الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne Démocratique et Populaire
وزارة التعليم العالي والبحث العلمي
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
جامعة سعيدة، د. مولاي الطاهر
University of Saida, Dr. MOULAY Tahar


كلية العلوم
Faculty of Sciences
قسم البيولوجيا
Department of Biologie

N° d'Ordre

**Thesis for obtaining the degree of Master**

In Biological Sciences

**Field: Biochemistry**

Theme

# Molecular Modelling and Prediction of Protein 3D-structure, Principals and Applications

Presented by:

- Miss: Ouafaa BAHLOUL

Submitted on: 02 July 2022

Before the jury composed by:

| | | |
|---|---|---|
| President | Mr. Yahia BELLIL | MCA University of Saida |
| Examiner | Mr. Noureddine HALLA | MCA University of Saida |
| Supervisor | Mr. Abdelkrim RACHEDI | MCA University of Saida |

**Année universitaire 2021/2022**


بسم الله الرحمن الرحيم

# Acknowledgment

First and foremost I am extremely grateful to my supervisor Dr. Abdelkrim Rachedi, for his invaluable advice, continuous support, and patience during doing this wonderful project on the topic (Molecular Modeling snd Prediction of 3D-structure of proteins, Principals and Applications). His guidance and advice carried me through all the stages of writing my project, which also helped me to get acquainted with a new world of bioinformatics science and made me love this field.

I am really thankful to have acquired them.

My thanks also go to the jury members my respected teachers, Dr. Yahia Bellil and Dr. Noureddine Halla for their brilliant remarks and ideas.

# Dedication

*All praises be to Allah, the Lord of the world, who guided me the right way, today I fold the days tiredness between the cover of this humble work, to my My Master and honoured prophet Muhammad- May peace and grace of allah be upon him.*

*To my great parents, who worked hard and made every effort to continue my education until I reached this precious moment, the spring of endless giving and love my mother Fatema. To who worked hard and made every effort and supported me financially, morally my father Chaabane.*

*My dedication also goes to who helped me a lot in this journey, and gave me the peace of my heart that no one gave me, and who was always my only motivation, My fiancé Iliasse.*

*To those who have demonstrated to me what is brotherhood and love mean, my brothers and sisters: Abdellatif, Nour elhouda, Hadjer, Abdelmajid , Youcef, And my dear friend Hadjer.*

*To everyone who helped me from near or far.*

*Wafaa*

# Summary

# Summary

# Summary

# List of Tables

## Chapter I



## Chapter II



## Chapter III

# List of Tables

## Appendix part II



## Appendix part III

# List of Tables

# List of Figures

## Chapter I

# List of Figures



## Chapter II

# List of Figures

# List of Figures



## Chapter III

# List of Abbreviations

| | |
|---|---|
| **2D :** | Two-Dimensional |
| **3D :** | Three-Dimensional |
| **3d-SS :** | Three-Dimensional Structure Superposition |
| **Å :** | Angstrom |
| **AA :** | Amino acid |
| **APHM:** | Auto Protein Homology Modeling |
| **BMRB :** | Biological Magnetic Resonance Data Bank |
| **CC :** | Common Core |
| **Cryo-EM :** | Cryo-electron microscopy |
| **Cα:** | Central carbon α atom of amino acids |
| **DHFR :** | Dihydrofolate reductase enzyme |
| **DNA :** | Deoxyribonucleic Acid |
| **EBI :** | The European Bioinformatics Institute |
| **EMA :** | European Medicines Agency |
| **FA :** | Folic acid |
| **FDA :** | U.S. Food and Drug Administration |
| **H :** | Alpha-Helix |
| **hDHFR :** | human DHFR |
| **ICM:** | Internal Coordinate Mechanics |
| **L :** | Loop |
| **NADP :** | nicotinamide adenine dinucleotide phosphate |
| **NCBI :** | National Center for Biotechnology Information |
| **NMR :** | Nuclear Magnetic Resonance |
| **PDB ID :** | Protein Data Bank identifier |
| **PDB :** | Protein Data Bank |
| **PDBe :** | Protein Data Bank Europe |
| **PDBj :** | Protein Data Bank Japan |
| **PIR :** | Protein Information Resource |
| **RCSB :** | Research Collaboratory for Structural Bioinformatics |
| **RMSD :** | Root Mean  Square Deviation |
| **S :** | Beta-Strand |
| **SIB :** | the Swiss Institute of Bioinformatics |
| **SIFTS :** | Structure Integration with Function, Taxonomy, and Sequence |

# List of Abbreviations

| | |
|---|---|
| **SSFS :** | Sequence, Structure and Function Server |
| **STamp** | Structural Alignment of Multiple Proteins |
| **TEM :** | Transmission Electron Microscopy |
| **THF :** | Tetrahydrofolate |
| **UniParc :** | The UniProt Archive |
| **UniProtKB :** | The UniProt Knowledgebase |
| **UniRef :** | The UniProt Reference Clusters |

## Abstract

The construction of DNA is a very important process for growth, development and cellular reproduction as well as carrying genetic information, in which processes, supportive and essential enzymes intervene. Healthy cells reproduce normally but cells may be diseased and undergo abnormal cellular division resulting in pathology cases such as cancer. The quest to understand the enzymes involved in DNA creation allows understanding of how to deal with disorders and diseases related to unlimited and abnormal cell growth and division. It would also help understand microbial infections as this reply on bacterial cell reproduction.

Many studies in the field of structural biology have proven that three-dimensional structure of a protein gives deeper understanding of how proteins work and function. It is not always possible to carry out experimental structural studies and hence predicting the structure of a protein is one of the methods for studying their structure and function relationships.

Molecular modelling of protein structure enables to study catalytic amino-acids in active sites and other important sites and thus allows for putting hypotheses forward with the aim of understanding their roles in normal and disease situations. Such studies allow even for doing what has become known as rational design of new drugs that that may proof effective with limited side effects against various incurable diseases.

The aim of this study is to apply the molecular modelling methodology to predict and create a reliable structural model for the human Dihydrofolate Reductase enzyme - DHFR, which is part of the pool of biosynthesis reactions involved in the production of nucleotides that constitute DNA essential in cell division and cell proliferation.

The study resulted in the creation of a strategy of clear steps that enabled the production of a three-dimensional model of the human dihydrofolate reductase enzyme, and investigated the general rules that enable the analysis of the structure and function of this enzyme.

# Abstract

This is in addition to creating an online bioinformatics application available from he bioinformatics-server page of the University of Saida: APHM & Viewer tool that contributes to important stages of the molecular modelling and visual exploration. The tool aims at applying the strategy to general case for predicting the structure of other proteins and is available at the following link:

https://bioinformatics.univ-saida.dz/bit2/?arg=APHM&ttl=Auto%20Homology%20Molecular%20Modeling

**Keywords:**

Proteins, Enzymes, Dihydrofolate Reduction, 3D Structure, Molecular Modelling, Structure Prediction, Biological Function, Databases, Bioinformatics.

# Abstract


## الملخص

بناء الحمض النووي عملية مهمة جدا للنمو والتطور والتكاثر في الخلية وأيضا حمل المعلومات الجينية، والتي تتدخل فيها انزيمات داعمة وأساسية، وقد تسلك الخلايا الطبيعية إما تكاثرا طبيعيا أو تكون هذه الخلايا مريضة وتبدي نتائج غير طبيعية كالسرطان. والسعي لفهم وظيفة الانزيمات المساعدة في بناء الحمض النووي تسمح بفهم كيفية التعامل مع الاختلالات والأمراض المتعلقة بالنمو والانقسام الخلوي الغير محدود والغير طبيعي.

وقد أثبتت الكثير من الدراسات في مجال البيولوجيات الهيكلية أن البنية ثلاثية الأبعاد للبروتين تمنح مستوى أعلى لفهم كيفية عمل البروتين وبالتالي وظيفته، ويعتبر توقع بنية البروتين من أحدث الطرق لدراسة البنية حيث يمكن من استعراض الأحماض الأمينية التحفيزية للموقع الفعال وبذلك يسمح بإنشاء فرضيات بهدف تغيير أو تعديل هذا الانزيم وحتى القيام بما أصبح يسمى بالتصميم العقلاني لأدوية جديدة أكثر فعالية ومحدودة الآثار الجانبية ضد مختلف الأمراض المستعصية.

الهدف من هذه الدراسة هو تطبيق منهجية النمذجة الجزيئية للتنبؤ وانشاء نموذج هيكلي ثلاثي موثوق به لإنزيم اختزال ثنائي هيدرو الفولات والذي هو جزء من سلسلة المراحل الداخلة في انتاج النيوكليوتيدات التي تشكل الحمض النووي الضروري في الإنقسام الخلوي و تكاثر الخلايا.

انتهت الدراسة بإنشاء استراتيجية من خطوات واضحة مكنت من إنتاج نموذج ثلاثي الأبعاد لإنزيم اختزال ثنائي هيدرو الفولات، وأرست قواعد عامة تمكن من تحليل بنية هذا الانزيم ووظيفته والآثار التركيبية والوظيفية للطفرات التي قد يحملها. هذا بالإضافة إلى إنشاء تطبيق متوفر على صفحة المعلوماتية_الحيوية لجامعة سعيدة يساهم في مراحل مهمة من عملية النمذجة الجزيئية و الإستكشاف يستهدف تعميم تطبيق الإستراتيجية و هو متوفر على الرابط التالي:.
https://bioinformatics.univ-saida.dz/bit2/?arg=APHM&ttl=Auto%20Homology%20Molecular%20Modeling



**الكلمات المفتاحية:**

البروتينات، الإنزيمات، اختزال ثنائي هيدرو الفولات، النمذجة الجزئية، التنبؤ بالبنية الفراغية، الوظيفة البيولوجية، قواعد البيانات، المعلوماتية الحيوية.

# Abstract


## Résumé

La construction de l'ADN est un processus très important pour la croissance, le développement et la reproduction cellulaire ainsi que le transport de l'information génétique, dans lequel interviennent des processus, des enzymes de soutien et essentielles. Les cellules saines se reproduisent normalement, mais les cellules peuvent être malades et subir une division cellulaire anormale entraînant des cas pathologiques tels que le cancer. La quête pour comprendre les enzymes impliquées dans la création de l'ADN permet de comprendre comment traiter les troubles et les maladies liés à la croissance et à la division cellulaires illimitées et anormales. Cela aiderait également à comprendre les infections microbiennes comme cette réponse sur la reproduction des cellules bactériennes.

De nombreuses études dans le domaine de la biologie structurale ont prouvé que la structure tridimensionnelle d'une protéine permet de mieux comprendre le fonctionnement et la fonction des protéines. Il n'est pas toujours possible de réaliser des études structurales expérimentales et donc la prédiction de la structure d'une protéine est l'une des méthodes pour étudier leurs relations structure et fonction.

La modélisation moléculaire de la structure des protéines permet d'étudier les acides aminés catalytiques dans les sites actifs et autres sites importants et permet ainsi d'émettre des hypothèses dans le but de comprendre leurs rôles dans des situations normales et pathologiques. De telles études permettent même de faire ce que l'on appelle la conception rationnelle de nouveaux médicaments qui peuvent s'avérer efficaces avec des effets secondaires limités contre diverses maladies incurables.

Le but de cette étude est d'appliquer la méthodologie de modélisation moléculaire pour prédire et créer un modèle structurel fiable pour l'enzyme humaine Dihydrofolate Reductase - DHFR, qui fait partie du pool de réactions de biosynthèse impliquées dans la production de nucléotides qui constituent l'ADN essentiel dans la division et prolifération cellulaire.

L'étude a abouti à la création d'une stratégie d'étapes claires qui a permis la production d'un modèle tridimensionnel de l'enzyme humaine dihydrofolate réductase, et a étudié les règles générales qui permettent l'analyse de la structure et de la fonction de cette enzyme.

# Abstract

Ceci s'ajoute à la création d'une application bioinformatique en ligne disponible sur la page du serveur bioinformatique de l'Université de Saida: l'outil APHM & Viewer qui contribue à des étapes importantes de la modélisation moléculaire et de l'exploration visuelle. L'outil vise à appliquer la stratégie au cas général pour prédire la structure d'autres protéines et est disponible au lien suivant :

https://bioinformatics.univ-saida.dz/bit2/?arg=APHM&ttl=Auto%20Homology%20Molecular%20Modeling

**Mots clés:**

Protéines, enzymes, réduction du dihydrofolate, structure 3D, modélisation moléculaire, prédiction de structure, fonction biologique, bases de données, bioinformatique.

# General Introduction

# General Introduction

Proteins are among the most important molecules that can play a wide array of significant, important roles in all living cells, and are more diverse in structure and function than other classes of macromolecules due to their several unique three-dimensional structures.

In the absence of experimental means for determining 3D-structure of macromolecules, prediction and modeling of protein structure are among the most important means of computational biology to help elucidate protein function chiefly depend on its 3D-structure. It would greatly expedite attempts to comprehend the building elements of cells and solve important medicinal queries through rational design for effective drug and development.

This study touches the objective of predicting a three-dimensional structure of a protein, the human dihydrofolate reductase by homology modeling, defining and characterizing its major steps and methods. The project resulted in the creation of a homology model based on similarity to known structures found in the Protein Databank (PDB). Important conclusion points have been drawn in relation to fundamentals of structure-function relationship preservation throughout the evolutionary history among different species. An online bioinformatics tool has also been implemented into executing some tasks relevant to homology model building.

The project is sub-divided along three major chapters:

- **The first chapter:** Theoretical background

  It talks about the general concepts and structure components of proteins also their structural levels, and several information about the chosen enzyme Dihydrofolate reductase for this study, as well as about the computational and experimental methods for determining the three-dimensional structure of protein.

- **The second chapter:** Material and methods

  The tools and material used in the homology modeling process steps are presented.

- **The third chapter:** Results and discussion

  This chapter presents the results reached during this study, and their discussion as well as some important conclusions.

- **General conclusion** of major concepts touching upon the fundamental rules behind the biology function have ended the report in this thesis.

# Chapter 1: Theoretical background

## I. Wide concept of proteins

A protein attracts a great deal of attention because it's one of the most widespread and complicated macromolecules within living organisms, proteins do most of the work in cells and are required for the structure, function, and regulation of the body's tissues and organs. they are made up of hundreds or thousands of smaller units called amino acids, There are 20 different types of amino acids that can be combined to make a protein (**see Appendix part I**), they are linked by a peptide bond in a linear chain called a polypeptide.

However, so little is known about how a protein folds into the specific three-dimensional structure from its one-dimensional sequence, Therefore, many researchers and experts specializing in the study of biological research have been devoted since the last half of the twentieth century. **(Deng H et al., 2018).**

### I.1.Protein structure

Amino acids are the building blocks of proteins, Within a protein, multiple amino acids are linked together by peptide bonds, thereby forming a long chain. Amino acids are organic compounds that contain an alpha (central) carbon atom (Cα) linked to an amino group, a carboxyl group, a hydrogen atom, and a variable component called a side chain specific to each amino acid, All AA have the same basic structure **(figure 01)**. Peptide bonds are formed by a biochemical reaction that extracts a water molecule when the carboxyl group of one molecule reacts with the amino group of the other molecule **(figure 02)**. Thus, we can say that the basic primary structure of proteins is the linear sequence of amino acids. ( **Reddy M-K., 2020).**

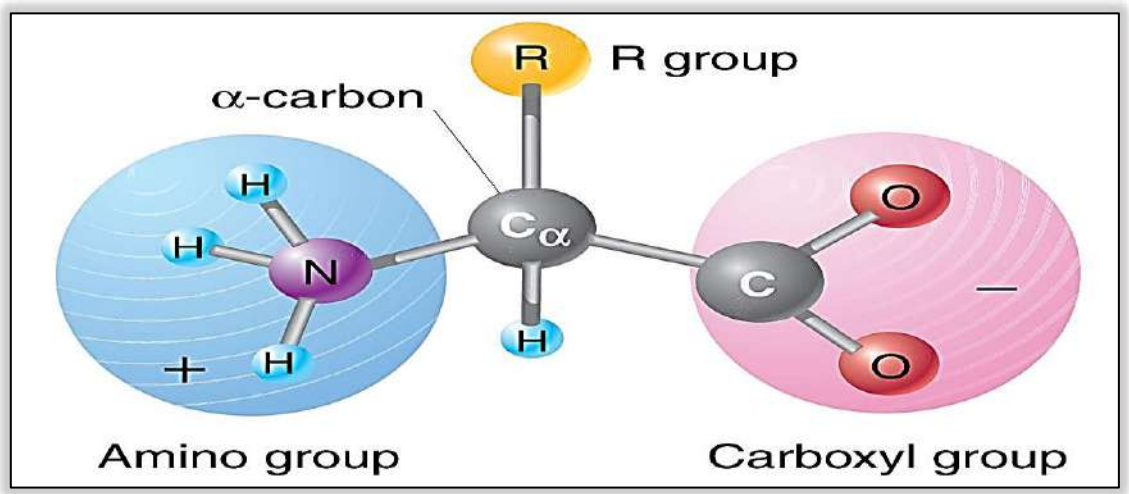



**Figure 01.** Amino Acid structure (**PJ Russell, 2010)**

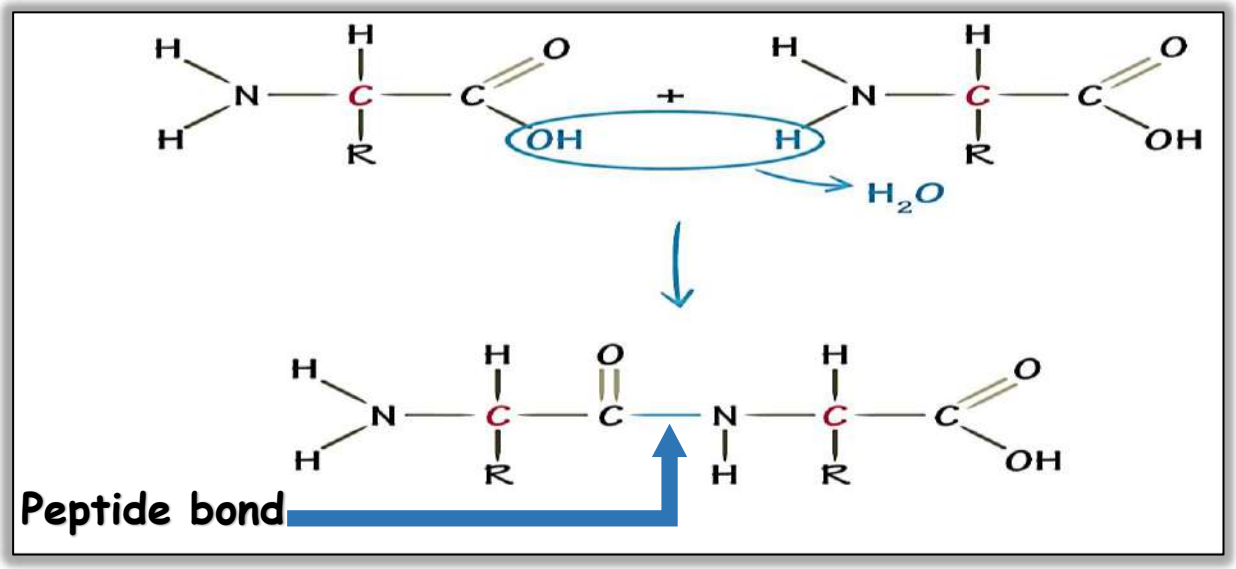



**Figure 02.** Peptide bond formation **(credit: Biochemistry Glossary).**

## II. Levels of protein structure

Protein structure is categorized in terms of four levels:

### II.1.Primary structure

It is the simplest level of protein structure, The primary structure is simply the sequence of amino acids that make up the polypeptide chain linked by peptide bonds as mentioned earlier. Because of the nature of the peptide bond, the backbone of a polypeptide will have a single primary amine at one end and a single carboxylic acid at the other end (they do not take part in a peptide bond), The two ends of each polypeptide chain are known as the amino terminus (N-terminus) and the carboxyl terminus (C-terminus) and The sequence of a polypeptide is always read from the N-terminus to the C-terminus **(Figure 03).** And the size (length), and specific amino acid sequence of a protein are major determinants of its shape, and the shape of a protein is critical to its function. (**OpenStax College, Biology**). (**Figure 04).**

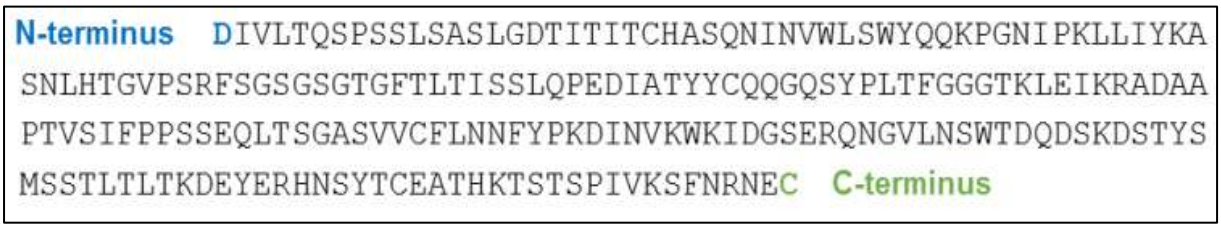



**Figure 03.** The sequence of a polypeptide chain from an antibody, with the N- and the C-terminus. (**credit: Structural Bioinformatics Group).**

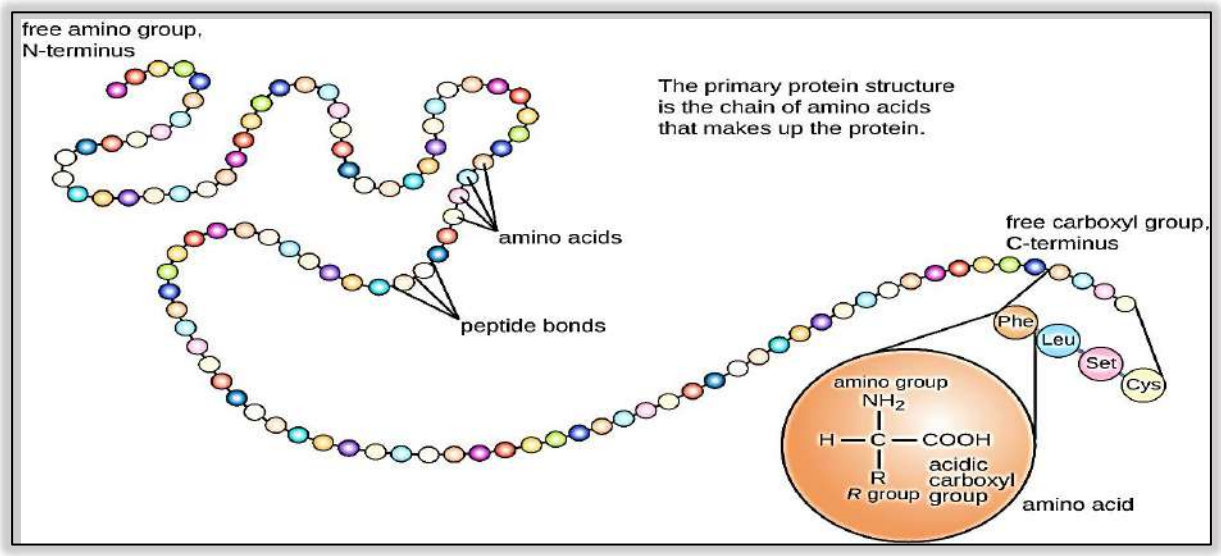



**Figure 04.** The primary structure of a protein is the sequence of amino acids. **(credit: modification of work by National Human Genome Research Institute)**

## II.2. Secondary structure

Proteins are dynamic entities, and they possess inherent flexibility because of the nature of the bonds that hold the amino acids together. Local folding of the polypeptide chain into helices and sheets occurs when the chain is sufficiently long, by forming a hydrogen bond between amine and carbonyl functional groups within the peptide backbone (except for the R side group), These shapes constitute a protein's secondary structure the most common secondary structures namely the α-helix and β-pleated sheet and turns**.**

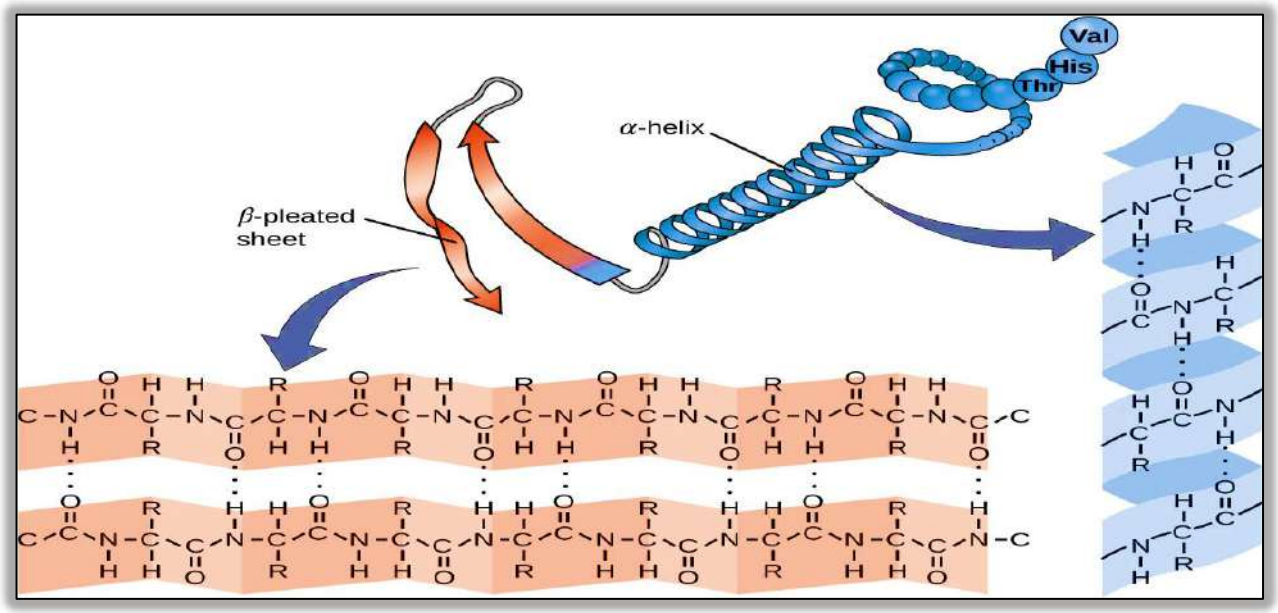



**Figure 05.** The secondary structure of a protein may be an α-helix or a β-pleated sheet or both

### II.2.1 Types of secondary structure

#### II.2.1.A The α-helix:

The alpha-helix is a right-handed helical coil, the helix is held by hydrogen bonds between the oxygen atom in a carbonyl group of one amino acid and the hydrogen atom of the amino group that is just four amino acid units farther along the chain. (**Ibraheem R et al,2021)**

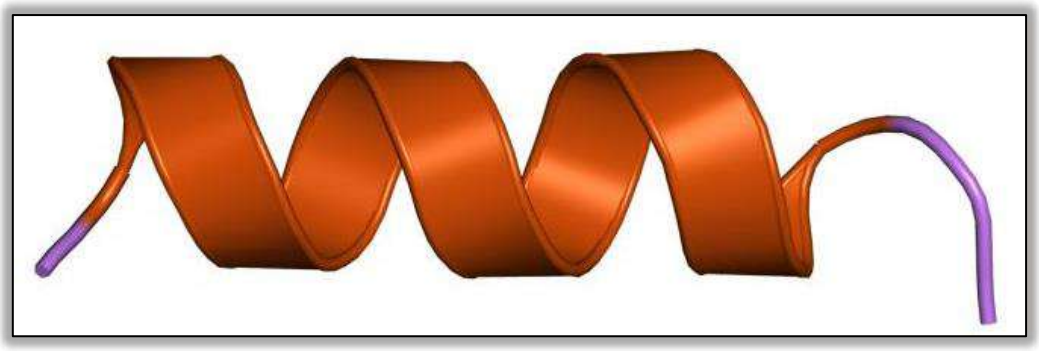


**Figure 06.** Alpha- Helix (**Jawahar Swaminathan, 2008)**

**Types of Alpha Helix:** There are three types, (Each of the three examples below in the table is a decapeptide fragment derived from a protein structure in the Protein Data Bank (PDB)):

**Table 01.** Three types of alpha-helix **(Lam et al., 2020)**

| **$3_{10}$ helix** | α-helix | π-helix |
|---|---|---|
|  |  |  |
| It has 3 residues per helical turn in a 10 atom ring**,** it provides insight into the initiation of α-helix folding. Because of the α-helices tendency. | It has 3.6 residues per helical turn and has 13 atoms in the ring formed by the hydrogen bond and so can also be called a 3.6-13 helix. | Is even less common. It is a wider helix with 4.6 residues per turn. |
| **F**rom **3l79: 514-525** | From **1HHO chain B: 5-16** | From **2QD3 chain A:346-357** |

### II.2.1.B The β-pleated sheet

William Astbury was the first to propose the first β-sheet structure in the 1930s, Beta sheets consist of beta strands (β-strands) connected laterally by at least two or three backbone hydrogen bonds, forming a generally twisted, pleated sheet. A β-strand is a stretch of polypeptide chain typically 3 to 10 amino acids long with backbone in an extended conformation. It is called the pleated sheet because of the wave like appearance.**(Manske M., 2001)**

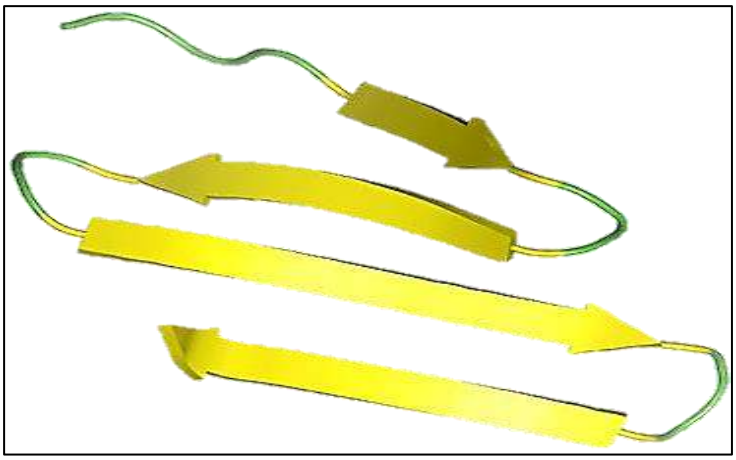


**Figure 07.** Beta-sheet **( Lever.G, 2015)**

**Types of Beta Sheets:** There are three types, parallel and anti-parallel sheets,β-Barrel moti, Parallel beta-sheets are chains of polypeptides that run in the same direction. Anti-parallel beta-sheets are chains of polypeptides that run in opposing and alternating directions. **figure 06**

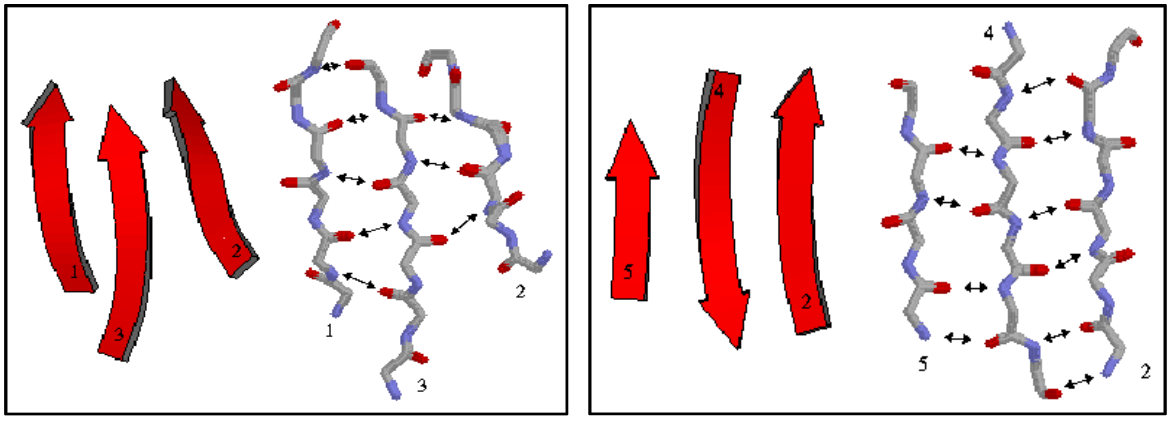



**Figure 08.** Parallel beta-sheet Antiparallel beta-sheet

The three parallel and anti-parallel strands are shown in both cartoon format (left) and in stick form containing backbone atoms N, CA, C, and O' (right). Hydrogen bonds are identified by

arrows connecting the donor nitrogen and acceptor oxygens. Strands are numbered according to their relative position in the polypeptide sequence. **(J.S. Richardson & D.C. Richardson, 1992)**

**Beta barrel:** It is made up of tandem repetitions that twist and coil to form a closed cylinder structure, in which the last strand is hydrogen bonded to the first strand.

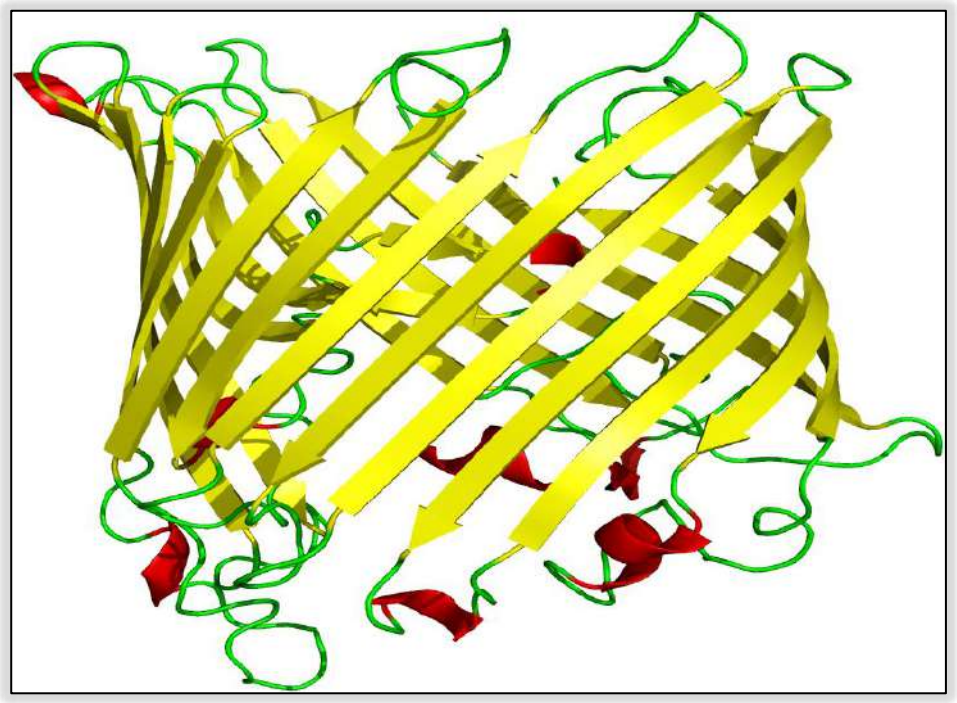


**Figure 09.** Beta Barrel **(PDB: 1A0S)**

### II.2.1.C Turns and Loops

A turn is a structural motif where the Cα atoms of two residues separated by a few (usually 1 to 5) peptide bonds are close (less than 7 Å [0.70 nm])**.** colored by element **figure 08**.

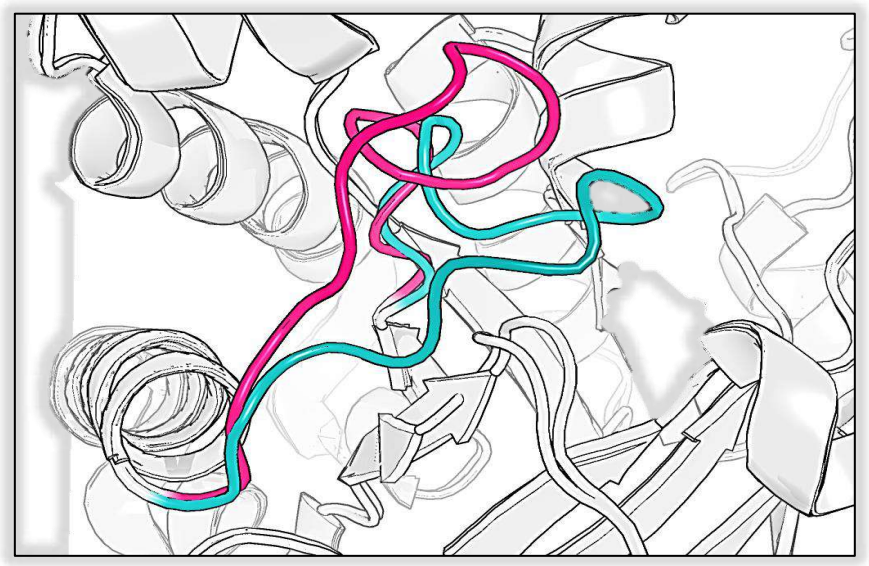


**Figure 10.** Turns and Loops : **(PDB- entry 1TPD)**

### II.2.2 Ramachandran Plot

Ramachandran plot is a two-dimensional (2D) plot of the torsional angles of amino acids φ (phi) and ψ (psi) in a protein sequence, with allowed regions for conformations where there is no steric interference. **(Pan A et al., 2021)**

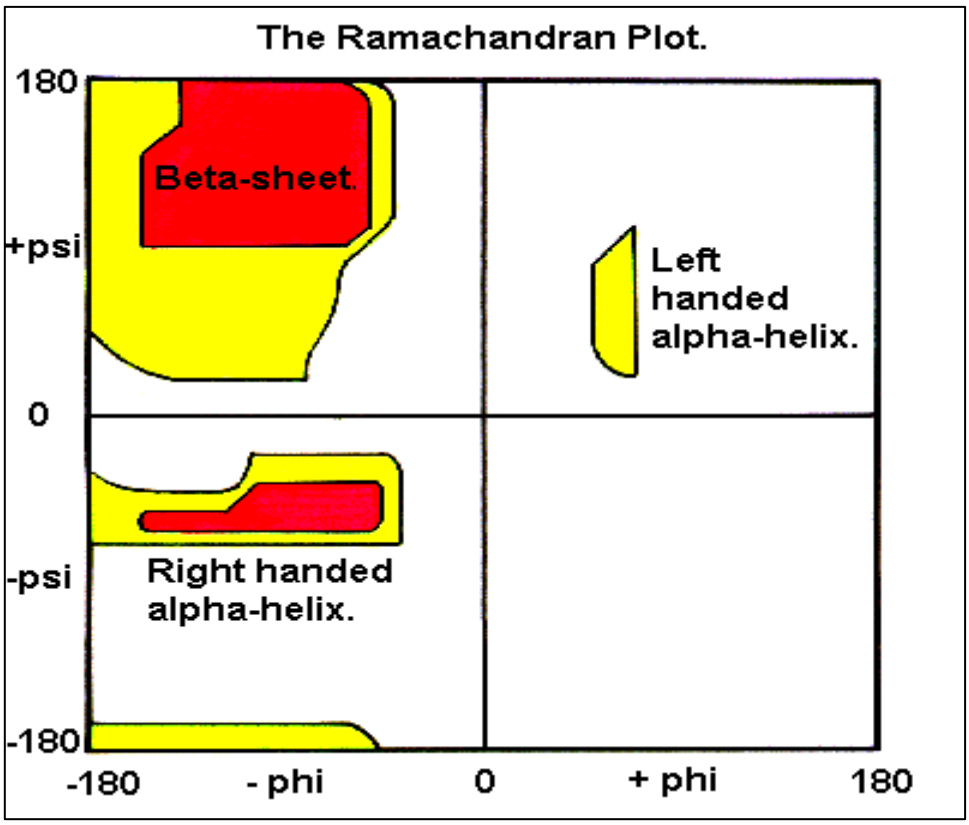



**Figure 11.** Importance and determinants of Secondary structure (Ramachandran plot) **(Pan A et al., 2021)**

### II.3.Tertiary structure

Is called three Dimensional structure of the polypeptide, it's due to the interactions between the R groups of the amino acids, in this level the polypeptide chains become functional by presenting a functional group on its outer surface which allows her to get a function of her own and a specific 3-Dimensional shape, it takes many forms (Globular Proteins which the Most proteins fall into this category, by forming a compact ball shape, and Fibrous Proteins which made of fibers often consisting of repeated sequences of amino acids), We can observe interactions that make up the tertiary structures of proteins that are covalent and non-covalent They guide the twisting and bending that help the protein molecule achieve a stable state. **(Figure 12). (Larry Li, 2014)**

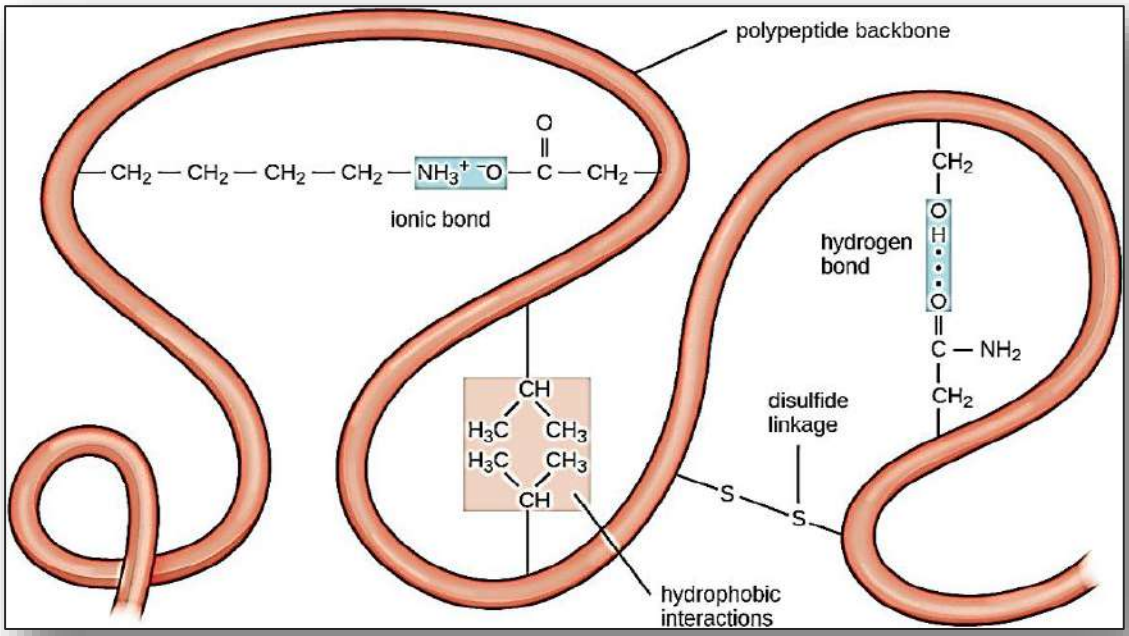



**Figure 12.** The tertiary structure (variety of attractive forces).(**Credit: OpenStax Biology).**

### II.4.Quaternary structure

Protein quaternary structure is the fourth (and highest) classification level of protein structure. Protein quaternary structure refers to the structure of proteins that are themselves composed of two or more smaller protein chains.

