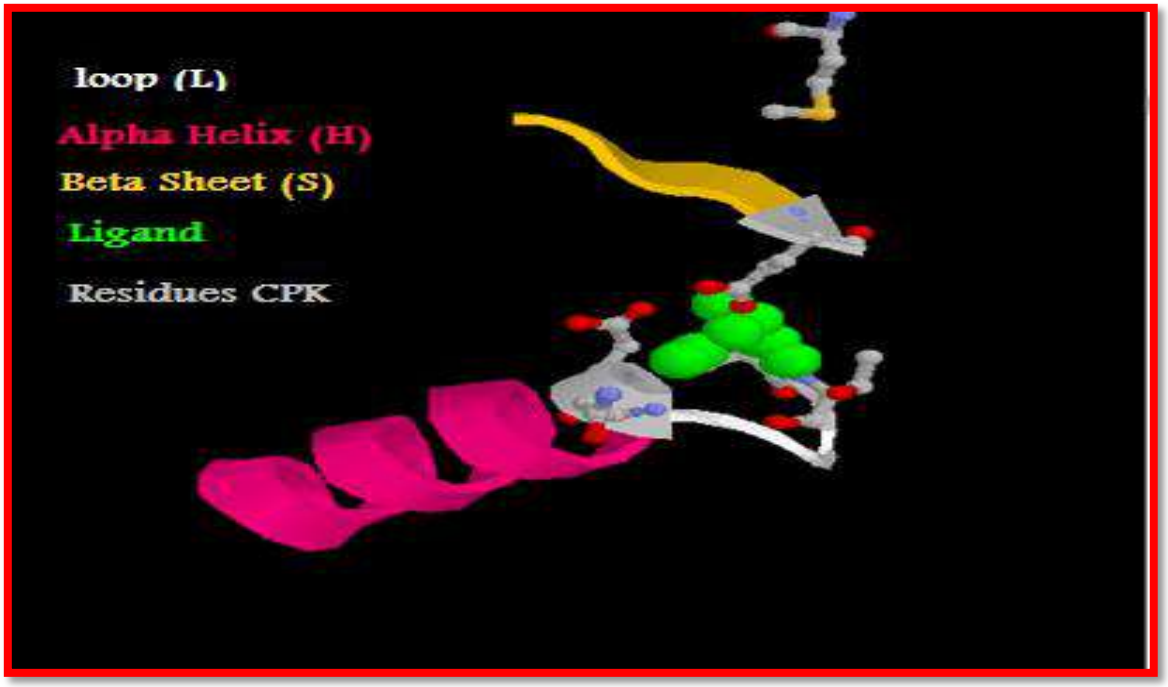


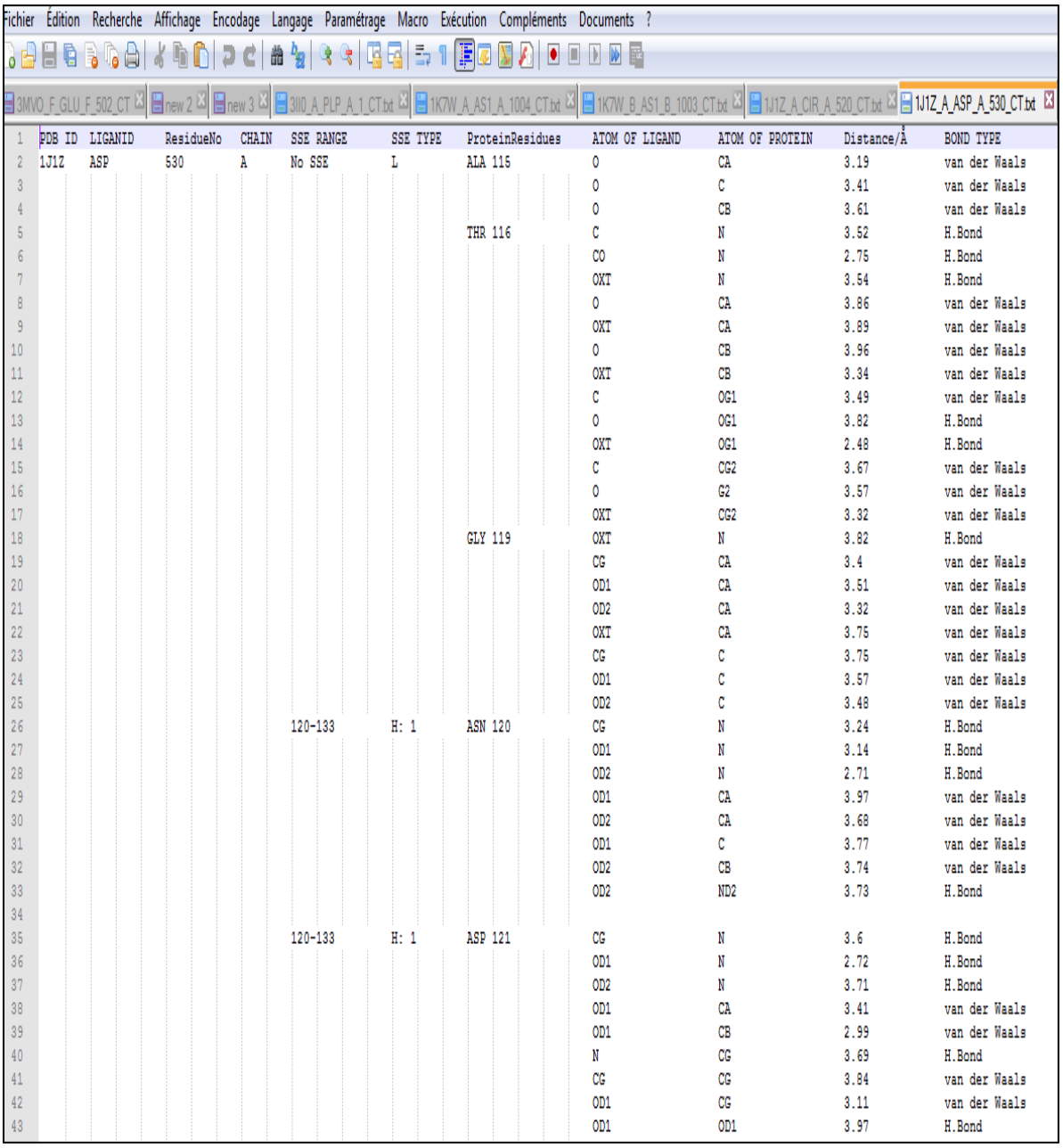
Figure n° 25-b: Capture of RasMol representation of the binding motifs where the binding residues and the ligand ASP are shown in the case of ASS (PDB id: 1J1Z, chain A).

It should be noted here that the three types of the graphical representations shown above for the case of the ligand ASP in complex with the enzyme ASS (PDB id: 1J1Z, chain A) are done for all of the ligand binding instances in all of the enzyme complexes studied in this project, see **Inde**x-III.

3. Data Storing and Flat-Files Database creation:

A Flat-File database , which is a simple schema type of a database, has been created, and to achieve this, the calculated binding details, seen above, have been stored into text based files (see **figure n°26**) and then stored in an arrangement that is based on the types of enzymes and PDB ids. The same treatment has been applied after storing the Rasmol graphical representations of the binding sites as shown in **Figure n°27.**

The binding details for all of the studied enzymes in this project have been stored in the same way as explained above, see **Index II**.



Fichier Édition Recherche Affichage Encodage Langage Paramétrage Macro Exécution Compléments Documents ?

3MVO_F_GLU_F_502_CT | new 2 | new 3 | 3II0_A_PLP_A_1_CT.txt | 1K7W_A_AS1_A_1004_CT.txt | 1K7W_B_AS1_B_1003_CT.txt | 1J1Z_A_CIR_A_520_CT.txt | 1J1Z_A_ASP_A_530_CT.txt

	PDB ID	LIGANID	ResidueNo	CHAIN	SSE RANGE	SSE TYPE	ProteinResidues	ATOM OF LIGAND	ATOM OF PROTEIN	Distance/Å	BOND TYPE
2	1J1Z	ASP	530	A	No SSE	L	ALA 115	O	CA	3.19	van der Waals
3								O	C	3.41	van der Waals
4								O	CB	3.61	van der Waals
5							THR 116	C	N	3.52	H.Bond
6								CO	N	2.75	H.Bond
7								OXT	N	3.54	H.Bond
8								O	CA	3.86	van der Waals
9								OXT	CA	3.89	van der Waals
10								O	CB	3.96	van der Waals
11								OXT	CB	3.34	van der Waals
12								C	OG1	3.49	van der Waals
13								O	OG1	3.82	H.Bond
14								OXT	OG1	2.48	H.Bond
15								C	CG2	3.67	van der Waals
16								O	G2	3.57	van der Waals
17								OXT	CG2	3.32	van der Waals
18							GLY 119	OXT	N	3.82	H.Bond
19								CG	CA	3.4	van der Waals
20								OD1	CA	3.51	van der Waals
21								OD2	CA	3.32	van der Waals
22								OXT	CA	3.75	van der Waals
23								CG	C	3.75	van der Waals
24								OD1	C	3.57	van der Waals
25								OD2	C	3.48	van der Waals
26					120-133	H: 1	ASN 120	CG	N	3.24	H.Bond
27								OD1	N	3.14	H.Bond
28								OD2	N	2.71	H.Bond
29								OD1	CA	3.97	van der Waals
30								OD2	CA	3.68	van der Waals
31								OD1	C	3.77	van der Waals
32								OD2	CB	3.74	van der Waals
33								OD2	ND2	3.73	H.Bond
34											
35					120-133	H: 1	ASP 121	CG	N	3.6	H.Bond
36								OD1	N	2.72	H.Bond
37								OD2	N	3.71	H.Bond
38								OD1	CA	3.41	van der Waals
39								OD1	CB	2.99	van der Waals
40								N	CG	3.69	H.Bond
41								CG	CG	3.84	van der Waals
42								OD1	CG	3.11	van der Waals
43								OD1	OD1	3.97	H.Bond

Figure n° 26: Notepad++ example of a file, named 1J1Z-A-ASP-A-530-ct.txt, which contains ligand binding details for the case of the ligand ASP and PDB id :1J1Z, Chain A.

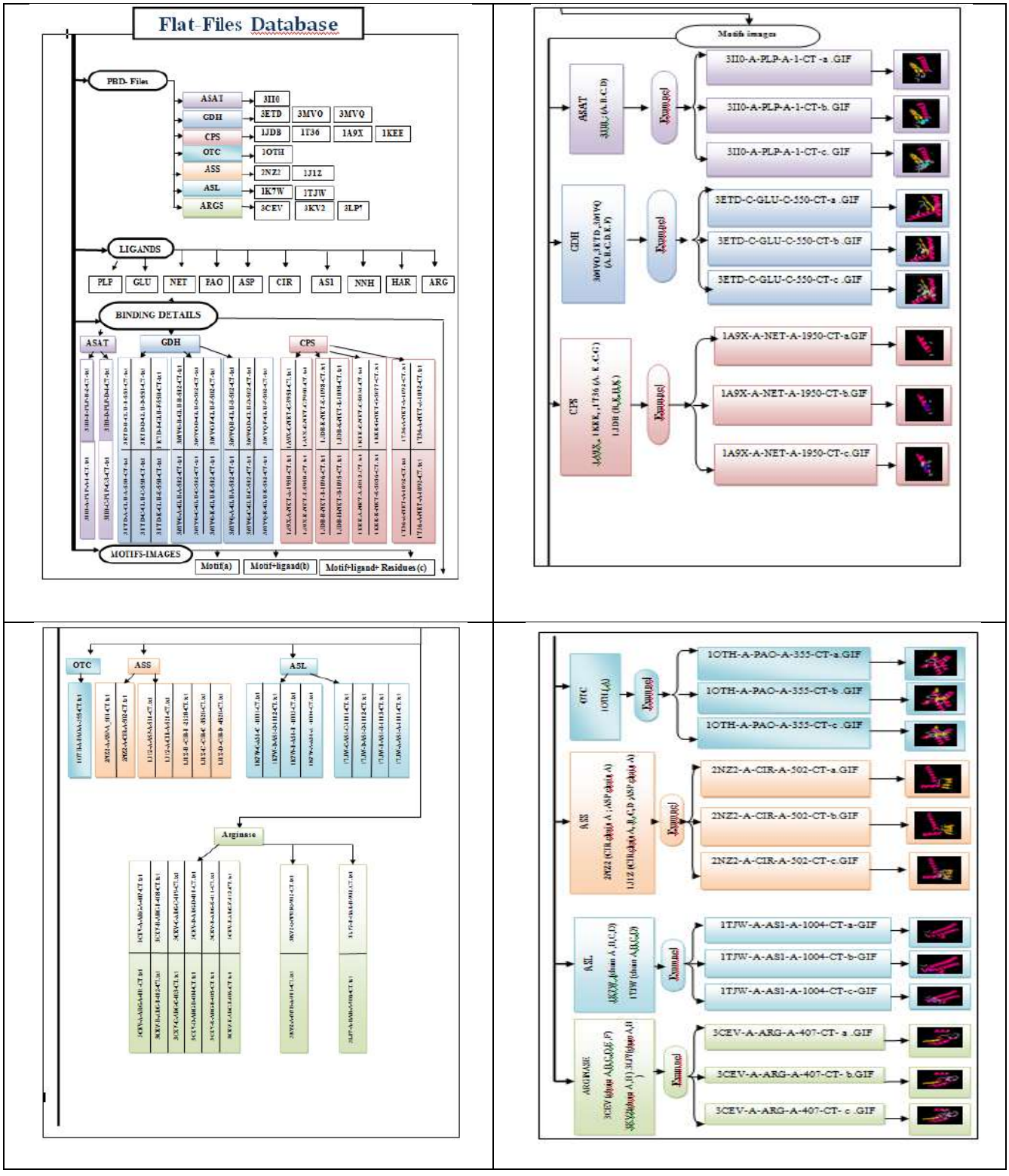


Figure n °27 : The database schema representing the architecture of the created Flat-File database**;** the **left side** show the arrangement and classification of the files containing the binding details while the **right side** of the figure show the arrangement of the graphics files containing the motifs in Rasmol representations.

4. World Wide Web Database:

In order to share the data and results with the scientific community both local and international, the Flat-Files database has been mounted on the server Bioinfornatics Tools[21] by the supervisor of this project, who also developed the programming scripts to make the database searchable.

This web version of the database has been named. Urea & Amino Acid Cycles Binding Structural & Functional Motifs (UadSFMs).

UadSFMs database has been created and made avaible at the web address:
http://bioinformaticstools.org/prjs/uadfms

1. Presentation of results:

1.1.Online Access and Database Querying:

The online version of the database "UadSFMs" can be uploaded by invoking the URL address shown in previous chapter, section: **World Wide Web Database.** The figure n°28 shows the web interface of the "UadSFMs" database developed specifically to render easy the experience of querying the data stored in the database.

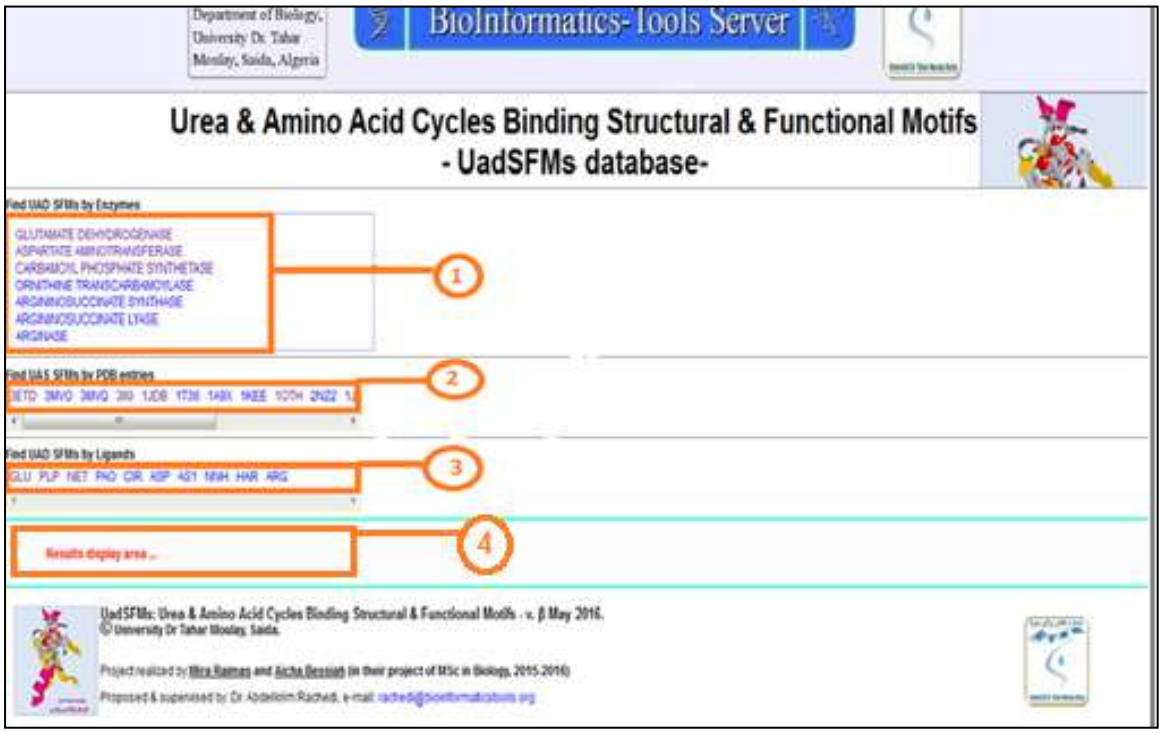


Figure n° 28: The main web interface of the **UadSFMs** database as captured from the web address, see next sections for explanations on the highlighted areas.

2.2. Database Methods of Querying and Results Display:

As shown above in Figure n°28, the interface of **UadSFMs** allows for 3 methods of searching the database content. For clarity these methods of querying are red colored and highlighted:

- **Area 1:** This list allows for querying by clicking on the amino acids degradation and urea cycles related enzymes.
- **Area 2:** This clickable list of PDB entries allows querying by PDB entry.
- **Area 3:** This clickable list of ligands allows querying by ligand id.

- **Area 4:** The "**Results display area …**" highlighted in green is the space area where querying results get displayed.

1. 2.1. Querying by Enzymes:

This method of query is allowed through selecting the enzyme to be explored by clicking on the hyperlinked enzymes list, **Area 1** as in **Figure n° 29**. The search produces an output page of results that display the ligand binding details that could be associated with more than one PDB entry.

1.2.2. Querying by PDB entries:

Ligands binding details per PDB entries can be searched by selecting the desired PDB entry and clicking on it from the hyperlinked PDB entries lists, **Area 2** as in **Figure n° 29**. The results page would display the binding details associated only with selected PDB entry.

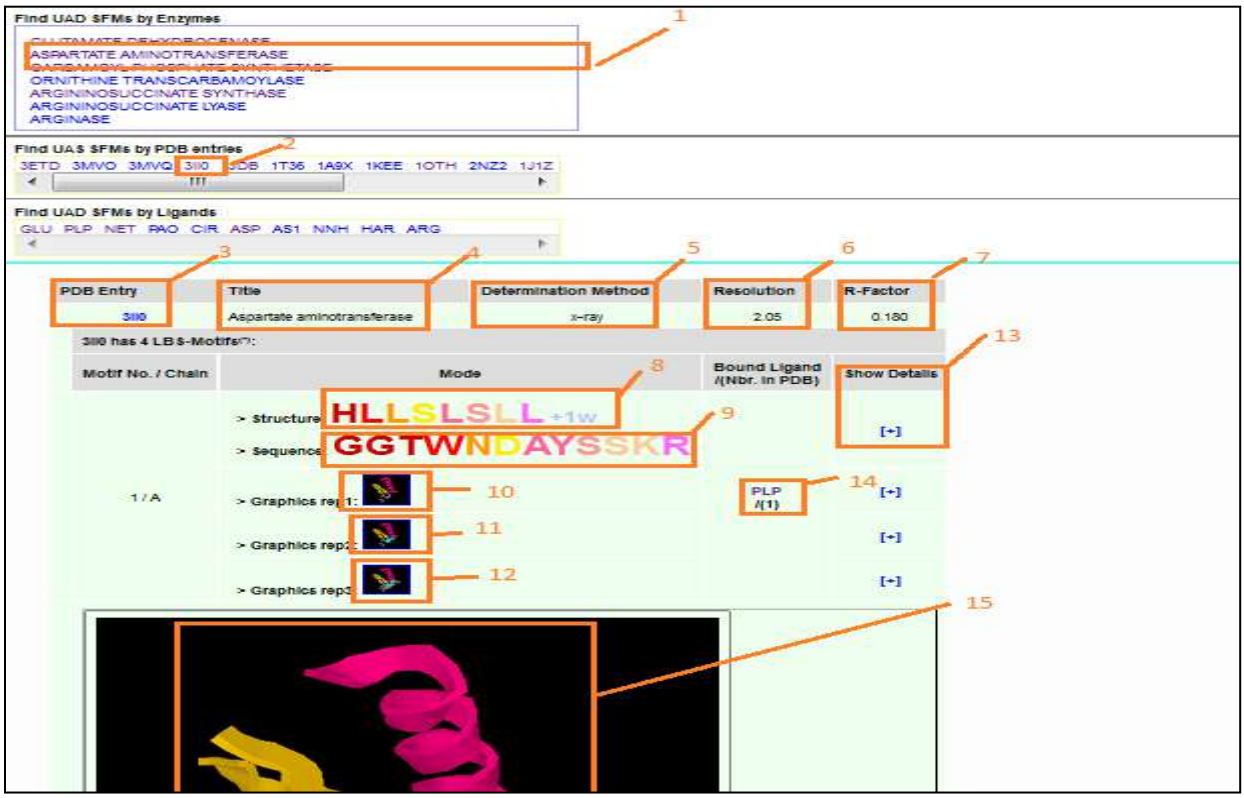


Figure n° 29: A screen-capture shows the two first method of searching the **UadSFMs** database **Area 1** and **Area 2**. The result displayed inside **Area 4** are the produced if the **search by PDB entries** method is implemented; **Area 3**: The

name of protein, **Area4:** Structure determination method, **Area** 5: Resolution, **Area6:** R-factor, **Area** 7**:** Motifs, **Area8:** Sequences., **Area** 9: presents picture of the motif, **Area** 10: Presents picture of the motif + ligand , **Area** 11: Presents picture of the motif + ligand+ Residue, **Area** 12: Link for details of contacts ,**Area13**: The contact between the ligand and the protein, **Area** 14: The image of the motif .

