PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA

MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH

UNIVERSITY OF SAIDA, DR. MOULAY TAHER

FACULTY OF SCIENCES

DEPARTMENT OF BIOLOGY

MEMORY

PRESENTED FOR OBTAINING THE MASTER'S DEGREE IN BIOLOGY

SPECIALITY: Biochemistry and Cell Physiology

Presented by: Hamidi Ikram and **Naimi Soumeya**

Structural Protein-Protein Interface Binding Motifs; A Contribution Study Towards Discovering The Basis Behind the Protein Structure/Function Relationship annotated in the online database: http://bioinformaticstools.org/prjs/pddsfms/

Presentation date: 10 October 2019

In front of the jury commission composed by:

Mr. Boumediene LASRI	**Prof.**	**President**
Mr. Abdelkrim BERROUKCHE	**Prof.**	**Examiner**
Mr. Abdelkrim RACHEDI	**MCA**	**Supervisor**

2018/2019

List of Abbreviations

3D: Three-dimensional.

AA: Amino acid.

BMRB: Biological Magnetic Resonance Bank

CATH: Class Architecture Topology and Homology.

Cryo-EM: Cryo-Electron Microscopy.

CSV: Comma Separated Values

NMR: Nuclear Magnetic Resonance spectroscopy.

PDB: Protein Data Bank.

PDBe: Protein Data Bank Europe.

PDBj: Protein Data Bank Japan.

PD:DI-SFIM: Protein Domain-Domain Interactions - Structural & Functional Interface Motifs

RCSB: Research Collaboratory for Structural Bioinformatics.

SCOP: Structural Classification OfProteins

H: Helix/Helical structural element

L: Loop region

S: Beta-strand structural element

HB: Hydrogen Bond

TEM: Transmission Electron Microscopy.

HT: Hetero

HM: Homo

PPI: Protein Protein Interaction

SSElM : Secondary Structral Element

Abstract

Abstract

Bioinformatics is a highly interdisciplinary field that uses techniques and concepts of informatics, genetics, biochemistry, chemistry, mathematics, statistics, etc. Research in this discipline consists in developing methods for the collection, stockage, organization, visualization and analysis of biological data.

Proteins play a major role in cellular processes; therefore, it is important to understand how they perform their functions. Proteins, however, do not as separate entities; they work together to create various biological processes.

Multiple protein units or chains physically bind together to form stoichiometrically stable and functional protein complexes. These complexes may also interact with each other to form functional modules and pathways that carry out most cellular processes.

In this project, some proteins have been selected for a structural-bioinformatics study to try and discover some of the underlying basis behind their biological function.

It was necessary to use spatial structures of proteins from the international database known as the Protein Data Bank or PDB which stores proteins in complex forms.

The study resulted in the detection, identification and description of a group of structural elements (called here Structural Motifs) that are deemed important in the function of the protein complexes.

Keywords:

Protein-protein Interface, Structure-function Relationship, Structural Motifs, PDB, Databases, Graphcal scenes, Interactions.

Résumé :

La bioinformatique est un domaine à forte interdisciplinarité qui utilise des techniques et des concepts de l'informatique, la génétique, la biochimie, la chimie, des mathématiques, des statistiques, etc…Les recherches dans cette discipline consistent à développer des méthodes pour la collection, le stockage, l'organisation, la visualisation et l'analyse des données biologiques.

Les protéines jouent un rôle majeur dans les processus cellulaires. par conséquent, il est important de comprendre comment ils remplissent leurs fonctions. Les protéines, cependant, n'agissent pas seules; ils travaillent ensemble pour créer divers processus biologiques.

de multiples protéines se lient physiquement pour former des complexes stables sur le plan stochiométrique. Ces complexes interagissent les uns avec les autres pour former des modules et des voies fonctionnels réalisant la plupart des processus cellulaires.

Dans le cadre de ce projet, certaines protéines ont été sélectionnées pour une étude structurale en bioinformatique dans le but de découvrir certaines des bases de leur fonction biologique.

Il était nécessaire d'utiliser les structures spatiales des protéines de la base de données internationale appelée Protein Data Bank ou PDB, qui stocke les protéines sous des formes complexes.

L'étude a abouti à la détection, à l'identification et à la description d'un groupe d'éléments structurels (appelés ici motifs structurels) jugés importants pour la fonction des complexes protéiques.

Mots clés: Protéines, structure, fonction, motifs structurels, BDP, bases de données, scènes, interactions.

ملخص

المعلوماتية الحيوية هي حقل متعدد التخصصات يستخدم تقنيات ومفاهيم علوم الكمبيوتر ، وعلم الوراثة ، والكيمياء الحيوية ، والكيمياء ، والرياضيات ، والإحصاء ، إلخ. يتكون البحث في هذا التخصص من تطوير طرق لجمع المعلومات. وتخزين وتنظيم وتصور وتحليل البيانات البيولوجية.

تلعب البروتينات دورًا رئيسيًا في العمليات الخلوية. لذلك ، من المهم أن نفهم كيف يؤدون واجباتهم. البروتينات ، ومع ذلك ، لا تعمل وحدها. انهم يعملون معا لخلق مختلف العمليات البيولوجية.

ترتبط البروتينات المتعددة جسديا لتشكيل مجمعات مستقرة. تتفاعل هذه المجمعات مع بعضها البعض لتشكيل وحدات ومسارات وظيفية تؤدي معظم العمليات الخلوية.

كجزء من هذا المشروع ، تم اختيار بعض البروتينات لدراسة هيكلية في المعلوماتية الحيوية من أجل اكتشاف بعض أسس وظيفتها البيولوجية.

PDBكان من الضروري استخدام هياكل البروتين المكاني لقاعدة البيانات الدولية المسماة بنك معلومات البروتين أو

، الذي يخزن البروتينات في أشكال معقدة.

أسفرت الدراسة عن اكتشاف وتحديد ووصف مجموعة من العناصر الإنشائية (يشار إليها هنا بالعناصر الهيكلية) التي تعتبر مهمة لوظيفة مجمعات البروتين.

الكلمات المفتاحية

، قواعد البيانات ، المشاهد ، التفاعلات.BDPالبروتينات ، التركيب ، الوظيفة ، الزخارف الهيكلية ،

Global Introduction

Proteins are required for the biological life to exist and continue. This necessitates precise three-structure starting from the simple monomeric tertiary structureprerequisit.Monomerics units of proteins often adopt a higher level of structure, the quaternary structure, resulting in polymericsProtein-Protein complex forms responsible for the vast variety of proteins and their associated biological functions.

Protein-protein interactions arecentralto the conservation and maintenance of the correct structure/function relationship of the protein complexes. Primary source of these complexes is the international database,the Protein Databank (PDB) created to share the 3-dimensional structures of proteins and many other types of macromoles made available to fundamental research as well as biotechnological implementation for the good of the global community.

The objective of this study is to try and try and discover some of the basics of the relationship between the structure and function in protein complexes. This goal is sought byfirst **selecting** a list of protein strutures, **identify**contact interface regions between the units making the quaternary structures,**calculate**contact interface existing between the inter-chains of the selected complex structures, **highlight** the major types of bonds found between the residues composing the contact interfaces, **define**and **characterise**, when possible, **structural motifs**responsible for the interfaces and explore their properties.

This project is divided into three chapters:

- ✓ **The first chapter:** Literature Review

This provides background information and concepts of protein and their structural levels as well as about protein-protein interaction, methods of protein structure determination, PDB and other useful databases and information required for understanding the subject treated in this project.

- ✓ **The second chapter:**Materials and Methods

The tools and materiasl used in the project are presented.

- ✓ **The Third chapter:** presents results obtained in this study, interpretation and discussion.
- ✓ At the end, the work is completed by a "**General Conclusion**".

Acknowledgment

There are a number of people without whom this thesis might not have been written, and to whom we are greatly indebted.

We would like to express our special thanks of gratitude to our supervisor Dr. Abdelkrim Rachedi, who gave us the golden opportunity to do this wonderful project, which also helped us in doing a lot of research and we came to know about so many new things we are really thankful to have acquired them and especially that he was very patient and understanding with us.

Our thanks also go to the jury members for the great honour they gave to examine this work.

Finally, we must express our very profound gratitude to our parents and family for providing us with unfailing support and continuous encouragement throughout our years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them. Thank you.

Dedication

This thesis is dedicated to: The sake of Allah, my Creator and my Master,

My great teacher and messenger, Mohammed (May Allah bless and grant him), who taught us the purpose of life,

All persons working in University of Saida, Dr. MOULAY TAHAR,

My great parents, who never stop giving of themselves in countless ways,

My beloved sisters: Samira, Manel, Imene, Ikram and Ghania who stand by me when things look bleak,

The symbol of love and giving, my friends who encourage and support me,

All the people in my life who touch my heart, I dedicate this research.

Dedication

I dedicate my dissertation work to my family.

A special feeling of gratitude to my loving parents, whose words of encouragement and push for tenacity ring in my ears.

My sisters Kaother, Sihem have never left my side and are very special.
I also dedicate this dissertation to my dear brother Aymen and mom and dad family who have supported me throughout the process. I will always appreciate all they have done, especially Grandfather ”Kadda" and my Aunt "Fatiha" for helping me develop my biology skills,

I dedicate this work and give special thanks to my best friend ImeneSlimani for helping me to master the leader dots.

And my wonderful friend and partner Soumeya Naimi who worked with me so hard in order to complete this work. All of you have been my best cheerleaders.

ChapterI: Literature review

Chapter II: Material and methods

List of Figures

Figure 61: The main web interface of the PD:DI-SFIM

Figure 62: Capture of the interface page of the" PD:DI-SFIM" database. Ellipse shapes and Numbers are for highlighting the four different ways made available to query the database

Figure 63: results for interface binding motifs types details.

Figure 64: capture represents the interface after clicking a PDBid "6ms4"

Figure 65a: Screen capture represents enlarged image of motifs after clicking [+] on the left.

Figure 65b: Screen capture represents enlarged image of motifs after clicking the [+] on the right.

Figure 65c: Screen capture represents enlarged interface bonding contacts details after clicking [+]

Figure 66: Screenshot shows the result page after the selection of a PDBid.

Figure 67: Represents the binding details to pdb id 6ms4.

Figure 68: Information about pdb id 6ms4 existing on PD:DI SFI database.

Figure 69: The 3D-structure of the interface binding motif HSLLLHHHL in red (chain A) / LLLLHL in yellow (chain B).

Figure 70: Double interface binding motif as shown in capture of the main page of the "PD:DI-SFIM".

Figure 71: Single ligand binding motif as shown in capture of the interface page of the "CacSFMs" database.

List of tables

I. Essential molecules of protein formation:

I.1. DNA (deoxyribonucleic):

The human body is estimated to contain 10 trillion cells and at some stage in its life cycle each contains a full complement of genes needed by the entire organism. Genes, composed of DNA in the nucleus, are clustered together in the chromosomes. In the chromosomes of all but the most primitive organism, DNA combines with protein. DNA, the molecular basis of heredity in higher organisms, is made up of a double helix held together by hydrogen bonds between purine and pyramidine bases i.e. between Adenine and Thymine and between Guanine and Cytosine. **(M. H. Fulekar,2009).**

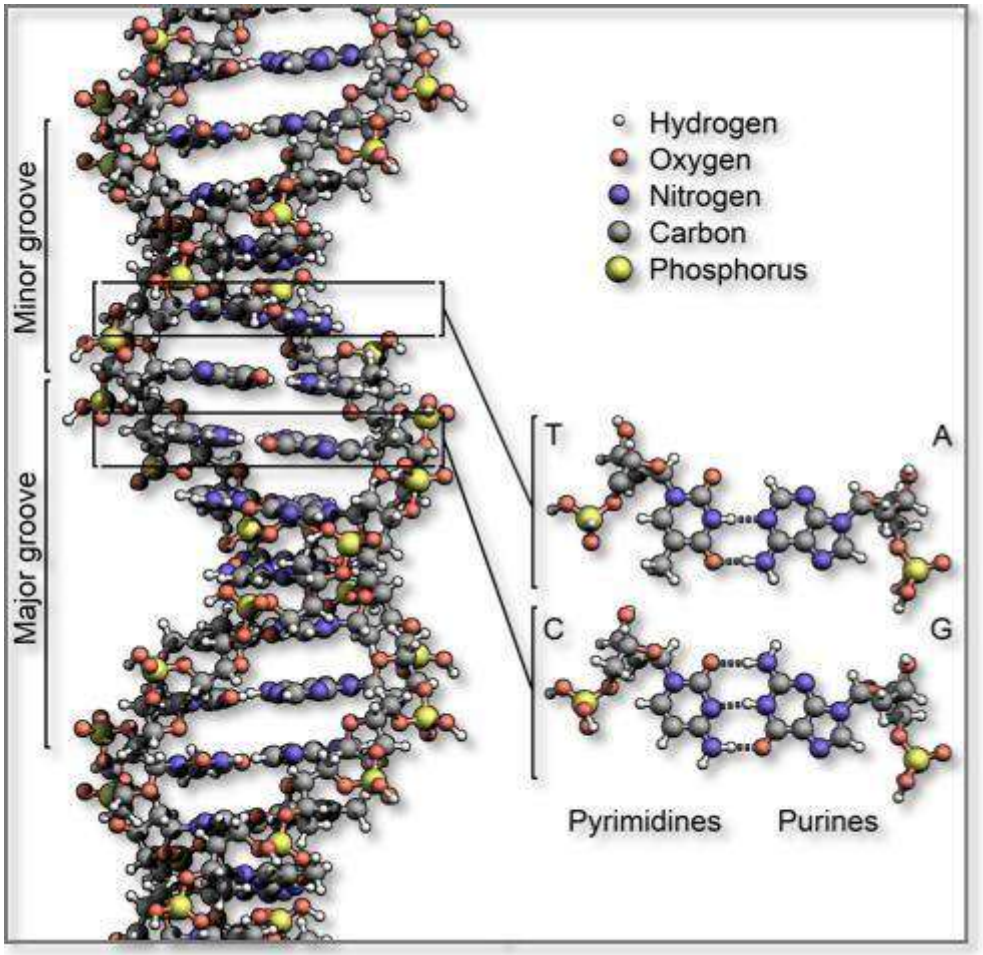


Figure 1: DNA STRUCTURE

(Jefferson,2018)

I.2. Genes:

A gene is a functional and inheritable element in the genome that usually codes for a protein. The protein coding sequence of most genes is interrupted many times by blocks of noncoding sequences**. (Graz, December 2005)**

I.3. RNA (ribonucleic acid):

RNA are similar to DNA. The major function of RNA is to copy information from DNA and bring it out of the nucleus to use wherever it is required to be used. The RNA in the cell has at least four different functions.Messenger RNA (RNAm) is used to direct the synthesis of specific proteins.

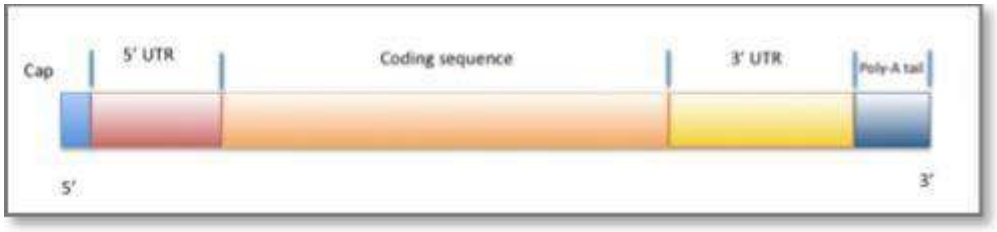


Figure 2: The structure of an mRNA (Jefferson, 2018)

Transfer RNA (RNAt) is used as an adapter molecule between the RNAm and the amino acids in the process of making proteins.

Ribosomal RNA (RNAr) is a structural component of a large complex of proteins and RNA known as the ribosome. The ribosome is responsible for binding to the RNAm and directing the synthesis of proteins.

The fourth class of RNA is a catch-all class. There are small, stable RNAs whose functions remain a mystery. Some small, stable RNAs have been shown to be involved in regulating expression of specific regions of the DNA. Other small, stable RNAs have been shown to be part of large complexes that play a specific role in the cell. In general, RNA is used to convey information from the DNA into proteins. **(M. H. Fulekar, 2009).**

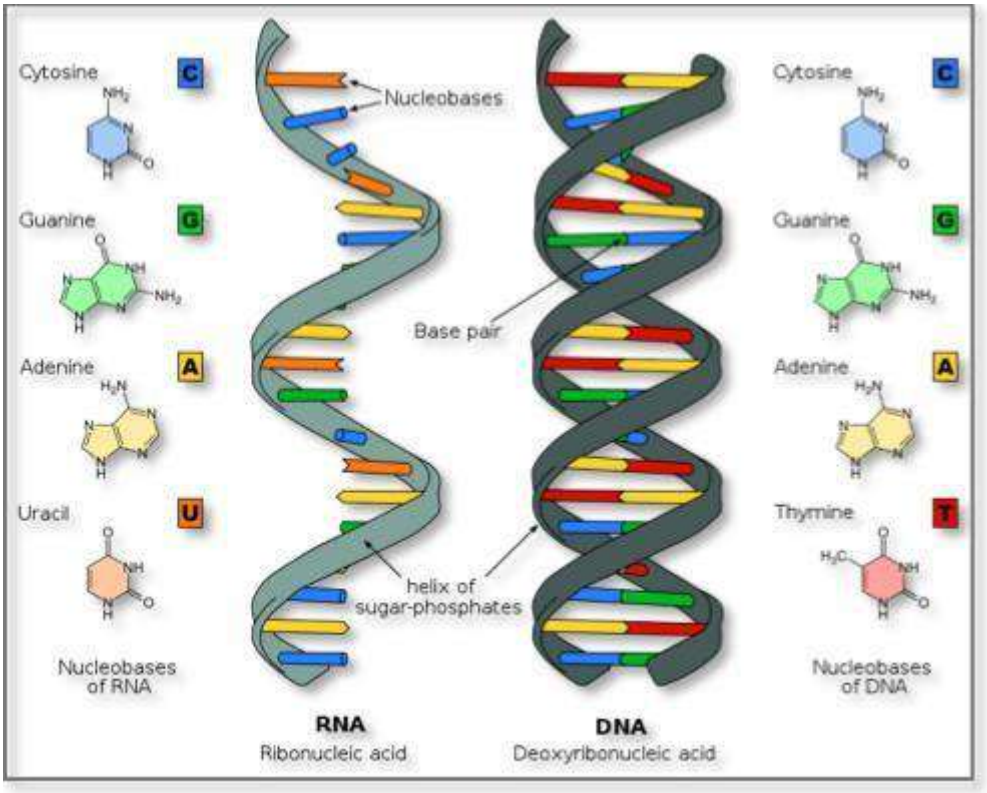


Figure 3: DNA and RNA similarities and differences

(Jefferson, 2018)

I.4. Amino Acids:

Any of a group of organic molecules that consist of a basic amino group (—NH2), an acidic carboxyl group (—COOH), and an organic Rgroup (or side chain) that is unique to each amino acid. The term amino acid is short for α-amino [alpha-amino] carboxylic acid. Each molecule contains a central carbon (C) atom, called the α-carbon, to which both an amino and a carboxyl group are attached. The remaining two bonds of the α-carbon atom are generally satisfied by a hydrogen (H) atom and the Rgroup. The formula of a general amino acid is:

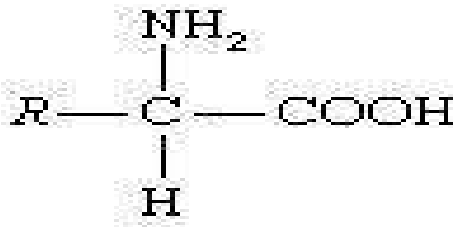


Figure 4 : Amino acid

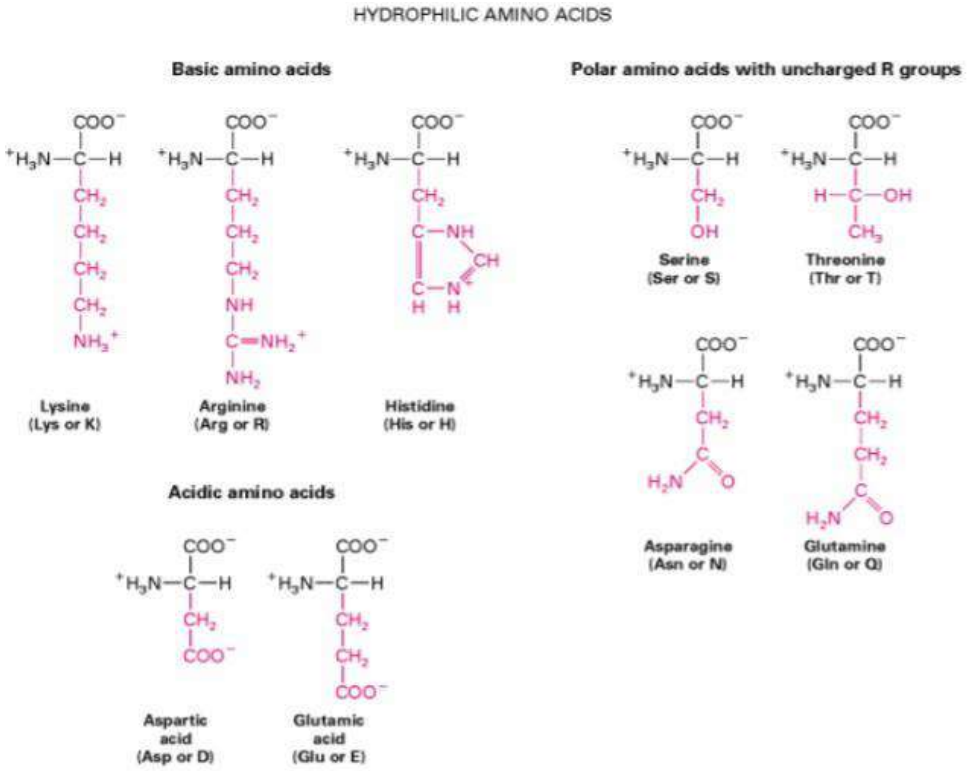
(Michael K. Reddy,2017)

The amino acids differ from each other in the particular chemical structure of the R group.

I.4.a. Properties of Amino Acids

Different amino acids have different structures and chemical properties. For example, hydrophobicity is how much an amino acid wants to avoid water, i.e. it would be in a high energy (unstable) state when in contact with water molecule. Therefore, a sequence of amino acid in the cell will fold into a specific 3-D structure that minimizes its energy. The hydrophobic, also called non-polar, amino acids will tend to be buried inside while the hydrophilic (polar) ones will be more likely on the surface.

Amino acids of opposite charge will tend to be close to each other while ones of the same charge are likely to be farther away. The size of the amino acid also puts a constraint on the possible configuration. In general, a protein will adopt certain 3-D structure based on its amino acid sequence although it is difficult to computationally decide its structure based on the sequence. **(Haidong Wang December 2008)**



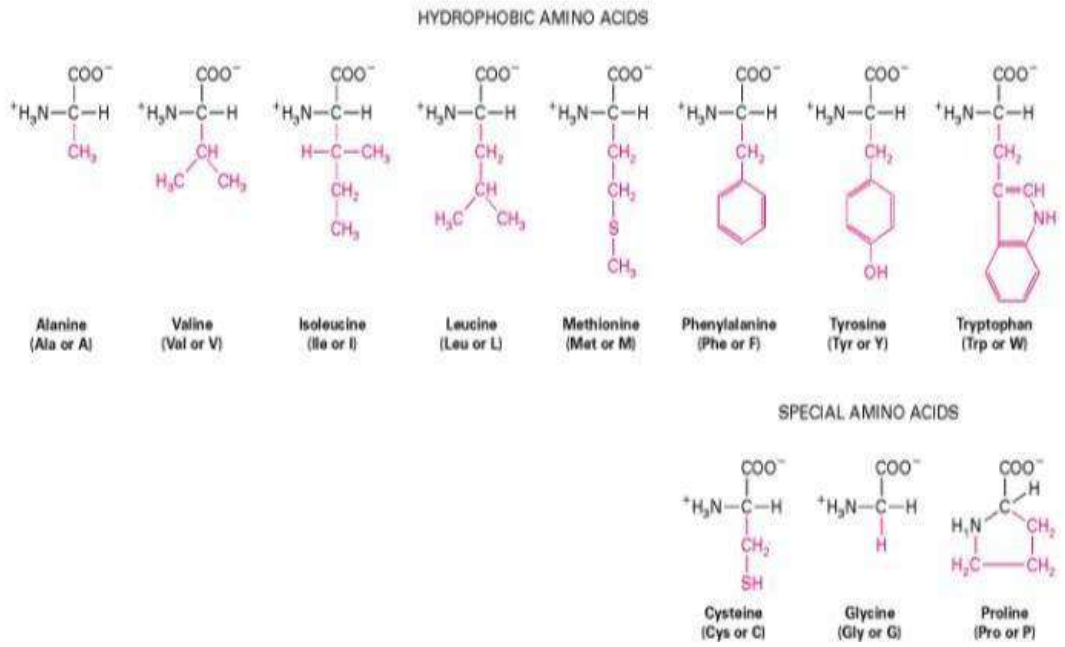


Figure 5: Molecular structure of the twenty essential amino acids that are present in all living organisms. A linear chain of amino acids linked by the amino group of one residue to the carboxyl group of the next makes up a protein. **(JESSICA YEE LAU January, 2005)**

I.5. Peptide Bond:

The carboxyl group of one amino acid interacts with the amino group of another to form a peptide bond by elimination of water. Amino acids are joined end-to-end during protein synthesis by the formation of such peptide bonds. The peptide bonds (C-N) has a partial double-bond character due to resonance, and hence there is no rotation about the peptide bond. In this figure, the peptide is represented as a planar unit with the C=O and N-H groups positioning in opposite directions in the plane. This is called tans-peptide. There is another form. Cis-peptide in which the C=O and N-H groups point in the same direction. To avoid steric hindrance, the trans form is frequently presented in protein structures for all amino acids except Pro, which has both tans and cis forms. The cisprolines are found in bends of the polypeptide chains. (**M. Michael Gromiha 2010**)

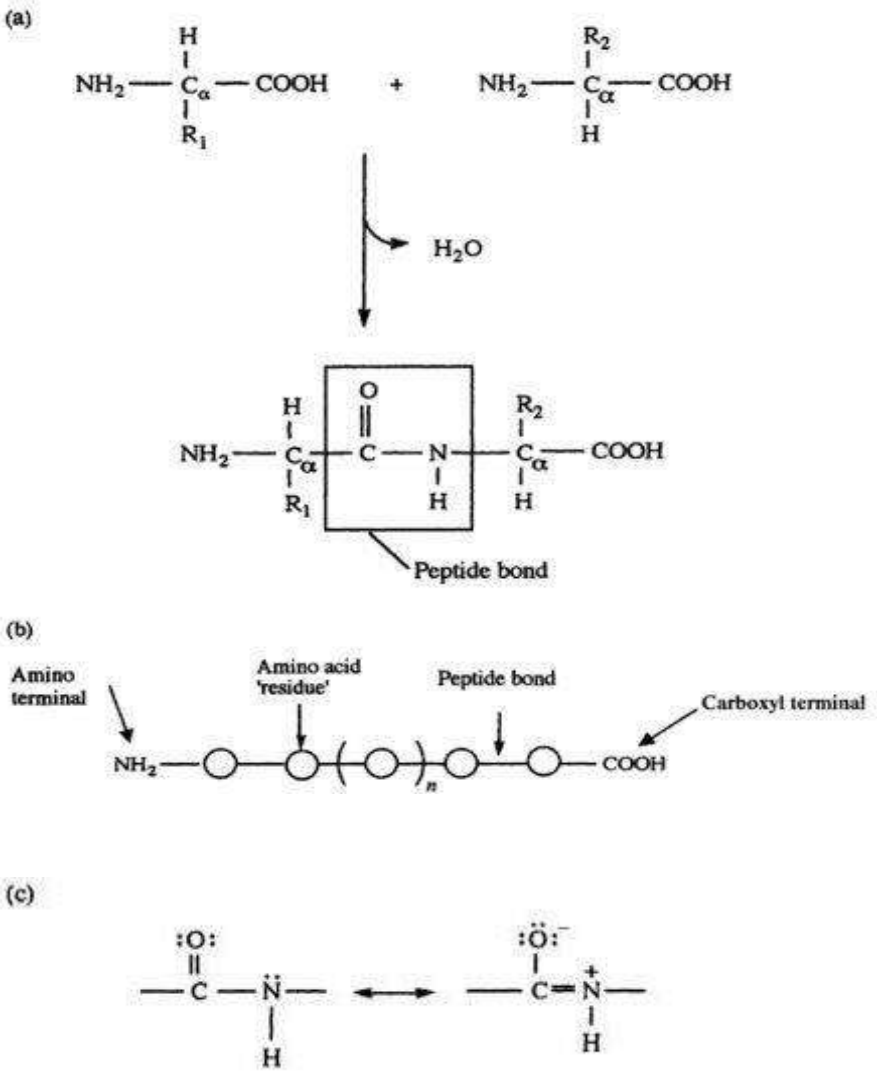


Figure 6 : a) peptide bond formation. b)Polypeptides consist of a linear chain of amino acids successively linked via peptid bonds . c) the peptide bond displays partial double bonded character **.(Gary walsh 2002)**

I.6. Proteins:

The word "protein" is derived from the Greek word prôtos, meaning "primary" or "first rank of importance."

Proteins form the very basis of life. They regulate a variety of activities in all known organisms, from replication of the genetic code to transporting oxygen, and are generally responsible for regulating the cellular machinery and consequently, the phenotype of an organism.

Knowing the structure of the protein, we can probe for its function and potentially apply the new knowledge to various genome projects**. (JESSICAYEELAU,January 2005)**

I.7. The Central Dogma:

The central dogma of molecular biology states that genetic information is stored in DNA and replicated through DNA, which is located inside the nucleus. DNA is a sequence of four different types of nucleotides. Certain segments of the DNA correspond to genes, which under appropriate condition would be used to make a molecule called mRNA, whose sequence directly corresponds to the DNA sequence. This process is called transcription. The resulting mRNA migrates out of the nucleus into cell cytoplasm. There, a protein is synthesized in a process called translation using the mRNA as template.(**Haidong Wang December 2008**)

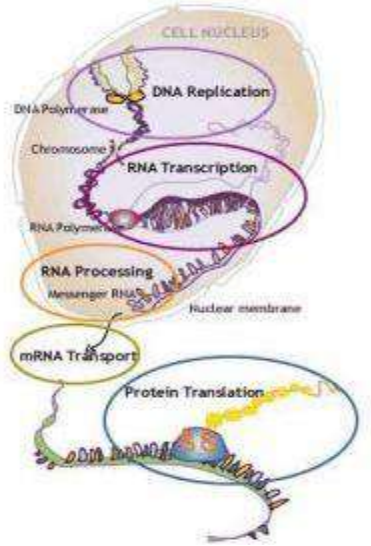


Figure 7: Central Dogma of Molecular Biology(Graz, December 2005)

II. Proteins

II.1. Proteins definition:

Proteins are biological macromolecules. They consist of chains of amino acid sequences, which fold into unique structures. **(Graz, December 2005)**

II.2. Properties:

Protein properties such as 3-D structure of the protein, the chemical properties of its amino acids, and its localization decide which other proteins or small molecules it physically binds to. Usually the binding happens at places of complementary 3- D structure. This kind of physical association enables multiple proteins to form stoichiometrically stable complexes. At the next level, the complexes interact with individual proteins or other complexes to form functional modules and pathways that carry out most cellular processes. Through this hierarchical structure, the limited

number of proteins is able to combine with each other to perform exponentially diverse kinds of cellular function. **(Haidong Wang December 2008)**

II.3. Level of Proteins Structure:

The structure of a proteins are determined by four aspects:

II.3.1. Primary Structure:

Proteins are composed of amino acid residues connected by covalent peptide bonds, which are planar and rigid (Figure 8). Protein primary structure is simply the amino acid sequence. It's linear, not branched and in one-dimension **(Rongkun Shen2006)**

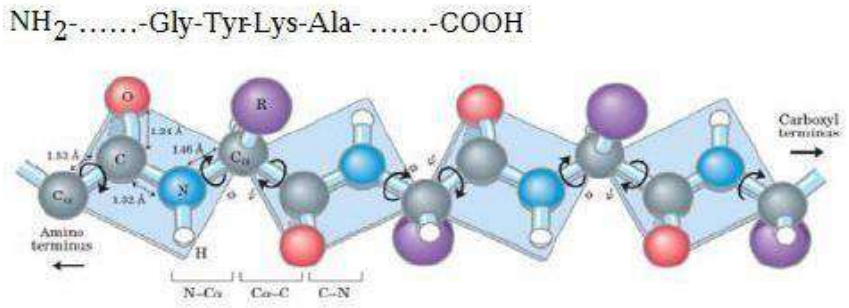


Figure 8: Primary structure of a partial protein sequence. **(Courtesy of Nelson and Cox (2004))**

II.3.2. Secondary Structure:

Because of the planar and rigid peptide pond plus the spatial restriction, the residues are not free to rotate and bend at all angles. Three basic local structures can be formed: α-helix (Figure9), β-strand (Figure10) and random coil **(Rongkun Shen2006)**

II.3.2.1. Alpha-Helix:

An alpha-helix is a right-handed structure, with 3.6 amino acid residues and 13 backbone atoms per turn. Therefore it is also sometimes called the 3.613 helix. The vertical distance between two neighboring turns is 5.4 Å.

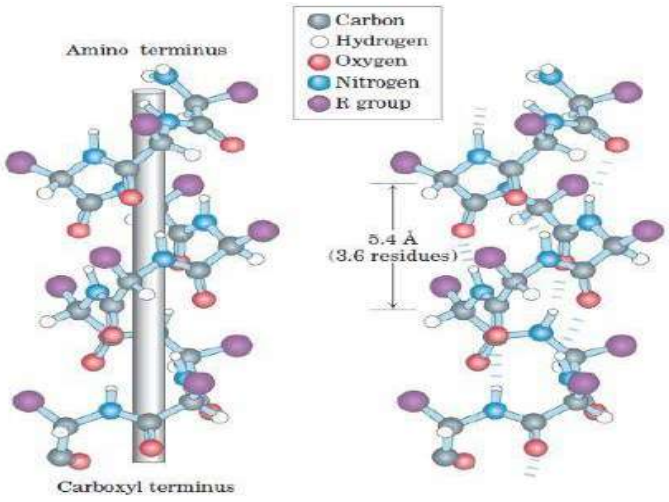


Figure 9: The structure of α-helix. **(Rongkun Shen2006)**

II.3.2.2. Beta-Sheet:

β-strands have a zigzag form and have two types: antiparallel **(Figure 10(a))** and parallel **(Figure 10(b)).**

In parallel strands, two sequence segments are in the same N-terminus to C-terminus direction; in antiparallel strands, they have opposite directions. Two sequence segments in remote positions in the sequence can form β-strands (Figure10 (c)). This is called a long-distance interaction, which is the most difficult part of secondary structure prediction.

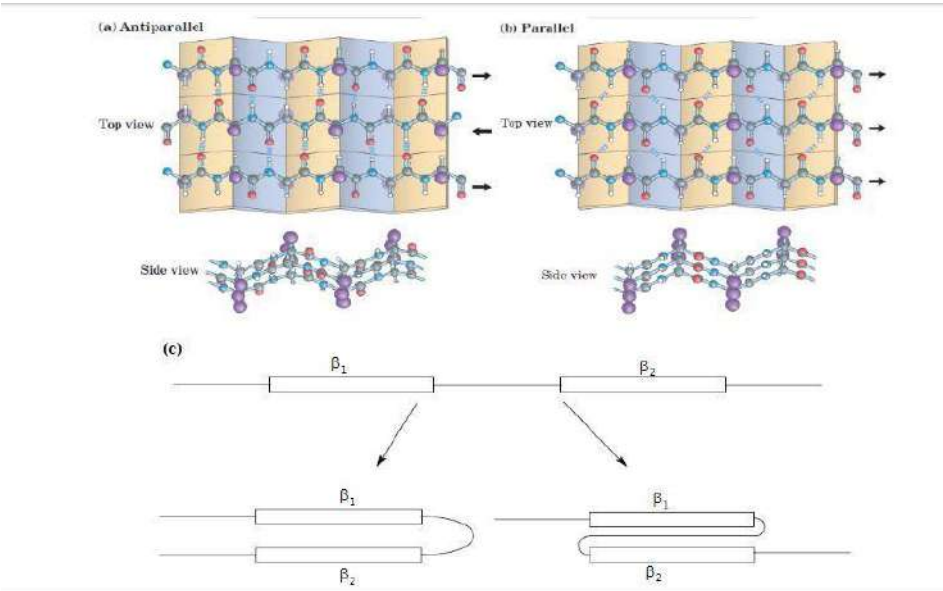


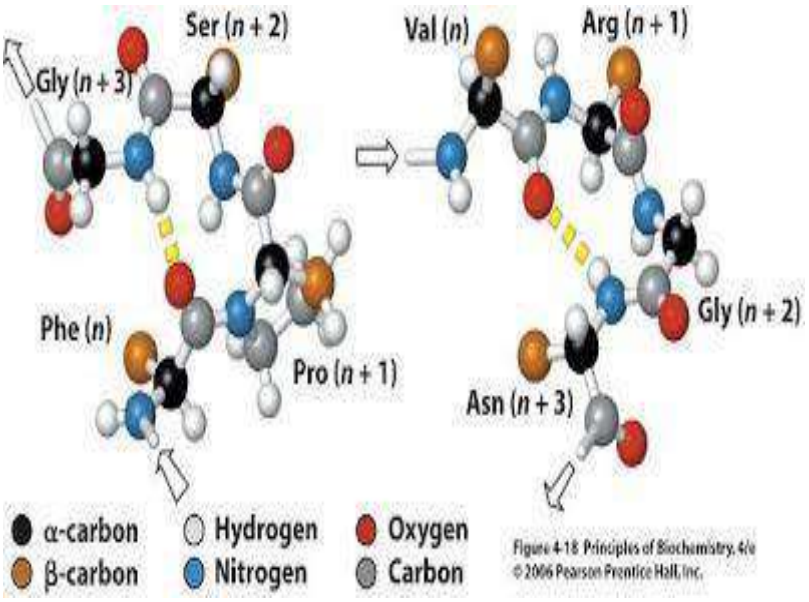
Figure 10: The structure of β-strands. (a) antiparallel β-strands; (b) parallel β-strands; (c) long-distance interaction between two β-strands. ((a) and (b), courtesy**(Rongkun Shen2006)**

II.3.2.3. Loops and Turns

In addition to α helices and β strands, a folded polypeptide chain contains two other types of secondary structure called loops and turns. (There are also regions of unordered structure, often called coils.)

Loops and turns connect α helices and β strands. The most common types cause a change in direction of the polypeptide chain allowing it to fold back on itself to create a more compact structure.

Loops are not well defined. They generally have hydrophilic residues and they are found on the surface of the protein. Loops that have only 4 or 5 amino acid residues are called turns when they have internal hydrogen bonds. Reverse turns are a form of tight turn where the polypeptide chain makes a 180° change in direction. Reverse turns are also called β turns because they usually connect adjacent β strands in a β sheet.



Reverse turns. (left) Type I β turn. The structure is stabilized by a hydrogen bond between the carbonyl oxygen of the first N-terminal residue (Phe) and the amide hydrogen of the fourth residue (Gly). Note the proline residue at position *n* + 1 (right) Type II β turn. This turn is also stabilized by a hydrogen bond between the carbonyl oxygen of the first N-terminal residue (Val) and the amide hydrogen of the fourth residue (Asn). Note the glycine residue at position *n* + 2 [PDB 1AHL (giant sea anemone neurotoxin)]. (Horton et al. 2006)

Figure 11: The two most common types of β turn are the type I and type II turns shown above. The key point about turns is that they are highly ordered structures stabilized by internal hydrogen bonds. This is why they are counted as the third form of secondary structure (along with the α helix and β strand).

(Larry Moran , MARCH 13, 2008)

II.3.2.4. The Ramachandran Plot :

The stereochemistry of the main chain folding can be investigated with a Ramachandran plot in which the dihedral angles φ (phi), ψ (psi)for each residue are plotted in a square matrix. It is customary to have the conformation of the fully extended chain in the corners of the square. Short contacts between atoms of adjacent residues prevent φ (phi) and ψ (psi) from taking on all possible angles between -180° and +180°.

They are clustered in regions in the Ramachandran matrix, with boundaries depending on the choice of the permitted van der waals distances and tetrahedral angles.**(Jan Drenth 1993)**.

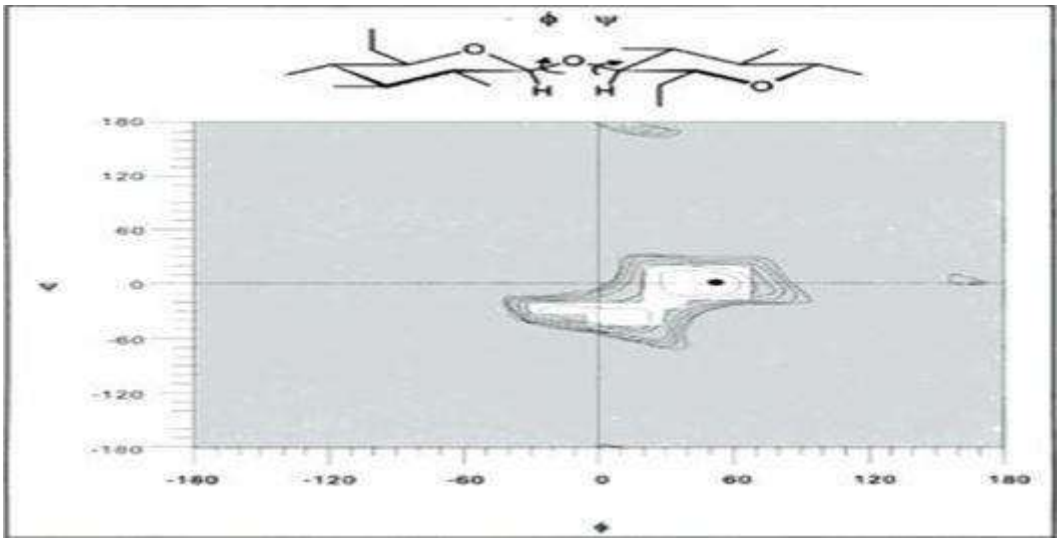


Figure 12: A ramachandran plot for cellulose **(Olav Smidsrød, Størker Moe, Størker T. Moe 2008).**

II.3.3. Tertiary Structure:

Protein tertiary structure is the folding in three-dimension. The regions consisting of secondary structures are folded into a specific compact structure for the entire polypeptide chain as exemplified in Figure bellow. **(Rongkun Shen2006)**

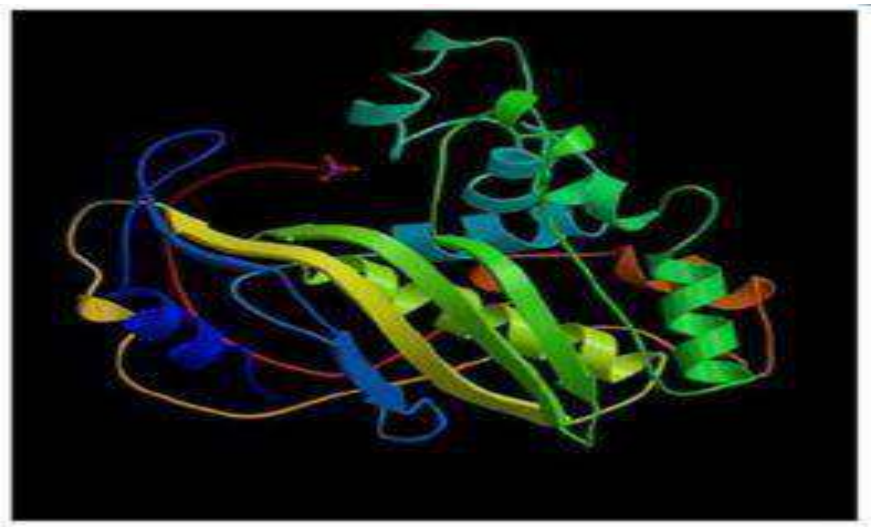


Figure 13: An example of 3-D structure of phage T4 thymidylate synthase. **(Rongkun Shen 2006)**

II.3.4. Quaternary Structure

A protein has quaternary structure with it is an assembly of multiple tertiary structures. We refer to such structures as oligomeric complexes and we describe these structures using labels that start with prefix representing the number of subunits followed by the suffix “mer”. The progression is monomer, dimer, trimer, tetramer, pentamer, ect. If the subunits are different, than an additional “hetero” prefix is used, as an “heterodimer” (Forbes J. Burkowski, 2009) and if the subunits are the same, than the “homo” prefix is used, as an “homotetramer”. **(Forbes J. Burkowski, 2009)**

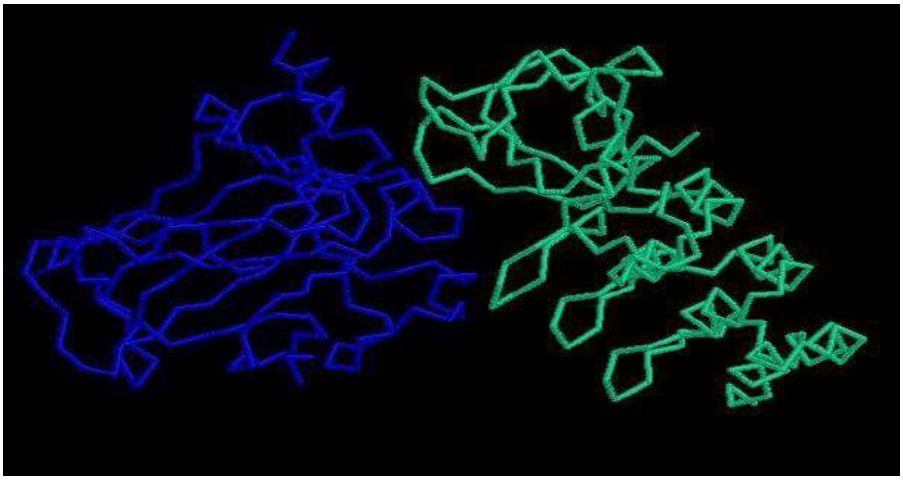


Figure 14: An example of 3-D structure: complex (anti-oncogene/ankyrin repeats) (pdb id: 1ycs)

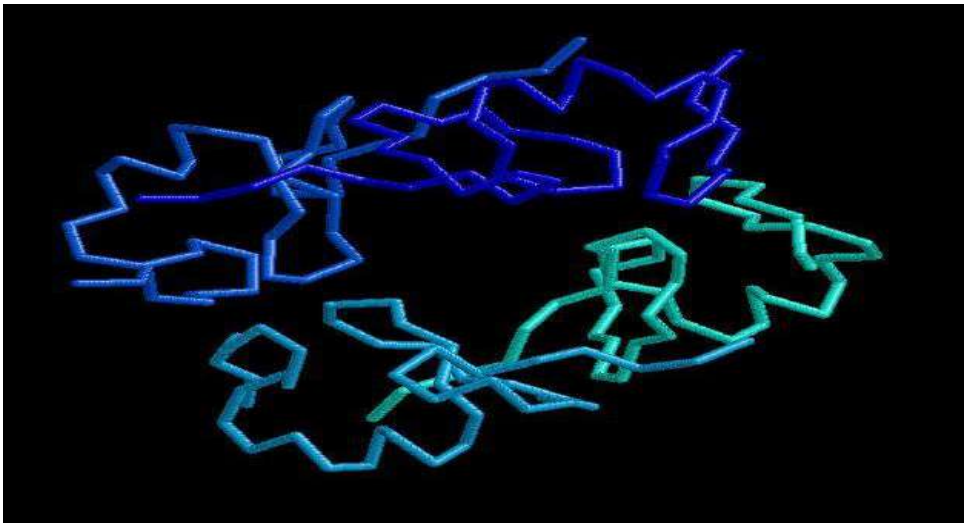


Figure15: An example of 3-D structure: crystal structure of pml ring tetramer (pdb id: 5yuf)

From an efficiency point of view, it is in the best interest of a cell to build complex structures from smaller building blocks that are all the same. Each unit will prescribed by the same DNA sequence, and so there is a parsimonious genetic representation of the full oligomeric structure.

In some cases, the quaternary structure can show a lower level of recursive assembly. For example three heterodimers maybe assembled to form a "trimer of dimers".**(Yolanda Smith, B.Pharm, 2018).**

III.Structural Motifs in proteins:

III. 1. Local Structural Motifs (Helices, Beta Strands & Sheets, Turns & Loops)

IV. 1. A. Alpha-Helices

Table 1: Average conformational parameters of the most commonly found helical secondary structure elements. Types of alpha helices **(*SALEM and MEBARKA, Master Thesis, 2016*)**

Conformation	Phi	Psi	Omega
Alpha helix	-57	-47	180
3-10 helix	-49	-26	180
Pi-helix	57	-70	180

III. 1. B. Beta-Strands & Sheets:

β-Sheets are structural configurations made of more than one β-stand held up by a number forces of which hydrogen bonds play major role. There are many types of β-Sheets:

- Parallel Beta-Sheet.
- Anti Parallel Beta-Sheet.
- Beta Barrel motif
- Parallel beta-Sheet: a beta sheet, formed from noncontiguous regions of thepolypeptide chain, in which every strand runs in the same direction.

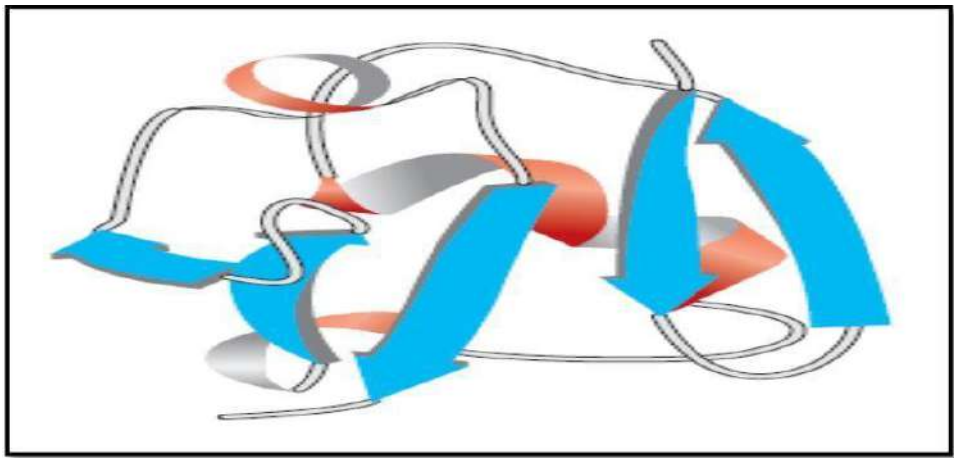


Figure 16: Parallel beta-Sheet.

- Anti parallel beta-Sheet: A beta sheet often formed from contiguous regions of the polypeptide chain, in which each strand runs in the opposite direction from its immediate neighbors **(Gellman S-H, 1998).**

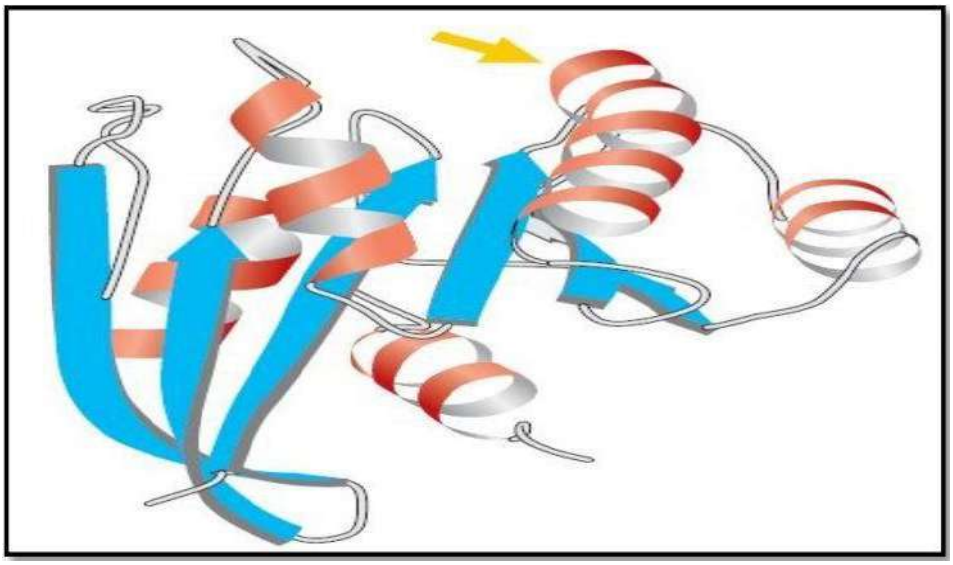


Figure 17: Anti parallel beta-Sheet in blue.

- **Beta Barrel motif:**

  A beta sheet in which the last strand is hydrogen bonded to the first strand, forming a closed cylinder (Gellman S-H, 1998).

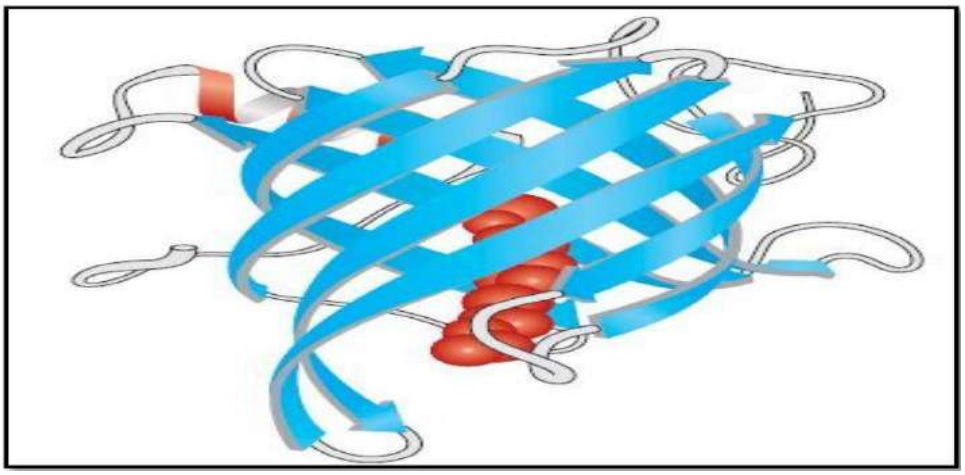


Figure 18: Beta Barrel motif.

III. 1. C. Turns

Turns are specific types of loops as they can be structurally described and classified. In general they can be distinct from loose loops in the cases when the closest Cα atoms are about 7A° apart **(Némethy, George, 1972).**

The following summarize the existing types of turns:

- ✓ α – Turn: a hydrogen bond(s) formed between residues (amino acids) are separated by
- ✓ four residues (x x+4).
- ✓ B – Turn: a hydrogen bond(s) between residues (amino acids) are separated by four
- ✓ residues (x x+3).
- ✓ γ – Turn: a hydrogen bond(s) between residues (amino acids) are separated by four
- ✓ residues (x x+2).
- ✓ π – Turn: a hydrogen bond(s) between residues (amino acids) are separated by four
- ✓ residues (x x+5).
- ✓ Beta Bulges: Beta bulge loops are commonly occurring motifs in proteins and
- ✓ polypeptides consisting of five to six amino acids **(Milner-White, EJ, 1987).**

III. 2. Functional Motifs:

Functional motifs are structural motifs that are usually associated with specific and definite biological function (**Rachedi A, 2013, Tyson and al, 2010**).

It is sequence or structural motif that is always associated with a particular biochemical function. An example is zinc finger motif and the helix-turn-helix.**(Aitken A, 1999).**

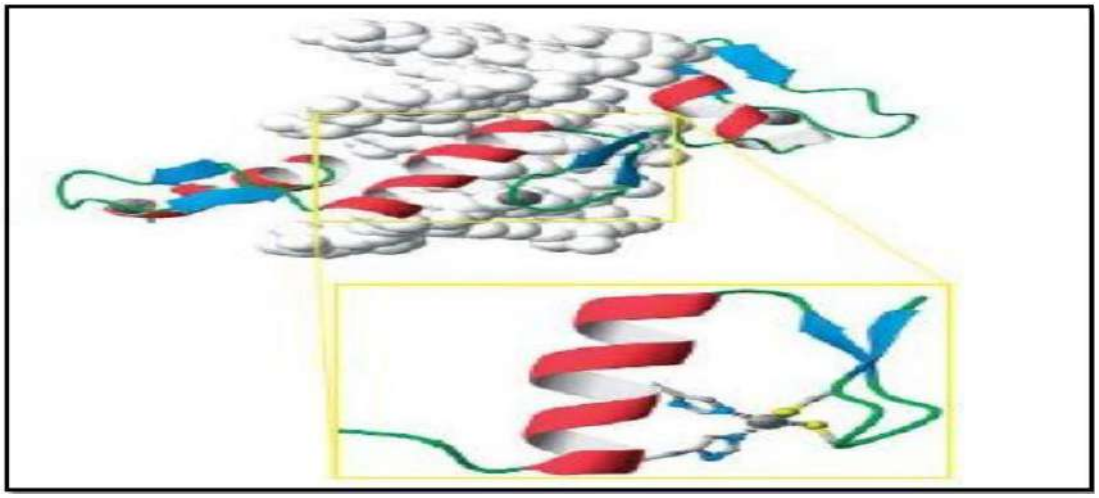


Figure 19: Zinc finger motif a fragment derive from a mouse gene regulatory protein is shown, with three zinc fingers bound spirally in the major groove of a DNA molecule. The inset shows the coordination of a zinc atom by characteristically spaced cysteine and histidine residues in a single zinc finger motif. The image is of Zif268**. (PDB 1aay).**

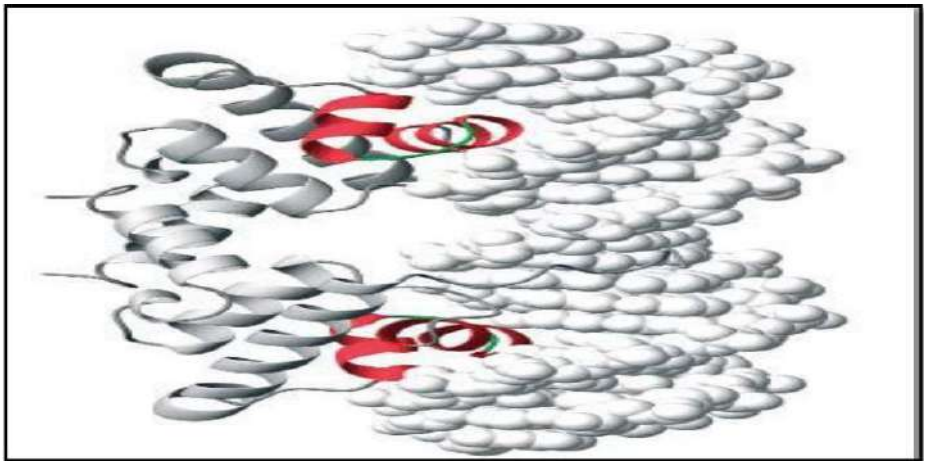


Figure20: Helix-turn-helix The DNA-binding domain of the bacterial gene regulatory protein lambda repressor, with the two helix-turn-helix motifs shown in color. The helices closest to the DNA are the reading or recognition helices, which bind in the major groove and recognize specific gene regulatory sequences in the DNA. **(PDB 1lmb)**

IV. Experimental methods for determining protein structures:

IV.1. X-ray Crystallographic method:

X-ray crystallography is a tool used to determine the 3D position of each atom present in the crystal lattice of the protein crystal. It is the only technique that is being used to solve the structure of the molecule at a resolution of better than1 Å. the major bottleneck in structure determination using X-ray is obtaining an optimum sized protein crystal. The buffers used for protein crystallization mainly consist of a buffering agent, precipitant, and salt. The most widely used precipitants include PEG (of varying molecular weight), ammonium sulfate, and some alcohols which when combined with other additives give various permutations and combination of buffers. For high-throughput crystallization, screening different robotic facilities are also available. The protein molecules in the crystals act as a signal amplifier as they are aligned in a crystal lattice and diffract the X-ray. The diffraction pattern obtained is analyzed for structure factor which is used to build the electron density of atom. The details thus obtained are based on all the complex calculations, probabilities, and assumption, and it needs to be established as the accurate or the closest to the accurate structure by refining the model at several steps. The accuracy of the model obtained after rigorous refinement is measured with regard to the R-value. **(Shailza Singh 2016**).