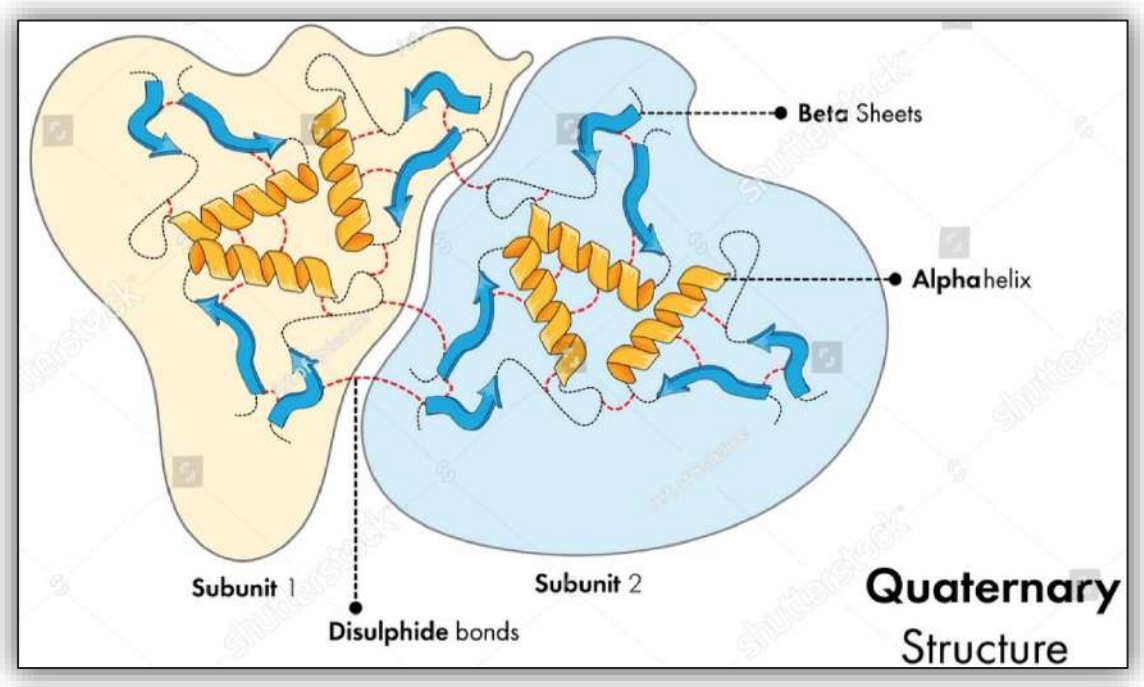



**Figure 13.** Quaternary Structure.**(Uday H,.n,d)**

### III. How Do Proteins Arrive at Their Final Shapes?

Proteins achieve their ultimate structures without any energy input once their amino acids are linked together. and there are proteins called chaperones that assist in protein folding and have the ability to prevent non-specific aggregation by binding to non-native proteins. **(Hartl, F. U& Hayer M,. 2009).**

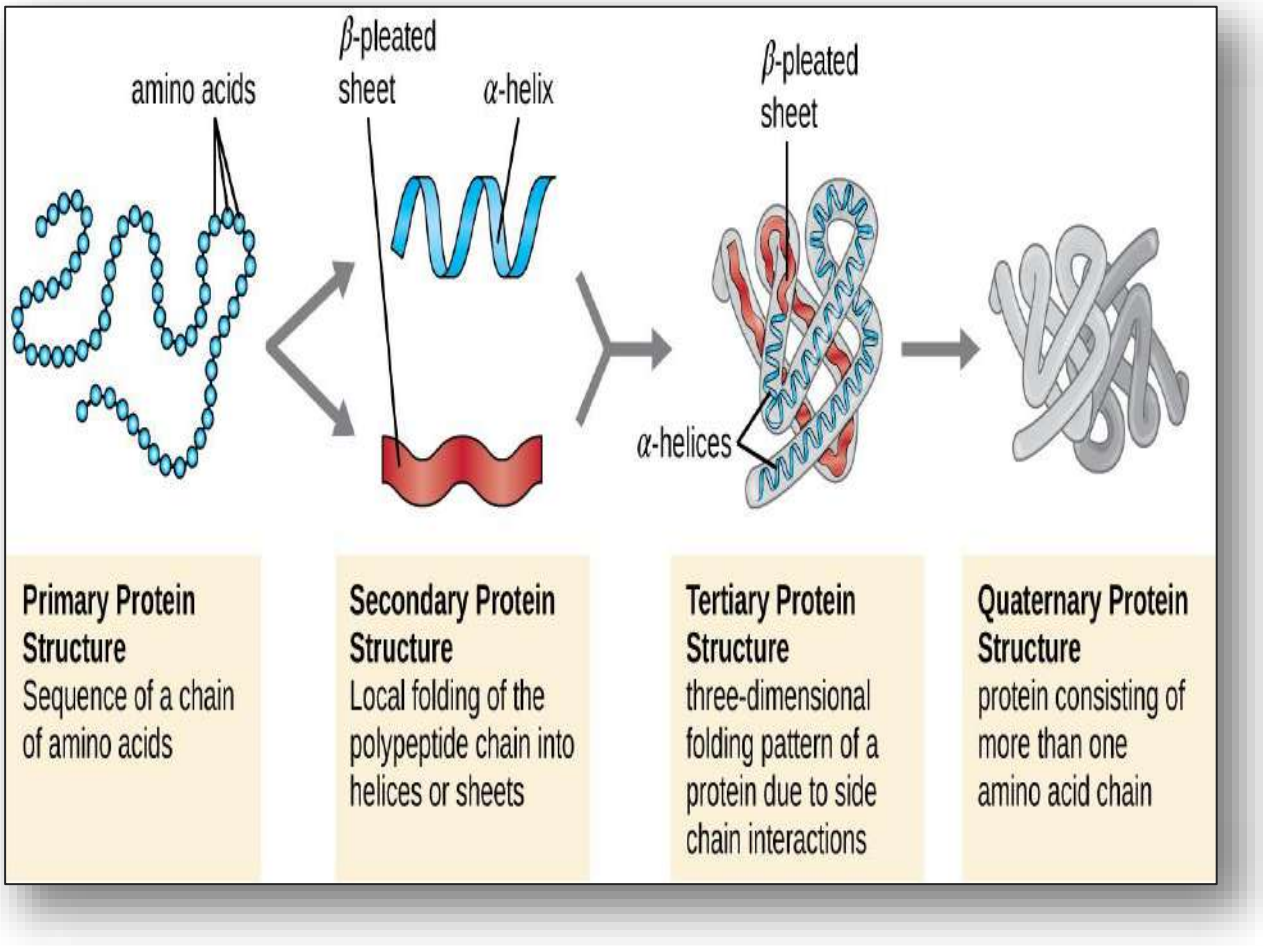
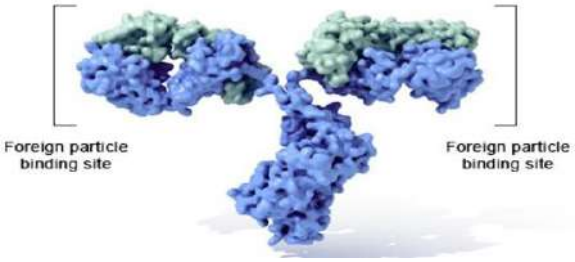
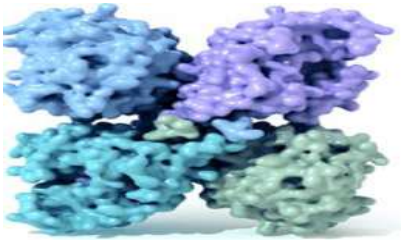



**Figure 14.** Protein structure has four levels of organization **(credit: CNX OpenStax).**

## IV. Protein classification:

Proteins can be described according to their large range of functions in the body, listed in alphabetical order:

**Table 02.** Examples of protein functions (**Marcella M, Meredith T., 2012 )**

| Function | Description | Example |
|---|---|---|
| **Antibody** | Antibodies are protective proteins produced by the immune system in response to the presence of a foreign substance, such as viruses and bacteria. | **Immunoglobulin G (IgG)**   |
| **Enzyme** | Enzymes help to make new molecules by reading the genetic information in DNA, and speed up almost the thousands of chemical reactions in cells. | **Phenylalanine hydroxylase (4sub units)**  |
| **Messenger** | Such as hormones, which transmit signals of biological processes between cells, tissues, and various organs. | **Growth hormone**  |
| **Structural component** | These proteins give cells shape and support. They also allow the body to move | **Actin (multiple sub units)**  |
| **Transport/ storage** | These proteins bind and carry atoms and small molecules within cells and throughout the body. | **Ferritin (24 subunits)**  |

## V. Dihydrofolate reductase enzyme (DHFR)

### V.1. Enzymes

Enzymes are macromolecular biological catalysts that catalyze chemical reactions. The molecules at the start of the process on which enzymes can function are known as substrates, and the enzyme changes these molecules into new molecules known as products. **(Stryer L et al., 2002)**

### V.2. Background of Dihydrofolate Reductase

(EC 1.5.1.3.) is a key enzyme in folate-mediated one-carbon metabolism, Dihydrofolate Reductase was discovered in the late 1950s. Because of DHFRs vital role in DNA synthesis, it was targeted for cancer chemotherapy; and it was chosen as the Protein Data Bank's Molecule of the Month for October of 2002. **(Futterman, S. 1957)**

DHFR is a small enzyme that plays a supporting role, such as producing cofactors that are necessary for DNA synthesis, and is found ubiquitously in all dividing cells of prokaryotes and eukaryotes but each being makes a slightly different version of the other, the version from bacteria, shown on the left is smaller and more streamlined than the version in our own cells, shown on the right with PDB entries, As seen in these structures, both bind similarly to NADPH, **(Smith SL et al., 1979)**

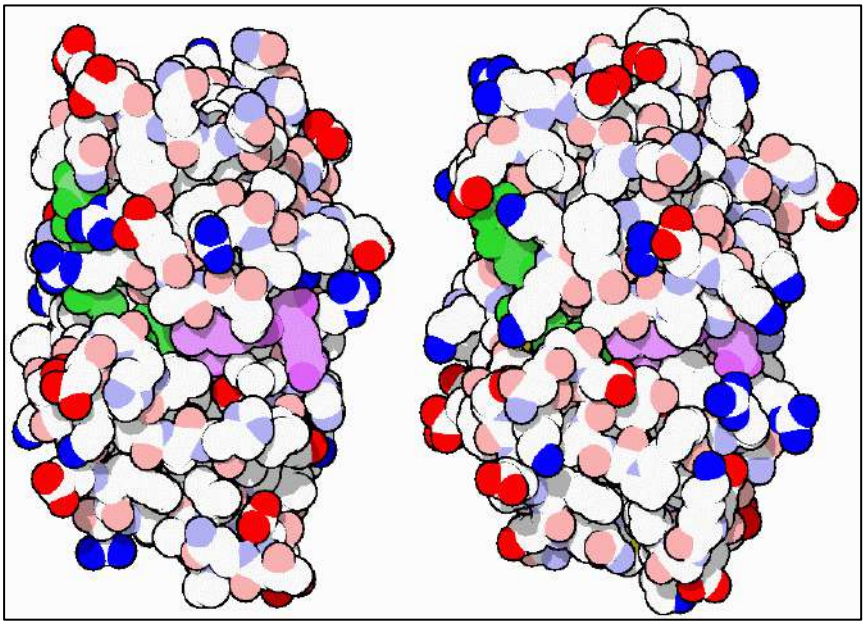
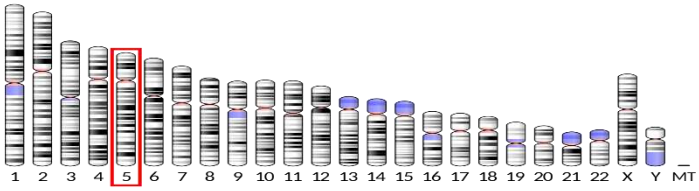


**Figure 15.** Bacterial left (3DFR) and human right (1DLS) dihydrofolate reductase **(Goodsell., 2002)**

### V.3. Human DHFR

The following table shows information about the dihydrofolate reductase enzyme in humans:

**Table 03.** General information of dihydrofolate reductase enzyme gene, structure, and active site. **(Abali EE et al., 2008)**

| | |
|---|---|
| **The enzyme gene** | DHFR gene, found in chromosome 5 (region: q11→q22).   |
| **Length and weight** | 186 amino acids, with a molecular weight of around 20 kDa. |
| **Structure studies** | A monomeric molecule with many secondary structural elements (eight-stranded beta-sheet and four alpha-helices with connecting loop regions). |
| **Subdomains** | Is divided into two:<br>• **the adenosine-binding subdomain:** is the larger one and binds the adenosine moiety of NADPH.<br>• **the loop subdomain:** contains three loops, the Met-20 loop, the F-G loop, and the G-H loop. |
| **Active site** | the active site is between the two subdomains along groove where folate and NADPH bind, and the movements of the two subdomains regulate the size of the active size. |

#### V.3.1. What is the function of dihydrofolate reductase DHFR?

DHFR has a critical role in regulating the amount of tetrahydrofolate in the cell. Tetrahydrofolate and its derivatives are essential for purine and thymidylate synthesis which are important for cell proliferation and cell growth, The physiological function of (DHFR) is converting dihydrofolate into tetrahydrofolate. **(Abali EE et al., 2008)**

### V.3.2. Mechanism of action

This enzyme has a steady-state random and stepwise mechanism, Specifically, the catalytic reaction begins with the NADPH, where DHFR catalyzes the transfer of a hydride from NADPH to dihydrofolate (its promote by the Met20 loop which helps stabilize the nicotinamide ring of the NADPH) **(figure 14)**, followed by protonation to produce tetrahydrofolate. In the end, dihydrofolate is reduced to tetrahydrofolate and NADPH is oxidized to NADP+. (**Osborne MJ et al., 2001)**

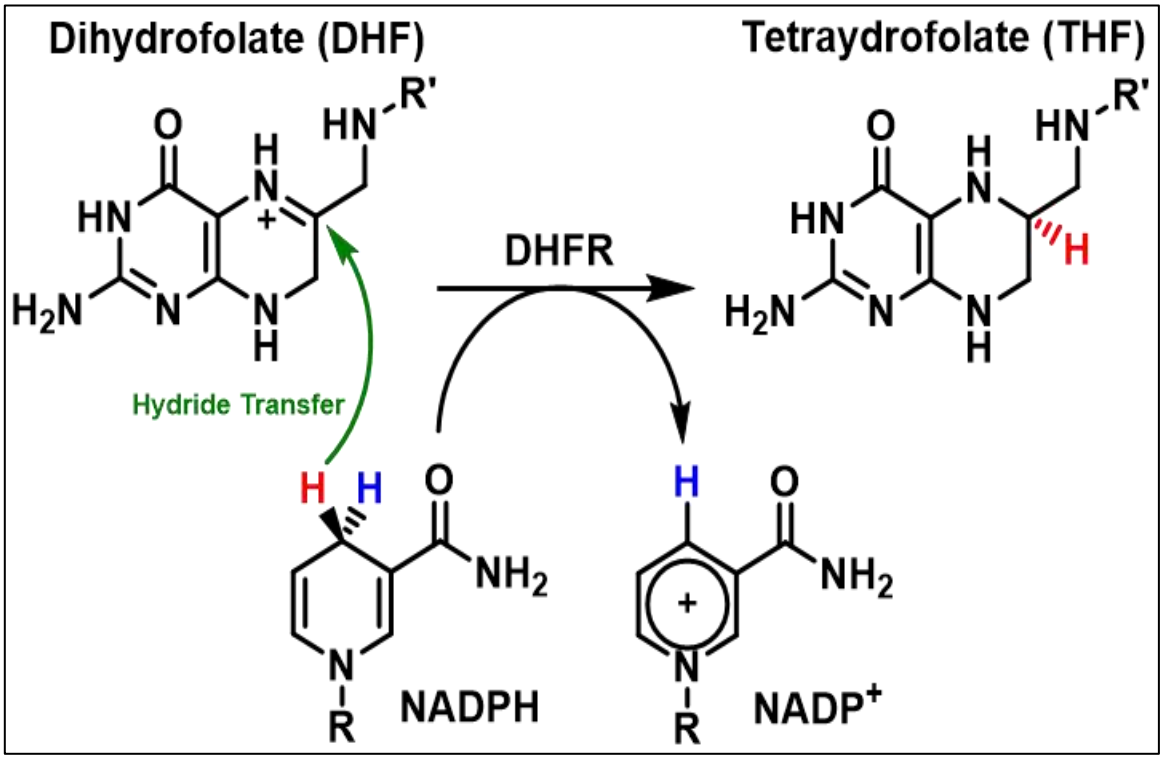



**Figure 16.** The reduction of dihydrofolate to tetrahydrofolate catalyzed by DHFR.

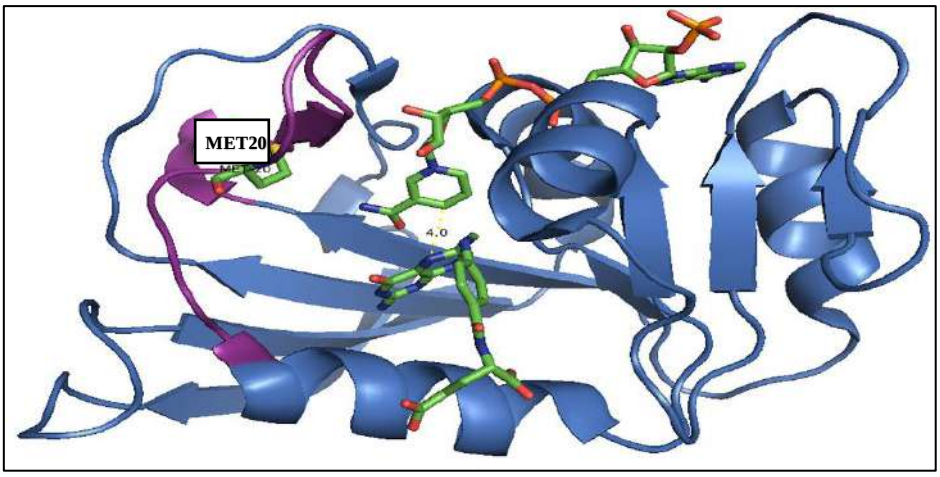



**Figure 17.** DHFR (Met20 loop) + NADPH + folate.

### V.3.3. DHFR uses in biotechnology

Dihydrofolate reductase is a crucial enzyme for the synthesis of purines, pyrimidines, and some amino acids. Biochemicals that act as substrates for DHFR have many applications in biochemical and physiological research, for example, Because of its critical role in DNA precursor synthesis, DHFR is an interesting pharmaceutical target for inhibition. **(Cowman AF & Lew AM,.1989)**

### V.3.4. Dihydrofolate reductase as a therapeutic target

For several years, the DHFR enzyme has been recognized as a therapeutic target in cancer treatment, and new compounds are being discovered having the ability to inhibit this enzyme continues to pique the interest of medicinal chemists. Recently discovered dual compounds that inhibit DHFR as well as other folate receptors (FRs) and have interesting antitumor properties, Aminopterin was the first drug used in cancer chemotherapy. It binds to dihydrofolate reductase 1,000 times stronger than folate and blocks its action. **(Goodsell, 2002).**

### V.3.4. Relevance of DHFR Inhibitors in therapy

DHFR inhibitors are used to treat fungal, bacterial, and mycobacterial infections, as well as malaria and other protozoal infections. **( Srinivasan B et al.,2019).**

The table below shows two examples of DHFR inhibitors that have been approved.

**Table 04.** DHFR inhibitors that have been approved. **(Raimondi M et al.,2019).**

| Antifolate | Status | Indication | Toxicity |
|---|---|---|---|
| **Methotrexate** | Approved in 1985 by FDA and EMA. | Treatment of lymphoma, acute lymphoblastic leukemia, osteosarcoma | Symptoms of overdose include bone marrow suppression and gastrointestinal side effects |
| **Pralatrexate** | Approved in 2009 by FDA and EMA . | Treatment of relapsed or refractory peripheral T-cell lymphoma (TCL) | Mucositis |

## VI.Methods for determining protein structures

This field includes many techniques, and it can be determined by either experimental methods or computational methods, Each method is characterized by certain controls, in this part, the various methods will be described briefly.

## VII. Experimental methods

X-ray crystallography and NMR spectroscopy and cryo-electron microscopy are currently the three major experimental techniques for protein structure determination all of them are, however, time- and manpower-consuming, the most used technique is X-ray crystallography with over 89% of structures deposited in Protein Data Bank, followed by NMR over 8%, and Cryo-EM about 2,5%, and the remaining 0,5 structures were solved using other methods. **(Mutharasappan N et al., 2020)**

### VII.1. X-Ray Crystallography

#### VII.1.1. Definition

The goal of x-ray crystallography is to produce a three-dimensional molecular structure from a crystal. Using the measured x-ray diffraction intensities, Based on the angles and intensity of those deflected rays, the crystallizer produces a three-dimensional picture of the electron density inside the crystal. Based on this electron density, the locations are known as an arithmetic mean of the atoms within the crystal, as well as their chemical bonds, entropy, and other information. . **(Smyth, M. S., Martin, J., 2000).** As first it used to determine the three-dimensional structures of inorganic materials, then small organic molecules deal with a small number of atoms, and finally, macromolecules like DNA and proteins determine a much larger number of atomic positions.**(Drenth J, 1994).**

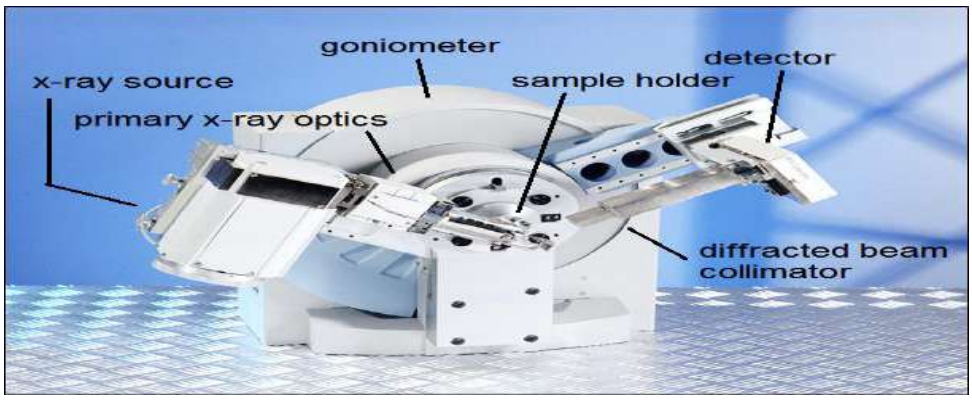



**Figure 18.** The main components of a X-Ray powder diffractometer including a goniometer **(courtesy of Malvern-PANalytical B.V.n,d)**

### VII.1.2. Protein X-ray crystallography:

Protein crystallography requires a consistent source of protein as well as a purification/concentration technique that yields high-quality, homogenous, soluble material. It starts with obtaining crystals this step remains a bottleneck for this technique including obtaining large amounts of pure protein through well-established molecular biology experiments such as molecular cloning and affinity chromatography. Then Once crystals of suitable size and composition are obtained, it is necessary to bombard the crystal with x-rays and observe the diffraction pattern, Dr. Ten Eyck said that we can use the fast fourier transform (FFT) for the calculation of structural factors after having the data of amplitude and phases.the finally improving the quality of the electron density map by the refinement step. **(Mutharasappan et al., 2020)** These steps are illustrated in the figure below

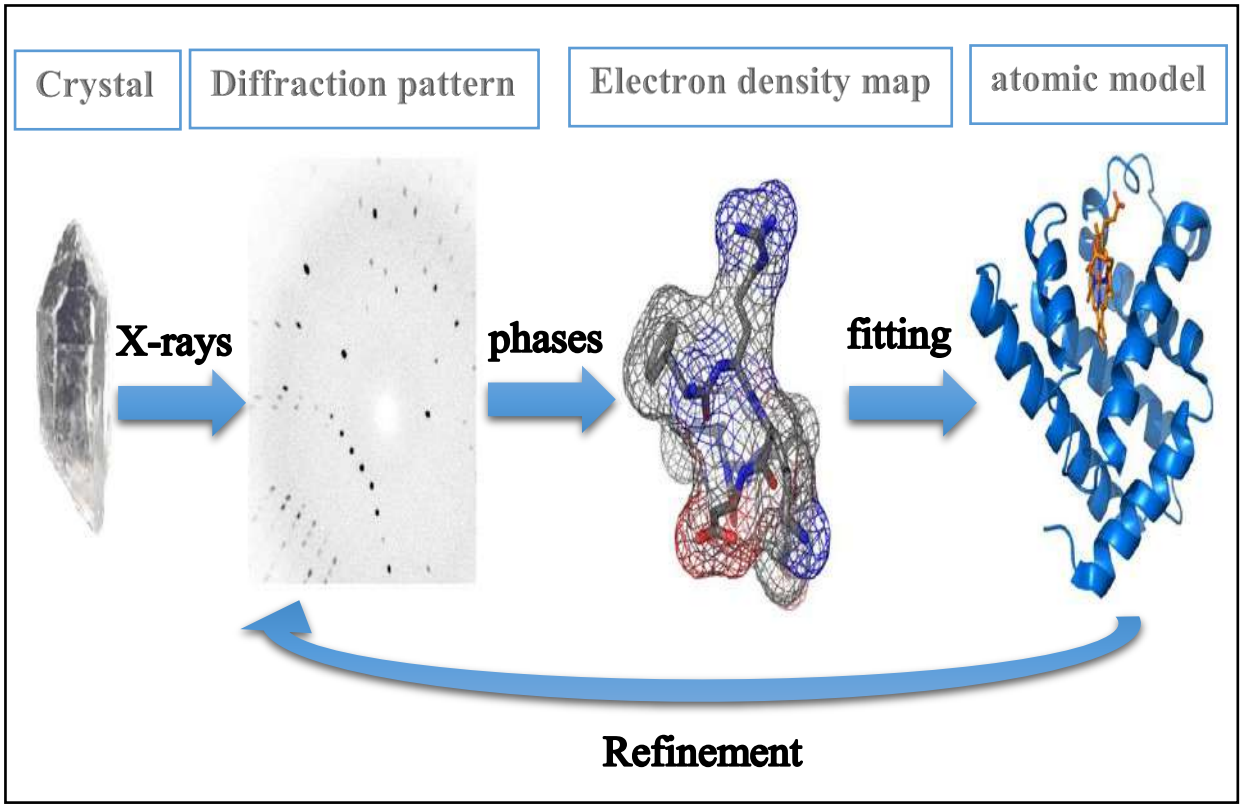



**Figure 19.** Workflow for solving the structure of a molecule by X-ray crystallography **(Drenth J, 1994).**

## VII.2 Nuclear Magnetic Resonance (NMR)

### VII.2.1 Definition

Nuclear magnetic resonance (NMR) spectroscopy is one of the most powerful physicochemical analysis techniques used in quality control and research for determining the content and purity of organic, organometallic, and biochemical molecules. Generally, the identification of compounds is complemented with data obtained with other techniques, NMR spectroscopy can be used as a quantitative analysis tool when the proportionality between the area of the signals and the number of nuclei that generate it. The time period of the NMR spectrometer is relatively long, so it is not suitable for observing fast phenomena and only gives an average spectrum. **(Héctor Zamora,C., 2021)**


**Figure 20.** Components of an NMR equipment**. (García Álvarez et al., 2016)**

### VII.2.2. Principle

According to the NMR principle, all nuclei have a charged electrically, and spin, when an external magnetic field is applied Transfer of energy is possible from base energy to higher energy levels, and The transfer of energy occurs at a wavelength that coincides with the radio frequency Also, energy is emitted at the same frequency when the spin returns to its base level. As a result, the processing of the NMR spectrum for the concerned nucleus is yielded by measuring the signal that fits this transfer. **(Aryal S., 2021)**

## VII.3. Cryo-electron Microscopy

### VII.3.1. Definition

Cryo-electron microscopy (cryo-EM) is an advanced structural molecular and cellular biology technique used to visualize the structural features of proteins, This technique complements x-ray crystallography because it reveals structural details without the need for a crystalline specimen. In this technique, specimen preparation is an important step where in the solution the biomolecules are rapidly frozen, Cryo-EM can determine structures with specimens starting molecule weight of around 50 KDa which is the upper limit of NMR. **(Mutharasappan et al., 2020).**

**Figure 21.** Machine of the Cryo-EM **(Elizaveta Galitckaia,.2019)**

### VII.3.2 Principle

Cryo-EM works on the principle of transmission electron microscopy (TEM) imaging the radiation-sensitive samples under low-temperature conditions, including the basic steps of sample preparation, and image processing, structure analysis. To Perform Cryo-EM start with the first step is t purify, in which a buffer solution must be used that maintains the biochemical activity of the sample to be pure and with high concentration particles, which makes it easy to study under the microscope. Then the stage of Plunge Freeze so that the sample freezes to be placed on a TEM grid, to be transferred to the stage of conversion to the frozen sample is placed on a specialized TEM holder to maintain the temperature of liquid nitrogen, then the transmitted electrons are detected and form recorded images that are magnified by factors. Computer programs then solve the detailed structure of the sample from the magnified images**. (Warwick,T,. 2018).**

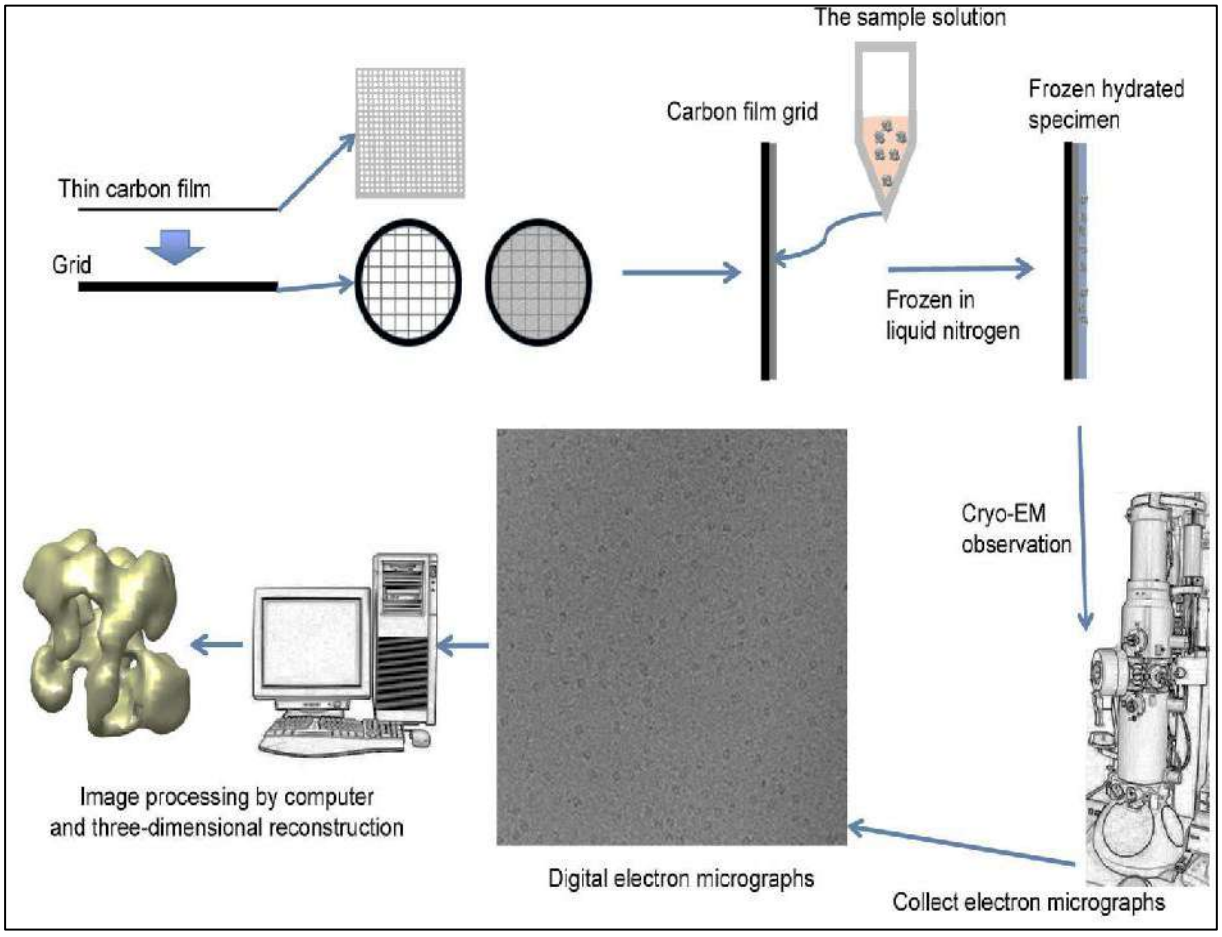



**Figure 22.** Workflow of cryo-EM structure determination. **(Wang H,. 2015)**

### VIII. Biological databases, and high-throughput data sources

Before addressing computational methods, this is an identification of some of the necessary sites and databases used in these approaches:

### VIII.1. Protein Data Bank (PDB)

It is the largest protein database that contains only experimentally resolved structures and is submitted by biologists and biochemists from around the world. used by various persons like scientists, bioinformatics, biologists and biochemists who use structure or sequence databases. Here we will learn protein databases (PDB), which have all known 3D structures of proteins. To find the PDB on the web search online by exploiting the webserver:

https://www.rcsb.org/

You can download the protein structure data (i.e, the PDB files) you need in your studies to your computer. On the PDB homepage type the keyword/name of your protein in the search bar, and in the left you can select the organism to filter out the proteins associated. Each PDB file has a special name and coordinates of all the atoms in the protein (x;y;z). **(Taimoor k, 2020)**

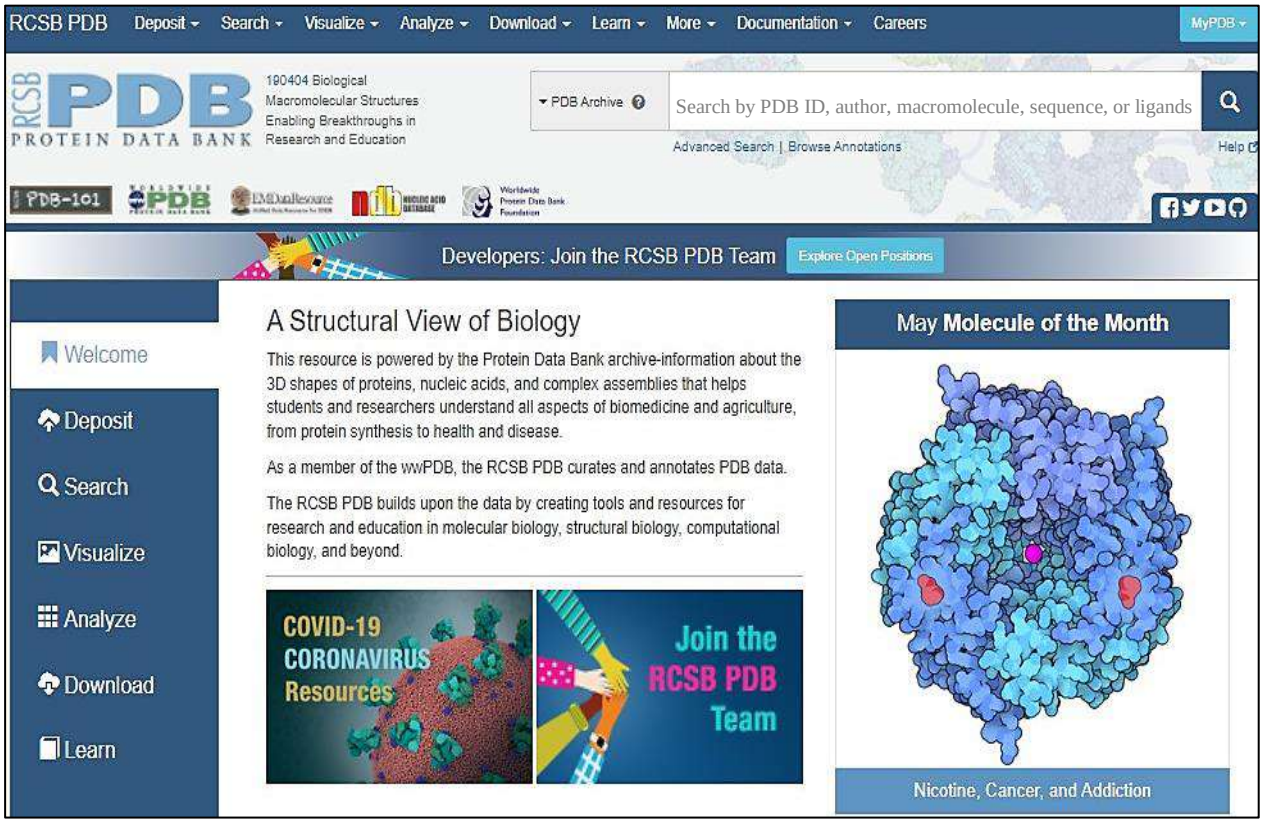



**Figure 23.** A screen shot of PDB – Protein Data Bank home page

Every structure in the PDB is given a four-character alphanumeric identification known as PDB ID (the PDB identifier). Eg: “1MVS”.

### VIII.1.2. PDB Members

The PDB is overseen by an organization called the Worldwide Protein Data Bank, wwPDB, and it has four members RCSB PDB, PDBe (Europe), PDBj (Japan), and BMRB (USA), to ensure the PDB archive is global and uniform. The RCSB PDB presently acts as the "archive keeper", The archive is updated once a week and it is distributed by wwPDB sites via FTP. **(Berman H et al.,2007).**

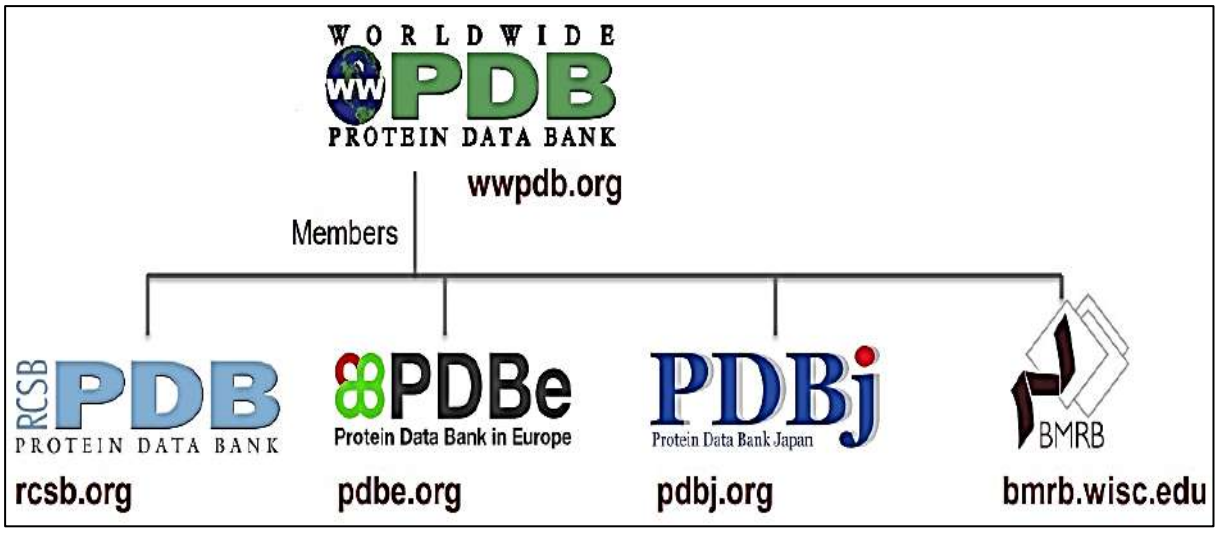



**Figure 24.** PDB Members**(Berman H et al., 2007).**

#### VIII.1.2.1 Protein Data Bank (RCSB)

Provides freely and publicly searching available to the global community for macromolecular structural data, Ligands, sequence-structure comparisons, 3D shapes of proteins, and nucleic acids. And also provides a renewed view of the molecule of the month, a feature for the phone called RCSB PDB Mobile, other educational resources at PDB-101, and more.

#### VIII.1.2.2 Protein Data Bank Europe (PDBe)

(PDBe http://www.ebi.ac.uk/pdbe/), is an important resource for high-quality molecular structures and related data and is also rich in information about all entries, It has many activities as an initiative "Structure Integration with Function, Taxonomy and Sequence" (SIFTS), and provides advanced visualization such Atlas pages, and validation of NMR and EM structures

tools for bioinformaticians, and different advanced services such as PDBe PISA, PDBe Fold, PDBe Motif. **(Velankar et al., 2009).**

### VIII.1.2.3 Protein Data Bank Japan (PDBj)

(PDBj, http://pdbj.org), provides a comprehensive range of services and tools for evaluating protein structures and functions, accepts and executes PDB entries mostly from Asia and Oceania, and browses in multiple languages, including Japanese, Chinese, and Korean. And more **(Kinjo A et al., 2011)**

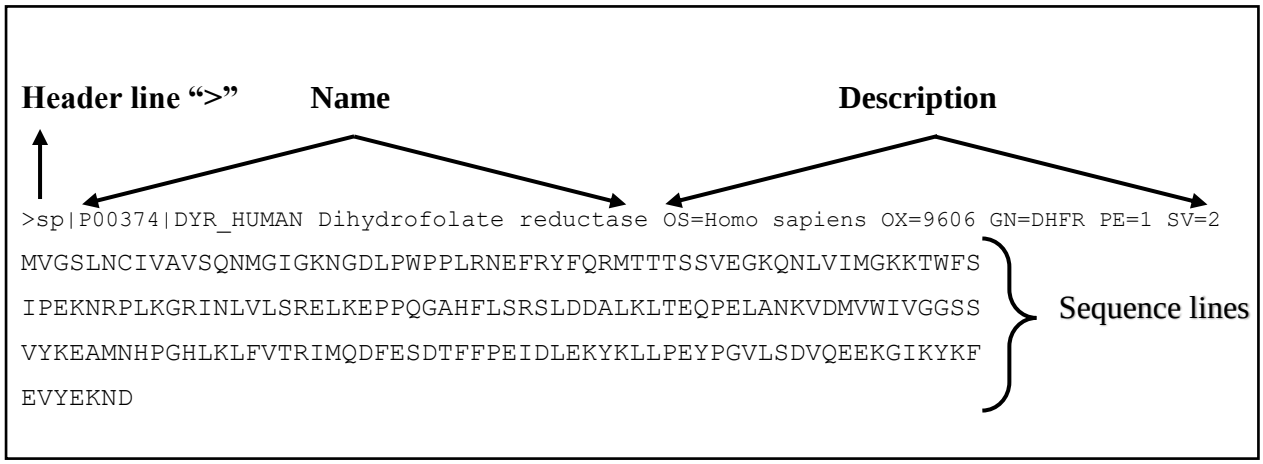
### VIII.1.2.4 Biological Magnetic Resonance Data Bank (BMRB)

BMRB gathers data from NMR spectroscopy studies of physiologically relevant compounds and distributes them ( in the public domain worldwide). The purpose is to help scientists analyze the structure, dynamics, and chemistry of biological systems. (**Eldon L et al., 2008).**

### VIII.2. What is a FASTA file?

A FASTA file is a text file for representing peptide or nucleotide sequences, The first line after the symbol ">" is a comment that provides relevant information about the sequence (the sequence name). and The remaining lines are the amino acid sequence in the single letter code**.** The amino acid codes supported (22 amino acids and 3 special codes) (**see Appendix part I)**.

#### VIII.2.1 Example of FASTA file:




#### VIII.2.2 How do you find FASTA sequence?

There are many ways, for example, can be obtained by accessing the Uniprot website, and after choosing the required protein page, there will be a button at the bottom that directs you to FASTA file as shown :

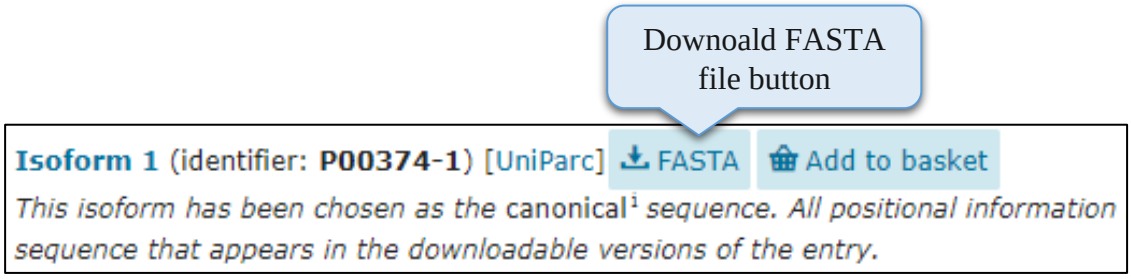



**Figure 25.** A screenshot of FASTA file from UniProt site.

And the other way is from the NCBI, by Accessing the Download option on the toolbar to download FASTA sequence and other data.


**Figure 26.** A screenshot of the FASTA file from NCBI site.

### VIII.3. BLAST (Basic Local Alignment Search Tool) – NCBI

Developed by the National Center for Biotechnology Information which is a part of the National Library of Medicine, It is an algorithm available on the site https://blast.ncbi.nlm.nih.gov/Blast.cgi, It compares the sequences of proteins and nucleotides, "query" with other databases of protein and nucleotides. so that it detects short matches between the two sequences, aligns, and gives the statistical information about them, this is called the "expect" value, When the E-value for sequence alignment is 0.05 indicates that this similarity has a 5 in 100 (1 in 20) probability of happening by chance alone.