1.2. 3 . Querying by ligand id:

The research for the binding details by ligand names is done through selecting and clicking on the hyperlinked list of Ligands, **Area 1**, Figure n°30-a.

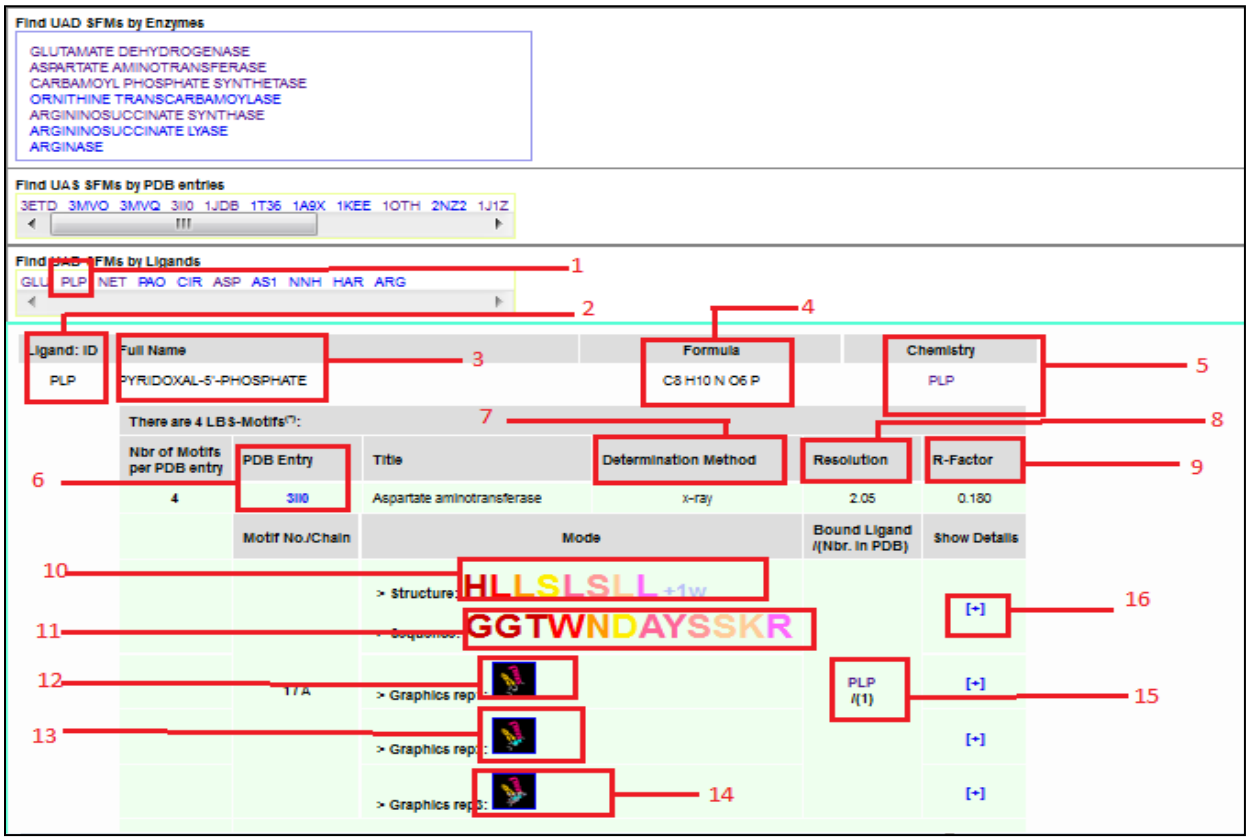
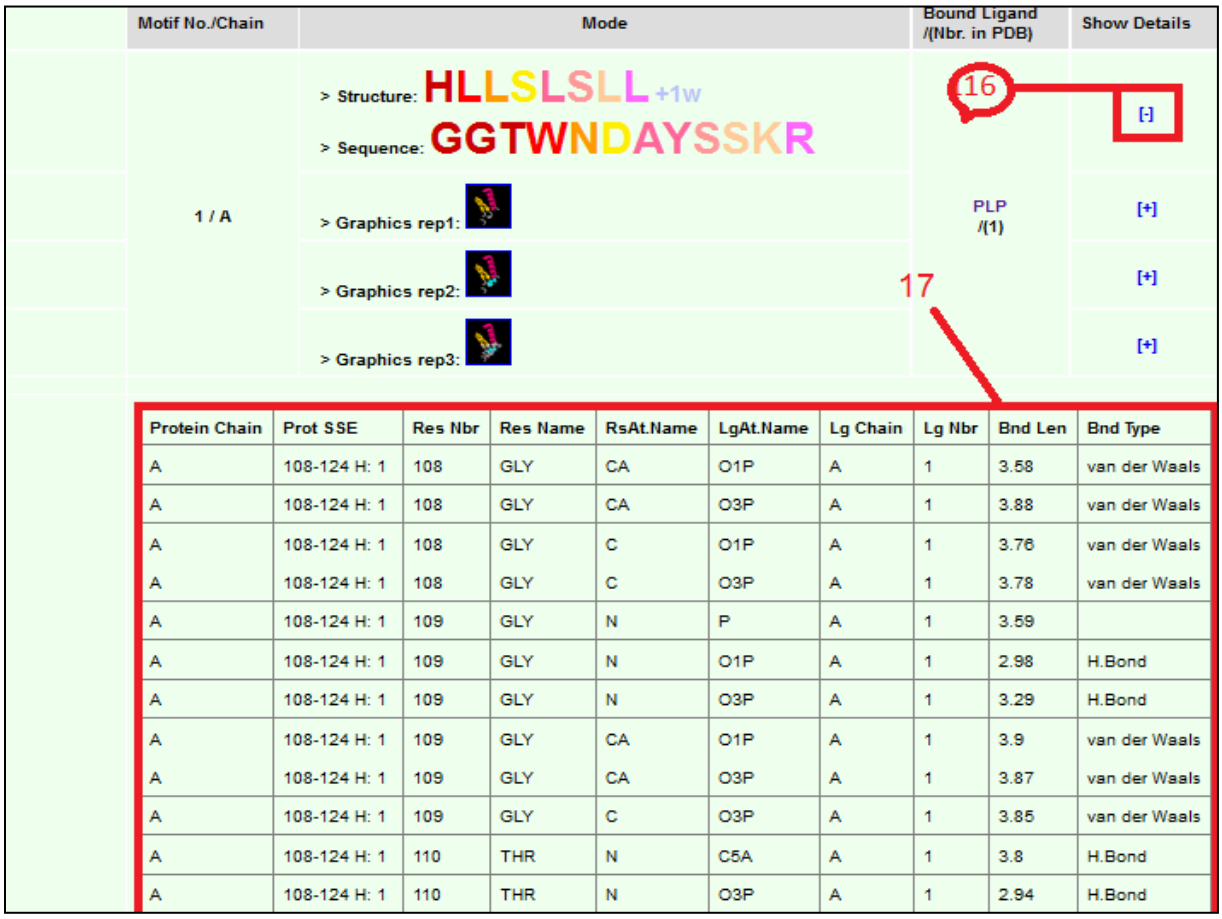


Figure n° 30-a: Screenshot shows the result page after the selecting and clicking on the ligand from the Ligands List, **Area 1**. The rest of the results display is represented as the following: **Area 2:** Ligand id, **Area 3**: Full name of the ligand, **Area 4:** Formula of the ligand, **Area 5:** Chemistry of the ligand .

It's here noted that this method of searching by ligands is more comprehensive compared to the other methods in that the ligand binding results are shown per pdb enteries and per enzyme types.

If clicked, the url link shown in highlighted **Area 16**, ligand binding details will be shown, as seen the highlighted **Area 17**. The ligand binding details are displayed in the manner that each existing contact shows the ligand atoms involved in the binding together with the residue atoms and the particular secondary structure element they belong to in the protein associated with the PDB entry depicted in the results. This is in addition of showing the types of bonds and their lengths, see Figure n°30-b.



Protein Chain	Prot SSE	Res Nbr	Res Name	RsAt.Name	LgAt.Name	Lg Chain	Lg Nbr	Bnd Len	Bnd Type
A	108-124 H: 1	108	GLY	CA	O1P	A	1	3.58	van der Waals
A	108-124 H: 1	108	GLY	CA	O3P	A	1	3.88	van der Waals
A	108-124 H: 1	108	GLY	C	O1P	A	1	3.76	van der Waals
A	108-124 H: 1	108	GLY	C	O3P	A	1	3.78	van der Waals
A	108-124 H: 1	109	GLY	N	P	A	1	3.59	
A	108-124 H: 1	109	GLY	N	O1P	A	1	2.98	H.Bond
A	108-124 H: 1	109	GLY	N	O3P	A	1	3.29	H.Bond
A	108-124 H: 1	109	GLY	CA	O1P	A	1	3.9	van der Waals
A	108-124 H: 1	109	GLY	CA	O3P	A	1	3.87	van der Waals
A	108-124 H: 1	109	GLY	C	O3P	A	1	3.85	van der Waals
A	108-124 H: 1	110	THR	N	C5A	A	1	3.8	H.Bond
A	108-124 H: 1	110	THR	N	O3P	A	1	2.94	H.Bond

Figure n° 30-b: Screenshot shows the result page after selecting a ligand form the Ligands List. The display show mainly the table of the binding details made by the ligand with the associated protein.

2. Binding Motifs and Properties:

The total of 10 ligands studied in this project bound to 70 protein chains associated with the 16 PDB entries. This resulted in the total number of 70 motif instances of which to 21 binding motifs are unique as shown below in Table n° 06.

Ligand ID	Number of contacts	Number of motifs	Motifs
PLP	4	1	HLSLSL
GLU	18	8	SLHSLH, LHLHLH, LHLHH, LHLHLHH, SHSSLH,SHSSLHLH , SSHSSLH, SLHSLHLH
NET	16	1	LHL
PAO	1	1	LHSLHLHLH
ASP	2	2	LHL, LHS
CIR	5	2	LHLSSLSSH , LHSSSSH
AS1	8	2	HHLHH, HHHLHH
ARG	12	3	HHL , HHHL , HLHLHSHL
NNH	2	1	LHHSHL
HAR	2	2	HLHHSHL , LHHSHL

Table n°6: The ligands and their binding motifs.

In the following section description of the ligands binding motifs, their properties and 3D representations as per the enzymes and the reaction involved:

2.1. Amino Acids Degradation Reactions:

2.1.1. Aspartate Aminotransferase (ASAT) Reaction:

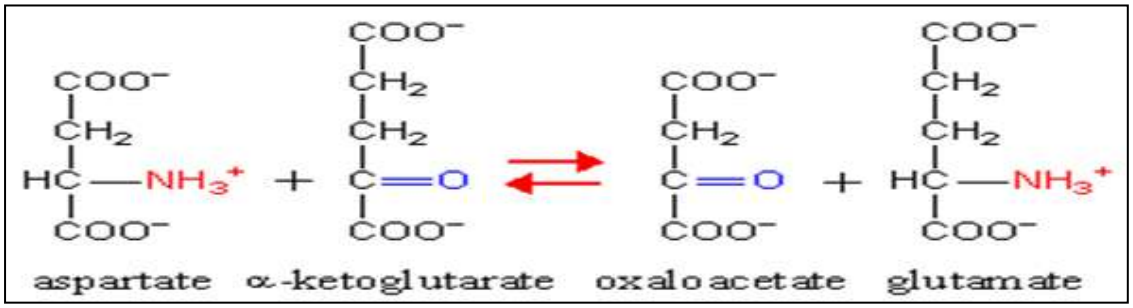


Figure n° 31: Aspartate transaminase reaction. Explanation of the reaction can be found Chapter 1, page 15.

2.1.1.1. Binding Motifs & their Properties for the ASAT:

❖ The **Aspartate Aminotransferase** in the PDB entry, 3II0, is represented by 04 protein chains where each is bound to an instant of the PLP cofactor, see table n° 7.

❖ The PLP ligand binds therefore only one type of motifs; **HLSLSL** repeated with each protein chain**.**

❖ The motif instances is made of a mixture of **α-helices** (H), **β-strand** (S) and **loop** regions (L).

LIGAND ID	PDB ID	CHAIN	MOTIF
PLP	3II0	A	HLSLSL
		B	
		C	
		D	

Table n °7: Types of motifs linked with the ligand PLP.

2.1.1.2. Graphical Representation of the Binding Motif for the ASAT:

❖ As seen above, the **ASAT** is associated with only one PDB entry, 3II0, which is in complex with only one ligand; the PLP.

❖ The images below represent the Rasmol 3D-graphical representation of the binding associated with the PDB entry ,3II0, Chains A and B, table n° 8.

❖ The 3D representation of the other ligand binding motif instances related to the chains C & D have been created and are stored in the Flat-Files database and in the online version **UadSFMs**, see also Index II.

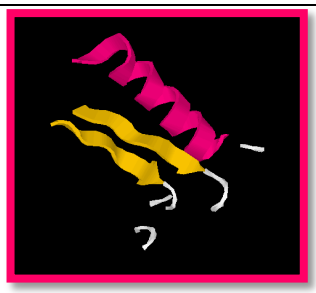
Codes	Chain	Ligand	Motifs	Motif+Ligand	Motif+Ligand+Residues
3II0	A	PLP			
	B	PLP			

Table n° 8: 3D representation of the binding motifs associated with ASAT from the PDB entry 3II0 (chains A & B).

2.1.2. Glutamate Dehydrogenase (GDH) Reaction:

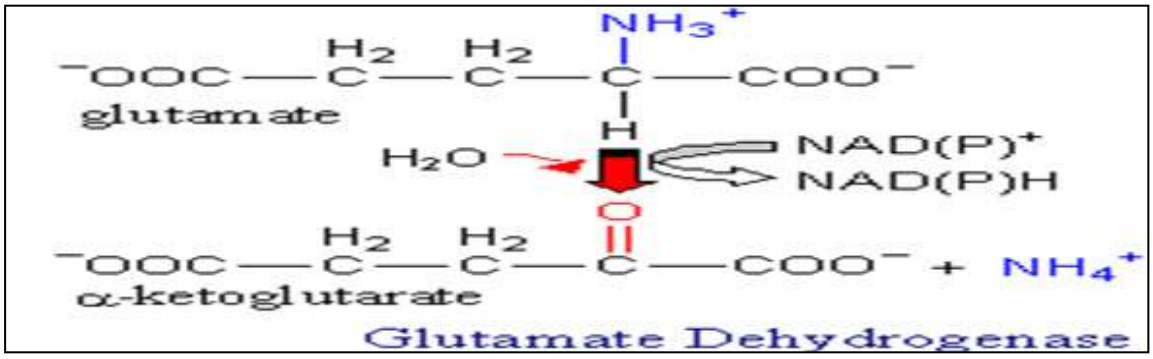


Figure n° 32: Reversible reaction catalyzed by GDH. Details of the reaction are given in Chapter 1, page 17.

2.1.2.1. Binding Motifs and their Properties for the GDH:

- The GDH in all of the related PDB entries is represented by 18 protein chains where each is bound to an instant of the ligand GLU which is the natural substrate, see table n° 9.
- The Ligand GLU bind eight (8) unique motifs which are: **SLHSLH, LHLHLH, LHLHH, LHLHLHH, SHSSLH, SHSSLHLH, SSHSSLH, SLHSLHLH.**
- Five (5) of these 8 motifs describe a mixture of **α-helices** (H), **β-strands** (S) and **loop** regions (L) and the other three (3) motifs are mixture of **α-helices** (H), and **loop** regions only (**L**), i.e. they lack β-strand element.

Ligand id	Pdb id	Chain	Motif
GLU	3ETD	A	SLHSLH
		B	SLHSLH
		C	SLHSLH
		D	SLHSLH
		E	SLHSLH
		F	SLHSLH
	3MVO	A	LHLHLH
		B	LHLHLH
		C	LHLHLH
		D	LHLHLH
		E	LHLHH
		F	LHLHLHH
	3MVQ	A	SHSSLH
		B	SHSSLH
		C	SLHSLH
		D	SHSSLHLH
		E	SSHSSLH
		F	SLHSLHLH

Table n °9: Types of motifs linked with ligand GLU.

2.1.2.2. Graphical Representation of Binding Motif for the GDH:

- As seen above, in the case of the GDH, there are 3 pdb entries: 3ETD, 3MVO, 3MVQ binding the natural substrate GLU.
- The images below, table 10, represent the RasMol 3D-graphical representation of the binding associated with the PDB entries: 3ETD (Chain:A), 3MVO (Chain:A), 3MVQ (Chain: D).

❖ The 3D representation of all the other binding motifs related to the other chains and their bound ligands have been created and are stored in the Flat-Files database and in the online version **UadSFMs**, see Index II.

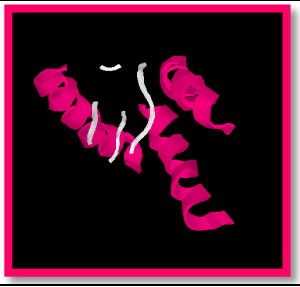
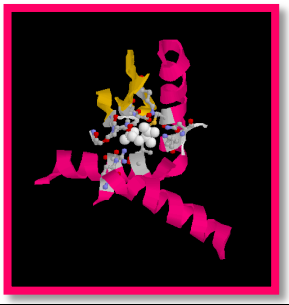
Codes	Chain	Ligand	Motifs	Motif+Ligand	Motif+Ligand+ Residues
3ETD	A	**GLU**			
3MVO	B	**GLU**			
3MVQ	D	**GLU**			

Table n°10: 3D representation of the binding motifs associated with GDH of the PDB entries 3ETD (chains A), 3MVO(chains B), 3MVQ (chain D).

2.2. Urea Cycle Reactions:

2.2.1. Carbamoyl-phosphate synthetase I (CPS) Reaction:

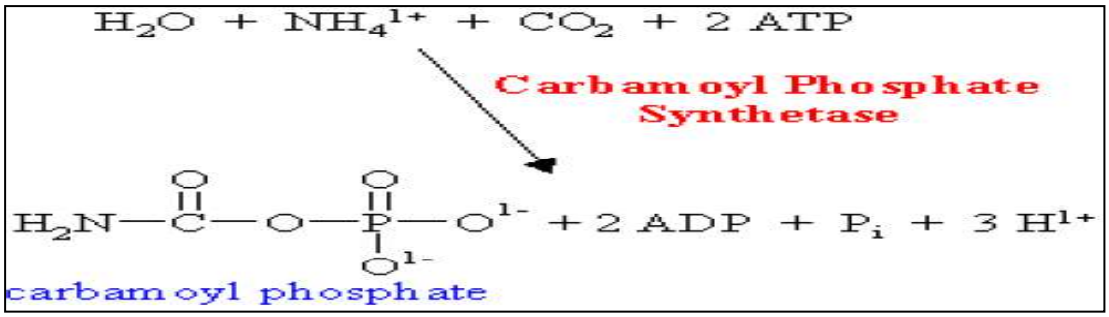


Figure n° 33: Formation of carbamoyl phosphate (step 1). Explanation of the reaction can be found in Chapter 1, page 19.

2.2.1.1. Binding Motifs & their Properties for the CPS:

- Carbamoyl-phosphate synthetase in all of the related PDB entries is represented by 16 protein chains where each is bound to an instant of the NET analogue to the natural substrate; showing one unique ligand binding motif, table n° 11.
- The only type of binding motif which bind the NET ligand is: **LHL**
- This motif is a mixture of a single **α-helix** (**H**) and **loop** regions **(L).**

Ligand id	Pdb id	Chain	Motif
NET	1JDB	B	LHL
		E	LHL
		H	LHL
		K	LHL
	1T36	A	LHL
		C	LHL
		E	LHL
		G	LHL
	1A9X	A	LHL
		C	LHL
		E	LHL
		G	LHL
	1KEE	A	LHL
		C	LHL
		E	LHL
		G	LHL

Table n° 11: Types of motifs linked with the NET ligand.

2.2.1.2. Graphical Representation of Binding Motif for the CPS:

- In the case of the Carbamoyl-phosphate synthetase, we have **04** PDB entries: 1JDB, 1T36, 1A9X, 1KEE binding a total of one ligand that is analogues to the natural substrate.
- The images below, table n°12, represent the RasMol 3D-graphical representation of the binding site associated with the PDB entries 1JDB (Chain E) and 1KEE (Chain A) .
- The 3D representation for all the other binding motifs related to the rest of chains composing the PDB entries 1JDB ,1KEE, 1A9X and 1T36 and their bound ligands have been created and are stored in the Flat-Files database and in the online version **UadSFMs** see also **Index. II.**

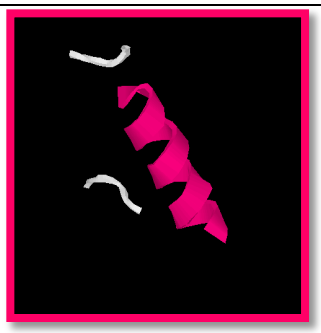
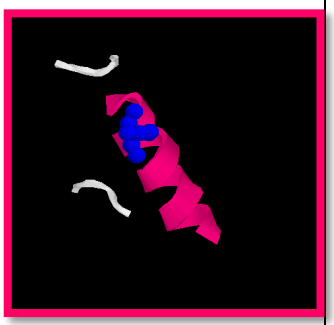
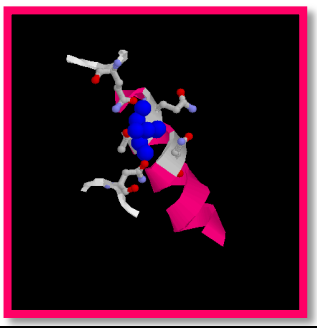
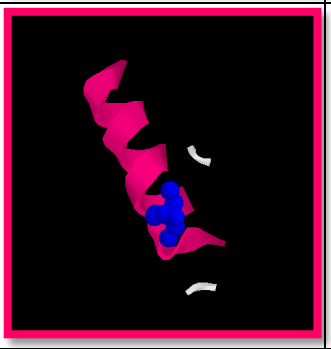
Codes	Chain	Ligand	Motifs	Motif+Ligand	Motif+Ligand+Residues
1JDB	**E**	**NET**			
1KEE	**A**	**NET**			

Table n°12: 3D representation of the binding motifs associated with CPS of the PDB entries 1JDB (chain E), 1KEE (chain A).