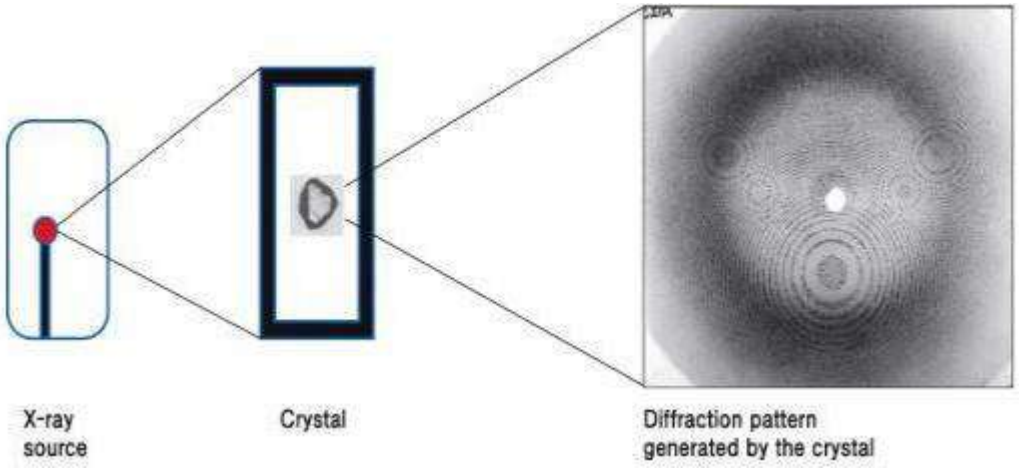


Figure 21: Principle of X-ray crystallography **(T.M.Picknett, S.Brenner,2001)**

IV.2. Cryo-Electron Microscopy Method :

Electrons when accelerated in vaccum are 100.000 times shorter in wavelength than visible range, which makes it possible to resolve the points of few hundred nanometers apart. TEM technique (transmission electron microscopy) uses this principle and has become a versatile tool in studying the protein structure at cryogenic temperature. Cryo-electron microscopy allows the observation of specimens in their native environment unlike X-ray crystallography. A thin film of a sample, aqueous solution, is rapidly frozen on a support grid and then placed in the high vaccum, where it is cooled with liquid nitrogen. Projection images of multiple copies of the molecule in random orientations are recorded, and 3D reconstructions of these images are performed using cryo-electron tomography. Transmission electron cryo microscopy was successful in determining the macromolecular structure considered too complex or large to be resolved by NMR or XRD(X-ray diffractometer).The first protein structure to be solved using electron microscopy was bacteriorhodopsin. Structures at near atomic resolution of viruses, ribosomes, mitochondria, and enzyme complexes has been determined using cryo-electron microscopy. **(Shailza Singh 2016)**

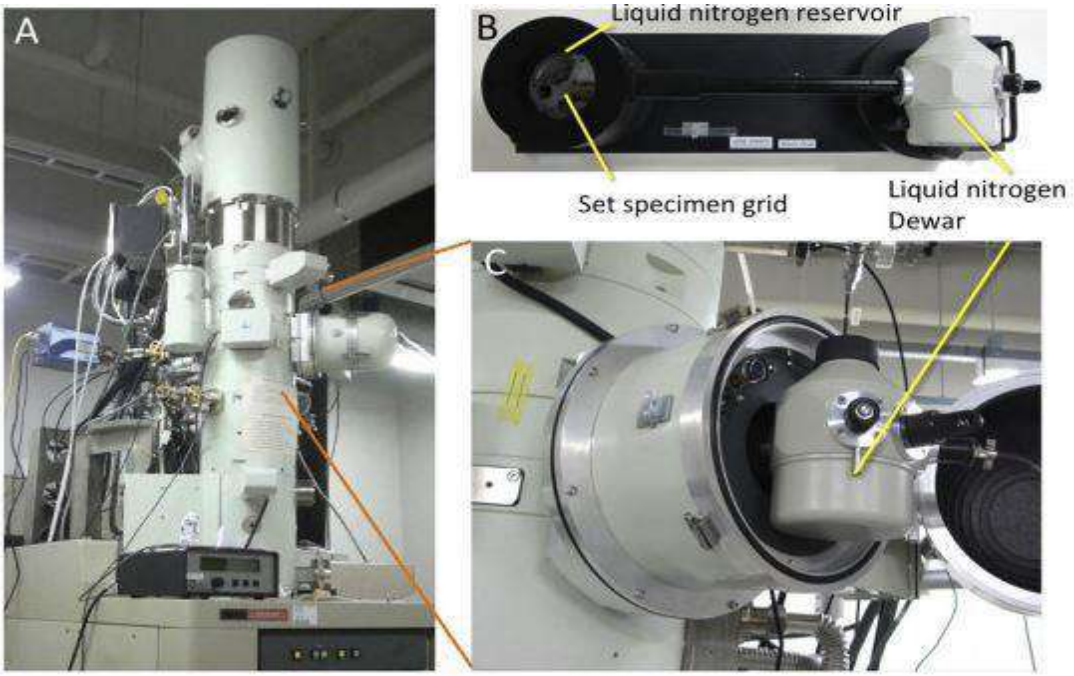


Figure 22: Representative cryo-EM equipment. A) Electron microscope using 200 kV field emission electron source (JEOL 2200FS). B) Side-entry type cryo-transfer specimen holder (Gatan 626). C) Frozen grid is mounted on a cryo-transfer holder and inserted into the EM column without ice-contamination.**(KazuyoshiMurata ,MatthiasWolf, February 2018)**

IV.3. Nuclear magnetic resonance spectroscopy (NMR) Method:

Nuclear magnetic resonance spectroscopy is another technique elucidate the solution structure of proteins. When a solution of labelled protein is placed in a magnetic field and subjected to different radio frequencies, then there is a change in the resonance of different atoms in the proteins. In an externally applied magnetic field, such atoms can flip between two states, against or aligned with the magnetic field. So when the atoms are aligned against the external magnetic field the energy state of the atom is higher and this energy is a function of the rate at which the atoms resonate. This resolution is used to interpret and deduce the structure of the protein. On the basis of atoms selected for labelling. NMR spectroscopy is commonly of two types:

1H NMR (to determine the type and number of H atoms in molecule)

13C NMR (determine the type of carbon atom in molecule)

NMR spectroscopy is better to determine the structure of proteins in the size ranging from 5 to 25 kDa by identifying carbon hydrogen frameworks within molecules **(Shailza Singh 2016.)**

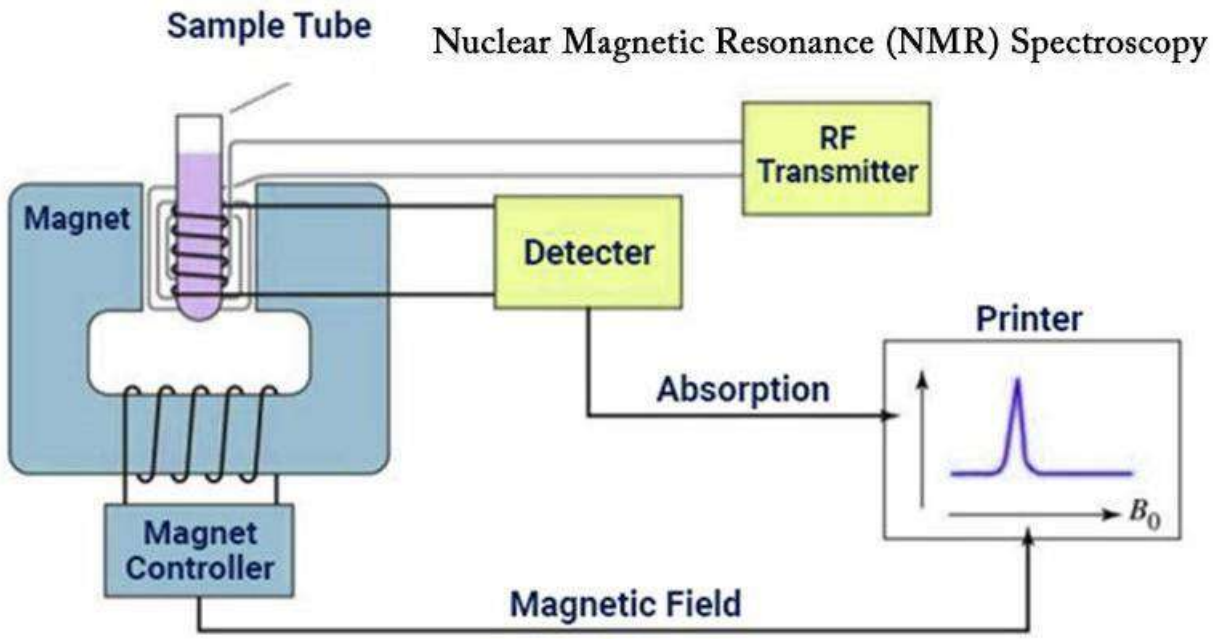


Figure 23: Principle of Nuclear Magnetic Resonance (NMR) Spectroscopy

(SAGAR ARYAL, november 2018)

V. Protein data Bank (PDB):

V. 1. PDB:

Biologists and biochimistes use sequence databases, structure databases, literature databases, etc. The databases we will learn here is called the Protein databases (PDB). The PDB has all known 3D structures of proteins, DNAs and RNAs. To find the PDB on the web,type 'PDB' into Google, and go to the first link returned, which is: http://www.rcsb.org/pdb/home/home.do

You need to download the Protein structures (i.e., the PDB files) that you are going to study, to your own computer. Each structure is in a PDB file with a name that does not carry much information. A PDB file is a simple text file with the(x; y; z) coordinates of all the atoms in the Protein **(Berman H-M and all, 2000).**

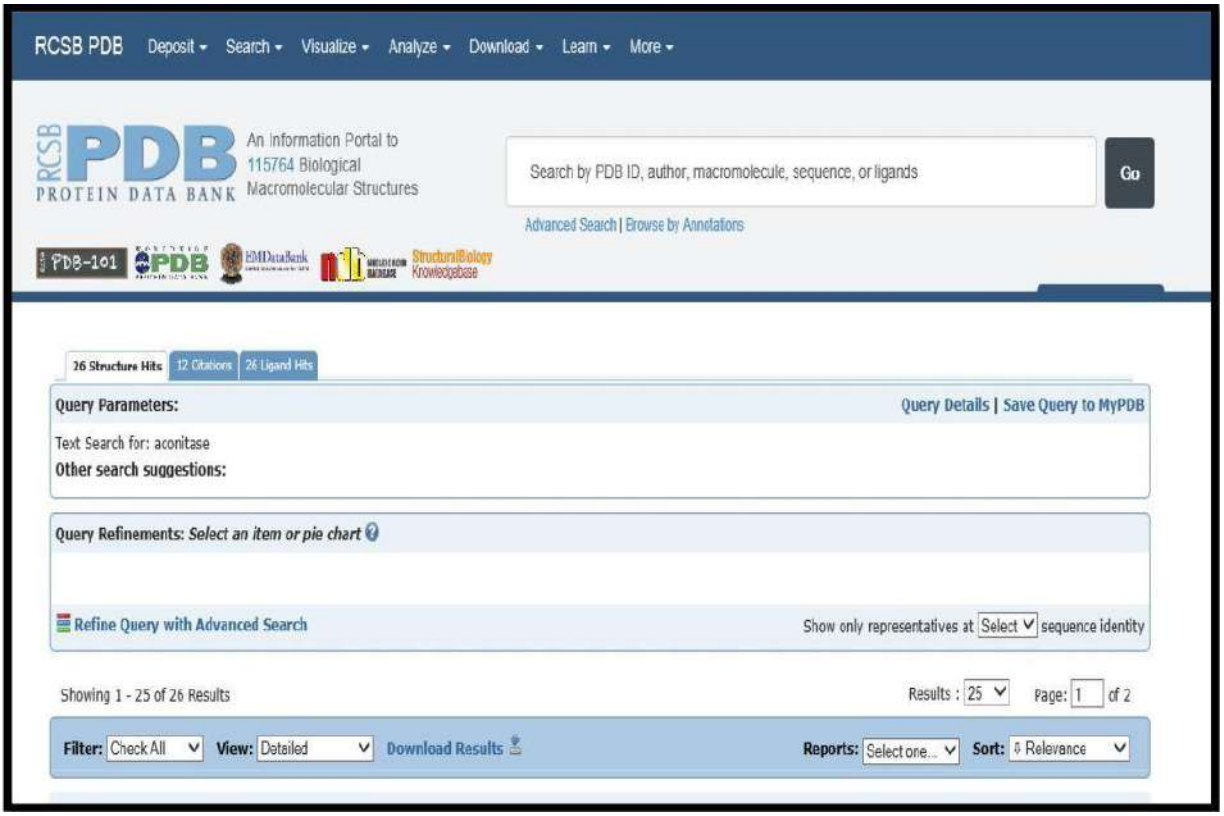


Figure 24: PDB – Protein Data Bank Page.

V. 2. Members PDB:

The RCSB PDB is a member of the wwPDB, a collaborative effort with PDBe (UK), PDBj (Japan), and BMRB (USA) to ensure the PDB archive is global and uniform.

As the wwPDB archive keeper, the RCSB PDB updates the PDB archive at ftp://ftp.wwpdb.org weekly. The structures included in each release are highlighted on the RCSB PDB home page and clearly defined on the FTP site. These sites are maintained 24 hours a day, seven days a week. A failover system automatically (Berman H-M and al, 2003).

Figure25: Members PDB.

V. 2. 1. Protein Data Bank (RCSB):

Simple and advanced searching for macromolecules and Ligands, tabular reports, specialized visualization tools, sequence-structure comparisons, RCSB PDB Mobile, Molecule of the Month and other educational resources at PDB-101, and more.

V. 2. 2. Biological Magnetic Resonance Bank (BMRB):

Collects NMR data from any experiment and captures assigned chemical shifts, coupling constants, and peak lists for a variety of macromolecules; contains derived annotations such as hydrogen exchange rates, pKa values, and relaxation parameters.

V. 2. 3. Protein Data Bank Europe (PDBe):

Rich information about all PDB entries, multiple search and browse facilities, advanced services including PDBePISA, PDBe Fold and PDBe Motif, advanced visualisation and validation of NMR and EM structures, tools for bioinformaticians.

V. 2. 4. Protein Data Bank Japan (PDBj):

Supports browsing in multiple languages such as Japanese, Chinese, and Korean; identifies functionally or evolutionarily conserved motifs by locating and annotating sequence and structural similarities, tools for bioinformaticians, and more **(Berman H-M and al, 2003).**

VI. The other protein structures of banks:

The classification and comparison of the more than 50'000 protein structures deposited in the PDB (Berman H-M and al) is an essential step to extract valuable knowledge from protein structure data **(Berman H-M and al, 2000).**

Today, the two most prominent protein structure classification schemes are SCOP **(Andreeva A and dell, 2008)** and CATH **(Greene L-H and al, 2007)** both partition proteins into domains. These domains are classified in a hierarchical manner:

VI.1. CATH database:

CATH is an acronym for the four main levels in the database hierarchy: Class (C), Architecture (A), Topology (T), and Homologous superfamily (H). There is also a fifth level, sequence family(S). Classification is carried out using sequence alignment methods, the structure comparison algorithm SSAP and human intervention where the automatic processes fail. An entry is assigned a number that correlates to its classification at each level. CATH is accessible on line at http://www.biochem.ucl.ac.uk/bsm/cath/. Users may search by the PDB code, a CATH number or text. **(David S.Moss, Sibila Jelaska, Sandor Pongor 2005)**

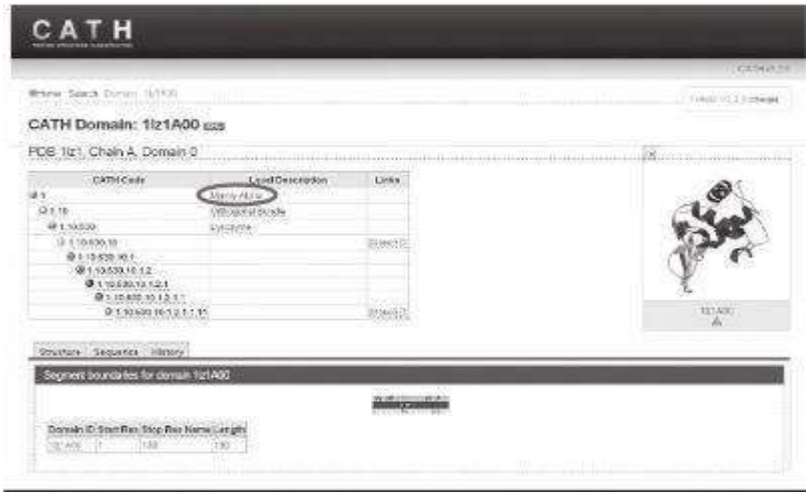


Figure 26: Structural classification of human lysozyme with CATH (**M. Michael Gromiha 2010**)

VI.2. The SCOP database:

The structural classification of proteins (SCOP) database provides a detailed description of the structural and evolutionary relationships of proteins of known structure and is accessible on the world wide web at **http://scop.mrc-lmb.cam.ac.uk/scop/**. There are two search facilities. One allows the user to enter a keyword to match text in the SCOP database and headers in the Protein Databank (PDB). SCOP protein classification is a mainly manual process using visual inspection to compare structures but it also employs sequence homology and a variety of automated procedures. The unit of classification is the domain, each being treated separately in multi-domain proteins.**(David S.Moss, Sibila Jelaska, Sandor Pongor 2005)**

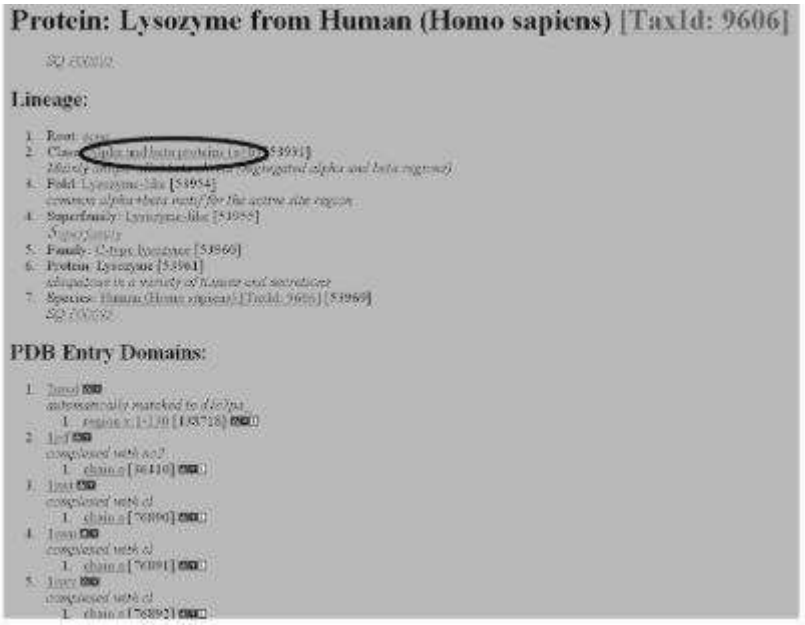


Figure 27: Sample entry for human lysozyme in SCOP database

(**M. Michael Gromiha 2010**)

VII. Rasmol:

Rasmol is a powerful educational tool for showing the Structure of proteins and smaller molecules. The program reads in molecular co-ordinate tiles and interactively displays the molecule on the screen in a variety of representations and colour schemes. The programme was initially written by Roger Sayle and E. J. Milnerwhite and developed at the University of Edinburgh's Bioconi- puting Research Unit and the Biomolecular Structure Department at Glaxo Research and Development, Greenford, UK. Information for getting and installing RasMol can be obtained at http://www.umass.edu/ microbio/rasmol/getras.htm.**(J. Tony Pembroke 2000)**

VIII. Protein -protein interactions:

Many binary protein-protein interactions comes from proteins within the same complex or from proteins between two interacting complexes, where a complex is a stoichiometrically stable set of proteins that permanently associate with each other to play its cellular role as a single unit.

Proteins do not function in isolation. They physically interact with each other or small molecules (ligands) to mediate biological processes or pathways. The interactions happen when the surface patches of the proteins or ligands complement to each other and form a number of non-covalent bonds such as hydrogen bond, ionic interactions, Van der Waal's forces, and hydrophobic packing. Therefore, protein structure, esp. its complementarity with other surface patch, plays an important role in facilitating protein-protein interactions. Those contacting surface patches, i.e. protein-protein interaction sites would be an important target when designing a drug to disrupt the interaction. **(Haidong Wang December 2008)**

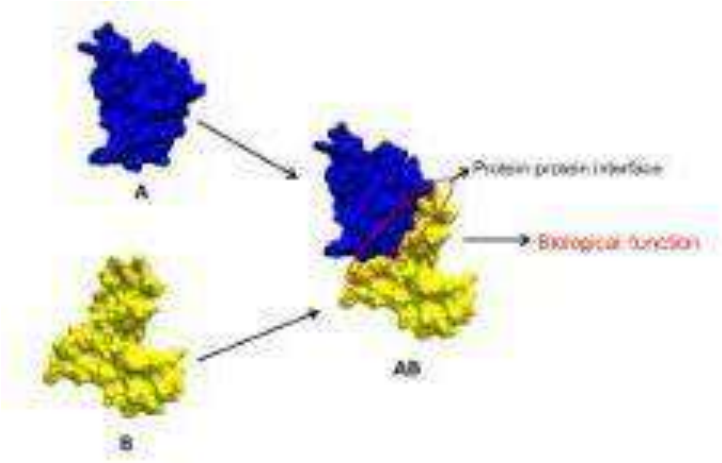


Figure 28: Protein-protein interaction

IX. Non-covalent bonds:

IX.1. Hydrogen Bond, HB:

It is a particular dipole-dipole interaction that takes place between an electronegative atom and a hydrogen atom linked to other electronegative atom or charge acceptor group. Usually, the HB is explained in function of the electron-donor or proton acceptor capability of involved groups. Based on this, it can be said that the hydrogen bond consists of a kind of dipole-dipole interaction between a functional group A (hydrogen acceptor) and an atom or atom groups DH (hydrogen donor), so that both A and D must have certain electronegative character. HB are directional and reversible interactions, whose force is variable (Bonding energy = 1- 30 kcal· mol - 1) and very dependent of the electronegative character of the involved atoms and their environment (bond distances and angles with the hydrogen atom).

The hydrogen bonding can be appearing in different conformations that can influence their strength. The most common can be observed in the Figure bellow and would be:

- ✓ Simple
- ✓ Bifurcated
- ✓ Trifurcated
- ✓ Bridge Cyclic
- ✓ Cyclic dimer

The hydrogen bond has a very important electrostatic component, but there are numerous cases where the hydrogen bonding has a very significant covalent nature. Very strong interactions take place when the donor and acceptor groups are highly electronegative and both have the same electronegativity; therefore the hydrogen atom is practically located halfway. Strong hydrogen bonds character is mainly electrostatic in nature and occurs when both the donor and acceptor systems are hard bases of very electronegative character (N, O, F). The weak hydrogen bond takes place when the donor and acceptor groups are not so strong bases and have a marked van der Waals character, although the electrostatic component that follows is predominant Two examples of weak hydrogen bond interactions with marked van der Waals character are the C-H···X and X-H···π contacts **(Carolina Estarellas Martín 2012)**

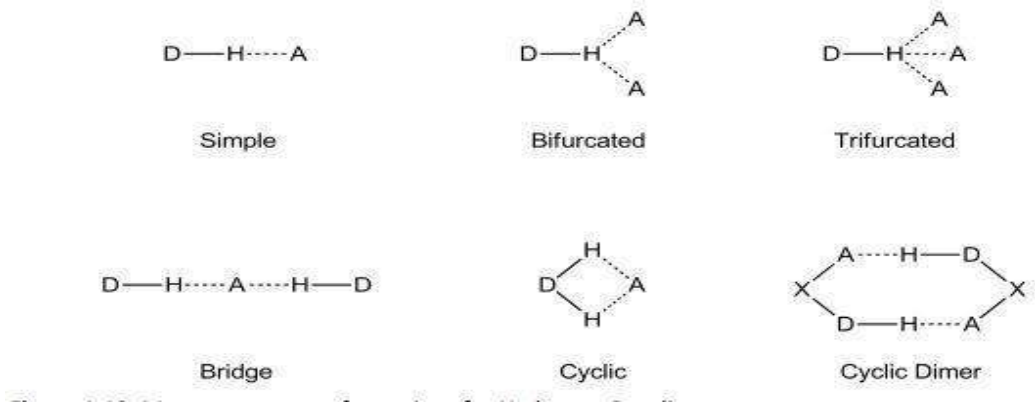


Figure 29: Most common conformations for Hydrogen-Bonding.

The hydrogen bond has a great importance in biological systems. For example, it plays a fundamental role in the stabilization and formation of tridimensional structures of proteins and nucleic acids. In these biological macromolecules, the coupling between different parts of the same macromolecule origins a specific structure that determines the biological and physiological role **(Carolina Estarellas Martín 2012)**

IX.2. Ionic Interactions:

IX.2.1. Ion-Ion Interaction

They are attractive interactions of electrostatic nature that occur between ions of different charge sign. The strength of these interactions are comparable to covalent bond. (Bonding energy = 25-85 kcal·**mol** - **1**). An example of a supramolecular system characterized by this interaction is observed in Figure 1.7. The interaction between the ligand tris(diazabicycleoctane) +3 and the anion [Fe(CN)6] - **3** is only observed in the solid state, since in solution solvation effects dominate and the complex is not formed.**(Carolina Estarellas Martín 2012)**

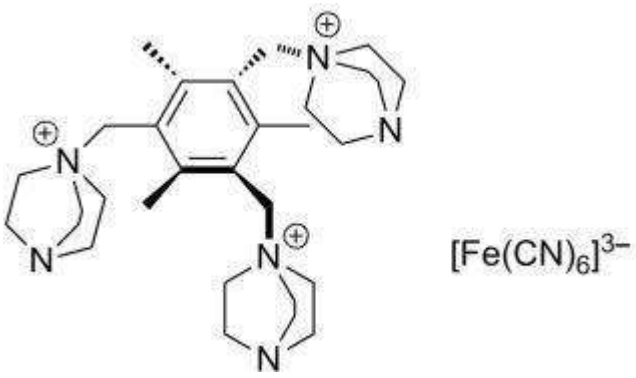


Figure 30: Ion-ion interaction between tris(diazabicycleoctane) +2 and [Fe(CN)6] +2

IX.2.2. Ion-Dipole Interaction:

This interaction takes place between an ion and a polar molecule. It exists in solid state and in solution and is weaker than the ion-ion interaction (bonding energy = 10-50 kcal·**mol** - **1**). A classic model is the solvation process of an ion, as for example, Na+ cation in water. The complexes formed by crown ethers and alkaline ions are other examples of ion-dipole interactions (see Figure31). This complex has a marked ionic character due to the interaction between a small polarising cation and the lone pairs of the oxygen atoms.**(Carolina Estarellas Martín 2012)**

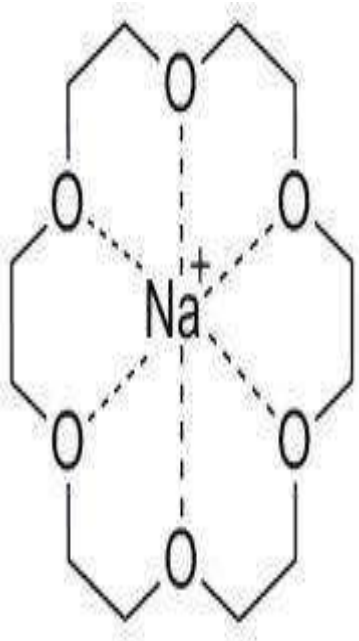


Figure 31: Crown ether within Na+ cation in its interior.

IX.3. Dipole-Dipole Interaction:

These types of interaction are of attractive nature due the electrostatic nature between dipoles, due to the alignment of the opposite poles of both. They are weak interactions (Bonding energy = 1-10 kcal·mol-1), especially in solution (see Figure32).**(Carolina Estarellas Martín 2012)**

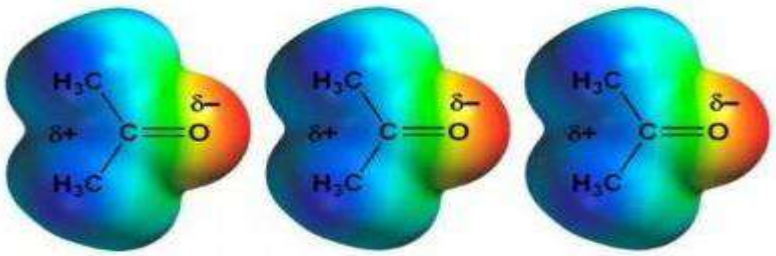


Figure 32: Dipole-dipole interactions in carbonylic compounds.

IX.4. Van der Waal's forces:

The electrostatic and inductive forces are of great importance when some of the molecules possess a dipolar permanent moment. However, other forces act between the molecules of the system. The combination of other forces, known as van der Waals forces, is responsible for the deviation of the behaviour of gases regarding an ideal system.

The van der Waals forces are generally not additive. Therefore, the force between two molecules is affected by the presence of other nearest molecules, and it is not possible to add all the potential pairs of one molecule to obtain its net interaction energy with other molecules. This absence of additivity is due to the field that emerges from any molecule, that reaches a second molecule in a direct and also indirect way, i.e., by reflection from other molecules that also are polarized by the field of the first one. The van der Waals forces can be repulsive or attractive, and the empiric expression most used is the Lennard-Jones potential present in Equation. **(Carolina Estarellas Martín 2012)**

$$V(r) = 4\varepsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right]$$

Equation 1: The parameter σ is the collision diameter, which is the distance where the energy is zero and is equal to 2–1/6rm (rm is the distance where the potential is minimum). The ε parameter is the depth of the potential well.

As we can see in the Lennard-Jones equation (Equation 1 and Figure33), the van der Waals forces are divided into two contributions. The attractive component comes from the dispersive effects and has the basis in the dominant term of Drude (r-6). However, the repulsive component emerges from the compliance of the exclusion principle of Pauli when the electronic clouds of two atoms interpenetrate, although theoretical arguments do not exist for r–12 term, since the quantum mechanics suggest an exponential law. Due to the previous discussion, it can be concluded that the van der Waals interactions are weak (bond energy < 1 kcal·mol–1) and not directional. These forces have been observed in supramolecular systems, especially in inclusion processes, where generally organic molecules (as solvents), are occluded in the crystalline packing or in the cavities of macrocycles.**(Carolina Estarellas Martín 2012)**

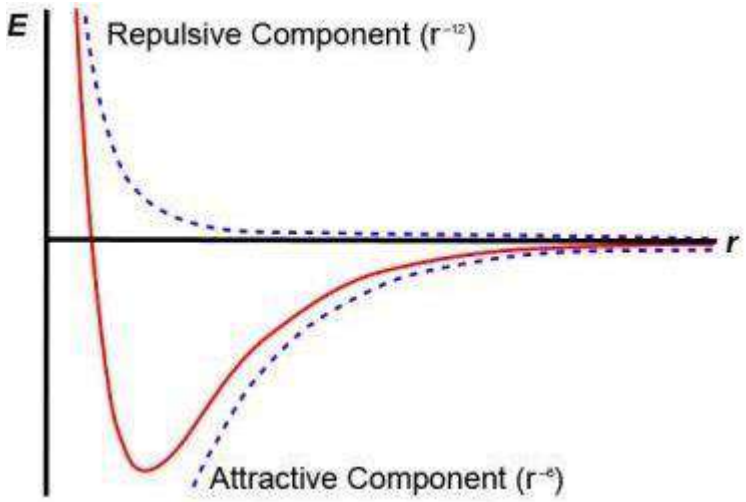


Figure 33: Lennard-Jones Potential Energy.

X. Hydrophobic effect:

As shown in Figure bellow, where the formation of these apolar aggregates results favorably from a thermodynamic point of view.The hydrophobic effect is a property that presents molecules or non-polar molecular fragments that have a tendency to form intermolecular aggregates when is found in a polar media, generally aqueous, hence the name. The hydrophobic effect is very important in biological systems in the creation and maintenance of protein and polynucleotide structure, among other functions. They are of crucial importance in the binding of organic guests by cyclodextrins and cyclophane hosts in water and may be divided into two energetic components: enthalpic and entropic.

The enthalpic hydrophobic effect involves the stabilization of water molecules that are driven from a host cavity upon guest binding. Because host cavities are often hydrophobic, intracavity water does not interact strongly with the host walls and is therefore of high energy. Upon release into the bulk solvent, it is stabilized by interactions with other water molecules. The entropic hydrophobic effect arises from the fact that the presence of two (often organic) molecules in solution (host and guest) creates “two holes” in the structure of bulk water. Combining host and guest to form a complex results in less disruption to the solvent structure and hence an entropic gain (resulting in a lowering of overall free energy). The process is represented schematically in Figure bellow, where the formation of these apolar aggregates results favorably from a thermodynamic point of view**(Carolina Estarellas Martín 2012).**

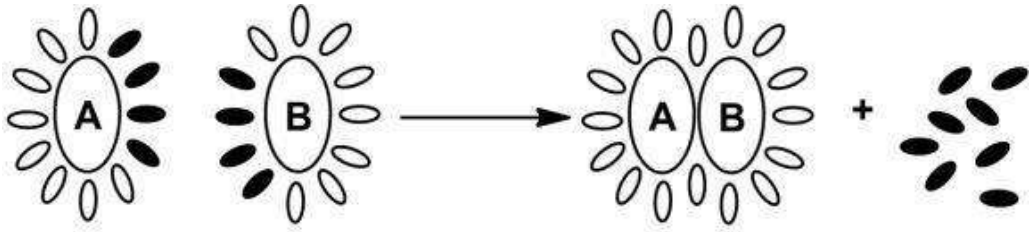


Figure34: Release of solvent molecules by association of two molecules of solute A and B.

XI. Structure-function relationship :

Proteins play an important role in many crucial biological processes and functions. They are very versatile and have many different functions in the body, as listed below:

Act as catalysts.

- ✓ Transport other molecules.
- ✓ Store other molecules.
- ✓ Provide mechanical support.
- ✓ Provide immune protection.
- ✓ Generate movement.
- ✓ Transmit nerve impulses.
- ✓ Control cell growth and differentiation.

The extent to which the structure of proteins has an impact on their function is shown by the effect of changes in the structure of a protein. Any change to a protein at any structural level, including slight changes in the folding and shape of the protein, may render it non-functional. **(Yolanda Smith, 2018)**

The Second Chapter: Materials and Methods

Objectives:

The main objective of this study is to try and discover some of the basics of the relationship between the structure and function in protein complexes. The related objectives include the following points:

1. Select contact interface regions between the units/chains making the quaternary structures (complex structures).
2. Calculate the bonds between the inter-chains of the selected complex structures
3. Highlight the major types of bonds found between the residues composing the contact interfaces.
4. Define, when possible, structural motifs characterizing the contact interfaces and explore their properties.

Material: Data Preparation & Programing Tools:

A number of software tools were implemented in the realization of this project included the following:

1. Microsoft Excel 2010 for data formating and preliminary storing.
2. Notpad++ v.7.5.9 for programming codes writing and editting.
3. XAMPP server vesion 3.1.0.3.1.9 for MySql in-house database creation.
4. PhpMyAdmin for PHP scripts interpretation and MySql in-house database handling, storing data and results retrieving
5. Rasmol v.2.7.5.2 for molecular graphics visualization and molecular scenes generation.

Software:

1. Microsoft Excel 2010:

Microsoft excel 2010 comes within Microsoft Windows software installed on own Personal Computer as part of the Microsoft Office software tools. This Excel version has been used store and format the various data generated by this project. All of the Excel files were formatted in the Comma-Separated Values of CSV format for easy handling by the PHP programming scripts, see Figure35.

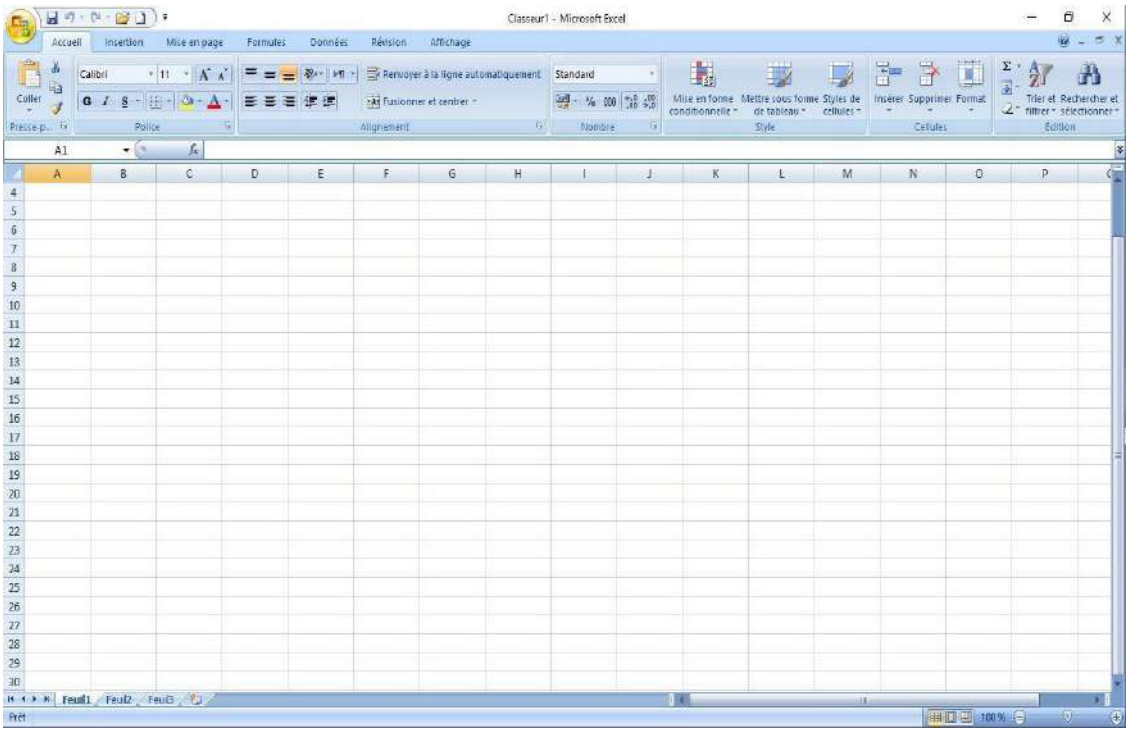


Figure 35: shows Excel's interface

2. XAMPP installation:

XAMPP is an easy to install Apache distribution containing MySql Database, PHP code interpretor, and Apache server platform. A free software version 3.1.0 3.1.0 was download and installed.

The XAMPP server is run in the background of the computer processes to be able to provide the framework of the tools mentioned above, see Figure36.

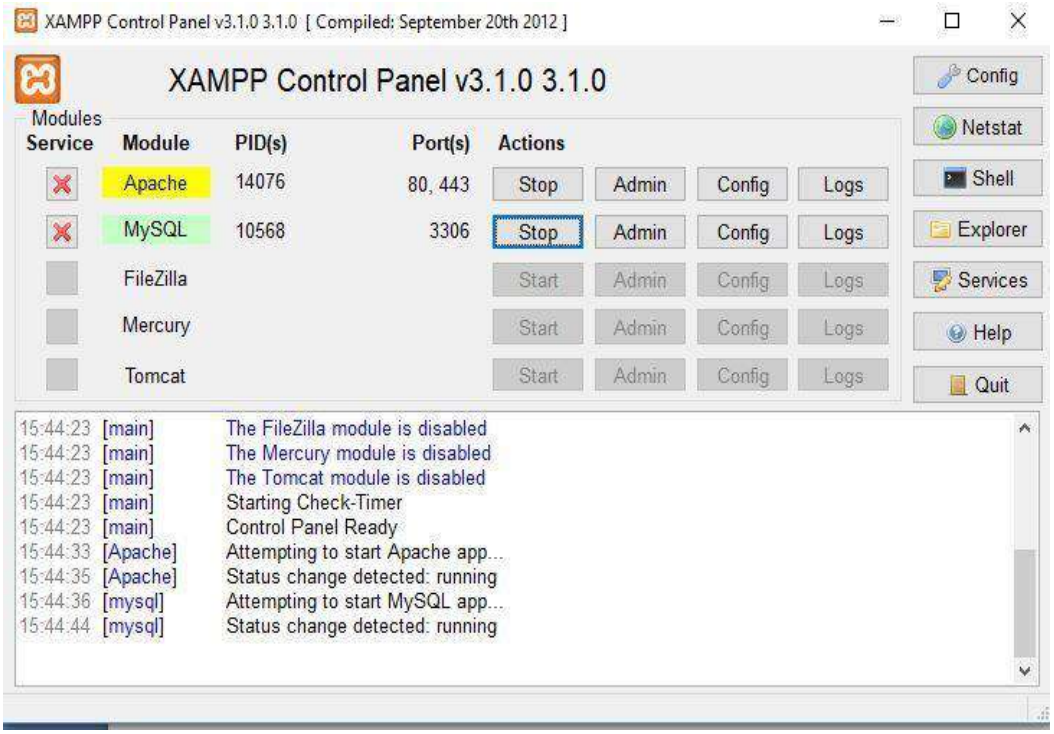


Figure 36: XAMPP started. The figure show that two essential processes are activated, the Apache server and MySql database necessary for the in-house database creation and PHP coding tasks.

3. Notepad++ Installation:

A free version of NotePad++ v.7.5.9. tool was downloaded from internet and installed. This tools has been used as the primary editor for creating the various PHP codes used in this project, see Figure37.

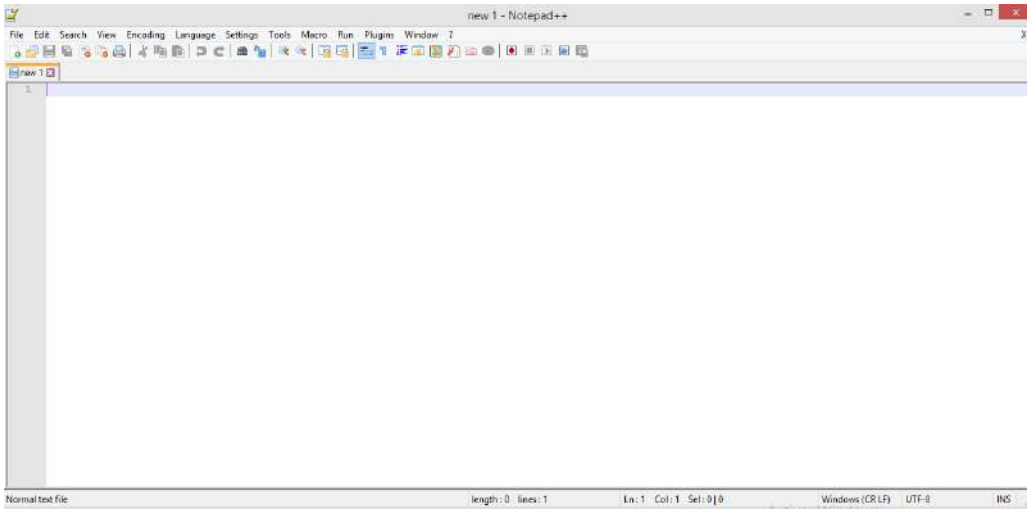


Figure 37: Note pad++ installed and run.

Methods: Data collection & Data Preparation:

II-1-A- PDB & SSFS

.

The Protein Data Bank (PDB) resource can be accessed via the public available address http://www.rcsb.org/pdb/. At the home page one can interrogate the database using the PDB ID of the molecule (if known) or utilising the Search Lite facility using simple keywords. See Figure 1.

The Sequence Structure Function Server tool – SSFS **(A. Rachedi, 2011),** refer to Figure 2**,** has also been used for data collection. Both resources constituted the main source of structures from which the chosen list of the mainly protein three-dimensional structures were download.

1) The Protein Data Bank - PDB

The PDB database is made available to public world wide via the the internet address:
https://www.rcsb.org

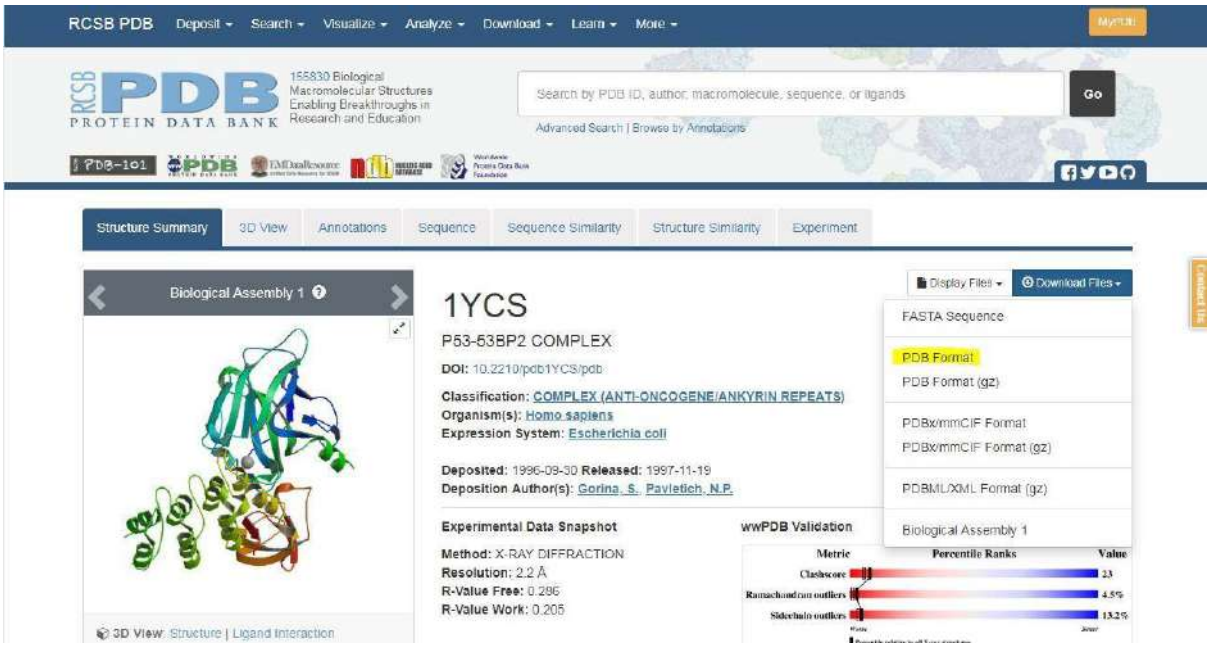


Figure38: The home page of the PDB database.

2) Sequence Structure Function Server tool – SSFS

The SSFS is a bioinformatics tool created at department of Biology, University of Saida, aimed at providing the same type of structural data found in the PDB but with a different approach

concentrating around highlighting the structure-function relationship mainly in the protein structure. It also provides light weight focus results pages as a response of users queries.

Figure39: The home page of the SSFS tool.

Structure Visualisation and Graphics scenes creation:

There are many software tools used to visualize the downloaded structures. Rasmol v2.7.5.2 tool is an easy to use software to explore 3D structures and create graphical scenes of regions of interest.

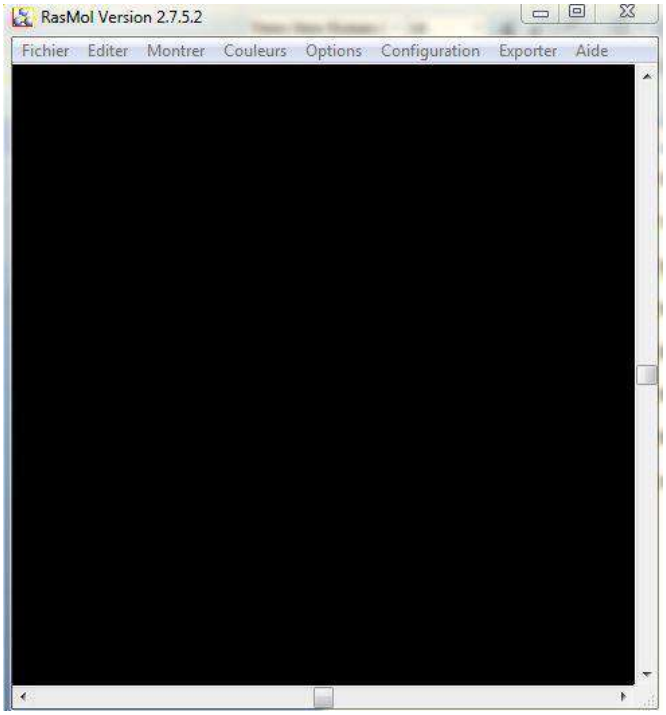
II-1-B- Rasmol (selection of interface contact regions):

To start RasMol under Microsoft Windows, double click on the RasMol icon in the program manager. When RasMol first starts, the program displays a single main window (the display window) with a black background on the screen and provides the command line window minimised as a small icon at the bottom of the screen. The command line or terminal window may be opened by double clicking on this RasMol icon. The Command Line allows the user to define various parts of the bio-molecular structure one wants to interact with and will be used in some of the examples described later.

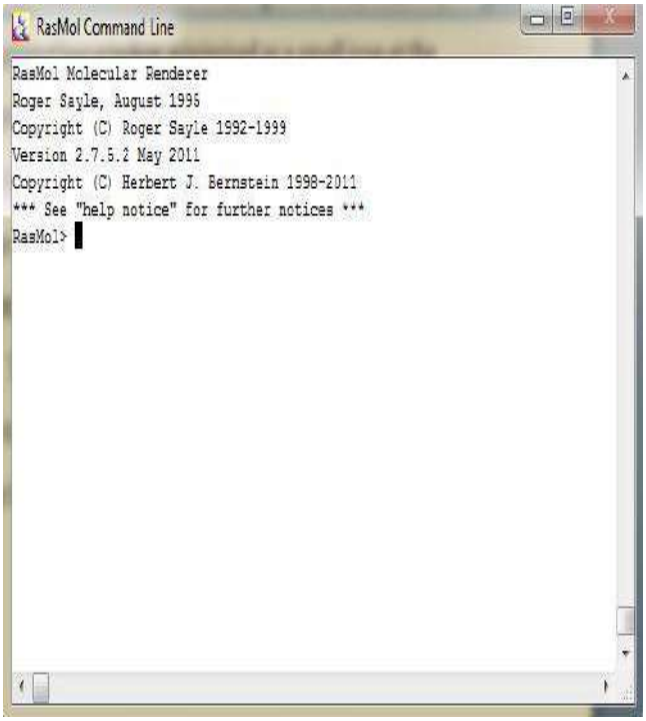
To open a pdb file within RasMol go to the file menu, click open. One can now open any of the PDB files one wishes having saved them from the PDB database.for the purpose of this example this will be the 1ycs.pdb1 file we saved earlier. The molecular structure should now be displayed within RasMol.

1)

The display window



The command line window



2)-

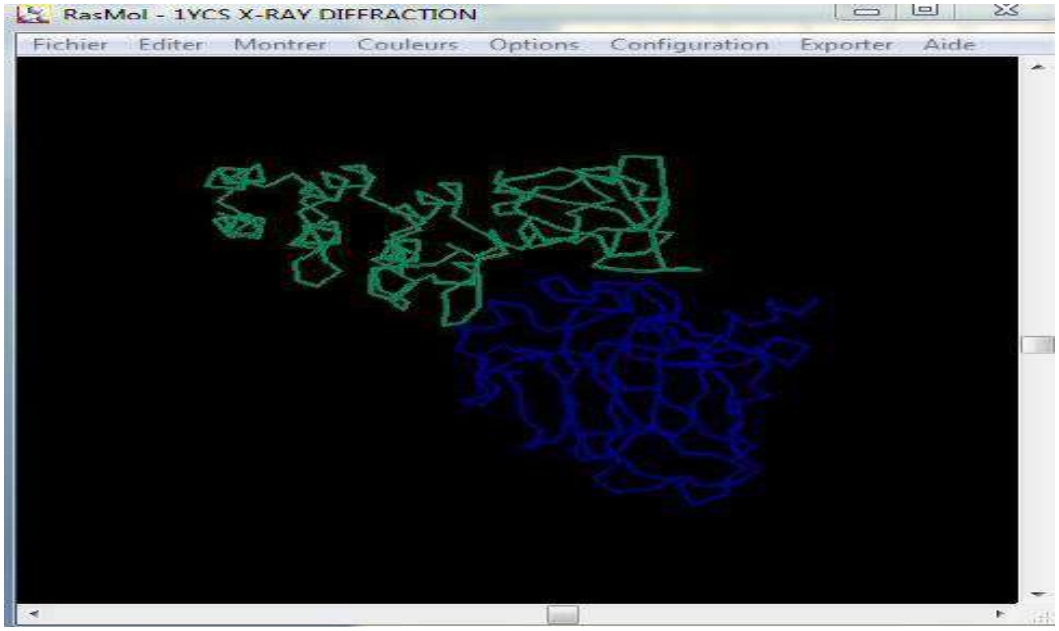


Figure 40: A RasMol generated Chain diagram of tho1ycs.pdh file

1 Holding down the left mouse button rotate the molecule in any and all directions to visualise different parts of the molecule.

2 In the Display Menu alter the molecular display by selecting Space fill, Ribbons, Cartoons, Wireframe or Ball and Stick.

3 Holding down the right mouse bottom allows one to move the molecule to different parts of the screen.

4 Holding down the shift-key and moving the mouse forward zoom-out the display while moving it back blows up the display allowing the viewer to zoom in on the molecule. This can be particularly useful if one selects wire-frame within the Display option and set- ting Labels On within the Options menu. The positions of particular amino acids are clearly labelled and one can zoom into the molecule at these positions.

5 Using ‘backbone’ and ‘colour’ by chains all chains should show up in Rasmol display coloured differently. Follow that with relatively quick examination of the regions in contact.

3) In this case the chosen regions is : 424-427, 470-475,493-498: chainB

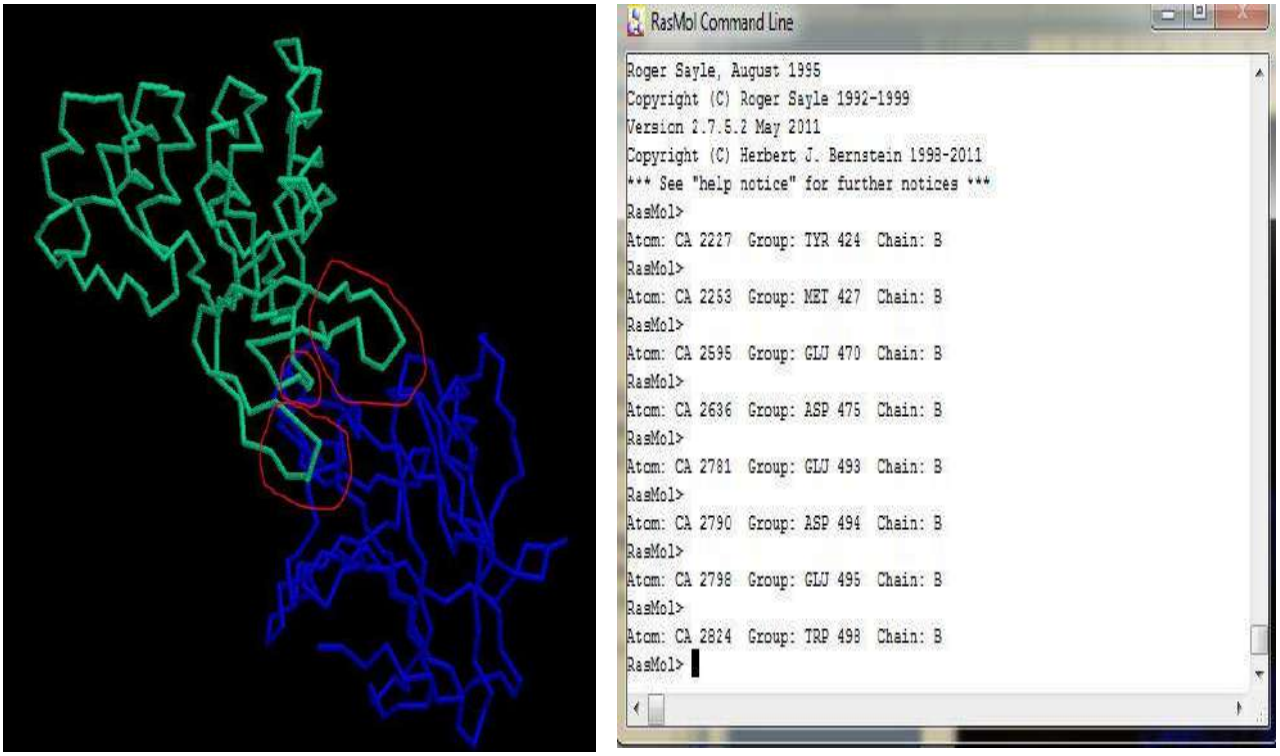


Figure 41: Capture represents the close contact of the selected regions.

II-1-C- Calculation of the Protein-Protein Interactions using the PPI web application:

The Protein-Protein Interactions system (PPI); **http://bioinformaticstools.org/prjs/ppi/** **(A. Rachedi, 2018**). The PPI web tool was used to calculate the inter-chain contacts and provide details of the binding environment existing between the units (chains) of protein complex structures selected in this project.