NCBI's WebBLAST offers four main search types: **(figure 23)**

- **BLASTn (Nucleotide-nucleotide BLAST):** It makes a comparison between DNA (nucleotide) "query" against a database of nucleotide "subject" sequences that the user specifies.
- **Blastx (Nucleotide 6-frame translation- protein):** Compares a nucleotide query sequence resulting in six-protein sequences against a database of protein.
- **tBlastn:** compares a protein query sequence against the six-frame translations of a database of nucleotide sequences.
- **Blastp (Protein BLAST):** compares one or more protein query sequences to a subject protein sequence or a database of protein sequences.

BLAST has adaptations in blastp such as PSI-BLAST (for iterative protein sequence similarity searches using a position-specific score matrix) and RPS- BLAST(for searching for protein domains in the Conserved Domains Database). **(Johnson M et al., 2008)**

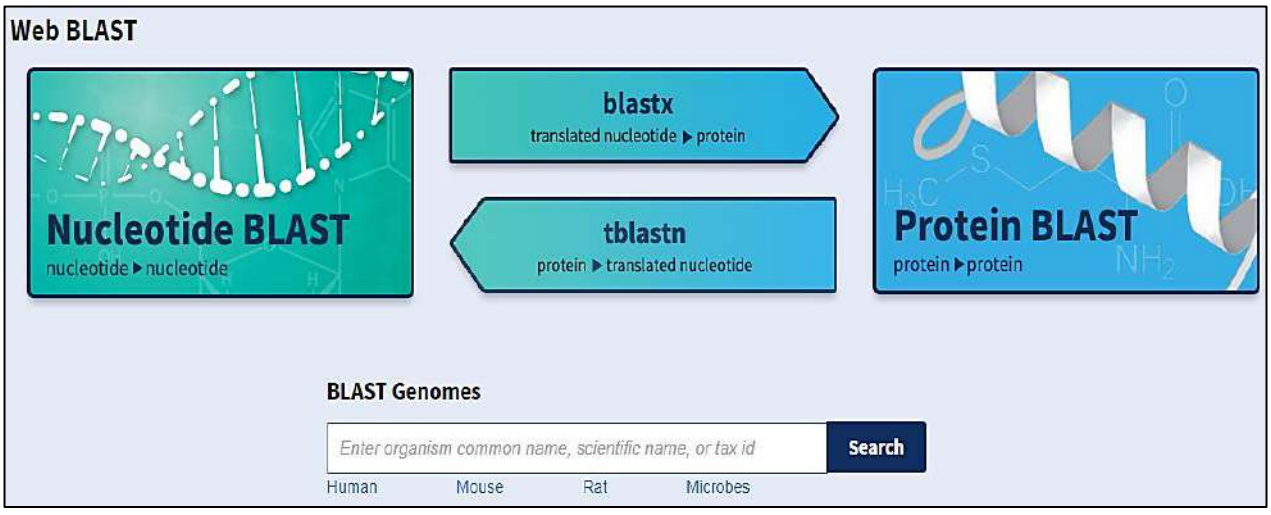



**Figure 27.** BLAST homepage showing the four options of BLAST.

### VIII.4 UniProt

UniProt is the universal protein resource, maintained and produced by the UniProt Consortium a collaboration between several European bioinformatics organizations (EBI), (SIB), (PIR). It provides a database containing large information about protein sequence and biological processes and obtained many entries from genome sequencing projects. And progress three key databases: UniProtKB (with sub-parts Swiss-Prot and TrEMBL), UniParc, and UniRef. **(Pundir S et al., 2016).**

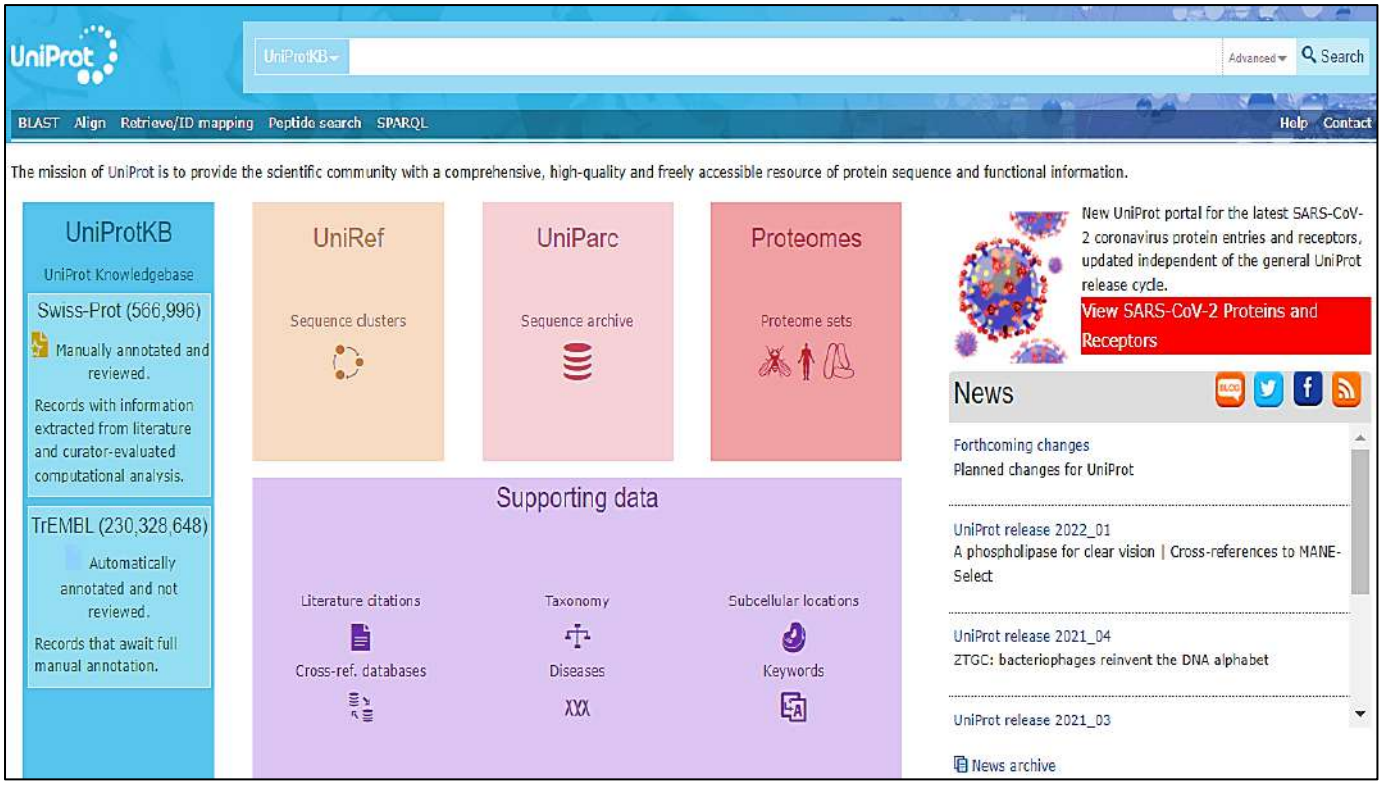



**Figure 28.** A screen shot of UniProt web page

#### VIII.4.1 Uniprot Tools

It has five major tools available on their own dedicated pages on the UniProt website :

#### VIII.4.1.a. Align (Multiple Sequence Alignment Tool)

Alignment means the writing process of two or more sequences, one on top of the other to find areas of similarity in the entries being aligned and make identities appear. Each alignment corresponds to an id% score, which can be calculated as the identity percentage (number of identities/length of alignment) during sequence editing. **(Deléage G et al., 2021)**

#### VIII.4.1.b. BLAST (Basic Local Alignment Search Tool)

Discovers areas (regions) of local similarity between sequences, The evolutionary and functional relationships between the sequences are identified by the BLAST results and more. Provides also a Running BLAST on a protein from the protein entry page.

#### VIII.4.1.c. Retrieve/ID Mapping

Allows to 'map' or convert identifiers from UniProt to Hundreds of external databases, to which UniProt is referred, It also contains many different databases such as Genome annotations, Sequence, 3D Structure, and more**.**

#### VIII.4.1.d. Peptide search

It is a tool that provides to input peptide sequences with at least three residues and identify all UniProtKB sequences with an exact match to the query sequence.

#### VIII.4.1.e. SPARQL

It is an endpoint that enables asking complex queries for analytical or data integration reasons that are difficult or hard to do using the Entry oriented web services. **(Pundir S et al., 2016).**

### VIII.5 3-Dimensional Structural Superposition server (3d-SS)

A software or web-based interactive computing server that Superpose two or many 3D protein structures, Where it finds the common and constant water molecules in all the superposed homologous protein structures, The molecular visualization tools RasMol or Mol-soft ICM browser can be used with the server to visualize the superposed 3D structures by saving the superposed 3D atomic coordinates in the client machine.

This process is mainly done by entering the PDB-ID (s) of each protein or its load the atomic coordinates in PDB format. **(Sumathi K et al., 2006).**

The web server is reachable on the website:

http://cluster.physics.iisc.ernet.in/3dss/


**Figure 29.** 3d- SS (3-Dimensional Structural Superposition server) homepage.

There are four major options provided in the proposed computing server:

a. Superpose only two structures.

b. Superpose several structures.

c. Superpose subunits within a structure.

d. Superpose different models present in NMR ensemble.

All the above options, Allow selecting structures available in PDB either by loading them from the 3D atomic coordinates (PDB format) from the local hard disk or by entering its unique PDB-ID. After downloading, it automatically displays all the details of the structure string in a specific format. It can also select the full file, set the whole string, or specify part of the mod.

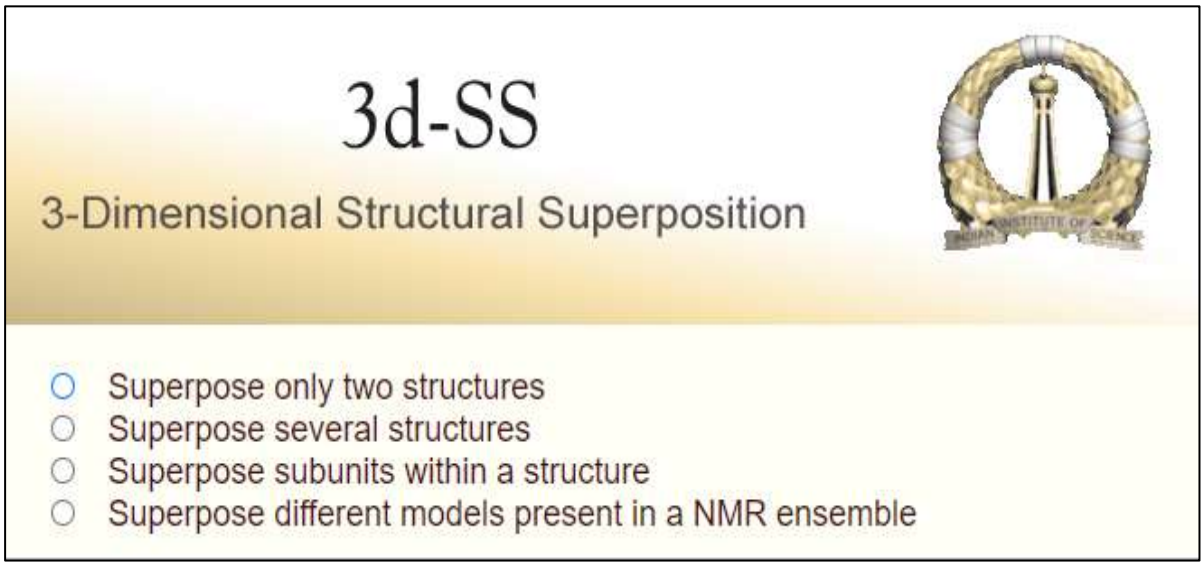



**Figure 30**. 3d- SS options from the web page **(Sumathi K et al., 2006).**

## IX. Computational and theoretical methods

Since the experimental method costs a lot and some proteins structure could not be found by this method, more practical and easier techniques have been resorted to by developing bioinformatics programs to predict the structures of proteins directly from a sequence Given an amino acid sequence (i.e., the primary structure), the objectives are to determine all helical segments and all beta-strands and residues.

The approaches are classified into three major theoretical methods for predicting the structure of proteins: Homology /comparative modeling, fold recognition, and ab initio prediction.

### IX.1. Homology modeling:

Also called Comparative modeling, it's one of the most used methods for protein structure prediction, it predicts the 3-D structure of the sequence of protein (amino acids) "target" based on its alignment to one or more proteins with a known structure as a "template" stored in the PDB . since proteins that have similar sequences usually have similar structures (50% or better sequence identity would be good), we can use the structural information of the known structure for modeling the target protein if we have a similarity. **(Dorn M et al.,2014)**

In this method, the 3D structure of a protein is obtained with the following six steps:

#### IX.1.1. Identification of target sequence

The sequence target is taken by experimental methods or from protein databases like UniProt.

#### IX.1.2. Sequence alignment

The second step is the alignment of the target and template sequences.

#### IX.1.3. Structure templates identification

Once optimal alignment is achieved, can copy the coordinates of the corresponding residues of the template proteins onto the target protein.

#### IX.1.4. Core structure calculation

At this stage, the selected sequences are used to find the invariant and common structural features present between their 3D-structures by superposition of the protein structures.

#### IX.1.5. Loop regions and residues building

This step includes the addition and optimization of side-chain atoms and loops.

### IX.1.6. Final model

The final step involves the overall quality of the model created.

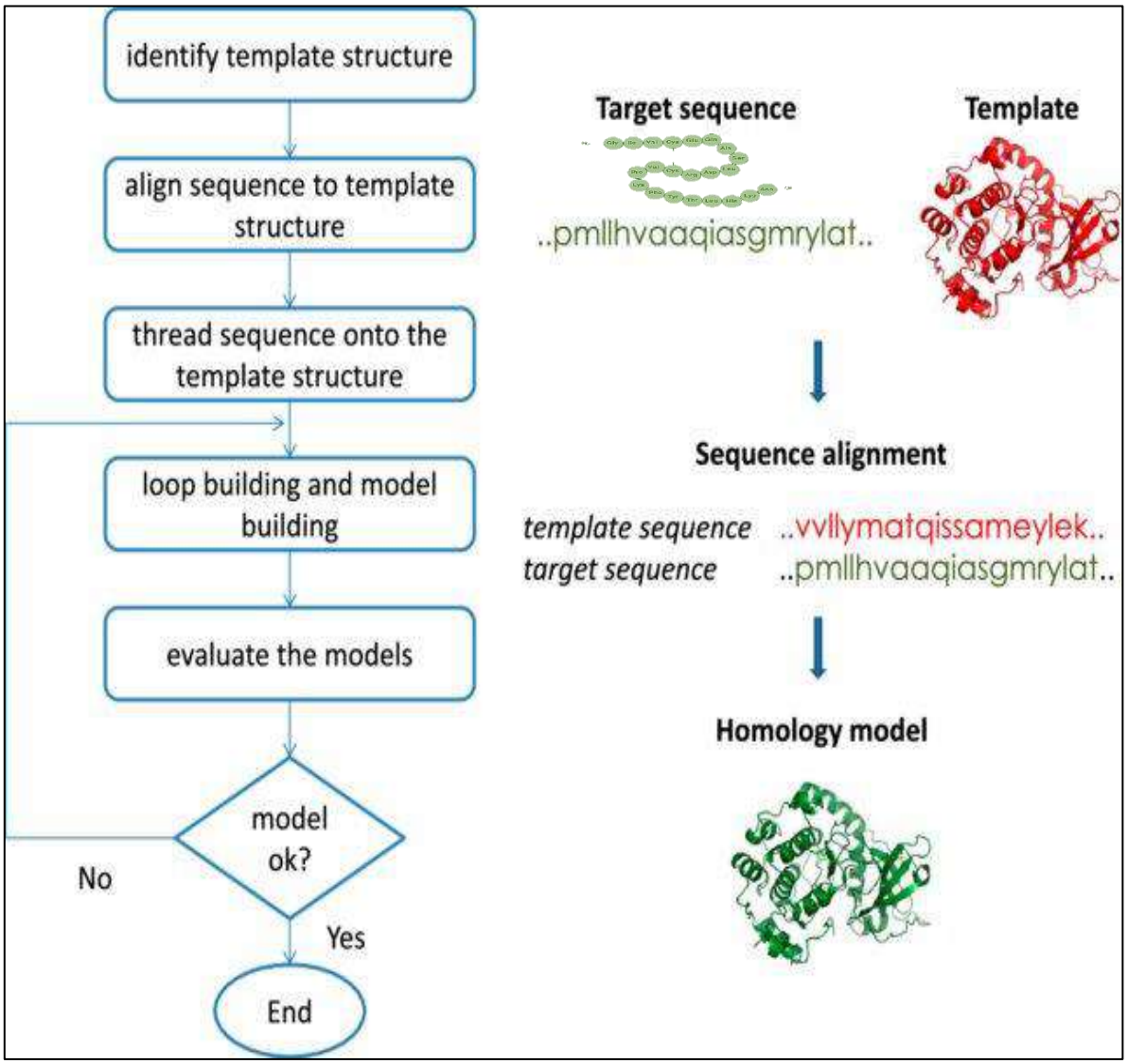



**Figure 31**. Steps in homology model building process**. (Sliwoski G et al., 2013)**

### IX.2. Fold recognition or "Threading"

This method is used to model those proteins which have the same fold as proteins of known structures, to match sequences without known structures with the protein folds. (Structural template detected from fitness to fold (threading)).

There are three main steps in this method: **(Figure 32)**

1) The construction of a structure target database based on templates.

2) Calculation of the quality of each model.

3) Selecting the best model.

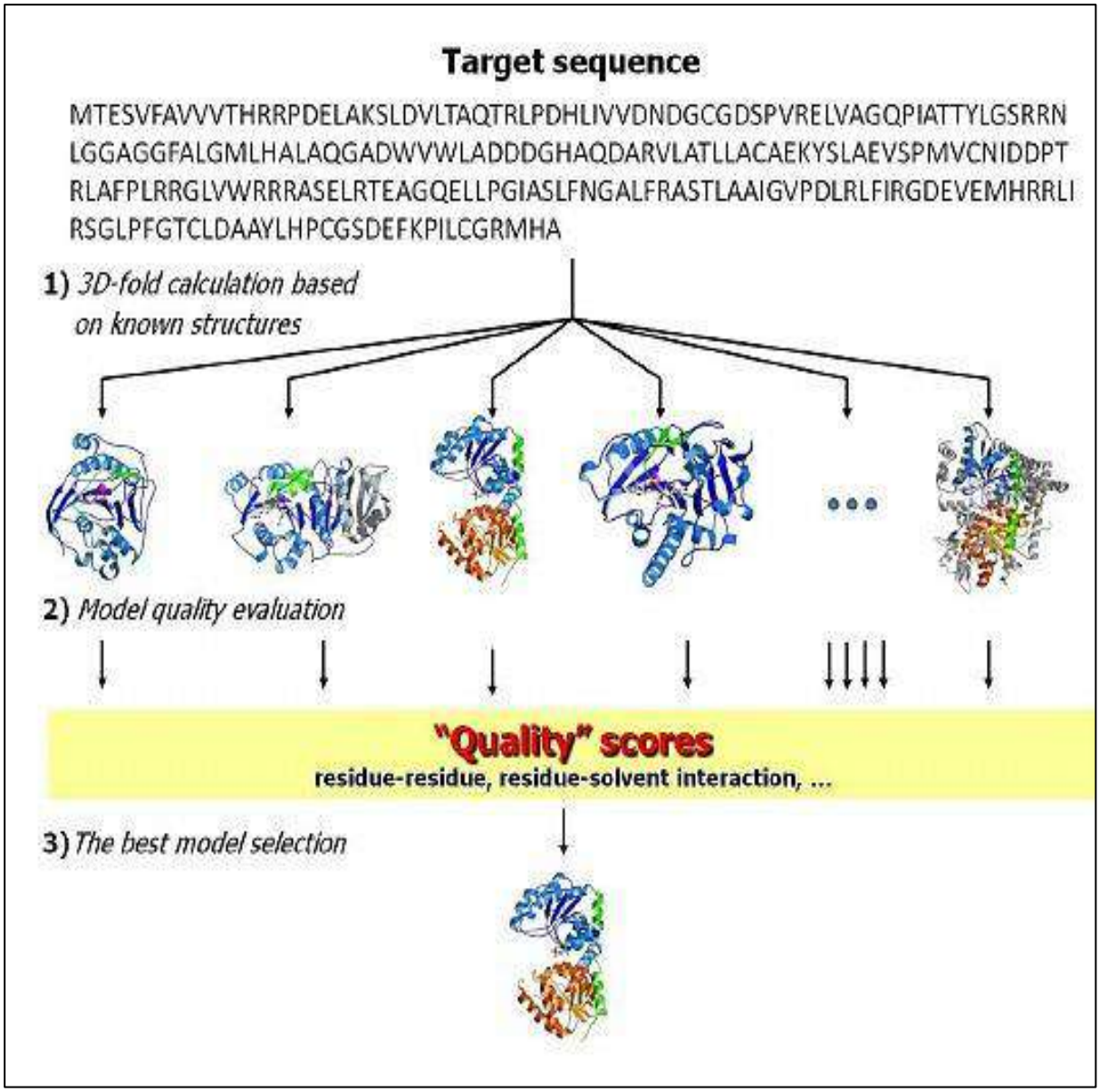



**Figure 32.** The three main steps of threading. **(Kumar S et al., 2012).**

### IX.3. Ab initio prediction

It's the prediction of protein structure without using a homolog (or analog) structure, Ab initio predict protein structure based on physical models, they are indispensable complementary methods to Knowledge-based approach.

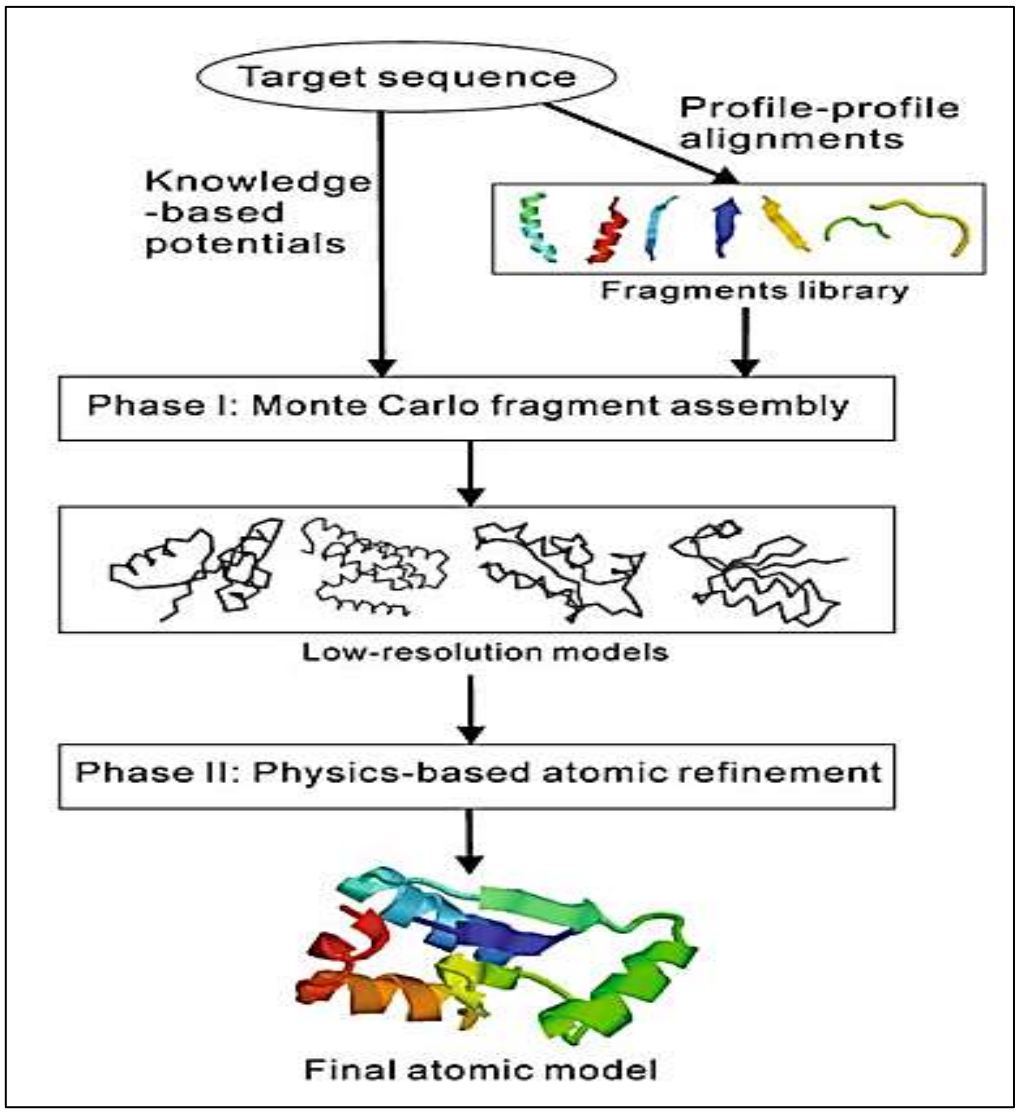



**Figure 33.** The ROSETTA protocol flowchart Fragments are generated from unrelated protein structures in the PDB before being used to assemble full-length models using simulated annealing simulations guided by a knowledge-based force field.

The selected models are refined at the atomic level in the second phase using a physics-based potentia. **(Lee J et al., 2017).**

**X. Why is it so important to understand a protein's three-dimensional structure?**

A protein's biological function is dictated by the arrangement of the atoms in the three-dimensional structure of the protein, and thus it is possible to know how the protein interacts with other proteins by recognizing the arrangement of the catalytic residues of the active site

The presence of the 3-D structure of the protein allows us to understand how it works in a high way and level, such as knowing how to control it or modify it by creating hypotheses, such as making special mutations for this protein to change its function. And there are more features.

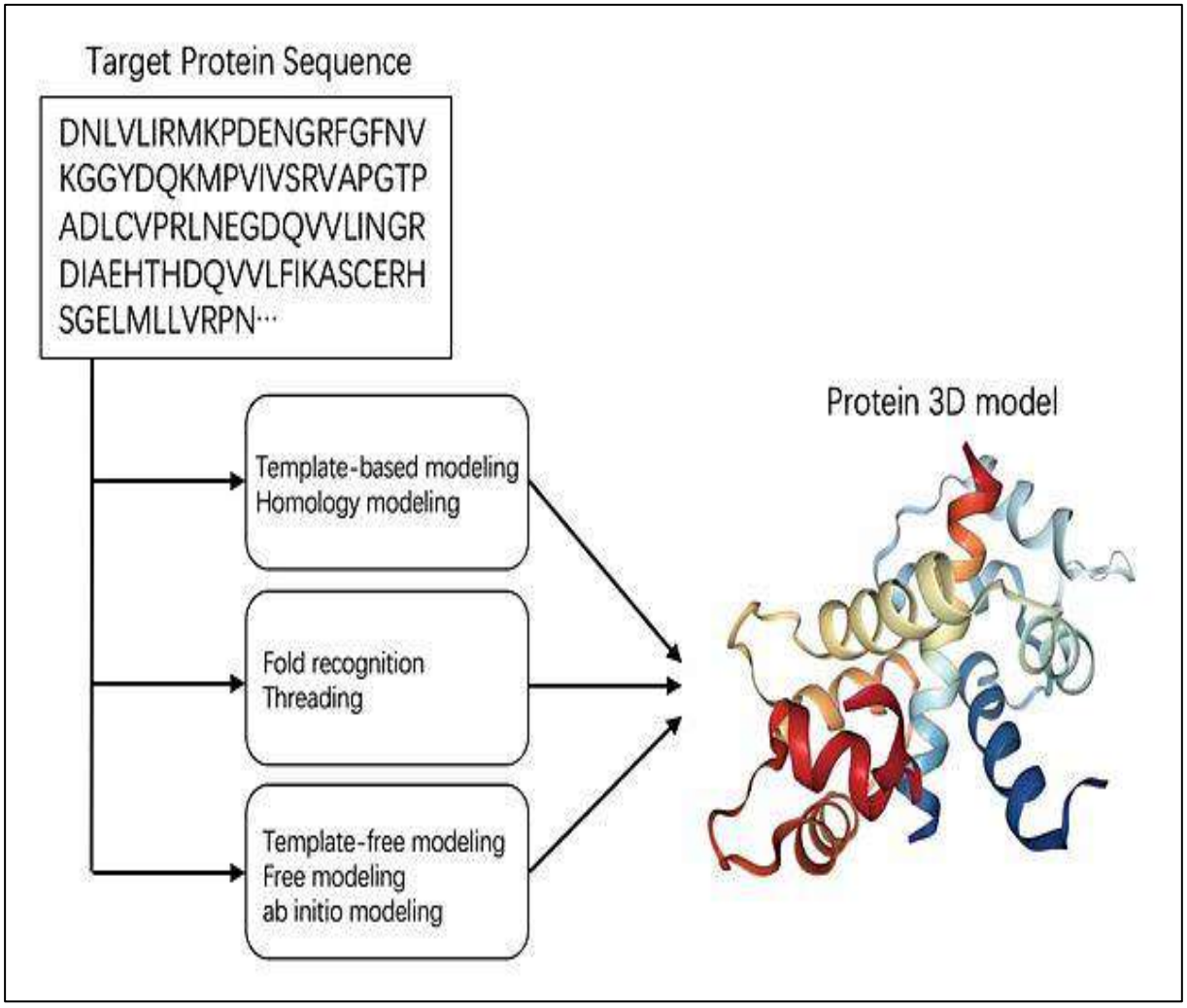



**Figure 34.** The three approaches of protein structure prediction. **(Darnell S, 2020).**

# Chapter II: Material & Methods

**Objectives:**

In this study, the attempt is to apply the Homology Molecular Modeling methodology, known also Comparative Modeling, to predict and create a reliable structural model for a query (target) protein starting from its amino-acids sequence.

For predicting the structural model, this method is based on using the high similarity (homology) of the amino-acids sequence of the target protein with other proteins of known 3D-structure.

**Comparative Modeling Steps:**

**I. Selection of a target protein and amino-acids sequence:**

The protein target selected for this structure prediction project is the human Dihydrofolate Reductase (DHFR) which is an enzyme involved in the synthesis of raw material necessary for cellular proliferation, found in both prokaryotic and eukaryotic cells.

DHFR has a critical role in the cellular regulation of tetrahydrofolate and its derivatives' amounts that are essential for the synthesis of purine and thymidylate in turn are important for cellular division and growth, see Chapter I, section DHFR Function.

Such an important role made DHFR a potential target for structural elucidation through a number of methods including structure prediction in studies aiming at rationally designing new effective drugs against cancer disease and virulent microbial infection both pathologies being based on cellular proliferation. This is in addition to better understanding of ligand binding details to DHFR and structure-function relationship fundamentals. The DHFR amino-acids sequence represented with single letter codes is shown displayed below in Figures 2 and 3.

**II. Homology molecular modeling:**

This is a complex of interconnected set of steps are realized in this project in manual to semi-automatic procedure to serve deep understanding of the rational behind the technique for academic (teaching) and research purposes. The modeling steps outlined next are followed in the prediction of 3D-structure starting from the knowledge of amino-acid sequence based on homology and sequence similarity as developed in the method "Auto Protein Homology Modeling" **(Rachedi A.,1994)**. Such methodology has become common and part of most methods of molecular modeling as discussed in reviews in the subject (**Haddad Y et al,.2020).**

The steps are summarized in the following:

**II.1. Identification and extraction of target sequence**

**II.2. Sequence alignment**

**II.3. Structure templates identification**

**II.4. Core structure calculation**

**II.5. Amino-acids side-chains building**

**II.6. Loop regions building**

**II.7. Final model**

## II.1. Identification and extraction of target sequence

Dihydrofolate Reductase sequence target has been selected as explained above and extracted from the Uniprot database (UNIversal PROTein database) in the following procedure:

Usually in this type of projects, the target's amino-acid sequence is obtained by experimental methods e.g. by extraction, purification and sequencing. However, due to nature of this project being purely bioinformatics based, a human DHFR sequence bearing the accession code P00374 has been extracted from the UniProt database via the use of the following website address: **https://www.uniprot.org/**

In the "Query" field representing the Protein Knowledgebase (UniProtKB) module of the database, the accession code P00374 has been provided to use for the search, **Figure 01**.

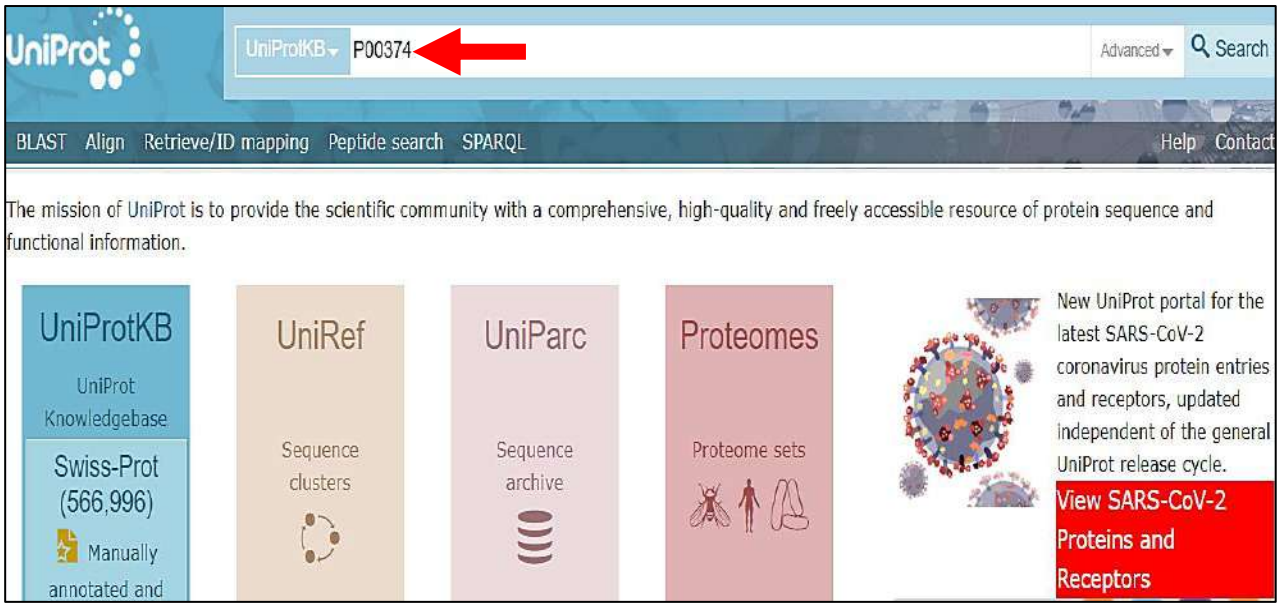



**Figure 01.** Screenshot of the process of obtaining the amino-acids sequence by entering the string for the DHFR in the Query field UniProtKB as indicated by the arrow. (**The UniProt Consortium,. 2021**)

After studying the result page and by scrolling down to the 'Sequence' section, the protein sequence is located as shown in **Figure 02** and the sequence is downloaded in FASTA format, **see Figure 03**. FASTA format is the standard format for representing amino-acids sequences and necessary for use in the subsequent steps.


**Figure 02.** The Open-format of the selected human DHFR amino-acids sequence. The sequence in FASTA format is downloaded by clicking on FASTA the link as pointed at by the red arrow.

```
>sp|P00374|DYR_HUMAN Dihydrofolate reductase OS=Homo sapiens OX=9606 GN=DHFR
PE=1 SV=2
MVGSLNCIVAVSQNMGIGKNGDLPWPPLRNEFRYFQRMTTTSSVEGKQNLVIMGKKTWFS
IPEKNRPLKGRINLVLSRELKEPPQGAHFLSRSLDDALKLTEQPELANKVDMVWIVGGSS
VYKEAMNHPGHLKLFVTRIMQDFESDTFFPEIDLEKYKLLPEYPGVLSDVQEEKGIKYKF
EVYEKND
```

**Figure 03.** The amino-acids sequence of the human DHFR in FASTA format was downloaded from the UniProt database as explained in text above. The first

line after the symbol ">" is a definition line that provides relevant information about the sequence. The remaining lines are the amino-acids sequence represented in the single letter code of amino-acids.

### II.2. Sequence alignment

This second step is used to find proteins of known 3D structures and sequences that are as similar as possible to the DHFR target. The Basic Local Alignment Search Tool (BLAST) is used, with default parameters within the application scope of this project, to compare pairwiely the target sequence with a database of sequences and identify a list of sequences that share high similarity with query sequence above a certain threshold (sequence identity percentage and similarity score value) as reported in the following:

**A.** BLAST tool, see **Figure 04**, is evoked by using the webserver address:

http://blast.ncbi.nlm.nih.gov/Blast.cgi

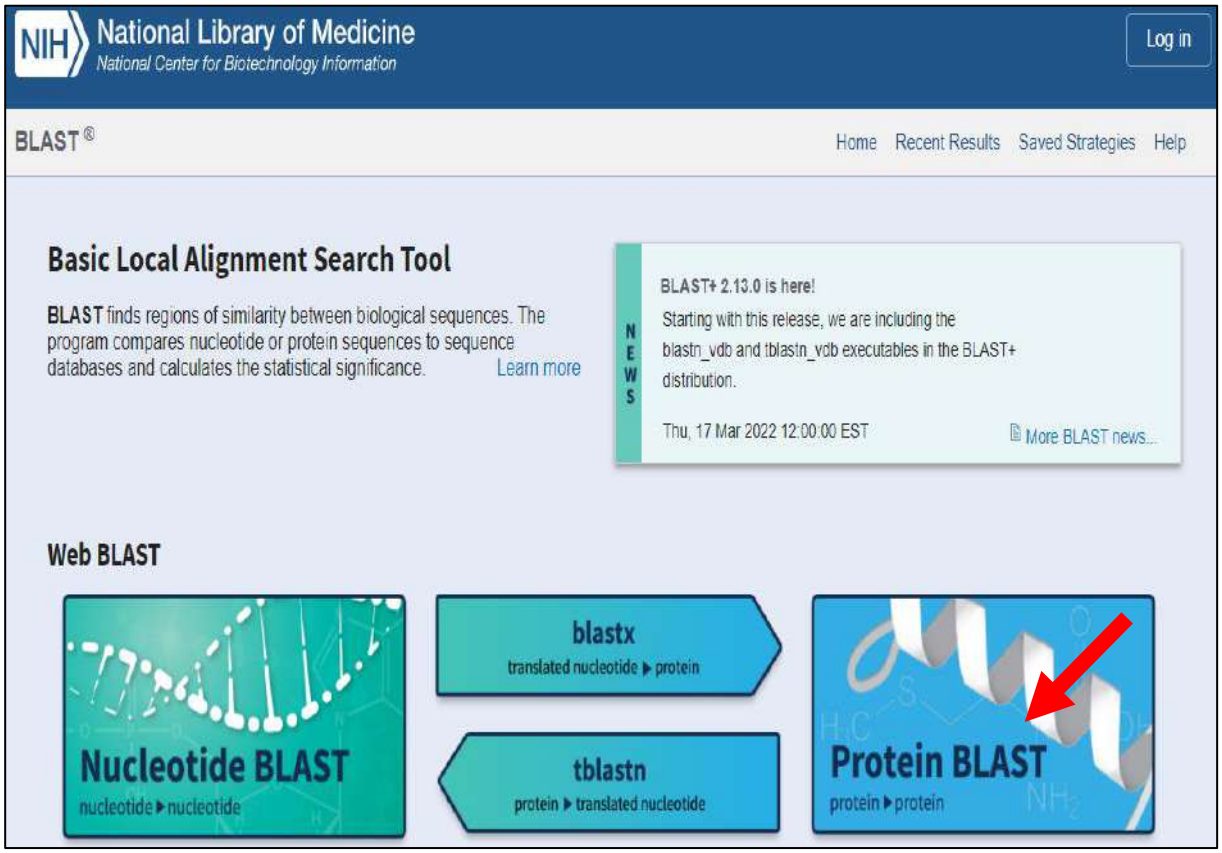



**Figure 04.** Screenshot image showing the home page of BLAST and the tools it carries. The option "protein blast", pointed to with a red arrow, is chosen to restrict the alignment procedure to using the protein database.

The procedure for running BLAST to search the protein sequences database aims at finding similar proteins to the target sequence. Since the prediction is for structural modeling, the PDB based sequences library is selected to be used in the BLAST search as highlighted in **Figure 05**.