2.2.2. Ornithine Transcarbamoylase (OTC) Reaction :

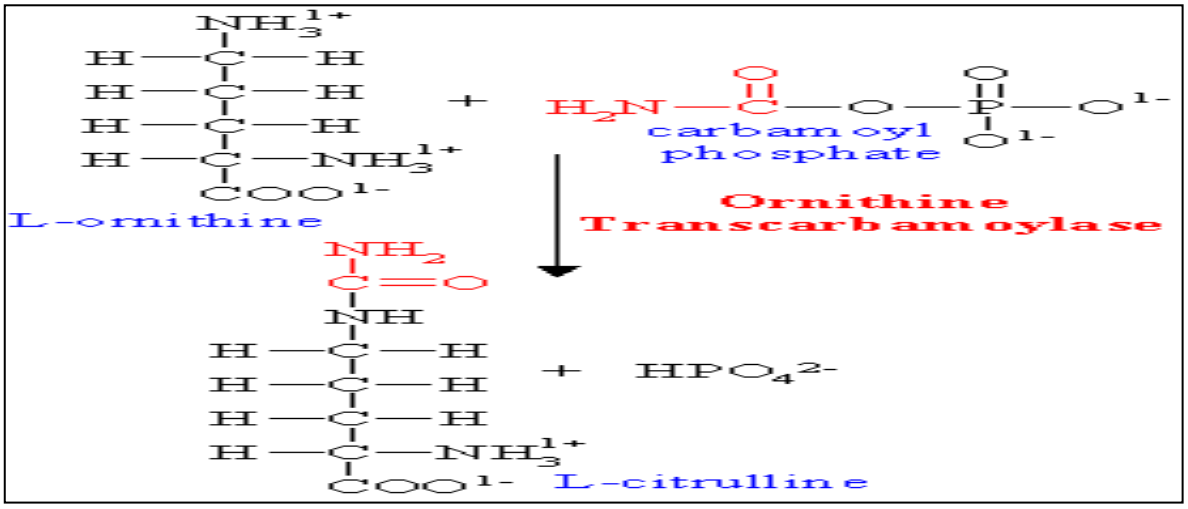


Figure n° 34: Formation of citrulline.. Details of the reaction are given in Chapter 1 page19.

2.2.2.1. Binding Motifs and their Properties for the OTC :

- The Ornithine Transcarbamoylase in all of the related one PDB entry is represented by one protein chain where it is bound to an instant of the PAO ligand which is an analogue of the natural substrate ornithine. Only binding motif exist in this case, table n°13.
- The PAO binds the only motif type: **LHSLHLHLH.**
- This motif describe a mixture of **α-helices (H)**, **loop** regions **(L)** and one **β-strand (S).**

Ligand id	Pdb id	Chain	Motif
PAO	1OTH	A	LHSLHLHLH

Table n° 13: Types of motifs linked to PAO.

2.2.2. 2.Graphical Representation of the Binding Motif for the OTC:

- As seen above, in the case of the OTC, there is **01** PDB entry: 1OTH binding a single ligand PAO which in turn binds only one type of binding motif.
- The images below, table n°14 represent the RasMol 3D-graphical representation of the binding motif associated with the PDB entry 1OTH (Chain A).

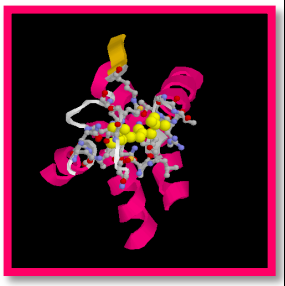
Codes	Chain	Ligand	Motifs	Motif+Ligand	Motif+Ligand+Residues
1OTH	**A**	**PAO**			

Table n°14: 3D representation of the binding motifs of the PAO ligand associated with the PDB entry 1OTH (chain A).

2.2. 3. Argininosuccinate Synthethase (ASS) Reaction :

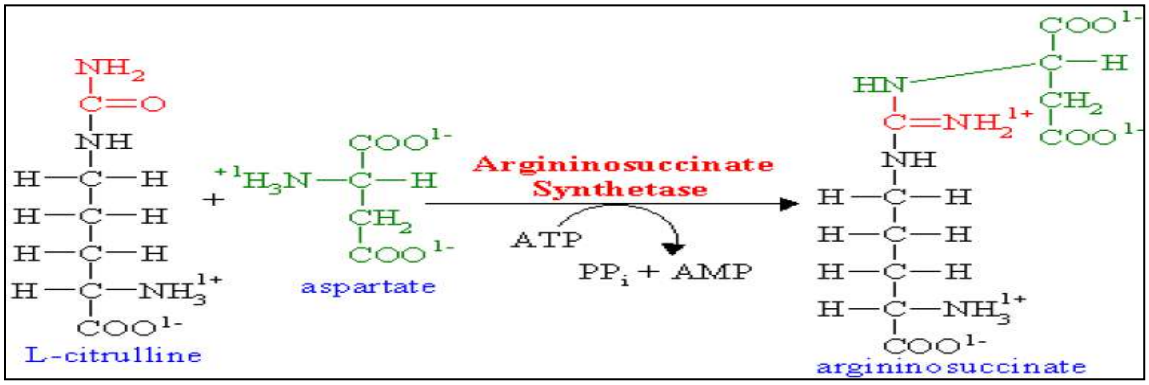


Figure n°35: Formation of arginosuccinate (Step 3). Explanation of the reaction can be found in Chapter 1, page 20.

2.2.3.1. Binding Motifs and their Properties for the ASS:

- The Argininosuccinate Synthethase in all of the related PDB entries which are represented by 07 protein chains where each is bound to an instant of the natural substrats Aspartate and Citruline thus the existence of 07 motifs, table n° 15.
- This enzyme binds to two (2) ligands: Aspartate (ASP) and Citrulline (CIR); the ASP bind 2 different motifs which are: **LHL** and **LHS.** The CIR binds also 2 different motifs which are: **LHLSSLSSH** and **LHSSSSH**.
- These motifs are of two types: one composed of an **α-helix (H),** and two **loop** regions **(L)** which is bind only the ASP ligand**.** The other type of motifs describe a mixture of **α-helices (H), β-strands (S)** and **loop** regions **(L)** that bind both of the ligands APS and CIR**.**

Ligand id	Pdb id	Chain	motif
ASP	2NZ2	A	LHL
	1J1Z	A	LHS
CIR	2NZ2	A	LHLSSLSSH
	1J1Z	A	LHSSSSH
		B	LHSSSSH
		C	LHSSSSH
		D	LHSSSSH

Table n° 15: Types of binding motifs linked to ASP and CIR.

2.2.3.2. Graphical Representation of the Binding Motif for the ASS:

- As seen above, in the case of the **ASS**, there are **02** PDB entries: **2NZ2, 1J1Z** binding a total of two ligands which are natural substrates.
- The images below, table n°16 represent the RasMol 3D-graphical representation of the binding motifs associated with the PDB entry **2NZ2** (Chain A) , ligand The
- 3D representation of all the other binding motifs related to the rest of chains associated with the PDB entry 1J1Z and their bound ligands have been created and are stored in the Flat-Files database and in the online version **UadSFMs** see also **Index. II.**

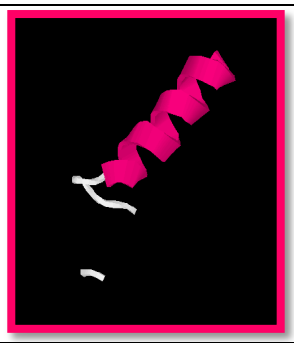
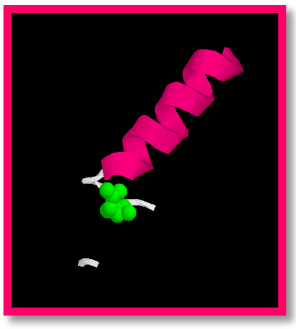
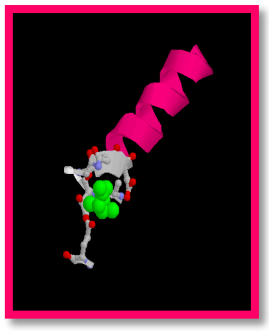
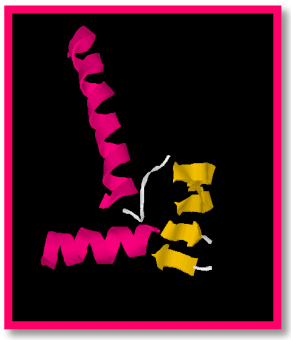
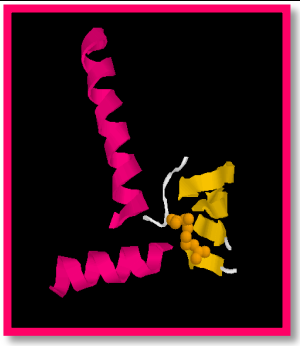
Codes	Chain	Ligand	Motifs	Motif+Ligand	Motif+Ligand+ Residues
2NZ2	A	ASP			
	A	CIR			

Table n° 16: 3D representation of the binding motifs associated with ASS of the PDB entries 2NZ2 (chain A). Ligand ASP,CIR..

2.2.4. Argininosuccinate lyase (ASL) Reaction :

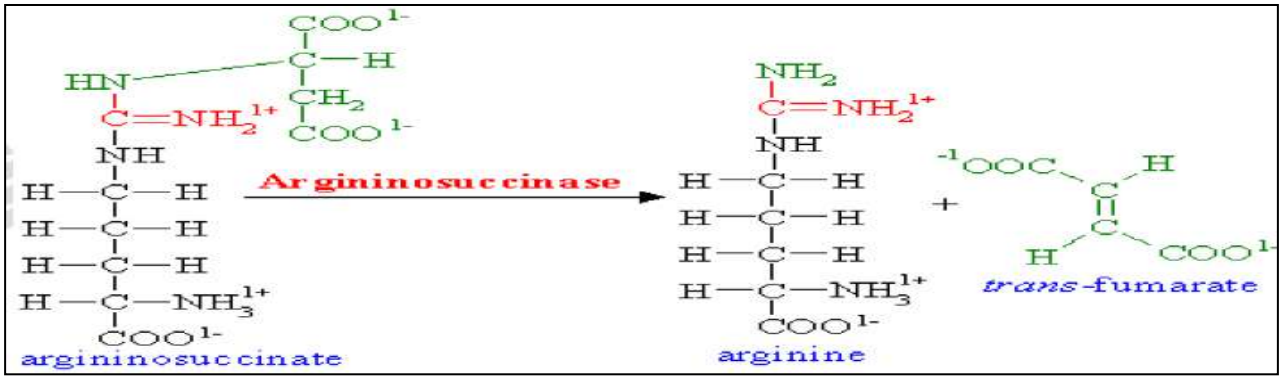


Figure n° 36: Formation of arginine (step 4). Details of the reaction are given in Chapter 1 page 20.

2.2.4.1. Binding motifs and Properties for the ASL :

- The Argininosuccinate lyase in all of the related PDB entries is represented by 08 protein chains where each is bound to an AS1 thus the existence of 06 motifs, see table n°17.
- The ligand AS1 binds two type of motifs : **HHLHH, HHHLHH.**
- This motifs describe a mixture only of **α-helices** and **loop** regions **(L).**

Ligand id	Pdb id	Chain	Motif
AS1	1K7W	A	HHLHH
		B	HHLHH
		C	HHHLHH
		D	HHHLHH
	1TJW	A	HHLHH
		B	HHLHH
		C	HHLHH
		D	HHLHH

Table n° 17: Types of motifs linked to AS1**.**

2.2.4.2. Graphical Representation of Binding Motif for the ASL:

- As seen above, in the case of the ASL there are **02** pdb entries: **1K7W , 1TJW** binding a the same ligand AS1which is natural substrate. see table n°18.
- The images below, table n°18 represent the RasMol 3D-graphical representation of the binding motif associated with the PDB entries : **1K7W (Chain D), 1TJW (ChainA).**
- The 3D representation of all the other binding Motifs related to the other chains and their bound ligands have been created and are stored in the Flat-Files database and in the online version **UadSFMs** see also **Index II**.

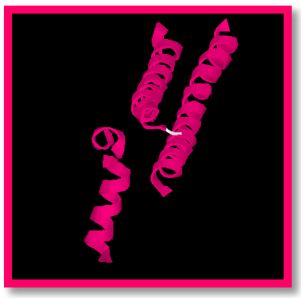
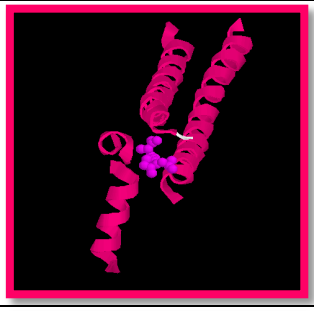
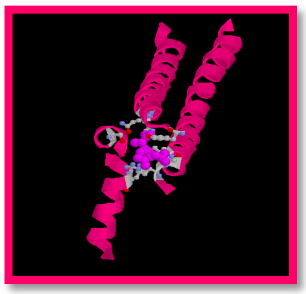
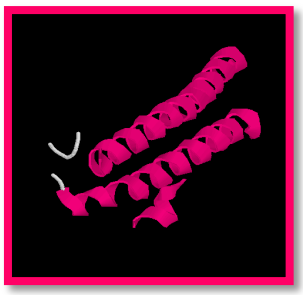
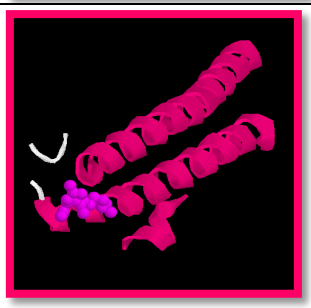
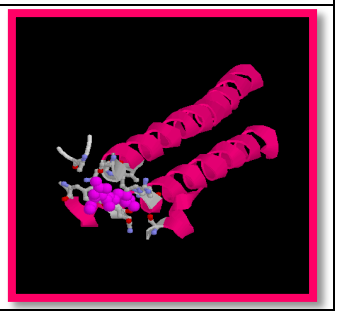
Codes	Chain	Ligand	Motifs	Motif+Ligand	Motif+Ligand+Residues
1K7W	D	AS1			
1TJW	A	AS1			

Table n° 18: 3D representation of the binding motifs associated with ASL of the PDB entries 1K7W (Chain D), 1TJW (Chain D).

2.2.5. Arginase (ARGS) Reaction:

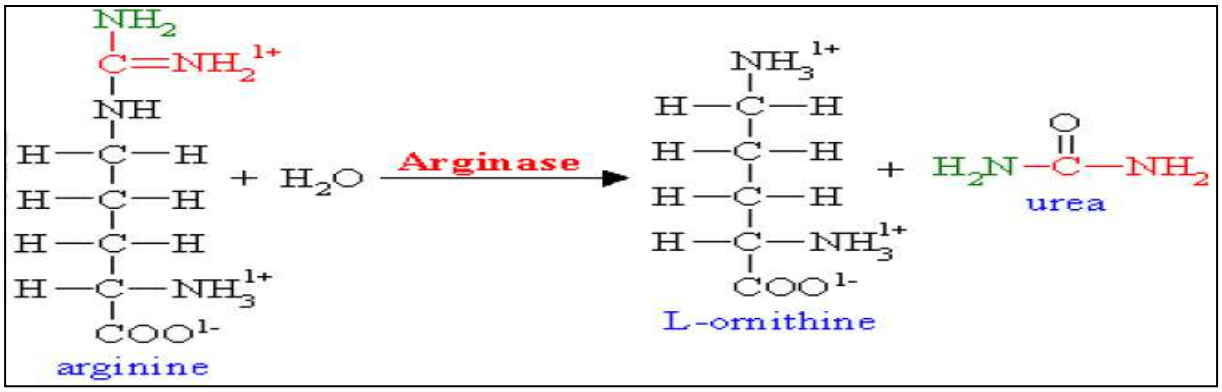


Figure n° 37: Formation of ornithine and urea (step 5). Details of the reaction are given in Chapter 1 page 21.

2.2.5.1. Binding motifs and Properties for the ARGS:

- The ARGS in all of the related PDB entries is represented by 12 protein chains where each is bound to an NNH, HAR , ARG thus the existence of 05 motifs, see table n°19

- NNH and HAR. bind the motif **LHHSHL** .
- HAR bind the motif **HLHHSHL** .
- ARG bind the motifs: **HHL** , **HHHL , HLHLHSHL** .
- Three (3) of these 5 motifs describe a mixture of **α-helices** (H), **β-strands** (S) and **loop** regions (L) and the other two (2) motifs are mixture of **α-helices** (H), and **loop** regions only

Ligand id	Pdb id	Chain	Motif
ARG	3KV2	A	LHHSHL
		B	LHHSHL
	3LP7	A	HLHHSHL
		B	LHHSHL
	3CEV	A	HHL
		B	HHL
		C	HHL
		D	HHHL
		E	HHHL
		F	HHL
		A	HLHLHSHL
		B	HLHLHSHL
		C	HLHLHSHL
		D	HLHLHSHL
		E	HLHLHSHL
		F	HLHLHSHL

Table n°19: Types of motifs linked to ARG and its analogs.

2.2.5.2. Graphical Representation of Binding Motif for the ARG:

- . In the case of the ARGS, we have **02** PDB entries: 3KV2, 3LP7, binding a total of two ligand that is analogues to the natural substrate and one pdb entry : 3CEV binding natural substrate .see table n°20.
- The images below represent the RasMol 3D-graphical representation of the binding associated with the PDB entries 3KV2 (Chain A) , 3LP7 (B) , and 3CEV (Chain C) .
- The 3D representation of all the other binding Motifs related to the other chains and their bound ligands have been created and are stored in the Flat-Files database and in the online version **UadSFMs** see also **Index II**.

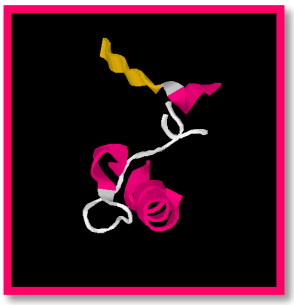
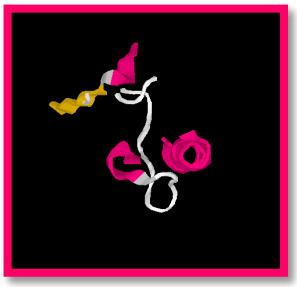
Codes	Chain	Ligand	Motifs	Motif+Ligand	Motif+Ligand+ Residues
3KV2	A	NNH			
3LP7	B	HAR			
3CEV	C	ARG			

Table n° 20: 3D representation of the binding motifs associated with ARGS of PDB entries 3KN2 (chain A), 3LP7 (Chain B),3CEV (Chain C).

3. Ligands and Binding Residue types:

The binding environment details calculated for the 10 ligands, as reported in Chapter II and Index II, have been used to explore the types of the residues (amino acids while peptide bound each other in the protein chain) that **actually do the act of binding the ligands** in the structural motifs described above.

After the analysis of the amino acid distribution at the ligands binding motifs, it was found that from the known 20 amino acids, only few amino acids were more frequent like ASN,GLY,ASP,SER, LEU,VAL,ARG,LYS,TYR,GLU,THR,PRO,GLU in this ligands

The types of residues linked to ligands are shown in table n°21.

Ligand ID	Proteins Residues
PLP	ASN ; GLY ;ARG ; ASP ; THR ; TYR ; SER ; LYS
GLU	LYS;GLY; MET; ALA; PRO; ASP; ARG; ASN; VAL; SER; THR
NET	VAL; GLN; THR; ASN
PAO	SER; THR; ARG; LEU; HIS; GLN; ASN; ILE; ASP; MET; CYS
CIR	TYR; THR; SER; ASN; ASP; ARG; MET; GLU
ASP	ALA; THR; GLY; ASN; ASP; GLU
AS1	SER; ARG; ASN; VAL; ALA ; TYR; GLN; LYS; HIS
NNH	ASP; HIS; ASN ; SER ; GLY; GLU; THR
HAR	HIS; ASP; ASN; SER;GLY; GLU; THR
ARG	GLU; SER; LEU; ARG ; PRO;GLY; ASP; THR; ASN

Table n° 21: Residues type shown arranged by their bound ligands. The residues are coloured per hydrophobicity, see figure n°38.

Water Affinity			Name
Highly Hydrophobic	1		Isoleucine
	2		Phenylalanine
	3		Valine
	4		Leucine
	5		Methionine
	6		Tryptophan
	7		Alanine
	8		Glycine
	9		Cysteine
	10		Tyrosine
	11		Proline
	12		Threonine
	13		Serine
	14		Histidine
	15		Glutamate
	16		Asparagine
	17		Glutamine
	18		Aspartate
Highly Hydrophilic	19		Lysine
	20		Arginine

Figure n° 38: Amino Acids colored per hydrophobicity [22].

As seen above, table n°21, it seems that the properties of binding residues governing the chemical environment of the binding motifs are mostly hydrophilic with some low level of hydrophobicity . See **General Conclusion**.

4. Ligands Binding Tendency and Motifs Classification:

As seen above, the properties analysis of the motifs suggest that the binding motifs associated with the enzymes involved in the Amino Acides degradation and urea cycle pathway can be classified into the following families:

4. 1. α/Loop family:

All Motifs belong to this family contain α-Helix (**H**) and Loop (**L**) elements such as:

LHLHLH, LHLHH , LHLHLHH , LHL, HHLHH, HHHLHH, HHL, HHHL

The different ligands bind this type of motifs are : GLU ; NET;ASP;AS1;ARG.

4.2. α/β/Loop family:

All Motifs belong to this family contain α-Helix (**H**) , β-strand (**S**) and Loop (**L**) elements such as:

HLSLSL, SLHSLH, SHSSLH, HLHHSHL, LHLSSLSSH, LHSSSSH,

LHHSHL, HLHHSHL, HLHLHSHL, LHS, SHSSLHLH

SSHSSLH, SLHSLHLH, L HSLHLHLH .