PPI web tool usage steps:

The following figures (Figure42, Figure43) depict the steps used to calculate the binding environment found between the interface contacts in the protein structure under this study:

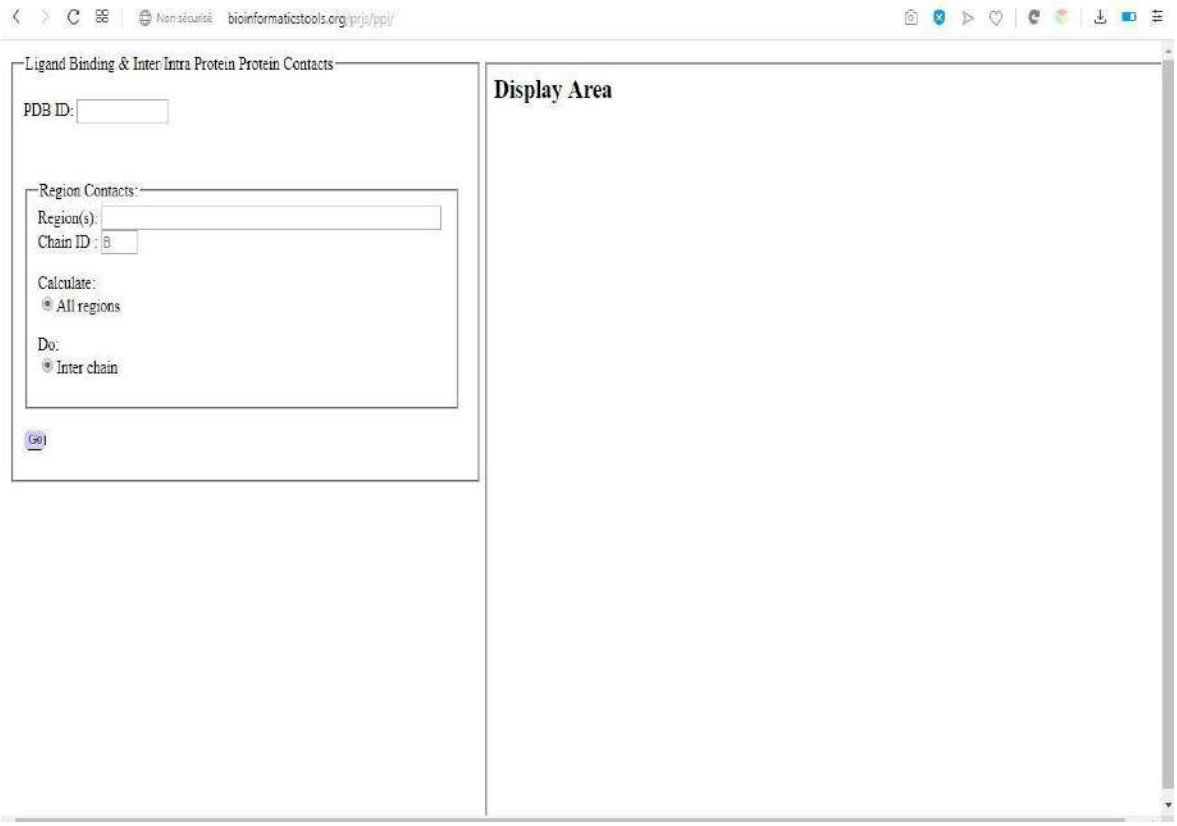


Figure 42: Capture the interface of the PPI site

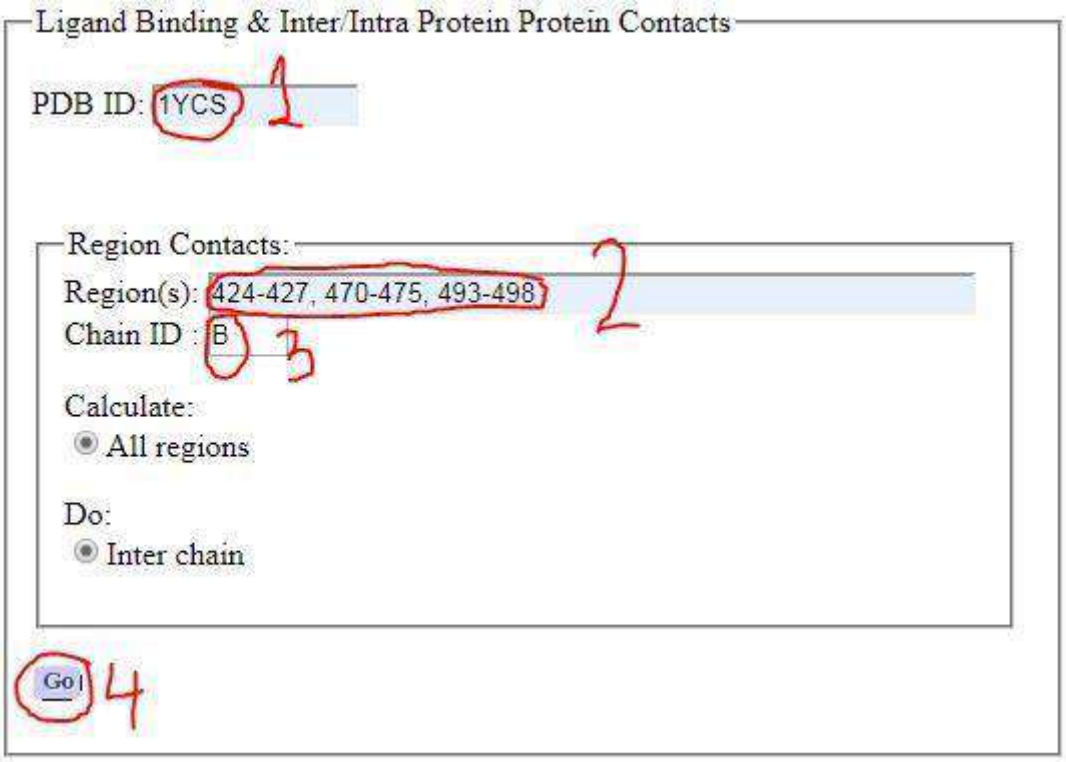


Figure43: the interface of ppi after filling the information

1: the area where we write the "PDBid"

2: the area where we write the closed regions selected

3: the area where we write the chain

4: finaly we click the bottom “Go” to obtain results

Figure 44: represents all information selected

5: we click “environment” to get the final table of contacts data.

Back

Entry: 1ycs	COMPLEX (ANTI-ONCOGENE/ANKYRIN REPEATS)

Protein-Protein Interchain Environment

Protein or NA Residues					Protein or NA Residues					Bonds	
Chain	SSelm	Name	Number	Atom	Chain	SSelm	Name	Number	Atom	Distance/ Å	Possible Bond Type
A	177-181 H: 1	HIS	178	C	B	No SSE	MET	427	SD	3.7	van der Waals
A	177-181 H: 1	HIS	178	O	B	No SSE	MET	427	SD	3.79	
A	177-181 H: 1	HIS	178	CB	B	No SSE	MET	427	SD	3.18	van der Waals
A	177-181 H: 1	HIS	178	CG	B	No SSE	MET	427	SD	3.8	van der Waals
A	177-181 H: 1	HIS	178	CD2	B	No SSE	TYR	424	CD1	3.74	van der Waals
A	177-181 H: 1	HIS	178	CD2	B	No SSE	MET	427	CG	3.83	van der Waals
A	177-181 H: 1	HIS	178	CD2	B	No SSE	MET	427	SD	3.83	van der Waals
A	177-181 H: 1	HIS	178	NE2	B	No SSE	TYR	424	CD1	3.39	
A	177-181 H: 1	HIS	179	CD2	B	No SSE	MET	427	CE	3.4	van der Waals
A	177-181 H: 1	HIS	179	NE2	B	No SSE	MET	427	CE	3.35	

A	177-181 H: 1	ARG	181	CG	B	No SSE	TYR	424	CB	3.56	van der Waals
A	177-181 H: 1	ARG	181	CD	B	No SSE	TYR	424	CB	3.54	van der Waals
A	177-181 H: 1	ARG	181	CD	B	No SSE	TYR	424	CG	3.64	van der Waals
A	177-181 H: 1	ARG	181	CD	B	No SSE	TYR	424	CD2	3.63	van der Waals
A	177-181 H: 1	ARG	181	NE	B	No SSE	TYR	424	CB	3.94	
A	177-181 H: 1	ARG	181	NE	B	No SSE	TYR	424	CG	3.6	
A	177-181 H: 1	ARG	181	NE	B	No SSE	TYR	424	CD1	3.85	
A	177-181 H: 1	ARG	181	NE	B	No SSE	TYR	424	CD2	3.67	
A	177-181 H: 1	ARG	181	NE	B	No SSE	TYR	424	CE2	3.98	
A	No SSE	CYS	182	CB	B	No SSE	MET	427	CG	3.96	van der Waals
A	No SSE	CYS	182	CB	B	No SSE	MET	427	CE	3.84	van der Waals
A	No SSE	SER	241	C	B	No SSE	GLU	497	OE1	3.93	van der Waals
A	No SSE	SER	241	C	B	498-503 S: 0	TRP	498	NE1	3.93	
A	No SSE	SER	241	O	B	No SSE	GLU	497	OE1	3.81	H.Bond
A	No SSE	SER	241	O	B	498-503 S: 0	TRP	498	NE1	3.08	H.Bond

A	No SSE	SER	241	O	B	498-503 S: 0	TRP	498	CE2	3.52	van der Waals
A	No SSE	SER	241	O	B	498-503 S: 0	TRP	498	CZ2	3.36	van der Waals
A	No SSE	SER	241	CB	B	No SSE	GLU	495	OE1	3.65	van der Waals
A	No SSE	SER	241	CB	B	498-503 S: 0	TRP	498	NE1	3.76	
A	No SSE	SER	241	OG	B	No SSE	GLU	497	CD	3.62	van der Waals
A	No SSE	SER	241	OG	B	No SSE	GLU	497	OE1	3.69	H.Bond
A	No SSE	SER	241	OG	B	No SSE	GLU	497	OE2	3.79	H.Bond
A	No SSE	CYS	242	N	B	No SSE	GLU	497	OE1	3.98	H.Bond
A	No SSE	CYS	242	CA	B	No SSE	GLU	497	OE1	3.93	van der Waals
A	No SSE	MET	243	CA	B	498-503 S: 0	TRP	498	CZ2	3.73	van der Waals
A	No SSE	MET	243	CB	B	498-503 S: 0	TRP	498	CZ2	3.54	van der Waals
A	No SSE	MET	243	CB	B	498-503 S: 0	TRP	498	CH2	3.96	van der Waals
A	No SSE	MET	243	CG	B	498-503 S: 0	TRP	498	CH2	3.95	van der Waals
A	No SSE	ASN	247	C	B	No SSE	ASN	473	ND2	4	
A	No SSE	ASN	247	O	B	No SSE	ASN	473	ND2	2.87	H.Bond
A	No SSE	ASN	247	CB	B	No SSE	GLN	472	CG	3.72	van der Waals
A	No SSE	ASN	247	CG	B	No SSE	GLN	472	CG	3.32	van der Waals

A	No SSE	ASN	247	CG	B	498-503 S: 0	TRP	498	CH2	3.79	van der Waals
A	No SSE	ASN	247	OD1	B	No SSE	GLN	472	CG	3.91	van der Waals
A	No SSE	ASN	247	OD1	B	498-503 S: 0	TRP	498	CH2	3.79	van der Waals
A	No SSE	ASN	247	ND2	B	No SSE	GLN	472	CB	3.68	
A	No SSE	ASN	247	ND2	B	No SSE	GLN	472	CG	3.02	
A	No SSE	ASN	247	ND2	B	498-503 S: 0	TRP	498	CZ2	3.77	
A	No SSE	ASN	247	ND2	B	498-503 S: 0	TRP	498	CH2	3.24	
A	No SSE	ARG	248	C	B	No SSE	ASN	473	ND2	3.5	
A	No SSE	ARG	248	O	B	No SSE	ASN	473	CG	3.48	van der Waals
A	No SSE	ARG	248	O	B	No SSE	ASN	473	OD1	3.3	H.Bond
A	No SSE	ARG	248	O	B	No SSE	ASN	473	ND2	3.1	H.Bond
A	No SSE	ARG	248	CB	B	No SSE	ASN	473	ND2	3.69	
A	No SSE	ARG	248	CG	B	498-503 S: 0	TRP	498	NE1	3.76	
A	No SSE	ARG	248	CG	B	498-503 S: 0	TRP	498	CE2	3.81	van der Waals
A	No SSE	ARG	248	CD	B	No SSE	GLU	495	CG	3.61	van der Waals
A	No SSE	ARG	248	CD	B	No SSE	GLU	495	CD	3.43	van der Waals
A	No SSE	ARG	248	CD	B	No SSE	GLU	495	OE1	3.66	van der Waals
A	No SSE	ARG	248	CD	B	No SSE	GLU	495	OE2	3.74	van der Waals

A	No SSE	ARG	248	CD	B	498-503 S: 0	TRP	498	CG	3.72	van der Waals
A	No SSE	ARG	248	CD	B	498-503 S: 0	TRP	498	CD1	3.66	van der Waals
A	No SSE	ARG	248	CD	B	498-503 S: 0	TRP	498	CD2	3.78	van der Waals
A	No SSE	ARG	248	CD	B	498-503 S: 0	TRP	498	NE1	3.68	
A	No SSE	ARG	248	CD	B	498-503 S: 0	TRP	498	CE2	3.75	van der Waals
A	No SSE	ARG	248	CZ	B	No SSE	GLU	495	CG	3.96	van der Waals
A	No SSE	ARG	248	CZ	B	No SSE	GLU	495	OE2	3.83	van der Waals
A	No SSE	ARG	248	NH1	B	No SSE	ASP	494	OD2	3.34	H.Bond
A	No SSE	ARG	248	NH1	B	No SSE	GLU	495	CG	3	
A	No SSE	ARG	248	NH1	B	No SSE	GLU	495	CD	3.09	
A	No SSE	ARG	248	NH1	B	No SSE	GLU	495	OE2	2.69	H.Bond
A	No SSE	ARG	248	NH2	B	No SSE	ASP	475	CG	3.32	
A	No SSE	ARG	248	NH2	B	No SSE	ASP	475	OD1	2.93	H.Bond
A	No SSE	ARG	248	NH2	B	No SSE	ASP	475	OD2	3.07	H.Bond
A	No SSE	ARG	249	CG	B	No SSE	ASN	473	CG	3.86	van der Waals
A	No SSE	ARG	249	CG	B	No SSE	ASN	473	OD1	3.48	van der Waals
A	No SSE	ARG	249	CG	B	No SSE	ASN	473	ND2	3.56	
A	No SSE	ARG	249	NH1	B	No SSE	GLN	472	NE2	3.94	H.Bond

Table 2: Details of binding interface between chains A and B of the dimer complex structure of P53 Tumor Suppressor represented by the PDB entry 1YCS.

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

The binding details shown in the table above and presented in the columns are explained in the following:

- **The set of columns "Protein residues":** These columns show the atoms of the protein residues (AA) that bind with other protein residues. The residues are also annotated in terms of what secondary structure elements (helix, b-sheet or loop) they may belong to.
- **The column "Bonds":** Types of possible bonds and their lengths.
- **File :** to get the contact file in form "EXCEL"

- **RasScene 1:**localized depiction of regions in contact between the moleculs

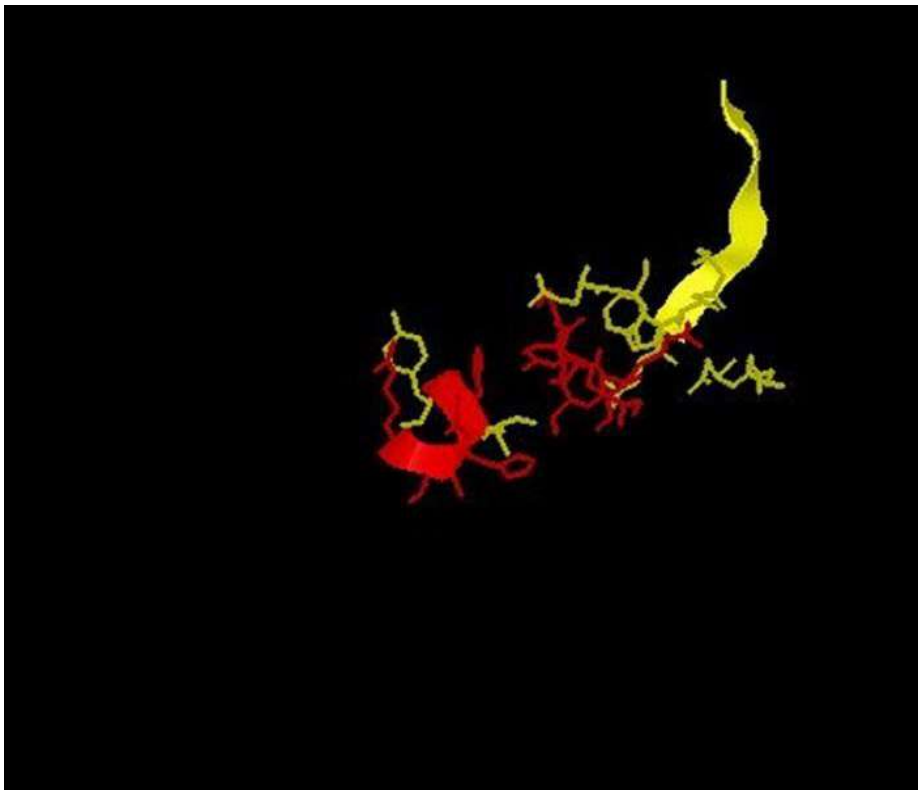


Figure45: represent capture of **RasScene 1**

- **RasScene 2:**localized depiction of regions in contact between the moleculs where amino acids are shown with van-der-walls spheres exposing the close contacts.

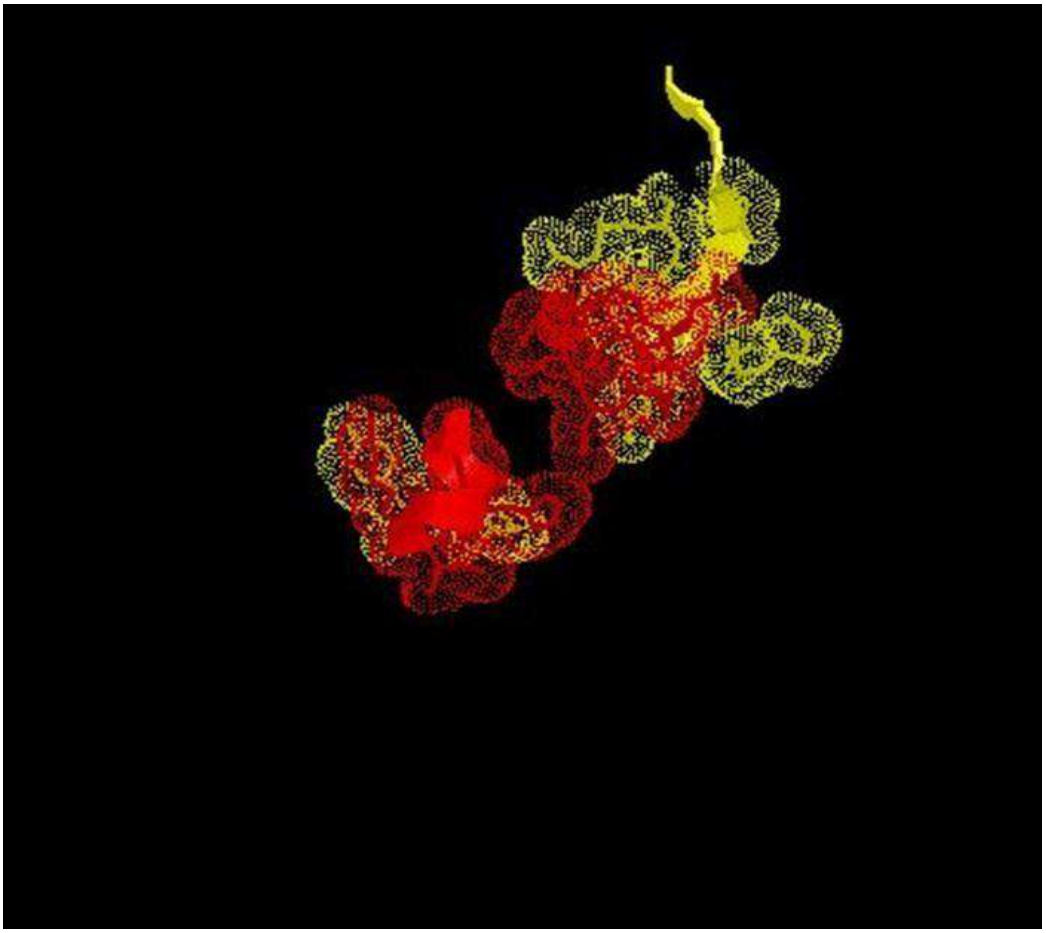


Figure46: Represent capture of **RasScene 2**

- **RasScene 3:**global depiction of regions in contact between the moleculs where chains are colored differently (chain A in red and B in yellow).

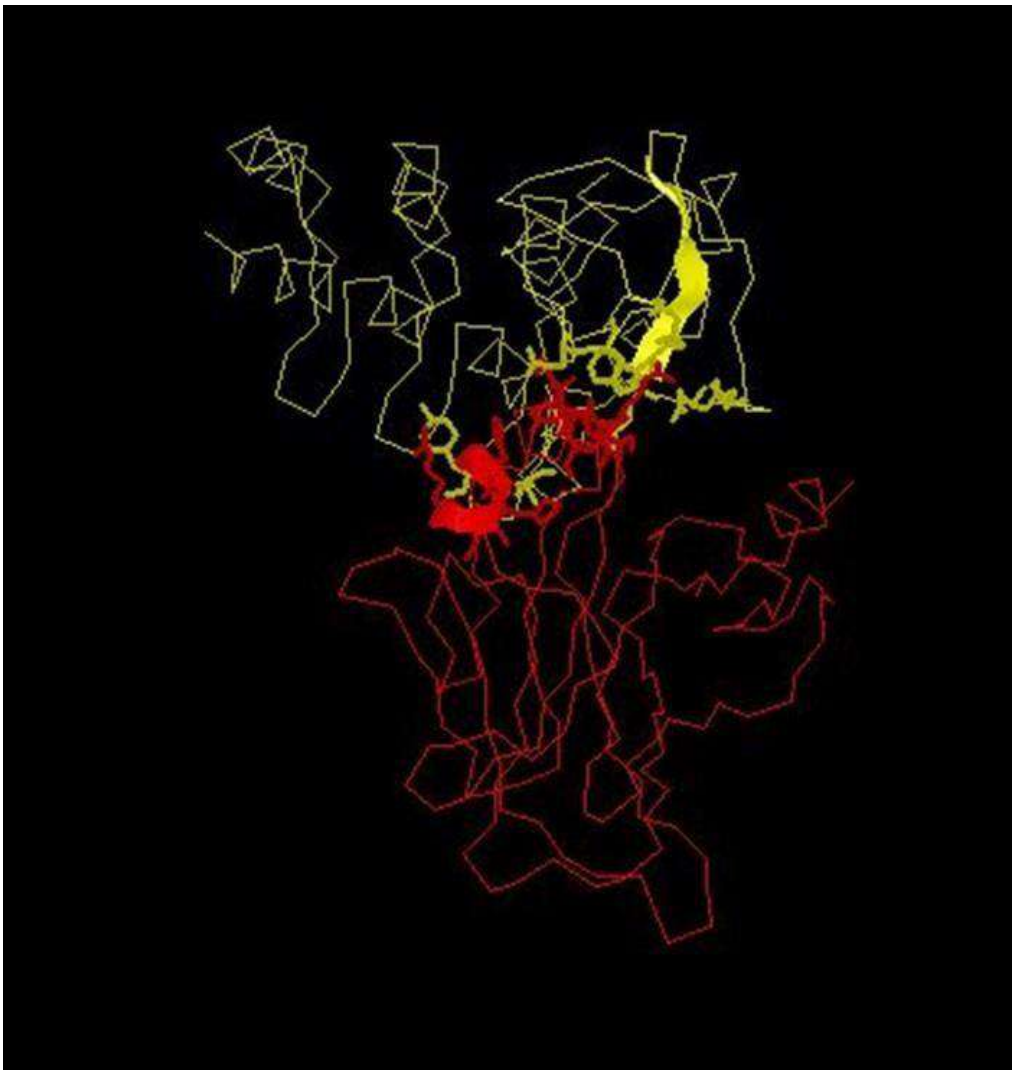


Figure 47: Represents capture of **RasScene 3**

- **RasScene 4:** same as RasScene 3 plus ver-der-walls representation of the AAs in contact

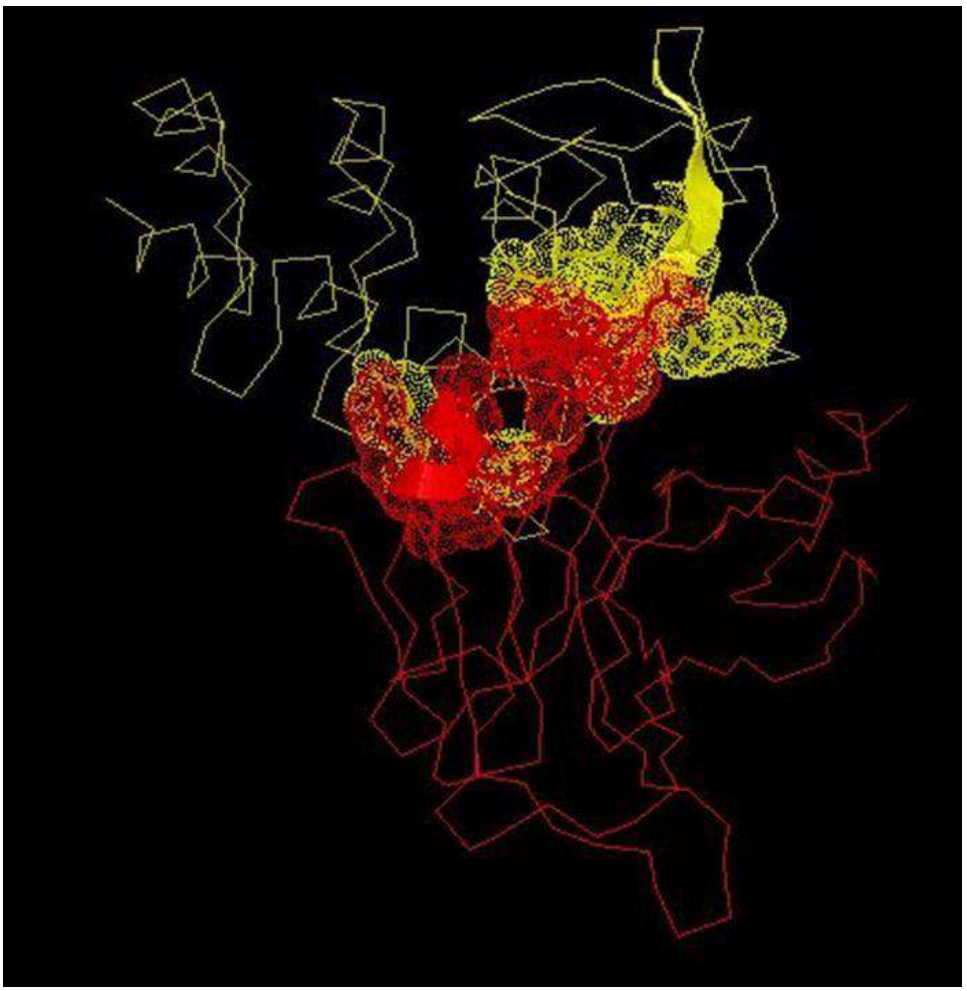


Figure 48: Represent capture of **RasScene 4**

- **RunFile:** An executable text file comtains rasmol script to generate rasmol scenes.

Motifs construction:

The secondary structure elements annotation, found in the table above, are used to create a patterns called motifs to describe the protein-protein interface binding in an abstract manner. The motifs are by default made of two parts since the interface binding, when exist, is between every two chains of the chains making the protein complex as explained via an example shown above in table 2, as follows:

Motif first part:

- (177-181 H: 5): Represents the secondary structure alpha-Helix and is denoted as H.
- (No SSE 182): Represents the lack of secondary structure which means it's a loop region and is denoted as L.

Motif second part

- (No SSE 424): Represents the lack of secondary structure which means it's a loop region and is denoted as L. This denotation is not as precise as the S and H

Definition since these latter have very well defined geometry (values for bond distances, bond angles and tetrahedral angles). The L is more relaxed its geometry which makes it had to give a fixed definition. Moreover, in this project, L could mean a long stretch of amino acids as it could also mean a short number of amino acids that are not bellowing to clear secondary structures elements (H or S).

- (498-503 S: 0): Represents the secondary structure beta-Sheet denoted as S.

The pattern representing the interactions of the PDBid: **1YCS** found in the table above is represented as follows: **HLLL/LLLLLLS** for the chains A and B respectively .

II-2- Main Data

II-2-A- List of protein structures (Master table)

A list of protein structures have been selected from the PDB database including dimmers, trimers, tetramers for achieving the goals set by this study.
TABLE3: some PDBids (the full list of master table in **"index"**)

PDB ID	Title	Determination method	Resolution	R-factor
6A0A	STRUCTURE OF A TRIPLE-HELIX REGION OF HUMAN COLLAGEN TYPE II	X-RAY DIFFRACTION	1.5	17.2
4CGB	CRYSTAL STRUCTURE OF THE TRIMERIZATION DOMAIN OF EML2	X-RAY DIFFRACTION	2.15	20.78
6BFS	THE MECHANISM OF GM-CSF INHIBITION BY HUMAN GM-CSF AUTO-ANTI	X-RAY DIFFRACTION	2	18.5
5OCK	CRYSTAL STRUCTURE OF ACPA E4 IN COMPLEX WITH CEP1	X-RAY DIFFRACTION	1.6	18.5
3HFE	A TRIMERIC FORM OF THE KV7.1 A DOMAIN TAIL	X-RAY DIFFRACTION	1.7	20.1

5YUF	CRYSTAL STRUCTURE OF PML RING TETRAMER	X-RAY DIFFRACTION	1.6	20.4

5LLV	CRYSTAL STRUCTURE OF DACM F87M/L110M TRANSTHYRETIN MUTANT	X-RAY DIFFRACTION	1.7	15.3
2WQI	CRYSTAL STRUCTURE OF THE HUMAN P73 TETRAMERIZATION DOMAIN	X-RAY DIFFRACTION	1.7	21.1
2KEI	REFINED SOLUTION STRUCTURE OF A DIMER OF LAC REPRESSOR DNA-BINDING DOMAIN COMPLEXED TO ITS NATURALOPERATOR O1	NMR	-	0

II-2-B- Quaternary structure types and classification:

The list of protein structures selected, in the above step, are all quaternary structure types which means they are constructed of more than one polypeptide chain. In this project, there are three types of structures classified based on the number of chains that the protein is made of and on the similarity folds of the chains; Dimers, Trimers and Tetramers and each class can be homo or hetero based on the level of folds similarity of the chains.

The class types are said Homo (HM) if the chains have the same fold, that is Homodimer, and are Hetero (HT) when they chains are different fold, that is Heterodimer, as shown in the tables below:

Dimers:

Table4: Partial list of the Dimers (the full list of dimers in **"index"**).

Pdb Id	Chain	Type
1YCS	A	HT
	B	
6MS4	A	HT
	B	
1M4I	A	HM
	B	

1AFW	A	H M
	B	

Trimers

Table5: Partial list of the Trimers (the full list of trimers in **"index"**)

Pdb Id	Chain	Type
3FWT	A	HM
	A	
	A	
5EIL	A	HM
	B	
	C	
5NF8	A	HT
	C	
	B	
2MKC	A	HT
	B	
	C	

TETRAMERS

TABLE6: Partial list of the Tetramers (the full list of tetramers in **"index"**)

PDB Id	Chain	Type
5YUF	A	HM
	B	
	C	
	D	
5LLV	A	HM
	B	
	C	
	D	
1S5Y	A	HM
	B	
	C	
	D	

All of the data where including images for the rasmol scenes where stored systematically in appropriate flat-files structure to be used for the in-house database creation as explained in the steps to follow.

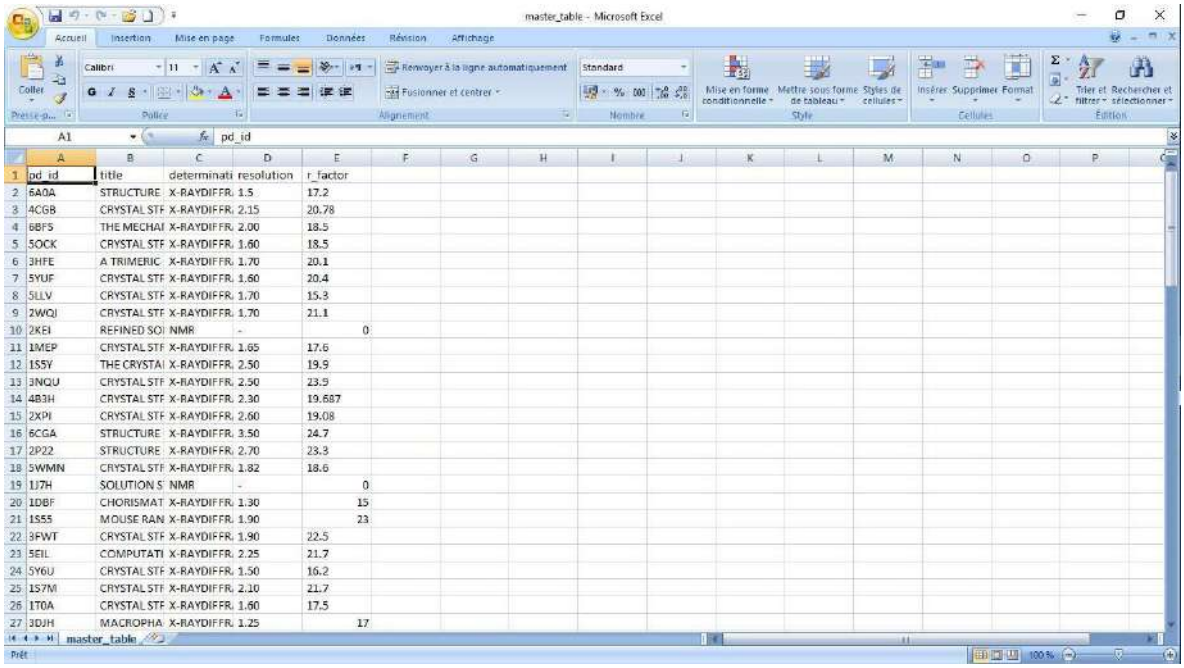
In-house Database Schema creation:

The database contained a number of tables dotted with primary keys and linked, when necessary, to each other via Foreign Keys.

The database work space which has been name as "**pp_interaction**" and tables were created via the use of PHP script codes written in this project for such a purpose.

Data Preparations

Using excel, data tables were created organized and saved into text files with the specific and simple format; Comma Separated Vales or CSV (Figures47, 48, 49).

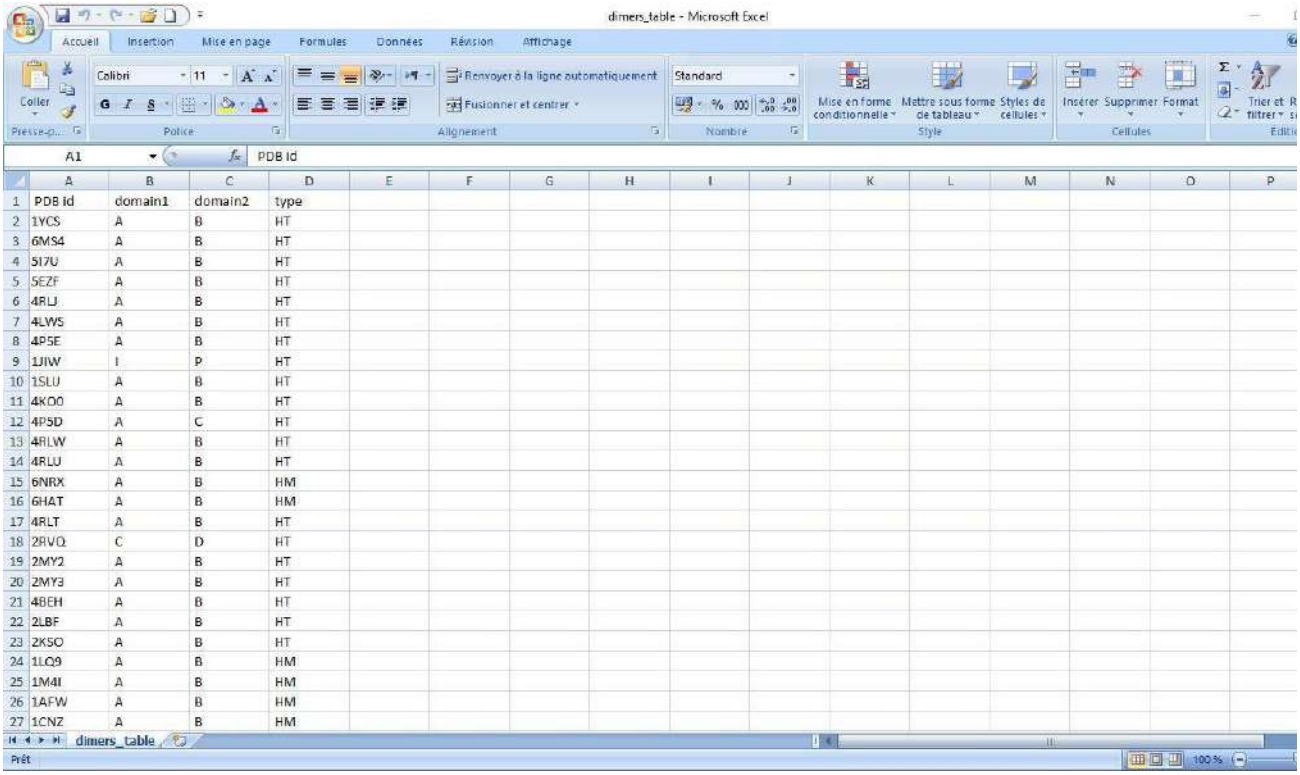


pd_id	title	determinati	resolution	r_factor
6A0A	STRUCTURE	X-RAYDIFFR	1.5	17.2
4CGB	CRYSTAL STF	X-RAYDIFFR	2.15	20.78
6BFS	THE MECHAI	X-RAYDIFFR	2.00	18.5
5OCK	CRYSTAL STF	X-RAYDIFFR	1.60	18.5
3HFE	A TRIMERIC	X-RAYDIFFR	1.70	20.1
5YUF	CRYSTAL STF	X-RAYDIFFR	1.60	20.4
5LLV	CRYSTAL STF	X-RAYDIFFR	1.70	15.3
2WQI	CRYSTAL STF	X-RAYDIFFR	1.70	21.1
2KEI	REFINED SOI	NMR	-	0
1MEP	CRYSTAL STF	X-RAYDIFFR	1.65	17.6
1S5Y	THE CRYSTAI	X-RAYDIFFR	2.50	19.9
3NQU	CRYSTAL STF	X-RAYDIFFR	2.50	23.9
4B3H	CRYSTAL STF	X-RAYDIFFR	2.30	19.687
2XPI	CRYSTAL STF	X-RAYDIFFR	2.60	19.08
6CGA	STRUCTURE	X-RAYDIFFR	3.50	24.7
2P22	STRUCTURE	X-RAYDIFFR	2.70	23.3
5WMN	CRYSTAL STF	X-RAYDIFFR	1.82	18.6
1J7H	SOLUTION S	NMR	-	0
1DBF	CHORISMAT	X-RAYDIFFR	1.30	15
1S55	MOUSE RAN	X-RAYDIFFR	1.90	23
3FWT	CRYSTAL STF	X-RAYDIFFR	1.90	22.5
5EIL	COMPUTATI	X-RAYDIFFR	2.25	21.7
5Y6U	CRYSTAL STF	X-RAYDIFFR	1.50	16.2
1S7M	CRYSTAL STF	X-RAYDIFFR	2.10	21.7
1T0A	CRYSTAL STF	X-RAYDIFFR	1.60	17.5
3DJH	MACROPHA	X-RAYDIFFR	1.25	17

Figure 49: Depiction from the Master-table containing selected protein structures in excel

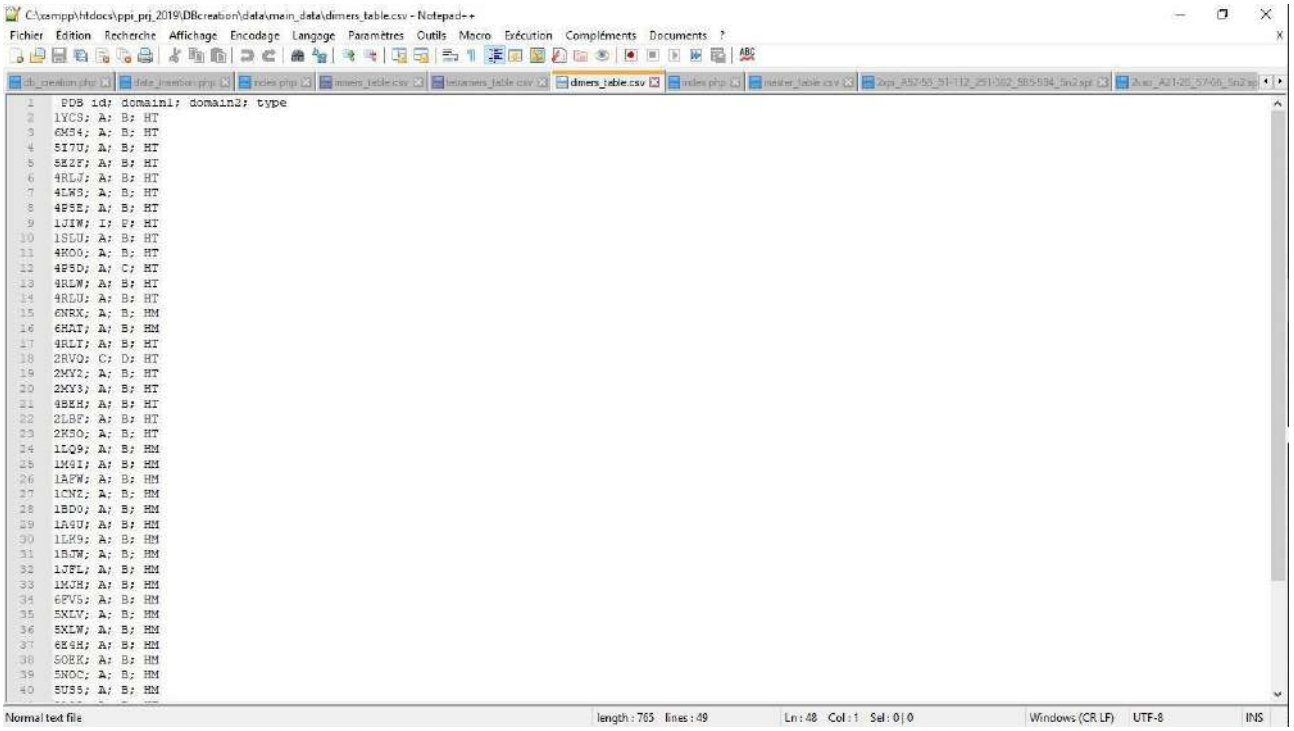
```
pd_id; title; determination_method; resolution; r_factor
6A0A; STRUCTURE OF A TRIPLE-HELIX REGION OF HUMAN COLLAGEN TYPE II; X-RAYDIFFRACTION; 1.5; 17.2
4CGB; CRYSTAL STRUCTURE OF THE TRIMERIZATION DOMAIN OF EML2; X-RAYDIFFRACTION; 2.15; 20.78
6BFS; THE MECHANISM OF GM-CSF INHIBITION BY HUMAN GM-CSF AUTO-ANTI; X-RAYDIFFRACTION; 2.00; 18.5
5OCK; CRYSTAL STRUCTURE OF ACPA E4 IN COMPLEX WITH CEP1; X-RAYDIFFRACTION; 1.60; 18.5
3HFE; A TRIMERIC FORM OF THE KV7.1 A DOMAIN TAIL; X-RAYDIFFRACTION; 1.70; 20.1
5YUF; CRYSTAL STRUCTURE OF PML RING TETRAMER; X-RAYDIFFRACTION; 1.60; 20.4
5LLV; CRYSTAL STRUCTURE OF DACM F87M/L110M TRANSTHYRETIN MUTANT; X-RAYDIFFRACTION; 1.70; 15.3
2WQI; CRYSTAL STRUCTURE OF THE HUMAN P73 TETRAMERIZATION DOMAIN; X-RAYDIFFRACTION; 1.70; 21.1
2KEI; REFINED SOLUTION STRUCTURE OF A DIMER OF LAC REPRESSOR DNA- BINDING DOMAIN COMPLEXED TO ITS NATURAL OPERATOR O1; NMR; -; 0
1MEP; CRYSTAL STRUCTURE OF STREPTAVIDIN DOUBLE MUTANT S45A/D128A WITH BIOTIN: COOPERATIVE HYDROGEN-BOND INTERACTIONS IN THE STREPTAVIDIN-BIOTIN SYSTEM; X-RAYDIFFI
1S5Y; THE CRYSTAL STRUCTURE OF TRENATOMUS BERNACCHII HEMOGLOBIN CX FERRICYANIDE; X-RAYDIFFRACTION; 2.50; 19.9
3NQU; CRYSTAL STRUCTURE OF PARTIALLY TRYPSINIZED (CENP-A/H4)2 HETE; X-RAYDIFFRACTION; 2.50; 23.9
4B3H; CRYSTAL STRUCTURE OF MYCOBACTERIUM TUBERCULOSIS FATTY ACID BETA-OXIDATION COMPLEX; X-RAYDIFFRACTION; 2.30; 19.687
2XPI; CRYSTAL STRUCTURE OF APC/C HETERO-TETRAMER CUT9-HCN1; X-RAYDIFFRACTION; 2.60; 19.08
6CGA; STRUCTURE OF THE PR-DUB COMPLEX; X-RAYDIFFRACTION; 3.50; 24.7
2P22; STRUCTURE OF THE YEAST ESCRT-I HETEROTETRAMER CORE; X-RAYDIFFRACTION; 2.70; 23.3
5WMN; CRYSTAL STRUCTURE OF HLA-B7 IN COMPLEX WITH SPI, AN INFLUEN2; X-RAYDIFFRACTION; 1.82; 18.6
1J7H; SOLUTION STRUCTURE OF HI0719, A HYPOTHETICAL PROTEIN FROM HAEMOPHILUS INFLUENZAE; NMR; -; 0
1DBF; CHORISMATE MUTASE FROM BACILLUS SUBTILLIS AT 1.30 ANGSTROM; X-RAYDIFFRACTION; 1.30; 15
1S55; MOUSE RANKL STRUCTURE AT 1.9A RESOLUTION; X-RAYDIFFRACTION; 1.90; 23
3FWT; CRYSTAL STRUCTURE OF LEISHMANIA MAJOR MIF2; X-RAYDIFFRACTION; 1.90; 22.5
5EIL; COMPUTATIONAL DESIGN OF A HIGH-AFFINITY METALLOPROTEIN HOMOT CONTAINING A METAL CHELATING NON-CANONICAL AMINO ACID; X-RAYDIFFRACTION; 2.25; 21.7
5Y6U; CRYSTAL STRUCTURE OF WILD-TYPE YABJ PROTEIN FROM BACILLUS SU (NATTO); X-RAYDIFFRACTION; 1.50; 16.2
1S7M; CRYSTAL STRUCTURE OF HIABD1; X-RAYDIFFRACTION; 2.10; 21.7
1T0A; CRYSTAL STRUCTURE OF 2C-METHYL-D-ERYTHRITOL-2,4-CYCLODIPHOSP SYNTHASE FROM SHEWANELLA ONEIDENSIS; X-RAYDIFFRACTION; 1.60; 17.5
3DJH; MACROPHAGE MIGRATION INHIBITORY FACTOR (MIF) AT 1.25 A RESOL; X-RAYDIFFRACTION; 1.25; 17
4J3H; RING CYCLE FOR DILATING AND CONSTRICTING THE NUCLEAR PORE: S OF A NUP54 HOMO-TETRAMER; X-RAYDIFFRACTION; 1.50; 19.5
4D2H; CRYSTAL STRUCTURE OF THE TETRAMERISATION DOMAIN OF HUMAN CTI; X-RAYDIFFRACTION; 1.90;  21.44
3KGT; V30M MUTANT HUMAN TRANSTHYRETIN (TTR) COMPLEXED WITH GENISTE (V30M:GEN) PH 7.5;  X-RAYDIFFRACTION; 1.95; 19.9
1CBM; THE 1.8 ANGSTROM STRUCTURE OF CARBONMONOXY-BETA4 HEMOGLOBIN: OF A HOMOTETRAMER WITH THE R QUATERNARY STRUCTURE OF LIGAND ALPHA2BETA2 HEMOGLOBIN; X-RAYDIFFRA
4MIF; PYRANOSE 2-OXIDASE FROM PHANEROCHAETE CHRYSOSPORIUM, WILD TY NATURAL SOURCE; X-RAYDIFFRACTION; 1.80; 15.3
4MIG; PYRANOSE 2-OXIDASE FROM PHANEROCHAETE CHRYSOSPORIUM, RECOMBI TYPE; X-RAYDIFFRACTION; 1.80; 17.4
1KKC; CRYSTAL STRUCTURE OF ASPERGILLUS FUMIGATUS MNSOD; X-RAYDIFFRACTION; 2.00; 19.4
1N7Y; STREPTAVIDIN MUTANT N23E AT 1.96A; X-RAYDIFFRACTION; 1.96; 16
1N9M; STREPTAVIDIN MUTANT S27A WITH BIOTIN AT 1.6A RESOLUTION; X-RAYDIFFRACTION; 1.60; 15.5
2NB1; P63/P73 HETERO-TETRAMERISATION DOMAIN; NMR; -; 0
4NUV; HETEROTETRAMER STRUCTURE OF REGION II FROM PLASMODIUM VIVAX BINDING PROTEIN (PVDBP) BOUND TO THE ECTODOMAIN OF THE DUFF RECEPTOR FOR CHEMOKINES (DARC); X-RA
5TCF; CRYSTAL STRUCTURE OF TRYPTOPHAN SYNTHASE FROM M. TUBERCULOSI LIGAND-FREE FORM; X-RAYDIFFRACTION; 2.46; 17.6
```

Figure 50: Master-table containing selected protein structures in Notepad++



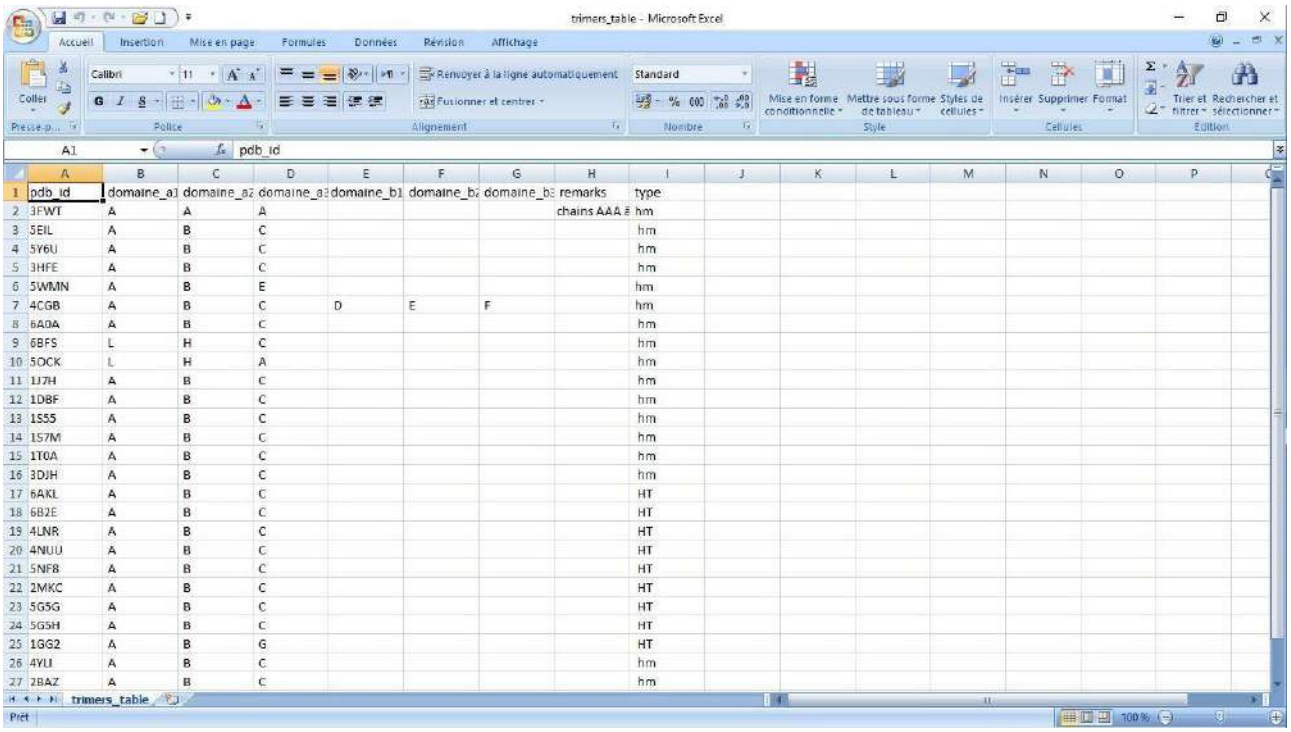
	A	B	C	D
1	PDB id	domain1	domain2	type
2	1YCS	A	B	HT
3	6MS4	A	B	HT
4	5I7U	A	B	HT
5	5EZF	A	B	HT
6	4RLJ	A	B	HT
7	4LWS	A	B	HT
8	4P5E	A	B	HT
9	1JIW	I	P	HT
10	1SLU	A	B	HT
11	4KO0	A	B	HT
12	4P5D	A	C	HT
13	4RLW	A	B	HT
14	4RLU	A	B	HT
15	6NRX	A	B	HM
16	6HAT	A	B	HM
17	4RLT	A	B	HT
18	2RVQ	C	D	HT
19	2MY2	A	B	HT
20	2MY3	A	B	HT
21	4BEH	A	B	HT
22	2LBF	A	B	HT
23	2K5O	A	B	HT
24	1LQ9	A	B	HM
25	1M4I	A	B	HM
26	1AFW	A	B	HM
27	1CNZ	A	B	HM

Figure 51: dimers table containing selected dimers_protein structures in excel



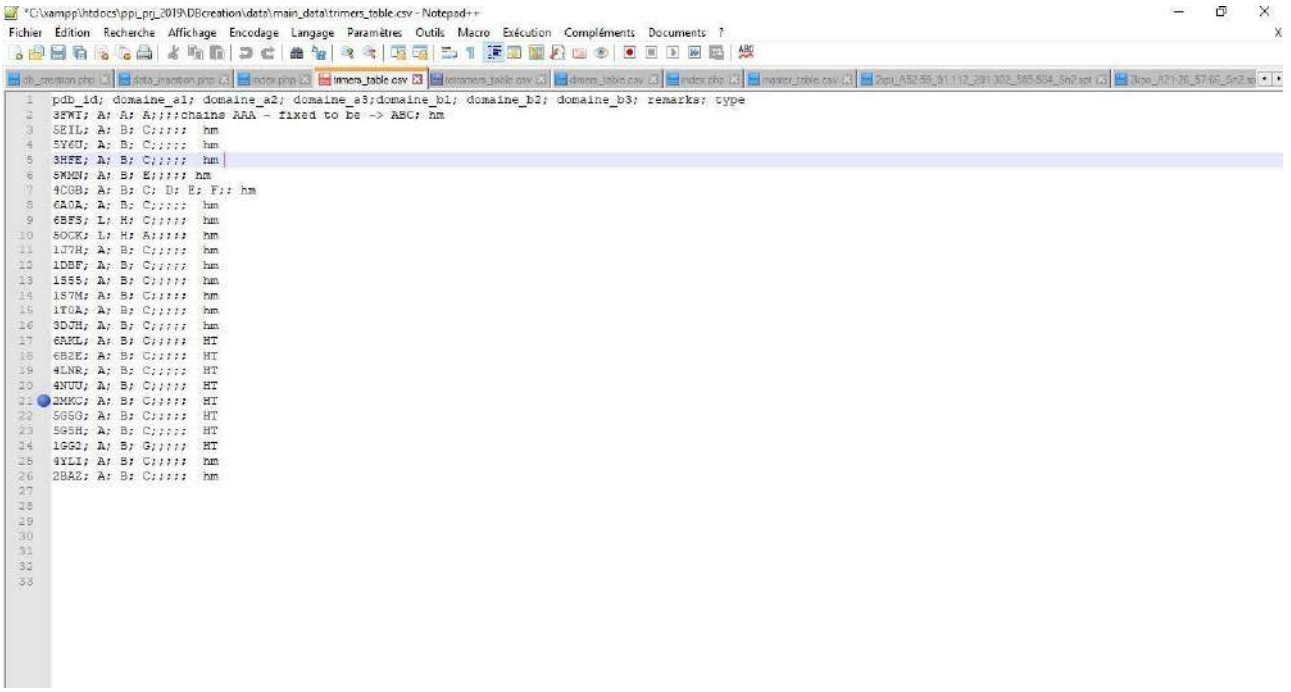
```
PDB id; domain1; domain2; type
1YCS; A; B; HT
6MS4; A; B; HT
5I7U; A; B; HT
5EZF; A; B; HT
4RLJ; A; B; HT
4LWS; A; B; HT
4P5E; A; B; HT
1JIW; I; P; HT
1SLU; A; B; HT
4KO0; A; B; HT
4P5D; A; C; HT
4RLW; A; B; HT
4RLU; A; B; HT
6NRX; A; B; HM
6HAT; A; B; HM
4RLT; A; B; HT
2RVQ; C; D; HT
2MY2; A; B; HT
2MY3; A; B; HT
4BEH; A; B; HT
2LBF; A; B; HT
2K5O; A; B; HT
1LQ9; A; B; HM
1M4I; A; B; HM
1AFW; A; B; HM
1CNZ; A; B; HM
1BD0; A; B; HM
1A4U; A; B; HM
1LR9; A; B; HM
1BJW; A; B; HM
1J8L; A; B; HM
1MJH; A; B; HM
6PVS; A; B; HM
5XLV; A; B; HM
5XLW; A; B; HM
6E4H; A; B; HM
5OEK; A; B; HM
5NOC; A; B; HM
5U55; A; B; HM
```

Figure 52: Dimers table containing selected dimers_protein structures in notpad++



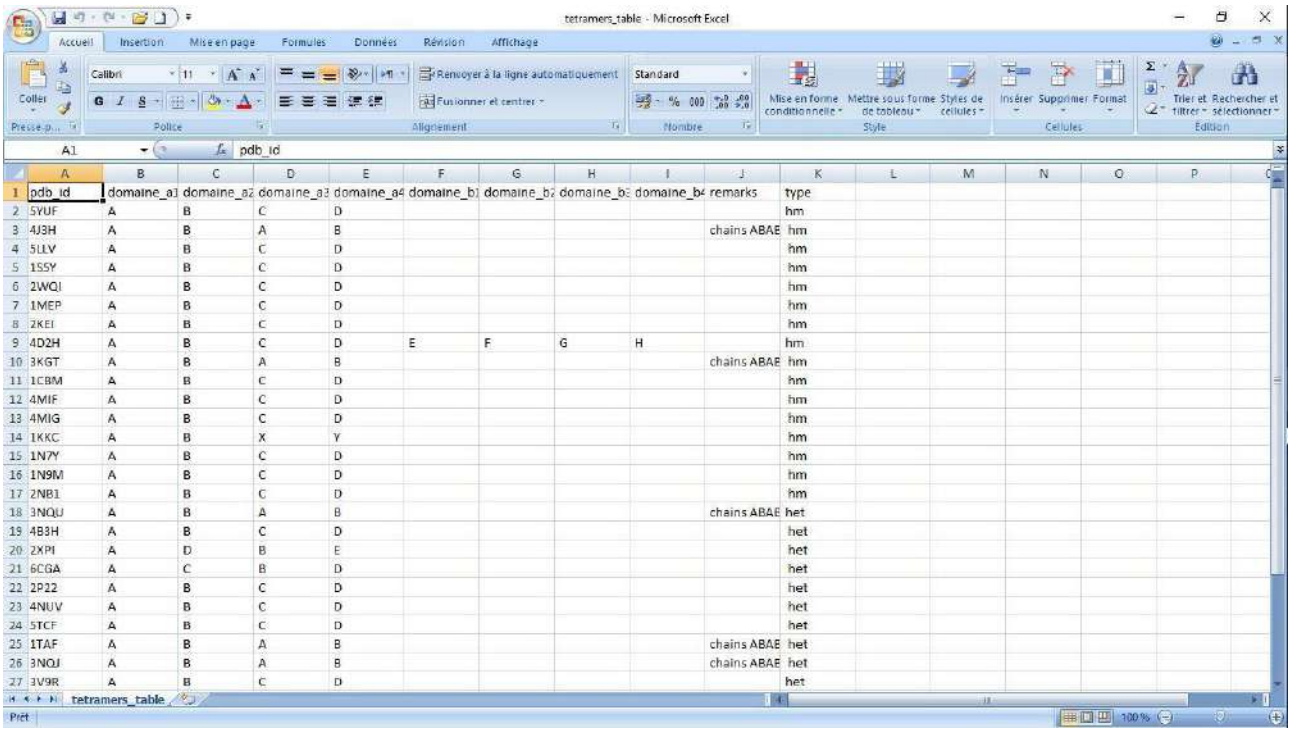
	pdb_id	domaine_a1	domaine_a2	domaine_a3	domaine_b1	domaine_b2	domaine_b3	remarks	type
2	3FWT	A	A	A				chains AAA à	hm
3	5EIL	A	B	C					hm
4	5Y6U	A	B	C					hm
5	3HFE	A	B	C					hm
6	5WMN	A	B	E					hm
7	4CGB	A	B	C	D	E	F		hm
8	6A0A	A	B	C					hm
9	6BFS	L	H	C					hm
10	5OCK	L	H	A					hm
11	1J7H	A	B	C					hm
12	1DBF	A	B	C					hm
13	1S55	A	B	C					hm
14	1S7M	A	B	C					hm
15	1T0A	A	B	C					hm
16	3DJH	A	B	C					hm
17	6AKL	A	B	C					HT
18	6B2E	A	B	C					HT
19	4LNR	A	B	C					HT
20	4NUU	A	B	C					HT
21	5NF8	A	B	C					HT
22	2MKC	A	B	C					HT
23	5G5G	A	B	C					HT
24	5G5H	A	B	C					HT
25	1GG2	A	B	G					HT
26	4YLI	A	B	C					hm
27	2BAZ	A	B	C					hm

Figure 53: Trimers table containing selected trimers_protein structures in excel



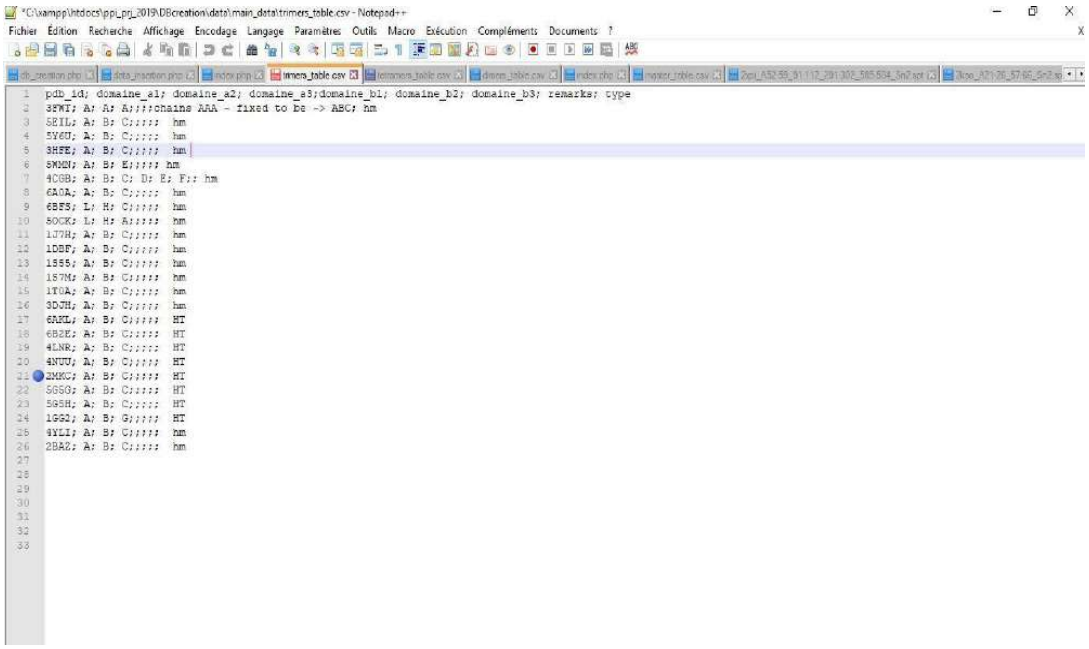
```
pdb_id; domaine_a1; domaine_a2; domaine_a3;domaine_b1; domaine_b2; domaine_b3; remarks; type
3FWT; A; A; A;;;;chains AAA - fixed to be -> ABC; hm
5EIL; A; B; C;;;;;  hm
5Y6U; A; B; C;;;;;  hm
3HFE; A; B; C;;;;;  hm
5WMN; A; B; E;;;;; hm
4CGB; A; B; C; D; E; F;; hm
6A0A; A; B; C;;;;;  hm
6BFS; L; H; C;;;;;  hm
5OCK; L; H; A;;;;;  hm
1J7H; A; B; C;;;;;  hm
1DBF; A; B; C;;;;;  hm
1S55; A; B; C;;;;;  hm
1S7M; A; B; C;;;;;  hm
1T0A; A; B; C;;;;;  hm
3DJH; A; B; C;;;;;  hm
6AKL; A; B; C;;;;;  HT
6B2E; A; B; C;;;;;  HT
4LNR; A; B; C;;;;;  HT
4NUU; A; B; C;;;;;  HT
5NF8; A; B; C;;;;;  HT
2MKC; A; B; C;;;;;  HT
5G5G; A; B; C;;;;;  HT
5G5H; A; B; C;;;;;  HT
1GG2; A; B; G;;;;;  HT
4YLI; A; B; C;;;;;  hm
2BAZ; A; B; C;;;;;  hm
```

Figure 54: Trimers table containing selected trimers_protein structures in notpad++



	pdb_id	domaine_a1	domaine_a2	domaine_a3	domaine_a4	domaine_b1	domaine_b2	domaine_b3	domaine_b4	remarks	type
2	5YUF	A	B	C	D						hm
3	4J3H	A	B	A	B					chains ABAB	hm
4	5LLV	A	B	C	D						hm
5	1SSY	A	B	C	D						hm
6	2WQI	A	B	C	D						hm
7	1MEP	A	B	C	D						hm
8	2KEI	A	B	C	D						hm
9	4D2H	A	B	C	D	E	F	G	H		hm
10	3KGT	A	B	A	B					chains ABAB	hm
11	1CBM	A	B	C	D						hm
12	4MIF	A	B	C	D						hm
13	4MIG	A	B	C	D						hm
14	1KKC	A	B	X	Y						hm
15	1N7Y	A	B	C	D						hm
16	1N9M	A	B	C	D						hm
17	2NB1	A	B	C	D						hm
18	3NQU	A	B	A	B					chains ABAB	het
19	4B3H	A	B	C	D						het
20	2XPI	A	D	B	E						het
21	6CGA	A	C	B	D						het
22	2P22	A	B	C	D						het
23	4NUV	A	B	C	D						het
24	5TCF	A	B	C	D						het
25	1TAF	A	B	A	B					chains ABAB	het
26	3NQJ	A	B	A	B					chains ABAB	het
27	3V9R	A	B	C	D						het

Figure 55: Tetramers table containing selected tetramers_protein structures in excel



```
pdb_id; domaine_a1; domaine_a2; domaine_a3;domaine_b1; domaine_b2; domaine_b3; remarks; type
3FWI; A; A;;;;chains AAA - fixed to bc -> ABC; hm
5EIL; A; B; C;;;;; hm
5V6U; A; B; C;;;;; hm
3HFE; A; B; C;;;;; hm
6MMN; A; B; E;;;;; hm
4CGB; A; B; C; D; E; F;; hm
6A0A; A; B; C;;;;; hm
6BF5; L; H; C;;;;; hm
5OCK; L; H; A;;;;; hm
1J7H; A; B; C;;;;; hm
1DBF; A; B; C;;;;; hm
1S55; A; B; C;;;;; hm
1S7M; A; B; C;;;;; hm
1T0A; A; B; C;;;;; hm
3DJH; A; B; C;;;;; hm
6AKL; A; B; C;;;;; HT
6B2E; A; B; C;;;;; HT
4LNR; A; B; C;;;;; HT
4NUU; A; B; C;;;;; HT
2MKC; A; B; C;;;;; HT
5GSG; A; B; C;;;;; HT
5GSH; A; B; C;;;;; HT
1GG2; A; B; G;;;;; HT
4YLI; A; B; C;;;;; hm
2BA2; A; B; C;;;;; hm
```

Figure 56: Tetramers table containing selected relevant protein structures Notpad++

All of the other data mentioned above including the selected PDB structures, Dimers, Trimers and Tetramers lists were saved in the CSV text files. The contacts data or the Interface Binding details calculate, mentioned in sections above, were generated by the PPI application ready in CSV format and downloaded and save in text files bearing systematic names according to the PDB ids and visually defined contact regions.