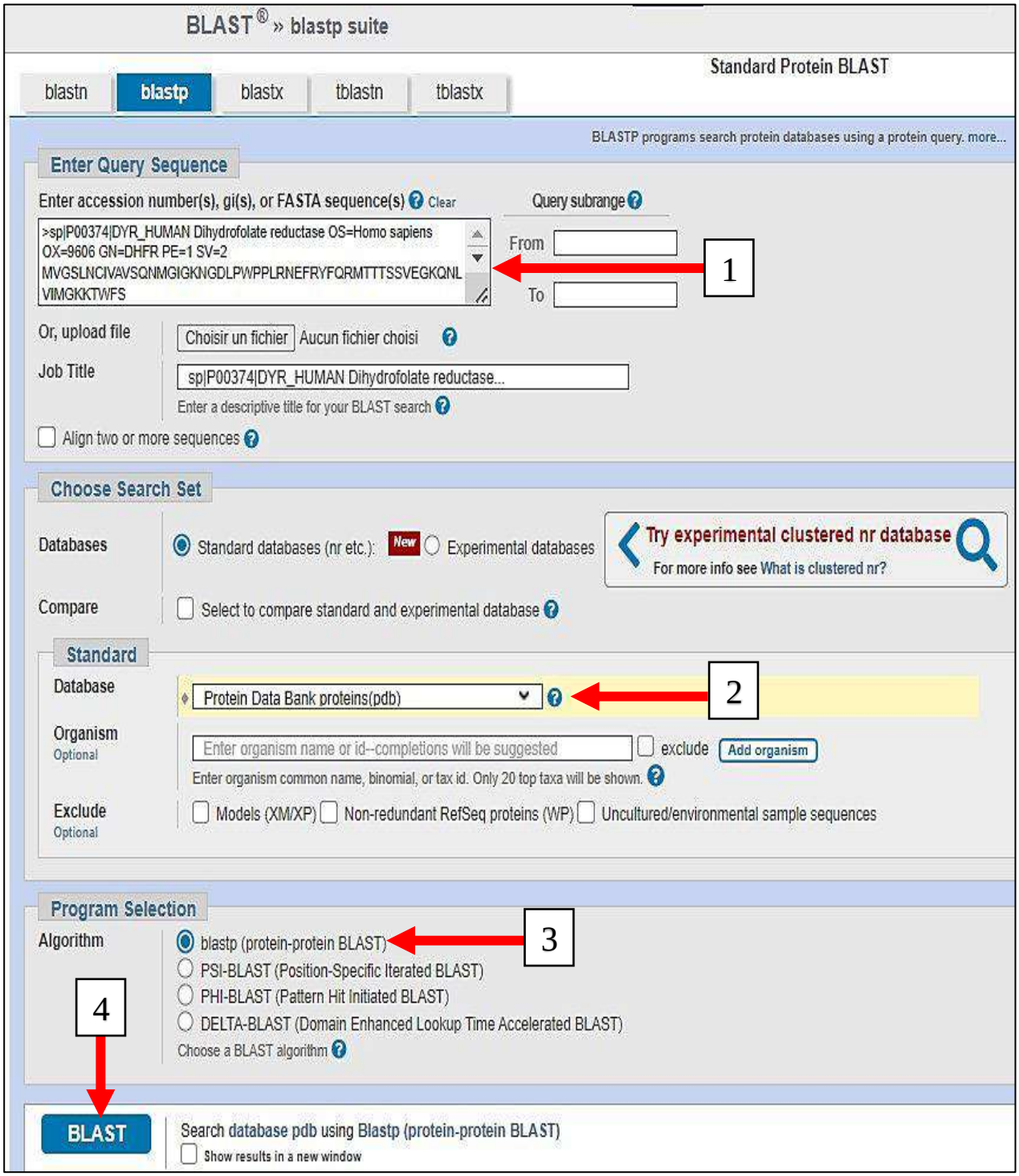


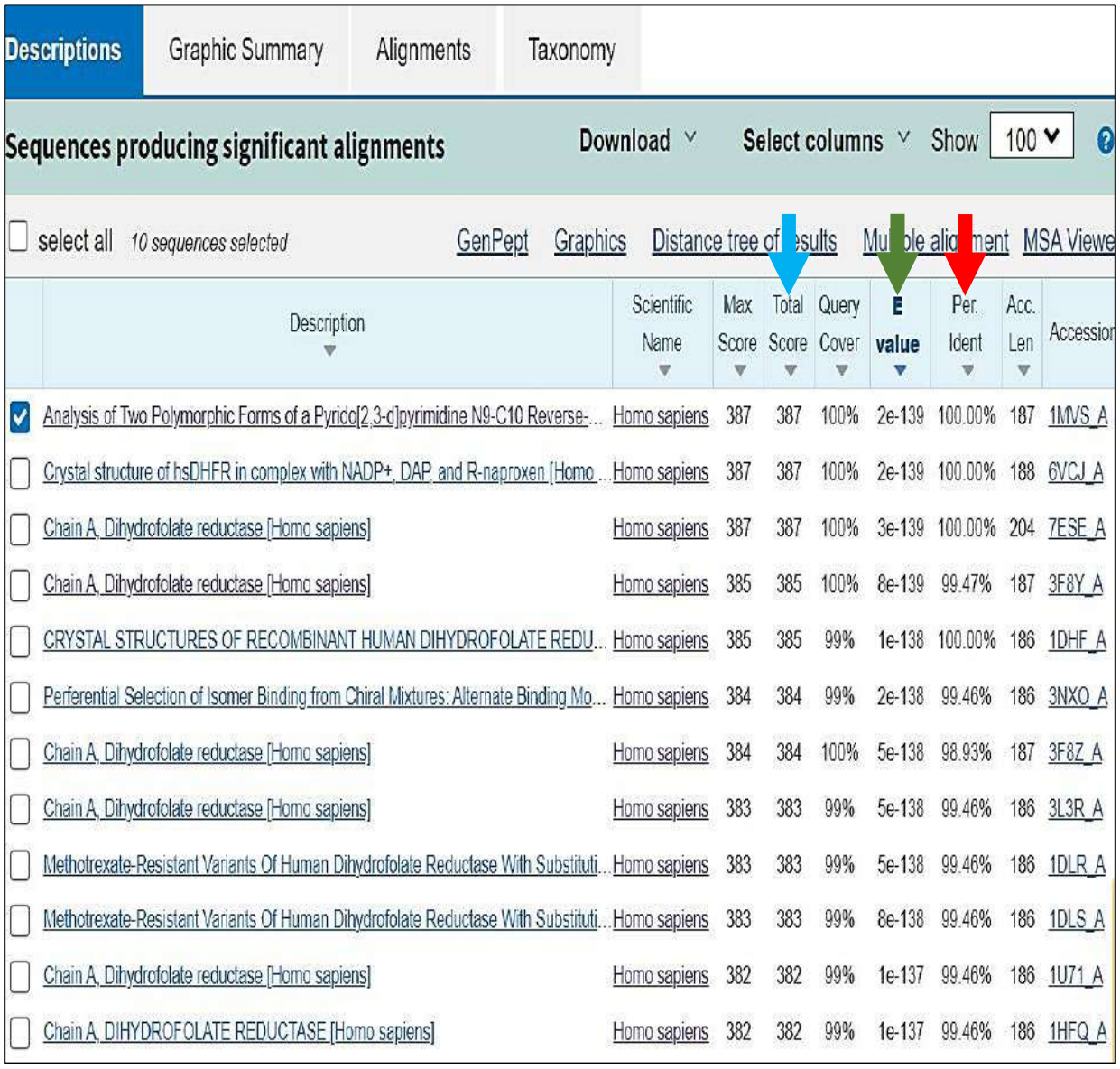

**Figure 05.** The target sequence in FASTA format is pasted in the window field [1] at the top of the page. The "Protein Data Bank proteins (PDB)" is selected a the database to search against [2]. Since the alignment is between protein

sequences, the algorithm **blastp (protein-protein BLAST)** is selected [3] . Finally the **BLAST button** is pressed to start the search.

BLAST server, after a few moments, produced the alignment and gave a list of matching sequences ordered primarily according to their "**sequence identity**" percentage with the target DHFR sequence **Figure 06**. The "**sequence identity**" percentage, pointed at by the **red arrow** in the figure, represents the estimation of how many identical amino-acids are shared between the aligned sequence where the larger the value mean the close the sequences to each other are.

Other alignment evaluation values such as:

- "**Score**", pointed at by the **blue arrow**. Represent the alignment score calculated based on rates of amino-acids substitutions the total value of which should be bigger than zero for related sequence. The larger the score the value the better the alignment is.
- "**Expected**" value (**E**), pointed at by the **green arrow**. Represent the expectation value of an alignment to happen by chance when searching a database of a certain size of sequences. The **E** value should be less than to closer to zero for good hits of sequence similarity.



Descriptions | Graphic Summary | Alignments | Taxonomy

Sequences producing significant alignments — Download — Select columns — Show 100

select all 10 sequences selected — GenPept | Graphics | Distance tree of results | Multiple alignment | MSA Viewe

| Description | Scientific Name | Max Score | Total Score | Query Cover | E value | Per. Ident | Acc. Len | Accession |
|---|---|---|---|---|---|---|---|---|
| Analysis of Two Polymorphic Forms of a Pyrido[2,3-d]pyrimidine N9-C10 Reverse-... | Homo sapiens | 387 | 387 | 100% | 2e-139 | 100.00% | 187 | 1MVS_A |
| Crystal structure of hsDHFR in complex with NADP+, DAP, and R-naproxen [Homo ... | Homo sapiens | 387 | 387 | 100% | 2e-139 | 100.00% | 188 | 6VCJ_A |
| Chain A, Dihydrofolate reductase [Homo sapiens] | Homo sapiens | 387 | 387 | 100% | 3e-139 | 100.00% | 204 | 7ESE_A |
| Chain A, Dihydrofolate reductase [Homo sapiens] | Homo sapiens | 385 | 385 | 100% | 8e-139 | 99.47% | 187 | 3F8Y_A |
| CRYSTAL STRUCTURES OF RECOMBINANT HUMAN DIHYDROFOLATE REDU... | Homo sapiens | 385 | 385 | 99% | 1e-138 | 100.00% | 186 | 1DHF_A |
| Perferential Selection of Isomer Binding from Chiral Mixtures: Alternate Binding Mo... | Homo sapiens | 384 | 384 | 99% | 2e-138 | 99.46% | 186 | 3NXO_A |
| Chain A, Dihydrofolate reductase [Homo sapiens] | Homo sapiens | 384 | 384 | 100% | 5e-138 | 98.93% | 187 | 3F8Z_A |
| Chain A, Dihydrofolate reductase [Homo sapiens] | Homo sapiens | 383 | 383 | 99% | 5e-138 | 99.46% | 186 | 3L3R_A |
| Methotrexate-Resistant Variants Of Human Dihydrofolate Reductase With Substituti... | Homo sapiens | 383 | 383 | 99% | 5e-138 | 99.46% | 186 | 1DLR_A |
| Methotrexate-Resistant Variants Of Human Dihydrofolate Reductase With Substituti... | Homo sapiens | 383 | 383 | 99% | 8e-138 | 99.46% | 186 | 1DLS_A |
| Chain A, Dihydrofolate reductase [Homo sapiens] | Homo sapiens | 382 | 382 | 99% | 1e-137 | 99.46% | 186 | 1U71_A |
| Chain A, DIHYDROFOLATE REDUCTASE [Homo sapiens] | Homo sapiens | 382 | 382 | 99% | 1e-137 | 99.46% | 186 | 1HFQ_A |

**Figure 06.** Depiction of top part from BLAST output sequence alignment run involving alignment search for the query sequence against the library of PDB sequences. Arrows in red, green and blue point to the evaluation terms on the basis of which related sequences of DHFR from different species were selected.

**B.** BLAST tool outputs a large list of aligned sequence hits from which nine (9) sequences were selected for the study, see **Table 01**, based on three characteristics:

- The value of sequence similarity with the target sequence[1].
- Sequences being from different species; Mammals (Homo sapiens, Mus musculus), Birds (Gallus gallus), Yeast *(Candida albicans)*, and bacteria *(Escherichia coli, Bacillus anthracis, Moritella profunda)*[2].
- All of the hit sequences have known 3D-structures[3].

**[1]** The sequence similarity value on which basis the selection is quite high, not less than 50% overall similarity (closely related amino-acids conservation) and above 30% identities (identical amino-acids conservation).

**[2]** Sequence diversity from the different species is a strategy followed in the study to verify if the structural fold of the DHFR is preserved over the very long evolutionary trajectory between the selected species. This would give high confidence in the final model of the DHFR generated by this procedure.

[3] The 3D-structures representing the sequences that are well sequence aligned with the target DHFR sequence are chosen as starting point for structural prediction for the target sequence.

**Table 01.** Selected nine (9) DHFR sequences from the BLAST sequence alignment shown here with their entry codes (first column) representing their access codes in the PDB databank each with the title of the particular study that generated the 3D-structure.

| PDB ID | Title |
|---|---|
| 3K45_A | Alternate Binding Modes Observed for the E- and Z-isomers of 2,4-Diaminofuro[2,3d]pyrimidines as Ternary Complexes with NADPH and Mouse Dihydrofolate Reductase [Mus musculus] |
| 1U70_A | Chain A, Dihydrofolate reductase [Mus musculus] |
| 1DR1_A | 2.2 ANGSTROMS CRYSTAL STRUCTURE OF CHICKEN LIVER DIHYDROFOLATE REDUCTASE COMPLEXED WITH NADP+ AND BIOPTERIN [Gallus gallus] |
| 4H95_A | Candida albicans dihydrofolate reductase complexed with NADPH and 6-ethyl-5-{3-[3-methoxy-5-(pyridin-4-yl)phenyl]but-1-yn-1-yl}pyrimidine-2,4-diamine (UCP1006) *[Candida albicans]* |
| 4GH8_A | Crystal structure of a 'humanized' E. coli dihydrofolate reductase *[Escherichia coli* K-12] |
| 1ZDR_A | DHFR from Bacillus Stearothermophilus *[Geobacillus stearothermophilus]* |
| 3JW3_A | Chain A, Dihydrofolate reductase *[Bacillus anthracis]* |
| 2QK8_A | Crystal structure of the anthrax drug target, *Bacillus anthracis* dihydrofolate reductase [*Bacillus anthracis* str. Sterne] |
| 2ZZA_A | *Moritella profunda* Dihydrofolate reductase complex with NADP+ and Folate *[Moritella profunda]* |

### II.3. Structure templates identification

Detailed of the sequence alignment, evaluation and the amino-acids sequences of the 9 DHFR sequences of the known structures are shown in the **Appendix Part II** of which three tables are shown below:

**Table 02.** Hit sequence with PDB structure id: 3K45

**Sequence ID: 3K45_A Length: 186 Number of Matches: 1**
**Range 1: 1 to 186**

| Score | Expect | Method | Identities | Positives | Gaps |
|---|---|---|---|---|---|
| 350 bits(898) | 8e-125 | Compositional matrix adjust. | 166/186(89%) | 179/186(96%) | 0/186(0%) |

```
Query  2    VGSLNCIVAVSQNMGIGKNGDLPWPPLRNEFRYFQRMTTTSSVEGKQNLVIMGKKTWFSI  61
            V  LNCIVAVSQNMGIGKNGDLPWPPLRNEF+YFQRMTTTSSVEGKQNLVIMG+KTWFSI
Sbjct  1    VRPLNCIVAVSQNMGIGKNGDLPWPPLRNEFKYFQRMTTTSSVEGKQNLVIMGRKTWFSI  60

Query  62   PEKNRPLKGRINLVLSRELKEPPQGAHFLSRSLDDALKLTEQPELANKVDMVWIVGGSSV  121
            PEKNRPLK RIN+VLSRELKEPP+GAHFL++SLDDAL+L EQP+LA+KVDMVWIVGGSSV
Sbjct  61   PEKNRPLKDRINIVLSRELKEPPRGAHFLAKSLDDALRLIEQPDLASKVDMVWIVGGSSV  120

Query  122  YKEAMNHPGHLKLFVTRIMQDFESDTFFPEIDLEKYKLLPEYPGVLSDVQEEKGIKYKFE  181
            Y+EAMN PGHL+LFVTRIMQ+FESDTFFPEIDL KYKLLPEYPGVLS+VQEEKGIKYKFE
Sbjct  121  YQEAMNQPGHLRLFVTRIMQEFESDTFFPEIDLGKYKLLPEYPGVLSEVQEEKGIKYKFE  180

Query  182  VYEKND  187
            VYEK D
Sbjct  181  VYEKKD  186
```

**Table 03.** Hit sequence with PDB structure id: 1U70

**Sequence ID: 1U70_A Length: 186 Number of Matches: 1**
**Range 1: 1 to 186**

| Score | Expect | Method | Identities | Positives | Gaps |
|---|---|---|---|---|---|
| 349 bits(896) | 2e-124 | Compositional matrix adjust. | 166/186(89%) | 178/186(95%) | 0/186(0%) |

```
Query  2    VGSLNCIVAVSQNMGIGKNGDLPWPPLRNEFRYFQRMTTTSSVEGKQNLVIMGKKTWFSI  61
            V  LNCIVAVSQNMGIGKNGD PWPPLRNEF+YFQRMTTTSSVEGKQNLVIMG+KTWFSI
Sbjct  1    VRPLNCIVAVSQNMGIGKNGDRPWPPLRNEFKYFQRMTTTSSVEGKQNLVIMGRKTWFSI  60

Query  62   PEKNRPLKGRINLVLSRELKEPPQGAHFLSRSLDDALKLTEQPELANKVDMVWIVGGSSV  121
            PEKNRPLK RIN+VLSRELKEPP+GAHFL++SLDDAL+L EQPELA+KVDMVWIVGGSSV
Sbjct  61   PEKNRPLKDRINIVLSRELKEPPRGAHFLAKSLDDALRLIEQPELASKVDMVWIVGGSSV  120

Query  122  YKEAMNHPGHLKLFVTRIMQDFESDTFFPEIDLEKYKLLPEYPGVLSDVQEEKGIKYKFE  181
            Y+EAMN PGHL+LFVTRIMQ+FESDTFFPEIDL KYKLLPEYPGVLS+VQEEKGIKYKFE
Sbjct  121  YQEAMNQPGHLRLFVTRIMQEFESDTFFPEIDLGKYKLLPEYPGVLSEVQEEKGIKYKFE  180

Query  182  VYEKND  187
            VYEK D
Sbjct  181  VYEKKD  186
```

In all of the alignments above, the **Query** sequence represents the target DHFR sequence. the **Sbjct** (subject) represent the sequence that is found to align well with the target sequence. **Identities** value represent the percentage of identical (same) residues between the aligned sequences. **Positives** value represent the percentage of identical + similar (close) residues found between the aligned sequences. **Gaps** value represent the percentage of introduced

gaps to achieve the particular sequence alignment. **Method**, which is Compositional Matrix Adjustment (part of the default parameters of BLAST the details of which are beyond the scope of this project), is the alignment algorithm used by BLAST tool to find the alignment between the sequences. The **Score** and **Expect** values are explained earlier in the section above.

**C.** Multiple alignment confirmation of sequences relatedness:

BLAST also offers the possibility to align multiple sequences (more than two) which is used hear as a mean of confirmation of the well alignment of the sequences selected above. Since the sequences selected in above steps have known 3D-structures, confirmation of a good alignment should show the alignment of the regions of secondary structure in those sequences to the best agreement possible.

It is clearly shown, in **Figure 07**, where most of the amino acids involved in the secondary structure elements align well for each and every sequence.

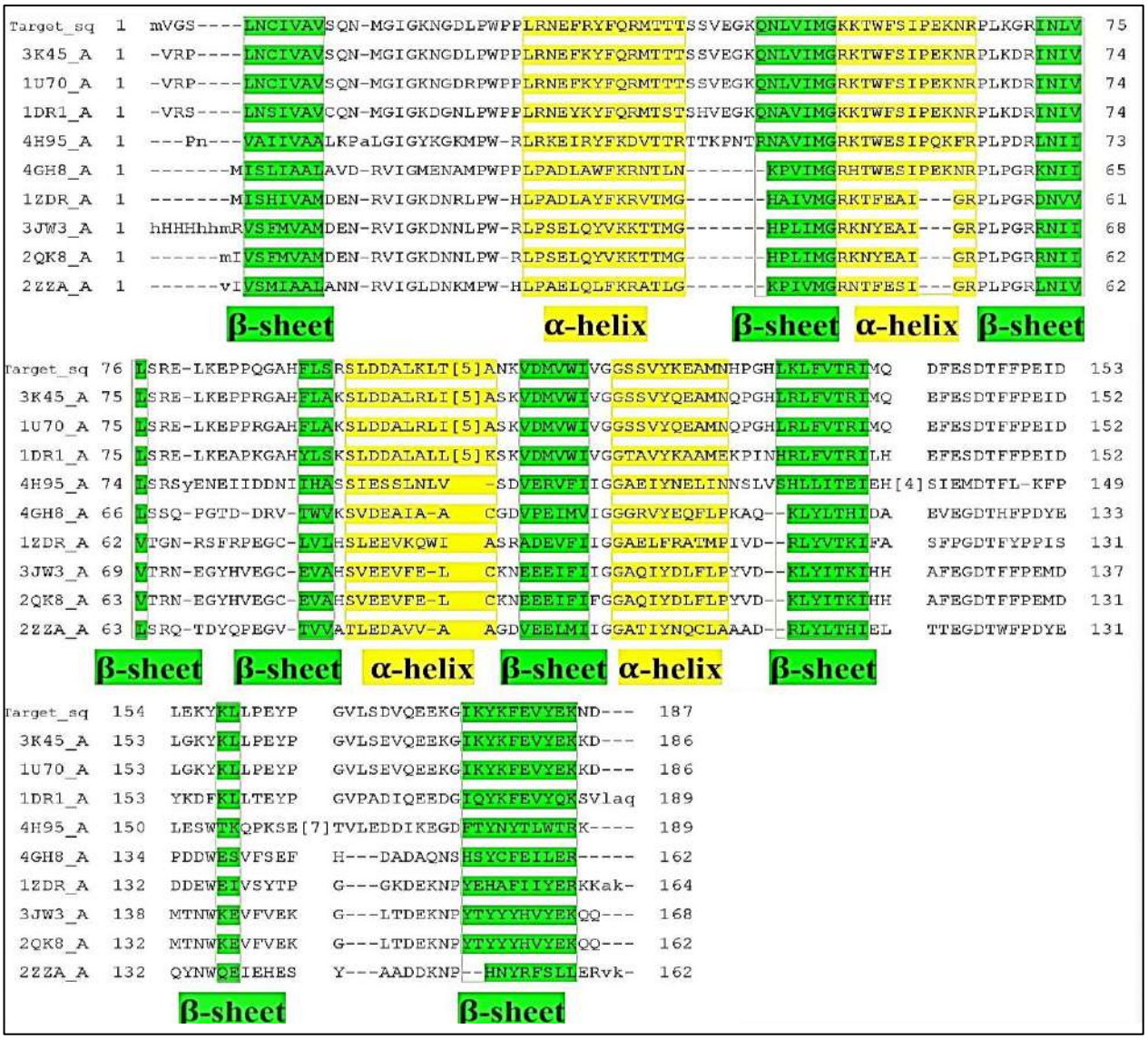



**Figure 07.** Multiple alignment of the selected sequences with the target sequence in top. Secondary structure is highlighted; α-Helices in yellow color and β-strands in green. The multiple alignment shows good agreement of the amino-acids representing secondary structure of the aligned sequences (of the known structures).

For secondary structure definition, the software tool Sequence, Structure & Function or SSFS (https://bioinformaticstools.org/ssfs/) **(Rachedi A, 2011, Goloving A et al., 2005)** has been used. See below the section entitled **Common Core and Secondary Structure Elements.**

### II.4. Common core calculation:

At this stage, the selected sequences are used to find the invariant and common structural features shared between their 3D-structures. This is does through use use of superposition of the protein structures.

One of the public tools for superposition is the 3D-SS (3Dimensional Structural Superposition) server **(K.Sumathi,et al 2005)** reachable at the website **http://cluster.physics.iisc.ernet.in/3dss/** , see **Figure 8**.

The superposition of the structures allows to find out the region of similar structural structure and this labeled here as the **common core** (CC) that is use as the starting point for creating a structural model for the target sequence the subject of this study.

In this study, the 3d-SS STAMP algorithm or Structural Alignment of Multiple Proteins

**(R. B. Russell, G. J. Barton,. 1992)** the cut-off distance of 2 Å (inter-chain) between the Cα atoms of the similar regions is used to qualify as suitable final CC. The reason of such a choice for defining structure similarity in this project is based on the fact that 2 Å is much less than the inter-atomic distance between atoms making in protein structure; e.g. average consecutive Cα -to- Cα distance is 3.8 Å and standard deviation of 0.04 Å **(Chakraborty S, et al 2013)**.


**Figure 08.** The 3D-SS starting page. Usage can click the button Option to proceed.

### II.4.1. Common Core calculation - overall superposition:

Using the 3d-SS tool, the nine (9) selected structures are superposed in an overall fushion. Firstly, the structure with PDB-id "3K45" is selected as the fixed molecule and the 8 other molecules are superposed one after the other on it. The PDB-ids of the other structure are supplied and the number of structures to superpose (8 non-fixed structures in this study case). To launch the procedure by clicking the button **Submit** as described in **Figures 09** and **10**.

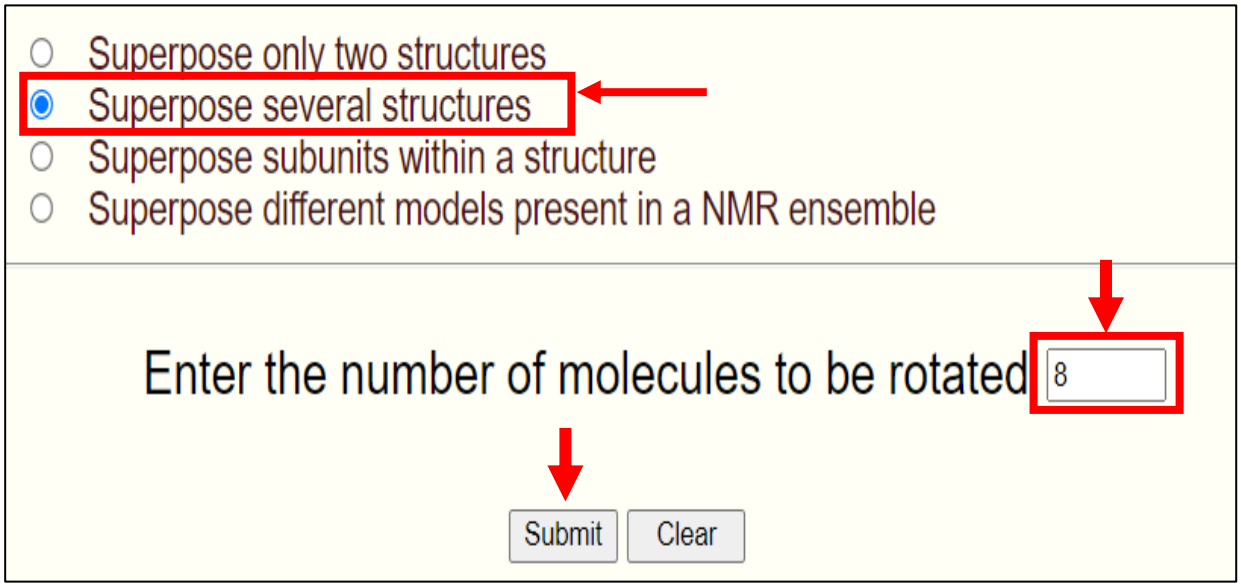



**Figure 09.** 3d-SS option list. The second option for sup as Highlighted.

The PDB-id 3K45 is used as a fixed molecule and the remaining 8 structures are treated as mobile molecules (molecules to be superposed on the fixed molecule). The output of a typical superposition of (1U70, 1DR1, 4H95, 4GH8,1ZDR, 3JW3, 2QK8, 2ZZA).

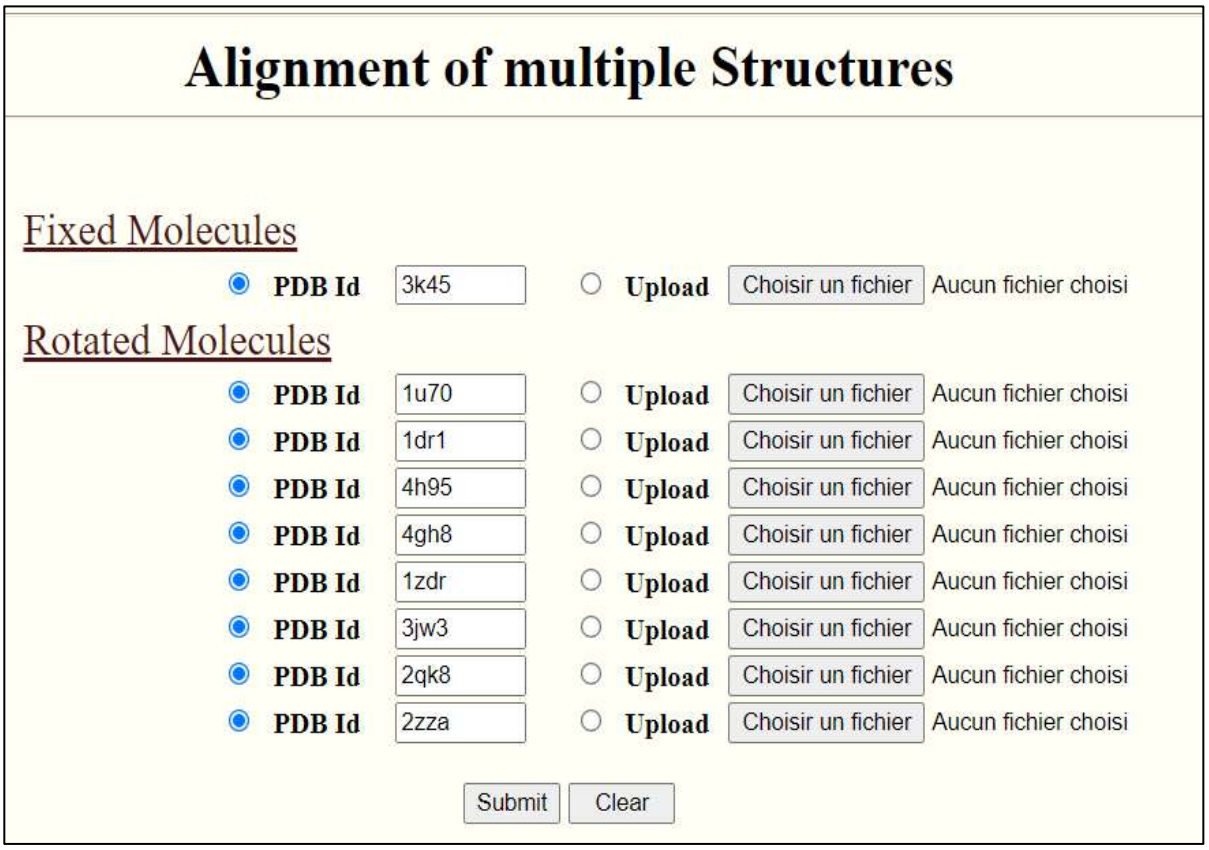



**Figure 10.** List of the fixed molecule and the other rotated molecules selected for superposition.

**Overall-Superposition procedure outcome:**

As shown in **Figure 11**, evaluation scores are shown to qualify the superposition of all the 8 non-fixed structures with the fixed structure 3K45. This table shows the superposition of 9 structures of dihydrofolate reductase .

| PDB ID | Chain ID | Superimposes | Sequence identity (%) | RMSD (Å) |
|---|---|---|---|---|
| 3K45 | A | [Fixed Molecule] | | |
| 1U70 | A | ✓ | 98.92 | 0.537 |
| 1DR1 | A | ✓ | 75.54 | 0.889 |
| 4H95 | A | ✓ | 40.67 | 1.295 |
| 4GH8 | A | ✓ | 39.29 | 1.272 |
| 1ZDR | A | ✓ | 36.64 | 1.118 |
| 3JW3 | A | ✓ | 38.24 | 1.276 |
| 2QK8 | A | ✓ | 37.68 | 1.412 |
| 2ZZA | A | ✓ | 35.56 | 1.241 |

**Figure 11.** superimposition of all selected structures. The standard deviation between the structures (RMSD) which is < 1.5 Å reflect a good quality of structure resemblance as this value is much less than the inter-atomic distance between heavy atoms (all atoms except hydrogen) in each of the involved structures.

The x, y and z coordinates of the superposed structures are then saved into a file and the Molsoft ICM-Browser molecular graphics **(Totrov M.M, Abagyan R.A.,1994)** system is used to inspect visually the quality of the superposition, see **Figure 12**.

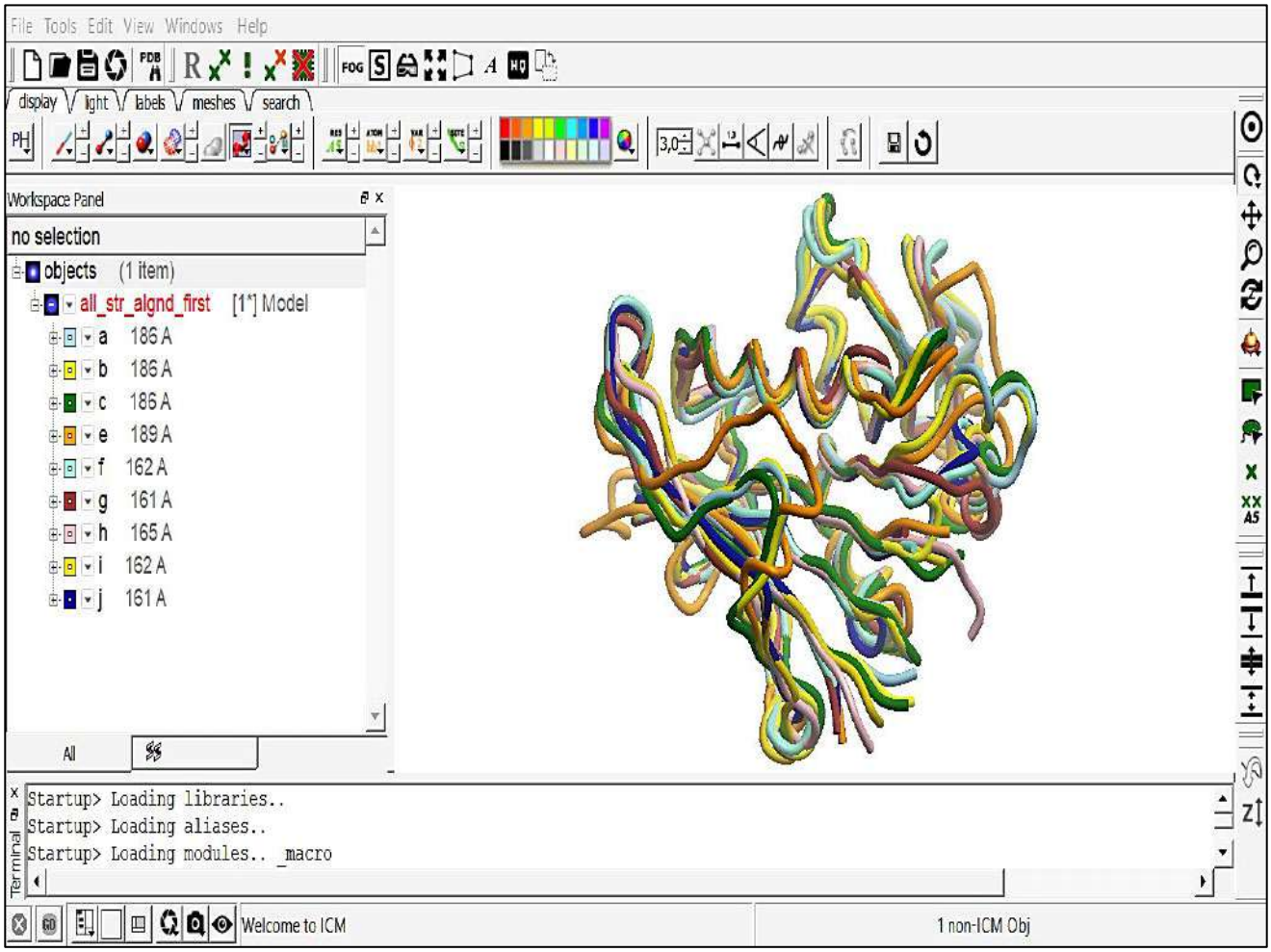



**Figure 12.** The molsoft graphics panel shows the superposition of all the structures in different colors.

For the overall superposition the molecules are chosen for the superposition without any other options which instructs the 3d-SS tool to use its default algorithm to find optimal superposition between the involved structures as shown below in Figure 13.

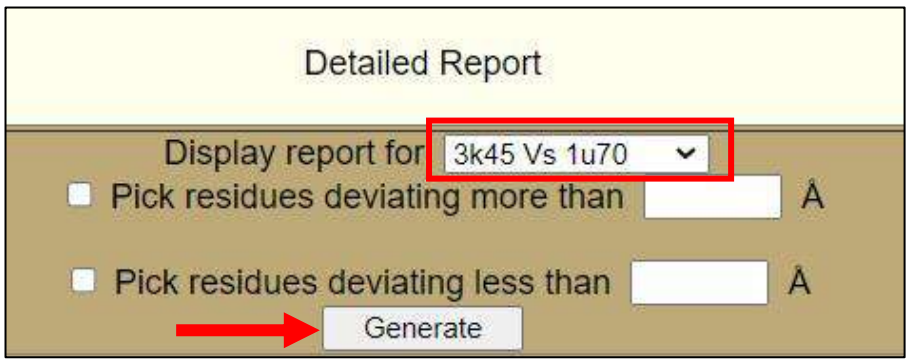



**Figure 13.** 3d-SS option list. The molecules are Highlighted in the field.

The display report for the fixed molecule and the other rotated molecules, with the name and the number of the residues, and the deviation in (Å) are shown in tables refer to **Appendix part III** overall superposition part.

### II.4.2. Identification of core regions:

At this point, the regions have been designated based on the cut-off distance less than 2 angstrom (i.e <= 2 Å). The 3k45 was chosen like a fixed molecule and the 8 other molecules are superposed one after the other on it, the results shown in tables: in **Appendix part III** core identification part.

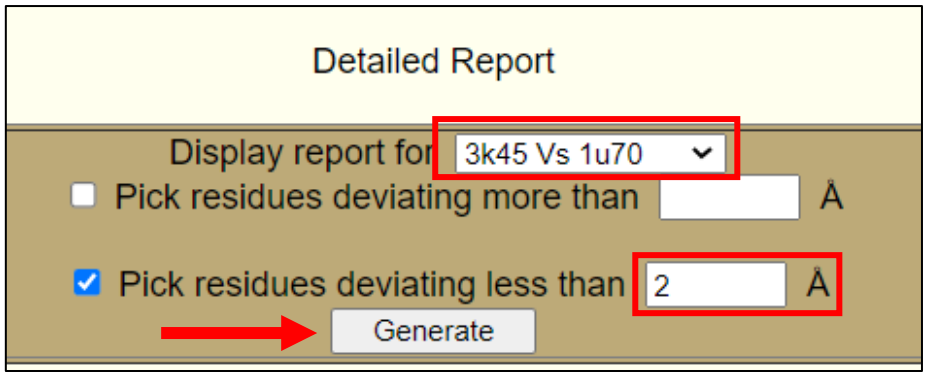



**Figure 14.** 3d-SS option list. The molecules are Highlighted in the field with picking residues deviating less than 2 **Å** as Highlighted.

"RMSD": Root Mean Score Distance (which is a statistical value reflect the quality of data fitting. In this case, RMSD value less than 2 **Å**, means good structural similarity).

- **Structural Alignment details between 3K45 and 1U70**

  In the following table, overall quality of superposition between the two structures, i.e. the RMSD value of 0.537 Å (<2 Å) indicates a very good quality of the fold configuration similarity between the two structures subject to this study, see Table 04.

**Table 04.** Structural Alignment details between 3K45 and 1U70

| Fixed Molecule | Rotated Molecule | RMSD (Å) | Alignment Length |
|---|---|---|---|
| **3K45** | **1U70** | **0.537** | **186** |

Details of the quality of superposition between these two structures (3K45 and 1U70) as shown in Table 09, refer to appendix. all the local deviation between CA atoms of every matching residues in both structures is less than 2 Å indicating quite good fit between the structures.

- **Structural Alignment details between 3K45 and 1DR1**

In the following table, overall quality of superposition between the two structures, i.e. the RMSD value of 0.889 Å (<2 Å) indicates a very good quality of the fold configuration similarity between the two structures subject to this study, see Table 05.

**Table 05.** Structural Alignment details between 3K45 and 1DR1

| Fixed Molecule | Rotated Molecule | RMSD (Å) | Alignment Length |
|---|---|---|---|
| **3K45** | **1DR1** | **0.889** | **186** |

Details of the quality of superposition between these two structures (3K45 and 1DR1) as shown in Table 10, refer to appendix. all the local deviation between Cα atoms of every matching residues in both structures is less than 2 Å indicating quite good fit between the structures.

- **Structural Alignment details between 3K45 and 4H95**

In the following table, overall quality of superposition between the two structures, i.e. the RMSD value of 1.295 Å (<2 Å) indicates a very good quality of the fold configuration similarity between the two structures subject to this study, see Table 06 .

**Table 06**. Structural Alignment details between 3K45 and 4H95.

| Fixed Molecule | Rotated Molecule | RMSD (Å) | Alignment Length |
|---|---|---|---|
| **3K45** | **4H95** | **1.295** | **198** |

Details of the quality of superposition between these two structures (3K45 and 4H95) as shown in Table 11, refer to appendix. all the local deviation between CA atoms of every matching residues in both structures is less than 2 Å indicating quite good fit between the structures.

- **Structural Alignment details between 3K45 and 4GH8**

In the following table, overall quality of superposition between the two structures, i.e. the RMSD value of 1.262 Å (<2 Å) indicates a very good quality of the fold configuration similarity between the two structures subject to this study, see Table 07.

**Table 07.** Structural Alignment details between 3K45 and 4GH8.

| Fixed Molecule | Rotated Molecule | RMSD (Å) | Alignment Length |
|---|---|---|---|
| **3K45** | **4GH8** | **1.262** | **187** |

Details of the quality of superposition between these two structures (3K45 and 4GH8) as shown in Table 12, refer to appendix. all the local deviation between CA atoms of every matching residues in both structures is less than 2 Å indicating quite good fit between the structures.

- **Structural Alignment details between 3K45 and 1ZDR**

In the following table, overall quality of superposition between the two structures, i.e. the RMSD value of 1.118 Å (<2 Å) indicates a very good quality of the fold configuration similarity between the two structures subject to this study, see Table 08.

**Table 08.** Structural Alignment details between 3K45 and 1ZDR.

| Fixed Molecule | Rotated Molecule | RMSD (Å) | Alignment Length |
|---|---|---|---|
| **3K45** | **1ZDR** | **1.118** | **185** |

Details of the quality of superposition between these two structures (3K45 and 1ZDR) as shown in Table 13, refer to appendix. all the local deviation between CA atoms of every matching residues in both structures is less than 2 Å indicating quite good fit between the structures.

- **Structural Alignment details between 3K45 and 3JW3**

In the following table, overall quality of superposition between the two structures, i.e. the RMSD value of 1.276 Å (<2 Å) indicates a very good quality of the fold configuration similarity between the two structures subject to this study, see Table 09.

**Table 09.** Structural Alignment details between 3K45 and 3JW3.

| Fixed Molecule | Rotated Molecule | RMSD (Å) | Alignment Length |
|---|---|---|---|
| **3K45** | **3JW3** | **1.276** | **191** |

Details of the quality of superposition between these two structures (3K45 and 3JW3) as shown in Table 14, refer to appendix. all the local deviation between CA atoms of every matching residues in both structures is less than 2 Å indicating quite good fit between the structures.

- **Structural Alignment details between 3K45 and 2QK8**

In the following table, overall quality of superposition between the two structures, i.e. the RMSD value of 1.412 Å (<2 Å) indicates a very good quality of the fold configuration similarity between the two structures subject to this study, see Table 10.

**Table 10.** Structural Alignment details between 3K45 and 2QK8.

| Fixed Molecule | Rotated Molecule | RMSD (Å) | Alignment Length |
|---|---|---|---|
| **3K45** | **2QK8** | **1.412** | **189** |

Details of the quality of superposition between these two structures (3K45 and 2QK8) as shown in Table 15, refer to appendix. all the local deviation between CA atoms of every matching residues in both structures is less than 2 Å indicating quite good fit between the structures.

- **Structural Alignment details between 3K45 and 2ZZA**

In the following table, overall quality of superposition between the two structures, i.e. the RMSD value of 1.241 Å (<1 Å) indicates a very good quality of the fold configuration similarity between the two structures subject to this study, see Table 11.

**Table 11.** Structural Alignment details between 3K45 and 2ZZA.

| Fixed Molecule | Rotated Molecule | RMSD (Å) | Alignment Length |
|---|---|---|---|
| **3K45** | **2ZZA** | **1.241** | **189** |

Details of the quality of superposition between these two structures (3K45 and 2ZZA) as shown in Table 16, refer to appendix. all the local deviation between CA atoms of every matching residues in both structures is less than 2 Å indicating quite good fit between the structures.

The details provided in the tables allowed for defining the the shared common structure regions (call here Common Core or CC) by keeping all regions where intra-chain Cα- Cα distance is <=

2 Å and discarding the rest. The CC is composed of 10 regions presenved in structure throughout all of the 9 initial structures, see Figure 15 and Table 12.

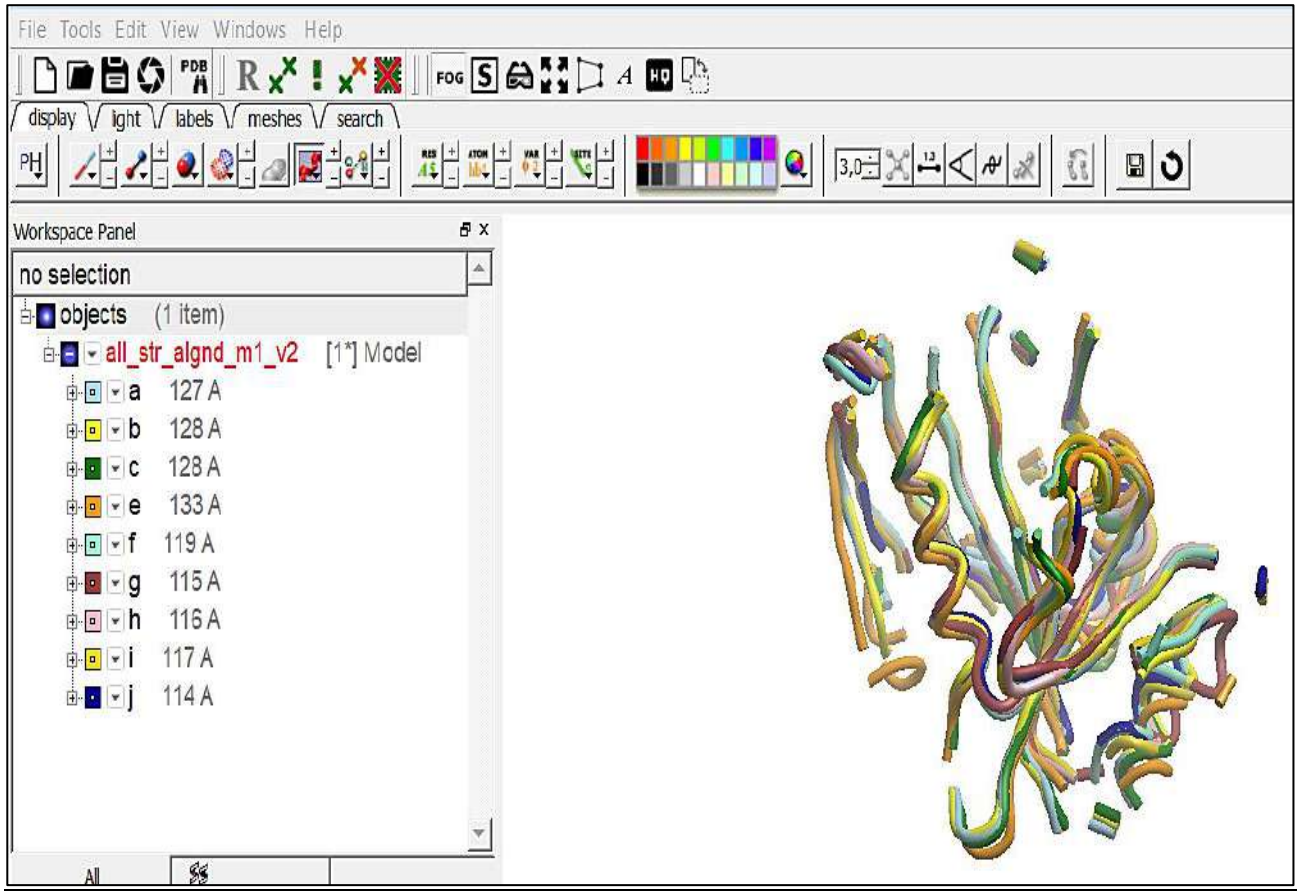



**Figure 15.** The molsoft graphics panel shows the superposition of all the structures in different colors after keepeing on the Common Core made of 10 regions.