The different ligands that bind these types of motifs are: PLP, GLU, PAO, ASP CIR, NNH, HAR, ARG.

The classification above declare that the types of ligands binding the enzymes involved in the Amino Acids degradation and Urea Cycle pathways bind only two classes of binding motifs; **α/β/Loop family** the **α/Loop family.**

It is important to note that this classification reinforces the notion that secondary structure elements do play major roles but with the contribution from non-secondary structure elements, loops, to fulfill the biological function carried out by the enzymes treated in this study.

In addition, the tendency of some type of amino acids to be responsible for binding ligands, shown by the analysis above, suggests that some amino acids may be inclined to run functional roles and other may be more associated with a structural role in the maintenance of the structural morphology of proteins.

The information provided by the motifs classification, done above, versus the ligands binding tendency and the properties of the residues responsible for the actual binding of the ligands, outlined also above, can be instrumental in the process of designing new drugs depending on the type of binding motif targets and the binding tendency of the drugs, for example in cases where *de novo* drug design is needed for treating pathogenic metabolism situations.

General Conclusion:

This project has set out to try and contribute in understanding the basis of Structure-Function relationship in macromolecules; protein is the case of this study. This relationship seems to be **coded by the amino acids** that compose enzymes and proteins in general though the **protein folding** process.

As shown in the various analysis and deductions made in the **Results and Discussions** (Chapter III), this project has identified, defined and characterized the protein structural elements dubbed here as the binding motifs (Structural & Functional Motifs) together with the binding residues (amino acids) that are directly involved in the ligand binding process and hence the function of the enzymes associated with the amino acids degradation and urea cycle.

The structural elements (a-helices, b-strand and loops) in the defined and characterized binding motifs are seen by this study as providing the structural support on which the functional elements, i.e. the residues, can reside to carry out the biological function of the enzymes.

The definition of the ligand binding sites in the form of structural motifs and storing them into a database can be very useful in finding similar motifs in other protein which may be of the same and/or different function. This may helps in a lot of studies for protein classifications, taxonomy, phylogenetics and homology molecular modeling.

This study has also reinforced the important role of secondary structure elements and more importantly the crucial role of the non-secondary structure regions (loops) in the biological function of protein.

However, the above conclusions related to the motifs' classification and the ligands binding tendency and their potential usages are, in this study, limited to the enzymes of the amino acids degradation and urea cycle pathway. In order to find out whether the conclusions can be generalized, there is the need to corroborate this kind of study by analyzing larger sets of enzymes associated with different types of metabolic pathways.

References

[1].Dissertation for obtaining the academic degree of the Doctor of natural sciences (Dr. rer. nat.) filed in the Department of biology, chemistry, pharmacy submitted by Francesco Bettella from Padua, Italy .Protein Secondary Structure Prediction Using Optimized Scoring Functions: A Comparative Statistical Method.(page 7) 28th November 2009 the free University of Berlin .

[2]. A thesis submitted to the Graduate School-New Brunswick Rutgers , The State University of New Jersey in partial fulfillment of the requirements for the degree of Master of Science Graduate Program in Biomedical Engineering submitted by Jessica yee lau. protein structure database for structural genomics group**.**
January 2005.

[3].Thesis submitted for the degree of doctor of the school Polytechnique, specialty Informatics. Submitted by Thuong Van Du TRAN." Modeling and Predicting Super-secondary Structures of Transmembrane β-barrel Proteins", 7 december 2011.

[4]. A dissertation submitted in partial satisfaction of the requirements for the degree of doctor of philosophy in computer science by Melissa Suzanne Cline ."Protein sequence alignment reliability: prediction and measurement" (Page 10) June 2000.

[5]. Gregory A. Petsko,Dagmar Ringe book Protein Structure and Function.

[6]. Pekka Mäntsälä and Jarmo Niemi University of Turku, Department of Biochemistry, Finland ;enzymes: the biological catalysts of life.

[7]. A thesis submitted to the graduate school of natural and applied sciences of middle east technical university in partial fulfillment of the requirements for the degree of master of science in chemistry by Ahu Arslan , Immobilization of tyrosinase in polysiloxane/polypyrrole copolymer matrices, January 2006.

[8]. A dissertation submitted to University College London for the degree of Doctor of Philosophy, UCL Research Department of Structural and Molecular Biology, by Anja Baresic´ , Structural analysis of single amino acid polymorphisms" ,September, 2011.

References

[9]. Dissertation is submitted for the degree of Doctor of Philosophy Alex Gutteridge_ Darwin College, Cambridge,Understanding the Relationship Between Enzyme Structure and Catalysis, October 15, 2005.

[10]. Arrel Toews (Biochemistry & Biophysics) , Amino Acid and Nitrogen Metabolism I: Overview; Elimination of N-waste ,Date: Wednesday, September 21, 2005.

[11]. Solomon Adugna, Lakshmi Ahuja Mekonnen Alemu, Tsehayneh Kelemu, Henok Tekola, Belayhun Kibret, Solomon Genet Gondar, Medical biochemistry, University, Jimma University, Debub University, In collaboration with the Ethiopia Public Health Training Initiative, The Carter Center, the Ethiopia Ministry of Health, and the Ethiopia Ministry of Education 2004.

[12]. Christopher Larbie, PhD,BCHEM 254 – METABOLISM IN HEALTH AND DISEASES II , Lecture 4 Protein Catabolism ,2014.

[13]. Donna S. Dimski, American College of Veterinary Internal Medicine, Ammonia Metabolism and the Urea Cycle: Function and Clinical Implications, J Vet Intern Med ?994;8:73-78. Copyright 0 1994.

[14]. Protein data bank under web site: http://www.rcsb.org/pdb/home/home.do.

[15]. Rachedi A., Masters Teaching Material at Saida University, 2013, http://www.bioinformaticstools.org/masters/

[16]. Orengo, C.A., Michie, A.D., Jones, S., Jones, D.T., Swindells, M.B., and Thornton, J.M. (1997) CATH- A Hierarchic Classification of Protein Domain Structures. Structure. Vol 5. No 8. p.1093-1108. Department of Biochemistry and Molecular Biology, University College London.

[17]. Cand. Scient. Thesis ,Narve Sætre, Department of Informatics, University of Bergen Cand classification of protein structures. December 1999.

References

[18]. Rachedi A., Sequence, Structure and Function Server, 2011, http://www.bioinformaticstools.org/ssfs/

[19]. Golovin A1, Dimitropoulos D, Oldfield T, Rachedi A, Henrick K., MSDsite: a database search and retrieval system for the analysis and viewing of bound ligands and active sites. Proteins. 2005 Jan 1;58(1):190-9.

[20]. Roger Sayle v2.6 features added May, 1997: RasMol v2.6Quick Reference Card(c) Copyright1994. http://www.bio.cmu.edu/Courses/BiochemMols/RasFrames/TOC.HTM.

[21]. Rachedi A. Bioinformatics Tools Sever, 2012, http://www.bioinformaticstools.org.

[22]. Kaiser E, Colescott R-L, Bossinger C-D, P.I. Analytical Biochemistry, Volume 34, Issue 2, April 1970, pages 595-598.

Index I

1. Amino Acids Degradation / 1.1.Aspartate Aminotransferase :

Protein Chain	Prot SSE	Res Nbr	Res Name	RsAt.Name	LgAt.Name	Lg Chain	Lg Nbr	Bnd Len	Bond Type
A	108-124 H: 1	108	GLY	CA	O1P	A	1	3.58	van der Waals
A	108-124 H: 1	108	GLY	CA	O3P	A	1	3.88	van der Waals
A	108-124 H: 1	108	GLY	C	O1P	A	1	3.76	van der Waals
A	108-124 H: 1	108	GLY	C	O3P	A	1	3.78	van der Waals
A	108-124 H: 1	109	GLY	N	P	A	1	3.59	
A	108-124 H: 1	109	GLY	N	O1P	A	1	2.98	H.Bond
A	108-124 H: 1	109	GLY	N	O3P	A	1	3.29	H.Bond
A	108-124 H: 1	109	GLY	CA	O1P	A	1	3.9	van der Waals
A	108-124 H: 1	109	GLY	CA	O3P	A	1	3.87	van der Waals
A	108-124 H: 1	109	GLY	C	O3P	A	1	3.85	van der Waals
A	108-124 H: 1	110	THR	N	C5A	A	1	3.8	H.Bond
A	108-124 H: 1	110	THR	N	O3P	A	1	2.94	H.Bond
A	108-124 H: 1	110	THR	CA	O3P	A	1	3.76	van der Waals
A	108-124 H: 1	110	THR	CB	O3P	A	1	3.44	van der Waals
A	108-124 H: 1	110	THR	OG1	C5A	A	1	3.54	van der Waals
A	108-124 H: 1	110	THR	OG1	P	A	1	3.99	
A	108-124 H: 1	110	THR	OG1	O3P	A	1	2.8	H.Bond
A	No SSE	141	TRP	CD2	N1	A	1	3.98	H.Bond
A	No SSE	141	TRP	CD2	C2	A	1	3.78	van der Waals
A	No SSE	141	TRP	CD2	C3	A	1	3.98	van der Waals
A	No SSE	141	TRP	CE2	C3	A	1	3.88	van der Waals
A	No SSE	141	TRP	CE2	C4	A	1	3.83	van der Waals
A	No SSE	141	TRP	CE3	N1	A	1	3.61	H.Bond
A	No SSE	141	TRP	CE3	C2	A	1	3.89	van der Waals
A	No SSE	141	TRP	CE3	C6	A	1	3.85	van der Waals
A	No SSE	141	TRP	CZ2	C4	A	1	3.77	van der Waals
A	No SSE	141	TRP	CZ2	C5	A	1	3.63	van der Waals
A	No SSE	141	TRP	CZ2	C5A	A	1	3.86	van der Waals

A	No SSE	141	TRP	CZ3	N1	A	1	3.72	H.Bond
A	No SSE	141	TRP	CZ3	C5	A	1	3.9	van der Waals
A	No SSE	141	TRP	CZ3	C6	A	1	3.47	van der Waals
A	No SSE	141	TRP	CH2	C5	A	1	3.51	van der Waals
A	No SSE	141	TRP	CH2	C6	A	1	3.59	van der Waals
A	No SSE	141	TRP	CH2	C5A	A	1	3.6	van der Waals
A	No SSE	195	ASN	ND2	O3	A	1	3.08	H.Bond
A	219-224 S: 1	223	ASP	CG	N1	A	1	3.27	H.Bond
A	219-224 S: 1	223	ASP	OD1	N1	A	1	3.08	H.Bond
A	219-224 S: 1	223	ASP	OD1	C2	A	1	3.91	van der Waals
A	219-224 S: 1	223	ASP	OD1	C6	A	1	3.72	van der Waals
A	219-224 S: 1	223	ASP	OD2	N1	A	1	2.77	H.Bond
A	219-224 S: 1	223	ASP	OD2	C2	A	1	3.67	van der Waals
A	219-224 S: 1	223	ASP	OD2	C2A	A	1	3.67	van der Waals
A	219-224 S: 1	223	ASP	OD2	C6	A	1	3.65	van der Waals
A	No SSE	225	ALA	CB	N1	A	1	3.73	H.Bond
A	No SSE	225	ALA	CB	C2	A	1	3.66	van der Waals
A	No SSE	225	ALA	CB	C3	A	1	3.65	van der Waals
A	No SSE	225	ALA	CB	C4	A	1	3.8	van der Waals
A	No SSE	225	ALA	CB	C5	A	1	3.9	van der Waals
A	No SSE	225	ALA	CB	C6	A	1	3.84	van der Waals
A	No SSE	226	TYR	CE2	C3	A	1	3.87	van der Waals
A	No SSE	226	TYR	CE2	O3	A	1	3.15	van der Waals
A	No SSE	226	TYR	CZ	O3	A	1	3.19	van der Waals
A	No SSE	226	TYR	CZ	O4A	A	1	3.92	van der Waals
A	No SSE	226	TYR	OH	C3	A	1	3.64	van der Waals
A	No SSE	226	TYR	OH	O3	A	1	2.53	H.Bond
A	No SSE	226	TYR	OH	C4A	A	1	3.65	van der Waals
A	No SSE	226	TYR	OH	O4A	A	1	3.53	H.Bond
A	251-256 S: -1	256	SER	CB	O1P	A	1	3.46	van der Waals
A	251-256 S: -1	256	SER	OG	O4P	A	1	3.77	H.Bond
A	251-256 S: -1	256	SER	OG	P	A	1	3.85	
A	251-256 S: -1	256	SER	OG	O1P	A	1	2.63	H.Bond
A	No SSE	258	SER	CB	O1P	A	1	3.5	van der Waals
A	No SSE	258	SER	OG	P	A	1	3.69	
A	No SSE	258	SER	OG	O1P	A	1	2.77	H.Bond
A	No SSE	258	SER	OG	O2P	A	1	3.32	H.Bond

A	No SSE	259	LYS	CD	O4A	A	1	3.27	van der Waals
A	No SSE	259	LYS	CE	O4A	A	1	3.12	van der Waals
A	No SSE	259	LYS	NZ	C4A	A	1	3.01	H.Bond
A	No SSE	259	LYS	NZ	O4A	A	1	2.26	H.Bond
A	No SSE	267	ARG	CZ	O2P	A	1	3.81	van der Waals
A	No SSE	267	ARG	CZ	O3P	A	1	3.72	van der Waals
A	No SSE	267	ARG	NH1	P	A	1	3.8	
A	No SSE	267	ARG	NH1	O1P	A	1	3.71	H.Bond
A	No SSE	267	ARG	NH1	O2P	A	1	3.18	H.Bond
A	No SSE	267	ARG	NH1	O3P	A	1	3.81	H.Bond
A	No SSE	267	ARG	NH2	P	A	1	3.65	
A	No SSE	267	ARG	NH2	O2P	A	1	3.56	H.Bond
A	No SSE	267	ARG	NH2	O3P	A	1	2.74	H.Bond

Table n°22: The binding environment details of the PLP bound the enzyme ASAT (PDB id : 3II0) (Chain :A) as calculated by the **Lgb** system.

1.2.Glutamate dehydrogenase:

Protein Chain	Prot SSE	Res Nbr	Res Name	RsAt.Name	LgAt.Name	Lg Chain	Lg Nbr	Bnd Len	Bnd Type
A	89-90 S: 1	90	LYS	CG	OE2	A	550	3.26	van der Waals
A	89-90 S: 1	90	LYS	CD	OE2	A	550	3.89	van der Waals
A	89-90 S: 1	90	LYS	CE	OE2	A	550	3.4	van der Waals
A	89-90 S: 1	90	LYS	NZ	CD	A	550	3.58	H.Bond
A	89-90 S: 1	90	LYS	NZ	OE1	A	550	3.97	H.Bond
A	89-90 S: 1	90	LYS	NZ	OE2	A	550	2.65	H.Bond
A	No SSE	91	GLY	N	OE2	A	550	3.73	H.Bond
A	No SSE	91	GLY	CA	OE2	A	550	3.99	van der Waals
A	No SSE	92	GLY	N	N	A	550	3.75	H.Bond
A	No SSE	92	GLY	N	CG	A	550	3.46	H.Bond
A	No SSE	92	GLY	CA	N	A	550	3.65	H.Bond
A	No SSE	92	GLY	CA	CG	A	550	3.94	van der Waals
A	100-118 H: 1	111	MET	SD	C	A	550	3.35	van der Waals
A	100-118 H: 1	111	MET	SD	O	A	550	3.39	
A	100-118 H: 1	114	LYS	CE	O	A	550	3.68	van der Waals
A	100-118 H: 1	114	LYS	NZ	C	A	550	3.81	H.Bond
A	100-118 H: 1	114	LYS	NZ	O	A	550	2.77	H.Bond

A	123-129 S: -1	126	LYS	NZ	N	A	550	3.59	H.Bond
A	123-129 S: -1	126	LYS	NZ	C	A	550	3.55	H.Bond
A	No SSE	166	ALA	CB	CG	A	550	3.46	van der Waals
A	No SSE	166	ALA	CB	CD	A	550	3.21	van der Waals
A	No SSE	166	ALA	CB	OE1	A	550	3.51	van der Waals
A	No SSE	166	ALA	CB	OE2	A	550	3.45	van der Waals
A	No SSE	167	PRO	C	N	A	550	3.81	H.Bond
A	No SSE	167	PRO	O	N	A	550	2.65	H.Bond
A	No SSE	167	PRO	O	CA	A	550	3.79	van der Waals
A	No SSE	168	ASP	CB	N	A	550	3.83	H.Bond
A	No SSE	168	ASP	CG	N	A	550	3.96	H.Bond
A	No SSE	168	ASP	OD1	N	A	550	3.22	H.Bond
A	No SSE	211	ARG	NH2	CB	A	550	3.92	H.Bond
A	No SSE	211	ARG	NH2	OE1	A	550	3.13	H.Bond
A	No SSE	349	ASN	ND2	O	A	550	3.76	H.Bond
A	375-391 H: 1	377	GLY	C	OE1	A	550	3.96	van der Waals
A	375-391 H: 1	377	GLY	O	OE1	A	550	3.72	H.Bond
A	375-391 H: 1	378	VAL	N	OE1	A	550	3.92	H.Bond
A	375-391 H: 1	378	VAL	CA	OE1	A	550	3.59	van der Waals
A	375-391 H: 1	381	SER	CB	OE1	A	550	3.23	van der Waals
A	375-391 H: 1	381	SER	OG	CD	A	550	3.61	van der Waals
A	375-391 H: 1	381	SER	OG	OE1	A	550	2.61	H.Bond

Table n°23: The binding environment details of the GLU bound the enzyme GDH (PDB id:3ETD) , (chain A) as calculated by the **Lgb** system.

Protein Chain	Prot SSE	Res Nbr	Res Name	RsAt.Name	LgAt.Name	Lg Chain	Lg Nbr	Bnd Len	Bnd Type
A	No SSE	90	LYS	CG	OE1	A	502	3.18	van der Waals
A	No SSE	90	LYS	CD	OE1	A	502	3.71	van der Waals
A	No SSE	90	LYS	CE	CD	A	502	3.97	van der Waals
A	No SSE	90	LYS	CE	OE1	A	502	3.04	van der Waals
A	No SSE	90	LYS	NZ	CD	A	502	3.58	H.Bond
A	No SSE	90	LYS	NZ	OE1	A	502	2.42	H.Bond
A	No SSE	91	GLY	N	OE1	A	502	3.57	H.Bond
A	No SSE	91	GLY	CA	OE1	A	502	3.9	van der Waals
A	No SSE	92	GLY	N	CG	A	502	3.92	H.Bond
A	100-119 H: 1	111	MET	CG	OXT	A	502	3.64	van der Waals
A	100-119 H: 1	111	MET	SD	C	A	502	3.42	van der Waals
A	100-119 H: 1	111	MET	SD	O	A	502	3.39	

A	100-119 H: 1	111	MET	SD	OXT	A	502	2.95	
A	100-119 H: 1	111	MET	CE	CG	A	502	3.34	van der Waals
A	100-119 H: 1	111	MET	CE	CD	A	502	3.86	van der Waals
A	100-119 H: 1	111	MET	CE	OXT	A	502	3.92	van der Waals
A	100-119 H: 1	114	LYS	CD	O	A	502	3.63	van der Waals
A	100-119 H: 1	114	LYS	CE	O	A	502	3.87	van der Waals
A	100-119 H: 1	114	LYS	NZ	C	A	502	3.51	H.Bond
A	100-119 H: 1	114	LYS	NZ	O	A	502	2.89	H.Bond
A	100-119 H: 1	114	LYS	NZ	OXT	A	502	3.31	H.Bond
A	No SSE	126	LYS	CE	OXT	A	502	3.83	van der Waals
A	No SSE	126	LYS	NZ	C	A	502	3.89	H.Bond
A	No SSE	126	LYS	NZ	OXT	A	502	2.75	H.Bond
A	No SSE	166	ALA	CB	CD	A	502	3.24	van der Waals
A	No SSE	166	ALA	CB	OE1	A	502	2.93	van der Waals
A	No SSE	166	ALA	CB	OE2	A	502	3.61	van der Waals
A	No SSE	167	PRO	C	N	A	502	3.98	H.Bond
A	No SSE	167	PRO	O	N	A	502	2.79	H.Bond
A	No SSE	167	PRO	O	CA	A	502	3.88	van der Waals
A	No SSE	168	ASP	CG	N	A	502	3.73	H.Bond
A	No SSE	168	ASP	OD2	N	A	502	2.71	H.Bond
A	No SSE	168	ASP	OD2	CA	A	502	3.78	van der Waals
A	No SSE	168	ASP	OD2	OXT	A	502	3.81	H.Bond
A	210-227 H: 1	211	ARG	NH2	CB	A	502	3.64	H.Bond
A	210-227 H: 1	211	ARG	NH2	OE2	A	502	3.46	H.Bond
A	No SSE	349	ASN	ND2	OXT	A	502	3.72	H.Bond
A	375-391 H: 1	377	GLY	C	OE2	A	502	3.56	van der Waals
A	375-391 H: 1	377	GLY	O	OE2	A	502	3.62	H.Bond
A	375-391 H: 1	378	VAL	N	OE2	A	502	3.51	H.Bond
A	375-391 H: 1	378	VAL	CA	OE2	A	502	3.41	van der Waals
A	375-391 H: 1	378	VAL	CG1	CD	A	502	3.83	van der Waals
A	375-391 H: 1	378	VAL	CG1	OE2	A	502	3.88	van der Waals
A	375-391 H:	381	SER	CB	OE2	A	502	3.17	van der Waals

	1								
A	375-391 H: 1	381	SER	OG	CD	A	502	3.44	van der Waals
A	375-391 H: 1	381	SER	OG	OE1	A	502	3.91	H.Bond
A	375-391 H: 1	381	SER	OG	OE2	A	502	2.32	H.Bond

Table n°24 : The binding environment details of the GLU bound the enzyme GDH (PDB id: 3MVO) , (chain A) as calculated by the **Lgb** system.