The Rasmol scenes for each and every case of found interface binding were created and stored in relevant JPG image files systematically named in the same fashion mentioned above.

The CSV format was found convenient for parsing by the PHP codes written for the purpose of data insertion into the database, see next steps in the methodology implemented in this project.

In-House Database Creation

The phpMyAdmin system allows for databases creation through the use of the standard menus provided for such puposes, see figure. Hoever, this is not very practical for complex kind of database creation. Through out this project, all database creation including tables generation and data insertion are created via the use of PHP programming.

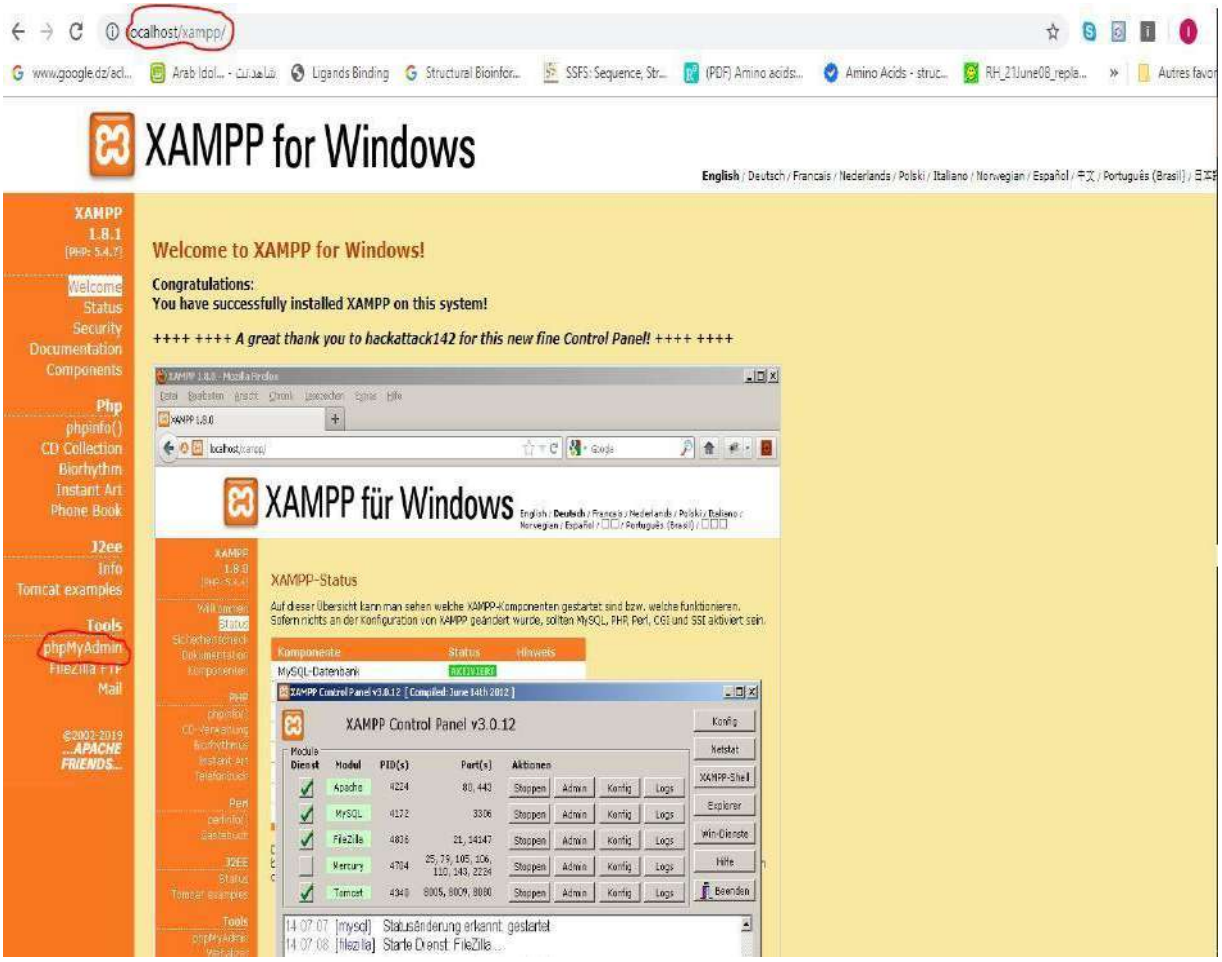


Figure 57: Create a new database named pp_interaction

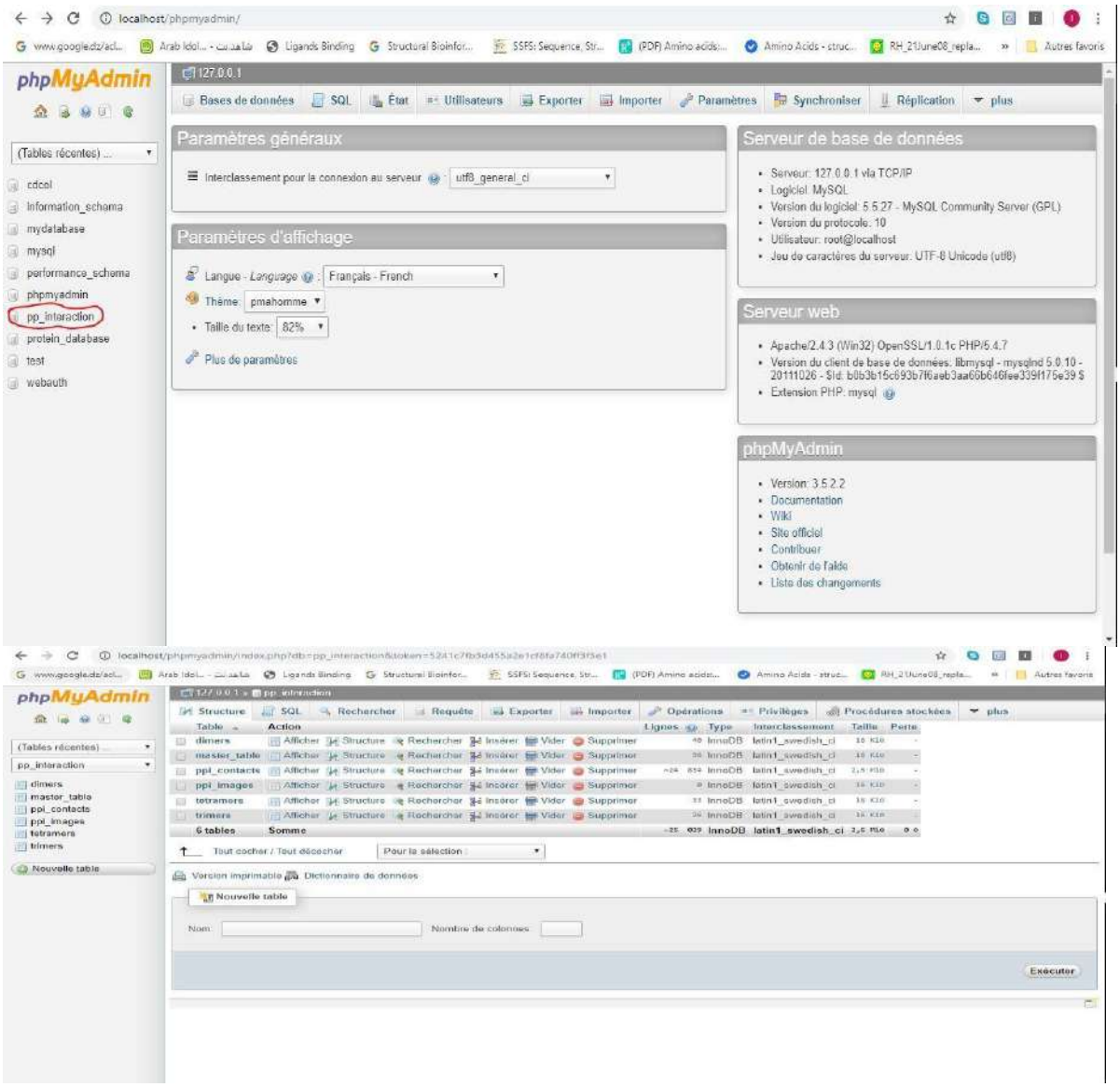


Figure 58: Represents the Interface of the main data in php my admin

A number of PHP codes have been written to create the database schema, tables definitions and necessary relational links between them.

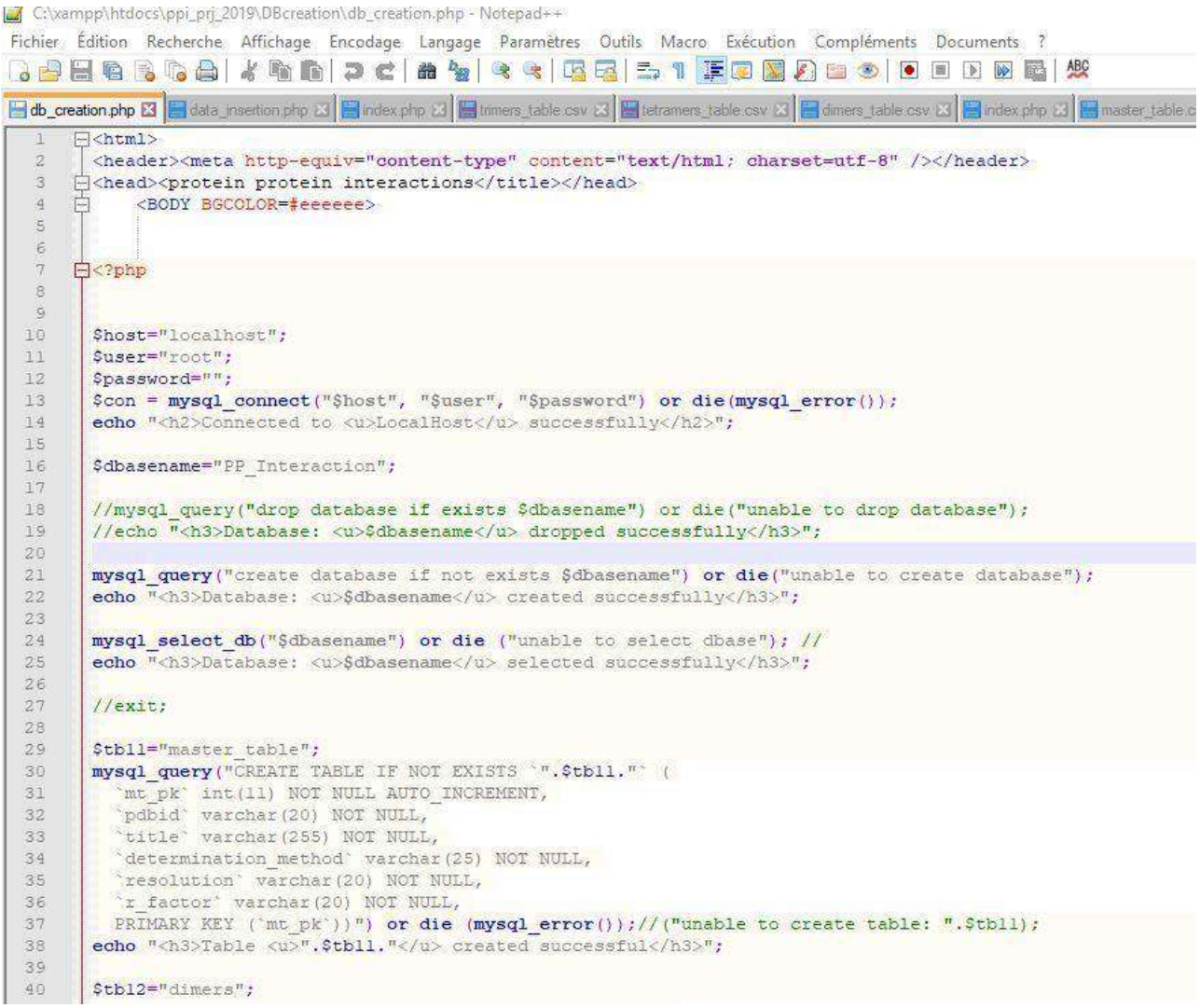
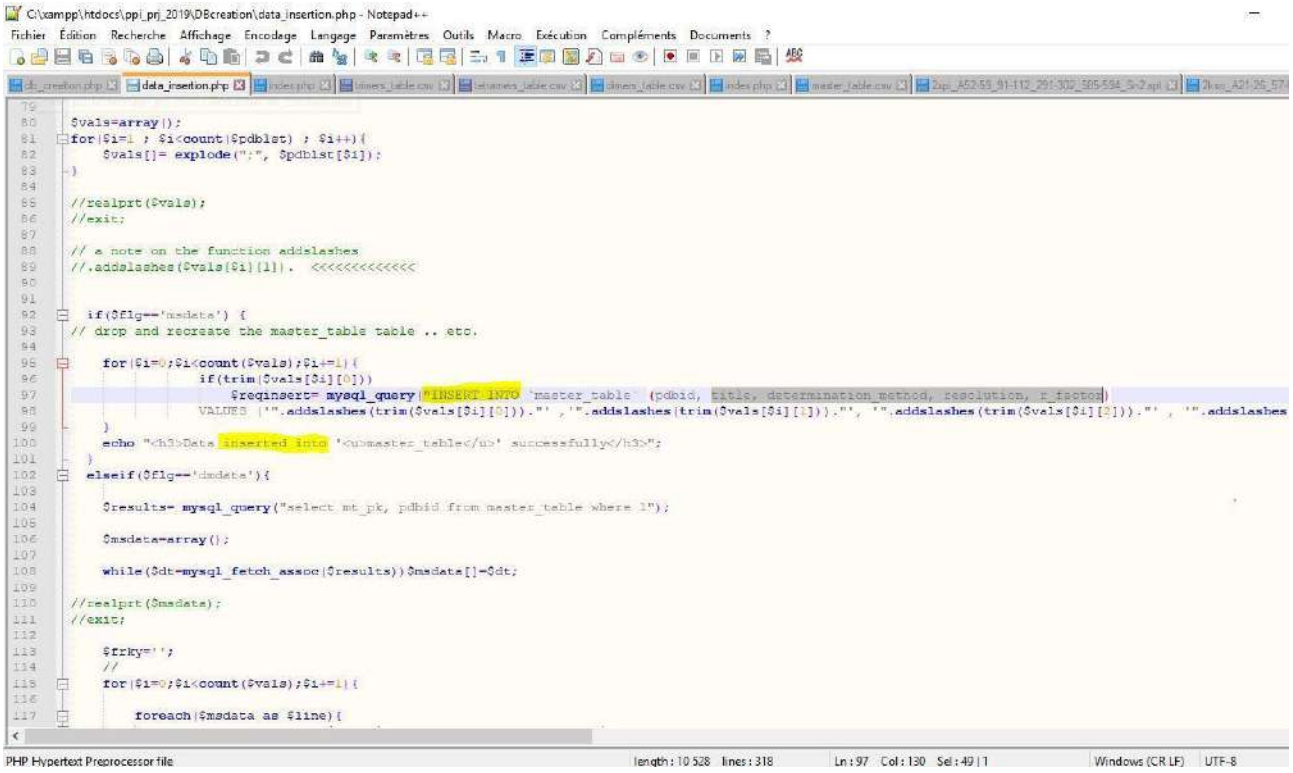


Figure 59: Depiction of one of the PHP codes implemented SQL functions developed to create the database and tables schema.

Data Insertion to the database and Retrieval

After the creation of database and tables schema in MySQL, the data Prepared Data inserted into each table via the use of SQL queries.

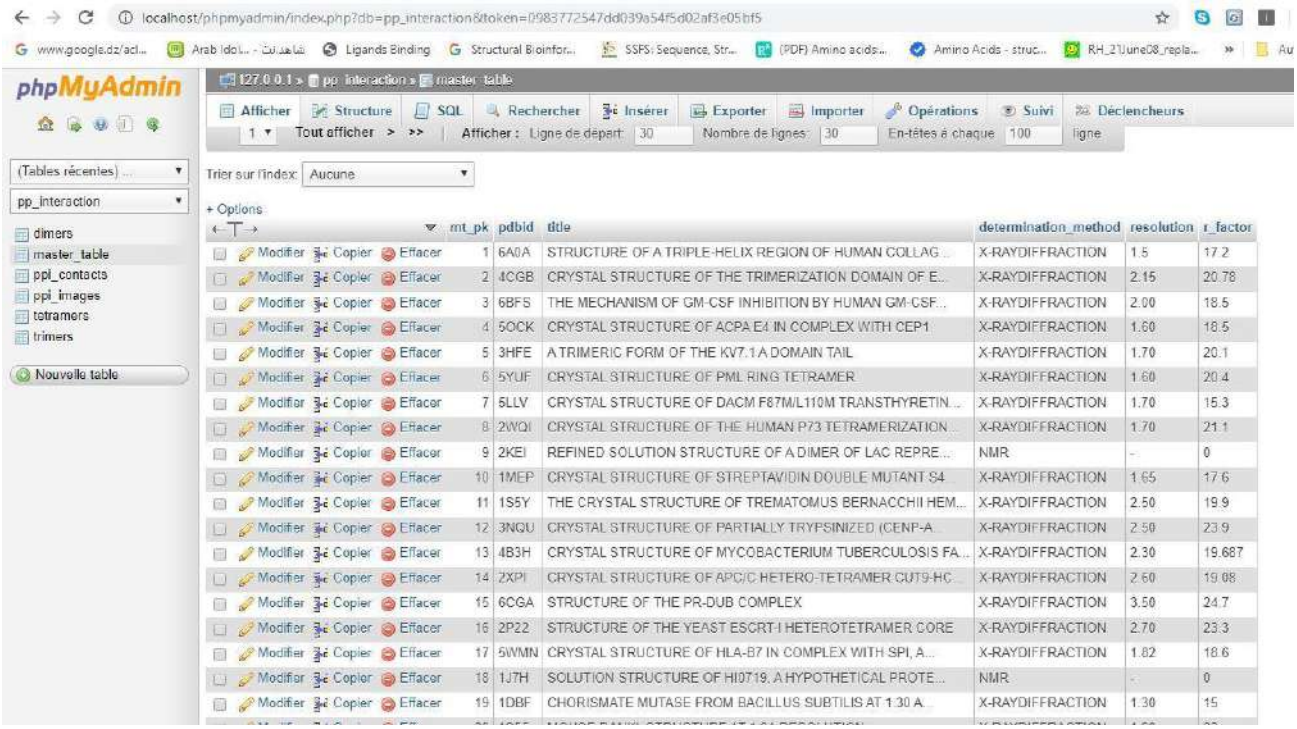
The **INSERT INTO** and related SQL statement have been used to insert new rows in the database tables **(Figure 60a** and **Figure 60b)**.



```
$vals=array();
for($i=1 ; $i<count($pdblst) ; $i++){
    $vals[]= explode(":", $pdblst[$i]);
}

//realprt($vals);
//exit;

// a note on the function addslashes
//.addslashes($vals[$i][1]).  <<<<<<<<<<<<<

 if($flg=='msdata') {
// drop and recreate the master_table table .. etc.

    for($i=0;$i<count($vals);$i+=1){
            if(trim($vals[$i][0]))
                $reqinsert= mysql_query("INSERT INTO `master_table` (pdbid, title, determination_method, resolution, r_factor)
              VALUES ('".addslashes(trim($vals[$i][0])."' ,'".addslashes(trim($vals[$i][1])."', '".addslashes(trim($vals[$i][2]))."' , '".addslashes
    }
    echo "<h3>Data inserted into '<u>master_table</u>' successfully</h3>";
 }
 elseif($flg=='msdata'){

    $results= mysql_query("select mt_pk, pdbid from master_table where 1");

    $msdata=array();

    while($dt=mysql_fetch_assoc($results))$msdata[]=$dt;

//realprt($msdata);
//exit;

    $frky='';
    //
    for($i=0;$i<count($vals);$i+=1){

        foreach($msdata as $line){
```

PHP Hypertext Preprocessor file — length : 10 528 lines : 318 — Ln : 97 Col : 130 Sel : 49 | 1 — Windows (CR LF) — UTF-8

Figure 60a: Depiction of an example PHP code implementing SQL statement data insertion.



phpMyAdmin — 127.0.0.1 » pp_interaction » master_table

Afficher | Structure | SQL | Rechercher | Insérer | Exporter | Importer | Opérations | Suivi | Déclencheurs

Tout afficher > >> | Afficher : Ligne de départ: 30 | Nombre de lignes: 30 | En-têtes à chaque 100 ligne

Trier sur l'index: Aucune

+ Options

(Tables récentes) ...

pp_interaction

- dimers
- master_table
- ppi_contacts
- ppi_images
- tetramers
- trimers

Nouvelle table

	mt_pk	pdbid	title	determination_method	resolution	r_factor
Modifier Copier Effacer	1	6A0A	STRUCTURE OF A TRIPLE-HELIX REGION OF HUMAN COLLAG...	X-RAYDIFFRACTION	1.5	17.2
Modifier Copier Effacer	2	4CGB	CRYSTAL STRUCTURE OF THE TRIMERIZATION DOMAIN OF E...	X-RAYDIFFRACTION	2.15	20.78
Modifier Copier Effacer	3	6BF5	THE MECHANISM OF GM-CSF INHIBITION BY HUMAN GM-CSF...	X-RAYDIFFRACTION	2.00	18.5
Modifier Copier Effacer	4	5OCK	CRYSTAL STRUCTURE OF ACPA E4 IN COMPLEX WITH CEP1	X-RAYDIFFRACTION	1.60	18.5
Modifier Copier Effacer	5	3HFE	A TRIMERIC FORM OF THE KV7.1 A DOMAIN TAIL	X-RAYDIFFRACTION	1.70	20.1
Modifier Copier Effacer	6	5YUF	CRYSTAL STRUCTURE OF PML RING TETRAMER	X-RAYDIFFRACTION	1.60	20.4
Modifier Copier Effacer	7	5LLV	CRYSTAL STRUCTURE OF DACM F87M/L110M TRANSTHYRETIN...	X-RAYDIFFRACTION	1.70	15.3
Modifier Copier Effacer	8	2WQI	CRYSTAL STRUCTURE OF THE HUMAN P73 TETRAMERIZATION...	X-RAYDIFFRACTION	1.70	21.1
Modifier Copier Effacer	9	2KEI	REFINED SOLUTION STRUCTURE OF A DIMER OF LAC REPRE...	NMR	-	0
Modifier Copier Effacer	10	1MEP	CRYSTAL STRUCTURE OF STREPTAVIDIN DOUBLE MUTANT S4	X-RAYDIFFRACTION	1.65	17.6
Modifier Copier Effacer	11	1S5Y	THE CRYSTAL STRUCTURE OF TREMATOMUS BERNACCHII HEM...	X-RAYDIFFRACTION	2.50	19.9
Modifier Copier Effacer	12	3NQU	CRYSTAL STRUCTURE OF PARTIALLY TRYPSINIZED (CENP-A...	X-RAYDIFFRACTION	2.50	23.9
Modifier Copier Effacer	13	4B3H	CRYSTAL STRUCTURE OF MYCOBACTERIUM TUBERCULOSIS FA...	X-RAYDIFFRACTION	2.30	19.687
Modifier Copier Effacer	14	2XPI	CRYSTAL STRUCTURE OF APC/C HETERO-TETRAMER CUT9-HC...	X-RAYDIFFRACTION	2.60	19.08
Modifier Copier Effacer	15	6CGA	STRUCTURE OF THE PR-DUB COMPLEX	X-RAYDIFFRACTION	3.50	24.7
Modifier Copier Effacer	16	2P22	STRUCTURE OF THE YEAST ESCRT-I HETEROTETRAMER CORE	X-RAYDIFFRACTION	2.70	23.3
Modifier Copier Effacer	17	5WMN	CRYSTAL STRUCTURE OF HLA-B7 IN COMPLEX WITH SPI, A...	X-RAYDIFFRACTION	1.82	18.6
Modifier Copier Effacer	18	1J7H	SOLUTION STRUCTURE OF HI0719, A HYPOTHETICAL PROTE...	NMR	-	0
Modifier Copier Effacer	19	1DBF	CHORISMATE MUTASE FROM BACILLUS SUBTILIS AT 1.30 A...	X-RAYDIFFRACTION	1.30	15

Figure 60b: Depiction from phpMyAmin showing an example of the Master-table data details after the automatic insertion by the relevamt PHP script.

Online Database Creation:

To share the data and the discoveries with the local and international scientific communities, the database has been made available online by mounting it on the server BioinfornaticsTools (**A. Rachedi, 2012**) by the supervisor of this project, who also wrote the necessary scripts to make the database searchable; that is to provide it with a web-interface for querying the database and displaying results in a meaningful way.The web version of the database has been named**: Protein Domain:Domain Interactions - Structural & Functional Interface Motifs Database -(PD:DI-SFIM)**

The **PD:DI-SFIM** database is accessible online via the following URL address:

http://bioinformaticstools.org/prjs/pddsfms/ (Figure 61)

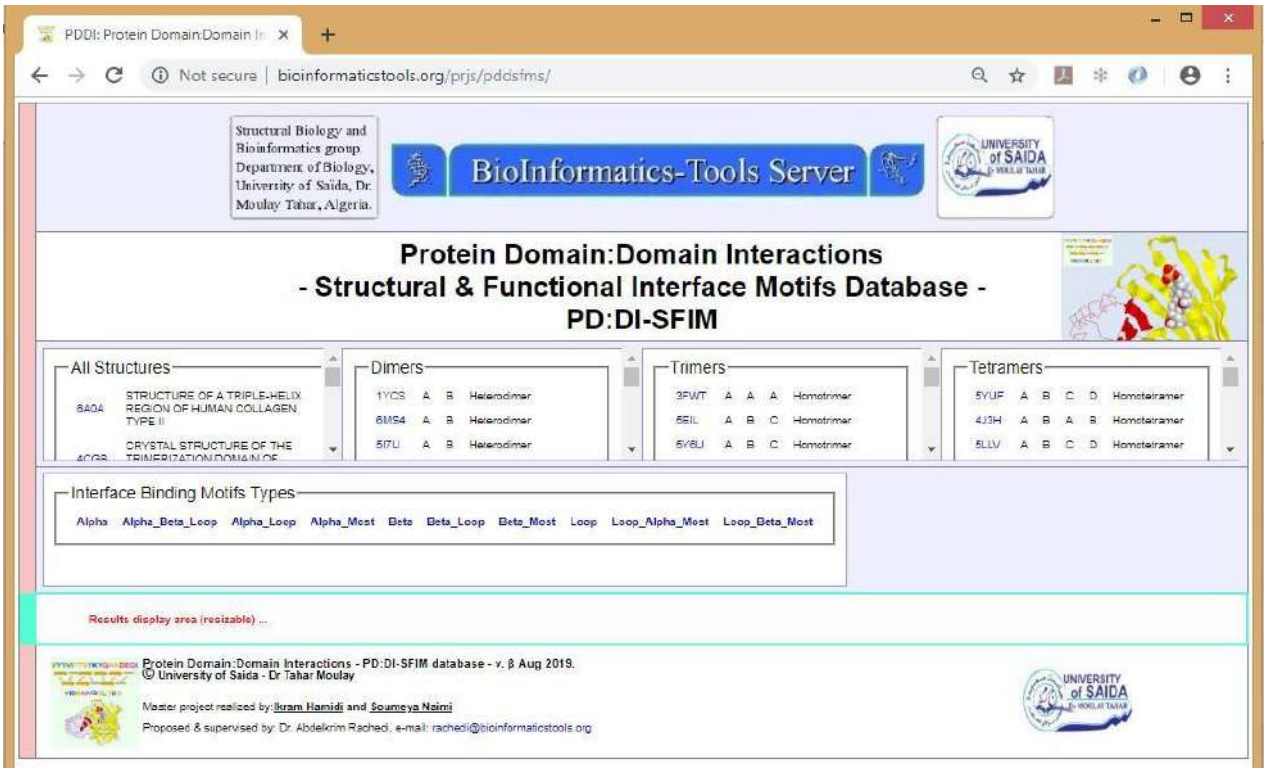


Figure 61: The main web interface of the PD:DI-SFIM

Results and Discussion

I. Presentation of the results

I. 1. Data Base web access and Querying:

The online version of the database "PD:DI-SFIM " can be uploaded by invoking the URL address shown in previous chapter. The figure below, shows the interface of the " PD:DI-SFIM " which has been developed to allow the easy querying of data recorded in the database.

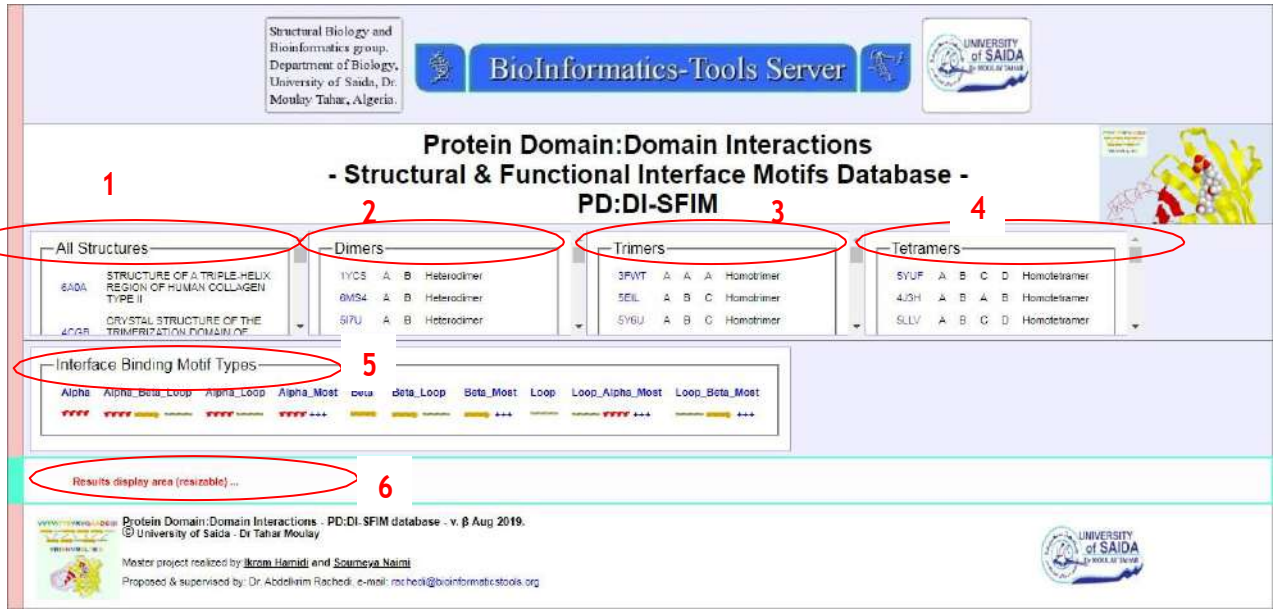


Figure 62: Capture of the interface page of the ”PD:DI-SFIM” database. Ellipse shapes and Numbers are for highlighting the four different ways made available to query the database.

I.2. Ways to Query the Database and Display Results:

As shown above in Figure 68, the PD:DI-SFIM interface provides 4 ways for searching the database content. For clarity these ways of searching are highlighted in yellow and numbered in red color:

1. All Structures:

This list can be used by users, by clicking, to directly explore any protein regardless of its type of structure.

A. Dimers:

This list provides **dimer** proteins together with their type of being Homo or Hetero.

B. Trimers:

This list provides **trimer** proteins and their type; Homo or Hetero.

C. Tetramers:

This is the area where users find only **tetramers** and their types; hetero-tetramers and homo-tetramers.

D. Interface binding motifs Types:

Results of any search action taken by clicking on one of the areas (**1, 2, 3** and **4**), shown above, will be shown in the "Results display area …" (pink rectangle) as highlighted above in **« 6 »**

As shown in (**Figure 63**) below a bar of the discovered motif typeslist. When users click on the types, like 'Alpha' or 'Alpha_ Beta _Loop' etc., a list of PDB ids,where the motifs types exist, is displayed in a box next to the motifs types list.

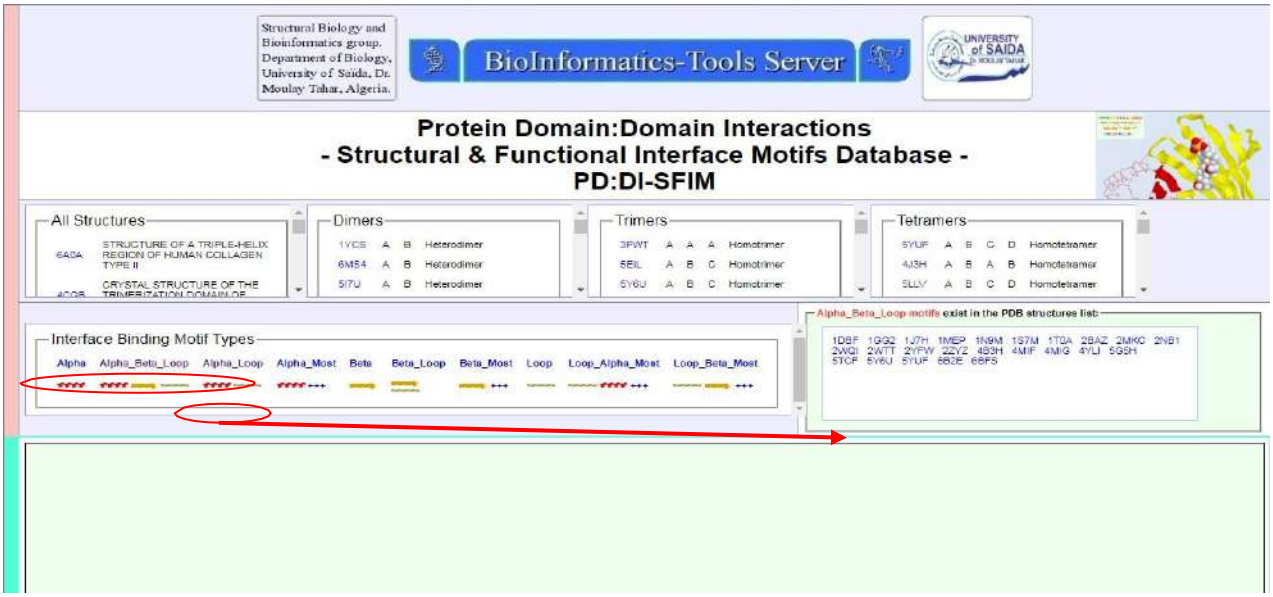


Figure 63: Results for interface binding motifs types details.

I.3. Search by PDB entries:

As explained above, if the user clicks PDB entries list (1,2, 3 or 4), the search engine of the database will search accordingly and display the results as follows in **Figure 64**:

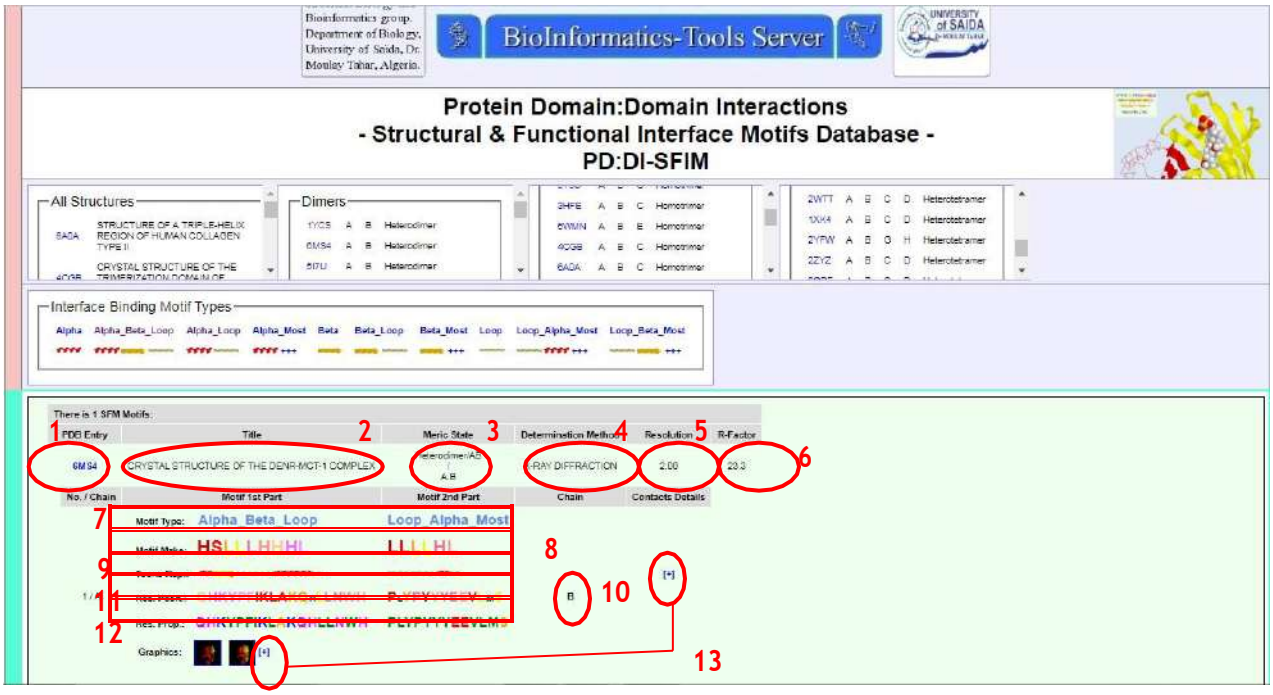


Figure 64: Capture represents the interface after clicking a PDBid "6ms4".

Where:

1: PDB identity code entry in the PDB database for the protein in question.

2: Protein name.

3: The polymeric state of the complex structure represented by the pdb id and associated chains.

4,5,6: Display respectively the method of determination, resolution and R-Factor

7: Motifs classes

8: Motif elements

9: Simplified representation of the motif

10: Chain identity code of the protein.

11 : Residues in contact coloured by position in relevant secondary structure element in the motif.

12: Residues chemical properties

13: the [+] link are to be clicked for displaying details of contacts and graphical Rasmol scenes of the binding holding the quaternary structure of the protein see Figures 65a, 65b and 65c

Figure 65a: Screen capture represents enlarged image of motifs after clicking [+] on the left .

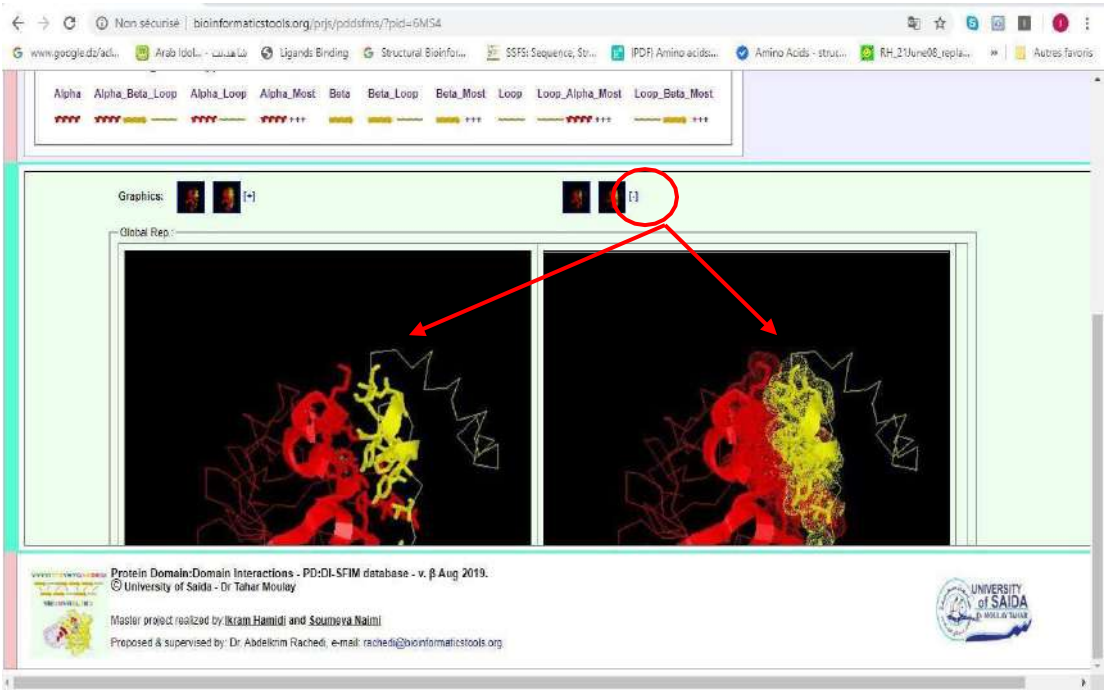


Figure 65b: Screen capture represents enlarged image of motifs after clicking on the [+] on the right.

I.3. Proteins Binding Details Display:

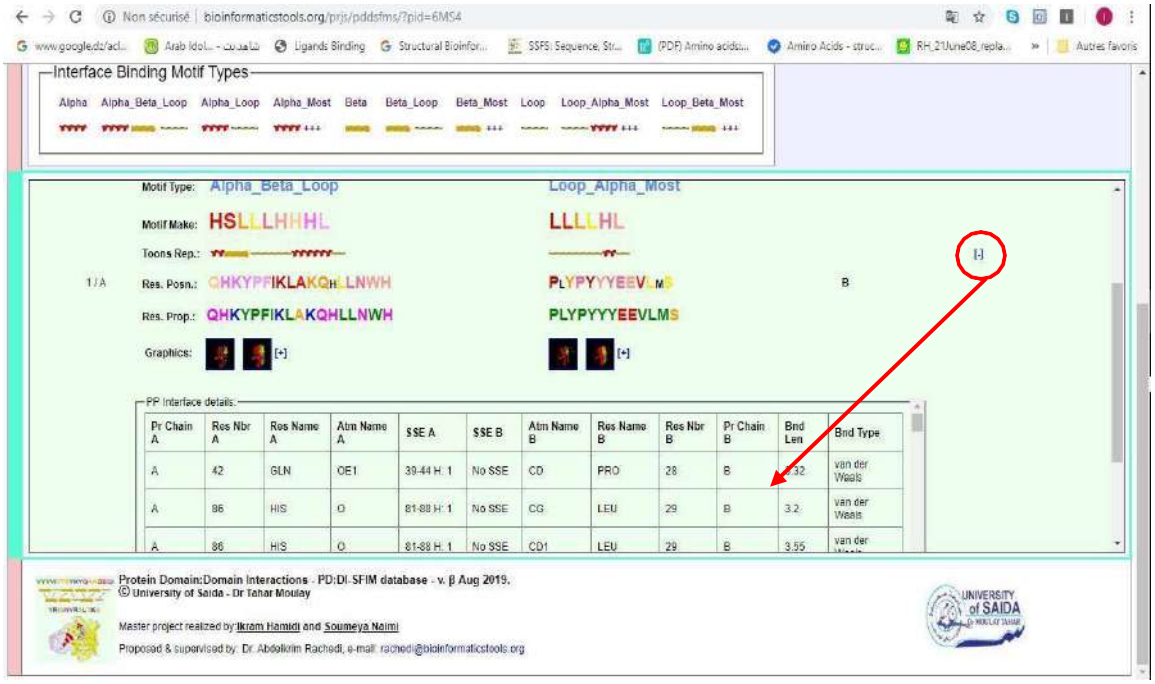


Figure 65c: Screen capture represents enlarged interface bonding contacts details. The mark [+] is to be clicked to show the table for the protein binding details. For details refer to Chapter II.

II. Properties of the interface binding environment:

II.1. Amino-acids types involved in the binding:

The properties of the amino acids (the hydrophobicity, polarity, ...) allow them to bind specifically with the secondary structures constituted each motif, which results a functional motif.

The table below shows types of amino acids are involved in the binding and each is characterized by specific properties the color of each amino-acid was chosen according to LESK classification

Table 11: Residues type shown arranged by their chemical properties. Orange = small non-polar.Green = hydrophobic.Magenta = polar. Red = negatively charge.Blue = positively charge.

PROPERTIES	AMINO-ACIDS
Small non polar	**G, A, S, T**
Hydrophobic	**C, V, I, L, P, F,Y, M, W**
Polar	**N, Q, H**
Negatively charge	**D, E**

Positively charge	**K**, **R**

II.2. Types of bonds in the binding:

The Van Der Waals interactions are not a hard binding system they arise from electron force attraction which allows to the molecules to be very close to one another.

The accumulative effect binding of the van der waals interactions (when the van der waals number > 4) give a strong system binding to the molecules.

Table12: Depiction of part of the**Bond types** characterizing the Interface Binding regions: where '++++++++++++'indicates most type of contacts and '+++' fewertype of contact. The rest of the data is found in the **index**.

PDB ID	HYDROGYN BOND	VAN_DER_WAALS
6A0A	+++	++++++++++++
4CGB	+++	++++++++++++
6BFS	+++	++++++++++++
5OCK	+++	++++++++++++
3HFE	+++	++++++++++++
5YUF	+++	++++++++++++
5LLV	+++	++++++++++++
2WQI	+++	++++++++++++
2KEI	+++	++++++++++++
1MEP	+++	++++++++++++
1S5Y	+++	++++++++++++
3NQU	+++	++++++++++++
4B3H	+++	++++++++++++
2XPI	+++	++++++++++++
6CGA	+++	++++++++++++
2P22	+++	++++++++++++
5WMN	+++	++++++++++++
1J7H	+++	++++++++++++

As it is noticed all the chosen proteins contains vander waals and hydrogen bonds for the following reasons:

- Hydrogen bonds provide many of the critical, life-sustaining properties of water and also stabilize the structures of proteins.
- Hydrogen bonds occur in inorganic molecules, such as water, and organic molecules (proteins).
- Van der Waals attractions can occur between any two or more molecules and are dependent on slight fluctuations of the electron densities.
- While hydrogen bonds and van der Waals interactions are weak individually, they are strong combined in vast numbers.
- As they are non-covalent bonds it allows the denaturation of the proteins the latter is one of the known properties of proteins.

II.3. Types of secondary elements in the interface binding regions:

On the interface each residues (Amino-acids) take the same color as the element of the motif with which it is linked in the example bellow the residues tyrosine and serine **(Y, S)** are linked with a loop **(L)** et tyrosine, threonine, Alanine **(Y, T, T, A)** are linked with a beta-sheet, the Methionine **(M)** is linked with a loop too and the Glutamic acid (E) is linked with a beta-sheet. So a functional motif is formed of a linkage of amino-acids and secondary structure elements **(H, L, S).**

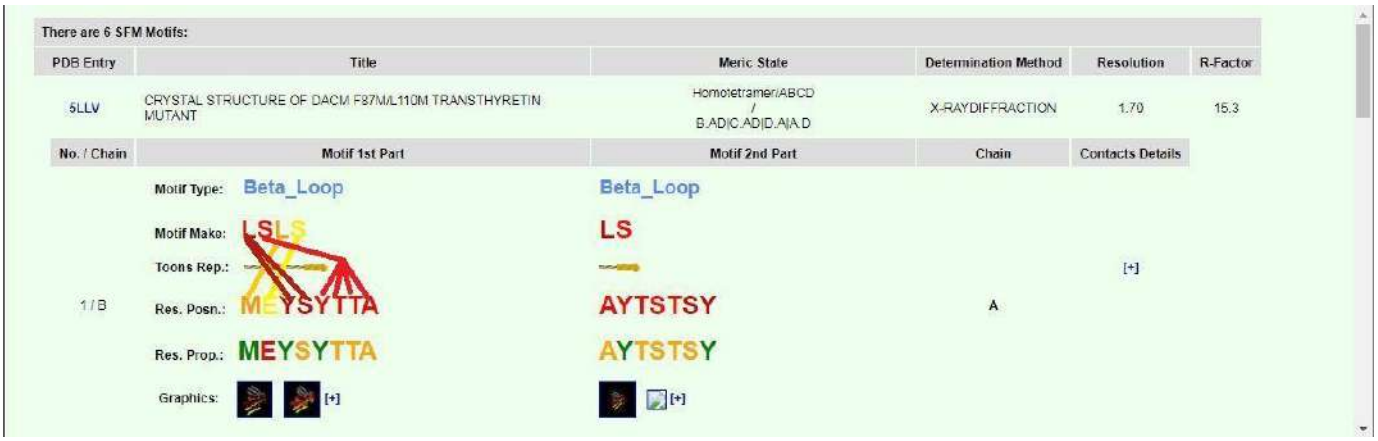


Figure 66: Screenshot shows the result page after the selection of a PDBid.

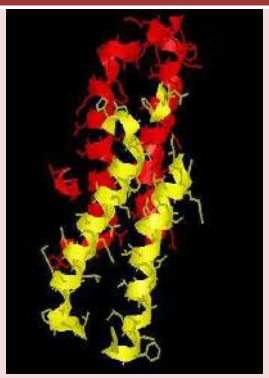
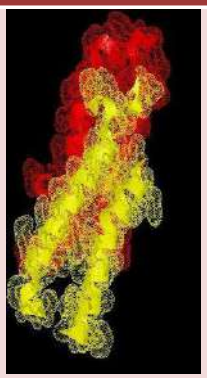
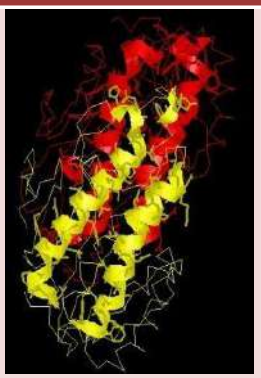
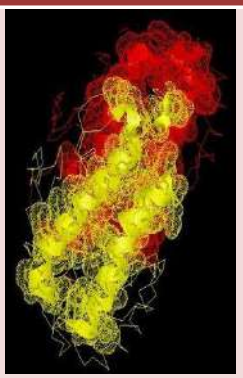
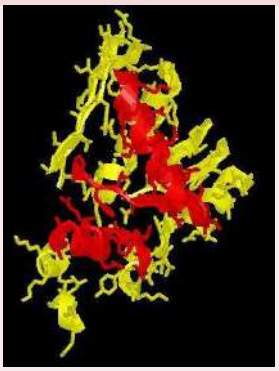
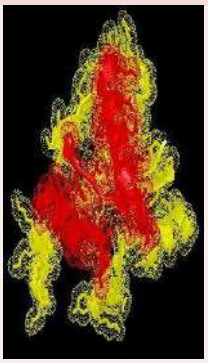
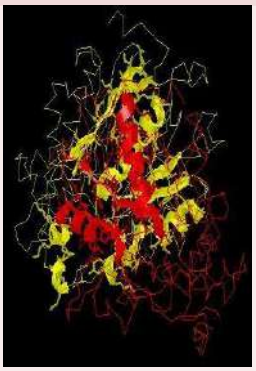
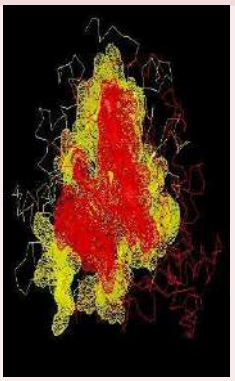
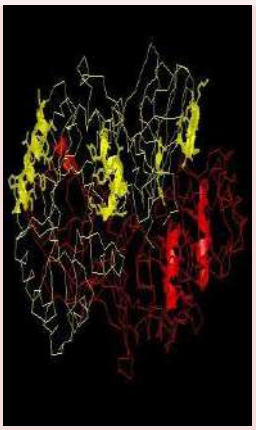
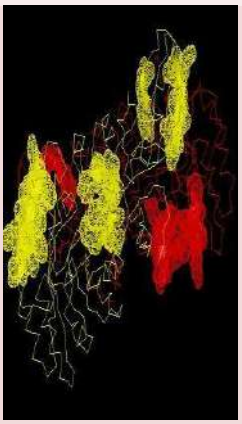
III. Interface Binding Motif Classes:

III.1 Motifs 3D-Graphical Representation:

The images below, **Table13**, represent the RasMol 3D-graphical representation (scenes) of some of the binding motifs associated with the PDB entries selected in this study. The full list of the scenes are listed in the Index.

• The 3D representation of all the scenes **(1, 2, 3, 4)** allow for the graphical visualization of the type of motifs in local depiction in scenes **1** & **2** and in global view in scenes **3** & **4**.