**Table 12.** The 10 regions/segments common core in each structure with represented by their PDB ids.

| PDBid / Region | 3K45 | 1U70 | 1DR1 | 4H95 | 4GH8 | 1ZDR | 3JW3 | 2QK8 | 2ZZA |
|---|---|---|---|---|---|---|---|---|---|
| **R-N°1** | 4→12 | 4→12 | 4→12 | 6→14 | 2→10 | 2→10 | 3→11 | 3→11 | 3→11 |
| **R-N°2** | 14→22 | 14→22 | 14→22 | 17→25 | 12→20 | 12→20 | 13→21 | 13→21 | 13→21 |
| **R-N°3** | 26→38 | 26→38 | 26→38 | 28→40 | 24→36 | 23→35 | 24→36 | 24→36 | 24→36 |
| **R-N°4** | 48→60 | 48→60 | 48→60 | 50→62 | 39→51 | 38→50 | 39→51 | 39→51 | 39→51 |
| **R-N°5** | 66→77 | 66→77 | 66→77 | 68→79 | 57→68 | 53→64 | 54→65 | 54→65 | 54→65 |
| **R-N°6** | 88→91 | 88→91 | 88→91 | 91→94 | 77→80 | 75→77 | 75→78 | 75→78 | 75→78 |
| **R-N°7** | 112→125 | 112→125 | 112→125 | 109→122 | 95→108 | 93→106 | 93→106 | 93→106 | 93→106 |
| **R-N°8** | 131→138 | 131→138 | 131→138 | 128→135 | 112→119 | 110→117 | 110→117 | 110→117 | 110→117 |
| **R-N°9** | 145→149 | 145→149 | 145→149 | 146→150 | 126→130 | 124→128 | 124→128 | 124→128 | 124→128 |
| **R-N°10** | 175→184 | 175→184 | 175→184 | 182→191 | 153→162 | 151→160 | 151→160 | 151→160 | 151→160 |

### II.4.3. Region guided superposition and Common Core enhancement:

In this step the superposition has been undertaken using more options provided by the 3d-SS in which case, regions for the superposition procedure were introduced.

For every pairwise superposition, the structure 3K45 is also used as a fixed molecule and the other as a rotated molecule. The case below of the 1U70, the superposition guided by the CC regions is shown in Figures16.a to c:

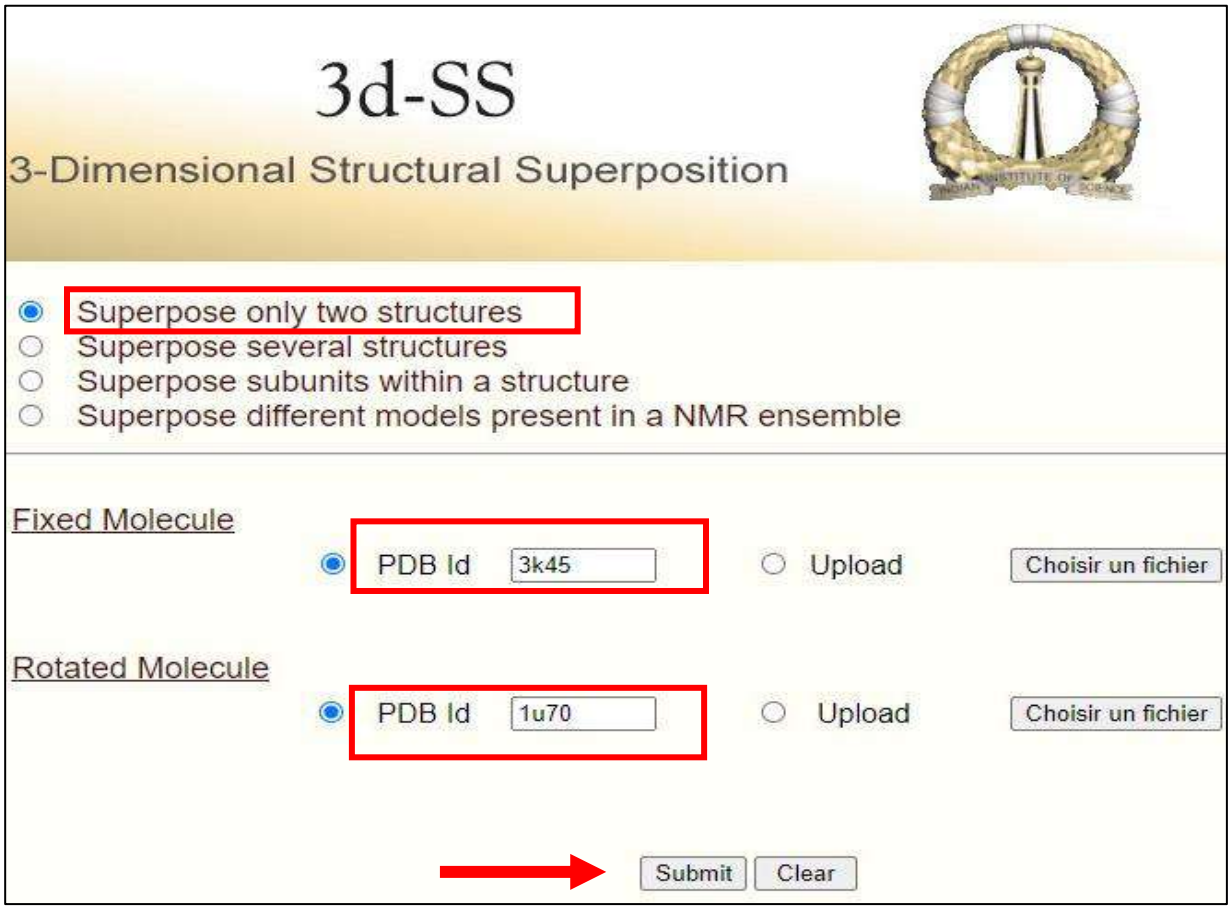



**Figure 16.a.** 3d-SS option list. Highlighted PDB ids are to superposed.

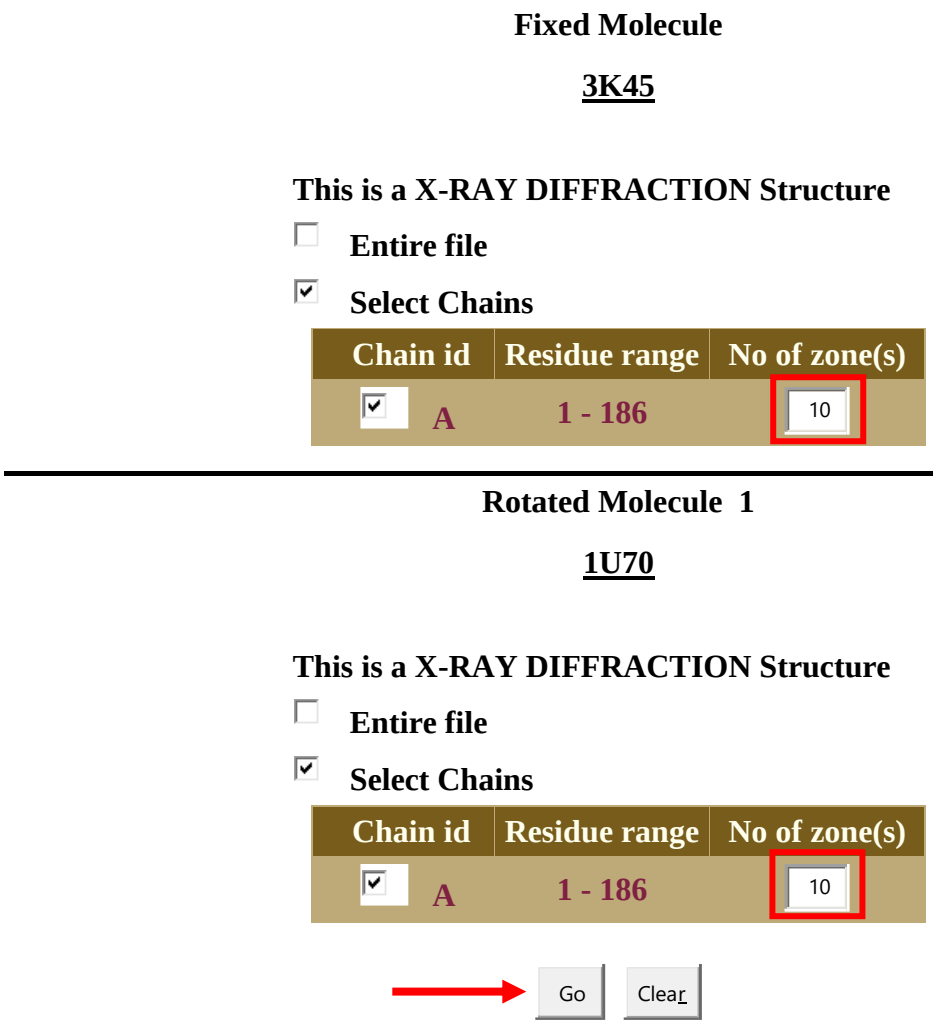



**Figure 16.b.** 3d-SS option list. Highlighted are the number of regions (10) to be used in the superposition.

Then Corresponding 10 regions from each structure regions are supplied in the 3d-SS option '**Zone ranges**', which simply means supply the structures' regions to be used in the superposition, see Figure 16.c.

**Fixed Molecule**

**3K45**

**Enter the Zone ranges**

| Chain id | Zone No. | Starting residue | Ending residue |
|---|---|---|---|
| A 1-186 | 1 | 4 | 12 |
| | 2 | 14 | 22 |
| | 3 | 26 | 38 |
| | 4 | 48 | 60 |
| | 5 | 66 | 77 |
| | 6 | 88 | 91 |
| | 7 | 112 | 122 |
| | 8 | 131 | 138 |
| | 9 | 145 | 149 |
| | 10 | 175 | 184 |

**Rotated Molecule 1**

**1U70**

**Enter the Zone ranges**

| Chain id | Zone No. | Starting residue | Ending residue |
|---|---|---|---|
| A 1-186 | 1 | 4 | 12 |
| | 2 | 14 | 22 |
| | 3 | 26 | 38 |
| | 4 | 48 | 60 |
| | 5 | 66 | 77 |
| | 6 | 88 | 91 |
| | 7 | 112 | 122 |
| | 8 | 131 | 138 |
| | 9 | 145 | 149 |
| | 10 | 175 | 184 |

Go

**Figure 16.c.** 3d-SS option 'Zone ranges'. Starting and Ending residues represent the numbers defining the corresponding regions between the structures to be superposed.

**Regions Guided Superposition outcome:**

Same method and details are followed with superposing the rest of the PDB structures. All the other structures. As shown in **Figure 17**, evaluation scores are shown to qualify the superposition of all the 8 non-fixed structures with the fixed structure 3K45 (the 9th structure). The RMSD values clearly show important an improvement in superposition compared to the overall superposition applied earlier seen in **Figure 11**.

| PDB ID | Chain ID | Superimposes | Sequence identity %) | RMSD (Å) |
|---|---|---|---|---|
| 3K45 | A | [Fixed Molecule] | 100.00 | - |
| 1U70 | A | ✓ | 98.94 | 0.464 |
| 1DR1 | A | ✓ | 84.04 | 0.710 |
| 4H95 | A | ✓ | 48.94 | 0.719 |
| 4GH8 | A | ✓ | 43.62 | 0.726 |
| 1ZDR | A | ✓ | 41.94 | 0.724 |
| 3JW3 | A | ✓ | 42.55 | 0.744 |
| 2QK8 | A | ✓ | 42.55 | 0.832 |
| 2ZZA | A | ✓ | 41.49 | 0.731 |

**Figure 17. S**uperposition of the 9 dihydrofolate reductase structures after the input of the CC region. RMSD values show even better agreement compare to overall superposition (seen in **Figure 11**).

The RSMD values shown in **Figure 17** reflect quite a good structural common core of 97 residues along which the 9 structures match within less than 1 Å from each other in structural space, though many of them belong to different species, as illustrated in **Figure 18**. The details of the inter-atomic distance of all superpositions are shown in Tables found in the **Appendix part III** under the section "**Common Core calculation**".

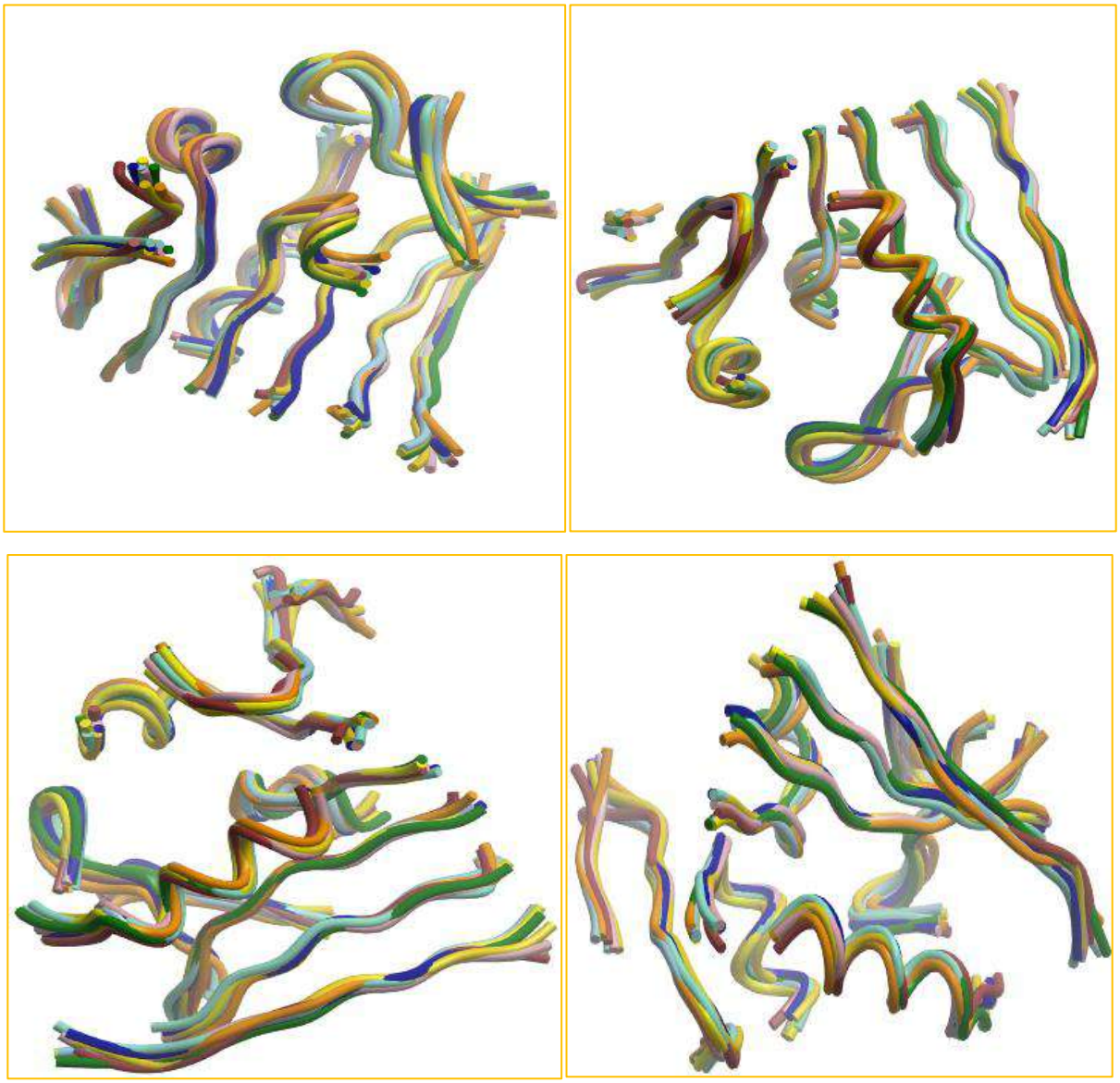


**Figure 18.** The molsoft graphics panel shows the superposition of all the shared common core structures each in a different color. The RMSD values reflect a quite high quality core region shared between the structure.

**Common Core and its Relation to the Secondary Structure Elements:**

The secondary structure elements in all of the 9 structures have been identified using the Sequence, Structure & Function server of SSFS tool ***(Rachedi A, 2011, Gloving A et al., 2005),*** https://www.bioinformaticstools.org/ssfs/ , refer to the SSFS section in the Appendix.

As mentioned above, the Common Core is made of 10 segments many of which encompose secondary structure elements but not all. Some CC regions longer or shorter than secondary structures elements and others are part of loop regions, see Figures 19 and 20 which are included below for vertebrate DHFR (PDB id 1U70) and bacterial one (PDB id 1ZDR).


**Figure 19.** Structure based alignment of the query (target) sequence with the vertebrate sequence (PDB id: 1U70). Alignment shows the correlation of CC regions with the secondary structure; Most of secondary elements, α-helices and β-sheet are contained in the CC regions, however there are CC regions that represent loop regions, like regions 12-22 and 145-149 and some secondary structure, like the α-helix 92-107 and β-strand 157-158 regions are not part of the CC.

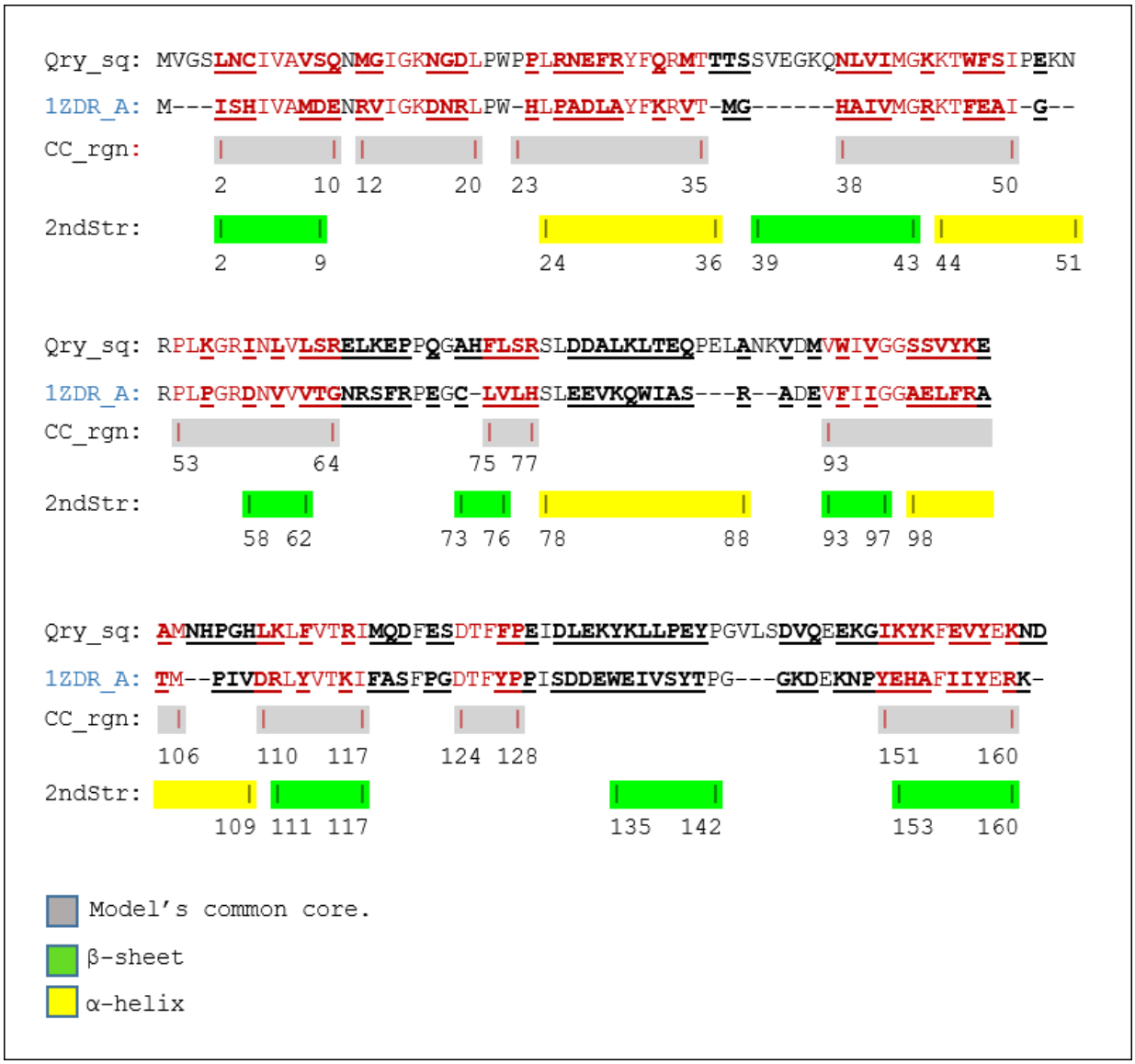



**Figure 20.** Structure based alignment of the query (target) sequence with the bacterial sequence (PDB id: 1ZDR). Alignment shows the correlation of CC regions with the secondary structure; Most of secondary elements, α-helices and β-sheet are contained in the CC regions, however there are CC regions that represent loop regions, like regions 12-20 and 124-128 and some secondary structure, like the α-helix 78-88 and β-strand 157-158 region is not part of the CC.

**Model template and amino-acids sequence of the model:**

In general methodology, any of the core regions from the 9 structures used in the calculation can be use a template for the **model structure** sought for the DHFR query sequence.

In this project two models are generated one based on the core regions from the vertebrate DHFR. the mouse **Mus musculus**, represented by the PDB structure 1U70 and the second based on the core regions from the bacterial DHFR, the bacteria ***Geobacillus stearothermophilus***, represented by the structure 1ZDR, refer to **Table 1**.

As seen in Figures 19 and 20, the template from vertebrate and bacterial DHFR structures is almost same in its main conformational structure, but differ at the level of amino-acids sequence, marked as underlined amino acids.

1. Amino-acids sequence representing the model in the core region, from 1U70, are thus the segments (highlighted in red in the **Figures 19**). The amino-acids sequence (Seq_1) to be modeled is shown next in **Figure 21**.

**LNCIVAVSQ – MGIGKNGDL – PLRNEFRYFQRMT – NLVIMGKKTWFSI – PLKGRINLVLSR - FLSR – VWIVGGSSVYKEAM – LKLFVTRI - DTFFP - IKYKFEVYEK**

**Figure 21**. Amino-acids sequence, Seq_1, for the query sequence representing the core region. The red highlighted are the amino-acids in the query sequence that different from the corresponding sequence of the core region in the template structure 1U70.

Highlighted in red are the amino-acids in the query sequence that different from the corresponding sequence of the core region in the template structure 1U70 and which are:

**4 – 12** , **14 – 22** , **26 – 38** , **48 – 60** , **66 – 77** , **88 – 91** , **112 – 125** , **131 – 138** , **145 – 149** , **175 – 184**

(see Table 12 and also Figures 19)

2. Amino-acids sequence representing the model in the core region, from 1ZDR, are the segments (highlighted in red in the **Figures 20**). The amino-acids sequence (Seq_2) to be modeled is shown next in **Figure 22**.

LNCIVAVSQ – MGIGKNGDL – PLRNEFRYFQRMT – NLVIMGKKTWFSI – PLKGRINLVLSR - FLSR – VWIVGGSSVYKEAM – LKLFVTRI - DTFFP - IKYKFEVYEK

**Figure 22.** Amino-acids sequence, Seq_2, for the query sequence representing the core region. The red highlighted are the amino-acids in the query sequence that different from the corresponding sequence of the core region in the template structure 1ZDR.

Highlighted in red are the amino-acids in the query sequence that different from the correspoding sequence of the core region in the template structure 1u70 and which are:

**2 – 10** , **12 – 20** , **23 – 35** , **38 – 50** , **53 – 64** , **75 – 77** , **93 – 106** , **110 – 117** , **124 – 128** , **151 - 160**

(see Table 12 and also **Figures 20**)

**II.5. Amino-acids side-chains building:**

To build the side-chains of the amino-acids that differ between the query sequence and the sequence of both core regions of 1U70 and 1ZDR outlined in the section above and **Figures 21** & **22**, the molecular modelling tool Auto Protein Homology Modeling or APHM **(Rachedi, A. 1994)** have been implimented. More information about the APHM and access link are descrbed in the next Chapter III.

**Model creation for the Seq_1 and Seq_2:**

The APHM precedure need to be supplied with the PDB id of the template structure like 1U70, the core region data and the core sequence Seq_1 build the model and the different amino-acids as outline in **Figures 23-25**.

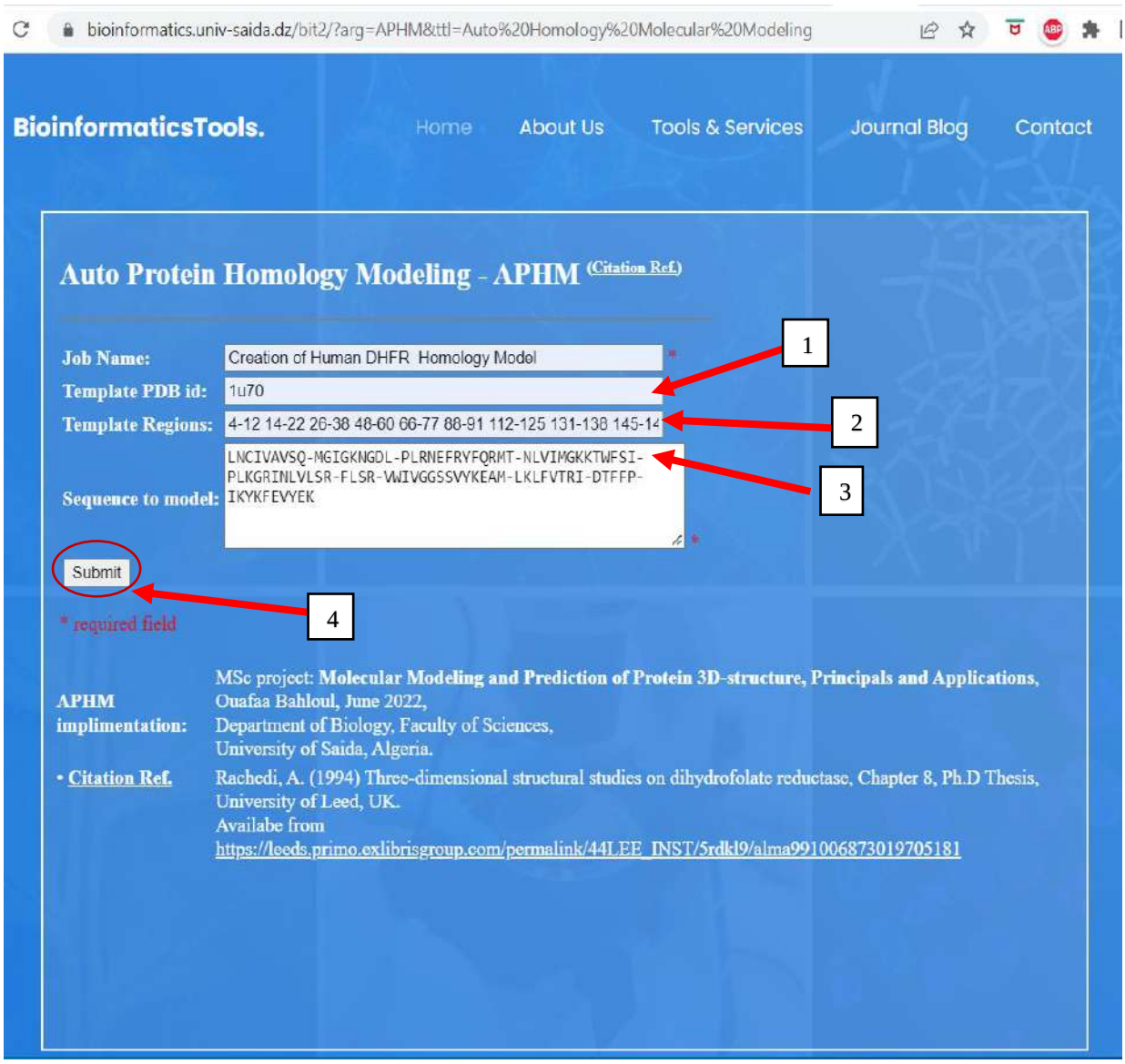



**Figure 23.** APHM interface with input data of 1 structure template 1U70 , 2 core regions and 3 Seq_1 amino-acids sequence. To builf the side-chain and model, the user need to click button Submit encircled in red 4.

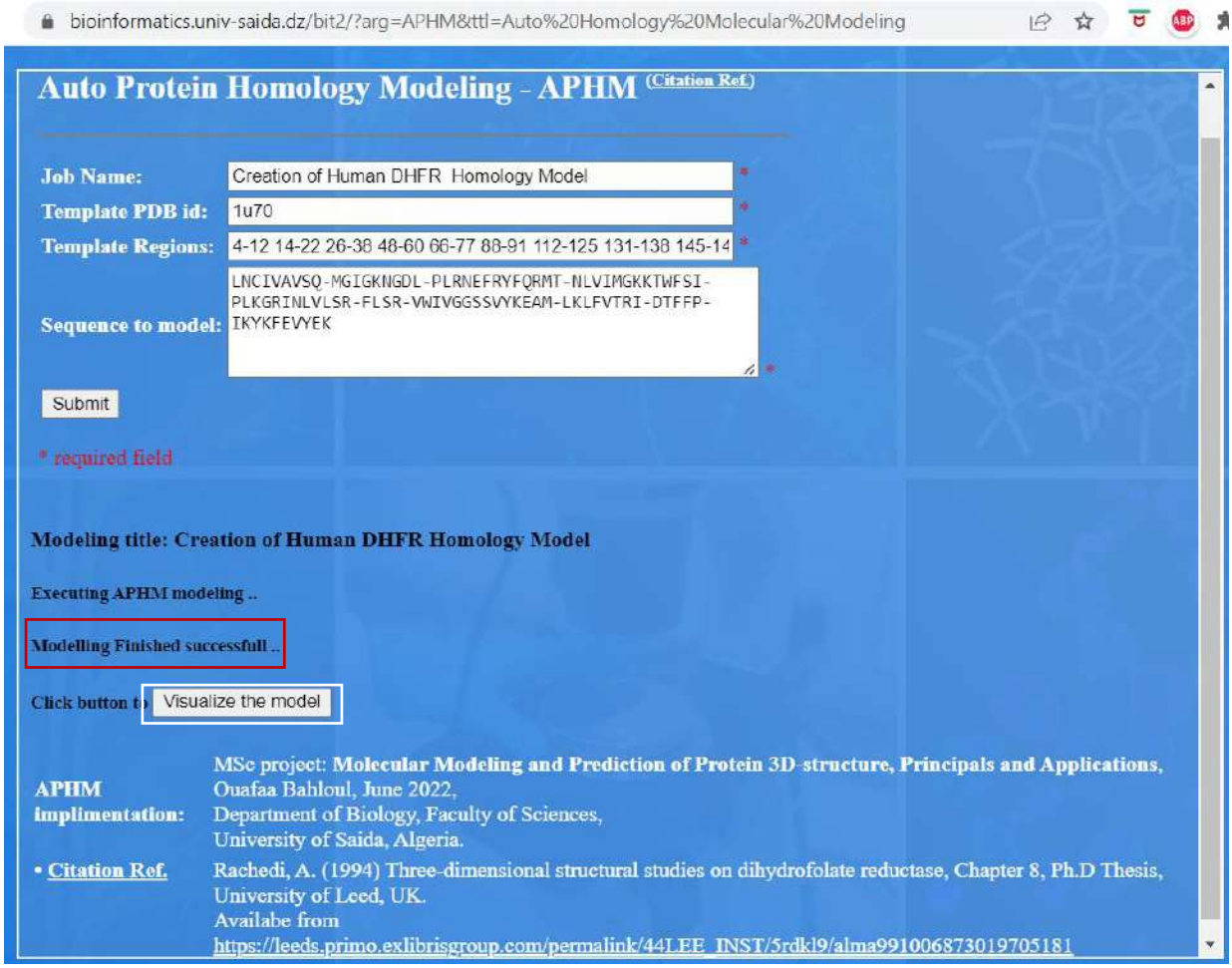



**Figure 24.** APHM after execution, if succefull as outlined in red rectangle, it provides a button for visualzing the created molecule, the button outlined in white rectangle which requires clicking to proceed to the APHM 3D-structure viewer.

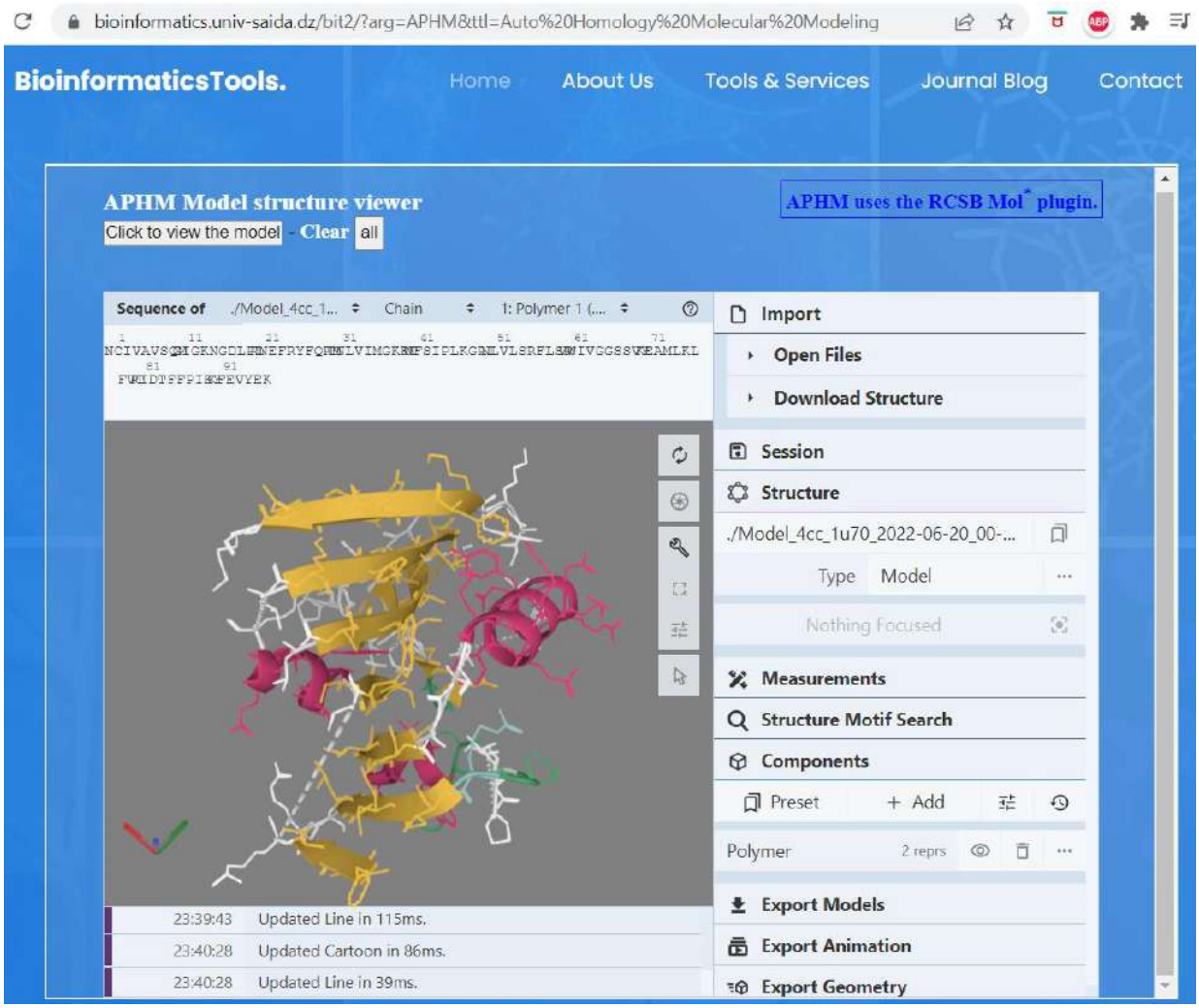



**Figure 25.** The APHM 3D-viewer displaying the Model created based Seq_1 and core regions suitable to the template structure 1U70.

Building the model using the core template 1DZR, follows the same procedure except that the structure template sould be the PDB id 1DZR and the related core regions mentioned above, see Figures 26 and 27.

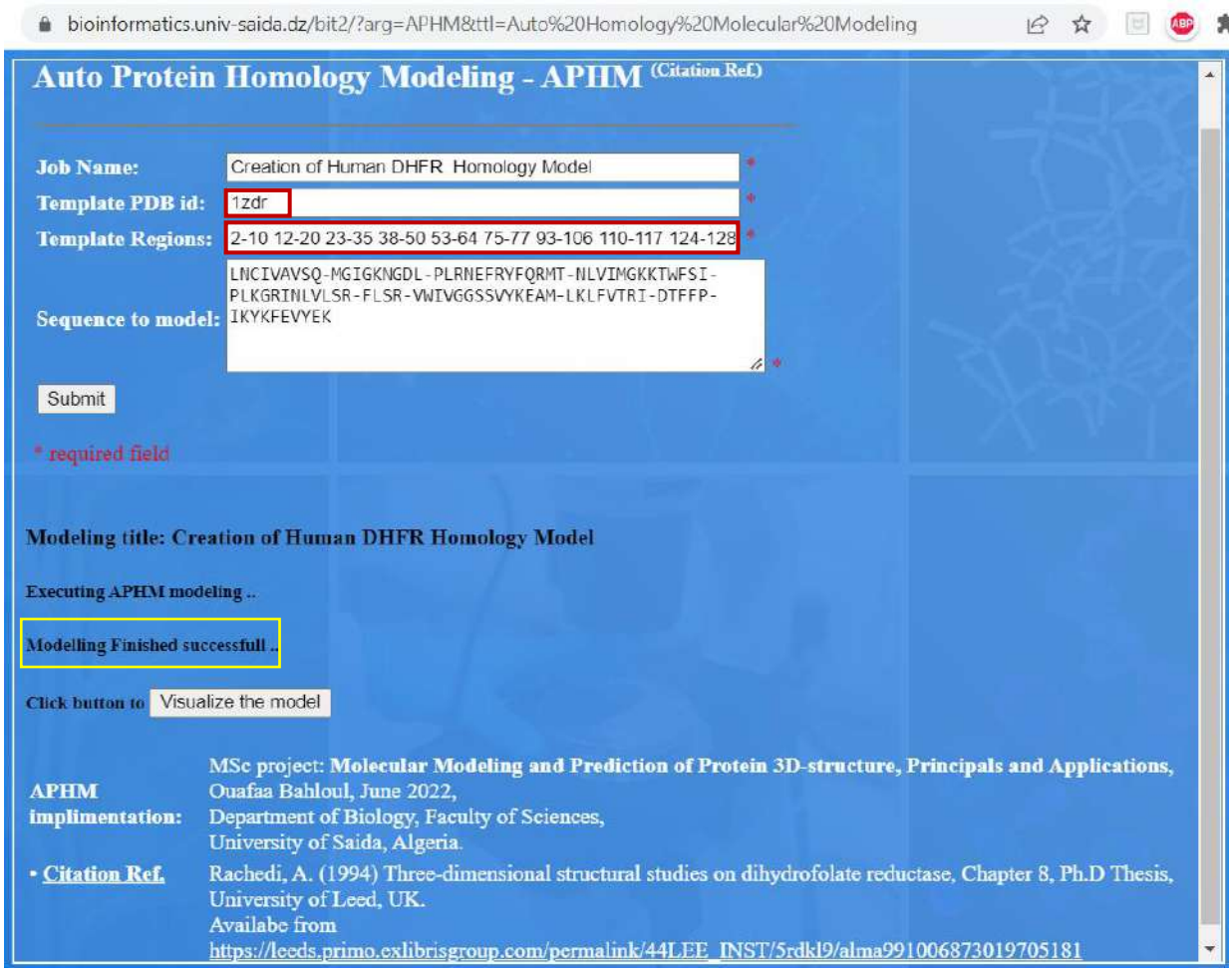



**Figure 26.** APHM after execution with the bacterial template structure 1ZDR and template regions (core regions) highlighted in red. Outlined in yellow show model building.

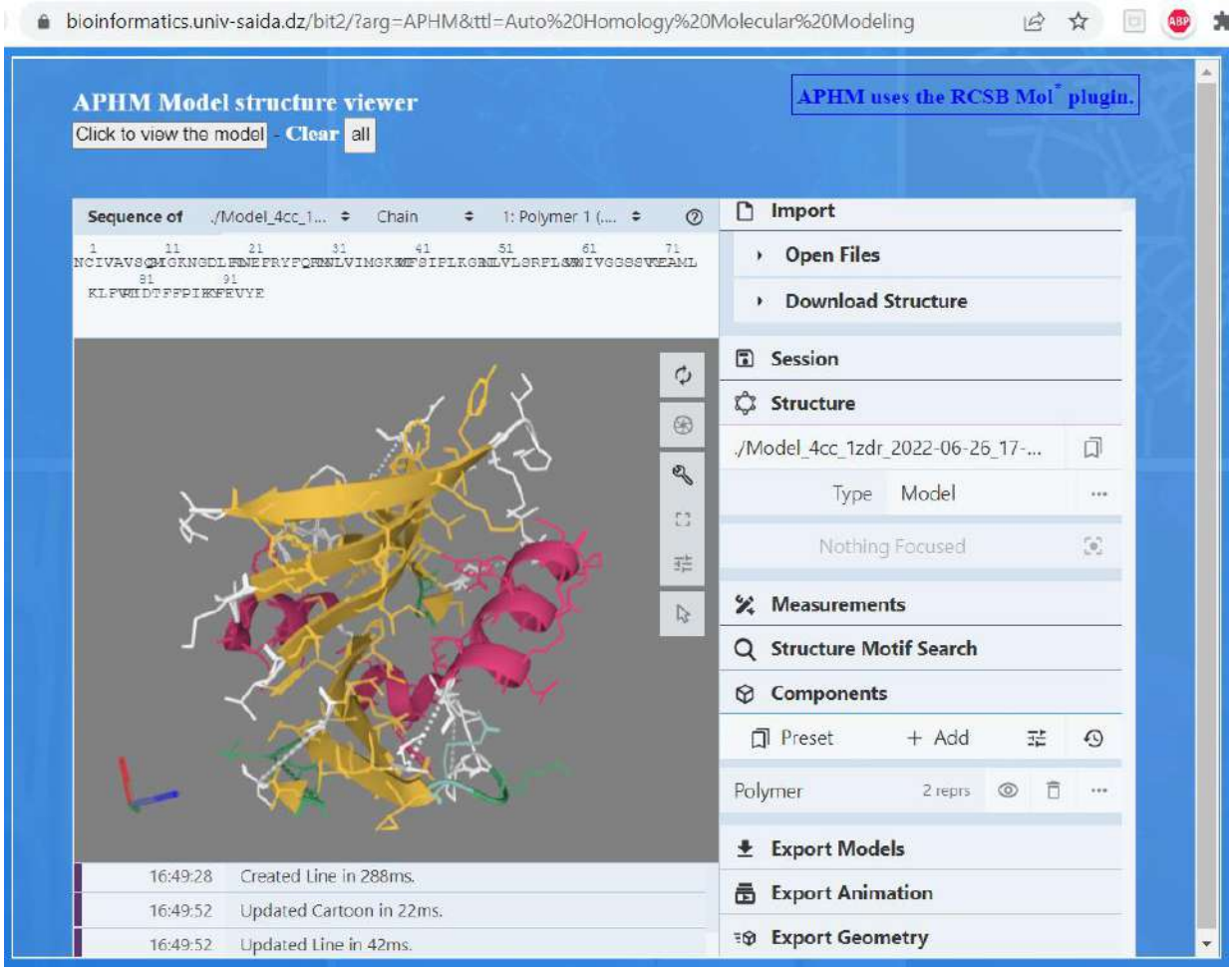



**Figure 27.** The APHM 3D-viewer displaying the Mode created based Seq_2 and core regions suitable to the template structure 1ZDR.