Protein Chain	Prot SSE	Res Nbr	Res Name	RsAt.Name	LgAt.Name	Lg Chain	Lg Nbr	Bnd Len	Bnd Type
A	89-96 S: 1	90	LYS	CG	OE1	A	502	3.02	van der Waals
A	89-96 S: 1	90	LYS	CD	OE1	A	502	3.8	van der Waals
A	89-96 S: 1	90	LYS	CE	OE1	A	502	3.35	van der Waals
A	89-96 S: 1	90	LYS	NZ	CD	A	502	3.74	H.Bond
A	89-96 S: 1	90	LYS	NZ	OE1	A	502	2.73	H.Bond
A	89-96 S: 1	91	GLY	N	OE1	A	502	3.56	H.Bond
A	89-96 S: 1	92	GLY	N	N	A	502	3.64	H.Bond
A	89-96 S: 1	92	GLY	N	CG	A	502	3.71	H.Bond
A	89-96 S: 1	92	GLY	CA	N	A	502	3.8	H.Bond
A	100-118 H: 1	111	MET	SD	C	A	502	3.31	van der Waals
A	100-118 H: 1	111	MET	SD	O	A	502	3.53	
A	100-118 H: 1	111	MET	SD	CG	A	502	3.88	van der Waals
A	100-118 H: 1	111	MET	SD	OXT	A	502	2.95	
A	100-118 H: 1	111	MET	CE	O	A	502	3.9	van der Waals
A	100-118 H: 1	114	LYS	CD	O	A	502	3.62	van der Waals
A	100-118 H: 1	114	LYS	CD	OXT	A	502	3.93	van der Waals
A	100-118 H: 1	114	LYS	CE	OXT	A	502	3.58	van der Waals
A	100-118 H: 1	114	LYS	NZ	C	A	502	3.26	H.Bond
A	100-118 H: 1	114	LYS	NZ	O	A	502	3.34	H.Bond
A	100-118 H: 1	114	LYS	NZ	OXT	A	502	2.42	H.Bond
A	123-130 S: -1	126	LYS	NZ	N	A	502	2.65	H.Bond
A	123-130 S: -1	126	LYS	NZ	CA	A	502	3.99	H.Bond
A	123-130 S: -1	126	LYS	NZ	OXT	A	502	3.84	H.Bond
A	163-166 S: 1	166	ALA	CB	CD	A	502	3.44	van der Waals
A	163-166 S: 1	166	ALA	CB	OE1	A	502	3.51	van der Waals
A	163-166 S: 1	166	ALA	CB	OE2	A	502	3.34	van der Waals
A	No SSE	167	PRO	O	N	A	502	3.57	H.Bond

A	No SSE	167	PRO	O	CB	A	502	3.75	van der Waals
A	No SSE	168	ASP	CG	N	A	502	3.91	H.Bond
A	No SSE	168	ASP	OD2	N	A	502	2.87	H.Bond
A	No SSE	211	ARG	CZ	OE2	A	502	3.97	van der Waals
A	No SSE	211	ARG	NH2	CB	A	502	3.72	H.Bond
A	No SSE	211	ARG	NH2	OE2	A	502	2.94	H.Bond
A	No SSE	349	ASN	ND2	OXT	A	502	3.36	H.Bond
A	375-391 H: 1	378	VAL	N	O	A	502	3.95	H.Bond
A	375-391 H: 1	378	VAL	N	OE2	A	502	3.82	H.Bond
A	375-391 H: 1	378	VAL	CA	OE2	A	502	3.86	van der Waals
A	375-391 H: 1	378	VAL	CG1	O	A	502	3.86	van der Waals
A	375-391 H: 1	378	VAL	CG1	CG	A	502	3.79	van der Waals
A	375-391 H: 1	378	VAL	CG1	CD	A	502	3.58	van der Waals
A	375-391 H: 1	378	VAL	CG1	OE1	A	502	3.56	van der Waals
A	375-391 H: 1	381	SER	CB	OE2	A	502	3.31	van der Waals
A	375-391 H: 1	381	SER	OG	CD	A	502	3.45	van der Waals
A	375-391 H: 1	381	SER	OG	OE1	A	502	3.68	H.Bond
A	375-391 H: 1	381	SER	OG	OE2	A	502	2.47	H.Bond

Table n°25: The binding environment details of the GLU bound the enzyme GDH (PDB id: 3MVQ) , (chain A) as calculated by the Lgb system.

2.Urea cycle:

2.1.Carbamoyl phosphate synthetase I:

Protein Chain	Prot SSE	Res Nbr	Res Name	RsAt.Name	LgAt.Name	Lg Chain	Lg Nbr	Bnd Len	Bnd Type
B	No SSE	18	VAL	CG2	C4	B	1096	3.97	van der Waals
B	No SSE	21	GLN	CD	C4	B	1096	3.95	van der Waals
B	No SSE	21	GLN	OE1	C3	B	1096	3.89	van der Waals
B	No SSE	21	GLN	OE1	C4	B	1096	3.7	van der Waals
B	No SSE	21	GLN	NE2	C3	B	1096	3.88	H.Bond
B	91-103 H: 1	92	GLN	C	C2	B	1096	3.73	van der Waals
B	91-103 H: 1	92	GLN	CB	C2	B	1096	3.99	van der Waals
B	91-103 H: 1	93	THR	N	C2	B	1096	3.4	H.Bond
B	91-103 H: 1	93	THR	CA	C2	B	1096	3.5	van der Waals
B	91-103 H: 1	93	THR	OG1	C2	B	1096	3.46	van der Waals
B	91-103 H: 1	93	THR	OG1	C3	B	1096	3.59	van der Waals
B	91-103 H: 1	93	THR	OG1	C7	B	1096	3.43	van der Waals

B	91-103 H: 1	96	ASN	ND2	C2	B	1096	3.84	H.Bond
B	No SSE	935	ASN	CG	C8	B	1096	3.7	van der Waals
B	No SSE	935	ASN	OD1	C8	B	1096	3.72	van der Waals
B	No SSE	935	ASN	ND2	C8	B	1096	3.67	H.Bond

Table n°26: The binding environment details of the NET bound the enzyme CPS (PDB id:1JDB) , (chainB) as calculated by the **Lgb** system.

Protein Chain	Prot SSE	Res Nbr	Res Name	RsAt.Name	LgAt.Name	Lg Chain	Lg Nbr	Bnd Len	Bnd Type
G	No SSE	6019	VAL	CG2	C4	G	7950	3.9	van der Waals
G	No SSE	6022	GLN	CD	C3	G	7950	3.88	van der Waals
G	No SSE	6022	GLN	CD	C4	G	7950	3.84	van der Waals
G	No SSE	6022	GLN	OE1	C3	G	7950	3.78	van der Waals
G	No SSE	6022	GLN	OE1	C4	G	7950	3.63	van der Waals
G	No SSE	6022	GLN	NE2	C3	G	7950	3.71	H.Bond
G	6092-6104 H: 1	6093	GLN	C	C2	G	7950	3.82	van der Waals
G	6092-6104 H: 1	6093	GLN	CB	C2	G	7950	3.68	van der Waals
G	6092-6104 H: 1	6093	GLN	CG	C2	G	7950	3.8	van der Waals
G	6092-6104 H: 1	6094	THR	N	C2	G	7950	3.59	H.Bond
G	6092-6104 H: 1	6094	THR	CA	C2	G	7950	3.78	van der Waals
G	6092-6104 H: 1	6094	THR	OG1	N1	G	7950	3.83	H.Bond
G	6092-6104 H: 1	6094	THR	OG1	C2	G	7950	3.54	van der Waals
G	6092-6104 H: 1	6094	THR	OG1	C3	G	7950	3.39	van der Waals
G	6092-6104 H: 1	6094	THR	OG1	C7	G	7950	3.28	van der Waals
G	6092-6104 H: 1	6097	ASN	ND2	C2	G	7950	3.46	H.Bond
G	No SSE	6936	ASN	CG	C8	G	7950	3.94	van der Waals
G	No SSE	6936	ASN	ND2	C8	G	7950	3.81	H.Bond

Table n°27: The binding environment details of the NET bound the enzyme CPS (PDB id:1T36) , (chain G) as calculated by the **Lgb** system.

2.2.Ornithine Transcarbomylase :

Protein Chain	Prot SSE	Res Nbr	Res Name	RsAt.Name	LgAt.Name	Lg Chain	Lg Nbr	Bnd Len	Bnd Type
A	No SSE	90	SER	CA	O1P	A	355	3.85	van der Waals
A	No SSE	90	SER	CA	O2P	A	355	3.83	van der Waals
A	No SSE	90	SER	C	O3P	A	355	3.89	van der Waals
A	No SSE	90	SER	CB	O2P	A	355	3.58	van der Waals
A	No SSE	90	SER	OG	P	A	355	3.93	
A	No SSE	90	SER	OG	O2P	A	355	2.73	H.Bond
A	No SSE	91	THR	N	P	A	355	3.81	
A	No SSE	91	THR	N	O2P	A	355	3.77	H.Bond
A	No SSE	91	THR	N	O3P	A	355	2.86	H.Bond
A	No SSE	91	THR	CA	O3P	A	355	3.54	van der Waals
A	No SSE	91	THR	C	O3P	A	355	3.62	van der Waals
A	No SSE	91	THR	CB	O3P	A	355	3.64	van der Waals
A	92-103 H: 1	92	ARG	N	P	A	355	3.78	
A	92-103 H: 1	92	ARG	N	O2P	A	355	3.56	H.Bond
A	92-103 H: 1	92	ARG	N	O3P	A	355	2.86	H.Bond
A	92-103 H: 1	92	ARG	CA	O2P	A	355	3.9	van der Waals
A	92-103 H: 1	92	ARG	CA	O3P	A	355	3.82	van der Waals
A	92-103 H: 1	92	ARG	C	O2P	A	355	3.75	van der Waals
A	92-103 H: 1	92	ARG	CB	C1P	A	355	3.91	van der Waals
A	92-103 H: 1	92	ARG	CB	O2P	A	355	3.97	van der Waals
A	92-103 H: 1	92	ARG	CB	O3P	A	355	3.74	van der Waals
A	92-103 H: 1	92	ARG	NE	C1P	A	355	3.4	H.Bond
A	92-103 H: 1	92	ARG	NE	P	A	355	3.77	
A	92-103 H: 1	92	ARG	NE	O3P	A	355	2.89	H.Bond
A	92-103 H: 1	92	ARG	CZ	O3P	A	355	3.3	van der Waals
A	92-103 H: 1	92	ARG	NH2	O3P	A	355	2.86	H.Bond
A	92-103 H: 1	93	THR	N	O2P	A	355	2.77	H.Bond
A	92-103 H: 1	93	THR	CA	O2P	A	355	3.55	van der Waals
A	92-103 H: 1	93	THR	CB	O2P	A	355	3.24	van der Waals
A	92-103 H: 1	93	THR	OG1	O1	A	355	3.33	H.Bond
A	92-103 H: 1	93	THR	OG1	O2P	A	355	2.78	H.Bond
A	137-141 S: 1	141	ARG	CZ	O1P	A	355	3.73	van der Waals
A	137-141 S: 1	141	ARG	CZ	O2P	A	355	3.71	van der Waals
A	137-141 S: 1	141	ARG	NH1	P	A	355	3.78	

A	137-141 S: 1	141	ARG	NH1	O1P	A	355	2.79	H.Bond
A	137-141 S: 1	141	ARG	NH1	O2P	A	355	3.52	H.Bond
A	137-141 S: 1	141	ARG	NH2	C1	A	355	3.68	H.Bond
A	137-141 S: 1	141	ARG	NH2	O1	A	355	2.92	H.Bond
A	137-141 S: 1	141	ARG	NH2	O1P	A	355	3.82	H.Bond
A	137-141 S: 1	141	ARG	NH2	O2P	A	355	3.41	H.Bond
A	No SSE	163	LEU	CD1	CB	A	355	3.8	van der Waals
A	No SSE	163	LEU	CD1	CG	A	355	3.8	van der Waals
A	No SSE	163	LEU	CD1	CD	A	355	3.66	van der Waals
A	No SSE	163	LEU	CD1	OXT	A	355	3.72	van der Waals
A	No SSE	168	HIS	CD2	CD	A	355	3.7	van der Waals
A	No SSE	168	HIS	CD2	O1	A	355	3.69	van der Waals
A	No SSE	168	HIS	CE1	O1	A	355	3.74	van der Waals
A	No SSE	168	HIS	NE2	CD	A	355	3.8	H.Bond
A	No SSE	168	HIS	NE2	C1	A	355	3.85	H.Bond
A	No SSE	168	HIS	NE2	O1	A	355	2.81	H.Bond
A	169-183 H: 1	171	GLN	NE2	O1	A	355	3.82	H.Bond
A	No SSE	198	ASN	ND2	N	A	355	3.31	H.Bond
A	199-205 H: 1	199	ASN	CG	N	A	355	3.82	H.Bond
A	199-205 H: 1	199	ASN	CG	OXT	A	355	3.76	van der Waals
A	199-205 H: 1	199	ASN	OD1	N	A	355	2.84	H.Bond
A	199-205 H: 1	199	ASN	OD1	CA	A	355	3.82	van der Waals
A	199-205 H: 1	199	ASN	OD1	CB	A	355	3.9	van der Waals
A	199-205 H: 1	199	ASN	OD1	OXT	A	355	3.61	H.Bond
A	199-205 H: 1	199	ASN	ND2	OXT	A	355	3.02	H.Bond
A	199-205 H: 1	200	ILE	CD1	CB	A	355	3.88	van der Waals
A	No SSE	263	ASP	CG	N	A	355	3.65	H.Bond
A	No SSE	263	ASP	CG	CA	A	355	3.96	van der Waals
A	No SSE	263	ASP	OD1	N	A	355	3.87	H.Bond
A	No SSE	263	ASP	OD1	CA	A	355	3.73	van der Waals
A	No SSE	263	ASP	OD1	CB	A	355	3.96	van der Waals
A	No SSE	263	ASP	OD2	N	A	355	2.7	H.Bond
A	No SSE	263	ASP	OD2	CA	A	355	3.35	van der Waals

A	No SSE	263	ASP	OD2	CB	A	355	3.67	van der Waals
A	No SSE	267	SER	CA	C	A	355	3.92	van der Waals
A	No SSE	267	SER	CA	O	A	355	3.49	van der Waals
A	No SSE	267	SER	C	O	A	355	3.71	van der Waals
A	No SSE	267	SER	CB	N	A	355	3.87	H.Bond
A	No SSE	267	SER	CB	C	A	355	3.54	van der Waals
A	No SSE	267	SER	CB	O	A	355	3.58	van der Waals
A	No SSE	267	SER	CB	OXT	A	355	3.61	van der Waals
A	No SSE	267	SER	OG	N	A	355	2.91	H.Bond
A	No SSE	267	SER	OG	CA	A	355	3.61	van der Waals
A	No SSE	267	SER	OG	C	A	355	3.48	van der Waals
A	No SSE	267	SER	OG	O	A	355	3.85	H.Bond
A	No SSE	267	SER	OG	OXT	A	355	3.64	H.Bond
A	No SSE	268	MET	N	C	A	355	3.86	H.Bond
A	No SSE	268	MET	N	O	A	355	2.99	H.Bond
A	No SSE	268	MET	CB	O	A	355	3.87	van der Waals
A	No SSE	268	MET	CG	O	A	355	3.78	van der Waals
A	No SSE	268	MET	CE	O	A	355	3.49	van der Waals
A	No SSE	303	CYS	O	C1P	A	355	3.75	van der Waals
A	No SSE	303	CYS	SG	CB	A	355	3.94	van der Waals
A	No SSE	303	CYS	SG	CG	A	355	3.88	van der Waals
A	No SSE	303	CYS	SG	CD	A	355	3.73	van der Waals
A	No SSE	304	LEU	O	CG	A	355	3.99	van der Waals
A	No SSE	304	LEU	O	CD	A	355	3.82	van der Waals
A	No SSE	304	LEU	O	NE	A	355	2.84	H.Bond
A	No SSE	304	LEU	O	C1	A	355	3.54	van der Waals
A	No SSE	304	LEU	O	C1P	A	355	3.29	van der Waals
A	323-342 H: 1	330	ARG	CZ	O1	A	355	3.99	van der Waals
A	323-342 H: 1	330	ARG	NH1	C1	A	355	3.57	H.Bond
A	323-342 H: 1	330	ARG	NH1	O1	A	355	2.95	H.Bond
A	323-342 H: 1	330	ARG	NH1	C1P	A	355	3.79	H.Bond
A	Water	381	HOH	O	OXT	A	355	2.8	H.Bond
A	Water	385	HOH	O	O	A	355	2.81	H.Bond

Table n°28: The binding environment details of the PAO bound the enzyme OTC (PDB id: 1OTH) , (chainA) as calculated by the **Lgb** system.

2.3.Argininosuccinic Acid Synthase:

Protein Chain	Prot SSE	Res Nbr	Res Name	RsAt.Name	LgAt.Name	Lg Chain	Lg Nbr	Bnd Len	Bnd Type
A	No SSE	118	ALA	CA	OXT	A	501	3.53	van der Waals
A	No SSE	118	ALA	C	OXT	A	501	3.68	van der Waals
A	No SSE	118	ALA	CB	OXT	A	501	3.74	van der Waals
A	No SSE	119	THR	N	C	A	501	3.8	H.Bond
A	No SSE	119	THR	N	O	A	501	3.83	H.Bond
A	No SSE	119	THR	N	OXT	A	501	2.9	H.Bond
A	No SSE	119	THR	CA	OXT	A	501	3.87	van der Waals
A	No SSE	119	THR	CB	O	A	501	3.41	van der Waals
A	No SSE	119	THR	CB	OXT	A	501	3.93	van der Waals
A	No SSE	119	THR	OG1	C	A	501	3.48	van der Waals
A	No SSE	119	THR	OG1	O	A	501	2.5	H.Bond
A	No SSE	119	THR	OG1	OXT	A	501	3.65	H.Bond
A	No SSE	119	THR	CG2	C	A	501	3.87	van der Waals
A	No SSE	119	THR	CG2	O	A	501	3.42	van der Waals
A	No SSE	119	THR	CG2	OXT	A	501	3.72	van der Waals
A	No SSE	122	GLY	N	O	A	501	3.99	H.Bond
A	No SSE	122	GLY	CA	O	A	501	3.77	van der Waals
A	No SSE	122	GLY	CA	CG	A	501	3.6	van der Waals
A	No SSE	122	GLY	CA	OD1	A	501	3.75	van der Waals
A	No SSE	122	GLY	CA	OD2	A	501	3.43	van der Waals
A	No SSE	122	GLY	C	OD1	A	501	3.88	van der Waals
A	No SSE	122	GLY	C	OD2	A	501	3.66	van der Waals
A	123-136 H: 1	123	ASN	N	CG	A	501	3.33	H.Bond
A	123-136 H: 1	123	ASN	N	OD1	A	501	2.98	H.Bond
A	123-136 H: 1	123	ASN	N	OD2	A	501	3.09	H.Bond
A	123-136 H: 1	123	ASN	CA	OD1	A	501	3.78	van der Waals
A	123-136 H: 1	123	ASN	CA	OD2	A	501	3.93	van der Waals
A	123-136 H: 1	123	ASN	C	OD2	A	501	3.89	van der Waals
A	123-136 H: 1	123	ASN	CB	OD1	A	501	3.51	van der Waals
A	123-136 H: 1	123	ASN	CG	OD1	A	501	3.95	van der Waals
A	123-136 H: 1	123	ASN	ND2	OD1	A	501	3.79	H.Bond
A	123-136 H: 1	124	ASP	N	CG	A	501	3.85	H.Bond

A	123-136 H: 1	124	ASP	N	OD1	A	501	3.98	H.Bond
A	123-136 H: 1	124	ASP	N	OD2	A	501	2.94	H.Bond
A	123-136 H: 1	124	ASP	CA	OD2	A	501	3.74	van der Waals
A	123-136 H: 1	124	ASP	CB	OD2	A	501	3.4	van der Waals
A	123-136 H: 1	124	ASP	CG	N	A	501	3.83	H.Bond
A	123-136 H: 1	124	ASP	CG	OD2	A	501	3.43	van der Waals
A	123-136 H: 1	124	ASP	OD2	N	A	501	2.86	H.Bond
A	123-136 H: 1	124	ASP	OD2	CG	A	501	3.85	van der Waals
A	123-136 H: 1	124	ASP	OD2	OD2	A	501	3.33	H.Bond
A	No SSE	191	GLU	OE2	CB	A	501	3.52	van der Waals
A	Water	569	HOH	O	OD1	A	501	2.63	H.Bond
A	Water	664	HOH	O	N	A	501	2.72	H.Bond

Table n°29: The binding environment details of the ASP bound the enzyme ASS (PDB id: 2NZ2) (chain A) as calculated by the **Lgb** system.