Table1 3: 3D representation of the binding motifs in four scenes (1, 2, 3, 4)

PDBid	Scene1	Scene2	Scene3	Scene4
1a4u				
1afw				
1bd0				

III.2 Motifs Classes:

This study has revealed 10 class types based on the types of secondary structure composition, see also Table below:

1. **Alpha [] :**
   This type of protein chains binding motifs is entirely made of helices (α-helices).

2. **Alpha_Beta_Loop [ ] :**
   This type of protein chain binding is a mixture of helices (α-helices), beta-strands (*) and loops.

3. **Alpha_Loop [ ] :**
   This type of protein chain binding motifs are made of helices (α-helices) and loops only.

4. **Alpha_most [+++] :**
   In this case the most secondary structure elements are helices (α-helices) but may contain few beta-strands and loop regions.

5. **Beta [ ] :**
   This type of protein chains binding motifs is entirely made of only beta-strands (*).

6. **Beta_Loop [ ] :**
   A couple of secondary structure element types enter into the constitution of this type of binding motifs which are: Beta-strands and loops (no helices)

7. **Beta_most [+++] :**
   Most of the protein chain binding motifs in this case are made of beta-strands (*). Few helices and loops may be part of.

8. **Loop [ ] :**
   This type of protein chains binding motifs is entirely made of loops.

9. **Loop_Alpha_Most [+++] :**
   In this case, most of secondary elements involved in making this type of binding motifs are loop regions and few helices (no beta-stands).

10. Loop_Beta_most [ +++] :

In this case the most of secondary elements involved in making this type of binding motifs are loop regions and few beta-stands (no helices).

(*) Beta-strand may exist individual or arranged in Beta-sheets or mix of.

Table 14: Classification of the Portein-Proyein Binding Motifs and associated examples of proteins structures (PDB ids). Names of classes are followed by cartoon representation of the secondary structure elements

Classes	Examples
Alpha [ ]	
HHHH / HH	**1xk4**
HH / HH	**2nb1**
H / H	**4yli**
Alpha_Beta_Loop [ ]	
HLHSH / LHLL	**4b3h**
HLHSLHLLHH / LLLLLH	**5tcf**
HLH / LSHSL	**6b2e**
HLSHLH / LSLH	**1t0a**
Alpha_Loop [ ] HHL /]	**5eil**
HH / LL	**4b3h**
HL / LH	**4nuv**

LHLLH / HLH	**1taf**

Alpha_most [vvvv

++ **2wqi**

+] HHH / HHHL

HHH / HLHL **2wqi**

HHH / LHL **4nuu**

HHLHLHHHLHH / HLHHLH **2xpi**

Beta [ ]

SSSS / SS **5llv**

SS / SS **1mep**

S / SS **2qpf**

S / S **1n9m**

Beta_Loop [ ]

LSLS / LS **5llv**

L/S **1kkc**

Loop_Beta_most [+++]

LLLSLLS / LSSLS **4lnr**

LLSLSSLLSSS / LLSLSLLL **3mez**

SL / LSL **5llv**

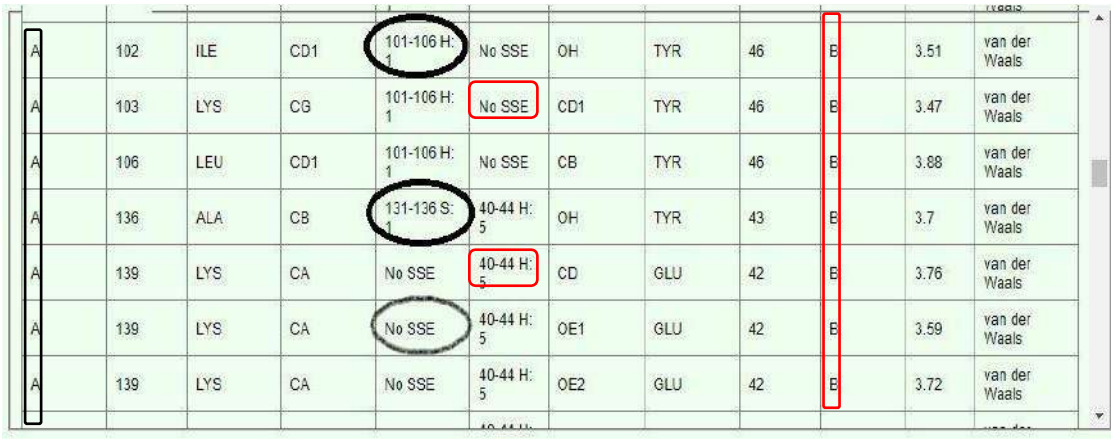
Beta_most [+++] :	
HLHSSLS / LS	**1dbf**
HSHHSLS / SLSS	**3djh**
LHLLLSSSLHHSS / LHSLSSHS	**6b2e**
LHSSLS / LSS	**1dbf**
Loop [ ]	
LLLLLL/LLLLLLLL	**6a0a**
LLLLL/LLLL	**6a0a**
LLLLL/LLL	**6a0a**
LLLL/L	**6bfs**
Loop_Alpha_most [+++]	
LHLHHLLLLHLLLL / HHLLLH	**5g5g**
LHL / HL	**1cbm**
LLHLHH / LLHLH	**1kkc**
LLHLLHLHHLLLHLL / HHLLLHHL	**5g5h**

IV. Properties of the motifs:

1. Motifs continues in space but not in sequence:

The secondary structural elements that form the motifs are continues in the 3D-space though they may belong to different regions in the sequence. To illustrate such a property, the PDB entry 6ms4 which represent the crystal structure of the density-regulated protein (DENR) and the malignant T cell-amplified sequence 1 (MCT-1/MCTS1) oncoprotein, as an example to understanding the situation.

Since this is a dimer made of two chains A and B there's one two-parts binding motif each associated with one of the chains A or B. It's the structural motif binding the interface and hence holing the chains together in the dimer, see **figure 67.**



A	102	ILE	CD1	101-106 H: 1	No SSE	OH	TYR	46	B	3.51	van der Waals
A	103	LYS	CG	101-106 H: 1	No SSE	CD1	TYR	46	B	3.47	van der Waals
A	106	LEU	CD1	101-106 H: 1	No SSE	CB	TYR	46	B	3.88	van der Waals
A	136	ALA	CB	131-136 S: 1	40-44 H: 5	OH	TYR	43	B	3.7	van der Waals
A	139	LYS	CA	No SSE	40-44 H: 5	CD	GLU	42	B	3.76	van der Waals
A	139	LYS	CA	No SSE	40-44 H: 5	OE1	GLU	42	B	3.59	van der Waals
A	139	LYS	CA	No SSE	40-44 H: 5	OE2	GLU	42	B	3.72	van der Waals

Figure 67: Represents the binding details of partial motif-1st-part1 HSL (chain A), in black, bound to partial motif-2nd-part LH, in red.

The amino acids numbering reported above show clearly that the motifs are distant in sequence context but close in space as shown in **figure 69** found below. The full motif-1st-part and motif-2nd-part are shown in **figure 68**.

Figure 68: Information about the binding motif associated with the PDB id 6ms4 annotated in the PD:DI SFI database.

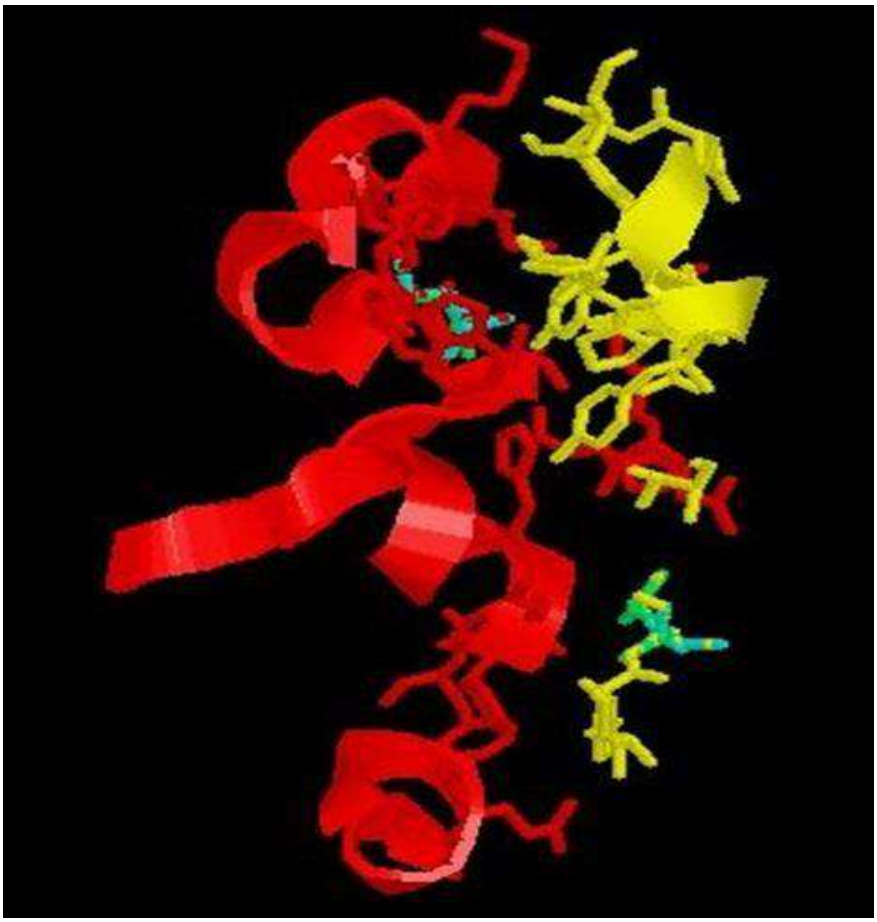


Figure 69: The 3D-structure of the interface binding motif HSLLLHHHL in red (chain A) / LLLLHL in yellow (chain B).

2. Multi-units/chains:

As also explained above, these motifs in their basic form are intrinsically made of two motif units which are bound to each other (**Figure70**) considering the quaternary structure of the proteins chosen. This is in contrast with the motifs created in a previous study "**Citric Acid Cycle Enzymes Research on the Function and Structure Relationship Bases**" (***Salem and Mebarka, 2016, SALEM Abdellahi and RACHEDI Abdelkrim, 2017***) and in related studies (***Abdellahi Ould Ahmed Salem et.al., 2018***) where the motifs for ligands binding are most of time single unit based, see the **figure71** below.

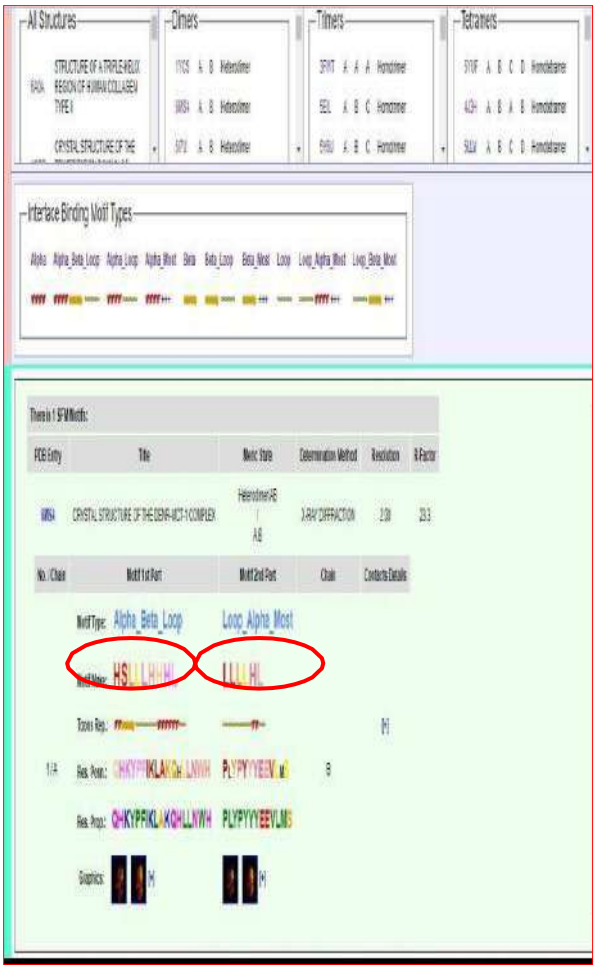


Figure 70: Double interface binding motif as shown in capture of the main page of "PD:DI-SFIM".

Figure 71: Single ligand binding motif as shown in capture of the interface page of the "CacSFMs" database.

Discussion:

The classification obtained in this study shows that the motifs consist of a single type of secondary structure (alpha family, beta family, loop family) are functional so that each of these protein binding can work individually there is no collaboration between the three secondary structure elements.

Contrary to what is common that loop regions are only connectors between secondary structure elements in the overall three-dimensional structure of proteins. The Loop family gives the indication that loop regions may play important roles in protein biological functions.

The presence of van der waals and hydrogen interactions in all the motifs has an interest in the dissociation and reassociation of the structures as in the case of hemoglobin or their units can dissociate easily thanks to the weak interactions.

Conclusion:

This study has set out to try and contribute in understanding the basis of Structure-Function relationship in macromolecules; protein is the case of this study, And also that complex proteins to carry out their function they rely on may aspect one is the binding regions which have been define in this study as motifs.

As shown in previous chapters, the contribution of this study has been to **define**, **characterize** and **classify** basic units of structural natural, called here as **binding motifs** that are responsible to some degree in **assembling** the proteins tertiary units to make the **quaternary** structure (dimmers, trimers, tetramers).

Doing such detailed study clearly contributes toward understanding and solving the outstanding structure-function relationship problem.

The structural elements (α-helices, β-strand and loops) in the define and characterized binding motifs are seen by this study as providing the structural support on which the functional elements.

The identification of motifs and their collection in a database can help widely in the discovery of proteins that have a similarity in relation to their motifs and therefore their functions.

Index

6A0A	STRUCTURE OF A TRIPLE-HELIX REGION OF HUMAN COLLAGEN TYPE II	X-RAYDIFFRACTION	1.5	17.2
4CGB	CRYSTAL STRUCTURE OF THE TRIMERIZATION DOMAIN OF EML2	X-RAY DIFFRACTION	2.15	20.78
6BFS	THE MECHANISM OF GM-CSF INHIBITION BY HUMAN GM-CSF AUTO-ANTI	X-RAYDIFFRACTION	2	18.5
5OCK	CRYSTAL STRUCTURE OF ACPA E4 IN COMPLEX WITH CEP1	X-RAY DIFFRACTION	1.6	18.5
3HFE	A TRIMERIC FORM OF THE KV7.1 A DOMAIN TAIL	X-RAYDIFFRACTION	1.7	20.1
5YUF	CRYSTAL STRUCTURE OF PML RING TETRAMER	X-RAY DIFFRACTION	1.6	20.4
5LLV	CRYSTAL STRUCTURE OF DACM F87M/L110M TRANSTHYRETIN MUTANT	X-RAYDIFFRACTION	1.7	15.3
2WQI	CRYSTAL STRUCTURE OF THE HUMAN P73 TETRAMERIZATION DOMAIN	X-RAY DIFFRACTION	1.7	21.1
2KEI	REFINED SOLUTION STRUCTURE OF A DIMER OF LAC REPRESSOR DNA- BINDING DOMAIN COMPLEXED TO ITS NATURAL OPERATOR O1	NMR	-	0
1MEP	CRYSTAL STRUCTURE OF STREPTAVIDIN DOUBLE	X-RAYDIFFRACTION	1.65	17.6

	MUTANT S45A/D128A WITH BIOTIN: COOPERATIVE HYDROGEN-BOND INTERACTIONS IN THE	N

	STREPTAVIDIN-BIOTIN SYSTEM			
1S5Y	THE CRYSTAL STRUCTURE OF TREMATOMUS BERNACCHII HEMOGLOBIN OX FERRICYANIDE	X-RAY DIFFRACTION	2.5	19.9
3NQU	CRYSTAL STRUCTURE OF PARTIALLY TRYPSINIZED (CENP-A/H4)2 HETE	X-RAYDIFFRACTION	2.5	23.9
4B3H	CRYSTAL STRUCTURE OF MYCOBACTERIUM TUBERCULOSIS FATTY ACID BETA-OXIDATION COMPLEX	X-RAYDIFFRACTION	2.3	19.687
2XPI	CRYSTAL STRUCTURE OF APC/C HETERO-TETRAMER CUT9-HCN1	X-RAY DIFFRACTION	2.6	19.08
6CGA	STRUCTURE OF THE PR-DUB COMPLEX	X-RAYDIFFRACTION	3.5	24.7
2P22	STRUCTURE OF THE YEAST ESCRT-I HETEROTETRAMER CORE	X-RAY DIFFRACTION	2.7	23.3
5WMN	CRYSTAL STRUCTURE OF HLA- B7 IN COMPLEX WITH SPI, AN INFLUENZ	X-RAYDIFFRACTION	1.82	18.6
1J7H	SOLUTION STRUCTURE OF HI0719, A HYPOTHETICAL PROTEIN FROM HAEMOPHILUS INFLUENZAE	NMR	-	0
1DBF	CHORISMATE MUTASE FROM BACILLUS SUBTILIS AT 1.30 ANGSTROM	X-RAY DIFFRACTION	1.3	15

1S55	MOUSE RANKL STRUCTURE AT 1.9A RESOLUTION	X-RAYDIFFRACTION	1.9	23
3FWT	CRYSTAL STRUCTURE OF LEISHMANIA MAJOR MIF2	X-RAY DIFFRACTION	1.9	22.5

5EIL	COMPUTATIONAL DESIGN OF A HIGH-AFFINITY METALLOPROTEIN HOMOT CONTAINING A METAL CHELATING NON-CANONICAL AMINO ACID	X-RAYDIFFRACTIO N	2.25	21.7
5Y6U	CRYSTAL STRUCTURE OF WILD-TYPE YABJ PROTEIN FROM BACILLUS SU (NATTO)	X-RAY DIFFRACTION	1.5	16.2
1S7M	CRYSTAL STRUCTURE OF HIABD1	X-RAY DIFFRACTION	2.1	21.7
1T0A	CRYSTAL STRUCTURE OF 2C-METHYL-D-ERYTHRITOL-2,4-CYCLODIPHOSP SYNTHASE FROM SHEWANELLA ONEIDENSIS	X-RAYDIFFRACTIO N	1.6	17.5
3DJH	MACROPHAGE MIGRATION INHIBITORY FACTOR (MIF) AT 1.25 A RESOL	X-RAYDIFFRACTIO N	1.25	17
4J3H	RING CYCLE FOR DILATING AND CONSTRICTING THE NUCLEAR PORE: S OF A NUP54 HOMO-TETRAMER	X-RAY DIFFRACTION	1.5	19.5
4D2H	CRYSTAL STRUCTURE OF THE TETRAMERISATION DOMAIN OF HUMAN CTI	X-RAYDIFFRACTIO N	1.9	21.44
3KGT	V30M MUTANT HUMAN TRANSTHYRETIN (TTR) COMPLEXED WITH GENISTE (V30M: GEN) PH 7.5	X-RAY DIFFRACTION	1.95	19.9
1CBM	THE 1.8 ANGSTROM STRUCTURE	X-RAYDIFFRACT	1.74	17.7

OF CARBONMONOXY-BETA4	IO
HEMOGLOBIN: OF A HOMOTETRAMER WITH THE R QUATERNARY STRUCTURE OF	N

	LIGAND ALPHA2BETA2 HEMOGLOBIN			
4MIF	PYRANOSE 2-OXIDASE FROM PHANEROCHAETE CHRYSOSPORIUM, WILD TY NATURAL SOURCE	X-RAY DIFFRACTION	1.8	15.3
4MIG	PYRANOSE 2-OXIDASE FROM PHANEROCHAETE CHRYSOSPORIUM, RECOMBI TYPE	X-RAYDIFFRACTIO N	1.8	17.4
1KKC	CRYSTAL STRUCTURE OF ASPERGILLUS FUMIGATUS MNSOD	X-RAY DIFFRACTION	2	19.4
1N7Y	STREPTAVIDIN MUTANT N23E AT 1.96A	X-RAYDIFFRACTIO N	1.96	16
1N9M	STREPTAVIDIN MUTANT S27A WITH BIOTIN AT 1.6A RESOLUTION	X-RAY DIFFRACTION	1.6	15.5
2NB1	P63/P73 HETERO-TETRAMERISATION DOMAIN	NMR	-	0
4NUV	HETEROTETRAMER STRUCTURE OF REGION II FROM PLASMODIUM VIVAX BINDING PROTEIN (PVDBP) BOUND TO THE ECTODOMAIN OF THE DUFF RECEPTOR FOR CHEMOKINES (DARC)	X-RAYDIFFRACTIO N	2.6	18.5
5TCF	CRYSTAL STRUCTURE OF TRYPTOPHAN SYNTHASE FROM M. TUBERCULOSI LIGAND-FREE FORM	X-RAY DIFFRACTION	2.46	17.6

1TAF	DROSOPHILA TBP ASSOCIATED FACTORS DTAFII42/DTAFII62 HETEROTE	X-RAYDIFFRACTION	2	19.8

3NQJ	CRYSTAL STRUCTURE OF (CENP-A/H4)2 HETEROTETRAMER	X-RAY DIFFRACTION	2.1	18.2
3V9R	CRYSTAL STRUCTURE OF SACCHAROMYCES CEREVISIAE MHF COMPLEX	X-RAYDIFFRACTION	2.4	24.5
2WTT	STRUCTURE OF THE HUMAN P73 TETRAMERIZATION DOMAIN (CRYSTAL FORM II)	X-RAY DIFFRACTION	2.3	23.33
1XK4	CRYSTAL STRUCTURE OF HUMAN CALPROTECTIN(S100A8/S100A9)	X-RAY DIFFRACTION	1.8	16
2YFW	HETEROTETRAMER STRUCTURE OF KLUYVEROMYCES LACTIS CSE4, H4	X-RAYDIFFRACTION	2.6	22.11
2ZYZ	PYROBACULUM AEROPHILUM SPLICING ENDONUCLEASE	X-RAY DIFFRACTION	1.7	21.2
2QPF	CRYSTAL STRUCTURE OF MOUSE TRANSTHYRETIN	X-RAYDIFFRACTION	2.05	20
3MEZ	X-RAY STRUCTURAL ANALYSIS OF A MANNOSE SPECIFIC LECTIN FROM CROCUS (CROCUS VERNUS)	X-RAY DIFFRACTION	1.94	18
1XOF	HETEROOLIGOMERIC BETA BETA ALPHA MINIPROTEIN	X-RAY DIFFRACTION	1.95	22.2
2KSO	STRUCTURE AND FUNCTION OF THE EPHA2:SHIP2 SAM DOMAI HETERODIMER	SOLUTION NMR	_	0
1YCS	STRUCTURE OF THE P53 TUMOR SUPPRESSOR BOUND TO THE	X-RAYDIFFRACTION	2.2	20.5

	ANKYRIN AND SH3 DOMAINS OF 53BP2	N		
2LBF	SOLUTION STRUCTURE OF THE DIMERIZATION DOMAIN OF HUMAN RIBOS PROTEIN P1/P2	SOLUTION NMR	_	0

	HETERODIMER			
2RVQ	SOLUTION STRUCTURE OF THE ISOLATED HISTONE H2A-H2B HETERODIM	SOLUTION NMR	_	0
2MY2	SNU17P-BUD13P STRUCTURE INTERMEDIATE DURING RES COMPLEX ASSE	SOLUTION NMR	_	0
2MY3	SNU17P-PML1P STRUCTURE INTERMEDIATE DURING RES COMPLEX ASSEM	SOLUTION NMR	_	0
4BEH	SOLUTION STRUCTURE OF HUMAN RIBOSOMAL PROTEIN P1.P2 HETERODI	SOLUTION NMR	_	0
4RLJ	CRYSTAL STRUCTURE OF (3R)-HYDROXYACYL-ACP DEHYDRATASE HADAB DIMER FROM MYCOBACTERIUM TUBERCULOSIS	X-RAY DIFFRACTION	1.75	15.9
4LWS	ESXA: ESXB (SEMET) HETERO-DIMER FROM THERMOMONOSPORA CURVAT	X-RAY DIFFRACTION	2	19
4P5E	CRYSTAL STRUCTURE OF HUMAN DNPH1 (RCL) WITH 6-NAPHTHYL-PURIN RIBOSIDE-MONOPHOSPHATE	X-RAY DIFFRACTION	1.35	15.6
4KO0	CRYSTAL STRUCTURE OF HIV-1 REVERSE TRANSCRIPTASE (RT) IN COM AN ANILINYLPYRIMIDINE DERIVATIVE (JLJ-135)	X-RAY DIFFRACTION	1.95	23.8
4P5D	CRYSTAL STRUCTURE OF RAT DNPH1 (RCL) WITH 6-NAPHTHYL-PURINE- MONOPHOSPHATE	X-RAY DIFFRACTION	2.11	17.6
4RL	CRYSTAL STRUCTURE OF (3R)-	X-RAY	2.2	16.5

W

HYDROXYACYL-ACP DIFFRACTION

	DEHYDRATASE HADAB DIMER FROM MYCOBACTERIUM TUBERCULOSIS COMPLEXED WITH BUTEIN			
4RLU	CRYSTAL STRUCTURE OF (3R)-HYDROXYACYL-ACP DEHYDRATASE HADAB DIMER FROM MYCOBACTERIUM TUBERCULOSIS COMPLEXED WITH 2',4,4 TRIHYDROXYCHALCONE	X-RAY DIFFRACTION	2.2	18.7
4RLT	CRYSTAL STRUCTURE OF (3R)-HYDROXYACYL-ACP DEHYDRATASE HADAB DIMER FROM MYCOBACTERIUM TUBERCULOSIS COMPLEXED WITH FISETI	X-RAY DIFFRACTION	2.05	15.9
5G5G	ESCHERICHIA COLI PERIPLASMIC ALDEHYDE OXIDASE	X-RAY DIFFRACTION	1.7	14
5G5H	ESCHERICHIA COLI PERIPLASMIC ALDEHYDE OXIDASE R440H MUTANT	X-RAY DIFFRACTION	2.3	16.7
4LNR	THE STRUCTURE OF HLA-B*35:01 IN COMPLEX WITH THE PEPTIDE (RP	X-RAY DIFFRACTION	2	19
4NUU	HETEROTRIMER STRUCTURE OF REGION II FROM PLASMODIUM VIVAX DU BINDING PROTEIN (PVDBP) BOUND TO THE	X-RAY DIFFRACTION	1.95	16.7

	ECTODOMAIN OF THE DUFF RECEPTOR FOR CHEMOKINES (DARC)			
2MKC	COOPERATIVE STRUCTURE OF THE HETEROTRIMERIC PRE-	SOLUTION NMR	_	0

	MRNA RETENTI SPLICING COMPLEX			
5OEK	PUTATIVE ACTIVE DIMERIC STATE OF GHR TRANSMEMBRANE DOMAIN	SOLUTION NMR	_	0
5OHD	PUTATIVE INACTIVE (DORMANT) DIMERIC STATE OF GHR TRANSMEMBRA	SOLUTION NMR	_	0
5NOC	SOLUTION NMR STRUCTURE OF THE C-TERMINAL DOMAIN OF PARB (SPO	SOLUTION NMR	_	0
5US5	SOLUTION STRUCTURE OF THE IREB HOMODIMER	SOLUTION NMR	_	0
6MS4	CRYSTAL STRUCTURE OF THE DENR-MCT-1 COMPLEX	X-RAY DIFFRACTION	2	23.3
5XLV	MYCOBACTERIUM TUBERCULOSIS PANTOTHENATE KINASE MUTANT F254A	X-RAY DIFFRACTION	1.8	17.9
6E4H	SOLUTION NMR STRUCTURE OF THE COLIED-COIL PALB2 HOMODIMER	SOLUTION NMR	_	0
6NRX	CRYSTAL STRUCTURE OF DIP-ETA IG1 HOMODIMER	X-RAY DIFFRACTION	1.9	23.4
6HAT	GLOBULAR DOMAIN OF HERPESVIRUS SAIMIRI ORF57	X-RAY DIFFRACTION	1.86	15
6FV5	QTRT2, THE NON-CATALYTIC SUBUNIT OF MURINE TRNA-GUANINE TRANSGLYCOSYLASE	X-RAY DIFFRACTION	2.18	17.1
6AKL	CRYSTAL STRUCTURE OF STRIATIN3 IN COMPLEX WITH SIKE1 COILED- DOMAIN	X-RAY DIFFRACTION	1.75	18.9
5XLW	MYCOBACTERIUM TUBERCULOSIS	X-RAY DIFFRACTION	2.26	19.3

	PANTOTHENATE KINASE MUTANT F247A			
1A4U	ALCOHOL DEHYDROGENASE	X-RAY	1.92	20.5

	FROM DROSOPHILA LEBANONENSIS	DIFFRACTION		
1AFW	THE 1.8 ANGSTROM CRYSTAL STRUCTURE OF THE DIMERIC PEROXISOMAL THIOLASE OF SACCHAROMYCES CEREVISIAE	X-RAY DIFFRACTION	1.8	18.74
1BD0	ALANINE RACEMASE COMPLEXED WITH ALANINE PHOSPHONATE	X-RAY DIFFRACTION	1.6	24
1BJW	ASPARTATE AMINOTRANSFERASE FROM THERMUS THERMOPHILUS	X-RAY DIFFRACTION	1.8	21.5
1CNZ	3-ISOPROPYLMALATE DEHYDROGENASE (IPMDH) FROM SALMONELLA TYPHIMURIUM	X-RAY DIFFRACTION	1.76	19.8
1JFL	CRYSTAL STRUCTURE DETERMINATION OF ASPARTATE RACEMASE FROM A	X-RAY DIFFRACTION	1.9	19.4
1LK9	THE THREE-DIMENSIONAL STRUCTURE OF ALLIINASE FROM GARLIC	X-RAY DIFFRACTION	1.53	19.3
1LQ9	CRYSTAL STRUCTURE OF A MONOOXYGENASE FROM THE GENE ACTVA-ORF STREPTOMYCES COELICOLOR STRAIN A3(2)	X-RAY DIFFRACTION	1.3	14.2
1M4I	AMINOGLYCOSIDE 2'-N-ACETYLTRANSFERASE FROM MYCOBACTERIUM TUB COMPLEX WITH COENZYME A	X-RAY DIFFRACTION	1.5	18.3

	AND KANAMYCIN A			
1MJH	STRUCTURE-BASED ASSIGNMENT OF THE BIOCHEMICAL FUNCTION OF	X-RAY DIFFRACTION	1.7	21

	HYPOTHETICAL PROTEIN MJ0577: A TEST CASE OF STRUCTURAL GENOMICS			
5NF8	SOLUTION STRUCTURE OF DETERGENT-SOLUBILIZED RCF1, A YEAST MITOCHONDRIAL INNER MEMBRANE PROTEIN INVOLVED IN RESPIRATOR III/IV SUPERCOMPLEX FORMATION	SOLUTION NMR	_	0
5I7U	HUMAN DPP4 IN COMPLEX WITH A NOVEL TRICYCLIC HETERO-CYCLE IN	X-RAY DIFFRACTION	1.95	0
5EZF	RACEMIC CRYSTAL STRUCTURES OF PRIBNOW BOX CONSENSUS PROMOTER (PBCA)	X-RAY DIFFRACTION	1.65	29.7
1jiw	CRYSTAL STRUCTURE OF THE APR-APRIN COMPLEX	X-RAY DIFFRACTION	1.74	18.4
1slu	RAT ANIONIC N143H, E151H TRYPSIN COMPLEXED TO A86H ECOTIN	X-RAY DIFFRACTION	1.8	19.3
2l35	STRUCTURE OF THE DAP12-NKG2C TRANSMEMBRANE HETEROTRIMER	SOLUTION NMR	-	0
1gg2	G PROTEIN HETEROTRIMER MUTANT GI_ALPHA_1(G203A) BETA_1 GAMMA_2 WITH GDP BOUND	X-RAY DIFFRACTION	2.4	20.5
4yli	CL-K1 TRIMER	X-RAY DIFFRACTION	2.45	18.9
6B2E	STRUCTURE OF FULL-LENGTH HUMAN AMPK (A2B2G1) IN COMPLEX WITH MOLECULE ACTIVATOR SC4	X-RAY DIFFRACTION	3.8	2.46

2baz	STRUCTURE OF YOSS, A PUTATIVE DUTPASE FROM	X-RAY DIFFRACTION	2.3	2.17

	BACILLUS SUBTILIS			
5kew	SIGNALING PROTEIN	X-RAY DIFFRACTION	2.1	19.8
5Kev	SIGNALING PROTEIN	X-RAY DIFFRACTION	2.7	26.5

The rest of master table of the proteins selected for project's study

Pdb Id	Chain	Type
1YCS	A	HT
	B	
6MS4	A	HT
	B	
5I7U	A	HT
	B	
5EZF	A	HT
	B	
4RLJ	A	HT
	B	
4LWS	A	HT
	B	
4PSE	A	HT
	B	
1JIW	I	HT
	P	
1SLU	A	HT
	B	
4KO0	A	HT
	B	
4P5D	A	HT
	C	
4RLW	A	HT
	B	
4RLU	A	HT
	B	
6NRX	A	HM
	B	
6HAT	A	HM
	B	
4RLT	A	HT

	B	

2RVQ	C	HT
	D	
2MY2	A	HT
	B	
2MY3	A	HT
	B	
4BEH	A	HT
	B	
2LBF	A	HT
	B	
2KSO	A	HT
	B	
1LQ9	A	HM
	B	
1M4I	A	HM
	B	
1AFW	A	HM
	B	
1CNZ	A	HM
	B	
1BD0	A	HM
	B	
1A4U	A	HM
	B	
1LK9	A	HM
	B	
1BJW	A	HM
	B	
1JFL	A	HM
	B	
1MJH	A	HM
	B	
6FV5	A	HM
	B	
5XLV	A	HM
	B	
5XLV W	A	HM
	B	
6E4H	A	HM
	B	
5OEK	A	HM
	B	
5NOC	A	HM
	B	
5US5	A	HM
	B	

2I35	A	HT
	B	
5OHD	A	HT
	B	
5kew	A	HT
	B	
5kev	A	HT
	B	
5NF8	A	HM
	B	

The rest of dimers table

Pdb Id	**Chain**	**Type**	**Remarks**
3FWT	A	HM	Chain AAA-fixed to be -> ABC
	A		
	A		
5EIL	A	HM	
	B		
	C		
5Y6U	A	HM	
	B		
	C		
3HFE	A	HM	
	B		
	C		
5WMN	A	HM	
	B		
	E		
4CGB	A	HM	
	B		
	C		
	D		
	E		
	F		
6A0A	A	HM	
	B		
	C		
6BFS	L	HM	
	H		
	C		
5OCK	L	HM	
	H		
	A		

1J7H	A	HM	
	B		
	C		
1DBF	A	HM	
	B		
	C		
1S55	A	HM	
	B		
	C		
1S7M	A	HM	
	B		
	C		
1T0A	A	HM	
	B		
	C		
3DJH	A	HM	
	B		
	C		
6AKL	A	HT	
	B		
	C		
6B2E	A	HT	
	B		
	C		
4LNR	A	HT	
	B		
	C		
4NUU	A	HT	
	B		
	C		
5NF8	A	HT	
	C		
	B		
2MKC	A	HT	
	B		
	C		
5G5G	A	HT	
	B		
	C		
5G5H	A	HT	
	B		
	C		
1GG2	A	HT	
	B		
	G		
4YLI	A	HM	

	B		
	C		
2BAZ	A	H	
	B		
	C		

All selected trimers

Pdb Id	Chain	Type	Remarks
5YUF	A	HM	
	B		
	C		
	D		
4J3H	A	HM	Chains ABAB fixed to be ABCD
	B		
	A		
	B		
5LLV	A	HM	
	B		
	C		
	D		
1S5Y	A	HM	
	B		
	C		
	D		
2WQI	A	HM	
	B		
	C		
	D		
1MEP	A	HM	
	B		
	C		
	D		
2KEI	A	HM	
	B		
	C		
	D		
4D2H	A	HM	
	B		
	C		
	D		
	E		
	D		
	G		

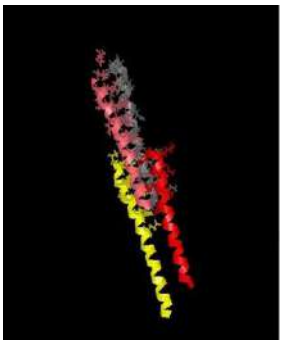
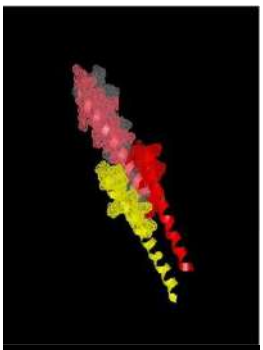
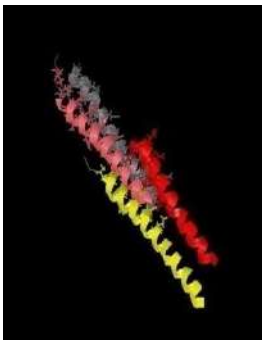
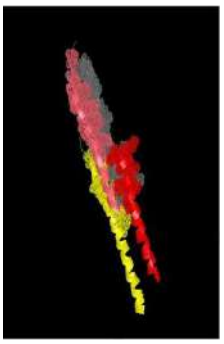
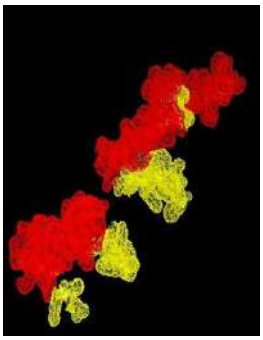
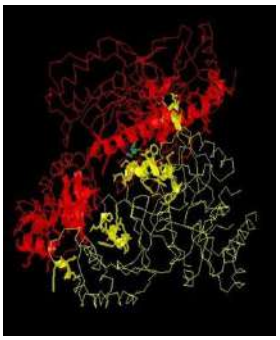
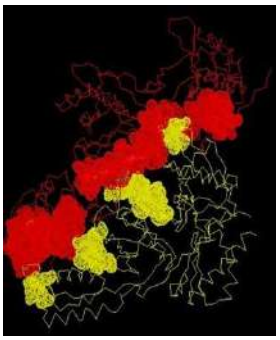
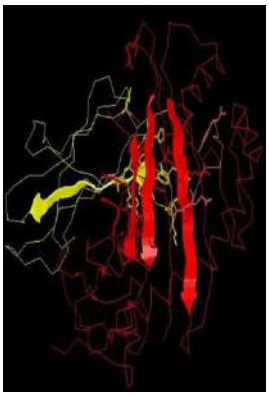
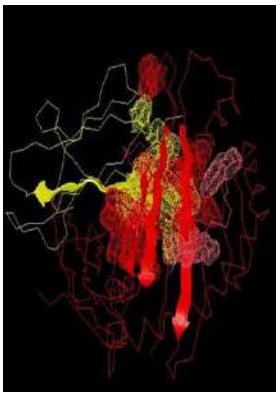
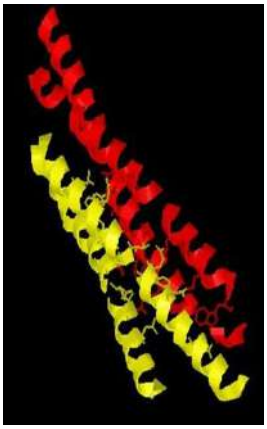
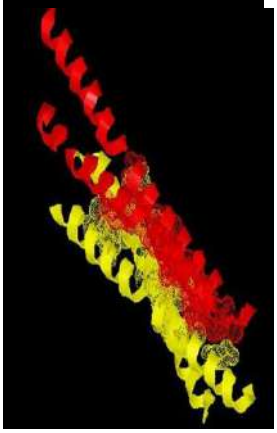
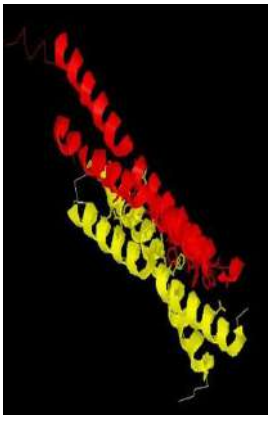
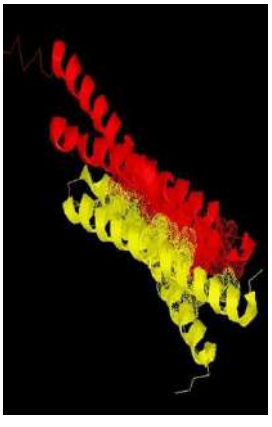
	H		

3KGT	A	HM	Chains ABAB fixed to be ABCD
	B		
	A		
	B		
1CBM	A	HM	
	B		
	C		
	D		
4MIF	A	HM	
	B		
	C		
	D		
4MIG	A	HM	
	B		
	C		
	D		
1KKC	A	HM	
	B		
	X		
	Y		
1N7Y	A	HM	
	B		
	C		
	D		
1N9M	A	HM	
	B		
	C		
	D		
2NB1	A	HM	
	B		
	C		
	D		
3NQU	A	HT	Chains ABAB fixed to be ABCD
	B		
	A		
	B		
4B3H	A	HT	
	B		
	C		
	D		
2XPI	A	HT	
	D		
	B		
	E		
6CGA	A	HT	
	C		

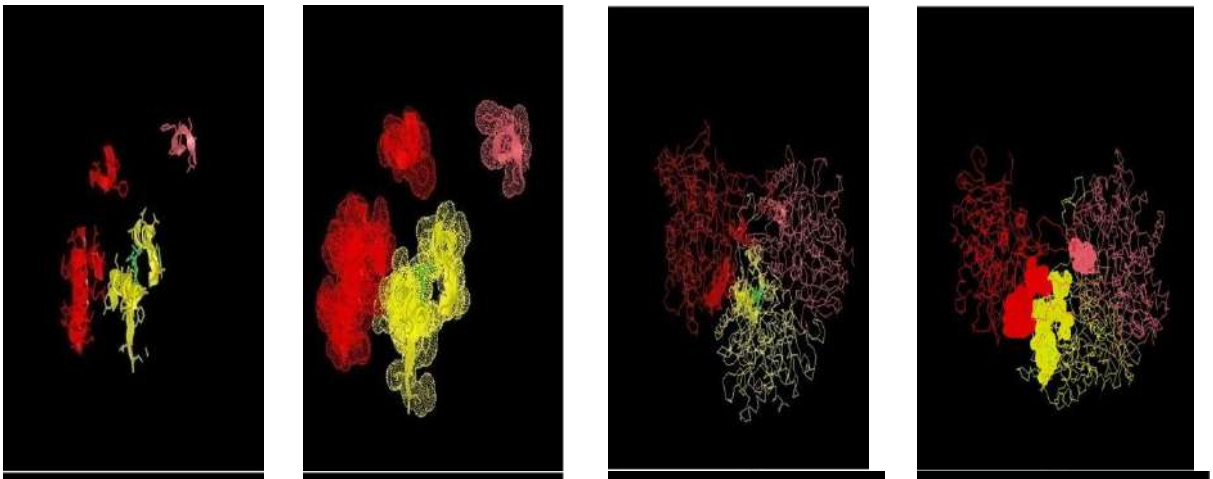
	B		
	D		
2P22	A	HT	
	B		
	C		
	D		
4NUV	A	HT	
	B		
	C		
	D		
5TCF	A	HT	
	B		
	C		
	D		
1TAF	A	HT	Chains ABAB fixed to be ABCD
	B		
	C		
	D		
3NGJ	A	HT	Chains ABAB fixed to be ABCD
	B		
	C		
	D		
3V9R	A	HET	
	B		
	C		
	D		
2WTT	A	HET	
	B		
	C		
	D		
	E		
	F		
	G		
	H		
1XK4	A	HET	
	B		
	C		
	D		
2YFW	A	HET	
	B		
	G		
	H		
2ZYZ	A	HET	
	B		
	C		
	D		

2QPF	A	HET	
	B		
	C		
	D		
3MEZ	A	HET	
	B		
	C		
	D		
1XOF	A	HET	Chains ABAB fixed to be ABCD
	B		
	A		
	B		

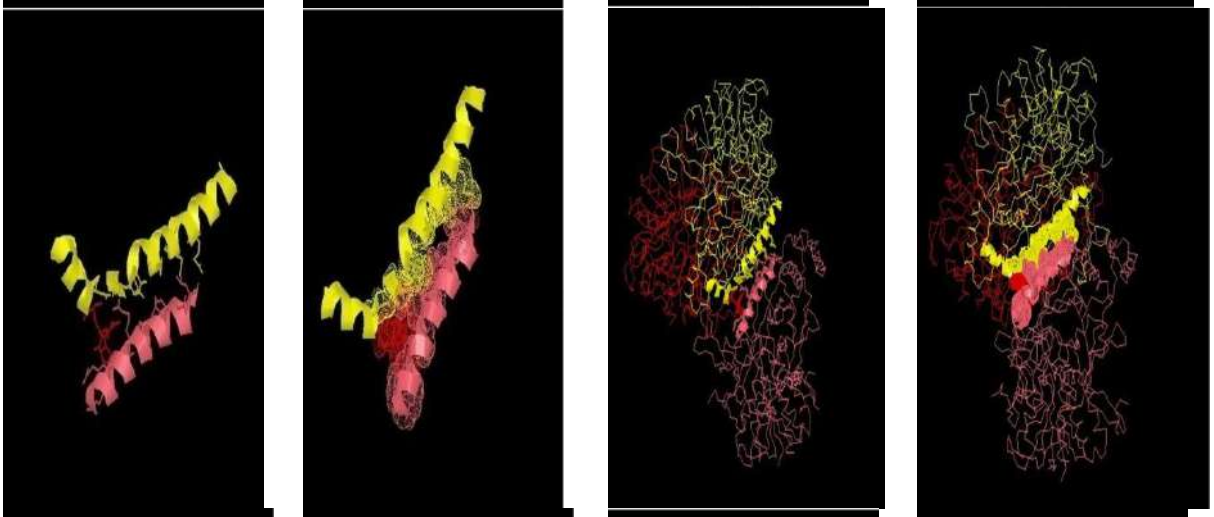
The rest of tetramers table

	Scene1	Scene2	Scene3	Scene4
4d2h « H »				
4j3h				
4ko0				
4lnr				
4lws				

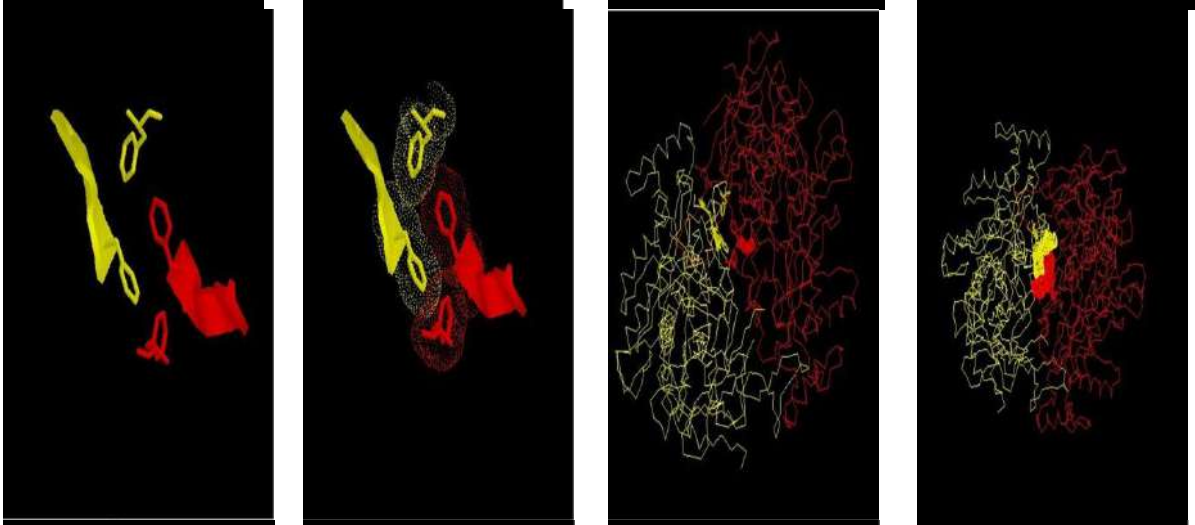
4mif « B »



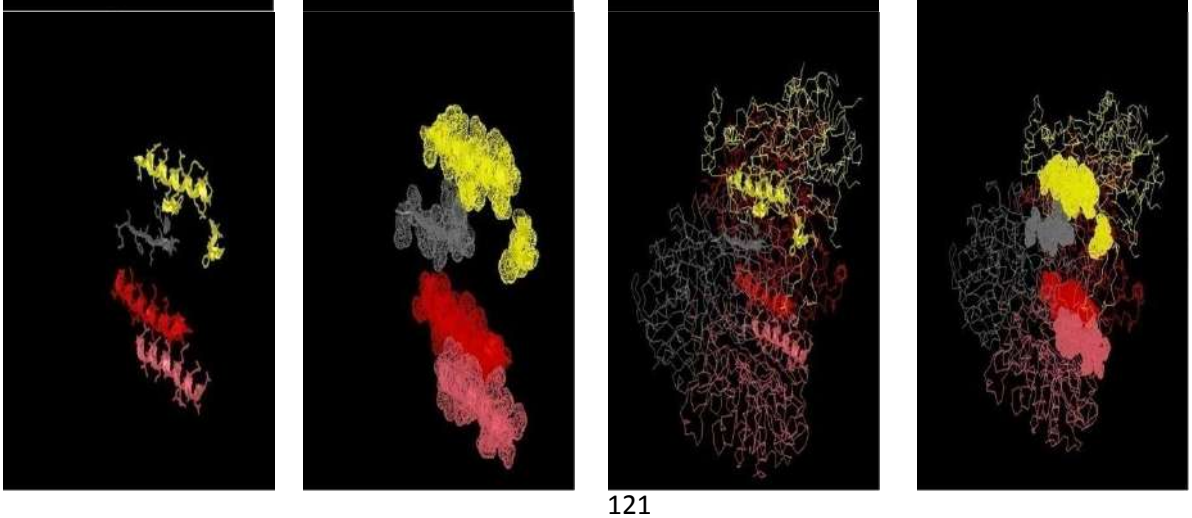
4mif « D»



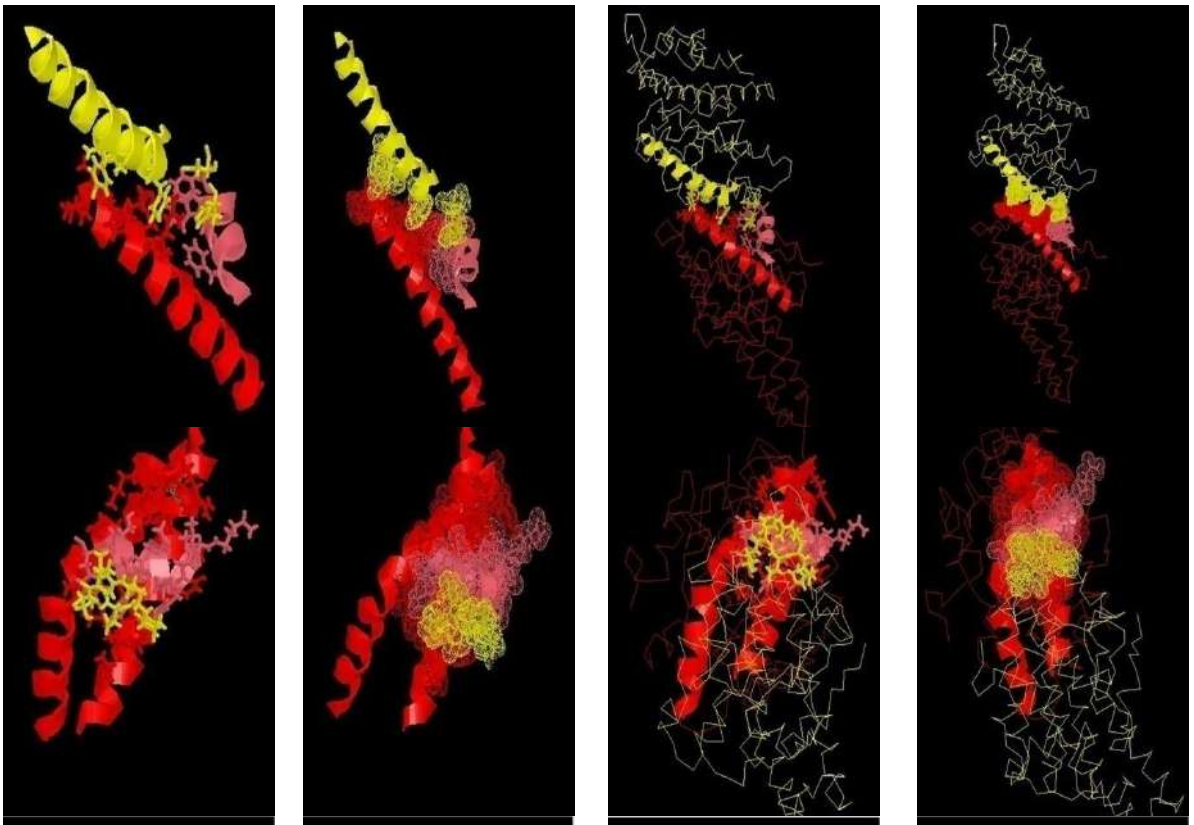
4mig « A »



4mig « C »

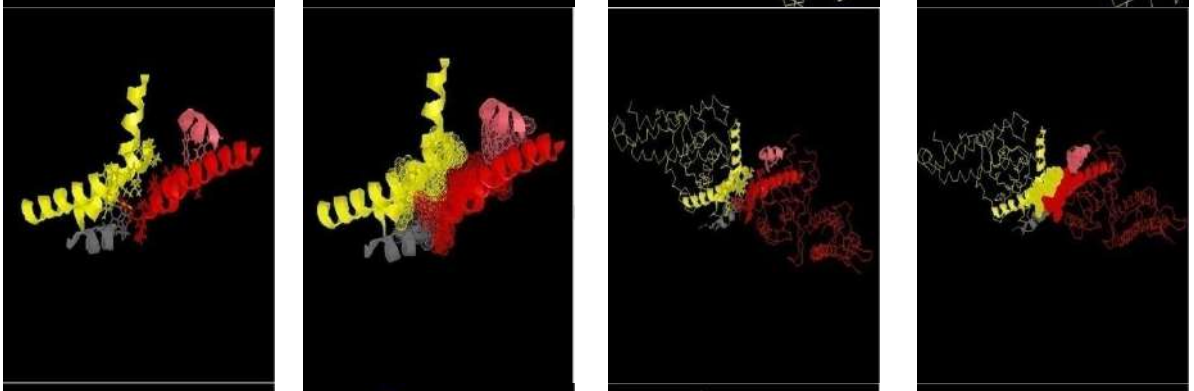


4nuu « A »

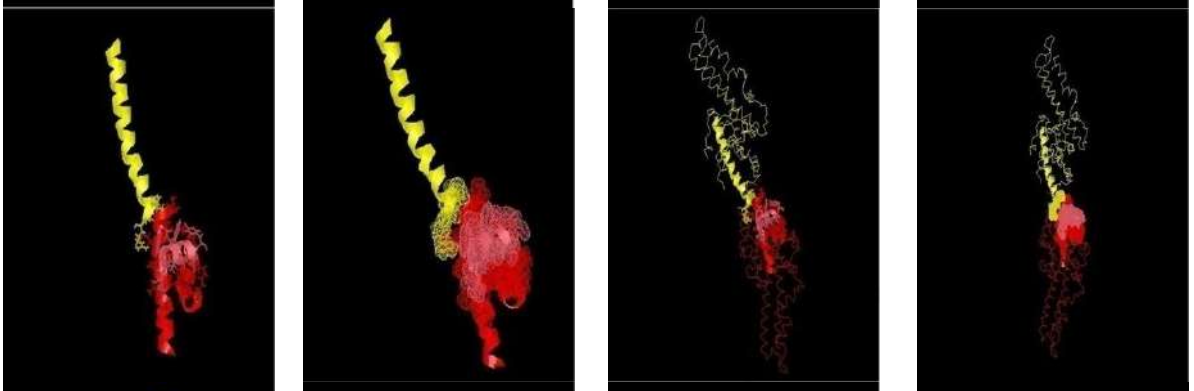


4nuu « C »

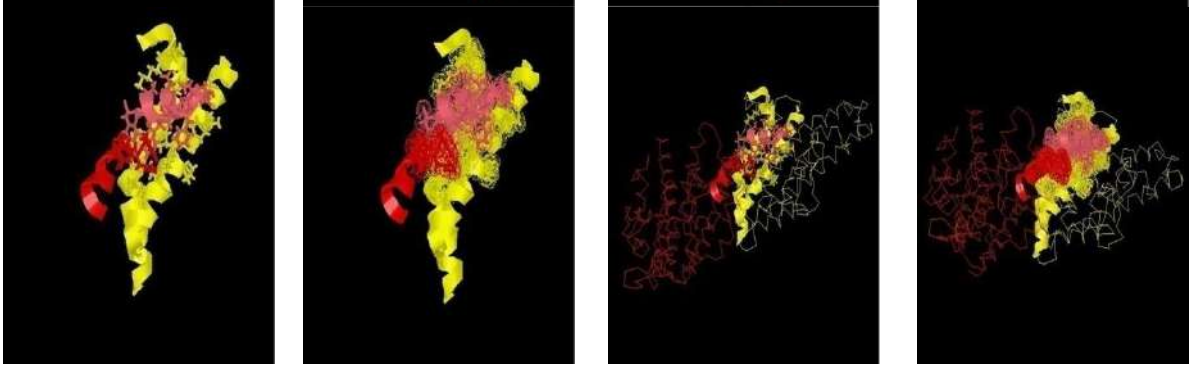
4nuv « A »



4nuv « C »



4nuv « D »