It should be noted that both models are the same in their conformational structure, however sabtle differences exist and are explained in the next Chapter III Results and Discussion.

**II.6. Loop regions building:**

As seen above, the calculated model is composed mainly made of commom coreoregions which mean linking regions known as loop regions and may be even some secondary structure elemnts or parts of.

However due to time and resources limitation, this step is left out. This is also based on the established knowledge in structural and functional biology, that the common core region would help the essecial of biological function of proteins.

**II.7. Final model:**

The final model in this project is predicted for the query sequence based on the common structural core preserved in 9 structures despite that different species and evoluationay history.

The schematic below, Figure 28, summarises the procedure followed into achieving the final model executed manually for academic and teaching purposes being the goals behind this project.

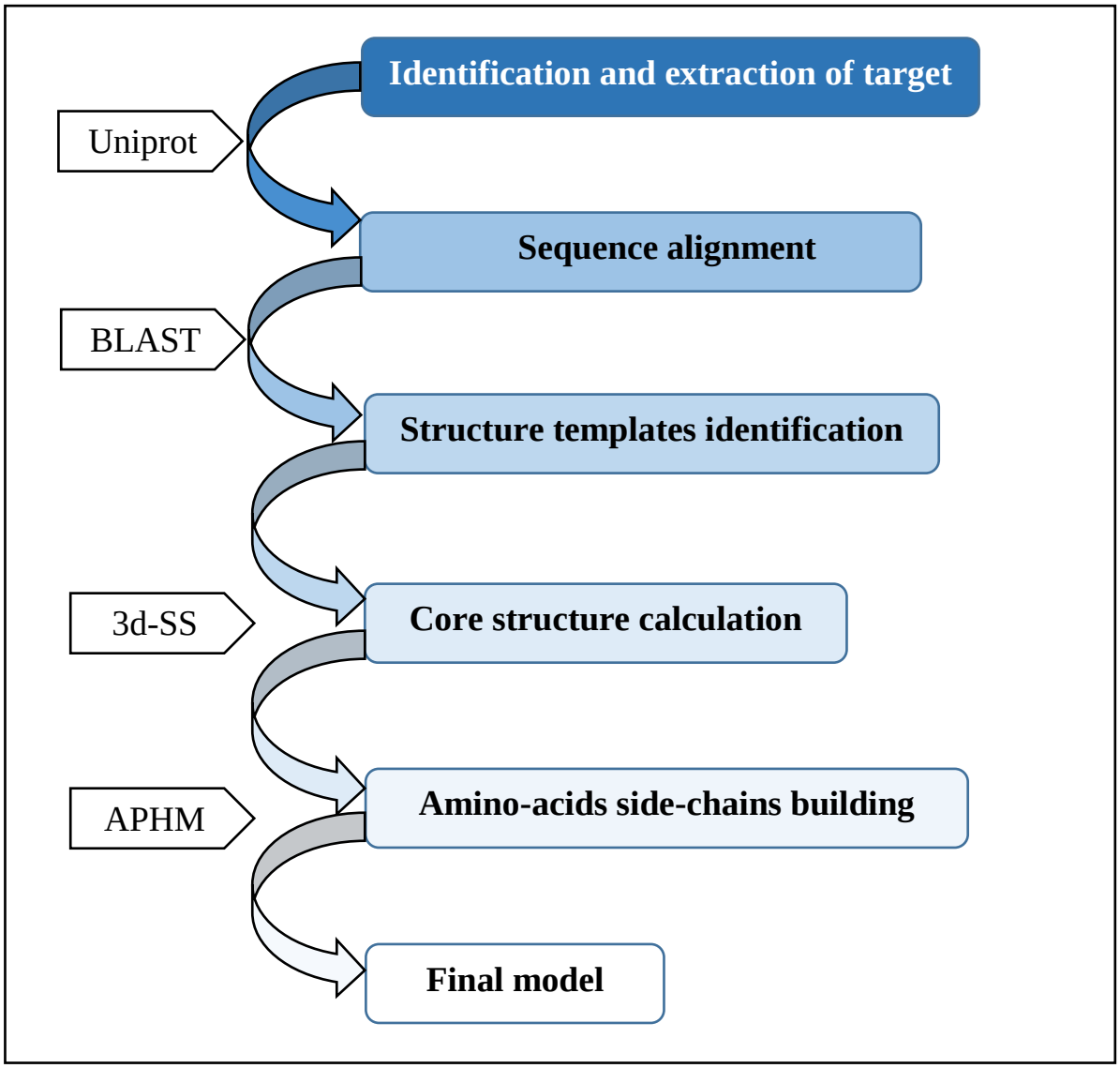



**Figure 28.** A schematic diagram of all the steps of homology modeling procedure implimented to reach the final model.

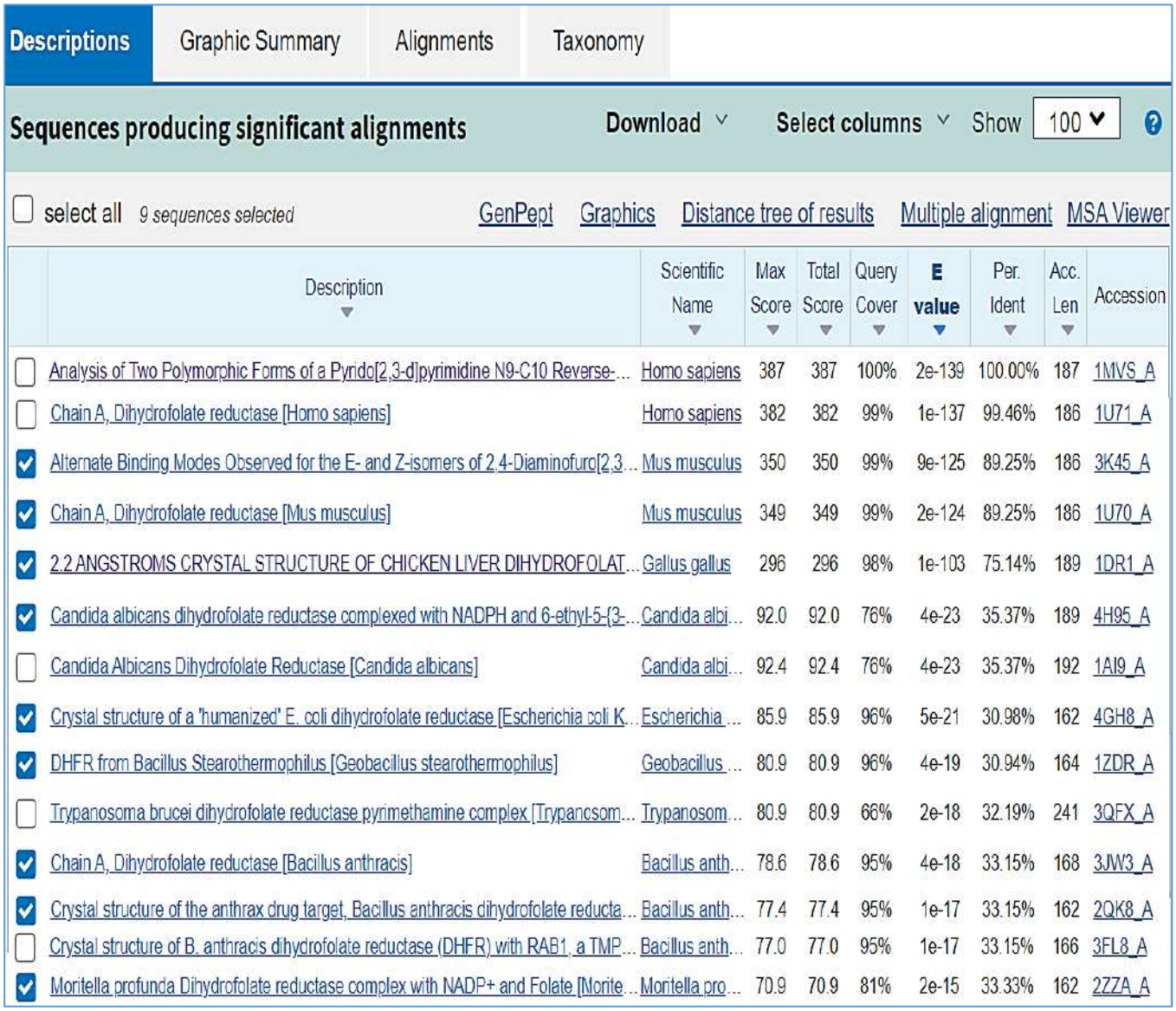
Discussion of the evaluation and application of such 3D-structure prediction is dealt with in the next Chapter III Results and Discussion.

# Chapter III: Results & Discussion

## I. Presentation of the results

### I.1. Structure similarity identification

The process of molecular modelling investigation towards structure prediction for protein sequences necessitate the identification of known protein structures the sequences of which share similarity with to the query sequence. In this project a given human Dihydrofolate reductase sequence has been selected and similarity has been discovered using BLAST alignment search against a library of sequences from known structures (PDB sequence library) each with a PDB id as shown in Figure 1; refer to Table 1 and Figure 3 in Chapter II for more details.



Descriptions | Graphic Summary | Alignments | Taxonomy

**Sequences producing significant alignments** Download Select columns Show 100

select all 9 sequences selected GenPept Graphics Distance tree of results Multiple alignment MSA Viewer

| Description | Scientific Name | Max Score | Total Score | Query Cover | E value | Per. Ident | Acc. Len | Accession |
|---|---|---|---|---|---|---|---|---|
| Analysis of Two Polymorphic Forms of a Pyrido[2,3-d]pyrimidine N9-C10 Reverse-... | Homo sapiens | 387 | 387 | 100% | 2e-139 | 100.00% | 187 | 1MVS_A |
| Chain A, Dihydrofolate reductase [Homo sapiens] | Homo sapiens | 382 | 382 | 99% | 1e-137 | 99.46% | 186 | 1U71_A |
| Alternate Binding Modes Observed for the E- and Z-isomers of 2,4-Diaminofuro[2,3... | Mus musculus | 350 | 350 | 99% | 9e-125 | 89.25% | 186 | 3K45_A |
| Chain A, Dihydrofolate reductase [Mus musculus] | Mus musculus | 349 | 349 | 99% | 2e-124 | 89.25% | 186 | 1U70_A |
| 2.2 ANGSTROMS CRYSTAL STRUCTURE OF CHICKEN LIVER DIHYDROFOLAT... | Gallus gallus | 296 | 296 | 98% | 1e-103 | 75.14% | 189 | 1DR1_A |
| Candida albicans dihydrofolate reductase complexed with NADPH and 6-ethyl-5-{3-... | Candida albi... | 92.0 | 92.0 | 76% | 4e-23 | 35.37% | 189 | 4H95_A |
| Candida Albicans Dihydrofolate Reductase [Candida albicans] | Candida albi... | 92.4 | 92.4 | 76% | 4e-23 | 35.37% | 192 | 1AI9_A |
| Crystal structure of a 'humanized' E. coli dihydrofolate reductase [Escherichia coli K... | Escherichia ... | 85.9 | 85.9 | 96% | 5e-21 | 30.98% | 162 | 4GH8_A |
| DHFR from Bacillus Stearothermophilus [Geobacillus stearothermophilus] | Geobacillus ... | 80.9 | 80.9 | 96% | 4e-19 | 30.94% | 164 | 1ZDR_A |
| Trypanosoma brucei dihydrofolate reductase pyrimethamine complex [Trypancsom... | Trypanosom... | 80.9 | 80.9 | 66% | 2e-18 | 32.19% | 241 | 3QFX_A |
| Chain A, Dihydrofolate reductase [Bacillus anthracis] | Bacillus anth... | 78.6 | 78.6 | 95% | 4e-18 | 33.15% | 168 | 3JW3_A |
| Crystal structure of the anthrax drug target, Bacillus anthracis dihydrofolate reducta... | Bacillus anth... | 77.4 | 77.4 | 95% | 1e-17 | 33.15% | 162 | 2QK8_A |
| Crystal structure of B. anthracis dihydrofolate reductase (DHFR) with RAB1, a TMP... | Bacillus anth... | 77.0 | 77.0 | 95% | 1e-17 | 33.15% | 166 | 3FL8_A |
| Moritella profunda Dihydrofolate reductase complex with NADP+ and Folate [Morite... | Moritella pro... | 70.9 | 70.9 | 81% | 2e-15 | 33.33% | 162 | 2ZZA_A |

**Figure 01.** Depiction from BLAST output sequence alignment run involving alignment search for the query sequence against the library of the nine chosen PDB sequences.

**I.2. Definition of the types of structures selected from BLAST results:**

The selected structures are from different species some show even lower sequence identity with the query sequence as low as 30%. This choice has been taken to test the theory that a particular biological function would be preserved over evolutionary history via the preservation of a particular structural configuration among other properties (***Rives A et al., 2021***).

The species of the sequences of the similar structures and function are briefly described along with sequence similarity in the following:

✓ **Mammals**

***Mus Musculus*:**

The house mouse, it has been domesticated as a pet, as well as an important laboratory mouse, and it is one of the most significant model animals in biology and medicine, It has characteristics like a pointed nose, wide rounded ears, and a long, almost hairless tail.

**Similarity percentage to the query sequence:** over 89%

**Structure in PDB identification codes**:

- 3K45 chain A, denoted in Figure 01 as 3K45_A (**Gangjee A et al., 2009).**

Details found here https://bioinformaticstools.org/ssfs/ssfs.php?qry=3k45

- 1U70 chain A. denoted in Figure 01 as 1U70_A (**Cody V et al., 2005).**

Details available here https://bioinformaticstools.org/ssfs/ssfs.php?qry=1u70

✓ **Birds**

***Gallus Gallus*:**

The red jungle fowl is a tropical bird, distinguished by its red color and lives in Southeast Asia. It is the species from which the chicken evolved.

**Similarity percentage to the query sequence:** over 75%

**Structure in PDB identification code**:

- 1DR1 chain A, denoted in Figure 01 as 1DR1_A (**McTigue M-A et al., (1992).**

Details found here https://bioinformaticstools.org/ssfs/ssfs.php?qry=1dr1

✓ **Yeast – (microbe)**

***Candida albicans*:**

Is an opportunistic pathogenic yeast that is found in the human gut flora.

**Structure in PDB identification code**:

- 4H95 chain A, denoted in Figure 01 as 4H95_A (**DeJarnette C et al., 2020)**

Details found here https://bioinformaticstools.org/ssfs/ssfs.php?qry=4h95

**Similarity percentage to the query sequence:** over 35%

✓ **Bacteria (Microbe)**

❖ ***Escherichia coli*:** Bacteria are found mainly in the intestines of healthy humans and animals.

**Structure in PDB identification code:**

- 4GH8 chain A, denoted in Figure 01 as 4GH8_A (**Srinivasan B et al., 2019)**

Details found here https://bioinformaticstools.org/ssfs/ssfs.php?qry=4gh8

❖ ***Bacillus*** species**:** These are Gram-positive bacteria, and causative agents of a number of diseases such as anthrax (a common disease of livestock and sometimes humans).

**Structure in PDB identification code:**

- 3JW3 chain A, denoted in Figure 01 as 3JW3_A (**Beierlein J-M et al., 2010)**.

Details found here https://bioinformaticstools.org/ssfs/ssfs.php?qry=3jw3

❖ ***Moritella profunda*:** from the Moritellaceae family these are marine bacteria. Most species grow at relatively low temperatures.

**Structure in PDB identification code:**

- 2ZZA chain A, denoted in Figure 01 as 2ZZA_A (**Penhallurick R-W & Ichiye T., 2021)**

Details found here https://bioinformaticstools.org/ssfs/ssfs.php?qry=2zza

**Similarity percentage to the query sequence:** over 30%

### I.3. Model template

The final model template can be calculated from vertebrate or bacterial based common core and in both cases it is composed of 10 regions.

- From vertebrate: represented here by the PDB id **1U70** of the mouse DHFR**:**
  As previously mentioned, in the second chapter, the sequence difference in the core region as compared to the query sequence does not contain many amino-acids different from the query sequence, Figure 2.

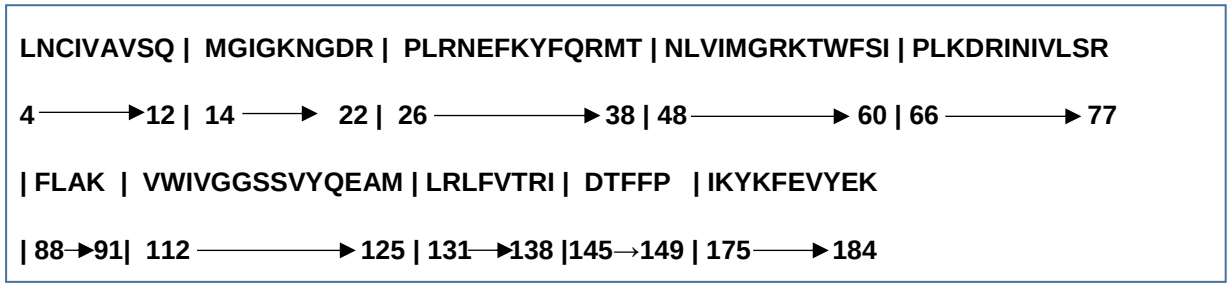



**Figure 02.** Amino-acids sequence from the *Mus musculus*, representing the core regions used in the model template.

- From bacteria: represented in the project by the **1ZDR.** Building a model based on template from microbes requires modelling the side chains of a large number of amino-acids that are different from the query sequence, Figure 3.

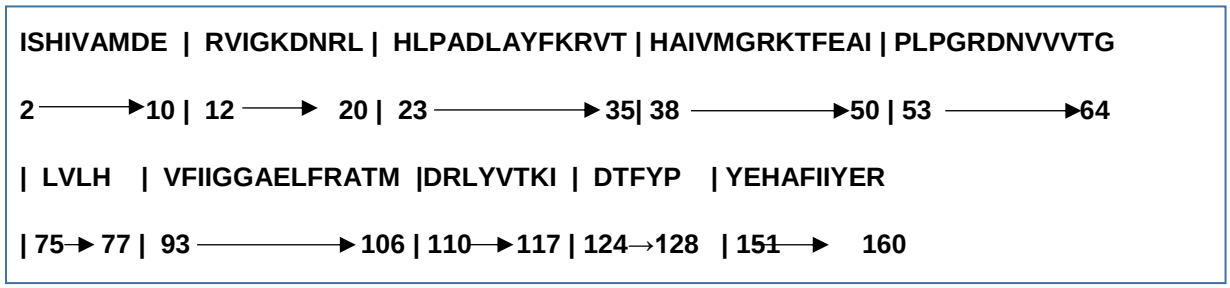



**Figure 03.** Amino-acids sequence from the *Geobacillus stearothermophilus*, representing the core regions used in the model template.

### I.4. Final model and completeness:

- Template has been completed as seen in chapter II by building side chains, Because the uniqueness of a protein structure is primarily defined by its side-chain packing conformation, side-chain modeling is crucial for protein structure prediction. which mainly contains the following secondary structures (seven β-sheet and 3 α-helices) see Figure 4.
- The final model in this project lacks the loop regions that were not possible to do due to the shortness for time and resources.
- Even that model lacks some non-conserved regions, it contains important regions for biological function and ligands binding site residues. Refer below to

**II.4 Catalytic Residues Preservation** for further details.

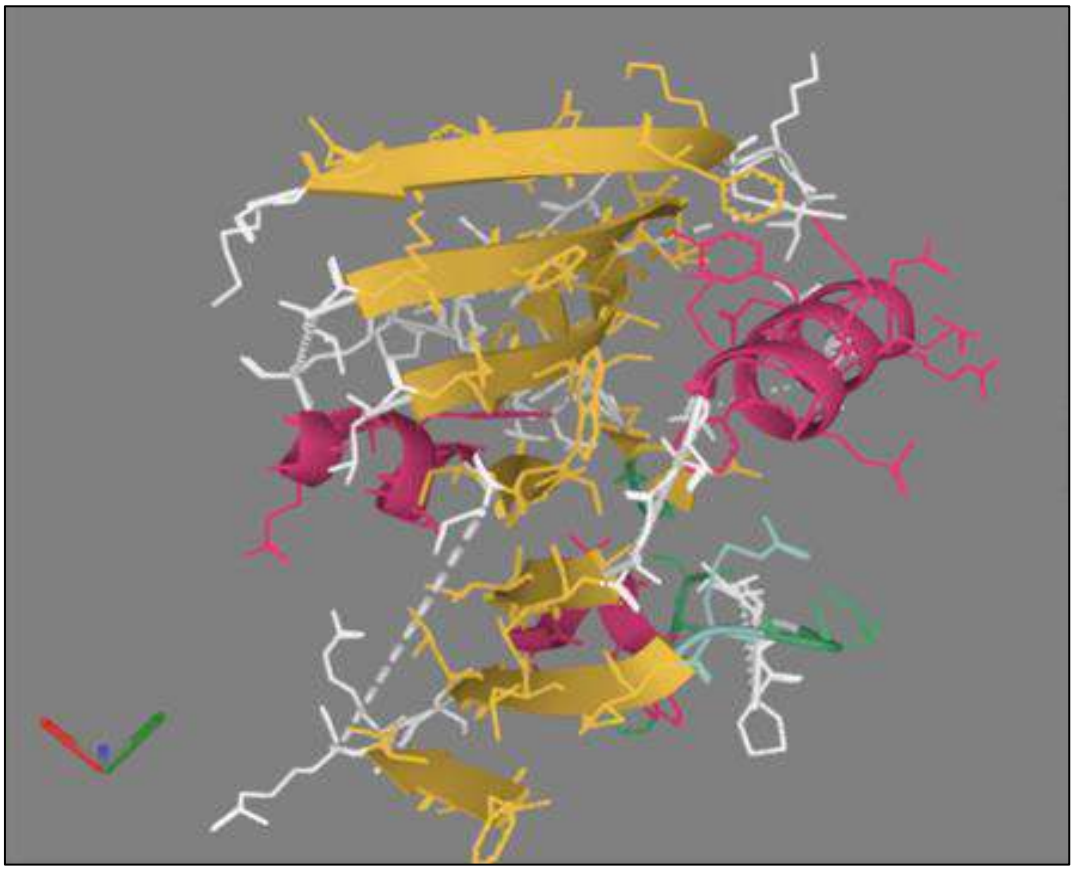


**Figure 04.** The APHM 3D-viewer displaying the Model created based Sequence and core regions suitable to the template structure 1U70.

## II. Modelling quality assessment and validity

The evaluating terms for assessing the quality and reliability of the final models are many, important ones are described in the following:

### II.1. RMSD overall superposition:

In overall superposition, the RMSD values, as in Table 1, are all less than 1.5Å, reflecting a good superposition fit for modelling and acceptable quality of structure resemblance since the RMSD value is much less than the inter-atomic distance between heavy atoms (all atoms except hydrogen) in each of the involved structures.

**Table 01.** RMSD values in overall superposition

| **PDB ID** | **1U70** | **1DR1** | **4H95** | **4GH8** | **1ZDR** | **3JW3** | **2QK8** | **2ZZA** |
|---|---|---|---|---|---|---|---|---|
| **RMSD (Å)** | 0,537 | 0,889 | 1,295 | 1,272 | 1,118 | 1,276 | 1,412 | 1,241 |

### II.2. RMSD core based superposition:

This RMSD are less than 1Å that reflect a very high quality for structural modeling since this distance is comparable to inter atomic distances seen in good proteins structures. Note that the value of RMSD decreased significantly compared to the value of the overalls, after choosing the similar and identical regions between the structures.

**Table 02.** RMSD directed per region values

| **PDB ID** | **1U70** | **1DR1** | **4H95** | **4GH8** | **1ZDR** | **3JW3** | **2QK8** | **2ZZA** |
|---|---|---|---|---|---|---|---|---|
| **RMSD (Å)** | 0.464 | 0.710 | 0.719 | 0.726 | 0.724 | 0.744 | 0.832 | 0.731 |

### II.3. Common core and secondary structure

As graphically shown above, Figure 4, the common core contains most of the secondary structures though an element or two are missing suggesting that their role in enzymatic function is in direct.

The number of secondary structures present in the common core are all between six and seven β-stands forming a single central β-sheet and between three and four α-helices in all nine selected structures see Figure 5 and Figure 6.


**Figure 05.** Structure based alignment of the query (target) sequence with the bacterial sequence (PDB id: 2ZZA). Alignment shows the correlation of CC regions with the secondary structure; Most of the secondary elements, four α-helices and six β-sheet are contained in the CC regions as highlighted.


**Figure 06.** Structure based sequence alignment of the query sequence with the vertebrate sequence (PDB id: 1DR1). Alignment shows the correlation of CC regions with the secondary structure; Most of the secondary elements, three α-helices and seven β-sheet are contained in the CC regions as highlighted.

It's here noted that sequences alignments of the query sequence with the PDB ids 1DR1 and 2ZZA from avian and bacterial sources respectively are used above. This is in contrast with mouse 1U70 and bacterial 1ZDR, described in Chapter II in Figures 19 and 20, to further show that the model template is the same in all these cases as a further demonstration for the notion of structure/function relationship conservation across the species.

### II.4. Catalytic Residues Preservation:

Based on the ligand binding environment of human DHFR in the structures PDB-id 1DHF **(Davies, J-F et al., 1990 )** and PDB-id 1U72 (**Cody Vet al., 2005**) calculate using the SSFS tool **(Rachedi A, 2011 et al., 2005)**, the residues compose catalytic sites for the ligands Folate (substrate) and NADPH (cofactor) are the following:

- Folate binding residues: Glutamate E-30 and Arginine R-70 highlighted as underlined green coloured amino-acids in Figure-7. Refer to online table accessed from: https://bioinformaticstools.org/ssfs/sitects.php?fl=1dhf_3643912495&ht=FOL:A187 For details on the 1DHF structure refer to: https://bioinformaticstools.org/ssfs/ssfs.php?qry=1dhf&stp=smr

- NADPH binding residues: Alanine A-9, Isoleucine I-16, Lysines K-54 & K-55, Threonine T-56, Serines S-59 & S-76, Arginine R-77 and Valine V-120 highlighted as underlined red coloured amino-acids in Figure-7. Refer to online table accessed from: https://bioinformaticstools.org/ssfs/sitects.php?fl=1u72_6413100842&ht=NDP:A187
- For details on the 1U70 structure refer to: https://bioinformaticstools.org/ssfs/ssfs.php?qry=1u72&stp=smr

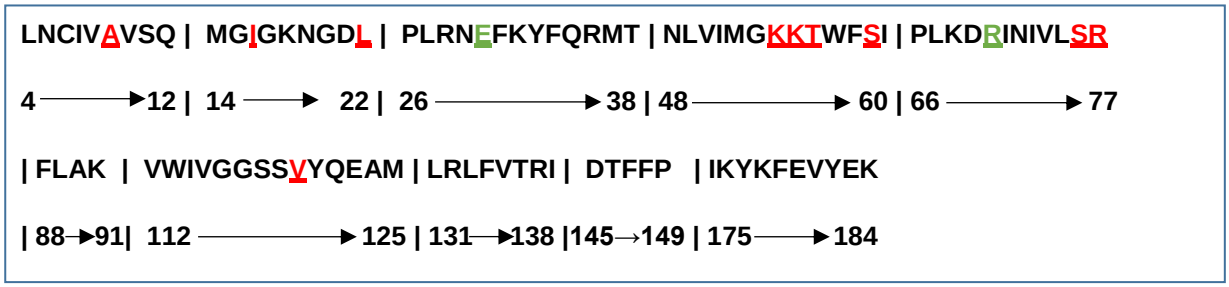



**Figure 07.** Amino-acids sequence representing the core regions of human DHFR. Substrate and cofactor binding residues preserved in the final predicted model.

As is clearly shown in Figure 7, all of the binding site residues for the substrate and cofactor are preserved in the core region of the created model of DHFR. This is a clear indication for the reliability of the modeling procedure implemented in this project. Such a homology based model can thus be quite instrumental in carrying important structural and functional studies on the effects of mutations on the activity of enzymes as also summarized in general conclusion below,

**III. Choice of model building template**

As per the methodology described in Chapter II the final model can be built from two common sources vertebrates and bacterial.

The most basic template selection rule is to choose the structure that has the greatest sequence similarity to the modelled sequence, availability and the quality of template.

In addition, the chosen sequence must be looked at, this study was done for a human dihydrofolate reductase sequence, which favours using the closest type which in this case would be the vertebrate template though it should not matter much if the bacterial template is use.

**IV. Bioinformatics tools development**

For the purpose of demonstrating the manual homology molecular procedure followed in this project, an online tool was created representing an implementation of the molecular modelling tool Auto Protein Homology Modeling or APHM **(Rachedi, A. 1994)** towards executing amino-acids side-chain atoms and graphical visualization of the predicted model and its structural details.

APHM & Molecular Graphic Viewer implementation is found on the Bioinformatics Server at the University of Saida, and is accessible online at the web-address:

https://bioinformatics.univ-saida.dz/bit2/?arg=APHM

See Figure 4 and for examples on how to use the implementation of APHM refer section "**II.5. Amino-acids side-chains building:"** and Figures 23 to 26 in Chapter II.

The APHM implementation, besides the visualization, offers also the ability to download the actual xyz coordinates of the final model to examine and/or visualize using other molecular graphic tools, see Figure 8.

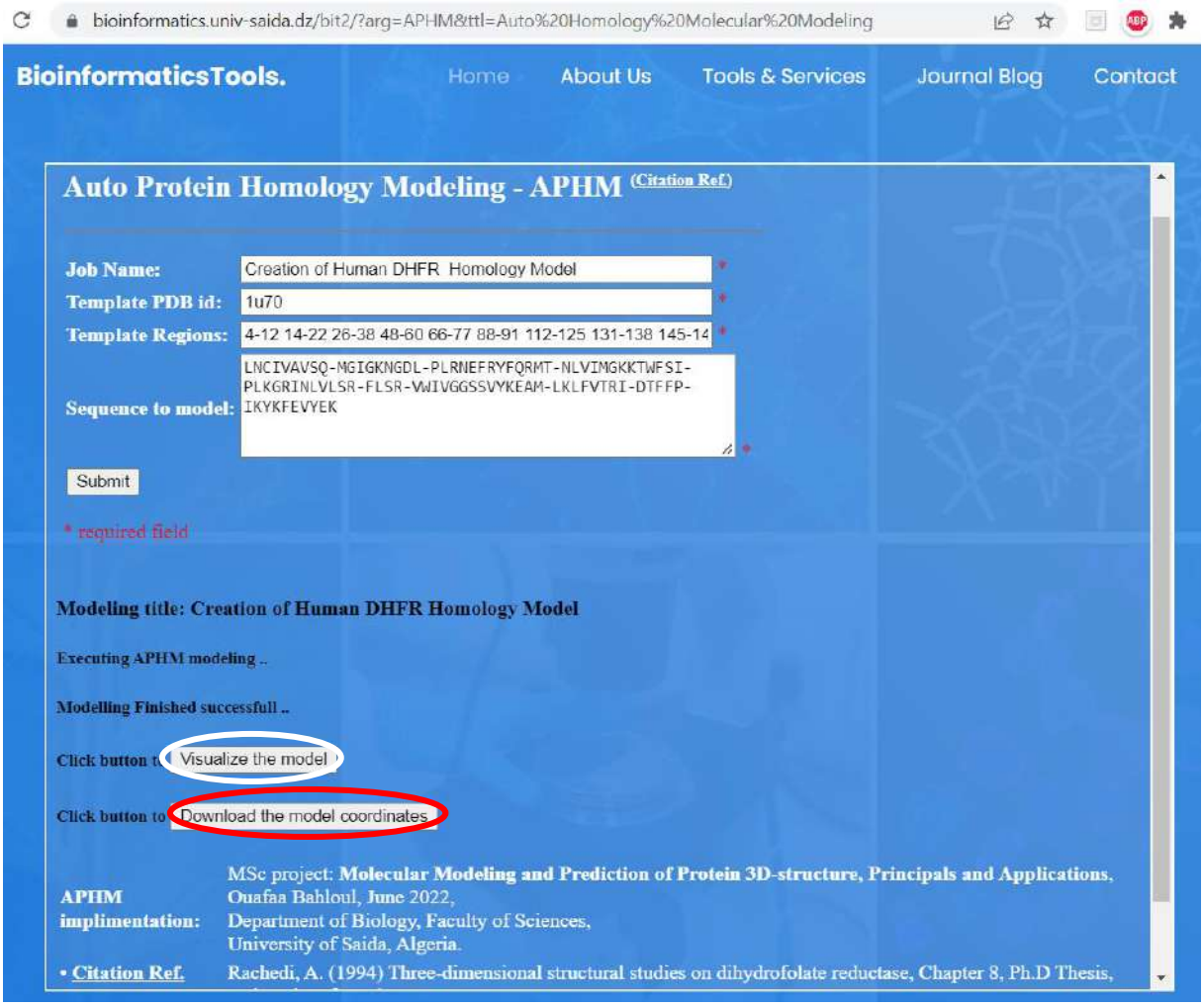



**Figure 08.** APHM interface page. An example showing model building successful execution that provide button for model structure visualization, white encircled, and button for downloading xyz coordinates, encircled in red.

# General Conclusion

# General conclusion

This project with the results obtained in it has demonstrated a number of important biology function related principals:

- The particular function like doing or mediating the action of reducing Dihydrofolate to Tetrahydrofolate, necessary for cellular division is preserved, throughout even thousand millions of evolutionary history. by a well defined three-dimensional architecture or arrangement of secondary structure elements and other elements not part of the secondary structure (as seen above.

- The modelling procedure requires side-chain building of dissimilar amino acids found between the query sequence and sequence of model templates. This mean that models created by homology modelling, amongst many things, can be used to predict the effects of mutations and their possible effect on the biological function of proteins.

- The results above showed that preserved common core between evolutionary diverted proteins, from different species, would contain regions that are not part of secondary structures which indicate that biological function is not necessarily associated with the secondary structure. This would call scientist to broaden the scope of their investigation for better understanding of the roots behind the rise of the biological function.

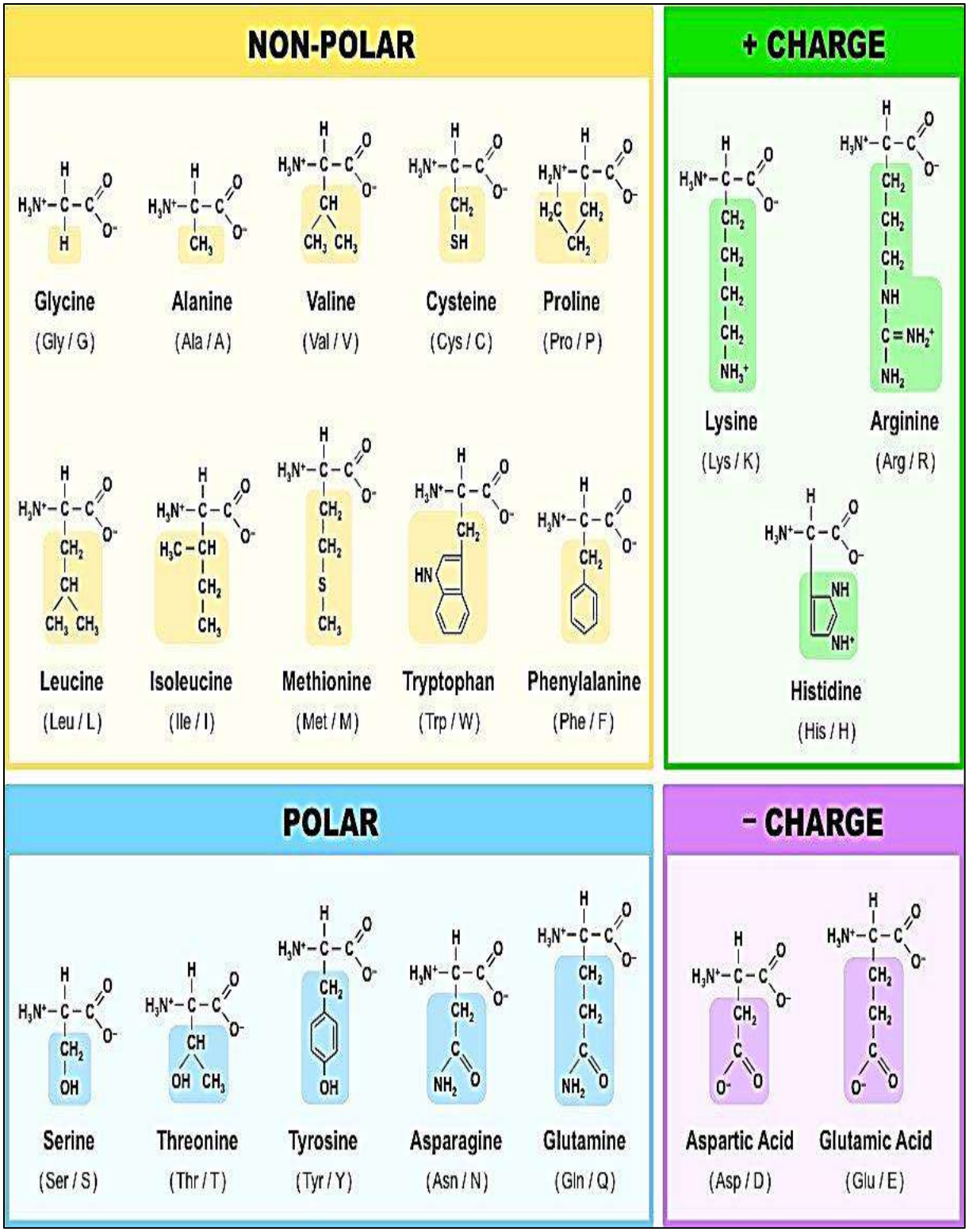
- Such work of creating homology based models can thus be used for rational design of new drugs that can be tested (used) against cancer disease and bacterial infections through the analysis of catalytic and ligand binding residues preserved on the common core in the final model.

# Appendix

# Appendix Part I

# Appendix part I

## 1) Structure of the 20 amino acids




The 20 amino acids that serve as the building blocks for proteins. These amino acids (shown with their respective three-letter abbreviations) can be divided into nonpolar, polar, and electrically charged groups. **(Cornell, B. 2016. Amino Acids. Bio ninja).**

# Appendix part I

**2) The amino acid codes supported (22 amino acids and 3 special codes) are:**

| Amino Acid Code | Meaning |
|---|---|
| **A** | Alanine |
| **B** | Aspartic acid (D) or Asparagine (N) |
| **C** | Cysteine |
| **D** | Aspartic acid |
| **E** | Glutamic acid |
| **F** | Phenylalanine |
| **G** | Glycine |
| **H** | Histidine |
| **I** | Isoleucine |
| **J** | Leucine (L) or Isoleucine (I) |
| **K** | Lysine |
| **L** | Leucine |
| **M** | Methionine/Start codon |
| **N** | Asparagine |
| **O** | Pyrrolysine (rare) |
| **P** | Proline |
| **Q** | Glutamine |
| **R** | Arginine |
| **S** | Serine |
| **T** | Threonine |
| **U** | Selenocysteine (rare) |
| **V** | Valine |
| **W** | Tryptophan |
| **Y** | Tyrosine |
| **Z** | Glutamic acid (E) or Glutamine (Q) |
| **X** | any |
| ***** | translation stop |
| **-** | gap of indeterminate length |

Sequences represented in the standard IUPAC amino acid codes; a single hyphen or dash can be used to represent a gap character **"**derived from **IUPAC code table**". NIAS DNA Bank. Archived from the original on **2011-08-11"**.