Protein Chain	Prot SSE	Res Nbr	Res Name	RsAt.Name	LgAt.Name	Lg Chain	Lg Nbr	Bnd Len	Bnd Type
A	No SSE	87	TYR	CE2	O2	A	502	3.54	van der Waals
A	No SSE	87	TYR	CZ	C1	A	502	3.98	van der Waals
A	No SSE	87	TYR	CZ	O2	A	502	3.46	van der Waals
A	No SSE	87	TYR	CZ	C2	A	502	3.95	van der Waals
A	No SSE	87	TYR	OH	C1	A	502	3.45	van der Waals
A	No SSE	87	TYR	OH	O2	A	502	2.69	H.Bond
A	No SSE	87	TYR	OH	C2	A	502	3.54	van der Waals
A	No SSE	91	THR	CB	O2	A	502	3.67	van der Waals
A	No SSE	91	THR	CG2	O2	A	502	3.78	van der Waals
A	No SSE	91	THR	CG2	C5	A	502	3.85	van der Waals
A	No SSE	92	SER	N	O2	A	502	3.86	H.Bond
A	123-136 H: 1	123	ASN	CB	O1	A	502	3.69	van der Waals
A	123-136 H: 1	123	ASN	CG	O1	A	502	3.69	van der Waals
A	123-136 H: 1	123	ASN	ND2	C1	A	502	3.98	H.Bond
A	123-136 H: 1	123	ASN	ND2	O1	A	502	2.77	H.Bond

A	123-136 H: 1	124	ASP	OD2	C5	A	502	3.66	van der Waals
A	123-136 H: 1	127	ARG	CZ	C1	A	502	3.93	van der Waals
A	123-136 H: 1	127	ARG	CZ	O1	A	502	3.58	van der Waals
A	123-136 H: 1	127	ARG	CZ	O2	A	502	3.4	van der Waals
A	123-136 H: 1	127	ARG	NH1	C1	A	502	3.31	H.Bond
A	123-136 H: 1	127	ARG	NH1	O1	A	502	3.17	H.Bond
A	123-136 H: 1	127	ARG	NH1	O2	A	502	2.73	H.Bond
A	123-136 H: 1	127	ARG	NH2	C1	A	502	3.6	H.Bond
A	123-136 H: 1	127	ARG	NH2	O1	A	502	3.2	H.Bond
A	123-136 H: 1	127	ARG	NH2	O2	A	502	3.22	H.Bond
A	No SSE	180	SER	O	N8	A	502	3.66	H.Bond
A	181-183 S: 0	181	MET	C	N8	A	502	3.69	H.Bond
A	181-183 S: 0	182	ASP	N	N8	A	502	3.47	H.Bond
A	181-183 S: 0	182	ASP	CB	N6	A	502	3.41	H.Bond
A	181-183 S: 0	182	ASP	CB	C7	A	502	3.55	van der Waals
A	181-183 S: 0	182	ASP	CB	N8	A	502	3.75	H.Bond
A	188-190 S: -1	189	SER	O	N8	A	502	3.57	H.Bond
A	188-190 S: -1	189	SER	CB	N6	A	502	3.89	H.Bond
A	188-190 S: -1	189	SER	OG	C3	A	502	3.62	van der Waals
A	188-190 S: -1	189	SER	OG	N6	A	502	3.24	H.Bond
A	188-190 S: -1	189	SER	OG	C7	A	502	3.83	van der Waals
A	188-190 S: -1	189	SER	OG	N8	A	502	3.48	H.Bond
A	No SSE	191	GLU	CG	N8	A	502	3.81	H.Bond
A	No SSE	191	GLU	CD	C4	A	502	3.93	van der Waals
A	No SSE	191	GLU	OE1	C2	A	502	3.99	van der Waals
A	No SSE	191	GLU	OE1	N2	A	502	3.23	H.Bond
A	No SSE	191	GLU	OE1	C3	A	502	3.57	van der Waals
A	No SSE	191	GLU	OE1	C4	A	502	3.65	van der Waals
A	No SSE	191	GLU	OE2	C4	A	502	3.96	van der Waals
A	265-271 S: -1	270	GLU	CD	N2	A	502	3.16	H.Bond
A	265-271 S: -1	270	GLU	OE1	N2	A	502	3.22	H.Bond
A	265-271 S: -1	270	GLU	OE2	C1	A	502	3.77	van der Waals
A	265-271 S: -1	270	GLU	OE2	O1	A	502	3.56	H.Bond
A	265-271 S: -1	270	GLU	OE2	C2	A	502	3.48	van der Waals
A	265-271 S: -	270	GLU	OE2	N2	A	502	2.54	H.Bond

	1								
A	277-283 S: -1	282	TYR	CE1	C2	A	502	3.91	van der Waals
A	277-283 S: -1	282	TYR	CE1	C3	A	502	3.93	van der Waals
A	277-283 S: -1	282	TYR	CZ	C2	A	502	3.88	van der Waals
A	277-283 S: -1	282	TYR	OH	C2	A	502	3	van der Waals
A	277-283 S: -1	282	TYR	OH	N2	A	502	2.81	H.Bond
A	277-283 S: -1	282	TYR	OH	C3	A	502	3.62	van der Waals
A	303-324 H: 1	322	TYR	CE1	O1	A	502	3.74	van der Waals
A	Water	569	HOH	O	N2	A	502	2.95	H.Bond
A	Water	647	HOH	O	O7	A	502	2.89	H.Bond

Table n°30: The binding environment details of the CIR bound the enzyme ASS (PDB id: 2NZ2) , (chain A) as calculated by the **Lgb** system.

2.4. Arginiosuccinate lyase:

Protein Chain	Prot SSE	Res Nbr	Res Name	RsAt.Name	LgAt.Name	Lg Chain	Lg Nbr	Bnd Len	Bnd Type
C	29-34 H: 1	29	SER	OG	N4	C	1001	3.33	H.Bond
C	89-103 H: 1	91	HIS	CE1	C3	C	1001	3.96	van der Waals
C	114-152 H: 1	115	ARG	O	C2	C	1001	3.86	van der Waals
C	114-152 H: 1	115	ARG	CB	C2	C	1001	3.98	van der Waals
C	114-152 H: 1	115	ARG	CB	C1	C	1001	3.96	van der Waals
C	114-152 H: 1	115	ARG	CZ	N3	C	1001	3.97	H.Bond
C	114-152 H: 1	115	ARG	NH2	N3	C	1001	3.16	H.Bond
C	114-152 H: 1	116	ASN	CG	N2	C	1001	3.91	H.Bond
C	114-152 H: 1	116	ASN	CG	OD2	C	1001	3.51	van der Waals
C	114-152 H: 1	116	ASN	OD1	C1	C	1001	3.83	van der Waals
C	114-152 H: 1	116	ASN	OD1	N2	C	1001	2.68	H.Bond
C	114-152 H: 1	116	ASN	OD1	C	C	1001	3.23	van der Waals
C	114-152 H: 1	116	ASN	OD1	N1	C	1001	3.18	H.Bond
C	114-152 H: 1	116	ASN	OD1	OD2	C	1001	3.73	H.Bond
C	114-152 H: 1	116	ASN	ND2	CD	C	1001	3.13	H.Bond
C	114-152 H: 1	116	ASN	ND2	OD1	C	1001	3.85	H.Bond
C	114-152 H: 1	116	ASN	ND2	OD2	C	1001	2.49	H.Bond

C	114-152 H: 1	119	VAL	CB	C3	C	1001	3.93	van der Waals
C	114-152 H: 1	119	VAL	CG2	O51	C	1001	3.81	van der Waals
C	114-152 H: 1	119	VAL	CG2	C3	C	1001	3.96	van der Waals
C	No SSE	205	ALA	CB	OD1	C	1001	3.97	van der Waals
C	No SSE	323	TYR	CB	OD2	C	1001	3.64	van der Waals
C	No SSE	323	TYR	CG	OD2	C	1001	3.89	van der Waals
C	No SSE	323	TYR	CD2	N1	C	1001	3.83	H.Bond
C	No SSE	323	TYR	CD2	OD2	C	1001	3.67	van der Waals
C	No SSE	323	TYR	CE1	O52	C	1001	3.45	van der Waals
C	No SSE	323	TYR	CE1	N2	C	1001	3.93	H.Bond
C	No SSE	323	TYR	CE2	N2	C	1001	3.43	H.Bond
C	No SSE	323	TYR	CE2	C	C	1001	3.52	van der Waals
C	No SSE	323	TYR	CE2	N1	C	1001	3.72	H.Bond
C	No SSE	323	TYR	CZ	O52	C	1001	3.49	van der Waals
C	No SSE	323	TYR	CZ	C2	C	1001	3.98	van der Waals
C	No SSE	323	TYR	CZ	C1	C	1001	3.63	van der Waals
C	No SSE	323	TYR	CZ	N2	C	1001	3.33	H.Bond
C	No SSE	323	TYR	CZ	C	C	1001	3.93	van der Waals
C	No SSE	323	TYR	OH	O52	C	1001	2.67	H.Bond
C	No SSE	323	TYR	OH	C5	C	1001	3.39	van der Waals
C	No SSE	323	TYR	OH	C4	C	1001	3.53	van der Waals
C	No SSE	323	TYR	OH	C2	C	1001	3.53	van der Waals
C	No SSE	323	TYR	OH	C1	C	1001	3.27	van der Waals
C	No SSE	323	TYR	OH	N2	C	1001	3.49	H.Bond
C	324-328 H: 5	328	GLN	N	O52	C	1001	3.88	H.Bond
C	324-328 H: 5	328	GLN	CA	O51	C	1001	3.57	van der Waals
C	324-328 H: 5	328	GLN	CA	O52	C	1001	3.78	van der Waals
C	324-328 H: 5	328	GLN	CA	C5	C	1001	3.93	van der Waals
C	324-328 H: 5	328	GLN	CB	O51	C	1001	3.55	van der Waals
C	324-328 H: 5	328	GLN	CB	O52	C	1001	3.84	van der Waals
C	324-328 H: 5	328	GLN	CB	C5	C	1001	3.74	van der Waals
C	324-328 H: 5	328	GLN	CG	O51	C	1001	3.52	van der Waals
C	324-328 H: 5	328	GLN	CD	N4	C	1001	3.69	H.Bond
C	324-328 H: 5	328	GLN	OE1	O51	C	1001	3.7	H.Bond
C	324-328 H: 5	328	GLN	OE1	C5	C	1001	3.7	van der Waals
C	324-328 H: 5	328	GLN	OE1	C4	C	1001	3.39	van der Waals
C	324-328 H: 5	328	GLN	OE1	N4	C	1001	2.68	H.Bond
C	329-355 H: 1	331	LYS	CE	O51	C	1001	3.1	van der Waals
C	329-355 H: 1	331	LYS	NZ	O51	C	1001	2.64	H.Bond
C	329-355 H: 1	331	LYS	NZ	C5	C	1001	3.67	H.Bond
C	Water	1023	HOH	O	O52	C	1001	2.61	H.Bond

C	Water	1086	HOH	O	N4	C	1001	2.51	H.Bond
C	Water	1123	HOH	O	N3	C	1001	2.66	H.Bond

Table n°31: The binding environment details of the AS1 bound the enzyme ASL (PDB id: 1K7W) ,(chain C) as calculated by the **Lgb** system.

2.5. Arginase:

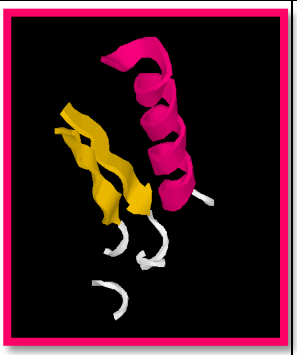
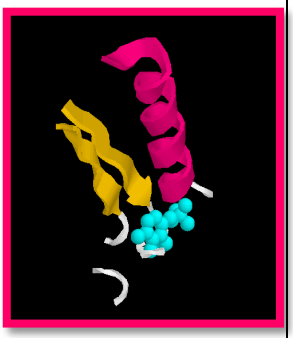
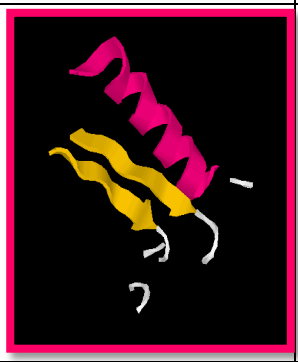
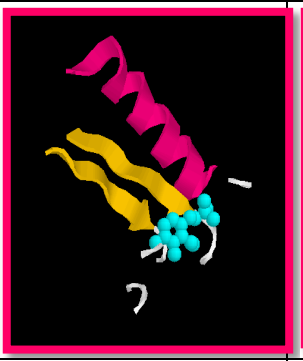
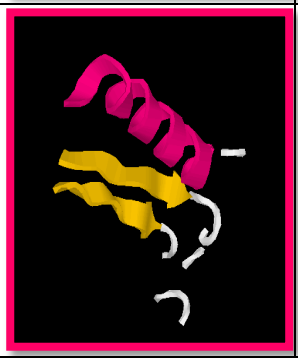
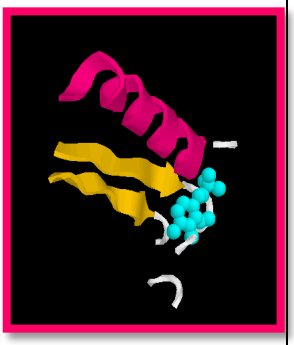
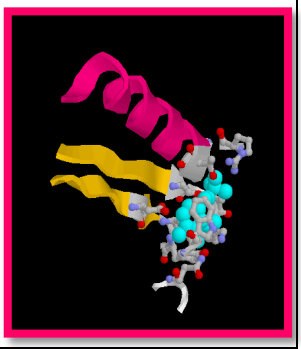
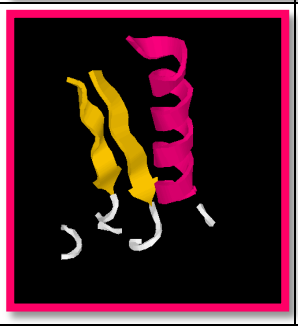
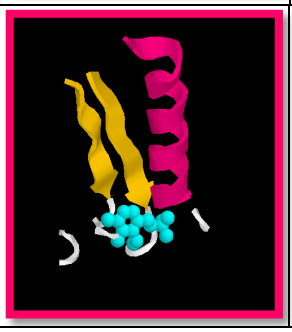
Protein Chain	Prot SSE	Res Nbr	Res Name	RsAt.Name	LgAt.Name	Lg Chain	Lg Nbr	Bnd Len	Bnd Type
A	No SSE	124	ASP	CG	OH1	A	901	3.77	van der Waals
A	No SSE	124	ASP	OD1	OH1	A	901	3.47	H.Bond
A	No SSE	124	ASP	OD2	OH1	A	901	3.27	H.Bond
A	No SSE	126	HIS	CB	NH1	A	901	3.45	H.Bond
A	No SSE	126	HIS	CB	OH1	A	901	3.8	van der Waals
A	No SSE	126	HIS	CG	CE	A	901	3.8	van der Waals
A	No SSE	126	HIS	CG	NH1	A	901	3.35	H.Bond
A	No SSE	126	HIS	CG	OH1	A	901	3.81	van der Waals
A	No SSE	126	HIS	ND1	CE	A	901	3.43	H.Bond
A	No SSE	126	HIS	ND1	NH1	A	901	3.14	H.Bond
A	No SSE	126	HIS	ND1	NH2	A	901	3.81	H.Bond
A	No SSE	126	HIS	ND1	OH1	A	901	3.29	H.Bond
A	No SSE	126	HIS	CE1	CE	A	901	3.78	van der Waals
A	No SSE	126	HIS	CE1	NH1	A	901	3.95	H.Bond
A	No SSE	126	HIS	CE1	NH2	A	901	3.92	H.Bond
A	No SSE	128	ASP	CG	NH1	A	901	3.44	H.Bond
A	No SSE	128	ASP	CG	OH1	A	901	3.51	van der Waals
A	No SSE	128	ASP	OD1	CG	A	901	3.69	van der Waals
A	No SSE	128	ASP	OD1	CE	A	901	3.83	van der Waals
A	No SSE	128	ASP	OD1	NH1	A	901	2.67	H.Bond
A	No SSE	128	ASP	OD1	OH1	A	901	3.16	H.Bond
A	No SSE	128	ASP	OD2	NH1	A	901	3.46	H.Bond
A	No SSE	128	ASP	OD2	OH1	A	901	3.09	H.Bond
A	No SSE	130	ASN	ND2	O	A	901	3.04	H.Bond
A	No SSE	137	SER	CB	OXT	A	901	3.25	van der Waals
A	No SSE	137	SER	OG	C	A	901	3.53	van der Waals
A	No SSE	137	SER	OG	O	A	901	3.82	H.Bond
A	No SSE	137	SER	OG	OXT	A	901	2.5	H.Bond
A	139-142 H: 5	141	HIS	ND1	CE	A	901	3.95	H.Bond

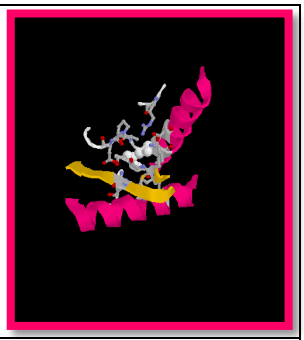
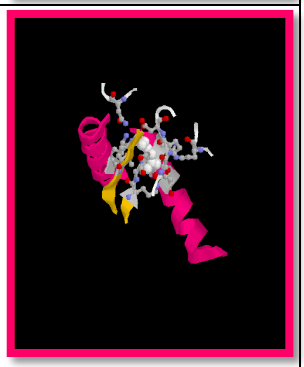
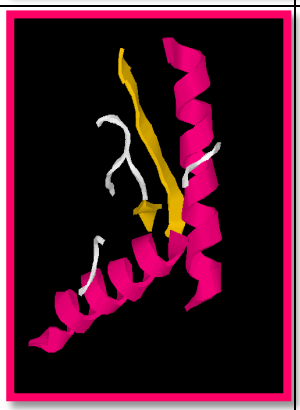
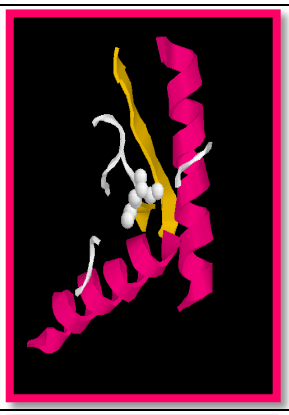
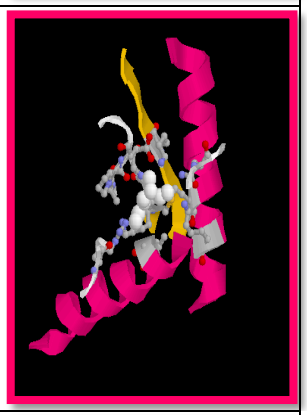
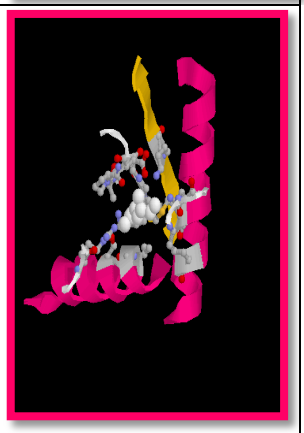
A	139-142 H: 5	141	HIS	ND1	NH2	A	901	3.64	H.Bond
A	139-142 H: 5	141	HIS	CD2	ND	A	901	3.88	H.Bond
A	139-142 H: 5	141	HIS	CE1	ND	A	901	3.52	H.Bond
A	139-142 H: 5	141	HIS	CE1	CE	A	901	3.55	van der Waals
A	139-142 H: 5	141	HIS	CE1	NH2	A	901	3.26	H.Bond
A	139-142 H: 5	141	HIS	NE2	CG	A	901	3.93	H.Bond
A	139-142 H: 5	141	HIS	NE2	ND	A	901	3.25	H.Bond
A	139-142 H: 5	141	HIS	NE2	CE	A	901	3.73	H.Bond
A	139-142 H: 5	141	HIS	NE2	NH2	A	901	3.88	H.Bond
A	139-142 H: 5	142	GLY	CA	O	A	901	3.71	van der Waals
A	139-142 H: 5	142	GLY	CA	CG	A	901	3.99	van der Waals
A	139-142 H: 5	142	GLY	O	O	A	901	3.73	H.Bond
A	183-194 H: 1	183	ASP	CG	N	A	901	3.52	H.Bond
A	183-194 H: 1	183	ASP	OD1	N	A	901	3.7	H.Bond
A	183-194 H: 1	183	ASP	OD2	N	A	901	2.81	H.Bond
A	183-194 H: 1	183	ASP	OD2	CA	A	901	3.51	van der Waals
A	183-194 H: 1	183	ASP	OD2	C	A	901	3.77	van der Waals
A	183-194 H: 1	183	ASP	OD2	O	A	901	3.57	H.Bond
A	183-194 H: 1	186	GLU	OE1	N	A	901	3.56	H.Bond
A	227-232 S: 1	232	ASP	CG	OH1	A	901	3.93	van der Waals
A	227-232 S: 1	232	ASP	OD2	NH1	A	901	3.93	H.Bond
A	227-232 S: 1	232	ASP	OD2	OH1	A	901	2.68	H.Bond
A	234-236 H: 5	234	ASP	OD1	NH2	A	901	3.36	H.Bond
A	234-236 H: 5	234	ASP	OD1	OH1	A	901	3.87	H.Bond
A	No SSE	246	THR	CB	NH2	A	901	3.99	H.Bond
A	No SSE	246	THR	OG1	ND	A	901	3.7	H.Bond
A	No SSE	246	THR	OG1	CE	A	901	3.43	van der Waals
A	No SSE	246	THR	OG1	NH2	A	901	2.69	H.Bond

Table n°32: The binding environment details of the NNH bound the enzyme ARGS (PDB id: 3KV2) ,(chain A) as calculated by the **Lgb** system.

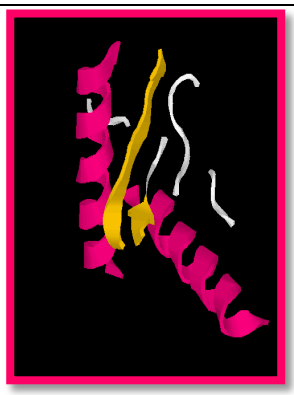
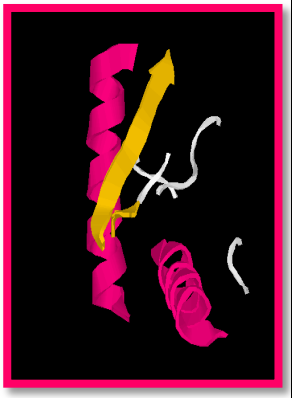
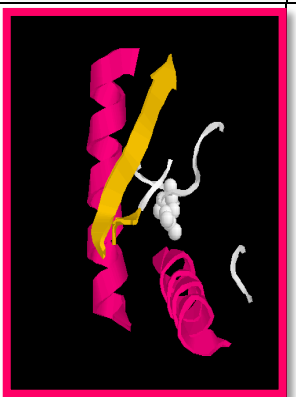
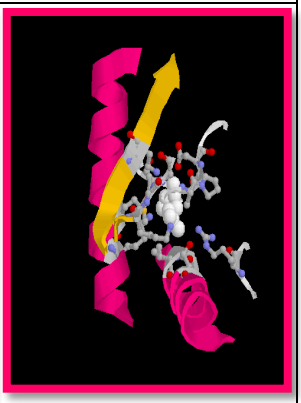
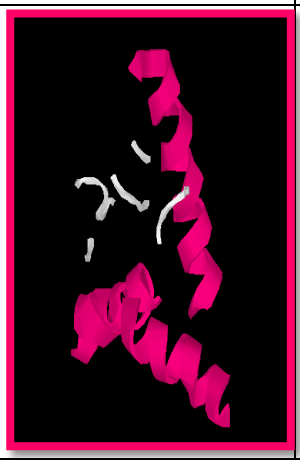
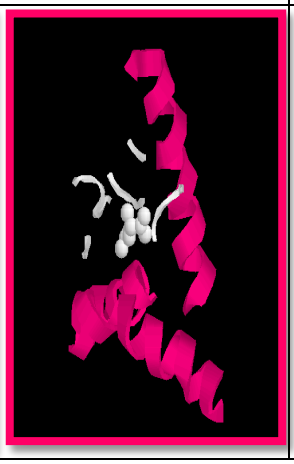
Note: The rest of the binding details for other chains or other pdb entries are stored in the online database.