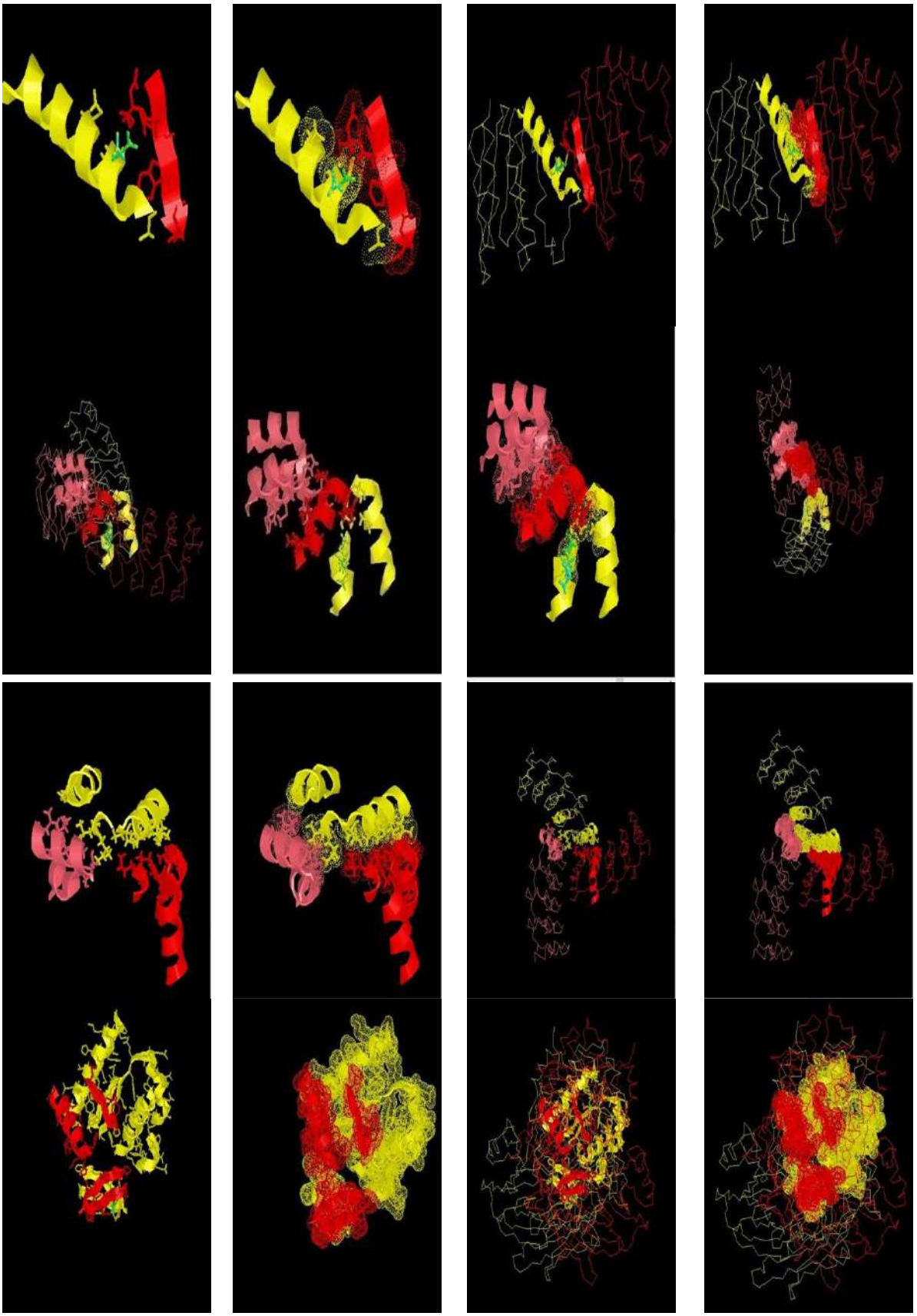


4p5d

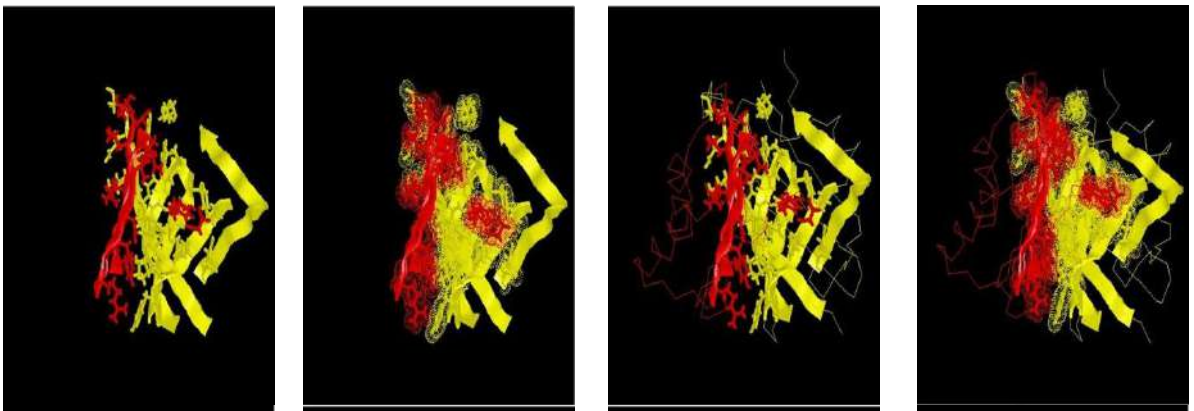
5eil « A »

5eil « B »

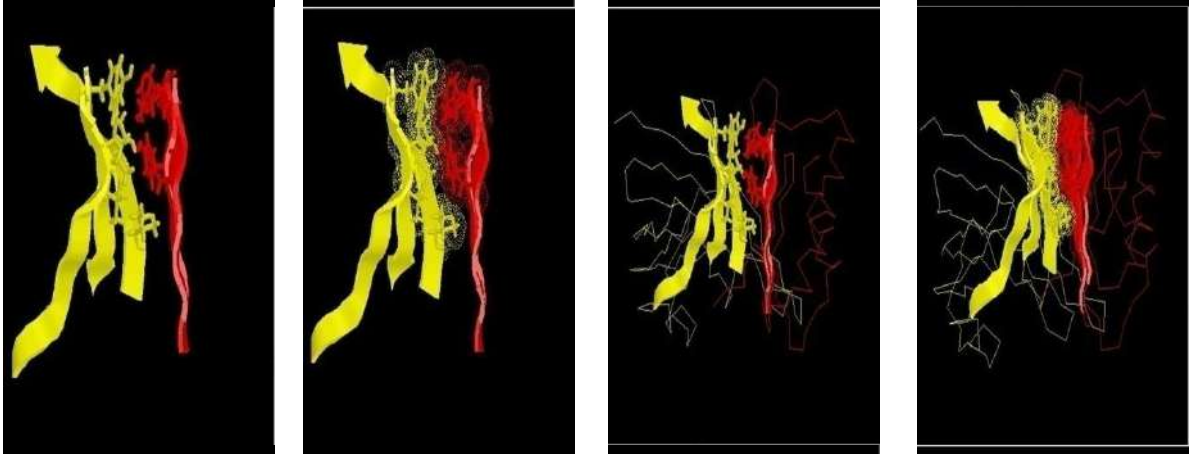
5i7u



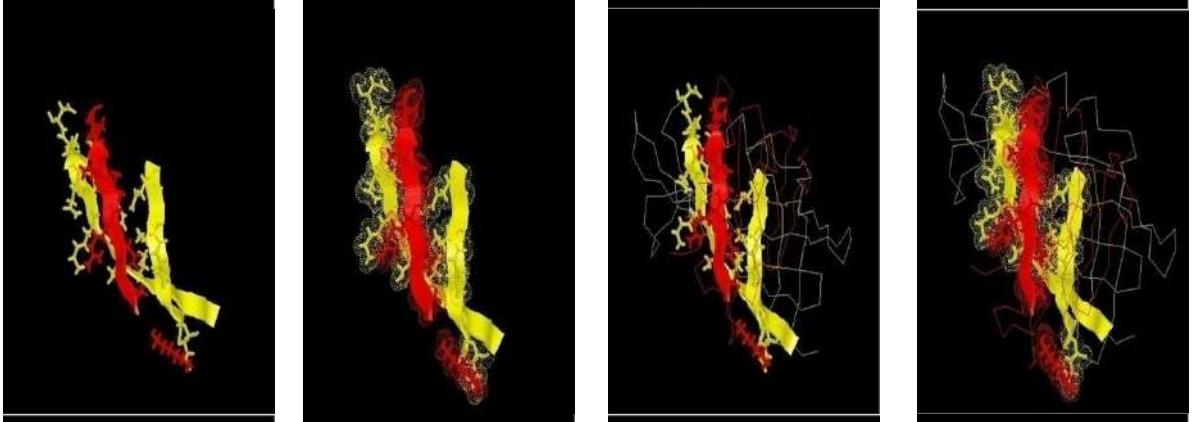
5ev



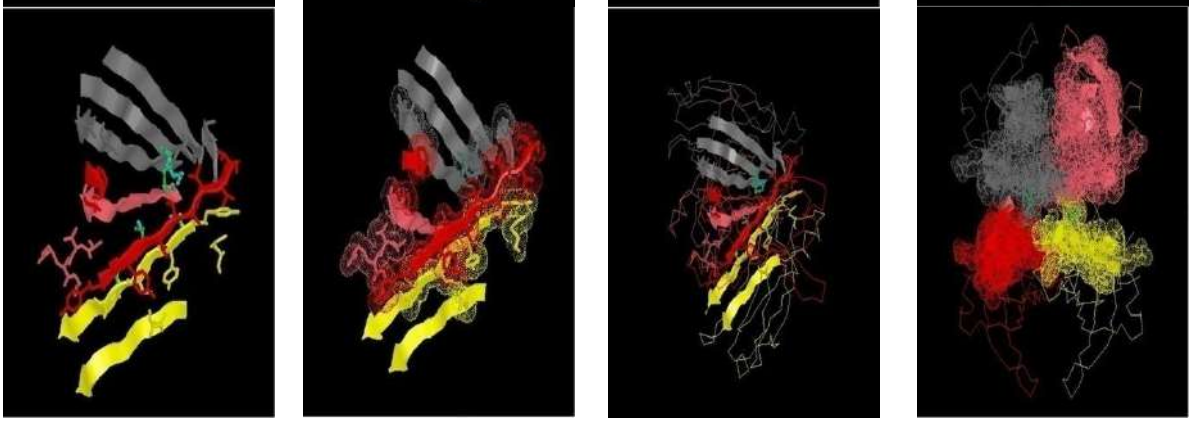
5kew_A200-209_218-221



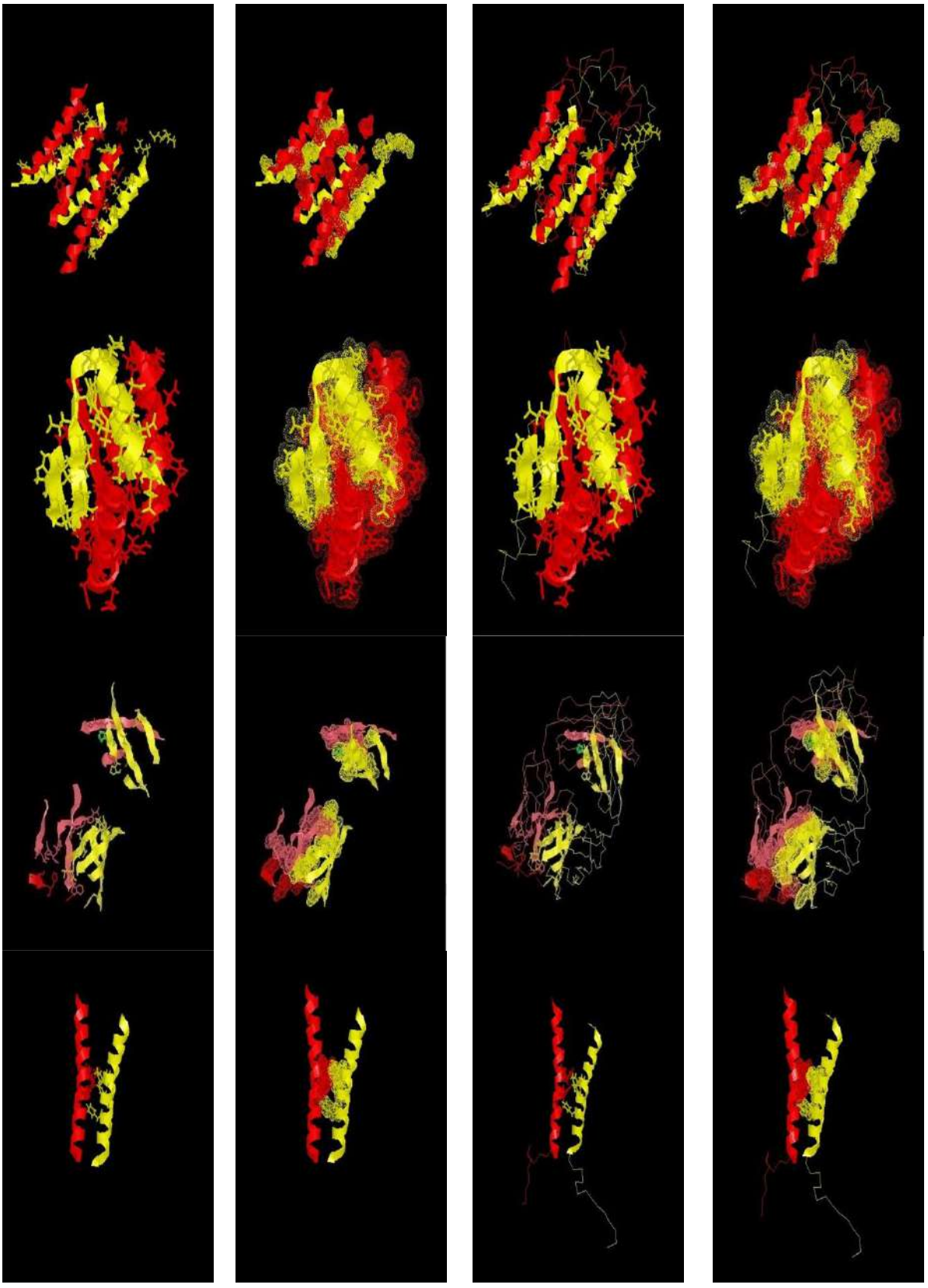
5kew_A222-234



5llv



5nf8

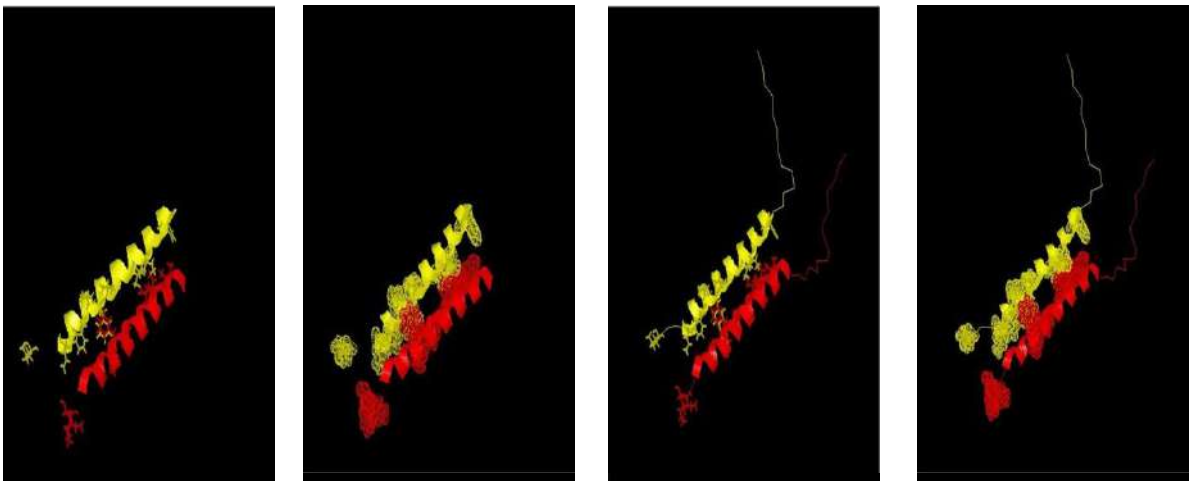


5noc

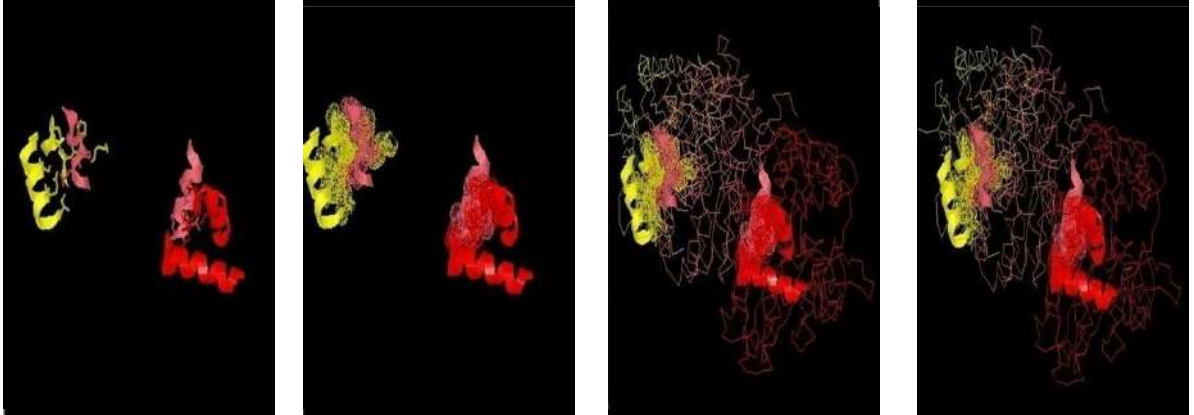
5ock

5oek

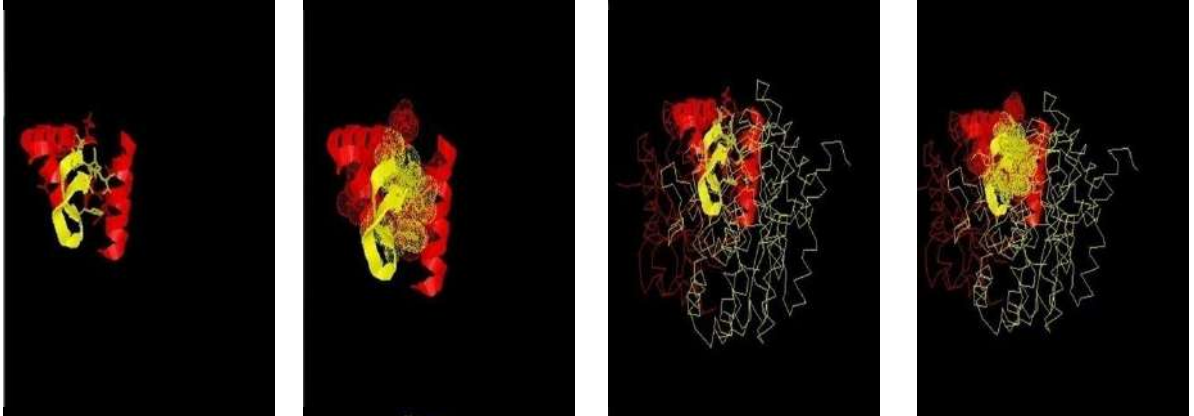
5ohd



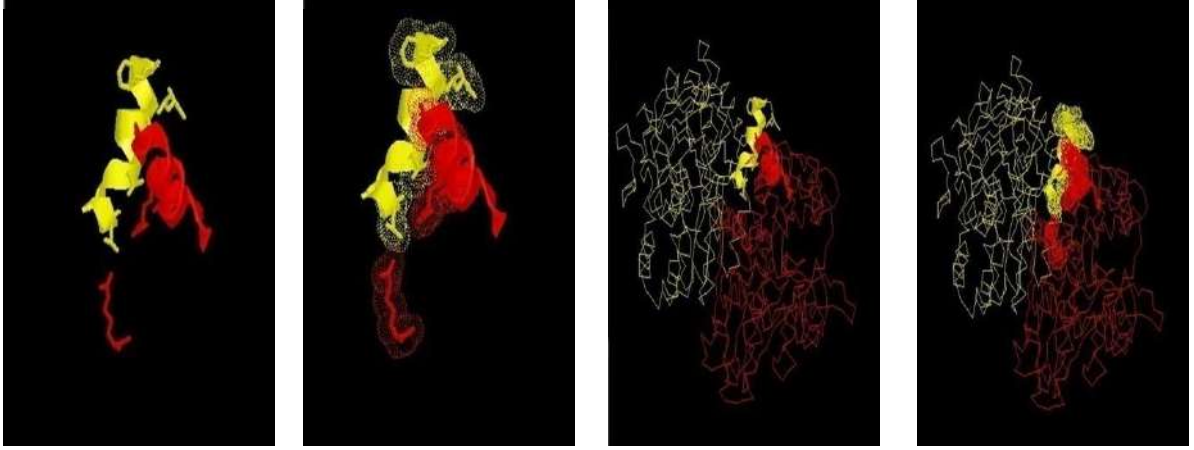
5tcf_D28-34_156-168



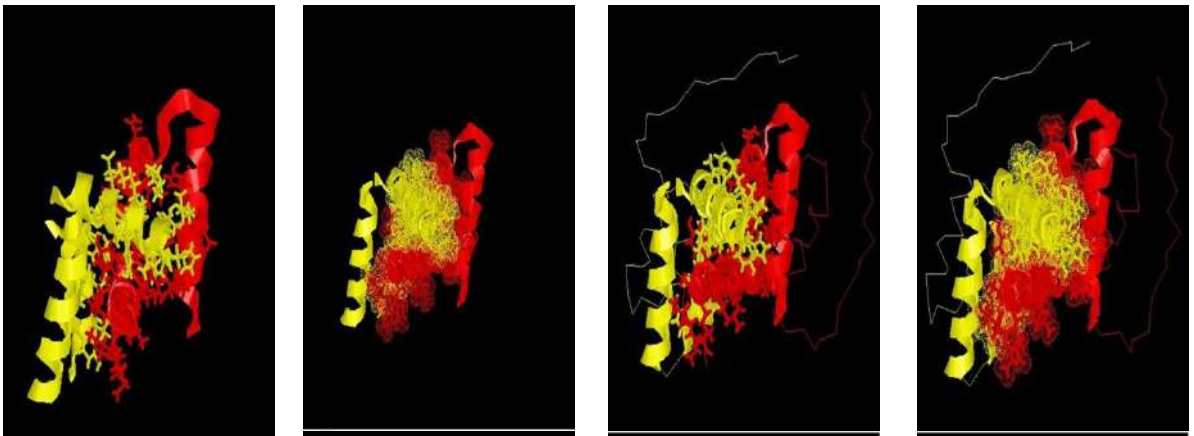
5tcf_D289-308_303-307



5tcf_D394-407

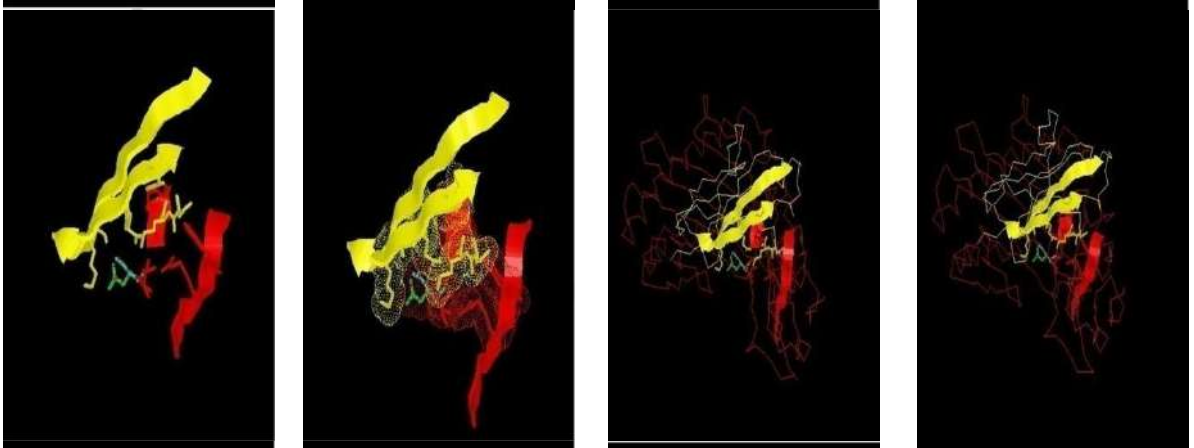


5us5

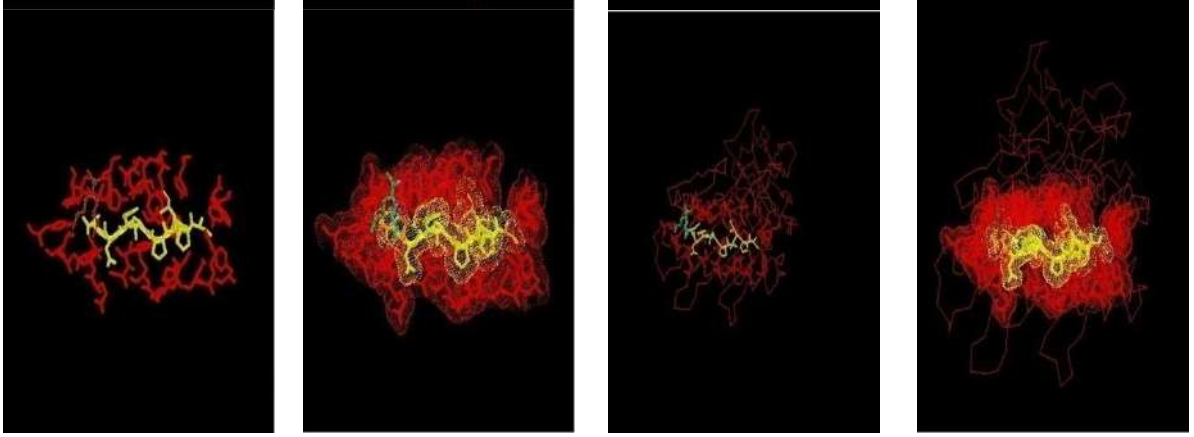


5wmn_A 30-119

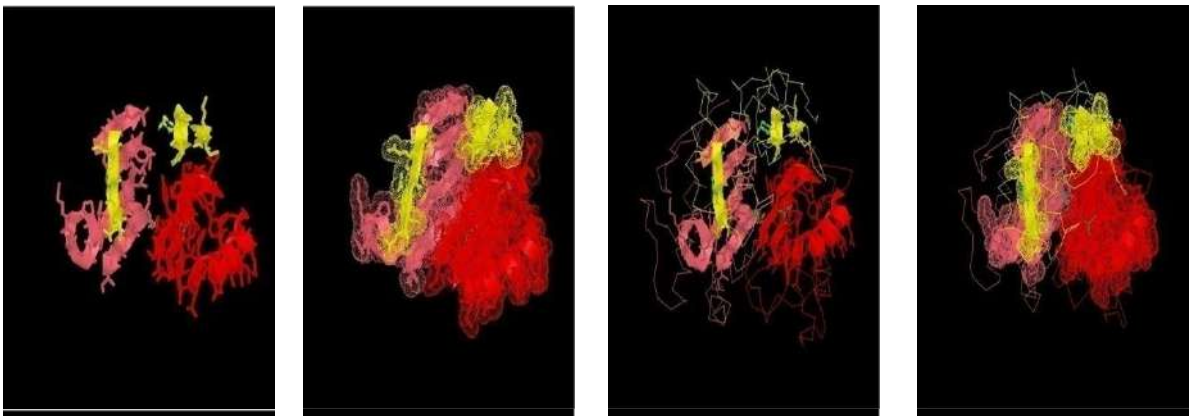
5wmn_A 194-236



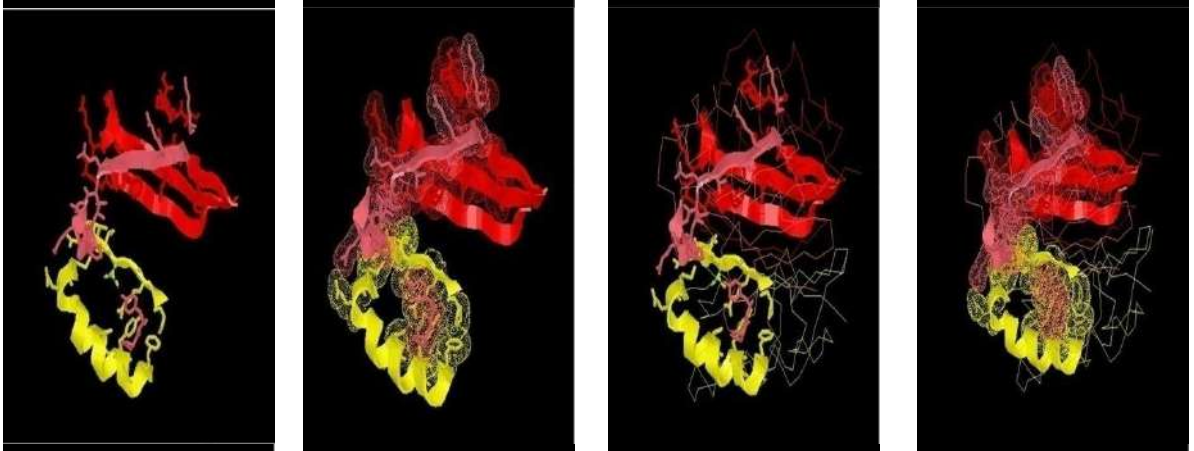
5wmn « C »



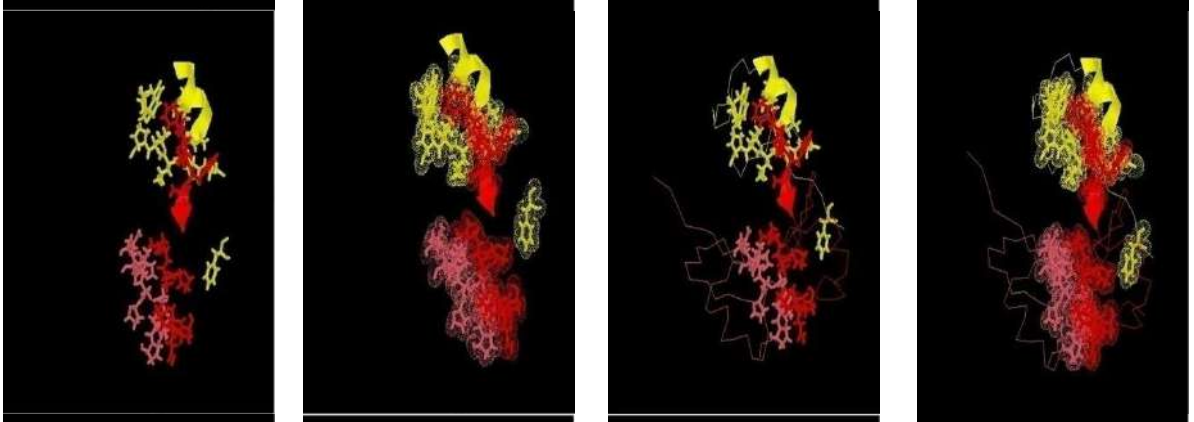
5y6u « B »



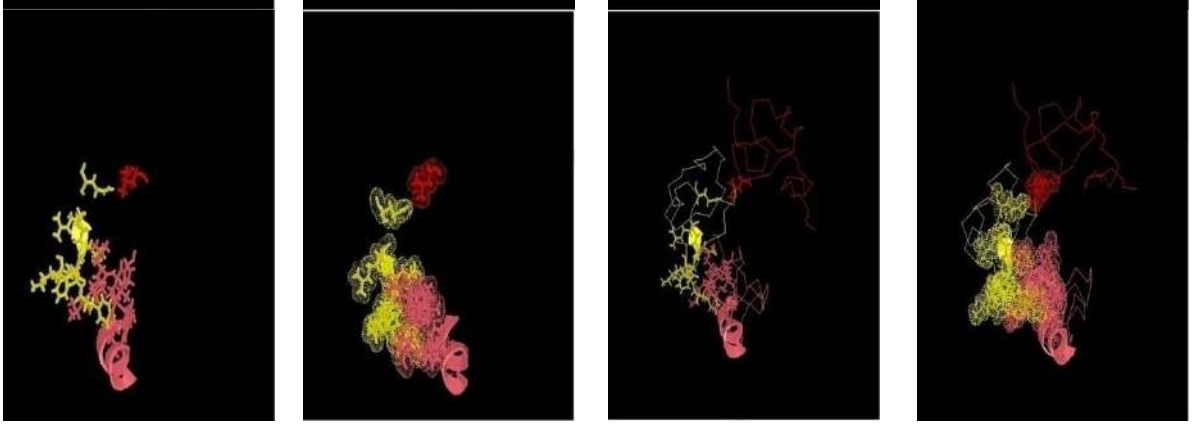
5y6u « C »



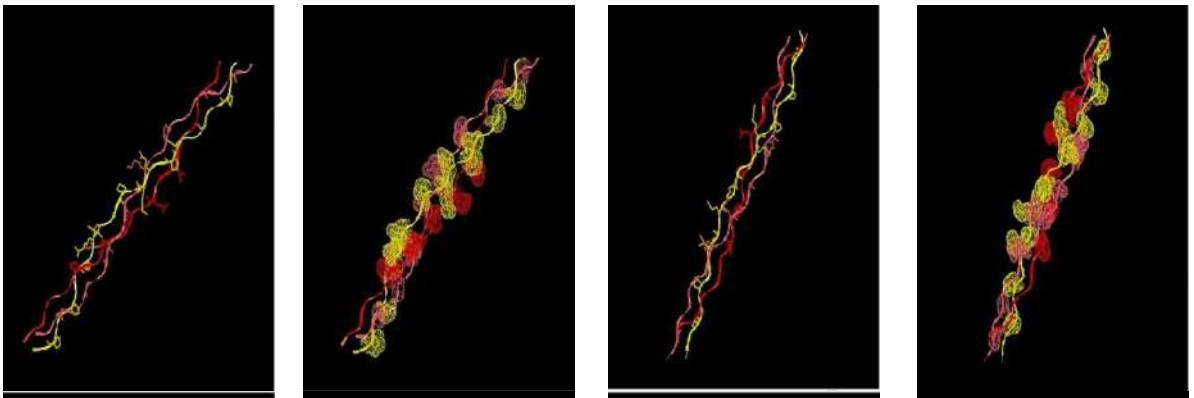
5yuf « A »



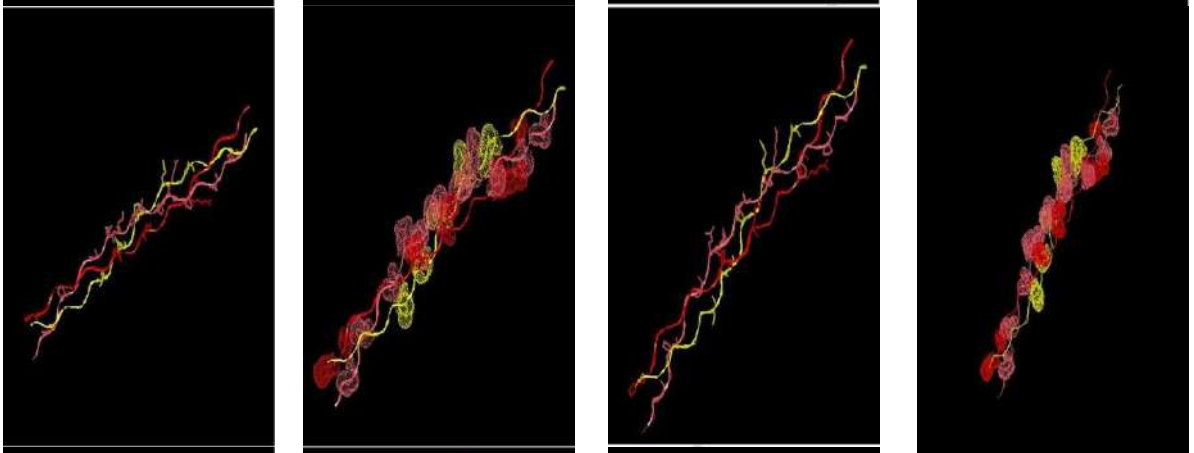
5yuf « C »



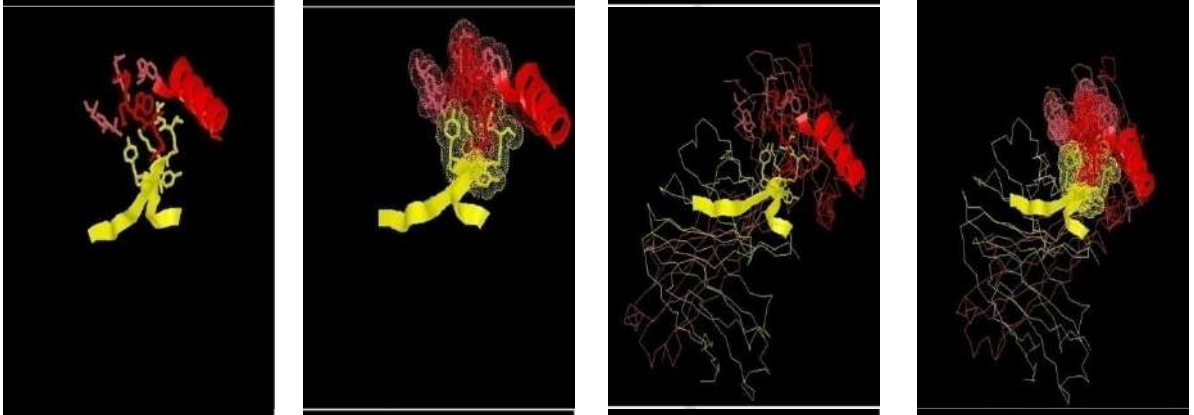
6a0a « B »



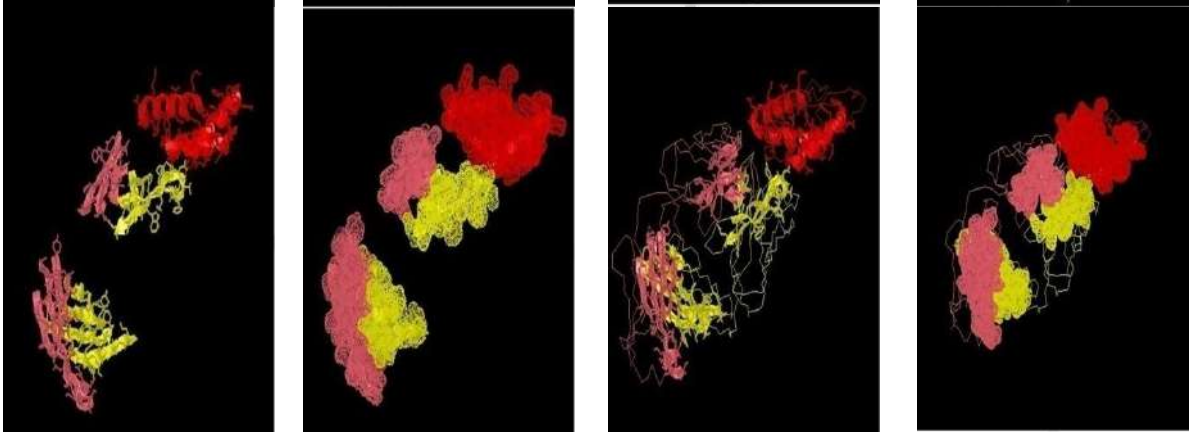
6a0a « C »



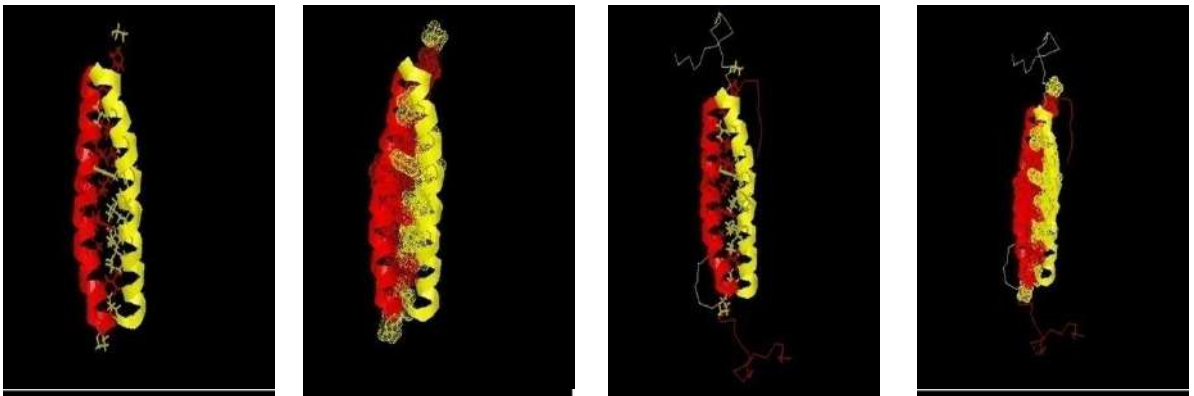
6bfs « C »



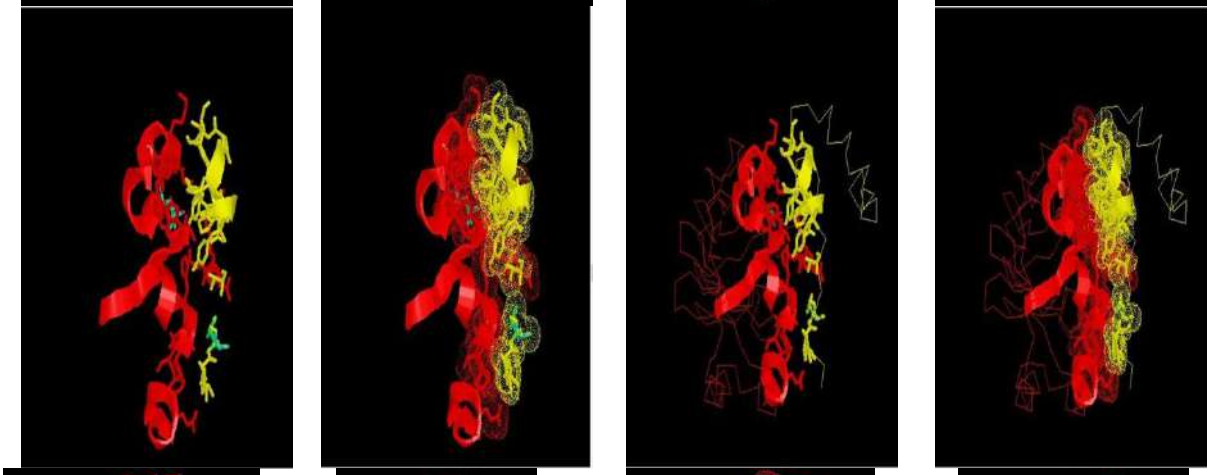
6bfs « H »



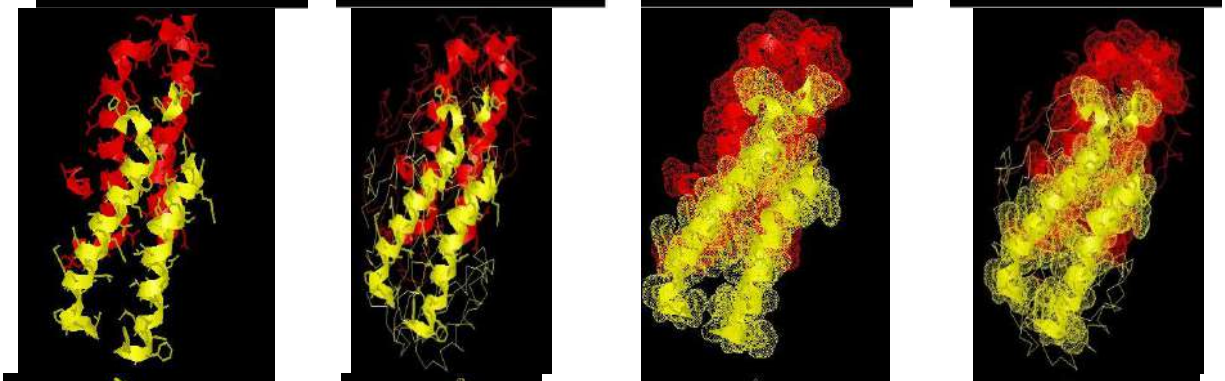
6E4H



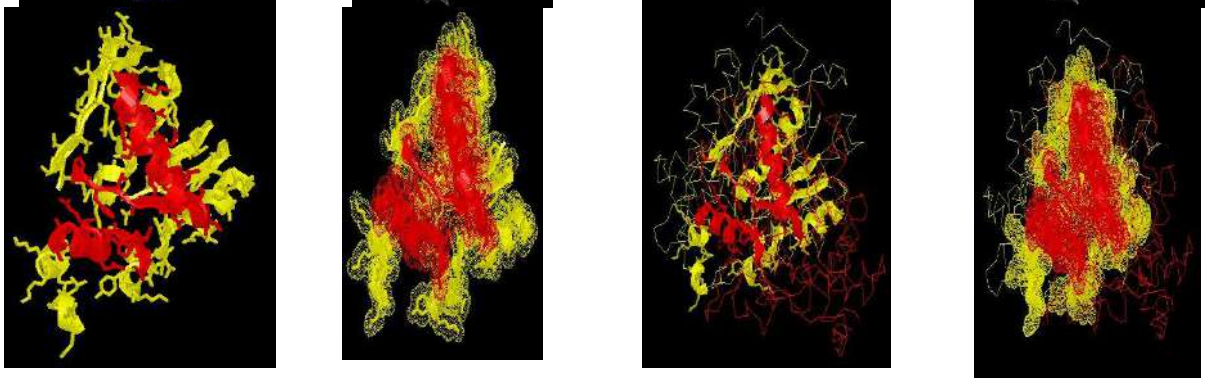
6MS4



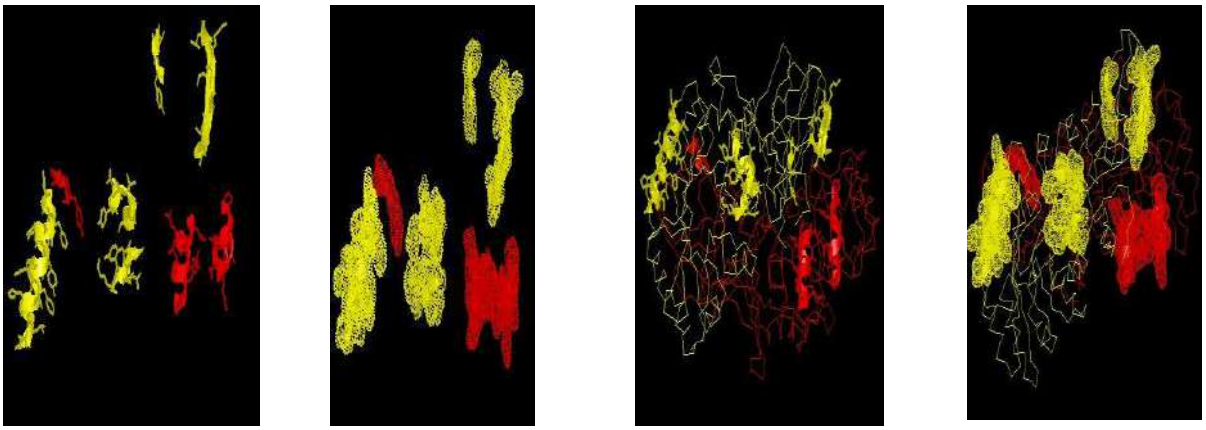
1a4u



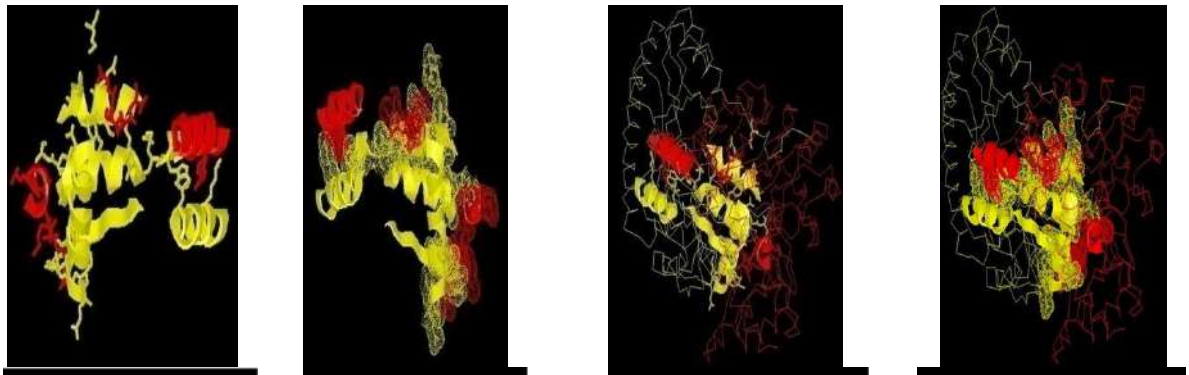
1afw



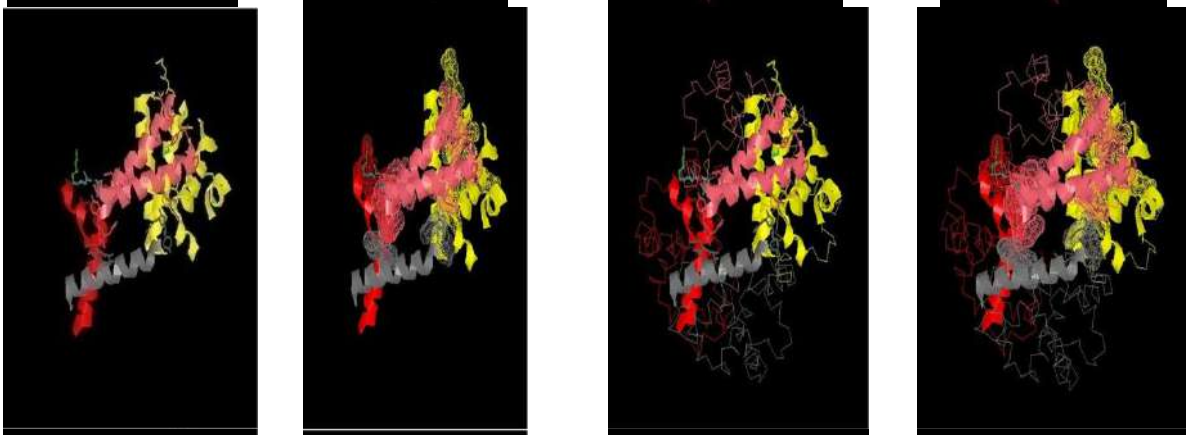
1bd0



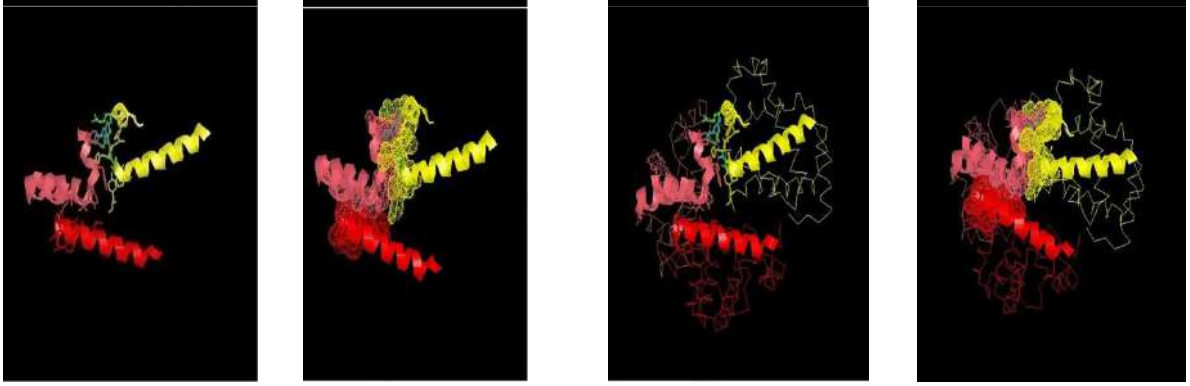
1bjw



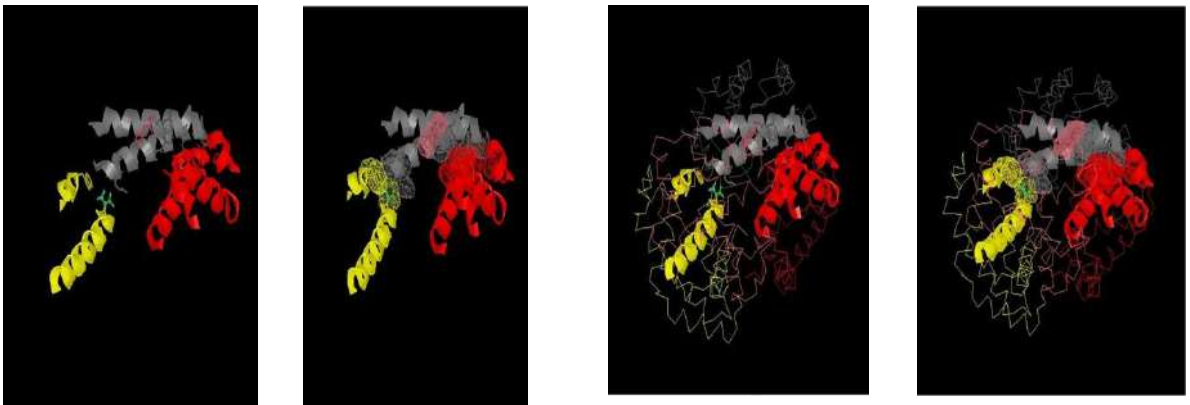
1cbm « C »



1cbm_D28-62_30-41_61-49

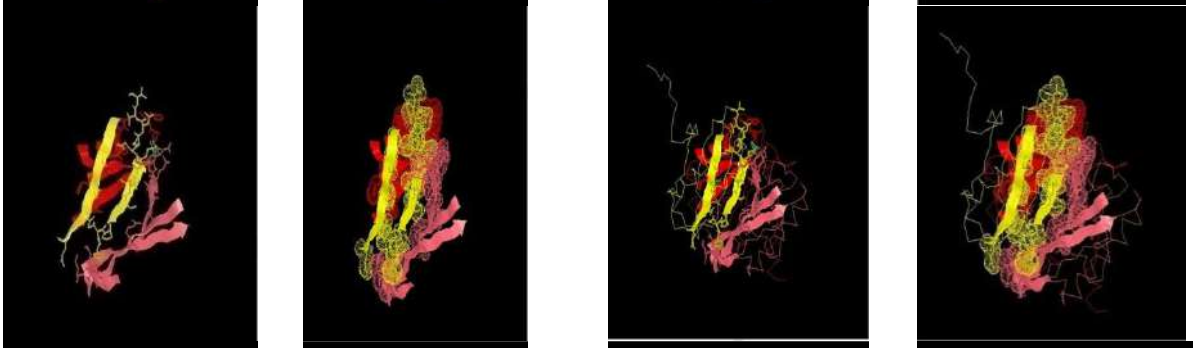


1cbm_D100-111_124-144

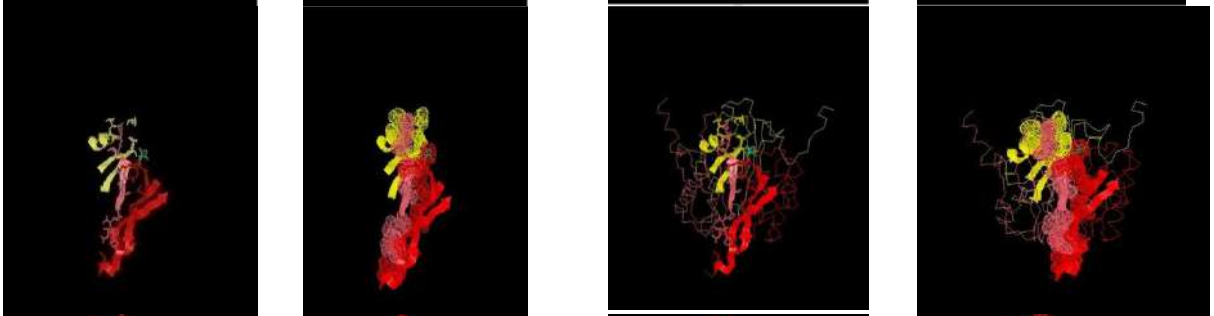


1cnz

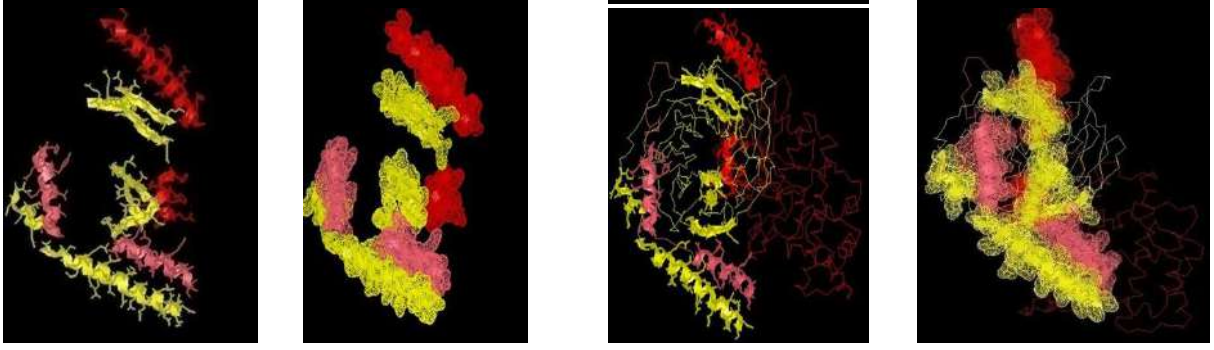
1dbf « B »



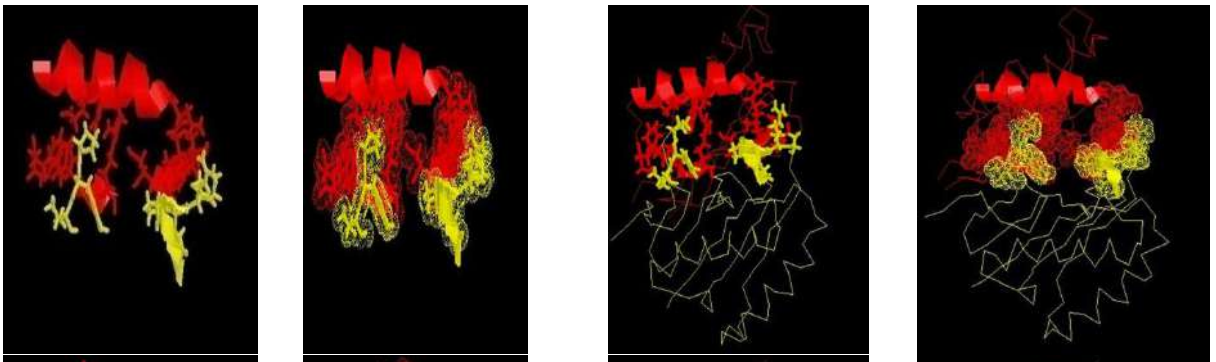
1dbf « C »



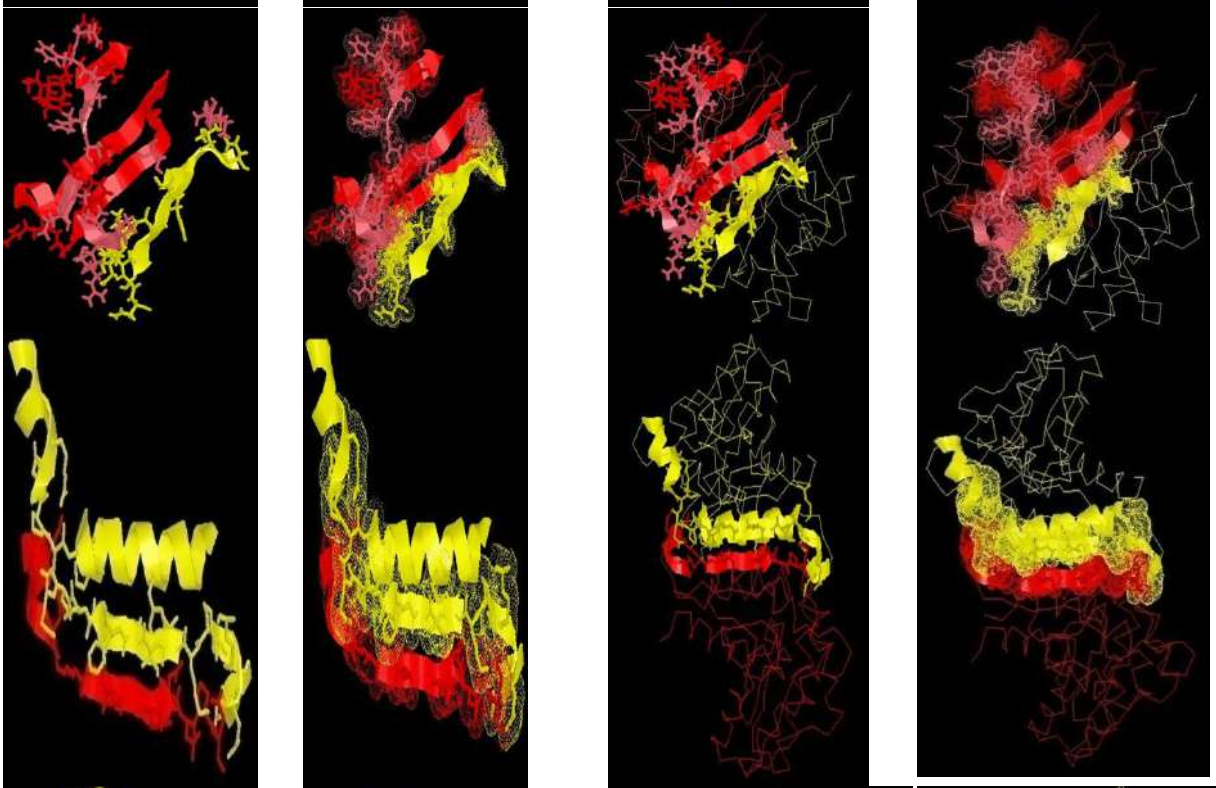
1gg2



1j7h « B »