# Appendix Part II

# Appendix part II

**Table 01.** Hit sequence with PDB structure id: 3K45

**Sequence ID: 3K45_A Length: 186 Number of Matches: 1**

**Range 1: 1 to 186**

| Score | Expect | Method | Identities | Positives | Gaps |
|---|---|---|---|---|---|
| 350 bits(898) | 8e-125 | Compositional matrix adjust. | 166/186(89%) | 179/186(96%) | 0/186(0%) |

```
Query  2    VGSLNCIVAVSQNMGIGKNGDLPWPPLRNEFRYFQRMTTTSSVEGKQNLVIMGKKTWFSI  61
            V  LNCIVAVSQNMGIGKNGDLPWPPLRNEF+YFQRMTTTSSVEGKQNLVIMG+KTWFSI
Sbjct  1    VRPLNCIVAVSQNMGIGKNGDLPWPPLRNEFKYFQRMTTTSSVEGKQNLVIMGRKTWFSI  60

Query  62   PEKNRPLKGRINLVLSRELKEPPQGAHFLSRSLDDALKLTEQPELANKVDMVWIVGGSSV  121
            PEKNRPLK RIN+VLSRELKEPP+GAHFL++SLDDAL+L EQP+LA+KVDMVWIVGGSSV
Sbjct  61   PEKNRPLKDRINIVLSRELKEPPRGAHFLAKSLDDALRLIEQPDLASKVDMVWIVGGSSV  120

Query  122  YKEAMNHPGHLKLFVTRIMQDFESDTFFPEIDLEKYKLLPEYPGVLSDVQEEKGIKYKFE  181
            Y+EAMN PGHL+LFVTRIMQ+FESDTFFPEIDL KYKLLPEYPGVLS+VQEEKGIKYKFE
Sbjct  121  YQEAMNQPGHLRLFVTRIMQEFESDTFFPEIDLGKYKLLPEYPGVLSEVQEEKGIKYKFE  180

Query  182  VYEKND  187
            VYEK D
Sbjct  181  VYEKKD  186
```

**Table 02.** Hit sequence with PDB structure id: 1U70

**Sequence ID: 1U70_A Length: 186 Number of Matches: 1**

**Range 1: 1 to 186**

| Score | Expect | Method | Identities | Positives | Gaps |
|---|---|---|---|---|---|
| 349 bits(896) | 2e-124 | Compositional matrix adjust. | 166/186(89%) | 178/186(95%) | 0/186(0%) |

```
Query  2    VGSLNCIVAVSQNMGIGKNGDLPWPPLRNEFRYFQRMTTTSSVEGKQNLVIMGKKTWFSI  61
            V  LNCIVAVSQNMGIGKNGD PWPPLRNEF+YFQRMTTTSSVEGKQNLVIMG+KTWFSI
Sbjct  1    VRPLNCIVAVSQNMGIGKNGDRPWPPLRNEFKYFQRMTTTSSVEGKQNLVIMGRKTWFSI  60

Query  62   PEKNRPLKGRINLVLSRELKEPPQGAHFLSRSLDDALKLTEQPELANKVDMVWIVGGSSV  121
            PEKNRPLK RIN+VLSRELKEPP+GAHFL++SLDDAL+L EQPELA+KVDMVWIVGGSSV
Sbjct  61   PEKNRPLKDRINIVLSRELKEPPRGAHFLAKSLDDALRLIEQPELASKVDMVWIVGGSSV  120

Query  122  YKEAMNHPGHLKLFVTRIMQDFESDTFFPEIDLEKYKLLPEYPGVLSDVQEEKGIKYKFE  181
            Y+EAMN PGHL+LFVTRIMQ+FESDTFFPEIDL KYKLLPEYPGVLS+VQEEKGIKYKFE
Sbjct  121  YQEAMNQPGHLRLFVTRIMQEFESDTFFPEIDLGKYKLLPEYPGVLSEVQEEKGIKYKFE  180

Query  182  VYEKND  187
            VYEK D
Sbjct  181  VYEKKD  186
```

# Appendix part II

**Table 03.** Hit sequence with PDB structure id: 1DR1

**Sequence ID:** 1DR1_A **Length:** 189 **Number of Matches:** 1

**Range 1: 1 to 185**

| Score | Expect | Method | Identities | Positives | Gaps |
|---|---|---|---|---|---|
| 296 bits(759) | 1e-103 | Compositional matrix adjust. | 139/185(75%) | 163/185(88%) | 0/185(0%) |

```
Query  2    VGSLNCIVAVSQNMGIGKNGDLPWPPLRNEFRYFQRMTTTSSVEGKQNLVIMGKKTWFSI  61
            V SLN IVAV QNMGIGK+G+LPWPPLRNE++YFQRMT+TS VEGKQN VIMGKKTWFSI
Sbjct  1    VRSLNSIVAVCQNMGIGKDGNLPWPPLRNEYKYFQRMTSTSHVEGKQNAVIMGKKTWFSI  60

Query  62   PEKNRPLKGRINLVLSRELKEPPQGAHFLSRSLDDALKLTEQPELANKVDMVWIVGGSSV  121
            PEKNRPLK RIN+VLSRELKE P+GAH+LS+SLDDAL L + PEL +KVDMVWIVGG++V
Sbjct  61   PEKNRPLKDRINIVLSRELKEAPKGAHYLSKSLDDALALLDSPELKSKVDMVWIVGGTAV  120

Query  122  YKEAMNHPGHLKLFVTRIMQDFESDTFFPEIDLEKYKLLPEYPGVLSDVQEEKGIKYKFE  181
            YK AM  P + +LFVTRI+ +FESDTFFPEID + +KLL EYPGV +D+QEE GI+YKFE
Sbjct  121  YKAAMEKPINHRLFVTRILHEFESDTFFPEIDYKDFKLLTEYPGVPADIQEEDGIQYKFE  180

Query  182  VYEKN  186
            VY+K+
Sbjct  181  VYQKS  185
```

**Table 04.** Hit sequence with PDB structure id: 4H95

**Sequence ID:** 4H95_A **Length:** 189 **Number of Matches:** 1

**Range 1: 6 to 146**

| Score | Expect | Method | Identities | Positives | Gaps |
|---|---|---|---|---|---|
| 92.0 bits(227) | 4e-23 | Compositional matrix adjust. | 52/147(35%) | 81/147(55%) | 10/147(6%) |

```
Query  7    CIVAVSQNMGIGKNGDLPWPPLRNEFRYFQRMTTTSSVEGKQNLVIMGKKTWFSIPEKNR  66
             + A+   +GIG  G +PW  LR E RYF+ +TT ++    +N VIMG+KTW SIP+K R
Sbjct  6    IVAALKPALGIGYKGKMPWR-LRKEIRYFKDVTTRTTKPNTRNAVIMGRKTWESIPQKFR  64

Query  67   PLKGRINLVLSRELKEPPQGAHFLSRSLDDALKLTEQPELANKVDMVWIVGGSSVYKEAM  126
            PL  R+N++LSR  +        +  ++  A  +     L + V+ V+I+GG+ +Y E +
Sbjct  65   PLPDRLNIILSRSYEN-----EIIDDNIIHASSIESSLNLVSDVERVFIIGGAEIYNELI  119

Query  127  NHPGHLKLFVTRIM----QDFESDTFF  149
            N+     L +T I     +  E DTF
Sbjct  120  NNSLVSHLLITEIEHPSPESIEMDTFL  146
```

# Appendix part II

**Table 05.** Hit sequence with PDB structure id: 4GH8

**Sequence ID:** 4GH8_A **Length:** 162 **Number of Matches:** 1

**Range 1: 2 to 162**

| Score | Expect | Method | Identities | Positives | Gaps |
|---|---|---|---|---|---|
| 85.9 bits(211) | 5e-21 | Compositional matrix adjust. | 57/184(31%) | 95/184(51%) | 26/184(14%) |

```
Query  5    LNCIVAVSQNMGIGKNGDLPWPPLRNEFRYFQRMTTTSSVEGKQNLVIMGKKTWFSIPEK  64
            ++ I A++ +  IG    +PWPPL  +  +F+R T    V       IMG+ TW SIPEK
Sbjct  2    ISLIAALAVDRVIGMENAMPWPPLPADLAWFKRNTLNKPV-------IMGRHTWESIPEK  54

Query  65   NRPLKGRINLVLSRELKEPPQGAHFLS--RSLDDALKLTEQPELANKVDMVWIVGGSSVY  122
            NRPL GR N++LS +    P     ++  +S+D+A+           V  + ++GG  VY
Sbjct  55   NRPLPGRKNIILSSQ----PGTDDRVTWVKSVDEAIA------ACGDVPEIMVIGGGRVY  104

Query  123  KEAMNHPGHLKLFVTRIMQDFESDTFFPEIDLEKYK-LLPEYPGVLSDVQEEKGIKYKFE  181
            ++ +  P   KL++T I  + E DT FP+ + + ++ +  E+     D   +    Y FE
Sbjct  105  EQFL--PKAQKLYLTHIDAEVEGDTHFPDYEPDDWESVFSEF----HDADAQNSHSYCFE  158

Query  182  VYEK  185
            + E+
Sbjct  159  ILER  162
```

**Table 06.** Hit sequence with PDB structure id: 1ZDR

**Sequence ID:** 1ZDR_A **Length:** 164 **Number of Matches:** 1

**Range 1: 2 to 160**

| Score | Expect | Method | Identities | Positives | Gaps |
|---|---|---|---|---|---|
| 80.9 bits(198) | 4e-19 | Compositional matrix adjust. | 56/181(31%) | 101/181(55%) | 22/181(12%) |

```
Query  5    LNCIVAVSQNMGIGKNGDLPWPPLRNEFRYFQRMTTTSSVEGKQNLVIMGKKTWFSIPEK  64
            ++ IVA+ +N  IGK+  LPW  L  +  YF+R+T   ++       +MG+KT+ +I
Sbjct  2    ISHIVAMDENRVIGKDNRLPWH-LPADLAYFKRVTMGHAI-------VMGRKTFEAI---  50

Query  65   NRPLKGRINLVLSRELKEPPQGAHFLSRSLDDALKLTEQPELANKVDMVWIVGGSSVYKE  124
             RPL GR N+V++      P+G   L  SL++  +      +A++ D V+I+GG+ +++
Sbjct  51   GRPLPGRDNVVVTGNRSFRPEGCLVL-HSLEEVKQW-----IASRADEVFIIGGAELFRA  104

Query  125  AMNHPGHLKLFVTRIMQDFESDTFFPEIDLEKYKLLPEYPGVLSDVQEEKGIKYKFEVYE  184
             M  P   +L+VT+I   F  DTF+P I  ++++++   PG      E+   ++ F +YE
Sbjct  105  TM--PIVDRLYVTKIFASFPGDTFYPPISDDEWEIVSYTPG---GKDEKNPYEHAFIIYE  159

Query  185  K  185
            +
Sbjct  160  R  160
```

# Appendix part II

**Table 07.** Hit sequence with PDB structure id: 3JW3

**Sequence ID: 3JW3_A Length: 168 Number of Matches: 1**

**Range 1: 12 to 166**

| Score | Expect | Method | Identities | Positives | Gaps |
|---|---|---|---|---|---|
| 78.6 bits(192) | 3e-18 | Compositional matrix adjust. | 59/178(33%) | 96/178(53%) | 23/178(12%) |

```
Query  8    IVAVSQNMGIGKNGDLPWPPLRNEFRYFQRMTTTSSVEGKQNLVIMGKKTWFSIPEKNRP  67
            +VA+ +N  IGK+ +LPW  L +E +Y ++ T    +       IMG+K + +I    RP
Sbjct  12   MVAMDENRVIGKDNNLPWR-LPSELQYVKKTTMGHPL-------IMGRKNYEAI---GRP  60

Query  68   LKGRINLVLSRELKEPPQGAHFLSRSLDDALKLTEQPELANKVDMVWIVGGSSVYKEAMN  127
            L GR N++++R      +G H      + A  + E  EL    + ++I+GG+ +Y   +
Sbjct  61   LPGRRNIIVTRN-----EGYHV--EGCEVAHSVEEVFELCKNEEEIFIIGGAQIYDLFL-  112

Query  128  HPGHLKLFVTRIMQDFESDTFFPEIDLEKYKLLPEYPGVLSDVQEEKGIKYKFEVYEK  185
             P   KL++T+I   FE DTFFPE+D+  +K +    G L+D  E+    Y + VYEK
Sbjct  113  -PYVDKLYITKIHHAFEGDTFFPEMDMTNWKEVFVEKG-LTD--EKNPYTYYYHVYEK  166
```

**Table 08.** Hit sequence with PDB structure id: 2QK8

**Sequence ID: 2QK8_A Length: 162 Number of Matches: 1**

**Range 1: 6 to 160**

| Score | Expect | Method | Identities | Positives | Gaps |
|---|---|---|---|---|---|
| 77.4 bits(189) | 1e-17 | Compositional matrix adjust. | 59/178(33%) | 95/178(53%) | 23/178(12%) |

```
Query  8    IVAVSQNMGIGKNGDLPWPPLRNEFRYFQRMTTTSSVEGKQNLVIMGKKTWFSIPEKNRP  67
            +VA+ +N  IGK+ +LPW  L +E +Y ++ T    +       IMG+K + +I    RP
Sbjct  6    MVAMDENRVIGKDNNLPWR-LPSELQYVKKTTMGHPL-------IMGRKNYEAI---GRP  54

Query  68   LKGRINLVLSRELKEPPQGAHFLSRSLDDALKLTEQPELANKVDMVWIVGGSSVYKEAMN  127
            L GR N++++R      +G H      + A  + E  EL    + ++I GG+ +Y   +
Sbjct  55   LPGRRNIIVTRN-----EGYH--VEGCEVAHSVEEVFELCKNEEEIFIFGGAQIYDLFL-  106

Query  128  HPGHLKLFVTRIMQDFESDTFFPEIDLEKYKLLPEYPGVLSDVQEEKGIKYKFEVYEK  185
             P   KL++T+I   FE DTFFPE+D+  +K +    G L+D  E+    Y + VYEK
Sbjct  107  -PYVDKLYITKIHHAFEGDTFFPEMDMTNWKEVFVEKG-LTD--EKNPYTYYYHVYEK  160
```

**Table 09.** Hit sequence with PDB structure id: 2ZZA

Sequence ID: 2ZZA_A Length: 162 Number of Matches: 1

Range 1: 3 to 133

| Score | Expect | Method | Identities | Positives | Gaps |
|---|---|---|---|---|---|
| 70.9 bits(172) | 2e-15 | Compositional matrix adjust. | 51/153(33%) | 85/153(55%) | 22/153(14%) |

```
Query  5    LNCIVAVSQNMGIGKNGDLPWPPLRNEFRYFQRMTTTSSVEGKQNLVIMGKKTWFSIPEK  64
            ++ I A++ N  IG +  +PW  L  E + F+R T      GK   ++MG+ T+ SI
Sbjct  3    VSMIAALANNRVIGLDNKMPWH-LPAELQLFKRATL-----GKP--IVMGRNTFESI---  51

Query  65   NRPLKGRINLVLSRELKEPPQGAHFLSRSLDDALKLTEQPELANKVDMVWIVGGSSVYKE  124
             RPL GR+N+VLSR+    P+G   ++ +L+DA+        A  V+ + I+GG+++Y +
Sbjct  52   GRPLPGRLNIVLSRQTDYQPEGVTVVA-TLEDAVV------AAGDVEELMIIGGATIYNQ  104

Query  125  AMNHPGHLKLFVTRIMQDFESDTFFPEIDLEKY  157
             +      +L++T I    E DT+FP  D E+Y
Sbjct  105  CLAAAD--RLYLTHIELTTEGDTWFP--DYEQY  133
```

# Appendix Part III

# Appendix part III

**a) Common Core calculation - overall superposition:**

As shown, all the local deviation between CA atoms of every matching residues in both structures is less and larger than 2 Å (the residues larger than 2 **Å** are highlighted in blue color).

**Table 01.** Residues used for superposition between 3K45 and 1U70

| Residue number (Fixed) | Residue name (Fixed) | Residue number (Rotated) | Residue name (Rotated) | Atom name | Deviation (Å) |
|---|---|---|---|---|---|
| 1A | V | 1A | V | CA | 0.597 |
| 2A | R | 2A | R | CA | 0.536 |
| 3A | P | 3A | P | CA | 0.457 |
| 4A | L | 4A | L | CA | 0.512 |
| 5A | N | 5A | N | CA | 0.159 |
| 6A | C | 6A | C | CA | 0.112 |
| 7A | I | 7A | I | CA | 0.130 |
| 8A | V | 8A | V | CA | 0.188 |
| 9A | A | 9A | A | CA | 0.257 |
| 10A | V | 10A | V | CA | 0.329 |
| 11A | S | 11A | S | CA | 0.111 |
| 12A | Q | 12A | Q | CA | 0.166 |
| 13A | N | 13A | N | CA | 0.061 |
| 14A | M | 14A | M | CA | 0.151 |
| 15A | G | 15A | G | CA | 0.332 |
| 16A | I | 16A | I | CA | 0.463 |
| 17A | G | 17A | G | CA | 1.343 |
| 18A | K | 18A | K | CA | 0.524 |
| 19A | N | 19A | N | CA | 0.564 |
| 20A | G | 20A | G | CA | 1.795 |
| 21A | D | 21A | D | CA | 0.981 |
| 22A | L | 22A | R | CA | 0.425 |
| 23A | P | 23A | P | CA | 0.662 |
| 24A | W | 24A | W | CA | 0.613 |
| 25A | P | 25A | P | CA | 0.744 |
| 26A | P | 26A | P | CA | 0.528 |
| 27A | L | 27A | L | CA | 0.363 |
| 28A | R | 28A | R | CA | 0.483 |
| 29A | N | 29A | N | CA | 0.175 |
| 30A | E | 30A | E | CA | 0.403 |
| 31A | F | 31A | F | CA | 0.377 |
| 32A | K | 32A | K | CA | 0.210 |
| 33A | Y | 33A | Y | CA | 0.255 |
| 34A | F | 34A | F | CA | 0.365 |
| 35A | Q | 35A | Q | CA | 0.326 |
| 36A | R | 36A | R | CA | 0.355 |
| 37A | M | 37A | M | CA | 0.457 |
| 38A | T | 38A | T | CA | 0.153 |
| 39A | T | 39A | T | CA | 0.274 |
| 40A | T | 40A | T | CA | 0.165 |
| 41A | S | 41A | S | CA | 0.170 |

| 42A | S | 42A | S | CA | 0.557 |
|---|---|---|---|---|---|
| 43A | V | 43A | V | CA | 0.303 |
| 44A | E | 44A | E | CA | 0.276 |
| 45A | G | 45A | G | CA | 0.261 |
| 46A | K | 46A | K | CA | 0.365 |
| 47A | Q | 47A | Q | CA | 0.271 |
| 48A | N | 48A | N | CA | 0.232 |
| 49A | L | 49A | L | CA | 0.171 |
| 50A | V | 50A | V | CA | 0.061 |
| 51A | I | 51A | I | CA | 0.136 |
| 52A | M | 52A | M | CA | 0.035 |
| 53A | G | 53A | G | CA | 0.137 |
| 54A | R | 54A | R | CA | 0.216 |
| 55A | K | 55A | K | CA | 0.210 |
| 56A | T | 56A | T | CA | 0.299 |
| 57A | W | 57A | W | CA | 0.162 |
| 58A | F | 58A | F | CA | 0.394 |
| 59A | S | 59A | S | CA | 0.476 |
| 60A | I | 60A | I | CA | 0.460 |
| 61A | P | 61A | P | CA | 0.457 |
| 62A | E | 62A | E | CA | 0.225 |
| 63A | K | 63A | K | CA | 0.450 |
| 64A | N | 64A | N | CA | 0.291 |
| 65A | R | 65A | R | CA | 0.233 |
| 66A | P | 66A | P | CA | 0.365 |
| 67A | L | 67A | L | CA | 0.423 |
| 68A | K | 68A | K | CA | 0.414 |
| 69A | D | 69A | D | CA | 0.354 |
| 70A | R | 70A | R | CA | 0.119 |
| 71A | I | 71A | I | CA | 0.154 |
| 72A | N | 72A | N | CA | 0.146 |
| 73A | I | 73A | I | CA | 0.091 |
| 74A | V | 74A | V | CA | 0.224 |
| 75A | L | 75A | L | CA | 0.238 |
| 76A | S | 76A | S | CA | 0.261 |
| 77A | R | 77A | R | CA | 0.449 |
| 78A | E | 78A | E | CA | 0.729 |
| 79A | L | 79A | L | CA | 0.337 |
| 80A | K | 80A | K | CA | 0.429 |
| 81A | E | 81A | E | CA | 0.804 |
| 82A | P | 82A | P | CA | 0.755 |
| 83A | P | 83A | P | CA | 0.666 |
| 84A | R | 84A | R | CA | 0.386 |
| 85A | G | 85A | G | CA | 0.156 |
| 86A | A | 86A | A | CA | 0.588 |
| 87A | H | 87A | H | CA | 0.238 |
| 88A | F | 88A | F | CA | 0.060 |
| 89A | L | 89A | L | CA | 0.194 |
| 90A | A | 90A | A | CA | 0.207 |
| 91A | K | 91A | K | CA | 0.251 |
| 92A | S | 92A | S | CA | 0.339 |
| 93A | L | 93A | L | CA | 0.063 |
| 94A | D | 94A | D | CA | 0.432 |

# Appendix part III

| 95A | D | 95A | D | CA | 0.558 |
|---|---|---|---|---|---|
| 96A | A | 96A | A | CA | 0.181 |
| 97A | L | 97A | L | CA | 0.344 |
| 98A | R | 98A | R | CA | 0.111 |
| 99A | L | 99A | L | CA | 0.222 |
| 100A | I | 100A | I | CA | 0.260 |
| 101A | E | 101A | E | CA | 0.270 |
| 102A | Q | 102A | Q | CA | 0.231 |
| 103A | P | 103A | P | CA | 0.138 |
| 104A | D | 104A | E | CA | 0.359 |
| 105A | L | 105A | L | CA | 0.765 |
| 106A | A | 106A | A | CA | 1.826 |
| 107A | S | 107A | S | CA | 1.763 |
| 108A | K | 108A | K | CA | 0.734 |
| 109A | V | 109A | V | CA | 0.319 |
| 110A | D | 110A | D | CA | 0.389 |
| 111A | M | 111A | M | CA | 0.100 |
| 112A | V | 112A | V | CA | 0.117 |
| 113A | W | 113A | W | CA | 0.031 |
| 114A | I | 114A | I | CA | 0.274 |
| 115A | V | 115A | V | CA | 0.284 |
| 116A | G | 116A | G | CA | 0.736 |
| 117A | G | 117A | G | CA | 1.555 |
| 118A | S | 118A | S | CA | 0.971 |
| 119A | S | 119A | S | CA | 0.975 |
| 120A | V | 120A | V | CA | 0.730 |
| 121A | Y | 121A | Y | CA | 0.481 |
| 122A | Q | 122A | Q | CA | 0.740 |
| 123A | E | 123A | E | CA | 0.694 |
| 124A | A | 124A | A | CA | 0.502 |
| 125A | M | 125A | M | CA | 0.875 |
| 126A | N | 126A | N | CA | 0.557 |
| 127A | Q | 127A | Q | CA | 1.454 |
| 128A | P | 128A | P | CA | 0.735 |
| 129A | G | 129A | G | CA | 0.761 |
| 130A | H | 130A | H | CA | 0.806 |
| 131A | L | 131A | L | CA | 0.542 |
| 132A | R | 132A | R | CA | 0.283 |
| 133A | L | 133A | L | CA | 0.315 |
| 134A | F | 134A | F | CA | 0.137 |
| 135A | V | 135A | V | CA | 0.189 |
| 136A | T | 136A | T | CA | 0.263 |
| 137A | R | 137A | R | CA | 0.309 |
| 138A | I | 138A | I | CA | 0.173 |
| 139A | M | 139A | M | CA | 0.121 |
| 140A | Q | 140A | Q | CA | 0.072 |
| 141A | E | 141A | E | CA | 0.185 |
| 142A | F | 142A | F | CA | 0.395 |
| 143A | E | 143A | E | CA | 0.619 |
| 144A | S | 144A | S | CA | 0.422 |
| 145A | D | 145A | D | CA | 0.556 |
| 146A | T | 146A | T | CA | 0.991 |
| 147A | F | 147A | F | CA | 0.792 |

# Appendix part III

| | | | | | |
|---|---|---|---|---|---|
| **148A** | **F** | **148A** | **F** | **CA** | **0.353** |
| **149A** | **P** | **149A** | **P** | **CA** | **0.340** |
| **150A** | **E** | **150A** | **E** | **CA** | **0.618** |
| **151A** | **I** | **151A** | **I** | **CA** | **0.409** |
| **152A** | **D** | **152A** | **D** | **CA** | **0.434** |
| **153A** | **L** | **153A** | **L** | **CA** | **0.306** |
| **154A** | **G** | **154A** | **G** | **CA** | **0.226** |
| **155A** | **K** | **155A** | **K** | **CA** | **0.425** |
| **156A** | **Y** | **156A** | **Y** | **CA** | **0.692** |
| **157A** | **K** | **157A** | **K** | **CA** | **0.294** |
| **158A** | **L** | **158A** | **L** | **CA** | **0.514** |
| **159A** | **L** | **159A** | **L** | **CA** | **0.396** |
| **160A** | **P** | **160A** | **P** | **CA** | **0.218** |
| **161A** | **E** | **161A** | **E** | **CA** | **0.397** |
| **162A** | **Y** | **162A** | **Y** | **CA** | **0.360** |
| **163A** | **P** | **163A** | **P** | **CA** | **0.285** |
| **164A** | **G** | **164A** | **G** | **CA** | **0.333** |
| **165A** | **V** | **165A** | **V** | **CA** | **0.300** |
| **166A** | **L** | **166A** | **L** | **CA** | **0.172** |
| **167A** | **S** | **167A** | **S** | **CA** | **0.201** |
| **168A** | **E** | **168A** | **E** | **CA** | **0.035** |
| **169A** | **V** | **169A** | **V** | **CA** | **0.098** |
| **170A** | **Q** | **170A** | **Q** | **CA** | **0.228** |
| **171A** | **E** | **171A** | **E** | **CA** | **0.242** |
| **172A** | **E** | **172A** | **E** | **CA** | **0.363** |
| **173A** | **K** | **173A** | **K** | **CA** | **0.333** |
| **174A** | **G** | **174A** | **G** | **CA** | **0.184** |
| **175A** | **I** | **175A** | **I** | **CA** | **0.193** |
| **176A** | **K** | **176A** | **K** | **CA** | **0.186** |
| **177A** | **Y** | **177A** | **Y** | **CA** | **0.263** |
| **178A** | **K** | **178A** | **K** | **CA** | **0.148** |
| **179A** | **F** | **179A** | **F** | **CA** | **0.141** |
| **180A** | **E** | **180A** | **E** | **CA** | **0.123** |
| **181A** | **V** | **181A** | **V** | **CA** | **0.205** |
| **182A** | **Y** | **182A** | **Y** | **CA** | **0.179** |
| **183A** | **E** | **183A** | **E** | **CA** | **0.238** |
| **184A** | **K** | **184A** | **K** | **CA** | **0.466** |
| **185A** | **K** | **185A** | **K** | **CA** | **0.630** |
| **186A** | **D** | **186A** | **D** | **CA** | **2.596** |

# Appendix part III

**Table 02.** Residues used for superposition between 3K45 and 1DR1

| Residue number (Fixed) | Residue name (Fixed) | Residue number (Rotated) | Residue name (Rotated) | Atom name | Deviation (Å) |
|---|---|---|---|---|---|
| 1A | V | 1A | V | CA | 5.148 |
| 2A | R | 2A | R | CA | 2.101 |
| 3A | P | 3A | S | CA | 0.844 |
| 4A | L | 4A | L | CA | 0.792 |
| 5A | N | 5A | N | CA | 0.568 |
| 6A | C | 6A | S | CA | 0.640 |
| 7A | I | 7A | I | CA | 0.263 |
| 8A | V | 8A | V | CA | 0.186 |
| 9A | A | 9A | A | CA | 0.263 |
| 10A | V | 10A | V | CA | 0.420 |
| 11A | S | 11A | C | CA | 0.309 |
| 12A | Q | 12A | Q | CA | 0.874 |
| 13A | N | 13A | N | CA | 1.066 |
| 14A | M | 14A | M | CA | 0.863 |
| 15A | G | 15A | G | CA | 0.905 |
| 16A | I | 16A | I | CA | 0.759 |
| 17A | G | 17A | G | CA | 1.181 |
| 18A | K | 18A | K | CA | 1.196 |
| 19A | N | 19A | D | CA | 1.529 |
| 20A | G | 20A | G | CA | 1.540 |
| 21A | D | 21A | N | CA | 1.114 |
| 22A | L | 22A | L | CA | 0.814 |
| 23A | P | 23A | P | CA | 0.868 |
| 24A | W | 24A | W | CA | 0.790 |
| 25A | P | 25A | P | CA | 0.798 |
| 26A | P | 26A | P | CA | 1.058 |
| 27A | L | 27A | L | CA | 0.921 |
| 28A | R | 28A | R | CA | 1.313 |
| 29A | N | 29A | N | CA | 0.817 |
| 30A | E | 30A | E | CA | 0.489 |
| 31A | F | 31A | Y | CA | 0.437 |
| 32A | K | 32A | K | CA | 0.122 |
| 33A | Y | 33A | Y | CA | 0.161 |
| 34A | F | 34A | F | CA | 0.153 |
| 35A | Q | 35A | Q | CA | 0.155 |
| 36A | R | 36A | R | CA | 0.175 |
| 37A | M | 37A | M | CA | 0.164 |
| 38A | T | 38A | T | CA | 0.098 |
| 39A | T | 39A | S | CA | 0.178 |
| 40A | T | 40A | T | CA | 0.215 |
| 41A | S | 41A | S | CA | 0.217 |
| 42A | S | 42A | H | CA | 0.788 |
| 43A | V | 43A | V | CA | 0.536 |
| 44A | E | 44A | E | CA | 0.884 |
| 45A | G | 45A | G | CA | 0.624 |
| 46A | K | 46A | K | CA | 0.279 |
| 47A | Q | 47A | Q | CA | 0.145 |
| 48A | N | 48A | N | CA | 0.385 |
| 49A | L | 49A | A | CA | 0.424 |

# Appendix part III

| 50A | V | 50A | V | CA | 0.125 |
|---|---|---|---|---|---|
| 51A | I | 51A | I | CA | 0.190 |
| 52A | M | 52A | M | CA | 0.237 |
| 53A | G | 53A | G | CA | 0.229 |
| 54A | R | 54A | K | CA | 0.569 |
| 55A | K | 55A | K | CA | 0.594 |
| 56A | T | 56A | T | CA | 0.342 |
| 57A | W | 57A | W | CA | 0.188 |
| 58A | F | 58A | F | CA | 0.543 |
| 59A | S | 59A | S | CA | 0.498 |
| 60A | I | 60A | I | CA | 0.432 |
| 61A | P | 61A | P | CA | 0.687 |
| 62A | E | 62A | E | CA | 0.803 |
| 63A | K | 63A | K | CA | 1.365 |
| 64A | N | 64A | N | CA | 1.009 |
| 65A | R | 65A | R | CA | 0.490 |
| 66A | P | 66A | P | CA | 0.252 |
| 67A | L | 67A | L | CA | 0.346 |
| 68A | K | 68A | K | CA | 0.150 |
| 69A | D | 69A | D | CA | 0.346 |
| 70A | R | 70A | R | CA | 0.280 |
| 71A | I | 71A | I | CA | 0.428 |
| 72A | N | 72A | N | CA | 0.429 |
| 73A | I | 73A | I | CA | 0.358 |
| 74A | V | 74A | V | CA | 0.447 |
| 75A | L | 75A | L | CA | 0.370 |
| 76A | S | 76A | S | CA | 0.635 |
| 77A | R | 77A | R | CA | 0.848 |
| 78A | E | 78A | E | CA | 0.982 |
| 79A | L | 79A | L | CA | 0.881 |
| 80A | K | 80A | K | CA | 0.640 |
| 81A | E | 81A | E | CA | 0.664 |
| 82A | P | 82A | A | CA | 0.574 |
| 83A | P | 83A | P | CA | 0.698 |
| 84A | R | 84A | K | CA | 0.732 |
| 85A | G | 85A | G | CA | 0.795 |
| 86A | A | 86A | A | CA | 0.548 |
| 87A | H | 87A | H | CA | 0.550 |
| 88A | F | 88A | Y | CA | 0.488 |
| 89A | L | 89A | L | CA | 0.472 |
| 90A | A | 90A | S | CA | 0.303 |
| 91A | K | 91A | K | CA | 0.776 |
| 92A | S | 92A | S | CA | 0.790 |
| 93A | L | 93A | L | CA | 0.726 |
| 94A | D | 94A | D | CA | 0.467 |
| 95A | D | 95A | D | CA | 0.292 |
| 96A | A | 96A | A | CA | 0.230 |
| 97A | L | 97A | L | CA | 0.482 |
| 98A | R | 98A | A | CA | 0.606 |
| 99A | L | 99A | L | CA | 0.549 |
| 100A | I | 100A | L | CA | 0.373 |
| 101A | E | 101A | D | CA | 0.853 |
| 102A | Q | 102A | S | CA | 1.257 |

# Appendix part III

| 103A | P | 103A | P | CA | 2.146 |
|---|---|---|---|---|---|
| 104A | D | 104A | E | CA | 1.674 |
| 105A | L | 105A | L | CA | 1.366 |
| 106A | A | 106A | K | CA | 2.125 |
| 107A | S | 107A | S | CA | 1.970 |
| 108A | K | 108A | K | CA | 0.459 |
| 109A | V | 109A | V | CA | 0.111 |
| 110A | D | 110A | D | CA | 0.271 |
| 111A | M | 111A | M | CA | 0.289 |
| 112A | V | 112A | V | CA | 0.221 |
| 113A | W | 113A | W | CA | 0.237 |
| 114A | I | 114A | I | CA | 0.453 |
| 115A | V | 115A | V | CA | 0.446 |
| 116A | G | 116A | G | CA | 1.128 |
| 117A | G | 117A | G | CA | 0.884 |
| 118A | S | 118A | T | CA | 1.305 |
| 119A | S | 119A | A | CA | 1.523 |
| 120A | V | 120A | V | CA | 0.813 |
| 121A | Y | 121A | Y | CA | 0.835 |
| 122A | Q | 122A | K | CA | 1.392 |
| 123A | E | 123A | A | CA | 1.232 |
| 124A | A | 124A | A | CA | 1.172 |
| 125A | M | 125A | M | CA | 1.405 |
| 126A | N | 126A | E | CA | 1.746 |
| 127A | Q | 127A | K | CA | 1.230 |
| 128A | P | 128A | P | CA | 1.191 |
| 129A | G | 129A | I | CA | 1.288 |
| 130A | H | 130A | N | CA | 0.880 |
| 131A | L | 131A | H | CA | 1.381 |
| 132A | R | 132A | R | CA | 1.163 |
| 133A | L | 133A | L | CA | 0.724 |
| 134A | F | 134A | F | CA | 0.423 |
| 135A | V | 135A | V | CA | 0.239 |
| 136A | T | 136A | T | CA | 0.226 |
| 137A | R | 137A | R | CA | 0.347 |
| 138A | I | 138A | I | CA | 0.426 |
| 139A | M | 139A | L | CA | 0.601 |
| 140A | Q | 140A | H | CA | 0.593 |
| 141A | E | 141A | E | CA | 0.950 |
| 142A | F | 142A | F | CA | 1.077 |
| 143A | E | 143A | E | CA | 1.016 |
| 144A | S | 144A | S | CA | 1.161 |
| 145A | D | 145A | D | CA | 0.867 |
| 146A | T | 146A | T | CA | 1.616 |
| 147A | F | 147A | F | CA | 1.158 |
| 148A | F | 148A | F | CA | 0.760 |
| 149A | P | 149A | P | CA | 1.082 |
| 150A | E | 150A | E | CA | 1.325 |
| 151A | I | 151A | I | CA | 1.901 |
| 152A | D | 152A | D | CA | 2.196 |
| 153A | L | 153A | Y | CA | 1.534 |
| 154A | G | 154A | K | CA | 1.056 |
| 155A | K | 155A | D | CA | 1.076 |

# Appendix part III

| | | | | | |
|---|---|---|---|---|---|
| **156A** | **Y** | **156A** | **F** | **CA** | **0.562** |
| **157A** | **K** | **157A** | **K** | **CA** | **0.706** |
| **158A** | **L** | **158A** | **L** | **CA** | **0.450** |
| **159A** | **L** | **159A** | **L** | **CA** | **0.829** |
| **160A** | **P** | **160A** | **T** | **CA** | **1.159** |
| **161A** | **E** | **161A** | **E** | **CA** | **1.221** |
| **162A** | **Y** | **162A** | **Y** | **CA** | **1.271** |
| **163A** | **P** | **163A** | **P** | **CA** | **1.421** |
| **164A** | **G** | **164A** | **G** | **CA** | **1.572** |
| **165A** | **V** | **165A** | **V** | **CA** | **1.189** |
| **166A** | **L** | **166A** | **P** | **CA** | **1.215** |
| **167A** | **S** | **167A** | **A** | **CA** | **1.261** |
| **168A** | **E** | **168A** | **D** | **CA** | **1.415** |
| **169A** | **V** | **169A** | **I** | **CA** | **0.990** |
| **170A** | **Q** | **170A** | **Q** | **CA** | **0.866** |
| **171A** | **E** | **171A** | **E** | **CA** | **0.842** |
| **172A** | **E** | **172A** | **E** | **CA** | **1.038** |
| **173A** | **K** | **173A** | **D** | **CA** | **0.566** |
| **174A** | **G** | **174A** | **G** | **CA** | **0.536** |
| **175A** | **I** | **175A** | **I** | **CA** | **0.777** |
| **176A** | **K** | **176A** | **Q** | **CA** | **0.484** |
| **177A** | **Y** | **177A** | **Y** | **CA** | **0.582** |
| **178A** | **K** | **178A** | **K** | **CA** | **1.026** |
| **179A** | **F** | **179A** | **F** | **CA** | **0.716** |
| **180A** | **E** | **180A** | **E** | **CA** | **0.628** |
| **181A** | **V** | **181A** | **V** | **CA** | **0.460** |
| **182A** | **Y** | **182A** | **Y** | **CA** | **0.289** |
| **183A** | **E** | **183A** | **Q** | **CA** | **0.611** |
| **184A** | **K** | **184A** | **K** | **CA** | **0.564** |
| **185A** | **K** | **185A** | **S** | **CA** | **1.014** |
| **186A** | **D** | **186A** | **V** | **CA** | **1.438** |

# Appendix part III

**Table 03.** Residues used for superposition between 3K45 and 4H95

| Residue number (Fixed) | Residue name (Fixed) | Residue number (Rotated) | Residue name (Rotated) | Atom name | Deviation (Å) |
|---|---|---|---|---|---|
| 1A | V | | | | |
| 2A | R | 4A | P | CA | 2.017 |
| 3A | P | 5A | N | CA | 2.358 |
| 4A | L | 6A | V | CA | 1.159 |
| 5A | N | 7A | A | CA | 0.702 |
| 6A | C | 8A | I | CA | 0.368 |
| 7A | I | 9A | I | CA | 0.262 |
| 8A | V | 10A | V | CA | 0.213 |
| 9A | A | 11A | A | CA | 0.307 |
| 10A | V | 12A | A | CA | 0.273 |
| 11A | S | 13A | L | CA | 0.597 |
| 12A | Q | 14A | K | CA | 1.557 |
| | | 15A | P | | |
| 13A | N | 16A | A | CA | 2.372 |
| 14A | M | 17A | L | CA | 0.961 |
| 15A | G | 18A | G | CA | 0.299 |
| 16A | I | 19A | I | CA | 0.583 |
| 17A | G | 20A | G | CA | 1.277 |
| 18A | K | 21A | Y | CA | 0.972 |
| 19A | N | 22A | K | CA | 1.078 |
| 20A | G | 23A | G | CA | 1.138 |
| 21A | D | 24A | K | CA | 0.942 |
| 22A | L | 25A | M | CA | 0.565 |
| 23A | P | 26A | P | CA | 1.925 |
| 24A | W | 27A | W | CA | 2.149 |
| 25A | P | | | | |
| 26A | P | 28A | R | CA | 0.988 |
| 27A | L | 29A | L | CA | 0.791 |
| 28A | R | 30A | R | CA | 1.353 |
| 29A | N | 31A | K | CA | 0.909 |
| 30A | E | 32A | E | CA | 0.696 |
| 31A | F | 33A | I | CA | 0.393 |
| 32A | K | 34A | R | CA | 0.457 |
| 33A | Y | 35A | Y | CA | 0.544 |
| 34A | F | 36A | F | CA | 0.462 |
| 35A | Q | 37A | K | CA | 0.415 |
| 36A | R | 38A | D | CA | 0.704 |
| 37A | M | 39A | V | CA | 0.634 |
| 38A | T | 40A | T | CA | 0.467 |
| 39A | T | 41A | T | CA | 0.709 |
| 40A | T | 42A | R | CA | 0.677 |
| 41A | S | 43A | T | CA | 1.165 |
| 42A | S | 44A | T | CA | 1.405 |
| 43A | V | 45A | K | CA | 1.622 |
| 44A | E | 46A | P | CA | 2.635 |
| 45A | G | 47A | N | CA | 2.169 |
| 46A | K | 48A | T | CA | 0.970 |
| 47A | Q | 49A | R | CA | 0.729 |

# Appendix part III

| | | | | | |
|---|---|---|---|---|---|
| 48A | N | 50A | N | CA | 0.550 |
| 49A | L | 51A | A | CA | 0.526 |
| 50A | V | 52A | V | CA | 0.331 |
| 51A | I | 53A | I | CA | 0.357 |
| 52A | M | 54A | M | CA | 0.255 |
| 53A | G | 55A | G | CA | 0.273 |
| 54A | R | 56A | R | CA | 0.294 |
| 55A | K | 57A | K | CA | 0.392 |
| 56A | T | 58A | T | CA | 0.414 |
| 57A | W | 59A | W | CA | 0.432 |
| 58A | F | 60A | E | CA | 0.572 |
| 59A | S | 61A | S | CA | 0.456 |
| 60A | I | 62A | I | CA | 0.469 |
| 61A | P | 63A | P | CA | 0.209 |
| 62A | E | 64A | Q | CA | 0.817 |
| 63A | K | 65A | K | CA | 0.680 |
| 64A | N | 66A | F | CA | 0.306 |
| 65A | R | 67A | R | CA | 0.519 |
| 66A | P | 68A | P | CA | 0.572 |
| 67A | L | 69A | L | CA | 0.707 |
| 68A | K | 70A | P | CA | 0.867 |
| 69A | D | 71A | D | CA | 0.315 |
| 70A | R | 72A | R | CA | 0.354 |
| 71A | I | 73A | L | CA | 0.469 |
| 72A | N | 74A | N | CA | 0.450 |
| 73A | I | 75A | I | CA | 0.534 |
| 74A | V | 76A | I | CA | 0.326 |
| 75A | L | 77A | L | CA | 0.325 |
| 76A | S | 78A | S | CA | 0.938 |
| 77A | R | 79A | R | CA | 1.874 |
| 78A | E | 80A | S | CA | 2.812 |
| 79A | L | 81A | Y | CA | 4.970 |
| 80A | K | 82A | E | CA | 5.738 |
| | | 83A | N | | |
| 81A | E | 84A | E | CA | 7.243 |
| 82A | P | 85A | I | CA | 7.050 |
| 83A | P | 86A | I | CA | 7.573 |
| 84A | R | 87A | D | CA | 8.414 |
| 85A | G | 88A | D | CA | 8.595 |
| 86A | A | 89A | N | CA | 5.265 |
| 87A | H | 90A | I | CA | 1.180 |
| 88A | F | 91A | I | CA | 0.369 |
| 89A | L | 92A | H | CA | 0.119 |
| 90A | A | 93A | A | CA | 0.067 |
| 91A | K | 94A | S | CA | 1.437 |
| 92A | S | 95A | S | CA | 2.022 |
| 93A | L | 96A | I | CA | 2.583 |
| 94A | D | 97A | E | CA | 2.137 |
| 95A | D | 98A | S | CA | 1.493 |
| 96A | A | 99A | S | CA | 1.697 |
| 97A | L | 100A | L | CA | 2.898 |
| 98A | R | 101A | N | CA | 2.794 |
| 99A | L | 102A | L | CA | 2.366 |

# Appendix part III

| | | | | | |
|---|---|---|---|---|---|
| 100A | I | 103A | V | CA | 2.950 |
| 101A | E | | | | |
| 102A | Q | | | | |
| 103A | P | | | | |
| 104A | D | | | | |
| 105A | L | | | | |
| 106A | A | 104A | S | CA | 4.044 |
| 107A | S | 105A | D | CA | 3.378 |
| 108A | K | | | | |
| 109A | V | 106A | V | CA | 1.363 |
| 110A | D | 107A | E | CA | 1.005 |
| 111A | M | 108A | R | CA | 0.706 |
| 112A | V | 109A | V | CA | 0.607 |
| 113A | W | 110A | F | CA | 0.495 |
| 114A | I | 111A | I | CA | 0.484 |
| 115A | V | 112A | I | CA | 0.355 |
| 116A | G | 113A | G | CA | 0.349 |
| 117A | G | 114A | G | CA | 0.420 |
| 118A | S | 115A | A | CA | 0.932 |
| 119A | S | 116A | E | CA | 1.079 |
| 120A | V | 117A | I | CA | 0.971 |
| 121A | Y | 118A | Y | CA | 1.138 |
| 122A | Q | 119A | N | CA | 1.786 |
| 123A | E | 120A | E | CA | 2.044 |
| 124A | A | 121A | L | CA | 1.391 |
| 125A | M | 122A | I | CA | 1.129 |
| 126A | N | 123A | N | CA | 1.760 |
| 127A | Q | 124A | N | CA | 1.109 |
| 128A | P | 125A | S | CA | 1.762 |
| 129A | G | 126A | L | CA | 4.429 |
| 130A | H | 127A | V | CA | 4.921 |
| 131A | L | 128A | S | CA | 1.301 |
| 132A | R | 129A | H | CA | 0.552 |
| 133A | L | 130A | L | CA | 0.517 |
| 134A | F | 131A | L | CA | 0.321 |
| 135A | V | 132A | I | CA | 0.276 |
| 136A | T | 133A | T | CA | 0.451 |
| 137A | R | 134A | E | CA | 0.692 |
| 138A | I | 135A | I | CA | 0.943 |
| 139A | M | 136A | E | CA | 0.459 |
| 140A | Q | 137A | H | CA | 2.027 |
| | | 138A | P | | |
| | | 139A | S | | |
| | | 140A | P | | |
| 141A | E | 141A | E | CA | 4.883 |
| | | 142A | S | | |
| 142A | F | 143A | I | CA | 3.227 |
| 143A | E | 144A | E | CA | 2.151 |
| 144A | S | 145A | M | CA | 1.399 |
| 145A | D | 146A | D | CA | 1.390 |
| 146A | T | 147A | T | CA | 0.746 |
| 147A | F | 148A | F | CA | 0.642 |
| 148A | F | 149A | L | CA | 0.986 |