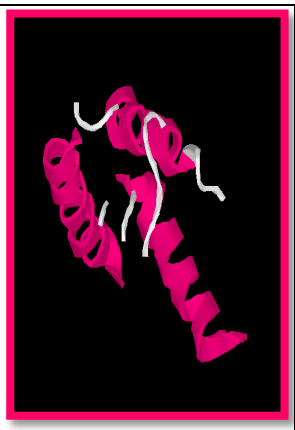
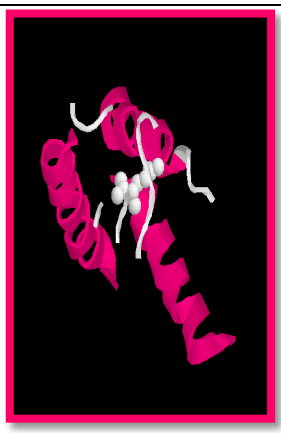
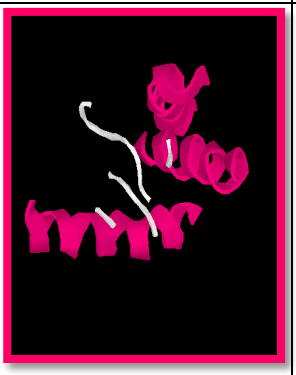
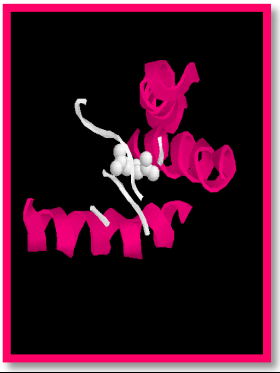
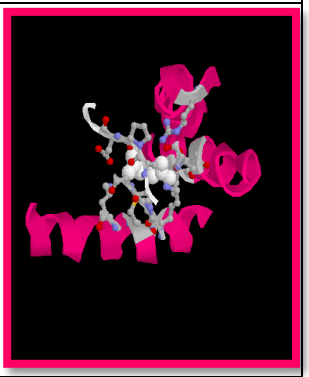
I- 3D representation of the ligands binding motifs

pdb id	ligand id	chain	Motif lineair	motif 3d representation structure.	motif +ligand	motif binding site
3II0	PLP	A	HLSLSL			
		B	HLSLSL			
		C	HLSLSH			
		D	HLSLSL			

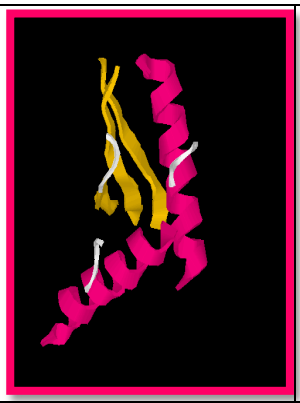
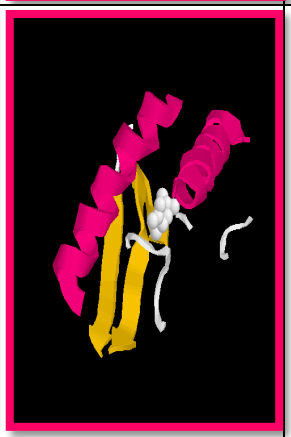
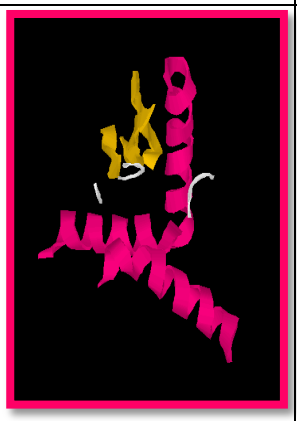
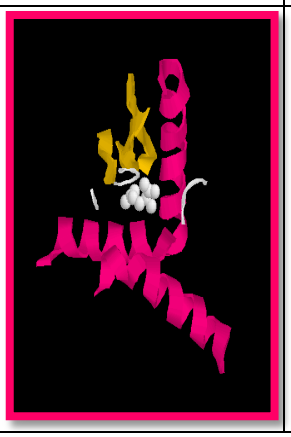
3ETD	**GLU**	**A**	**SLHSLH**			
	GLU	**B**	**SLHSLH**			
	GLU	**C**	**SLHSLH**			
	GLU	**D**	**SLHSLH**			

Index II

		E	**SLHSLH**			
		F	**SLHSLH**			
3MVO	**GLU**	**A**	**LHLHLH**			
	GLU	**B**	**LHLHLH**			

	GLU	**C**	**LHLHLH**			
	GLU	**D**	**LHLHLH**			
		E	**LHLHH**			
		F	**LHLHLHH**			

Index II

3MVQ	**GLU**	**A**	**SHSSLH**			
	GLU	**B**	**SHSSLH**			
	GLU	**C**	**SLHSLH**			
	GLU	**D**	**SHSSLHLH**			

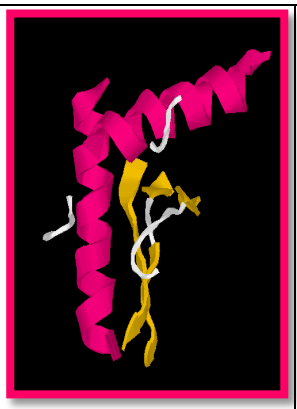
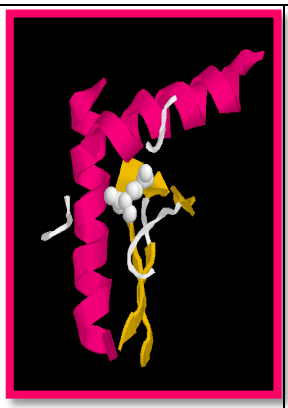
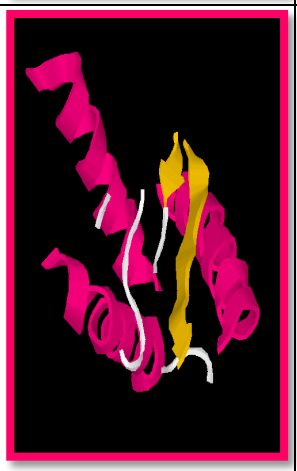
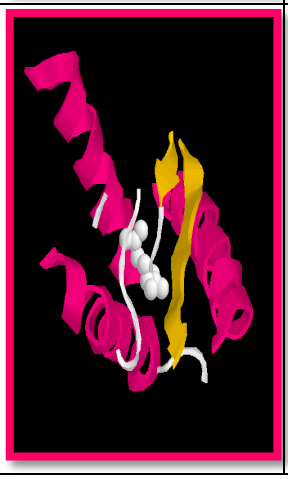
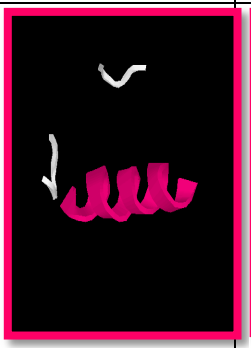
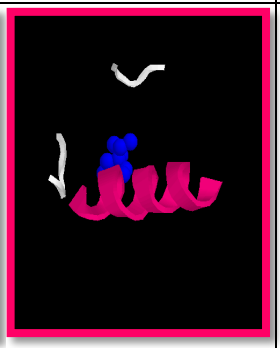
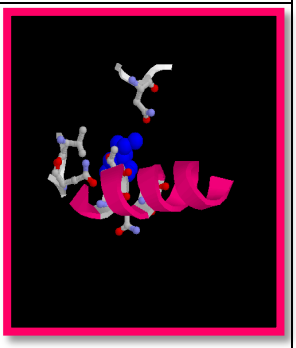
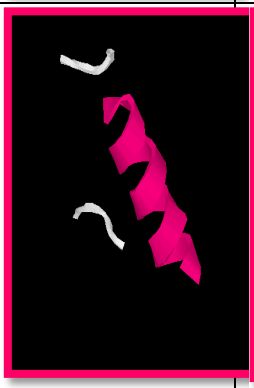
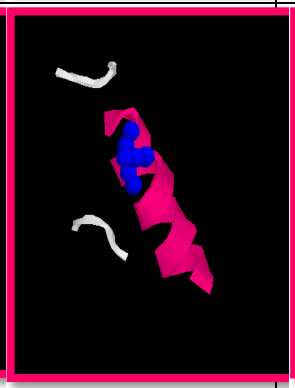
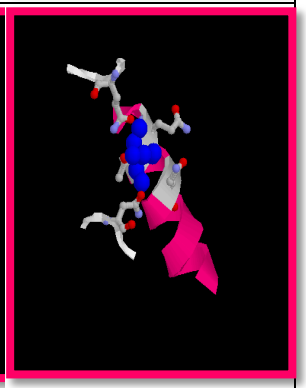
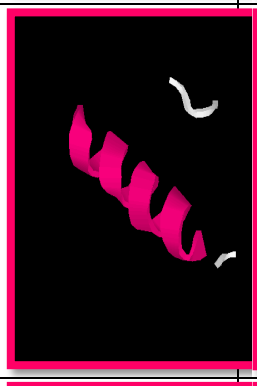
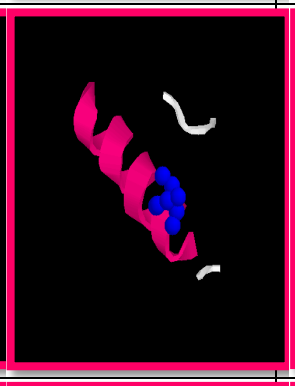
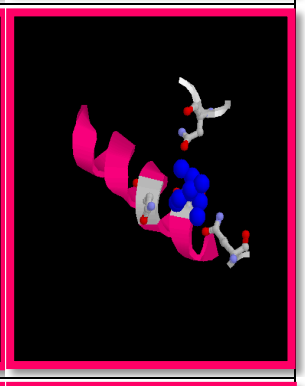
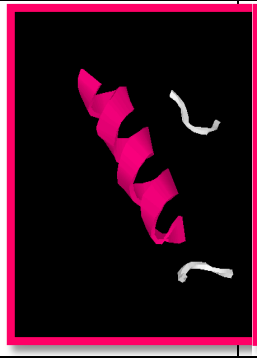
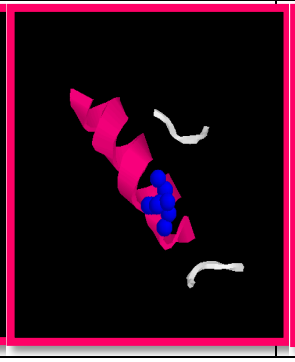
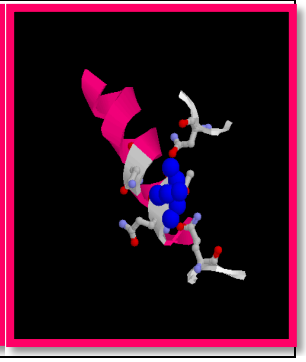
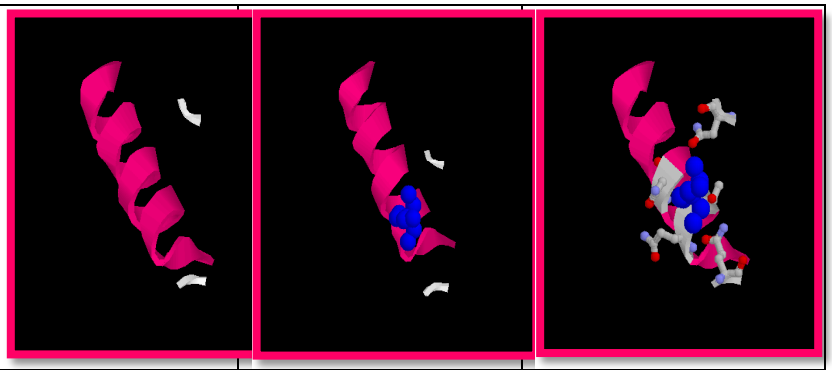
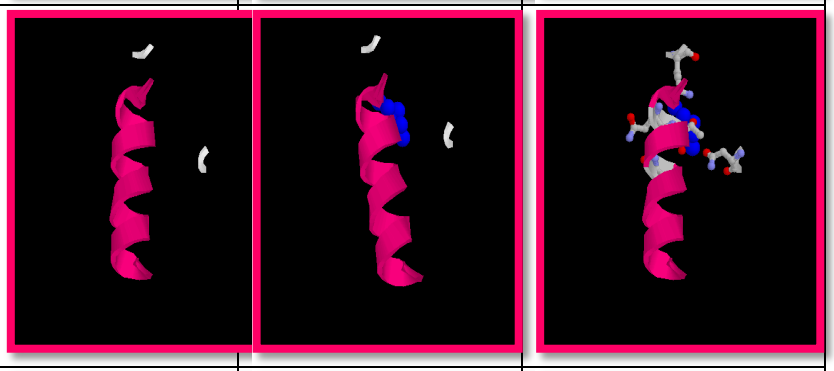
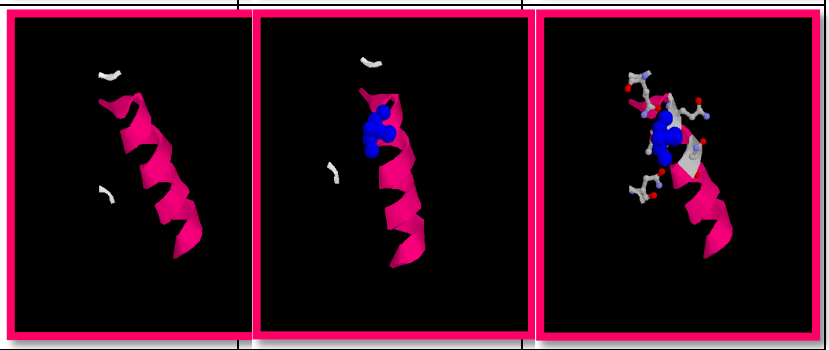
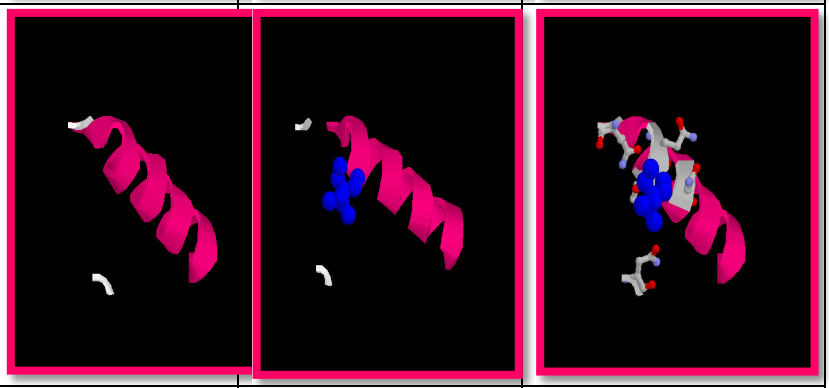
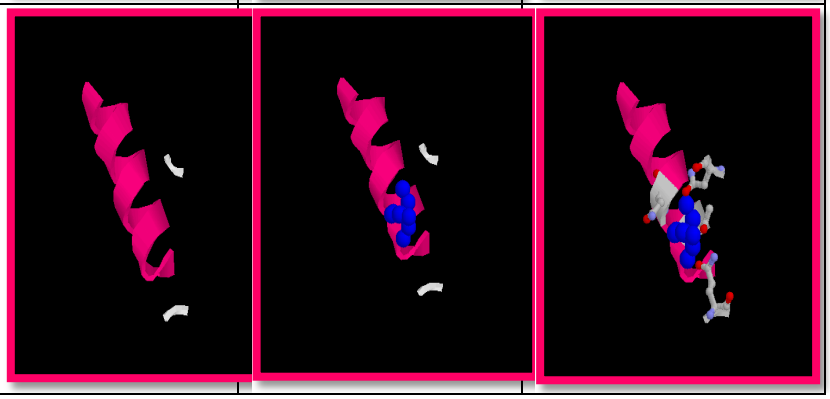
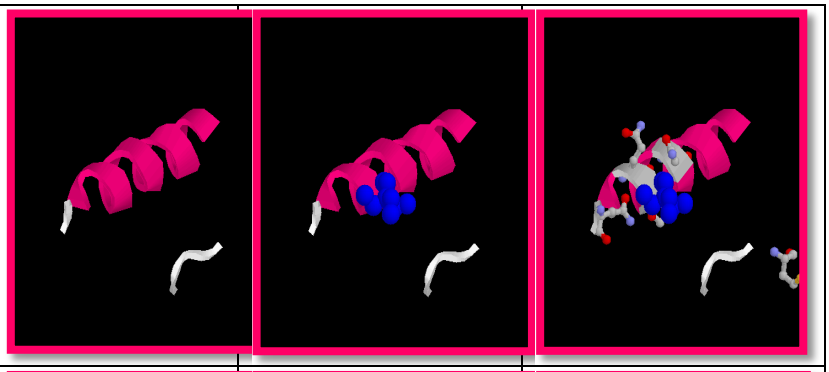
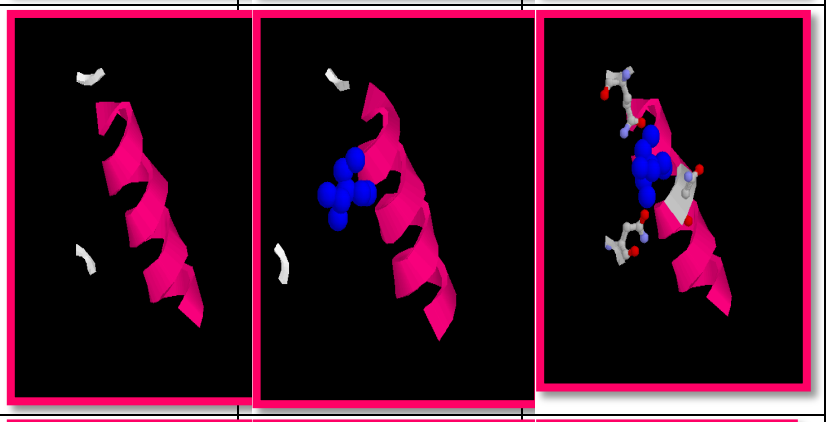
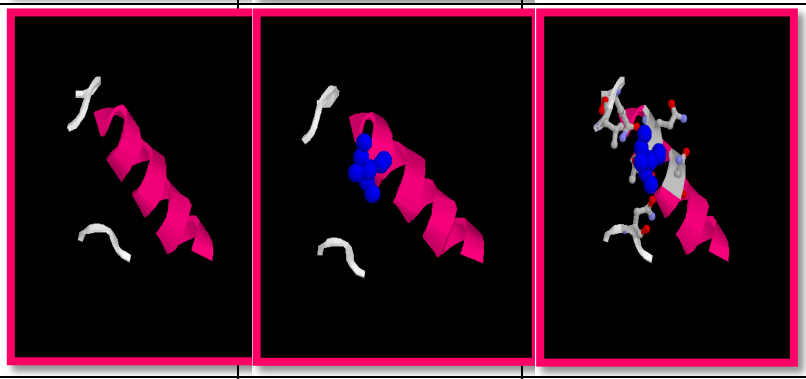
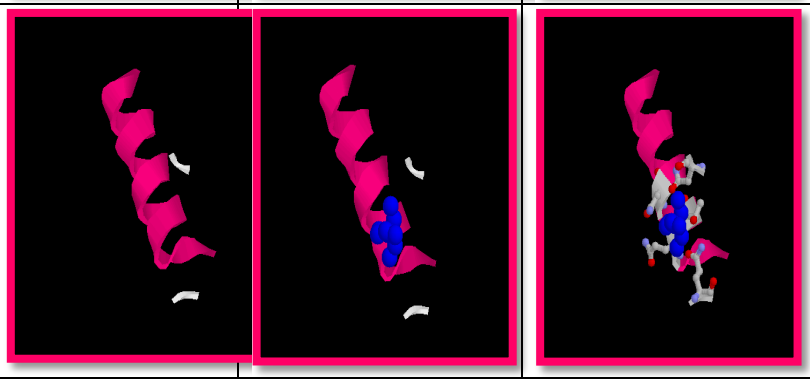
		E	**SSHSSLH**			
		F	**SLHSLHLH**			

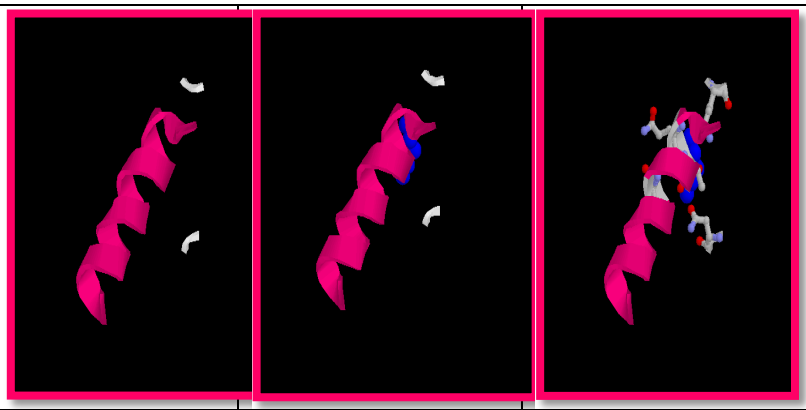
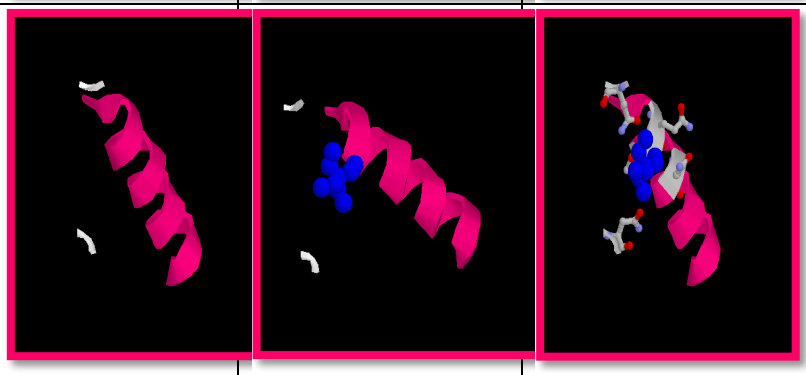
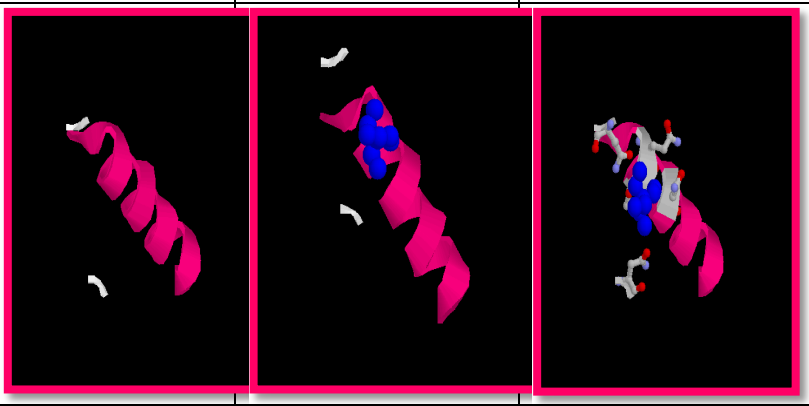
Table n°34 : 3D representation of the ligands binding motifs associated with the enzymes of amino acids degradation from the PDB entries :3II0,3ETD,3MVO,3MVQ.