1j7h « C »

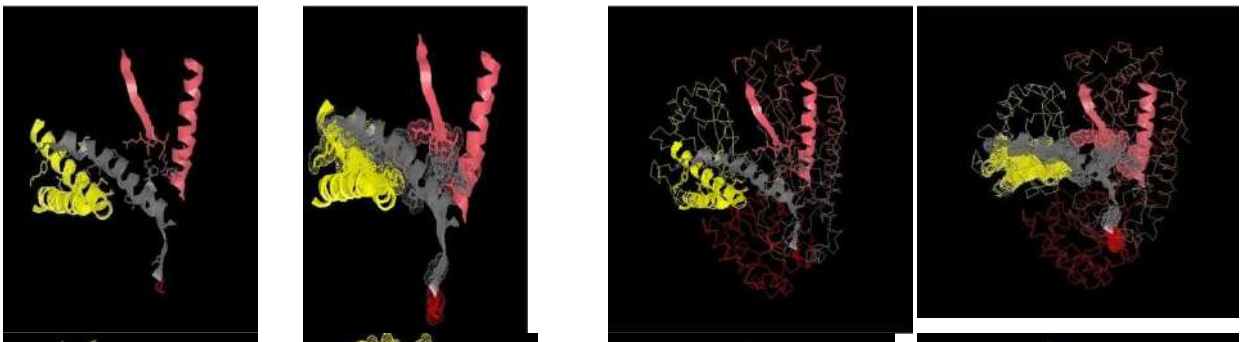


1jfl

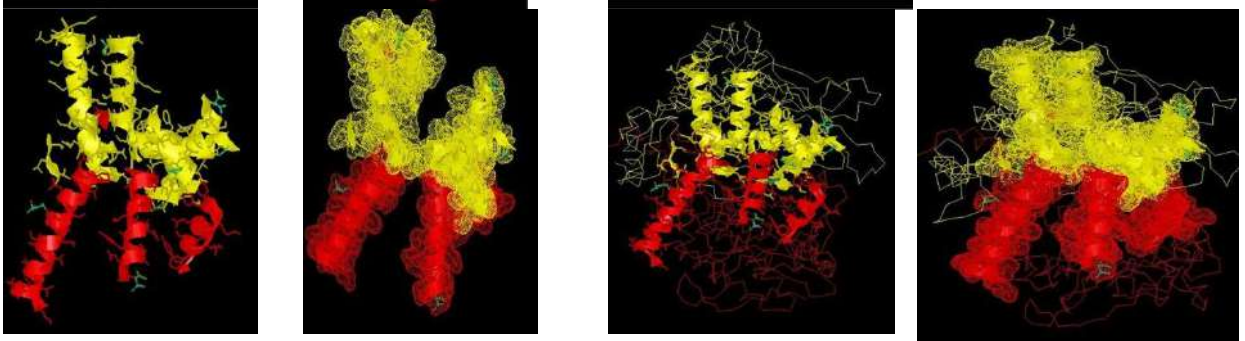
1jiw

1kkc « A »

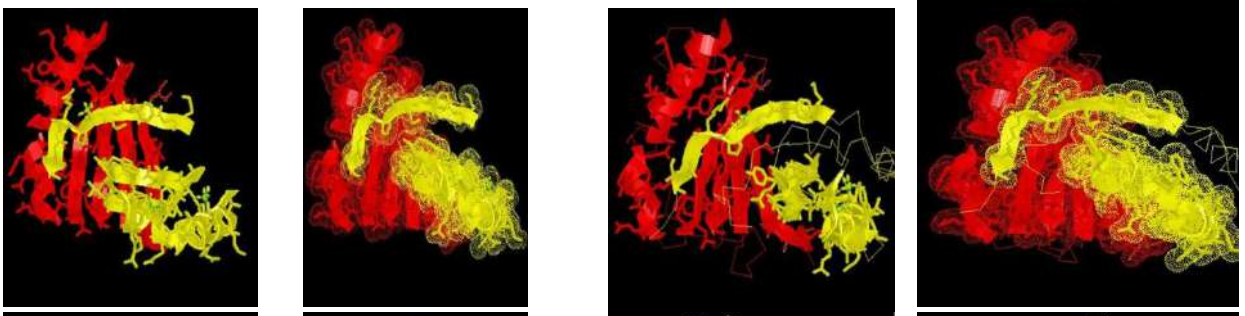
1kkc « Y »



1lk9

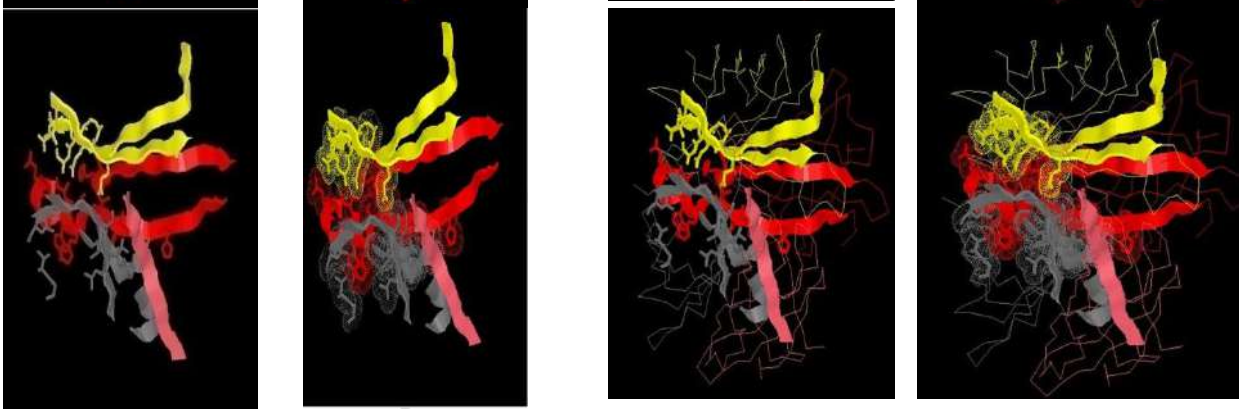


1lq9

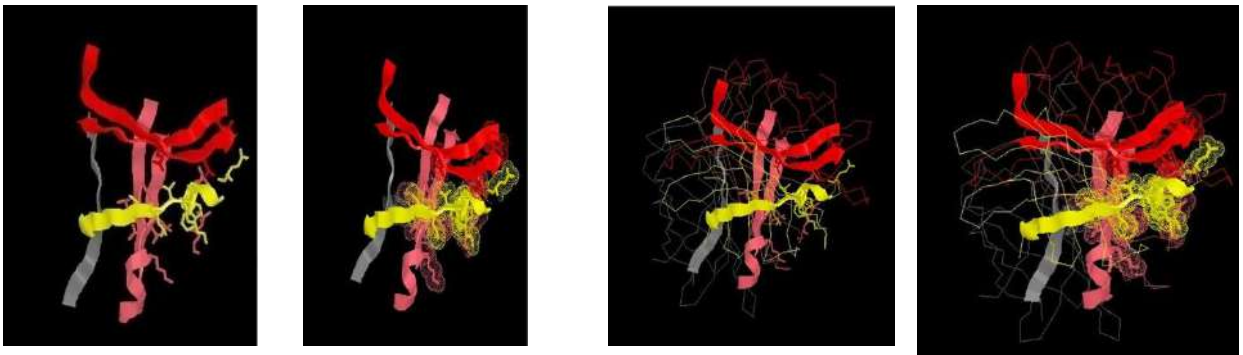


1m4i

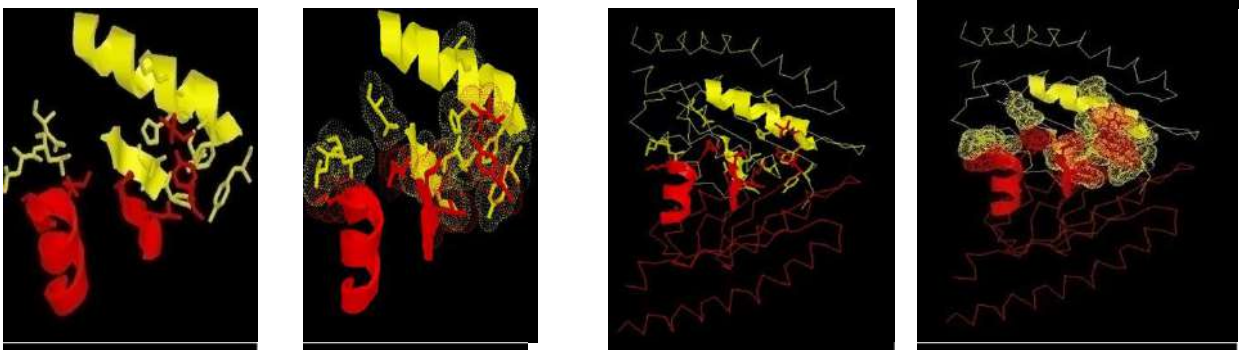
1mep « A »



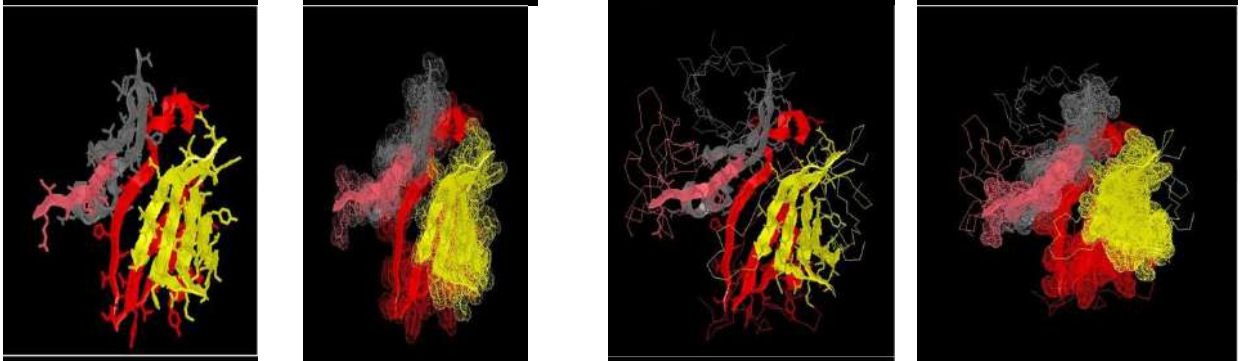
1mep « B »



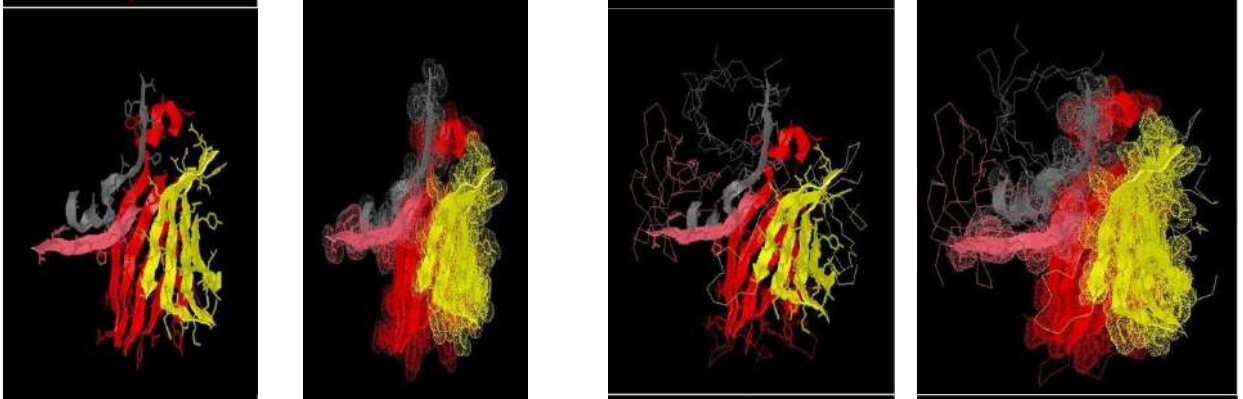
1mjh



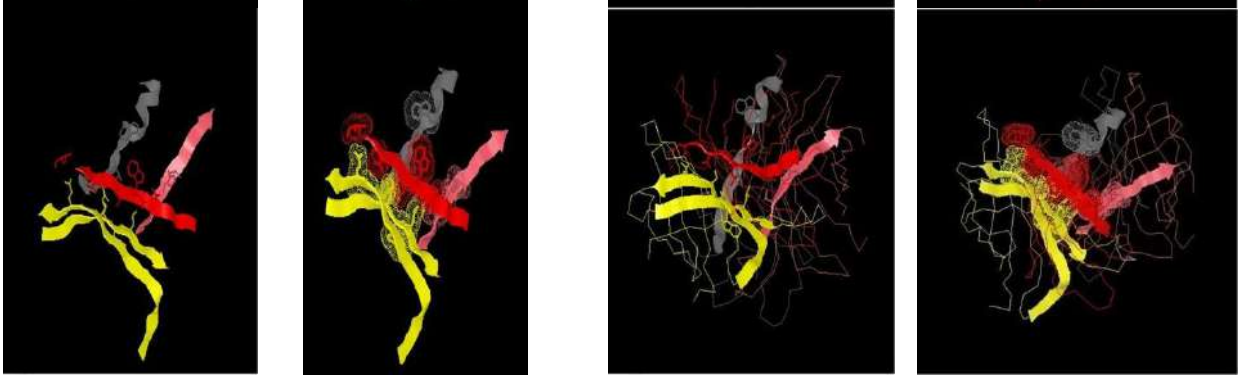
1n7y_A35-105_100-104_115-127_117-129



1n7y_A44-123_53-95

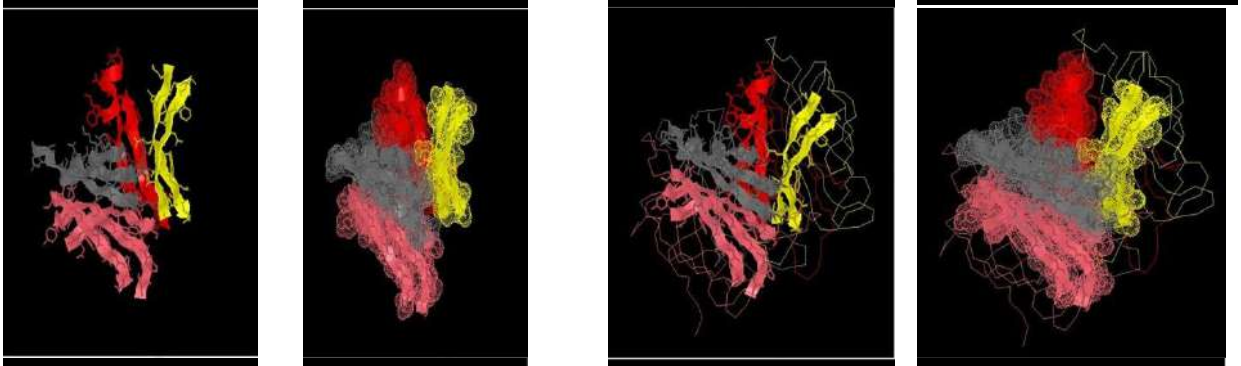


1n7y_A107-115_117-127



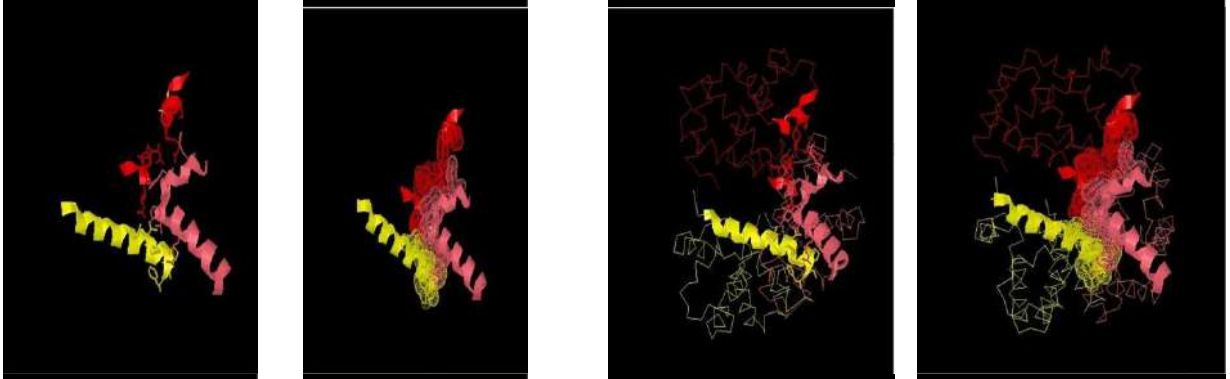
1n9m « A »

1n9m « D »

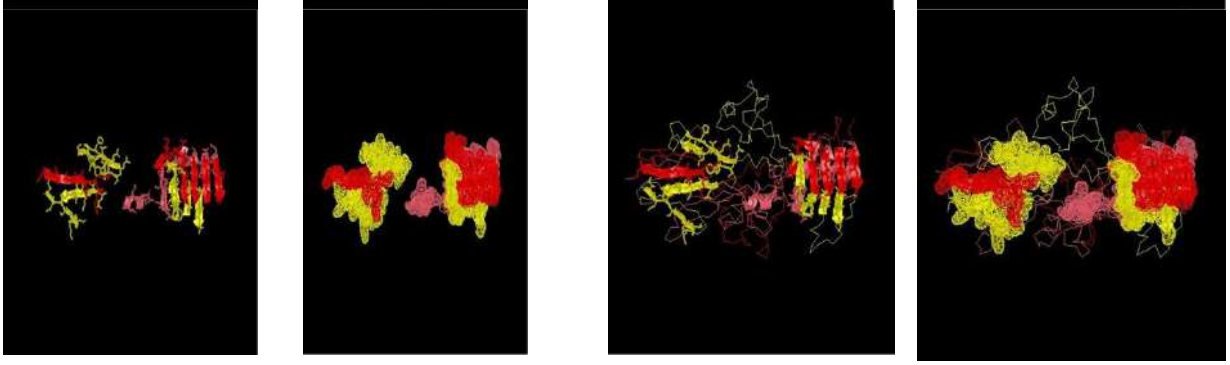


1s5y « B »

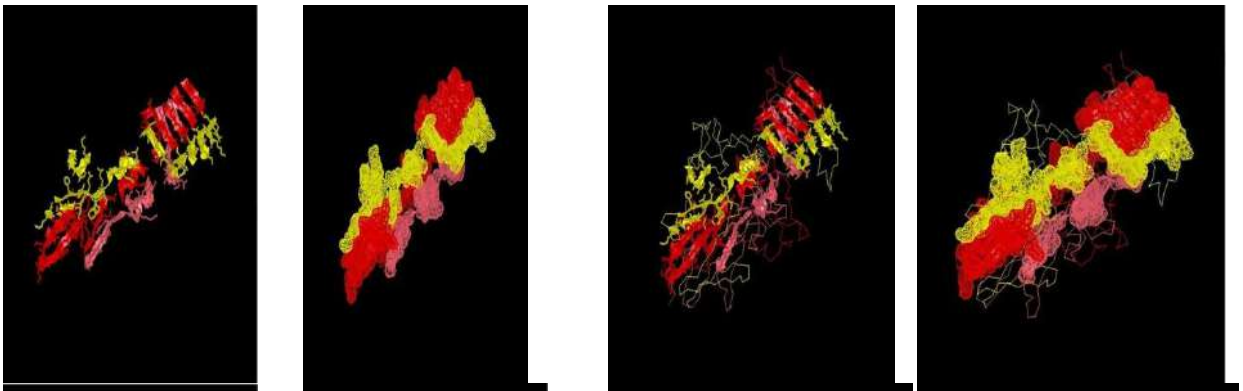
1s5y « D »



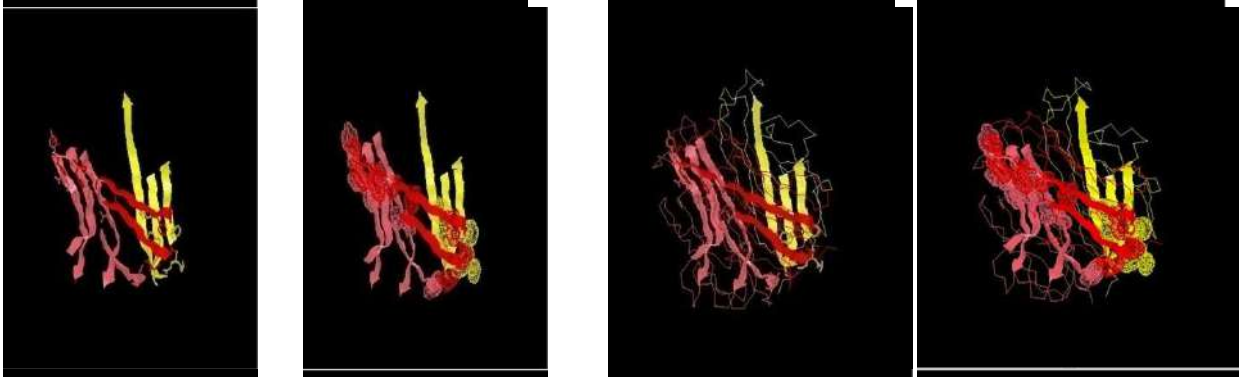
1s7m « A »



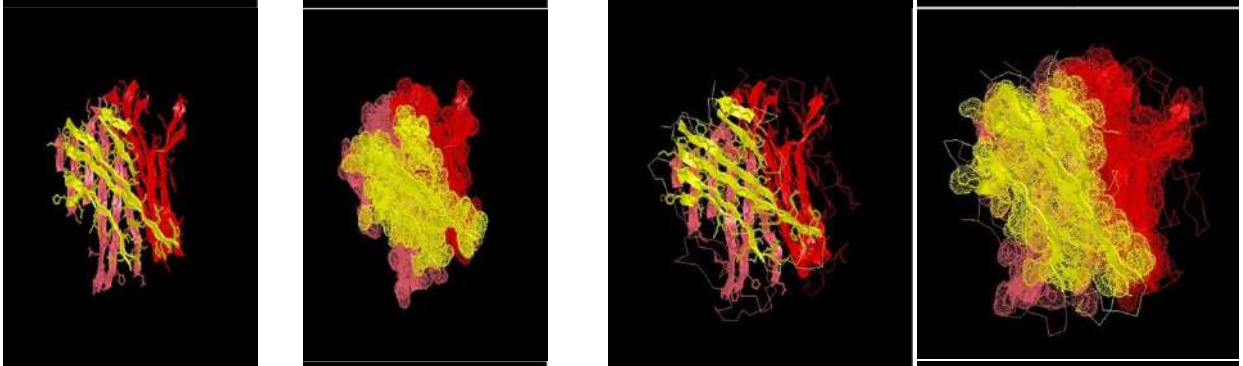
1s7m « B »



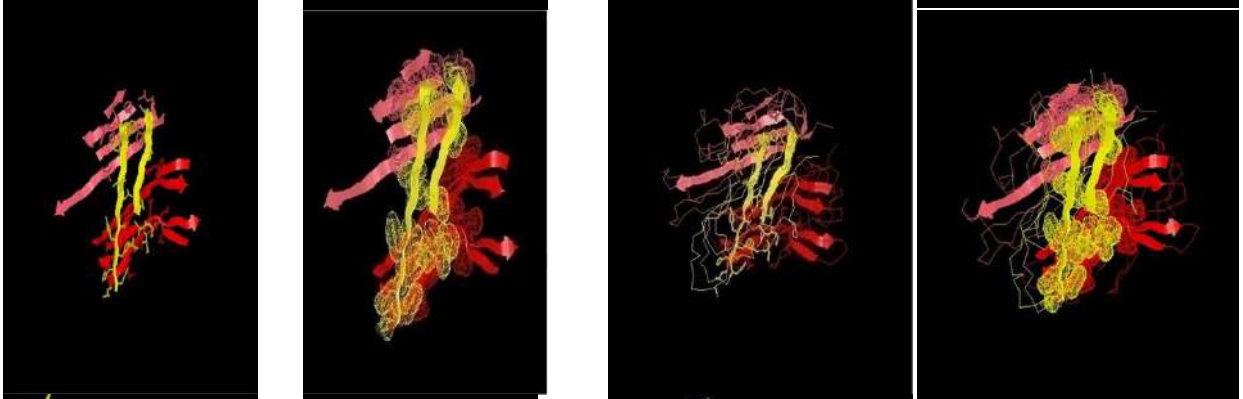
1s55



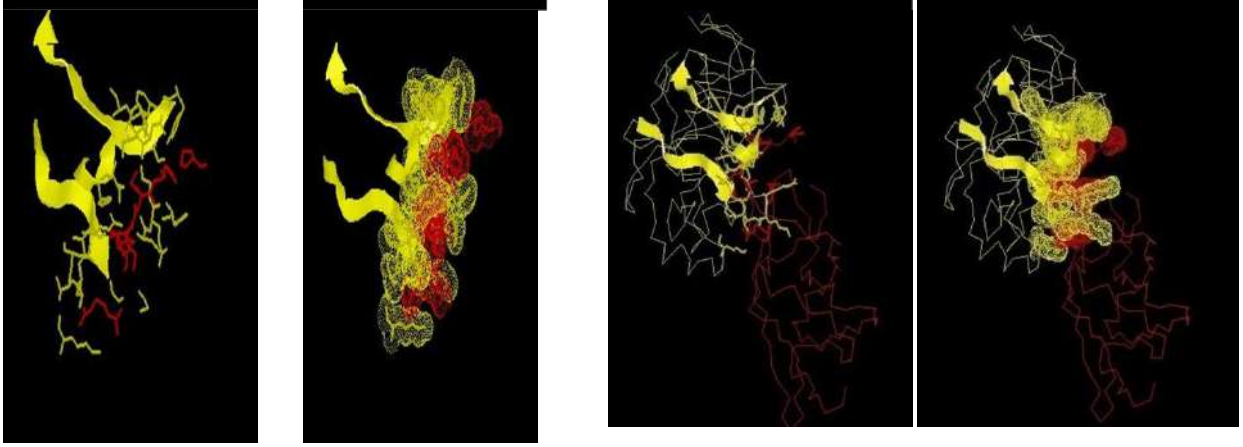
1s55_B162-281_194-189_267-273



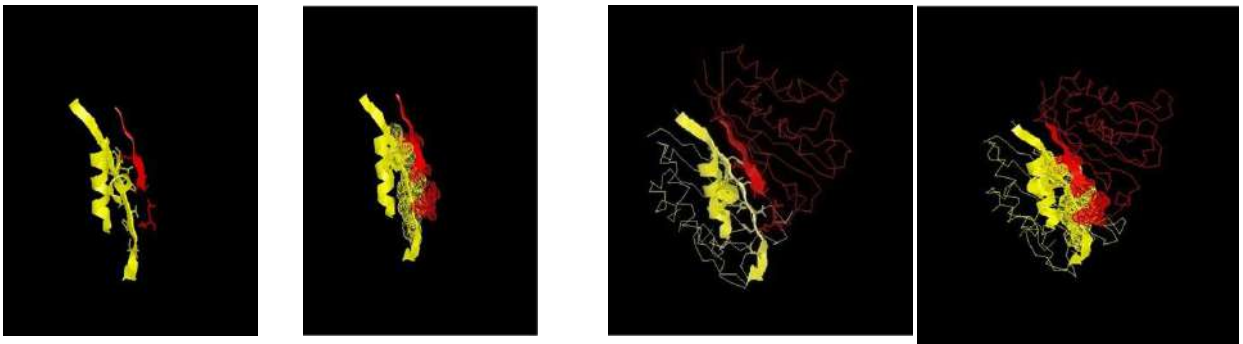
1s55_B267-313_300-304



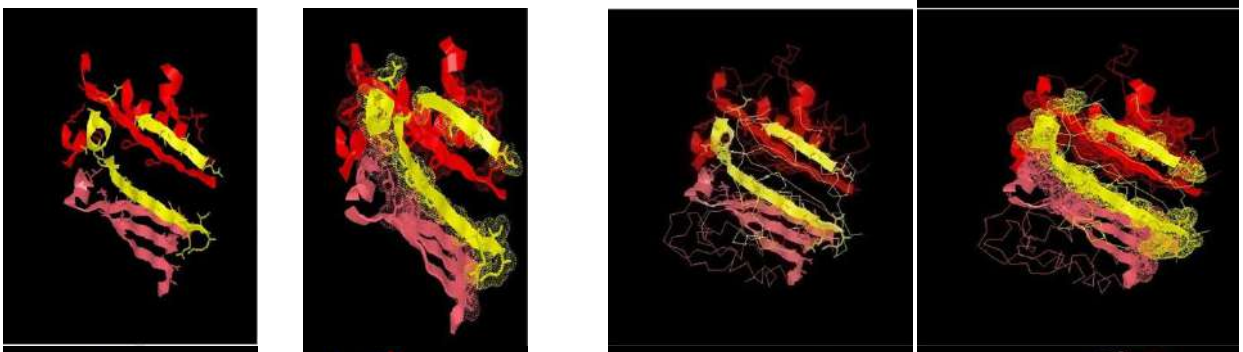
1slu



1t0a « A »

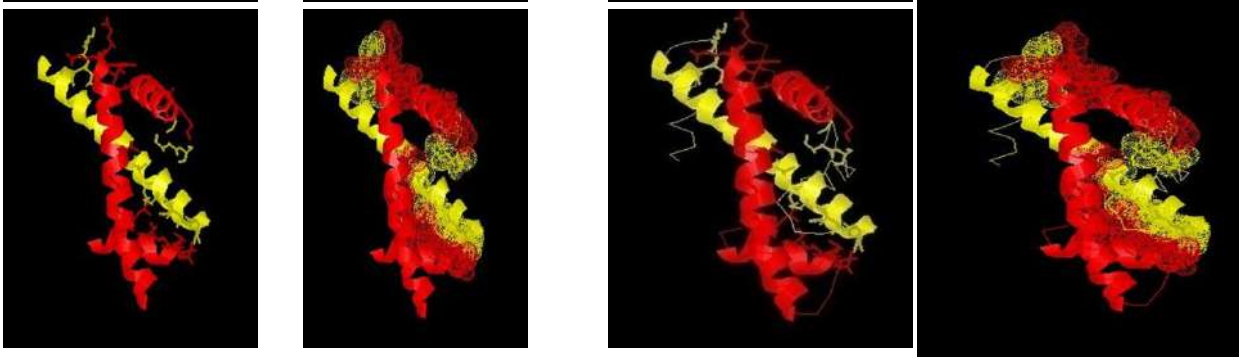


1t0a « B»

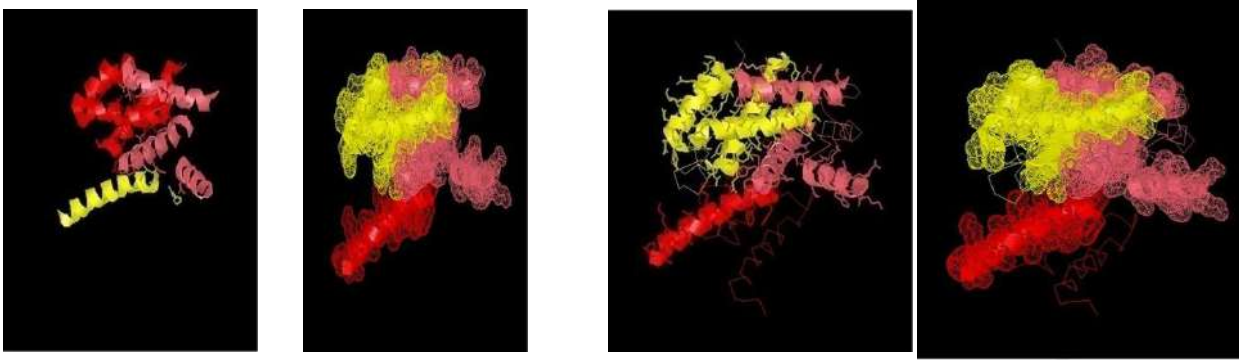


1taf « A »

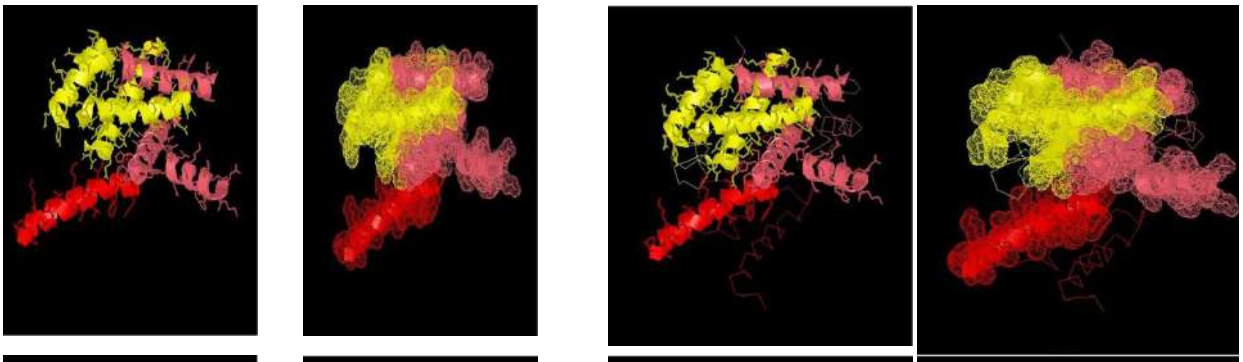
1taf « B »



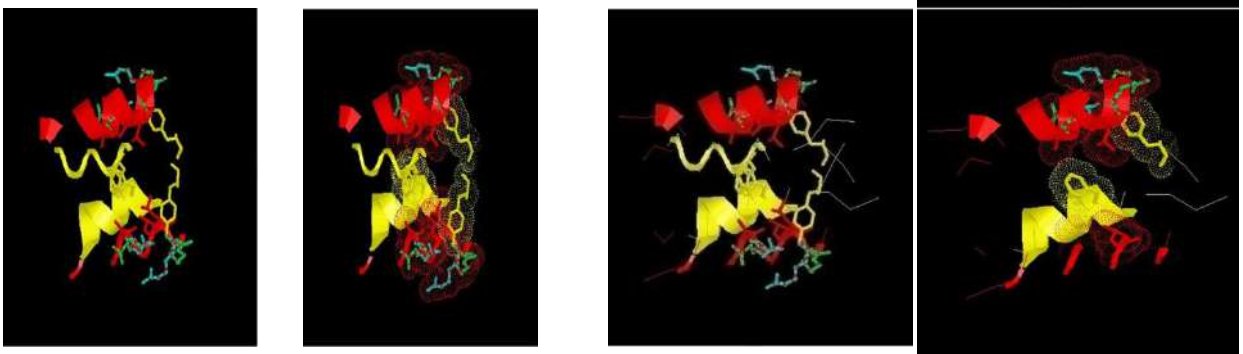
1xk4 « C »



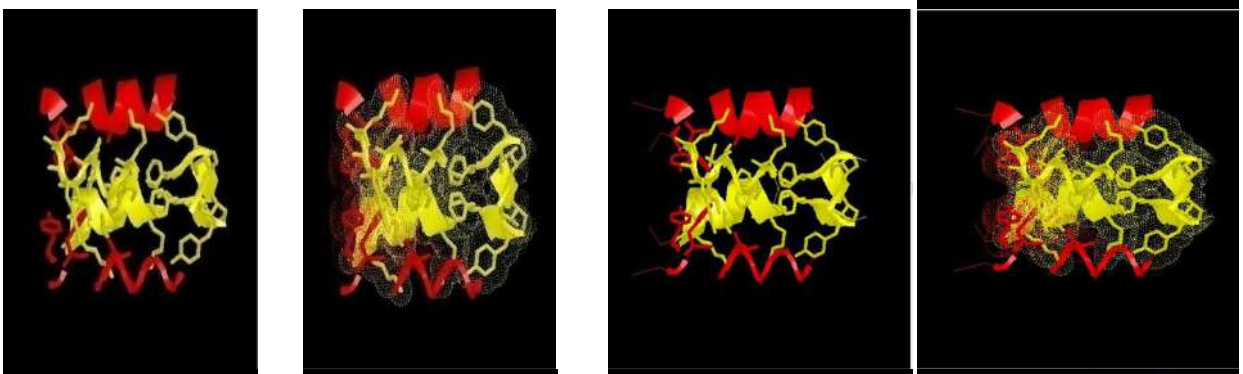
1xk4 « D »



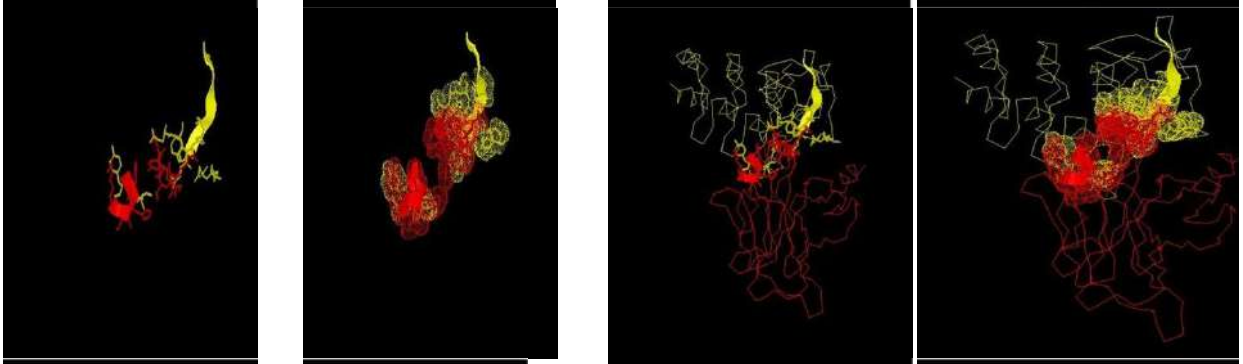
1xof « A »



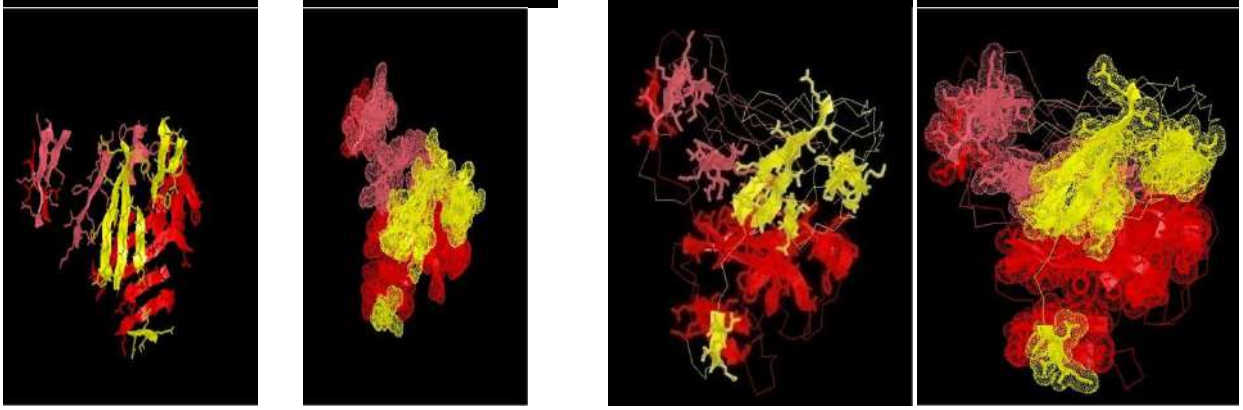
1xof « B »



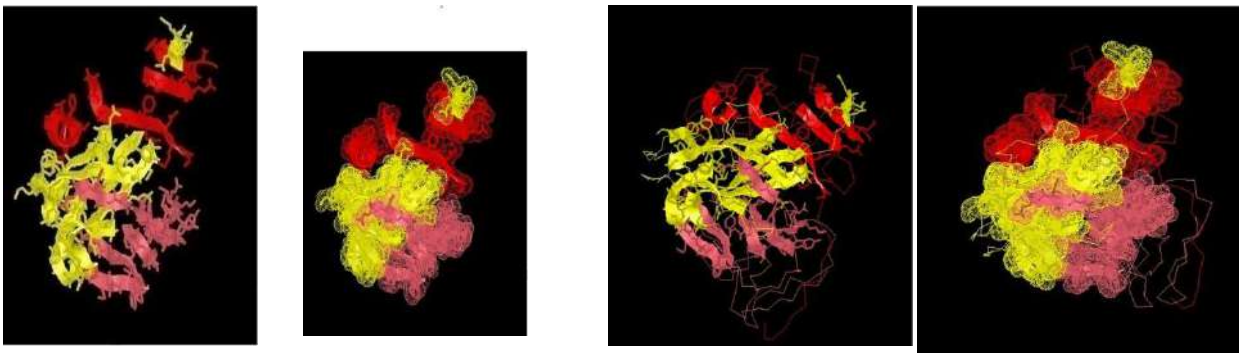
1ycs



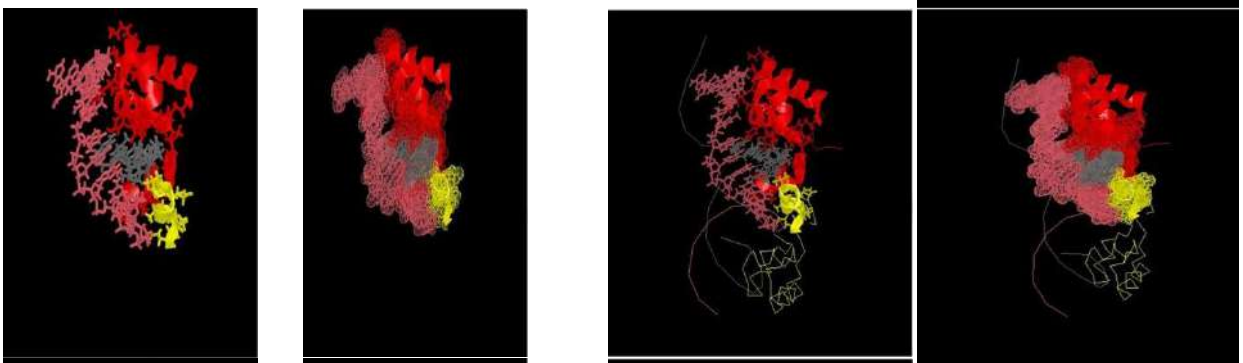
2baz « A »



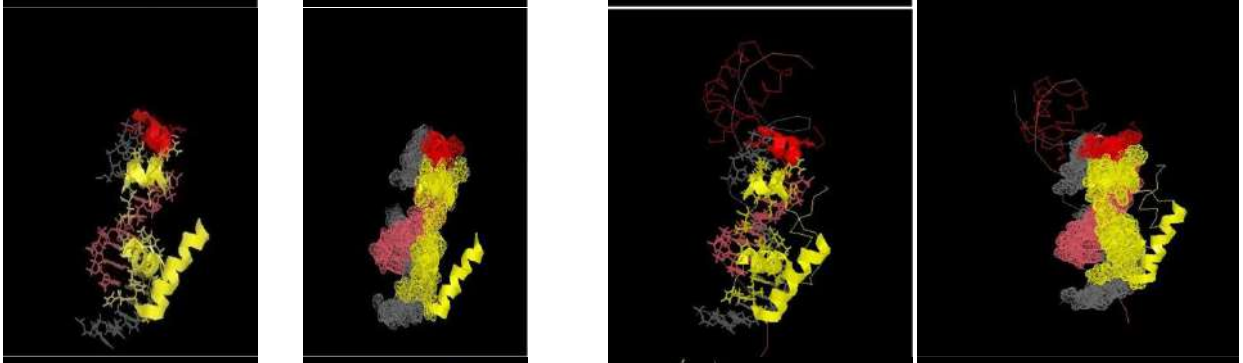
2baz « B »



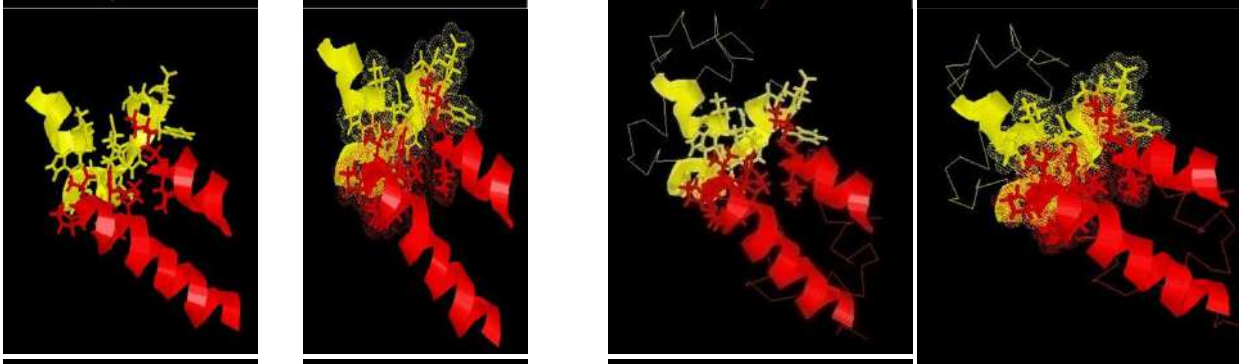
2kei « A »



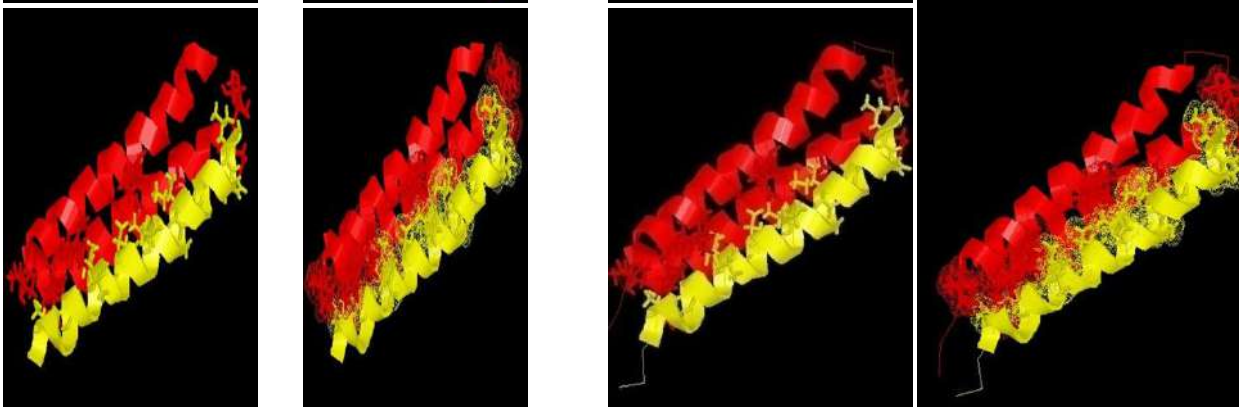
2kei « B »



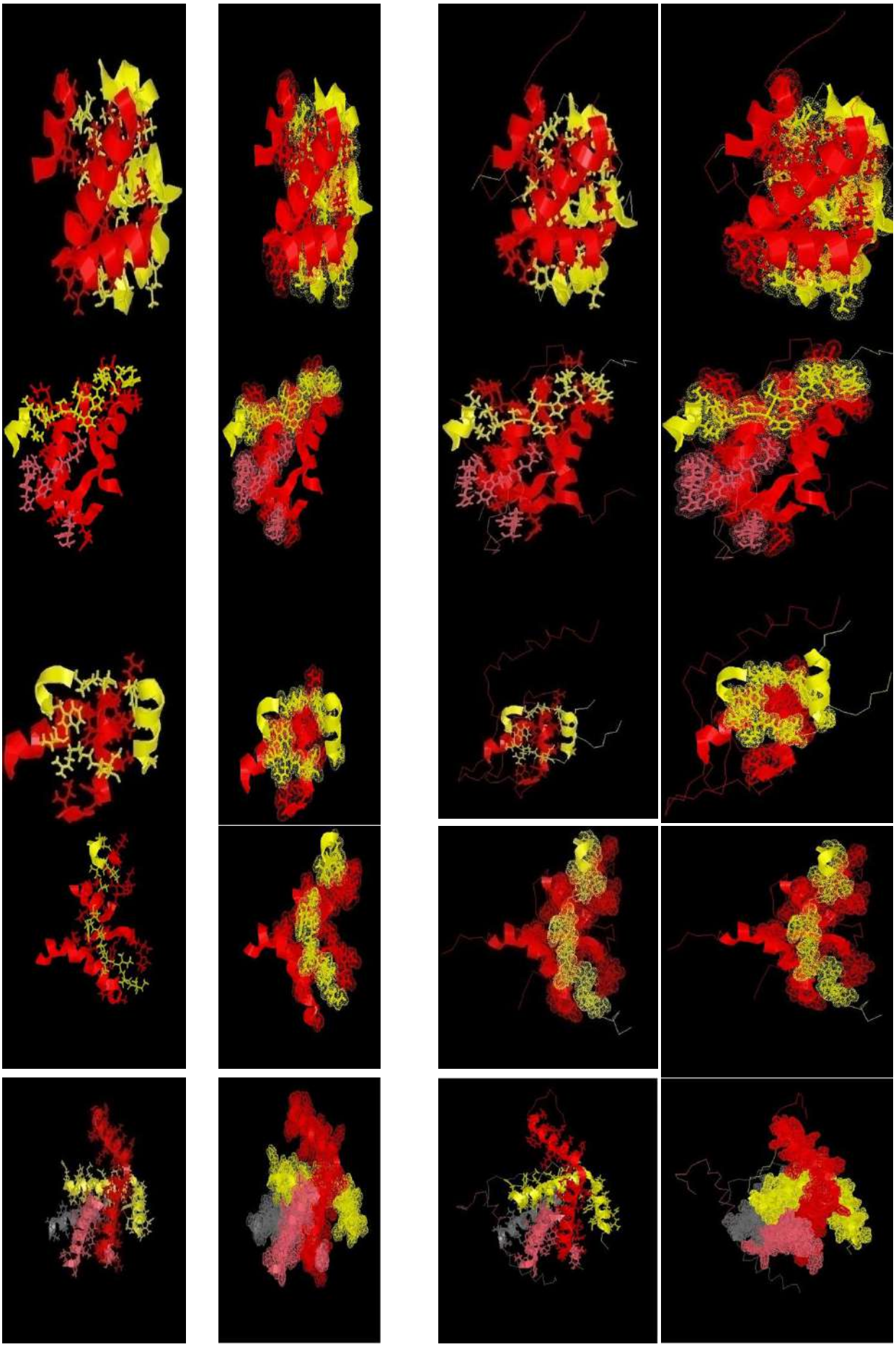
2kso



2l35



2lbf



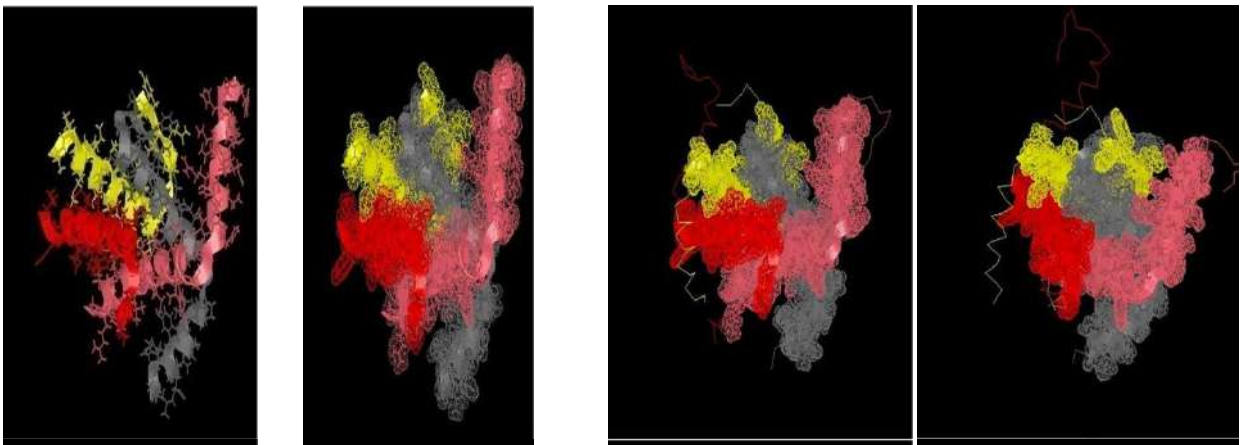
2mkc

2my2

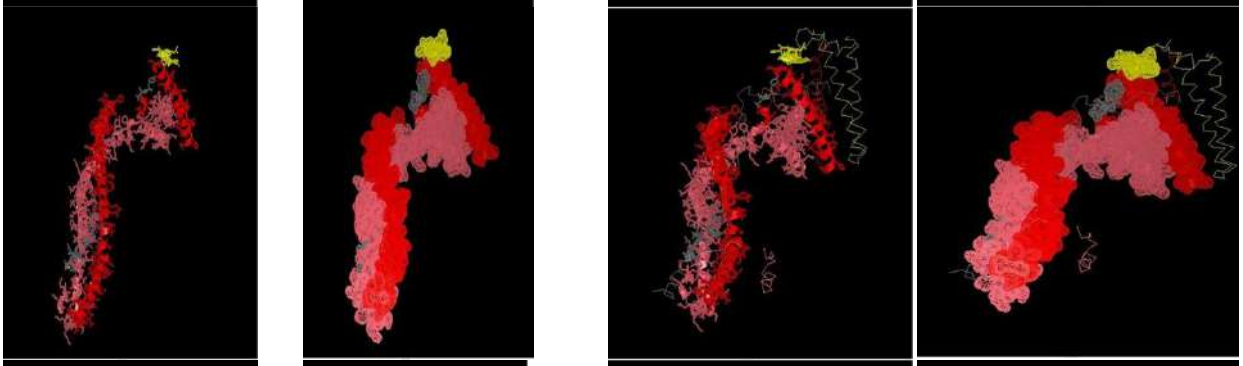
2my3

2nb1 « A »

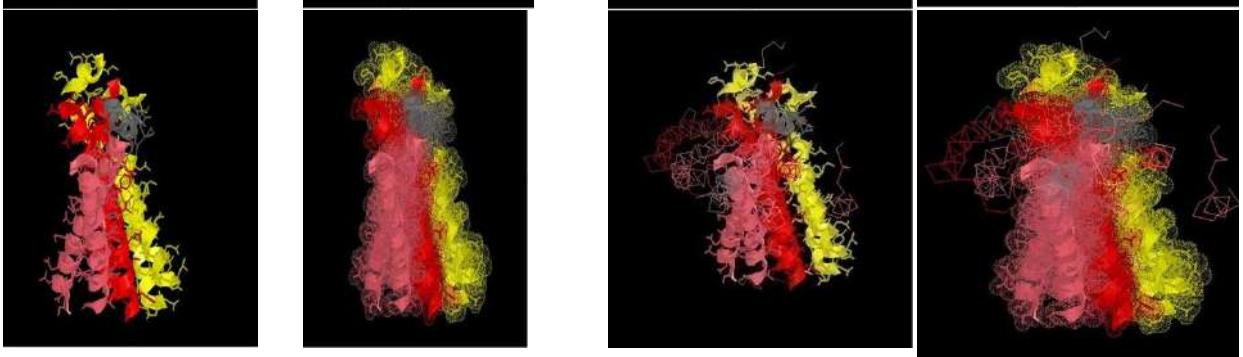
2nb1 « D »



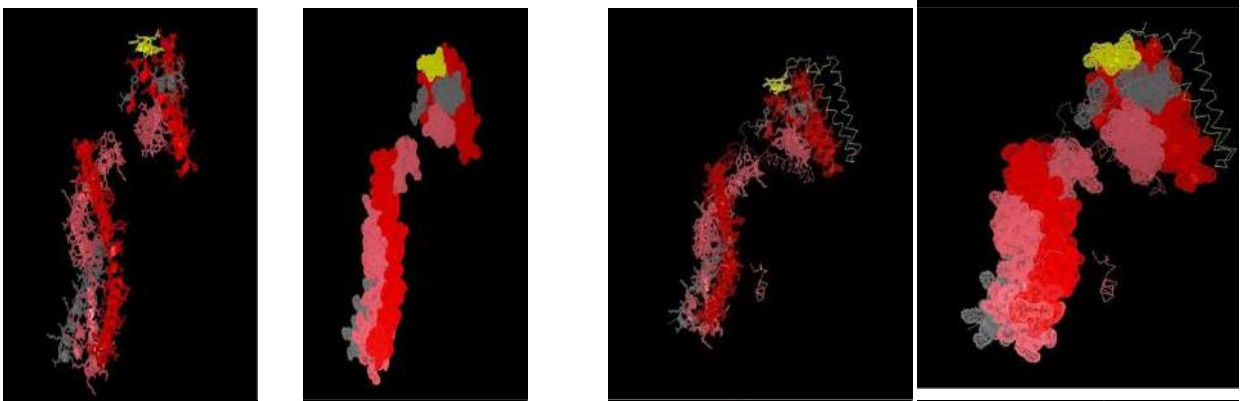
2p22_A2 19-240_ 238-277_ 280-323



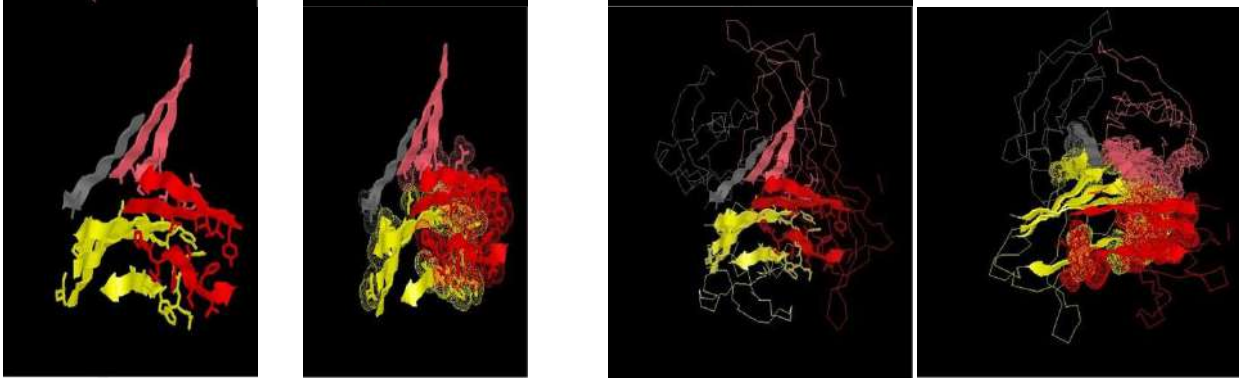
2p22_A3 26-352_ 358-385



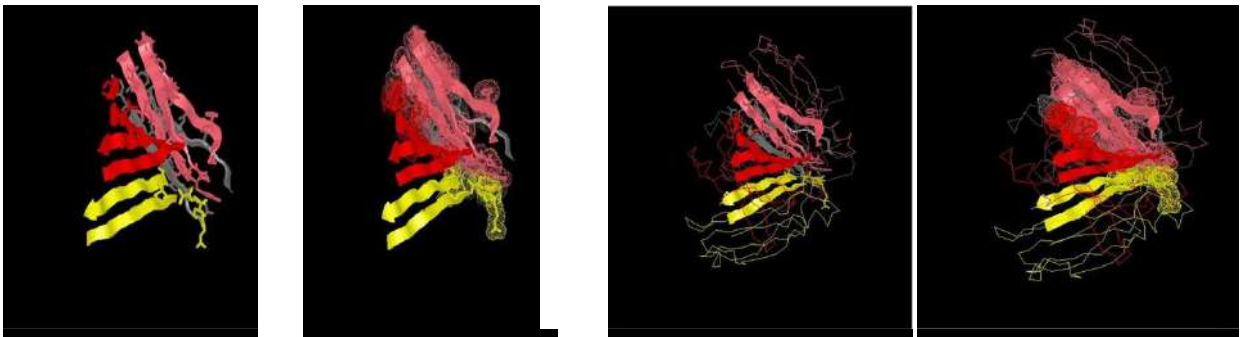
2p22 « D »



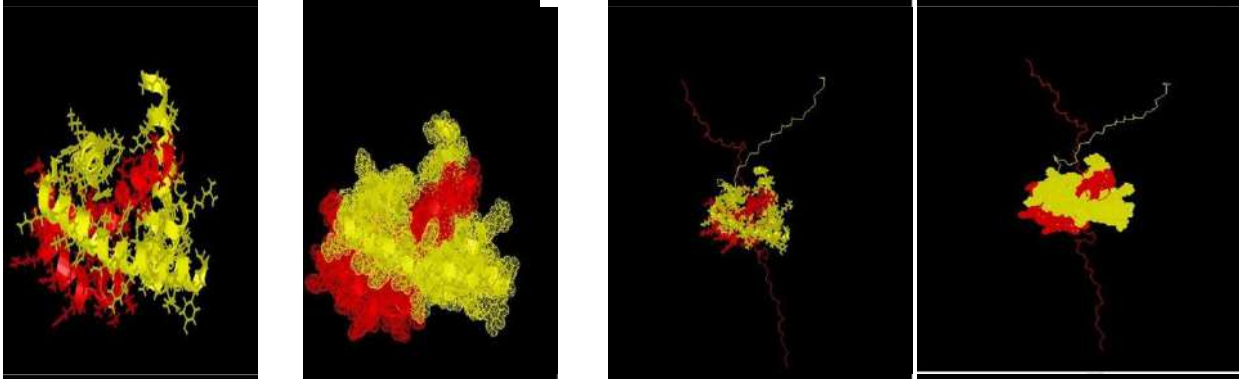
2qpf « B »



2qpf « C »

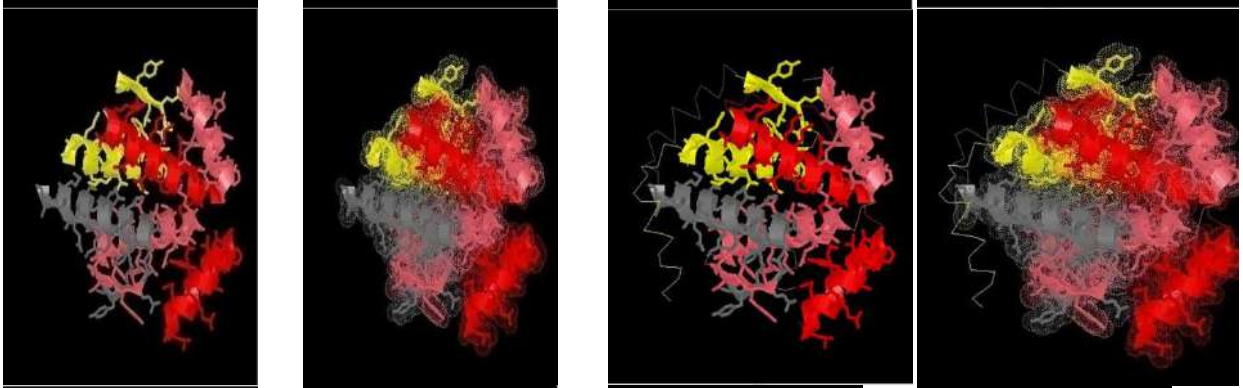


2rvq

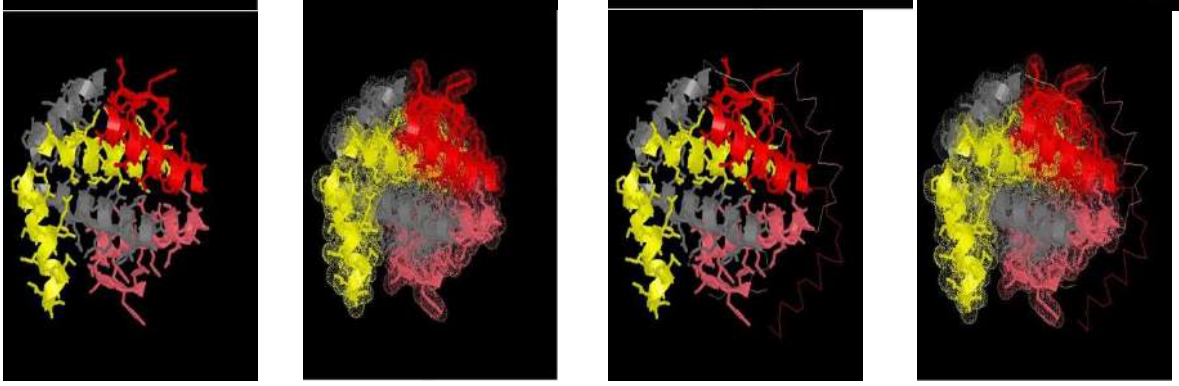


2wqi_B3 54-360

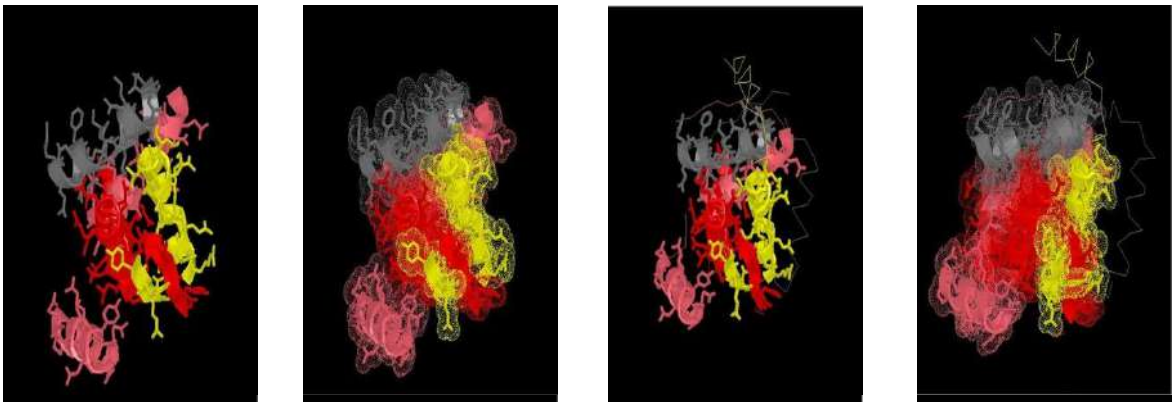
2wqi_C3 53-361_ 361-380_ 380-397_ 397-353



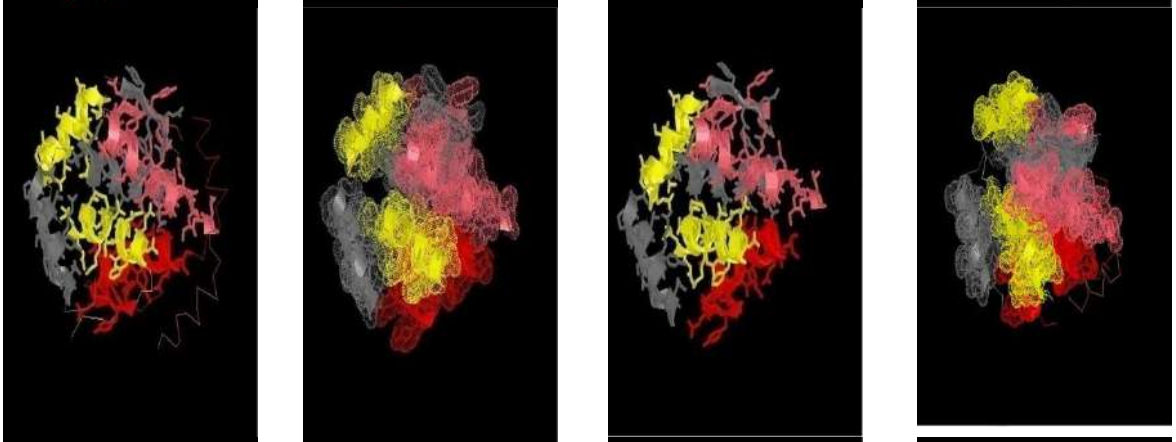
2wqi_D3 64-380_ 383-398



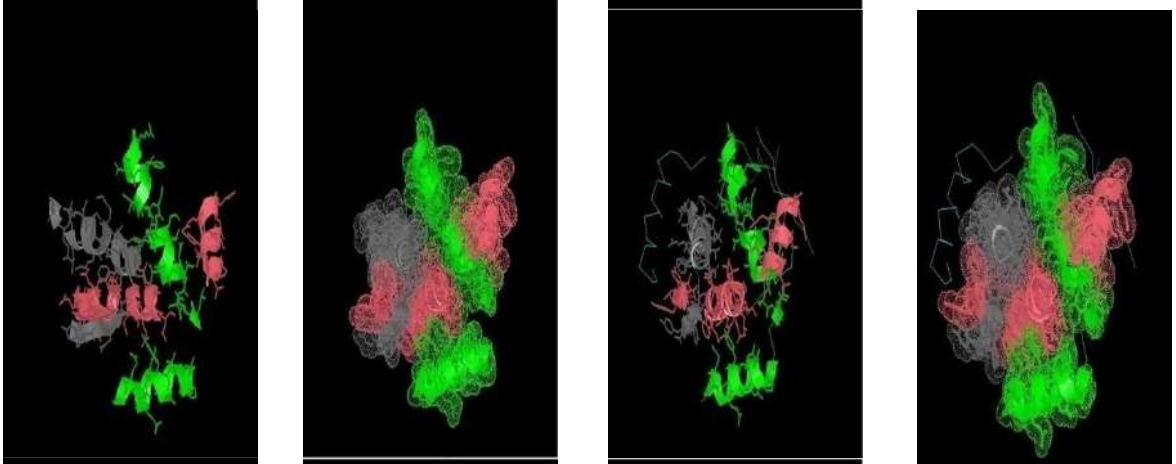
2wtt « A »



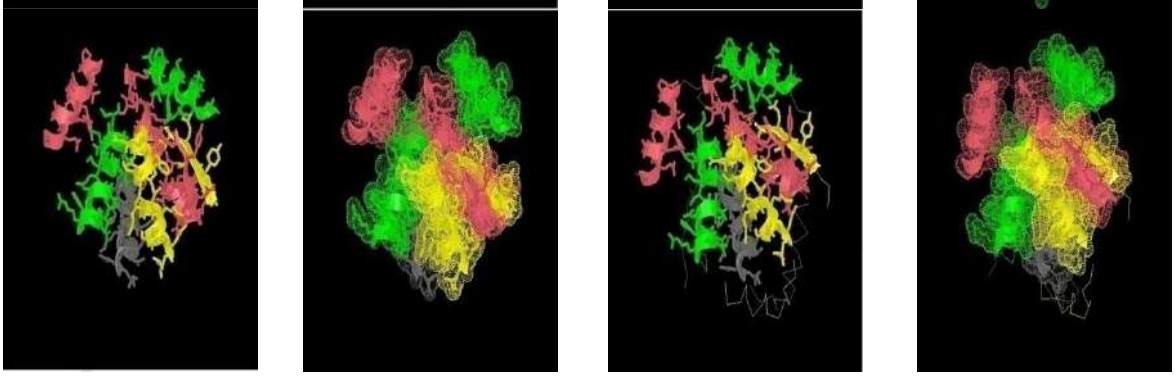
2wtt « D»



2wtt « E»



2wtt « F»



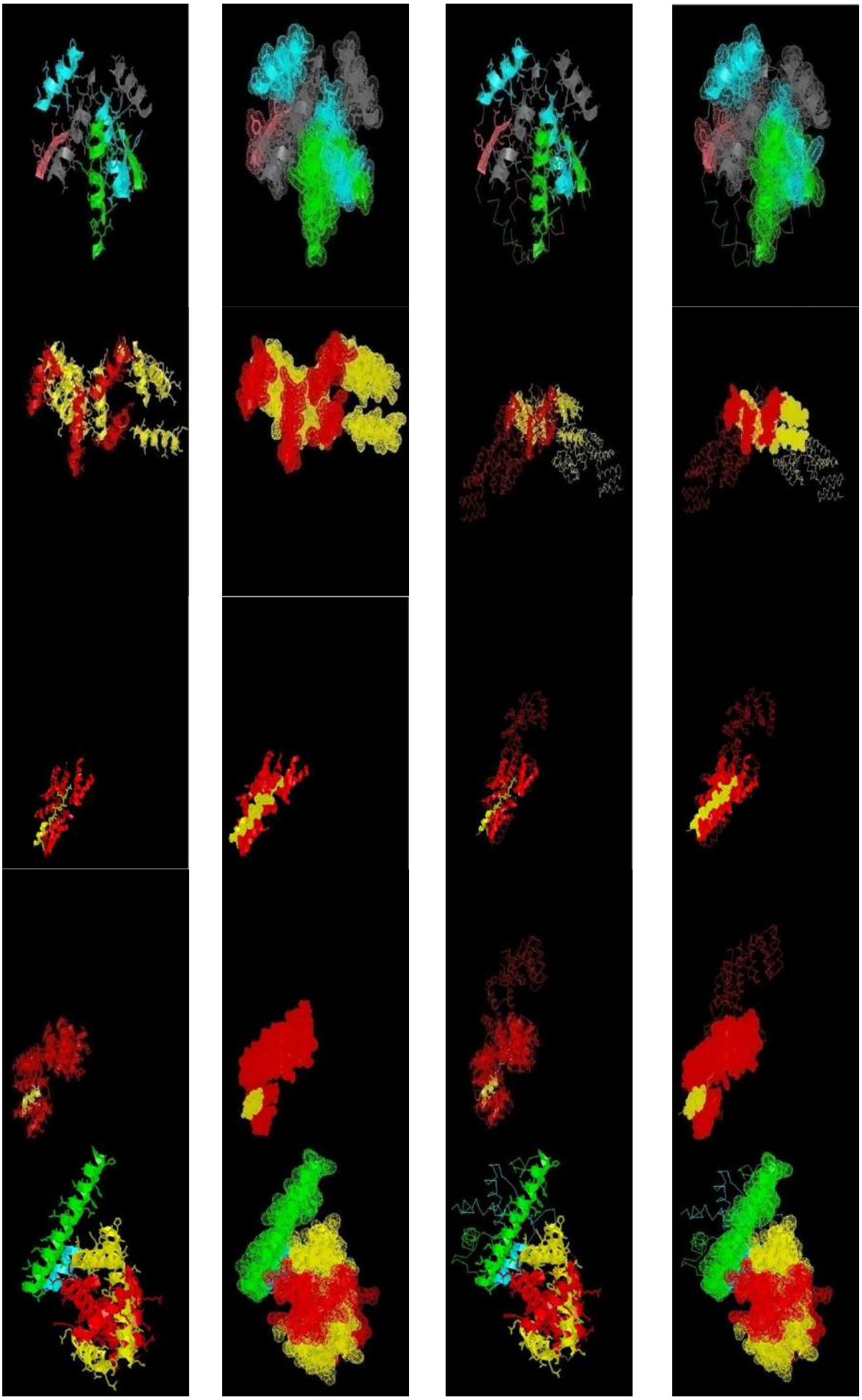
2wtt « H»

2xpi « A »

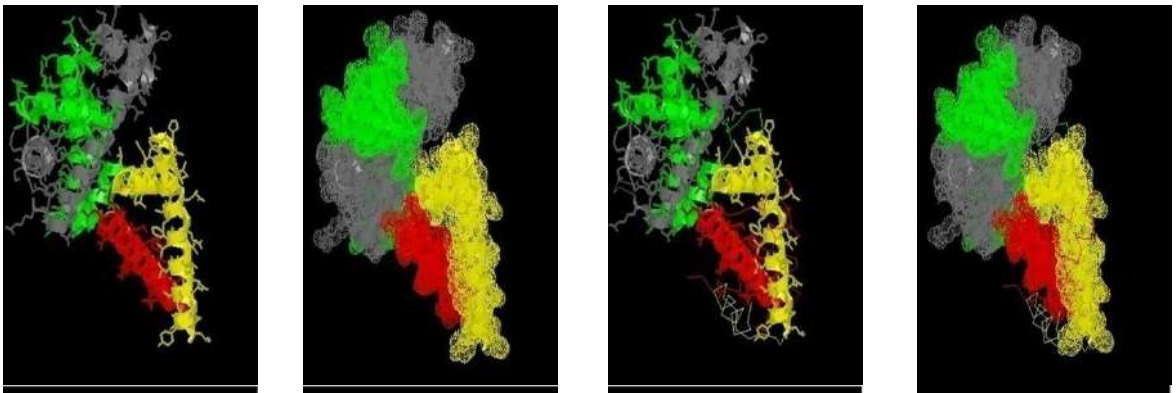
2xpi « B »

2xpi « E »

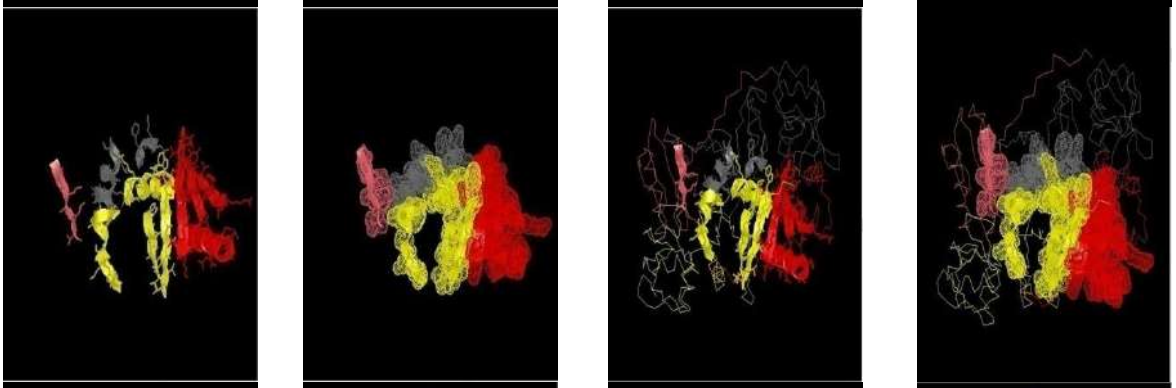
2yfw « B »



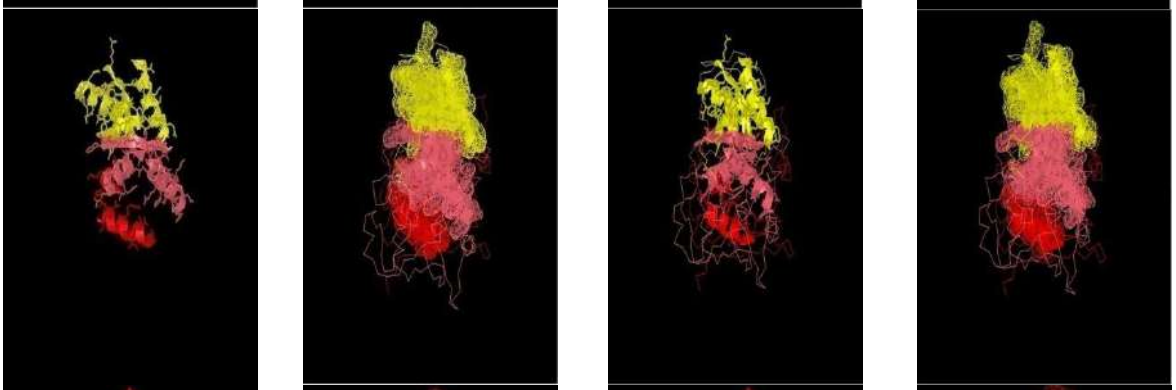
2yfw « H »



2zyz « B »

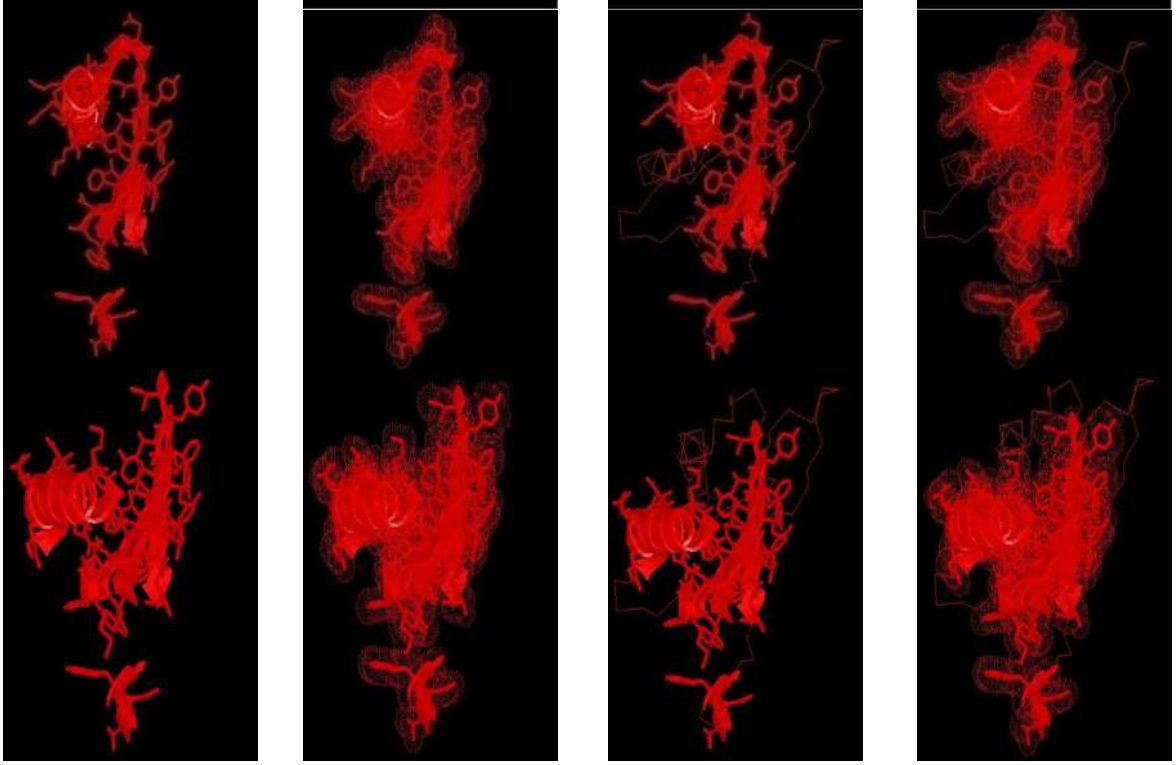


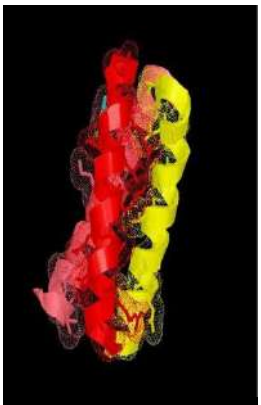
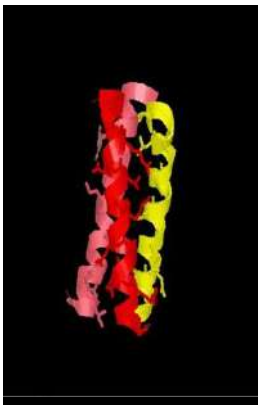
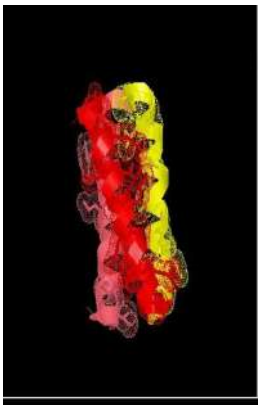
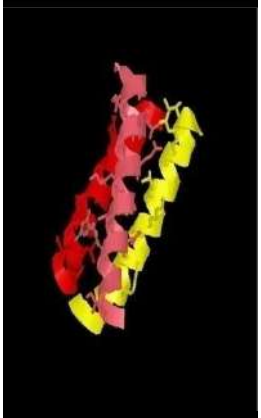
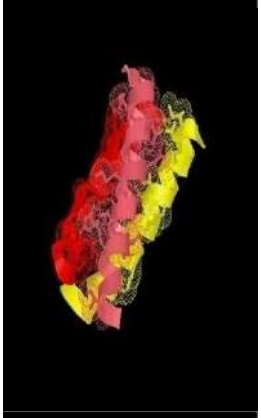
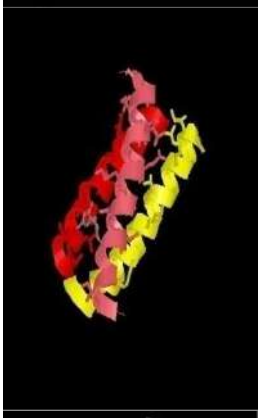
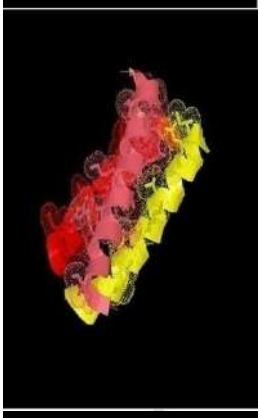
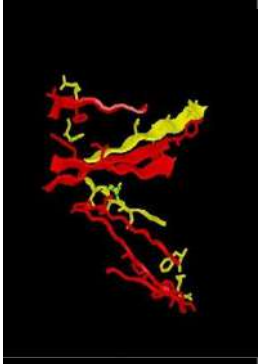
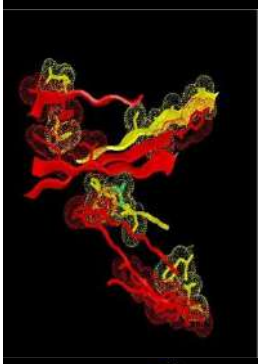
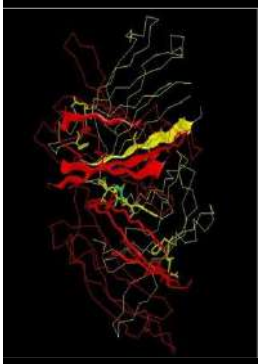
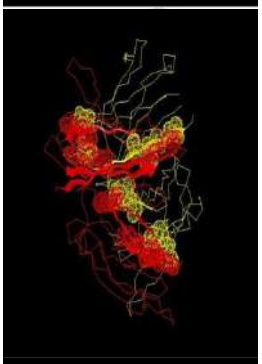
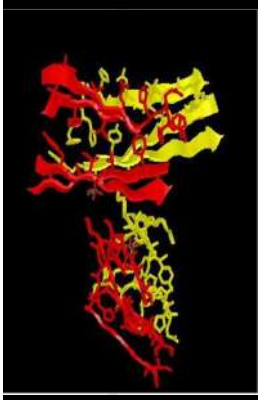
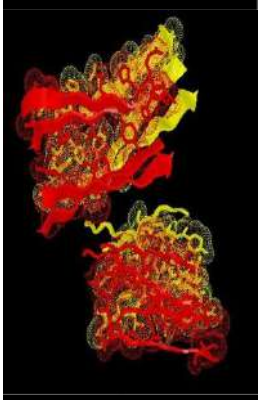
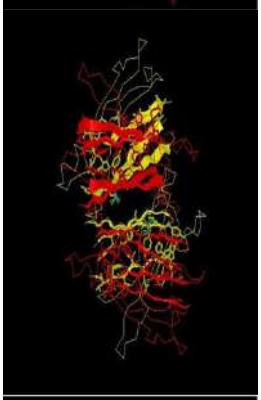
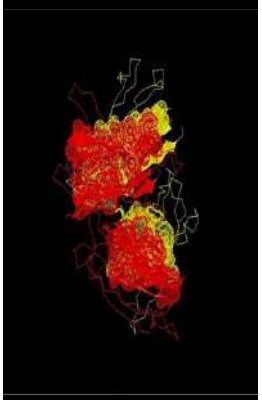
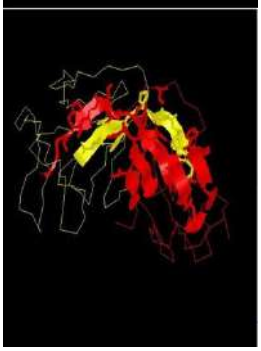
2zyz « D»



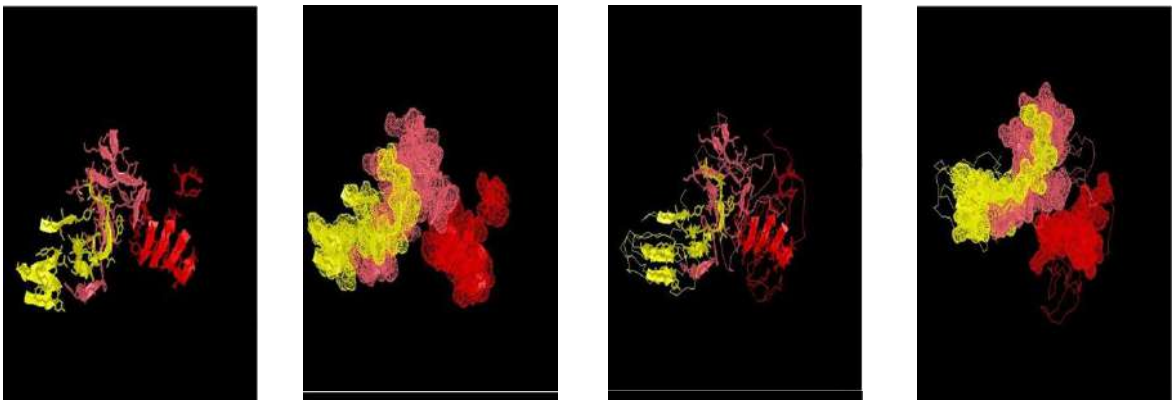
3fwt « A »

3fwt « C »

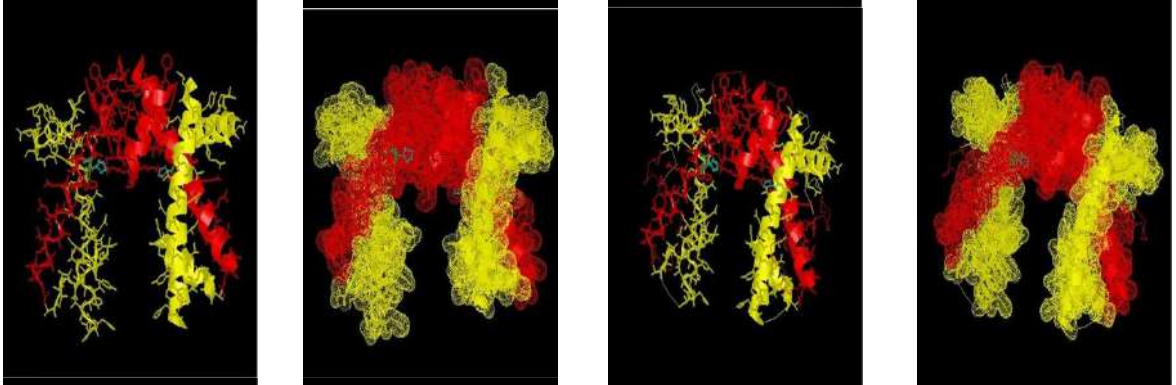


3hfe « A »				
3hfe « C »				
3kgt « A »				
3kgt « B »				
3mez « B »				

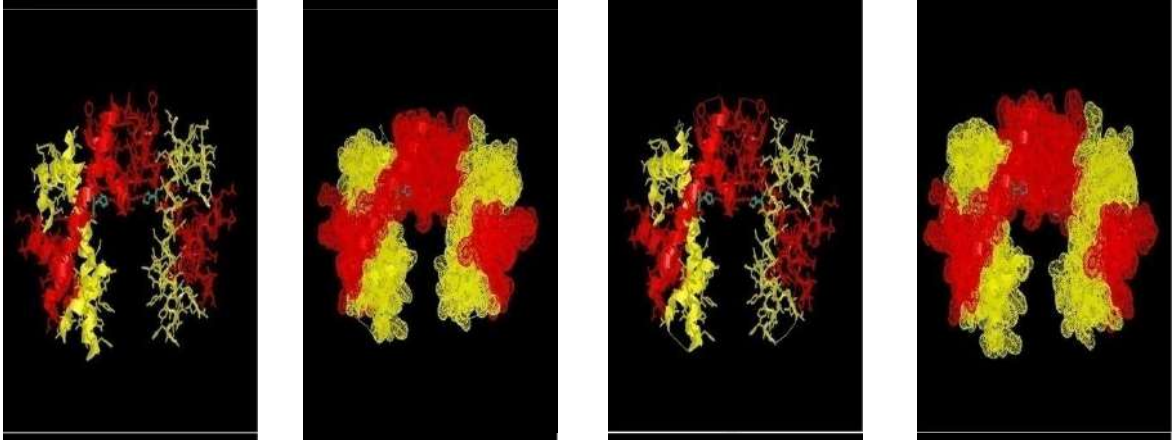
3mez « D»



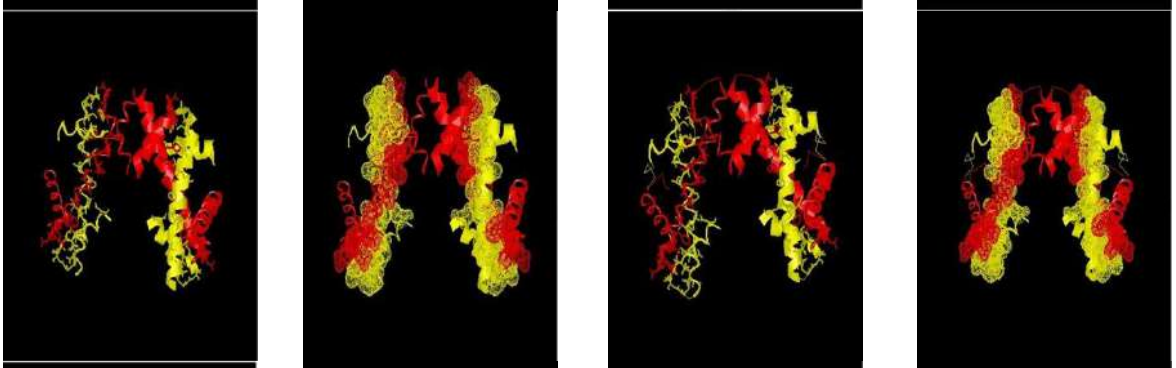
3nqj « A »



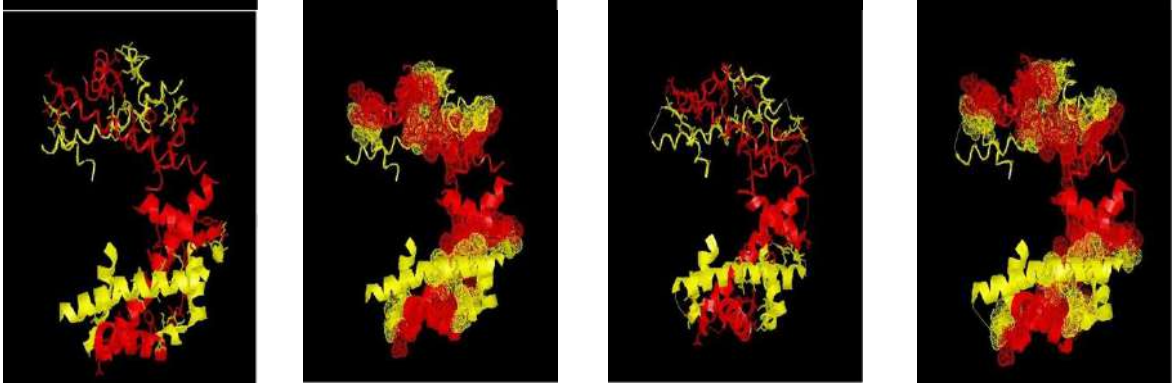
3nqj « B »



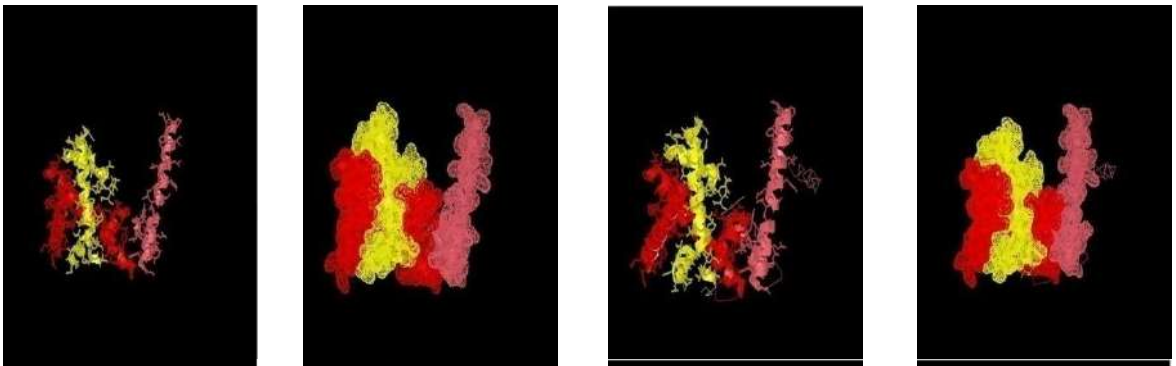
3nqu « A »



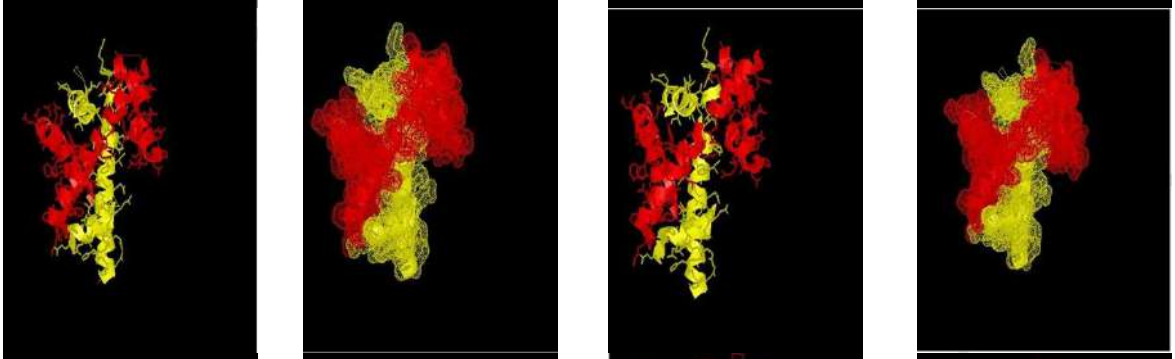
3nqu « B »



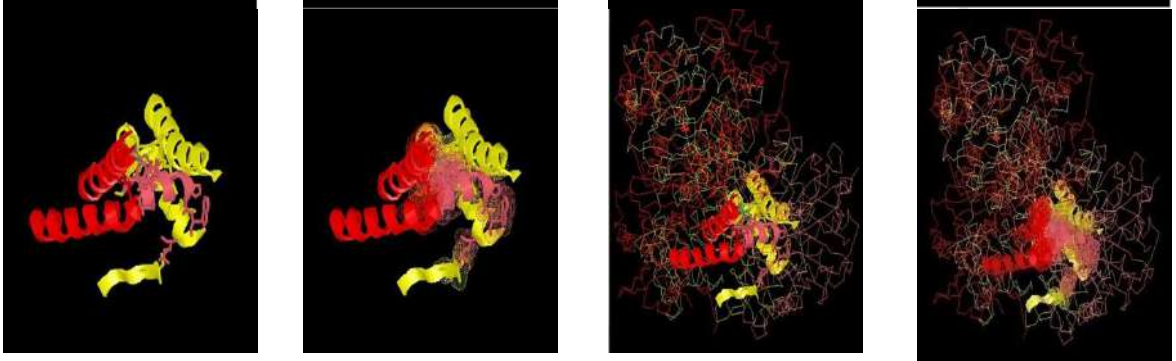
3v9r « A »



3v9r « D »

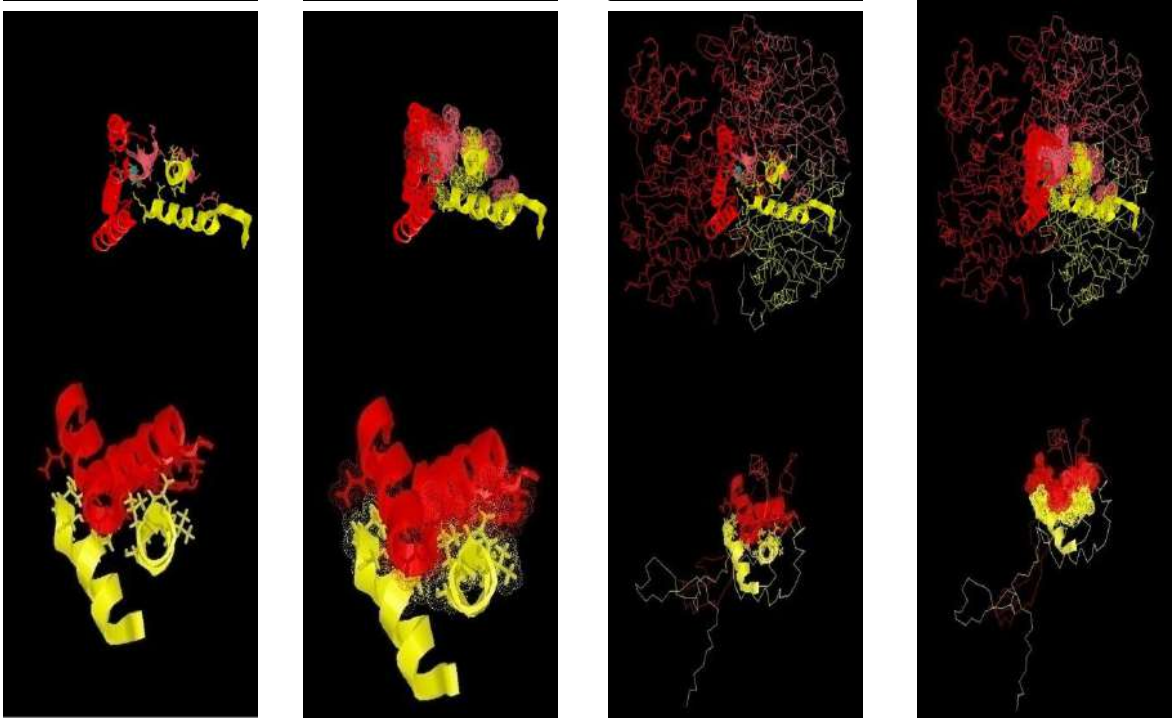


4b3h « C »

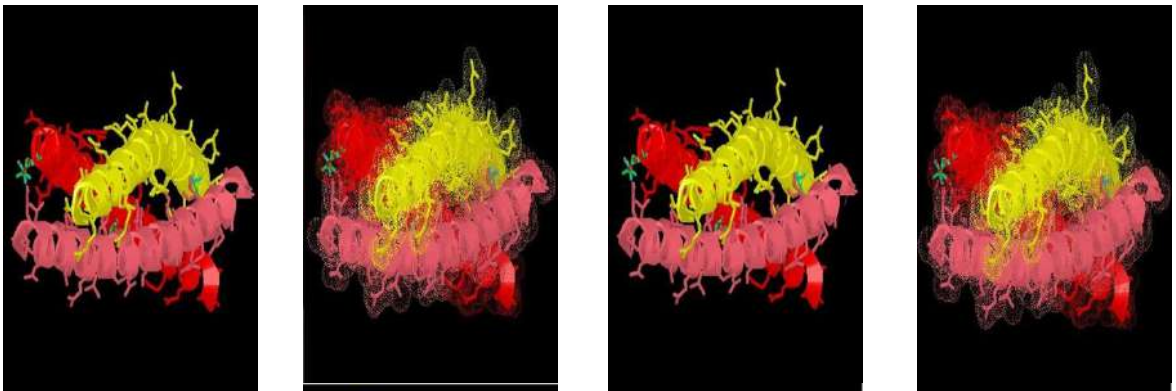


4b3h « D »

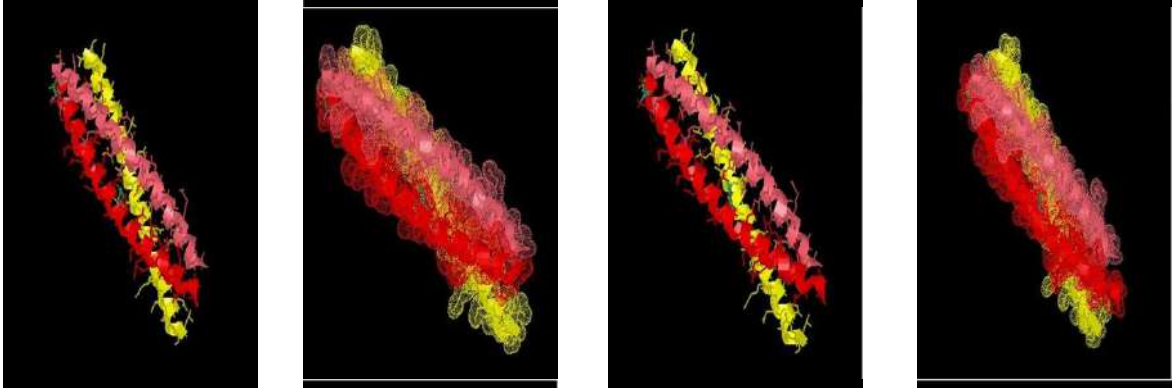
4beh



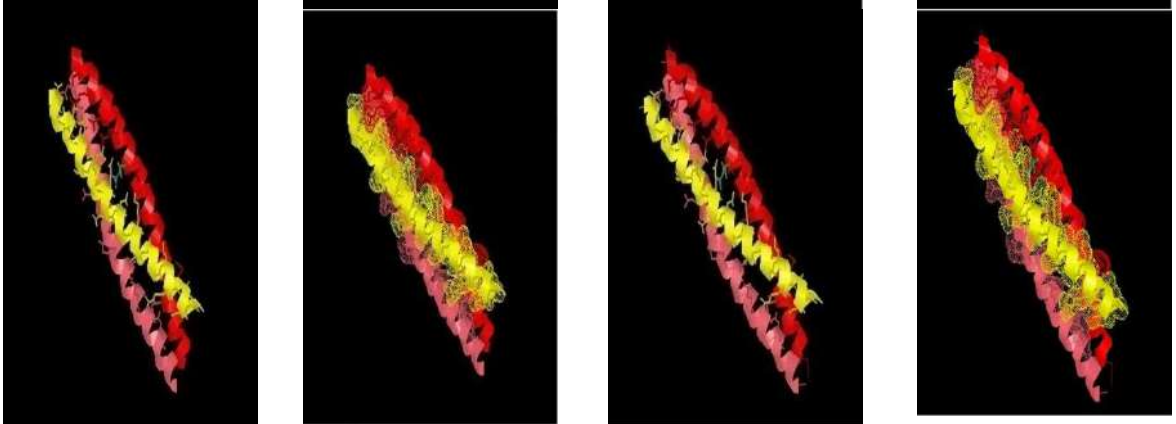
4cgb « B »



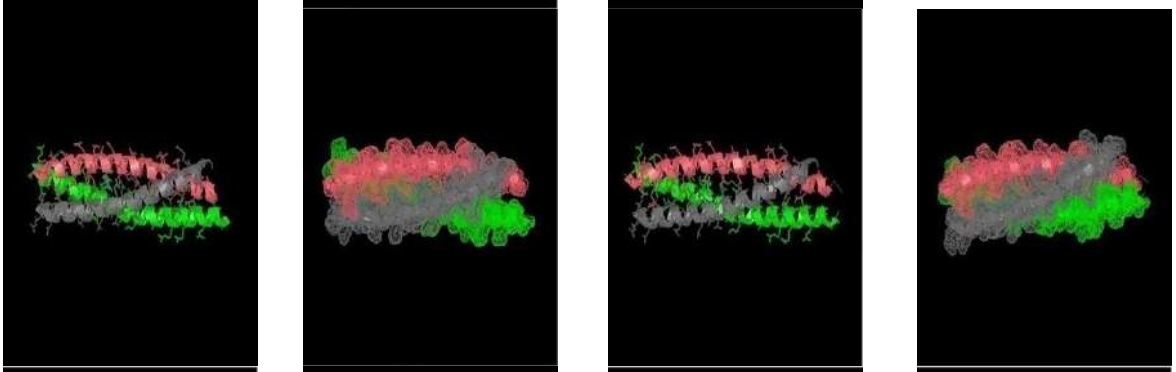
4cgb « C »



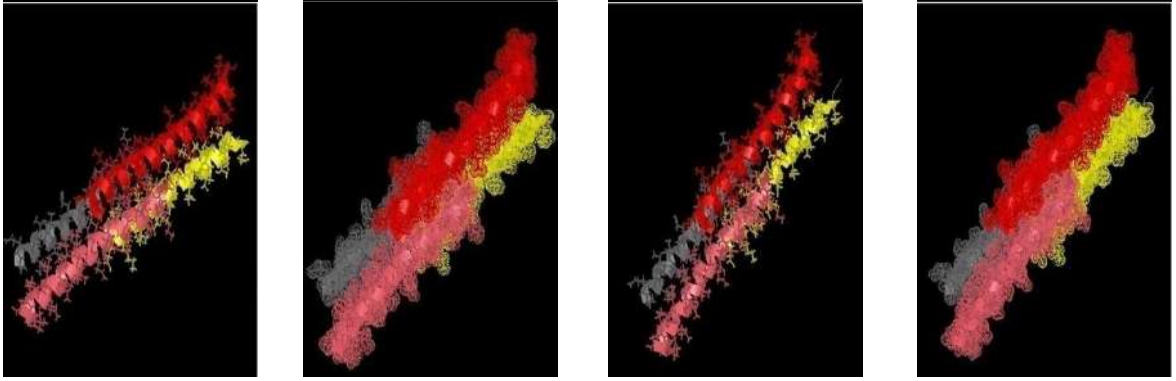
4cgb « E »



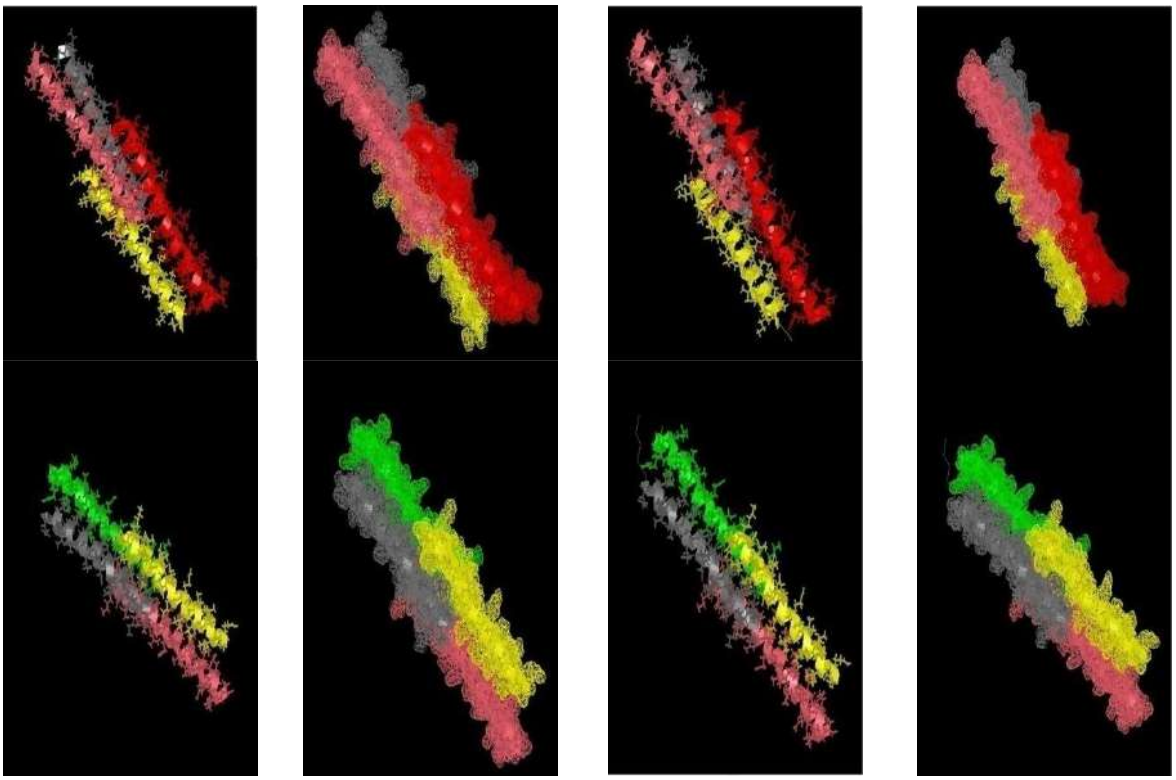
4cgb « F »



4d2h « A »



4d2h « C »



4d2h « F»

Level of bonds in each structure

PDB ID	HYDROGYN BOND	VAN_DER_WAALS
6A0A	+++	++++++++++++
4CGB	+++	++++++++++++
6BFS	+++	++++++++++++
5OCK	+++	++++++++++++
3HFE	+++	++++++++++++
5YUF	+++	++++++++++++
5LLV	+++	++++++++++++
2WQI	+++	++++++++++++
2KEI	+++	++++++++++++
1MEP	+++	++++++++++++
1S5Y	+++	++++++++++++
3NQU	+++	++++++++++++
4B3H	+++	++++++++++++
2XPI	+++	++++++++++++
6CGA	+++	++++++++++++
2P22	+++	++++++++++++
5WMN	+++	++++++++++++
1J7H	+++	++++++++++++
1DBF	+++	++++++++++++
1S55	+++	++++++++++++
3FWT	+++	++++++++++++
5EIL	+++	++++++++++++
5Y6U	+++	++++++++++++
1S7M	+++	++++++++++++

1T0A	+++	++++++++++++
3DJH	+++	++++++++++++
4J3H	+++	++++++++++++
4D2H	+++	++++++++++++
3KGT	+++	++++++++++++
CBM	+++	++++++++++++
4MIF	+++	++++++++++++
4MIG	+++	++++++++++++
1KKC	+++	++++++++++++
1N7Y	+++	++++++++++++
1N9M	+++	++++++++++++
2NB1	+++	++++++++++++
4NUV	+++	++++++++++++
5TCF	+++	++++++++++++
1TAF	+++	++++++++++++
3NQJ	+++	++++++++++++
3V9R	+++	++++++++++++
2WTT	+++	++++++++++++
1XK4	+++	++++++++++++
2YFW	+++	++++++++++++
2ZYZ	+++	++++++++++++
2QPF	+++	++++++++++++
3MEZ	+++	++++++++++++
1XOF	+++	++++++++++++
KSO	+++	++++++++++++
1YCS	+++	++++++++++++
2LBF	+++	++++++++++++
2RVQ	+++	++++++++++++
2MY2	+++	++++++++++++
2MY3	+++	++++++++++++
4BEH	+++	++++++++++++
4RLJ	+++	++++++++++++
4LWS	+++	++++++++++++

4P5E	+++	++++++++++++
4KO0	+++	++++++++++++
4P5D	+++	++++++++++++
4RLW	+++	++++++++++++
4RLU	+++	++++++++++++
4RLT	+++	++++++++++++
5G5G	+++	++++++++++++
5G5H	+++	++++++++++++
4LNR	+++	++++++++++++
4NUU	+++	++++++++++++
5OEK	+++	++++++++++++
5OHD	+++	++++++++++++
5NOC	+++	++++++++++++
5US5	+++	++++++++++++
6MS4	+++	++++++++++++
5XLV	+++	++++++++++++
6E4H	+++	++++++++++++
6NRX	+++	++++++++++++
6HAT	+++	++++++++++++
6FV5	+++	++++++++++++
6AKL	+++	++++++++++++
5XLW	+++	++++++++++++
1A4U	+++	++++++++++++
1AFW	+++	++++++++++++
1BD0	+++	++++++++++++
1BJW	+++	++++++++++++
1CNZ	+++	++++++++++++
1JFL	+++	++++++++++++
1LK9	+++	++++++++++++
1LQ9	+++	++++++++++++
1M4I	+++	++++++++++++
1MJH	+++	++++++++++++
5NF8	+++	++++++++++++

5I7U	+++	++++++++++++
5EZF	+++	++++++++++++
1jiw	+++	++++++++++++
1slu	+++	++++++++++++
2l35	+++	++++++++++++
1gg2	+++	++++++++++++
4yli	+++	++++++++++++
6B2E	+++	++++++++++++
2baz	+++	++++++++++++
5kew	+++	++++++++++++
5Kev	+++	++++++++++++

References

A. Rachedi, 2018 , Protein-Protein Interactions system, http://bioinformaticstools.org/prjs/ppi/
A. Rachedi, 2011,SSFS: Sequence, Structure & Function Server,

http://bioinformaticstools.org/SSFS

Abdellahi Ould Ahmed Salem, Mira Raimes, Aicha Bessaih, Mohamed Brahimi, Mohamed Amine Deghem, Mehdi Zein and Abdelkrim Rachedi, 28-29 October 2018, Oran, Algeria, Structural & Functional Motifs in Metabolic Enzymes . A new set of Functional Motifs. 2eme Colloques International, Biosciences 2018 "Genie Enzymatique et Production de Metabolites"

Carolina Estarellas Martín,September 2012 ,Theoretical and Experimental Study of Cooperativity Effects in Noncovalent Interactions

Forbes J. Burkowski, 2009, structural bioinformatics an algorithmic approach

Graz, December 2005,Design and Development of a Database for Protein-Protein Interaction in Crystals,

Haidong Wang Proteins, Interactions, December 2008, And Complexes: A Computational Approach

Horton, H.R., Moran, L.A., Scrimgeour, K.G., perry, M.D. and Rawn, J.D,2006, Principles of Biochemisty.

Pearson/Prentice Hall in Thursday, March 13, 2008, Upper Saddle River N.J. (USA) (published by Larry Moran

J. Tony Pembroke, 2000,Bio-molecular modelling utilising RasMol and PDB resources: a tutorial with HEW lysozyme

Jan Drenth, 1993 , Principles of Protein X-ray Crystallography

Jefferson; philadelphia university and Thomas Jefferson university web site, 2018, DNA and RNA.

Jessica Yee Lau, January 2005,Protein Structure Database For Structural Genomics Group

Kazuyoshi Murata ,Matthias Wolf, February 2018, Cryo-electron microscopy for structural analysis of dynamic biological macromolecules

M. H. Fulekar,2009,Bioinformatics: Applications in Life and Environmental Sciences.

Michael Gromiha,2010 ,Protein Bioinformatics: From Sequence To Function Larry Moran, MARCH 13, 2008 ,turns and loops

Michael K. Reddy, 2017, Amino acids

Olav Smidsrød, Størker Moe, Størker T. Moe,2008 , biopolymer chemistry

Rongkun Shen,March 9, 2006,Protein Secondary Structure Prediction Using Conditional Random Fields and Profiles

SALEM Abdellahi, RACHEDI Abdelkrim, April 04-06, 2017, Blida, Algeria., Structural/Functional Motifs in the Enzymes of the Krebs Cycle, 1er Seminaire de Bioinformatique

Salem and Mebarka, 2016, Citric Acid Cycle Enzymes Research on the Function and Structural Relationship Bases, Master Thesis

Shailza Singh, 2016 , Systems Biology Application in Synthetic Biology

Shuai Cheng Li, 2009, New Approaches to Protein Structure Prediction

T.M.Picknett, S.Brenner, november 201, Encyclopedia of Genetics SAGAR ARYAL_ Microbiology note

Yolanda Smith, 2018, B.Pharm, protein structure and function