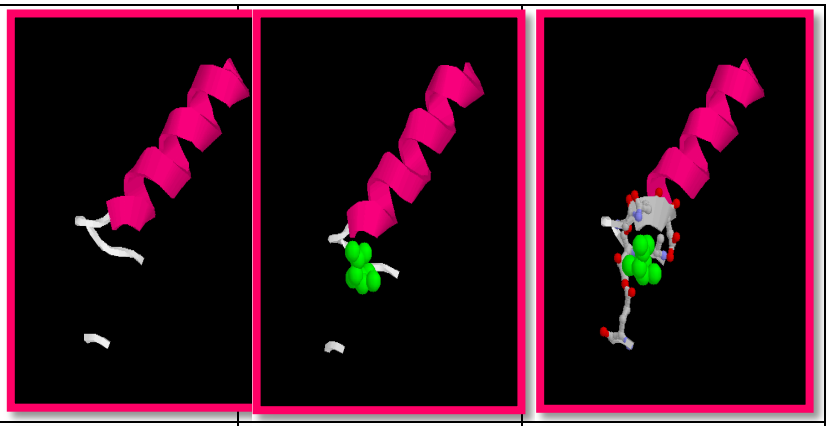
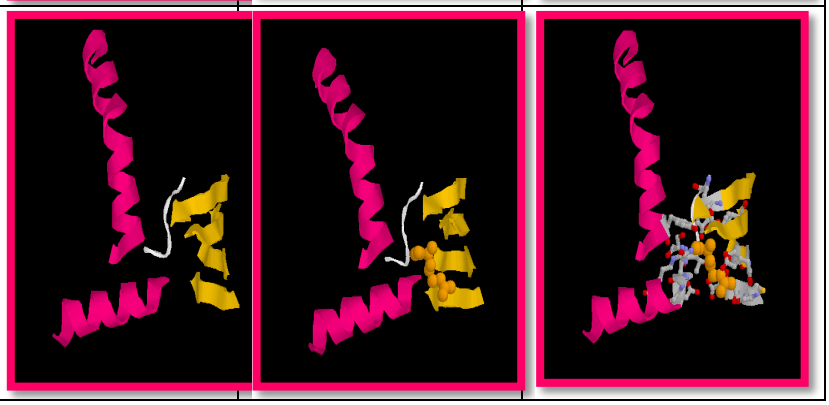
Index II

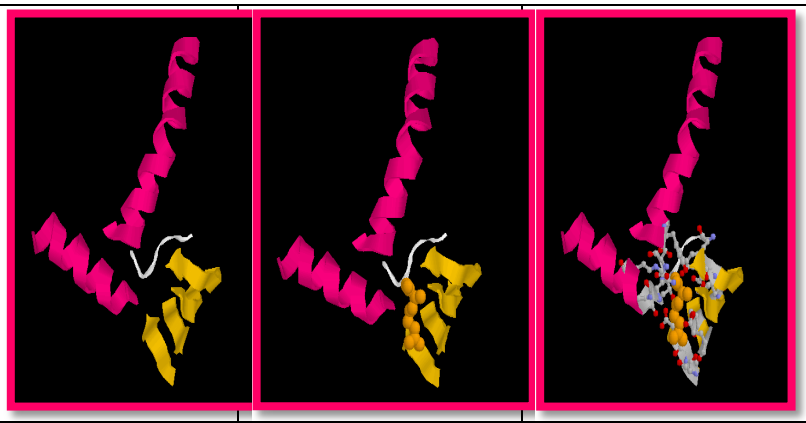
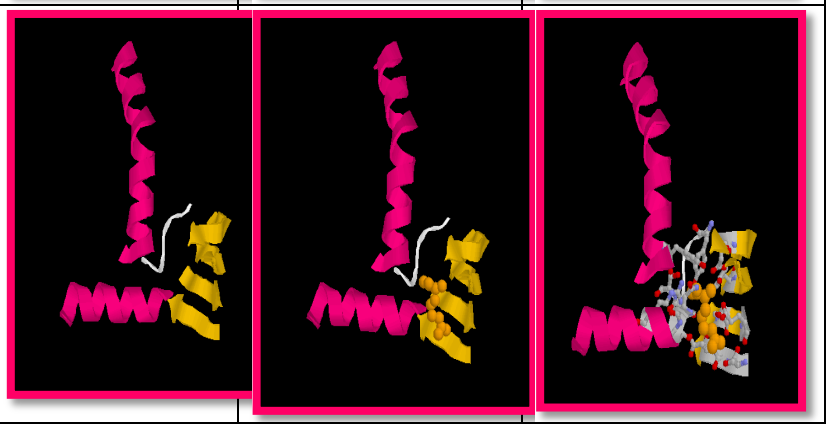
pdb id	ligand id	Chain	motif lineaire	motif 3d representation structure	motif +ligand	motif binding site
1JDB	NET	B	LHL			
	NET	E	LHL			
	NET	H	LHL			
	NET	K	LHL			

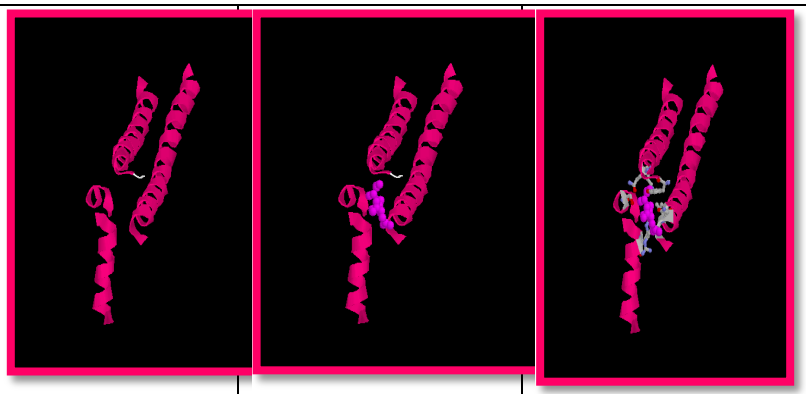
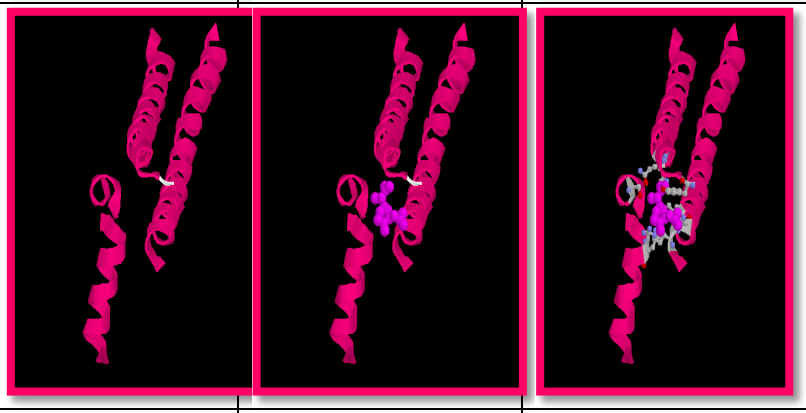
1T36	NET	A	LHL	
	NET	C	LHL	
	NET	E	LHL	
	NET	G	LHL	
1A9X	NET	A	LHL	

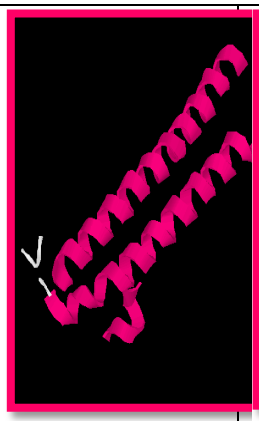
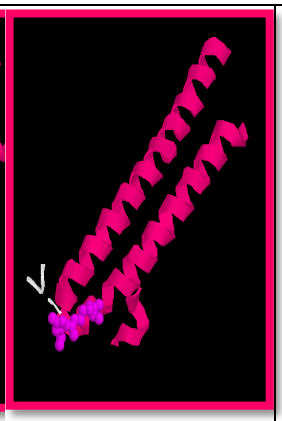
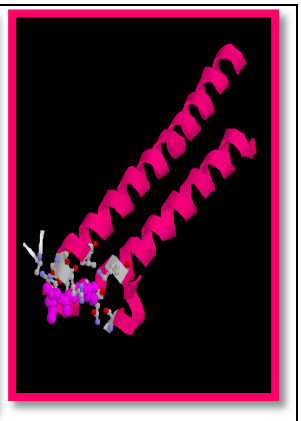
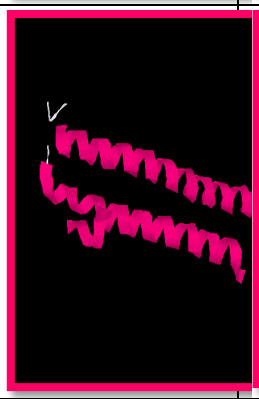
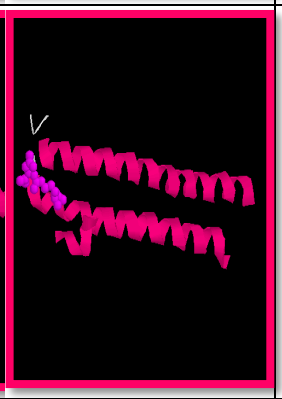
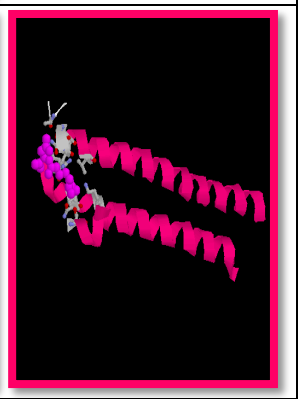
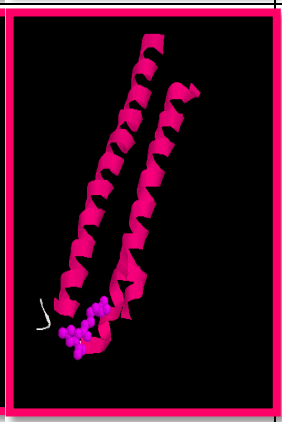
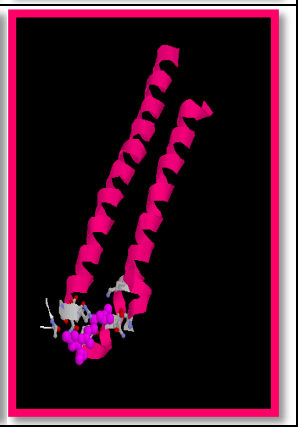
	NET	**C**	**LHL**	
	NET	**E**	**LHL**	
	NET	**G**	**LHL**	
1KEE	**NET**	**A**	**LHL**	

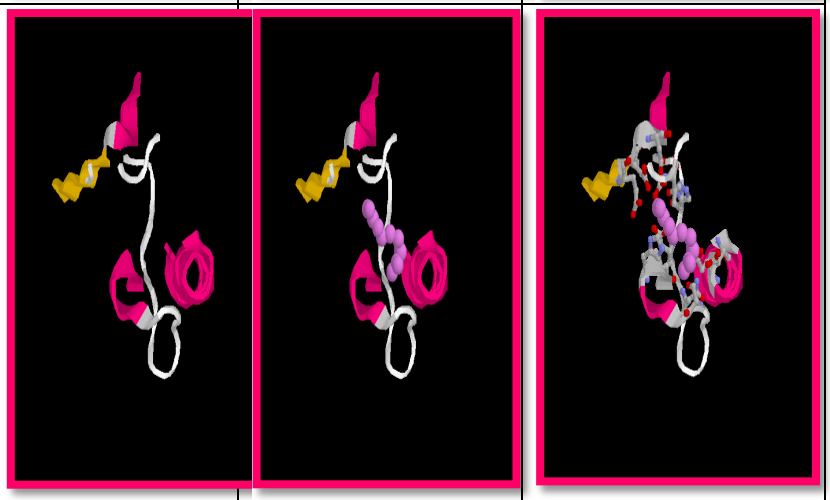
	NET	**C**	**LHL**	
	NET	**E**	**LHL**	
	NET	**G**	**LHL**	
1OTH	**PAO**	**A**	**LHSLHLHH**	

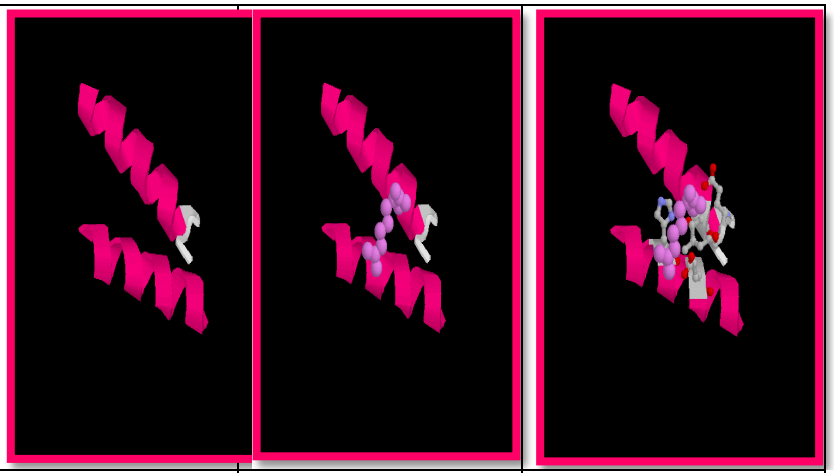
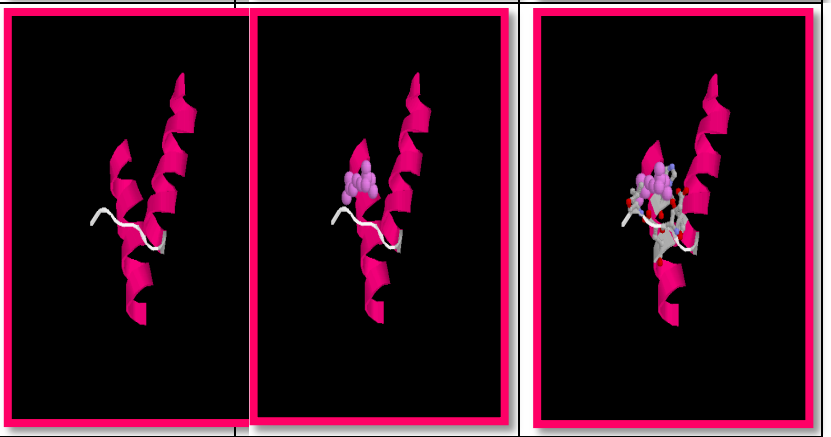
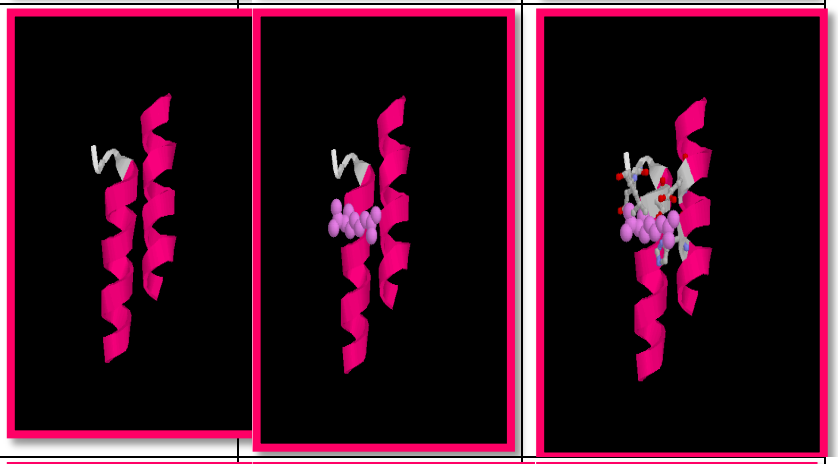
2NZ2	**ASP**	**A**	**LHL**	
	CIR	**A**	**LHLSSLSH**	
1J1Z	**ASP**	**A**	**LHS**	
	CIR	**A**	**LHSSSSH**	

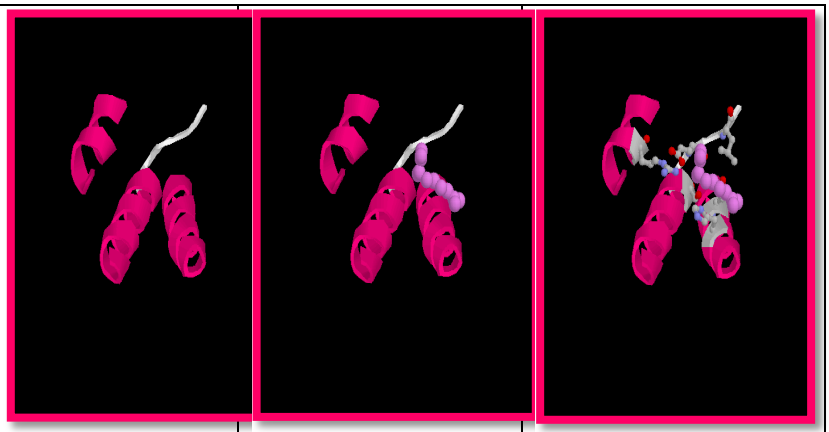
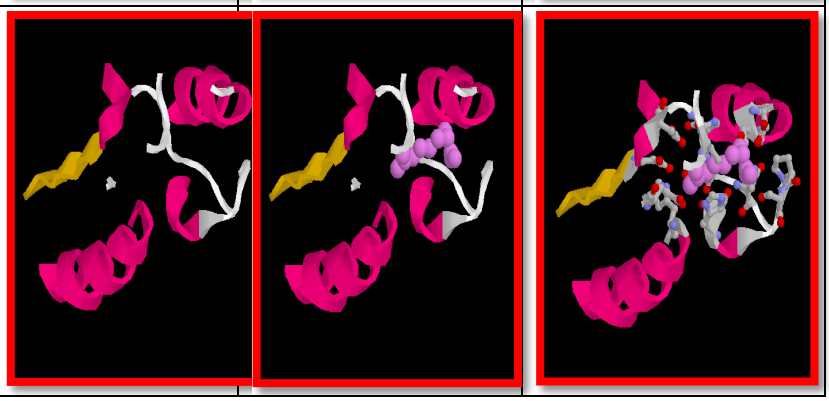
		B	**LHSSSSSH**	
		C	**LHSSSSH**	
		D	**LHSSSSH**	
1K7W	**AS1**	**A**	**HHLHH**	

		B	**HHHLHH**	
		C	**HHHLHH**	
		D	**HHHLHH**	
1TJW	**AS1**	**A**	**HHLHH**	

		B	**HHLHH**			
		C	**HHLHH**			
		D	**HHLHH**			
3KV2	**NNH**	**A**	**LHHSHL**			

		B	**LHHSHL**			
3LP7	**HAR**	**A**	**HLHHSHL**			
		B	**LHHSHL**			

3CEV	ARG	A	HHL			
		B	HHL			
		C	HHL			
		D	HHHL			

		E	**HHHL**	
		F	**HHL**	
		A	**HLHLHSHL**	
		B	**HLHLHSHL**	

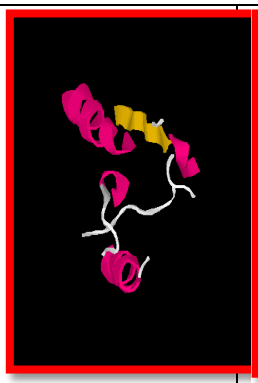
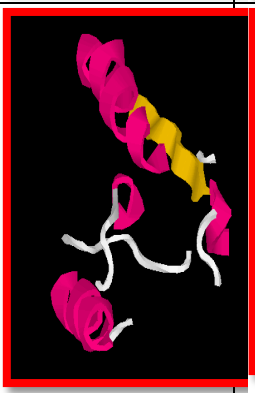
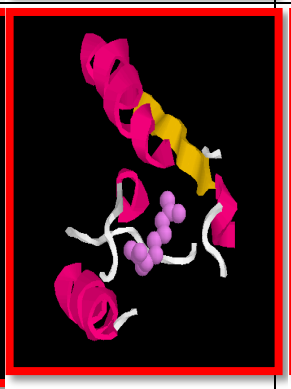
		C	**HLHLHSHL**			
		D	**HLHLHSHL**			
		E	**HLHLHSHL**			
		F	**HLHLHSHL**			

Table n°35:: 3D representation of the binding motifs associated with the enzymes involved in urea cycle from the PDB entries .