# Appendix part III

| | | | | | |
|---|---|---|---|---|---|
| 149A | P | 150A | K | CA | 1.511 |
| 150A | E | 151A | F | CA | 2.314 |
| 151A | I | 152A | P | CA | 3.418 |
| 152A | D | 153A | L | CA | 2.453 |
| 153A | L | 154A | E | CA | 2.494 |
| 154A | G | | | | |
| 155A | K | 155A | S | CA | 2.148 |
| 156A | Y | 156A | W | CA | 1.104 |
| 157A | K | 157A | T | CA | 1.154 |
| 158A | L | 158A | K | CA | 0.491 |
| 159A | L | 159A | Q | CA | 0.493 |
| 160A | P | 160A | P | CA | 0.746 |
| 161A | E | 161A | K | CA | 1.788 |
| | | 162A | S | | |
| | | 163A | E | | |
| 162A | Y | 164A | L | CA | 1.444 |
| 163A | P | 165A | Q | CA | 1.911 |
| | | 166A | K | | |
| | | 167A | F | | |
| | | 168A | V | | |
| | | 169A | G | | |
| 164A | G | 170A | D | CA | 4.688 |
| | | 171A | T | | |
| | | 172A | V | | |
| 165A | V | 173A | L | CA | 4.194 |
| 166A | L | 174A | E | CA | 4.338 |
| 167A | S | | | | |
| 168A | E | 175A | D | CA | 1.261 |
| 169A | V | 176A | D | CA | 1.003 |
| 170A | Q | 177A | I | CA | 0.725 |
| 171A | E | 178A | K | CA | 1.039 |
| 172A | E | 179A | E | CA | 1.419 |
| 173A | K | 180A | G | CA | 2.560 |
| 174A | G | 181A | D | CA | 3.038 |
| 175A | I | 182A | F | CA | 1.481 |
| 176A | K | 183A | T | CA | 0.750 |
| 177A | Y | 184A | Y | CA | 0.663 |
| 178A | K | 185A | N | CA | 0.665 |
| 179A | F | 186A | Y | CA | 0.631 |
| 180A | E | 187A | T | CA | 0.697 |
| 181A | V | 188A | L | CA | 0.626 |
| 182A | Y | 189A | W | CA | 0.409 |
| 183A | E | 190A | T | CA | 0.912 |
| 184A | K | 191A | R | CA | 1.588 |
| 185A | K | 192A | K | CA | 6.345 |
| 186A | D | | | | |

# Appendix part III

**Table 04.** Residues used for superposition between 3K45 and 4GH8

| Residue number (Fixed) | Residue name (Fixed) | Residue number (Rotated) | Residue name (Rotated) | Atom name | Deviation (Å) |
|---|---|---|---|---|---|
| 1A | V | | | | |
| 2A | R | | | | |
| 3A | P | 1A | M | CA | 0.684 |
| 4A | L | 2A | I | CA | 0.533 |
| | | | | | |
| | | | | | |
| 5A | N | 3A | S | CA | 0.947 |
| 6A | C | 4A | L | CA | 0.422 |
| 7A | I | 5A | I | CA | 0.190 |
| 8A | V | 6A | A | CA | 0.162 |
| 9A | A | 7A | A | CA | 0.398 |
| 10A | V | 8A | L | CA | 0.218 |
| 11A | S | 9A | A | CA | 1.070 |
| 12A | Q | 10A | V | CA | 1.739 |
| 13A | N | 11A | D | CA | 1.022 |
| 14A | M | 12A | R | CA | 0.843 |
| 15A | G | 13A | V | CA | 0.374 |
| 16A | I | 14A | I | CA | 0.395 |
| 17A | G | 15A | G | CA | 0.315 |
| 18A | K | 16A | M | CA | 0.826 |
| 19A | N | 17A | E | CA | 1.246 |
| 20A | G | 18A | N | CA | 0.847 |
| 21A | D | 19A | A | CA | 0.284 |
| 22A | L | 20A | M | CA | 0.372 |
| 23A | P | 21A | P | CA | 1.485 |
| 24A | W | 22A | W | CA | 0.902 |
| 25A | P | 23A | P | CA | 0.984 |
| 26A | P | 24A | P | CA | 0.905 |
| 27A | L | 25A | L | CA | 0.694 |
| 28A | R | 26A | P | CA | 1.221 |
| 29A | N | 27A | A | CA | 1.389 |
| 30A | E | 28A | D | CA | 0.855 |
| 31A | F | 29A | L | CA | 0.619 |
| 32A | K | 30A | A | CA | 0.877 |
| 33A | Y | 31A | W | CA | 0.816 |
| 34A | F | 32A | F | CA | 0.337 |
| 35A | Q | 33A | K | CA | 0.843 |
| 36A | R | 34A | R | CA | 1.055 |
| 37A | M | 35A | N | CA | 0.935 |
| 38A | T | 36A | T | CA | 0.474 |
| 39A | T | | | | |
| 40A | T | 37A | L | CA | 3.724 |
| 41A | S | 38A | N | CA | 4.033 |
| 42A | S | | | | |
| 43A | V | | | | |
| 44A | E | | | | |
| 45A | G | | | | |
| 46A | K | | | | |
| 47A | Q | | | | |

# Appendix part III

| | | | | | |
|---|---|---|---|---|---|
| 48A | N | 39A | K | CA | 1.020 |
| 49A | L | 40A | P | CA | 0.560 |
| 50A | V | 41A | V | CA | 0.421 |
| 51A | I | 42A | I | CA | 0.381 |
| 52A | M | 43A | M | CA | 0.182 |
| 53A | G | 44A | G | CA | 0.222 |
| 54A | R | 45A | R | CA | 0.381 |
| 55A | K | 46A | H | CA | 0.622 |
| 56A | T | 47A | T | CA | 0.437 |
| 57A | W | 48A | W | CA | 0.278 |
| 58A | F | 49A | E | CA | 1.008 |
| 59A | S | 50A | S | CA | 0.776 |
| 60A | I | 51A | I | CA | 0.665 |
| 61A | P | 52A | P | CA | 0.984 |
| 62A | E | 53A | E | CA | 1.482 |
| 63A | K | 54A | K | CA | 2.072 |
| 64A | N | 55A | N | CA | 1.259 |
| 65A | R | 56A | R | CA | 0.745 |
| 66A | P | 57A | P | CA | 0.602 |
| 67A | L | 58A | L | CA | 0.325 |
| 68A | K | 59A | P | CA | 0.425 |
| 69A | D | 60A | G | CA | 0.652 |
| 70A | R | 61A | R | CA | 0.404 |
| 71A | I | 62A | K | CA | 0.504 |
| 72A | N | 63A | N | CA | 0.425 |
| 73A | I | 64A | I | CA | 0.315 |
| 74A | V | 65A | I | CA | 0.298 |
| 75A | L | 66A | L | CA | 0.303 |
| 76A | S | 67A | S | CA | 0.431 |
| 77A | R | 68A | S | CA | 0.360 |
| 78A | E | 69A | Q | CA | 1.410 |
| 79A | L | | | | |
| 80A | K | 70A | P | CA | 1.456 |
| 81A | E | | | | |
| 82A | P | 71A | G | CA | 3.467 |
| 83A | P | 72A | T | CA | 3.605 |
| 84A | R | 73A | D | CA | 2.215 |
| | | 74A | D | | |
| 85A | G | 75A | R | CA | 2.158 |
| 86A | A | 76A | V | CA | 1.094 |
| 87A | H | | | | |
| 88A | F | 77A | T | CA | 1.286 |
| 89A | L | 78A | W | CA | 1.056 |
| 90A | A | 79A | V | CA | 0.873 |
| 91A | K | 80A | K | CA | 0.512 |
| 92A | S | 81A | S | CA | 0.901 |
| 93A | L | 82A | V | CA | 1.208 |
| 94A | D | 83A | D | CA | 1.324 |
| 95A | D | 84A | E | CA | 1.274 |
| 96A | A | 85A | A | CA | 1.306 |
| 97A | L | 86A | I | CA | 0.922 |
| 98A | R | 87A | A | CA | 2.511 |
| 99A | L | 88A | A | CA | 3.109 |

# Appendix part III

| | | | | | |
|---|---|---|---|---|---|
| 100A | I | 89A | C | CA | 2.541 |
| 101A | E | 90A | G | CA | 3.596 |
| 102A | Q | | | | |
| 103A | P | | | | |
| 104A | D | | | | |
| 105A | L | | | | |
| 106A | A | 91A | D | CA | 4.319 |
| 107A | S | | | | |
| 108A | K | | | | |
| 109A | V | 92A | V | CA | 4.607 |
| 110A | D | 93A | P | CA | 4.728 |
| 111A | M | 94A | E | CA | 2.883 |
| 112A | V | 95A | I | CA | 1.188 |
| 113A | W | 96A | M | CA | 0.470 |
| 114A | I | 97A | V | CA | 0.398 |
| 115A | V | 98A | I | CA | 0.329 |
| 116A | G | 99A | G | CA | 0.610 |
| 117A | G | 100A | G | CA | 0.456 |
| 118A | S | 101A | G | CA | 0.427 |
| 119A | S | 102A | R | CA | 0.411 |
| 120A | V | 103A | V | CA | 0.371 |
| 121A | Y | 104A | Y | CA | 0.335 |
| 122A | Q | 105A | E | CA | 0.249 |
| 123A | E | 106A | Q | CA | 0.173 |
| 124A | A | 107A | F | CA | 0.154 |
| 125A | M | 108A | L | CA | 0.949 |
| 126A | N | 109A | P | CA | 2.871 |
| 127A | Q | | | | |
| 128A | P | | | | |
| 129A | G | 110A | K | CA | 5.057 |
| 130A | H | 111A | A | CA | 4.784 |
| 131A | L | 112A | Q | CA | 1.794 |
| 132A | R | 113A | K | CA | 0.819 |
| 133A | L | 114A | L | CA | 0.706 |
| 134A | F | 115A | Y | CA | 0.290 |
| 135A | V | 116A | L | CA | 0.614 |
| 136A | T | 117A | T | CA | 0.503 |
| 137A | R | 118A | H | CA | 0.698 |
| 138A | I | 119A | I | CA | 0.808 |
| 139A | M | 120A | D | CA | 2.153 |
| 140A | Q | 121A | A | CA | 2.892 |
| 141A | E | 122A | E | CA | 1.457 |
| 142A | F | 123A | V | CA | 1.505 |
| 143A | E | 124A | E | CA | 1.189 |
| 144A | S | 125A | G | CA | 0.995 |
| 145A | D | 126A | D | CA | 0.989 |
| 146A | T | 127A | T | CA | 0.936 |
| 147A | F | 128A | H | CA | 0.620 |
| 148A | F | 129A | F | CA | 0.592 |
| 149A | P | 130A | P | CA | 0.715 |
| 150A | E | 131A | D | CA | 1.077 |
| 151A | I | 132A | Y | CA | 1.459 |
| 152A | D | 133A | E | CA | 0.284 |

# Appendix part III

| | | | | | |
|---|---|---|---|---|---|
| 153A | L | 134A | P | CA | 0.505 |
| 154A | G | 135A | D | CA | 0.792 |
| 155A | K | 136A | D | CA | 0.778 |
| 156A | Y | 137A | W | CA | 1.106 |
| 157A | K | 138A | E | CA | 1.378 |
| 158A | L | 139A | S | CA | 1.250 |
| 159A | L | 140A | V | CA | 1.167 |
| 160A | P | | | | |
| 161A | E | 141A | F | CA | 3.043 |
| 162A | Y | | | | |
| 163A | P | | | | |
| 164A | G | | | | |
| 165A | V | | | | |
| 166A | L | | | | |
| 167A | S | 142A | S | CA | 3.527 |
| 168A | E | 143A | E | CA | 3.052 |
| 169A | V | 144A | F | CA | 1.348 |
| 170A | Q | 145A | H | CA | 1.192 |
| 171A | E | 146A | D | CA | 2.885 |
| 172A | E | 147A | A | CA | 3.522 |
| | | 148A | D | | |
| | | 149A | A | | |
| | | 150A | Q | | |
| | | 151A | N | | |
| 173A | K | 152A | S | CA | 1.503 |
| 174A | G | | | | |
| 175A | I | 153A | H | CA | 0.884 |
| 176A | K | 154A | S | CA | 0.851 |
| 177A | Y | 155A | Y | CA | 0.596 |
| 178A | K | 156A | C | CA | 0.718 |
| 179A | F | 157A | F | CA | 0.571 |
| 180A | E | 158A | E | CA | 0.569 |
| 181A | V | 159A | I | CA | 0.727 |
| 182A | Y | 160A | L | CA | 1.058 |
| 183A | E | 161A | E | CA | 1.273 |
| 184A | K | 162A | R | CA | 1.614 |
| 185A | K | | | | |
| 186A | D | | | | |

# Appendix part III

**Table 05.** Residues used for superposition between 3K45 and 1ZDR

| Residue number (Fixed) | Residue name (Fixed) | Residue number (Rotated) | Residue name (Rotated) | Atom name | Deviation (Å) |
|---|---|---|---|---|---|
| 1A | V | | | | |
| 2A | R | | | | |
| 3A | P | 1A | M | CA | 1.446 |
| 4A | L | 2A | I | CA | 0.991 |
| 5A | N | 3A | S | CA | 0.760 |
| 6A | C | 4A | H | CA | 0.363 |
| 7A | I | 5A | I | CA | 0.188 |
| 8A | V | 6A | V | CA | 0.154 |
| 9A | A | 7A | A | CA | 0.274 |
| 10A | V | 8A | M | CA | 0.554 |
| 11A | S | 9A | D | CA | 0.410 |
| 12A | Q | 10A | E | CA | 0.428 |
| 13A | N | 11A | N | CA | 0.817 |
| 14A | M | 12A | R | CA | 0.830 |
| 15A | G | 13A | V | CA | 0.384 |
| 16A | I | 14A | I | CA | 0.371 |
| 17A | G | 15A | G | CA | 0.792 |
| 18A | K | 16A | K | CA | 0.394 |
| 19A | N | 17A | D | CA | 0.569 |
| 20A | G | 18A | N | CA | 0.671 |
| 21A | D | 19A | R | CA | 0.810 |
| 22A | L | 20A | L | CA | 1.065 |
| 23A | P | 21A | P | CA | 2.178 |
| 24A | W | 22A | W | CA | 1.471 |
| 25A | P | | | | |
| 26A | P | 23A | H | CA | 1.968 |
| 27A | L | 24A | L | CA | 1.276 |
| 28A | R | 25A | P | CA | 1.484 |
| 29A | N | 26A | A | CA | 1.526 |
| 30A | E | 27A | D | CA | 0.796 |
| 31A | F | 28A | L | CA | 0.263 |
| 32A | K | 29A | A | CA | 0.416 |
| 33A | Y | 30A | Y | CA | 0.052 |
| 34A | F | 31A | F | CA | 0.141 |
| 35A | Q | 32A | K | CA | 0.338 |
| 36A | R | 33A | R | CA | 0.465 |
| 37A | M | 34A | V | CA | 0.509 |
| 38A | T | 35A | T | CA | 0.771 |
| 39A | T | 36A | M | CA | 0.973 |
| 40A | T | 37A | G | CA | 5.192 |
| 41A | S | | | | |
| 42A | S | | | | |
| 43A | V | | | | |
| 44A | E | | | | |
| 45A | G | | | | |
| 46A | K | | | | |
| 47A | Q | | | | |
| 48A | N | 38A | H | CA | 1.340 |
| 49A | L | 39A | A | CA | 0.705 |

# Appendix part III

| | | | | | |
|---|---|---|---|---|---|
| 50A | V | 40A | I | CA | 0.796 |
| 51A | I | 41A | V | CA | 0.574 |
| 52A | M | 42A | M | CA | 0.349 |
| 53A | G | 43A | G | CA | 0.380 |
| 54A | R | 44A | R | CA | 0.707 |
| 55A | K | 45A | K | CA | 0.714 |
| 56A | T | 46A | T | CA | 0.236 |
| 57A | W | 47A | F | CA | 0.467 |
| 58A | F | 48A | E | CA | 1.041 |
| 59A | S | 49A | A | CA | 0.458 |
| 60A | I | 50A | I | CA | 0.804 |
| 61A | P | | | | |
| 62A | E | 51A | G | CA | 3.653 |
| 63A | K | | | | |
| 64A | N | | | | |
| 65A | R | 52A | R | CA | 2.741 |
| 66A | P | 53A | P | CA | 1.364 |
| 67A | L | 54A | L | CA | 1.256 |
| 68A | K | 55A | P | CA | 1.234 |
| 69A | D | 56A | G | CA | 1.284 |
| 70A | R | 57A | R | CA | 1.415 |
| 71A | I | 58A | D | CA | 0.890 |
| 72A | N | 59A | N | CA | 0.804 |
| 73A | I | 60A | V | CA | 0.730 |
| 74A | V | 61A | V | CA | 0.435 |
| 75A | L | 62A | V | CA | 0.727 |
| 76A | S | 63A | T | CA | 0.557 |
| 77A | R | 64A | G | CA | 1.083 |
| 78A | E | 65A | N | CA | 1.897 |
| 79A | L | 66A | R | CA | 3.934 |
| 80A | K | 67A | S | CA | 2.886 |
| 81A | E | 68A | F | CA | 4.654 |
| 82A | P | 69A | R | CA | 2.432 |
| 83A | P | 70A | P | CA | 1.266 |
| 84A | R | 71A | E | CA | 1.973 |
| 85A | G | 72A | G | CA | 2.435 |
| 86A | A | 73A | C | CA | 1.420 |
| 87A | H | | | | |
| 88A | F | 74A | L | CA | 0.742 |
| 89A | L | 75A | V | CA | 0.556 |
| 90A | A | 76A | L | CA | 0.424 |
| 91A | K | 77A | H | CA | 1.213 |
| 92A | S | 78A | S | CA | 3.326 |
| 93A | L | 79A | L | CA | 5.193 |
| 94A | D | 80A | E | CA | 4.835 |
| 95A | D | 81A | E | CA | 1.817 |
| 96A | A | 82A | V | CA | 2.483 |
| 97A | L | 83A | K | CA | 3.752 |
| 98A | R | 84A | Q | CA | 1.373 |
| 99A | L | 85A | W | CA | 1.891 |
| 100A | I | 86A | I | CA | 4.340 |
| 101A | E | 87A | A | CA | 3.660 |
| 102A | Q | 88A | S | CA | 3.836 |

# Appendix part III

| | | | | | |
|---|---|---|---|---|---|
| 103A | P | | | | |
| 104A | D | | | | |
| 105A | L | 89A | R | CA | 7.323 |
| 106A | A | 90A | A | CA | 6.126 |
| 107A | S | | | | |
| 108A | K | | | | |
| 109A | V | | | | |
| 110A | D | 91A | D | CA | 2.147 |
| 111A | M | 92A | E | CA | 1.644 |
| 112A | V | 93A | V | CA | 0.467 |
| 113A | W | 94A | F | CA | 0.506 |
| 114A | I | 95A | I | CA | 0.386 |
| 115A | V | 96A | I | CA | 0.495 |
| 116A | G | 97A | G | CA | 0.293 |
| 117A | G | 98A | G | CA | 1.242 |
| 118A | S | 99A | A | CA | 0.698 |
| 119A | S | 100A | E | CA | 0.520 |
| 120A | V | 101A | L | CA | 0.714 |
| 121A | Y | 102A | F | CA | 0.531 |
| 122A | Q | 103A | R | CA | 0.503 |
| 123A | E | 104A | A | CA | 0.754 |
| 124A | A | 105A | T | CA | 1.066 |
| 125A | M | 106A | M | CA | 0.924 |
| 126A | N | 107A | P | CA | 3.230 |
| 127A | Q | | | | |
| 128A | P | | | | |
| 129A | G | 108A | I | CA | 4.967 |
| 130A | H | 109A | V | CA | 4.938 |
| 131A | L | 110A | D | CA | 1.254 |
| 132A | R | 111A | R | CA | 0.414 |
| 133A | L | 112A | L | CA | 0.494 |
| 134A | F | 113A | Y | CA | 0.175 |
| 135A | V | 114A | V | CA | 0.135 |
| 136A | T | 115A | T | CA | 0.144 |
| 137A | R | 116A | K | CA | 0.340 |
| 138A | I | 117A | I | CA | 0.520 |
| 139A | M | 118A | F | CA | 0.551 |
| 140A | Q | 119A | A | CA | 0.295 |
| 141A | E | 120A | S | CA | 0.440 |
| 142A | F | 121A | F | CA | 0.593 |
| 143A | E | 122A | P | CA | 0.942 |
| 144A | S | 123A | G | CA | 0.318 |
| 145A | D | 124A | D | CA | 0.141 |
| 146A | T | 125A | T | CA | 0.799 |
| 147A | F | 126A | F | CA | 0.731 |
| 148A | F | 127A | Y | CA | 0.599 |
| 149A | P | 128A | P | CA | 0.210 |
| 150A | E | 129A | P | CA | 0.330 |
| 151A | I | 130A | I | CA | 0.501 |
| 152A | D | 131A | S | CA | 0.430 |
| 153A | L | 132A | D | CA | 0.948 |
| 154A | G | 133A | D | CA | 0.808 |
| 155A | K | 134A | E | CA | 0.023 |

| | | | | | |
|---|---|---|---|---|---|
| 156A | Y | 135A | W | CA | 0.477 |
| 157A | K | 136A | E | CA | 1.138 |
| 158A | L | 137A | I | CA | 0.873 |
| 159A | L | 138A | V | CA | 0.833 |
| 160A | P | | | | |
| 161A | E | 139A | S | CA | 2.775 |
| 162A | Y | | | | |
| 163A | P | | | | |
| 164A | G | | | | |
| 165A | V | 140A | Y | CA | 7.503 |
| 166A | L | 141A | T | CA | 4.085 |
| 167A | S | | | | |
| 168A | E | 142A | P | CA | 2.480 |
| 169A | V | 143A | G | CA | 3.115 |
| 170A | Q | 144A | G | CA | 2.023 |
| 171A | E | 145A | K | CA | 1.427 |
| 172A | E | 146A | D | CA | 4.237 |
| | | 147A | E | | |
| | | 148A | K | | |
| 173A | K | 149A | N | CA | 3.571 |
| 174A | G | 150A | P | CA | 2.095 |
| 175A | I | 151A | Y | CA | 1.121 |
| 176A | K | 152A | E | CA | 1.290 |
| 177A | Y | 153A | H | CA | 0.774 |
| 178A | K | 154A | A | CA | 0.590 |
| 179A | F | 155A | F | CA | 0.320 |
| 180A | E | 156A | I | CA | 0.226 |
| 181A | V | 157A | I | CA | 0.316 |
| 182A | Y | 158A | Y | CA | 0.443 |
| 183A | E | 159A | E | CA | 0.780 |
| 184A | K | 160A | R | CA | 1.737 |
| 185A | K | 161A | K | CA | 6.138 |
| 186A | D | | | | |

# Appendix part III

**Table 06.** Residues used for superposition between 3K45 and 3JW3

| Residue number (Fixed) | Residue name (Fixed) | Residue number (Rotated) | Residue name (Rotated) | Atom name | Deviation (Å) |
|---|---|---|---|---|---|
| | | -2A | H | | |
| | | -1A | H | | |
| | | 0A | H | | |
| 1A | V | 1A | M | CA | 2.845 |
| 2A | R | | | | |
| 3A | P | 2A | R | CA | 1.056 |
| 4A | L | 3A | V | CA | 0.262 |
| 5A | N | 4A | S | CA | 0.675 |
| 6A | C | 5A | F | CA | 0.699 |
| 7A | I | 6A | M | CA | 0.228 |
| 8A | V | 7A | V | CA | 0.445 |
| 9A | A | 8A | A | CA | 0.164 |
| 10A | V | 9A | M | CA | 0.463 |
| 11A | S | 10A | D | CA | 0.173 |
| 12A | Q | 11A | E | CA | 0.316 |
| 13A | N | 12A | N | CA | 0.832 |
| 14A | M | 13A | R | CA | 0.750 |
| 15A | G | 14A | V | CA | 0.449 |
| 16A | I | 15A | I | CA | 0.430 |
| 17A | G | 16A | G | CA | 0.737 |
| 18A | K | 17A | K | CA | 0.594 |
| 19A | N | 18A | D | CA | 0.627 |
| 20A | G | 19A | N | CA | 0.691 |
| 21A | D | 20A | N | CA | 0.636 |
| 22A | L | 21A | L | CA | 1.094 |
| 23A | P | 22A | P | CA | 1.815 |
| 24A | W | 23A | W | CA | 1.753 |
| 25A | P | | | | |
| 26A | P | 24A | R | CA | 1.245 |
| 27A | L | 25A | L | CA | 0.129 |
| 28A | R | 26A | P | CA | 0.412 |
| 29A | N | 27A | S | CA | 0.464 |
| 30A | E | 28A | E | CA | 0.447 |
| 31A | F | 29A | L | CA | 0.511 |
| 32A | K | 30A | Q | CA | 0.463 |
| 33A | Y | 31A | Y | CA | 0.187 |
| 34A | F | 32A | V | CA | 0.441 |
| 35A | Q | 33A | K | CA | 0.441 |
| 36A | R | 34A | K | CA | 0.345 |
| 37A | M | 35A | T | CA | 0.425 |
| 38A | T | 36A | T | CA | 0.796 |
| 39A | T | | | | |
| 40A | T | 37A | M | CA | 3.022 |
| 41A | S | 38A | G | CA | 4.369 |
| 42A | S | | | | |
| 43A | V | | | | |
| 44A | E | | | | |
| 45A | G | | | | |
| 46A | K | | | | |

# Appendix part III

| | | | | | |
|---|---|---|---|---|---|
| 47A | Q | | | | |
| 48A | N | 39A | H | CA | 1.230 |
| 49A | L | 40A | P | CA | 0.811 |
| 50A | V | 41A | L | CA | 0.800 |
| 51A | I | 42A | I | CA | 0.842 |
| 52A | M | 43A | M | CA | 0.629 |
| 53A | G | 44A | G | CA | 0.592 |
| 54A | R | 45A | R | CA | 0.540 |
| 55A | K | 46A | K | CA | 0.893 |
| 56A | T | 47A | N | CA | 0.949 |
| 57A | W | 48A | Y | CA | 0.934 |
| 58A | F | 49A | E | CA | 1.052 |
| 59A | S | 50A | A | CA | 1.036 |
| 60A | I | 51A | I | CA | 1.177 |
| 61A | P | | | | |
| 62A | E | 52A | G | CA | 3.985 |
| 63A | K | | | | |
| 64A | N | | | | |
| 65A | R | 53A | R | CA | 2.688 |
| 66A | P | 54A | P | CA | 1.396 |
| 67A | L | 55A | L | CA | 1.318 |
| 68A | K | 56A | P | CA | 0.968 |
| 69A | D | 57A | G | CA | 0.334 |
| 70A | R | 58A | R | CA | 0.569 |
| 71A | I | 59A | R | CA | 0.776 |
| 72A | N | 60A | N | CA | 0.888 |
| 73A | I | 61A | I | CA | 0.902 |
| 74A | V | 62A | I | CA | 0.709 |
| 75A | L | 63A | V | CA | 0.642 |
| 76A | S | 64A | T | CA | 0.866 |
| 77A | R | 65A | R | CA | 0.591 |
| 78A | E | 66A | N | CA | 1.020 |
| | | 67A | E | | |
| 79A | L | 68A | G | CA | 3.625 |
| 80A | K | 69A | Y | CA | 3.826 |
| 81A | E | 70A | H | CA | 3.282 |
| 82A | P | | | | |
| 83A | P | 71A | V | CA | 1.359 |
| 84A | R | 72A | E | CA | 1.980 |
| 85A | G | 73A | G | CA | 2.439 |
| 86A | A | 74A | C | CA | 0.739 |
| 87A | H | | | | |
| 88A | F | 75A | E | CA | 1.397 |
| 89A | L | 76A | V | CA | 0.739 |
| 90A | A | 77A | A | CA | 0.705 |
| 91A | K | 78A | H | CA | 0.705 |
| 92A | S | 79A | S | CA | 0.770 |
| 93A | L | 80A | V | CA | 1.674 |
| 94A | D | 81A | E | CA | 1.627 |
| 95A | D | 82A | E | CA | 0.979 |
| 96A | A | 83A | V | CA | 0.730 |
| 97A | L | 84A | F | CA | 0.959 |
| 98A | R | 85A | E | CA | 2.025 |

# Appendix part III

| | | | | | |
|---|---|---|---|---|---|
| 99A | L | 86A | L | CA | 2.135 |
| 100A | I | 87A | C | CA | 1.126 |
| 101A | E | | | | |
| 102A | Q | 88A | K | CA | 4.236 |
| 103A | P | | | | |
| 104A | D | | | | |
| 105A | L | | | | |
| 106A | A | 89A | N | CA | 3.425 |
| 107A | S | | | | |
| 108A | K | | | | |
| 109A | V | 90A | E | CA | 3.870 |
| 110A | D | 91A | E | CA | 3.729 |
| 111A | M | 92A | E | CA | 2.445 |
| 112A | V | 93A | I | CA | 0.815 |
| 113A | W | 94A | F | CA | 0.678 |
| 114A | I | 95A | I | CA | 0.700 |
| 115A | V | 96A | I | CA | 0.818 |
| 116A | G | 97A | G | CA | 0.939 |
| 117A | G | 98A | G | CA | 0.717 |
| 118A | S | 99A | A | CA | 0.739 |
| 119A | S | 100A | Q | CA | 0.624 |
| 120A | V | 101A | I | CA | 0.723 |
| 121A | Y | 102A | Y | CA | 0.857 |
| 122A | Q | 103A | D | CA | 1.020 |
| 123A | E | 104A | L | CA | 1.171 |
| 124A | A | 105A | F | CA | 0.950 |
| 125A | M | 106A | L | CA | 0.822 |
| 126A | N | 107A | P | CA | 2.387 |
| 127A | Q | | | | |
| 128A | P | | | | |
| 129A | G | 108A | Y | CA | 4.978 |
| 130A | H | 109A | V | CA | 5.373 |
| 131A | L | 110A | D | CA | 1.419 |
| 132A | R | 111A | K | CA | 0.784 |
| 133A | L | 112A | L | CA | 1.061 |
| 134A | F | 113A | Y | CA | 0.257 |
| 135A | V | 114A | I | CA | 0.218 |
| 136A | T | 115A | T | CA | 0.241 |
| 137A | R | 116A | K | CA | 0.391 |
| 138A | I | 117A | I | CA | 0.590 |
| 139A | M | 118A | H | CA | 1.051 |
| 140A | Q | 119A | H | CA | 1.054 |
| 141A | E | 120A | A | CA | 0.726 |
| 142A | F | 121A | F | CA | 0.540 |
| 143A | E | 122A | E | CA | 0.794 |
| 144A | S | 123A | G | CA | 1.006 |
| 145A | D | 124A | D | CA | 0.135 |
| 146A | T | 125A | T | CA | 1.049 |
| 147A | F | 126A | F | CA | 0.802 |
| 148A | F | 127A | F | CA | 0.492 |
| 149A | P | 128A | P | CA | 0.853 |
| 150A | E | 129A | E | CA | 1.117 |
| 151A | I | 130A | M | CA | 2.847 |

# Appendix part III

| | | | | | |
|---|---|---|---|---|---|
| 152A | D | 131A | D | CA | 2.080 |
| 153A | L | 132A | M | CA | 2.056 |
| 154A | G | 133A | T | CA | 0.931 |
| 155A | K | 134A | N | CA | 3.787 |
| 156A | Y | 135A | W | CA | 2.262 |
| 157A | K | 136A | K | CA | 1.835 |
| 158A | L | 137A | E | CA | 1.539 |
| 159A | L | 138A | V | CA | 1.597 |
| 160A | P | | | | |
| 161A | E | 139A | F | CA | 2.940 |
| 162A | Y | | | | |
| 163A | P | | | | |
| 164A | G | | | | |
| 165A | V | | | | |
| 166A | L | | | | |
| 167A | S | 140A | V | CA | 3.249 |
| 168A | E | 141A | E | CA | 3.133 |
| 169A | V | 142A | K | CA | 1.992 |
| 170A | Q | 143A | G | CA | 2.484 |
| | | 144A | L | | |
| 171A | E | 145A | T | CA | 1.766 |
| 172A | E | 146A | D | CA | 3.854 |
| | | 147A | E | | |
| | | 148A | K | | |
| | | 149A | N | | |
| 173A | K | 150A | P | CA | 1.542 |
| 174A | G | | | | |
| 175A | I | 151A | Y | CA | 0.410 |
| 176A | K | 152A | T | CA | 0.338 |
| 177A | Y | 153A | Y | CA | 0.338 |
| 178A | K | 154A | Y | CA | 0.515 |
| 179A | F | 155A | Y | CA | 0.534 |
| 180A | E | 156A | H | CA | 0.601 |
| 181A | V | 157A | V | CA | 0.755 |
| 182A | Y | 158A | Y | CA | 1.247 |
| 183A | E | 159A | E | CA | 1.764 |
| 184A | K | 160A | K | CA | 1.956 |
| 185A | K | 161A | Q | CA | 4.283 |
| 186A | D | 162A | Q | CA | 5.253 |

# Appendix part III

**Table 07.** Residues used for superposition between 3K45 and 2QK8

| Residue number (Fixed) | Residue name (Fixed) | Residue number (Rotated) | Residue name (Rotated) | Atom name | Deviation (Å) |
|---|---|---|---|---|---|
| 1A | V | | | | |
| 2A | R | 1A | M | CA | 0.279 |
| 3A | P | 2A | I | CA | 0.724 |
| 4A | L | 3A | V | CA | 0.589 |
| 5A | N | 4A | S | CA | 1.011 |
| 6A | C | 5A | F | CA | 0.875 |
| 7A | I | 6A | M | CA | 0.637 |
| 8A | V | 7A | V | CA | 0.520 |
| 9A | A | 8A | A | CA | 0.345 |
| 10A | V | 9A | M | CA | 0.787 |
| 11A | S | 10A | D | CA | 0.459 |
| 12A | Q | 11A | E | CA | 0.412 |
| 13A | N | 12A | N | CA | 0.548 |
| 14A | M | 13A | R | CA | 0.645 |
| 15A | G | 14A | V | CA | 0.550 |
| 16A | I | 15A | I | CA | 0.418 |
| 17A | G | 16A | G | CA | 0.588 |
| 18A | K | 17A | K | CA | 0.502 |
| 19A | N | 18A | D | CA | 0.287 |
| 20A | G | 19A | N | CA | 0.261 |
| 21A | D | 20A | N | CA | 1.000 |
| 22A | L | 21A | L | CA | 0.891 |
| 23A | P | 22A | P | CA | 1.295 |
| 24A | W | 23A | W | CA | 1.664 |
| 25A | P | | | | |
| 26A | P | 24A | R | CA | 1.070 |
| 27A | L | 25A | L | CA | 0.336 |
| 28A | R | 26A | P | CA | 0.202 |
| 29A | N | 27A | S | CA | 0.259 |
| 30A | E | 28A | E | CA | 0.373 |
| 31A | F | 29A | L | CA | 0.453 |
| 32A | K | 30A | Q | CA | 0.351 |
| 33A | Y | 31A | Y | CA | 0.448 |
| 34A | F | 32A | V | CA | 0.438 |
| 35A | Q | 33A | K | CA | 0.413 |
| 36A | R | 34A | K | CA | 0.393 |
| 37A | M | 35A | T | CA | 0.647 |
| 38A | T | 36A | T | CA | 0.485 |
| 39A | T | | | | |
| 40A | T | 37A | M | CA | 3.620 |
| 41A | S | 38A | G | CA | 4.438 |
| 42A | S | | | | |
| 43A | V | | | | |
| 44A | E | | | | |
| 45A | G | | | | |
| 46A | K | | | | |
| 47A | Q | | | | |
| 48A | N | 39A | H | CA | 1.359 |
| 49A | L | 40A | P | CA | 1.038 |

# Appendix part III

| | | | | | |
|---|---|---|---|---|---|
| 50A | V | 41A | L | CA | 1.025 |
| 51A | I | 42A | I | CA | 0.785 |
| 52A | M | 43A | M | CA | 0.635 |
| 53A | G | 44A | G | CA | 0.528 |
| 54A | R | 45A | R | CA | 0.972 |
| 55A | K | 46A | K | CA | 0.961 |
| 56A | T | 47A | N | CA | 0.845 |
| 57A | W | 48A | Y | CA | 1.152 |
| 58A | F | 49A | E | CA | 1.696 |
| 59A | S | 50A | A | CA | 1.502 |
| 60A | I | 51A | I | CA | 1.174 |
| 61A | P | 52A | G | CA | 2.894 |
| 62A | E | | | | |
| 63A | K | | | | |
| 64A | N | | | | |
| 65A | R | 53A | R | CA | 2.452 |
| 66A | P | 54A | P | CA | 1.418 |
| 67A | L | 55A | L | CA | 0.664 |
| 68A | K | 56A | P | CA | 0.183 |
| 69A | D | 57A | G | CA | 0.757 |
| 70A | R | 58A | R | CA | 0.673 |
| 71A | I | 59A | R | CA | 0.990 |
| 72A | N | 60A | N | CA | 1.162 |
| 73A | I | 61A | I | CA | 0.811 |
| 74A | V | 62A | I | CA | 0.808 |
| 75A | L | 63A | V | CA | 0.613 |
| 76A | S | 64A | T | CA | 0.982 |
| 77A | R | 65A | R | CA | 1.398 |
| 78A | E | 66A | N | CA | 1.715 |
| | | 67A | E | | |
| 79A | L | 68A | G | CA | 3.373 |
| 80A | K | 69A | Y | CA | 4.015 |
| 81A | E | 70A | H | CA | 3.673 |
| 82A | P | | | | |
| 83A | P | 71A | V | CA | 1.227 |
| 84A | R | 72A | E | CA | 1.495 |
| 85A | G | 73A | G | CA | 1.618 |
| 86A | A | 74A | C | CA | 0.732 |
| 87A | H | | | | |
| 88A | F | 75A | E | CA | 1.774 |
| 89A | L | 76A | V | CA | 1.018 |
| 90A | A | 77A | A | CA | 0.773 |
| 91A | K | 78A | H | CA | 0.633 |
| 92A | S | 79A | S | CA | 1.328 |
| 93A | L | 80A | V | CA | 1.793 |
| 94A | D | 81A | E | CA | 2.192 |
| 95A | D | 82A | E | CA | 1.677 |
| 96A | A | 83A | V | CA | 0.843 |
| 97A | L | 84A | F | CA | 1.380 |
| 98A | R | 85A | E | CA | 2.689 |
| 99A | L | 86A | L | CA | 2.675 |
| 100A | I | 87A | C | CA | 1.268 |
| 101A | E | | | | |

# Appendix part III

| | | | | | |
|---|---|---|---|---|---|
| 102A | Q | 88A | K | CA | 3.816 |
| 103A | P | | | | |
| 104A | D | | | | |
| 105A | L | | | | |
| 106A | A | 89A | N | CA | 3.023 |
| 107A | S | | | | |
| 108A | K | | | | |
| 109A | V | 90A | E | CA | 4.028 |
| 110A | D | 91A | E | CA | 3.614 |
| 111A | M | 92A | E | CA | 2.612 |
| 112A | V | 93A | I | CA | 1.391 |
| 113A | W | 94A | F | CA | 0.886 |
| 114A | I | 95A | I | CA | 0.910 |
| 115A | V | 96A | F | CA | 0.562 |
| 116A | G | 97A | G | CA | 1.074 |
| 117A | G | 98A | G | CA | 1.054 |
| 118A | S | 99A | A | CA | 0.892 |
| 119A | S | 100A | Q | CA | 0.601 |
| 120A | V | 101A | I | CA | 0.581 |
| 121A | Y | 102A | Y | CA | 0.982 |
| 122A | Q | 103A | D | CA | 1.442 |
| 123A | E | 104A | L | CA | 1.481 |
| 124A | A | 105A | F | CA | 1.286 |
| 125A | M | 106A | L | CA | 1.606 |
| 126A | N | | | | |
| 127A | Q | 107A | P | CA | 1.513 |
| 128A | P | | | | |
| 129A | G | 108A | Y | CA | 4.731 |
| 130A | H | 109A | V | CA | 4.928 |
| 131A | L | 110A | D | CA | 1.642 |
| 132A | R | 111A | K | CA | 0.968 |
| 133A | L | 112A | L | CA | 0.983 |
| 134A | F | 113A | Y | CA | 0.369 |
| 135A | V | 114A | I | CA | 0.276 |
| 136A | T | 115A | T | CA | 0.409 |
| 137A | R | 116A | K | CA | 0.517 |
| 138A | I | 117A | I | CA | 0.445 |
| 139A | M | 118A | H | CA | 1.127 |
| 140A | Q | 119A | H | CA | 1.328 |
| 141A | E | 120A | A | CA | 1.187 |
| 142A | F | 121A | F | CA | 1.326 |
| 143A | E | 122A | E | CA | 1.613 |
| 144A | S | 123A | G | CA | 2.113 |
| 145A | D | 124A | D | CA | 0.486 |
| 146A | T | 125A | T | CA | 0.814 |
| 147A | F | 126A | F | CA | 0.640 |
| 148A | F | 127A | F | CA | 0.408 |
| 149A | P | 128A | P | CA | 1.237 |
| 150A | E | 129A | E | CA | 1.339 |
| 151A | I | 130A | M | CA | 3.000 |
| 152A | D | 131A | D | CA | 1.158 |
| 153A | L | 132A | M | CA | 1.798 |
| 154A | G | 133A | T | CA | 1.384 |

| | | | | | |
|---|---|---|---|---|---|
| 155A | K | 134A | N | CA | 2.798 |
| 156A | Y | 135A | W | CA | 2.060 |
| 157A | K | 136A | K | CA | 1.675 |
| 158A | L | 137A | E | CA | 1.800 |
| 159A | L | 138A | V | CA | 1.695 |
| 160A | P | | | | |
| 161A | E | 139A | F | CA | 2.574 |
| 162A | Y | | | | |
| 163A | P | | | | |
| 164A | G | | | | |
| 165A | V | | | | |
| 166A | L | | | | |
| 167A | S | 140A | V | CA | 3.450 |
| 168A | E | 141A | E | CA | 3.459 |
| 169A | V | 142A | K | CA | 2.391 |
| 170A | Q | 143A | G | CA | 2.382 |
| | | 144A | L | | |
| 171A | E | 145A | T | CA | 2.139 |
| 172A | E | 146A | D | CA | 3.644 |
| | | 147A | E | | |
| 173A | K | 148A | K | CA | 3.972 |
| | | 149A | N | | |
| 174A | G | 150A | P | CA | 2.764 |
| 175A | I | 151A | Y | CA | 0.954 |
| 176A | K | 152A | T | CA | 0.970 |
| 177A | Y | 153A | Y | CA | 0.489 |
| 178A | K | 154A | Y | CA | 0.873 |
| 179A | F | 155A | Y | CA | 0.443 |
| 180A | E | 156A | H | CA | 0.452 |
| 181A | V | 157A | V | CA | 0.470 |
| 182A | Y | 158A | Y | CA | 1.375 |
| 183A | E | 159A | E | CA | 1.493 |
| 184A | K | 160A | K | CA | 1.490 |
| 185A | K | 161A | Q | CA | 2.660 |
| 186A | D | 162A | Q | CA | 4.933 |

# Appendix part III

**Table 08.** Residues used for superposition between 3K45 and 2ZZA

| Residue number (Fixed) | Residue name (Fixed) | Residue number (Rotated) | Residue name (Rotated) | Atom name | Deviation (Å) |
|---|---|---|---|---|---|
| 1A | V | | | | |
| 2A | R | 1A | V | CA | 0.574 |
| 3A | P | 2A | I | CA | 0.517 |
| 4A | L | 3A | V | CA | 0.991 |
| 5A | N | 4A | S | CA | 1.286 |
| 6A | C | 5A | M | CA | 0.564 |
| 7A | I | 6A | I | CA | 0.513 |
| 8A | V | 7A | A | CA | 0.135 |
| 9A | A | 8A | A | CA | 0.152 |
| 10A | V | 9A | L | CA | 0.219 |
| 11A | S | 10A | A | CA | 0.618 |
| 12A | Q | 11A | N | CA | 1.417 |
| 13A | N | 12A | N | CA | 1.414 |
| 14A | M | 13A | R | CA | 0.918 |
| 15A | G | 14A | V | CA | 0.424 |
| 16A | I | 15A | I | CA | 0.616 |
| 17A | G | 16A | G | CA | 0.848 |
| 18A | K | 17A | L | CA | 0.836 |
| 19A | N | 18A | D | CA | 1.222 |
| 20A | G | 19A | N | CA | 1.016 |
| 21A | D | 20A | K | CA | 0.707 |
| 22A | L | 21A | M | CA | 0.974 |
| 23A | P | 22A | P | CA | 1.659 |
| 24A | W | 23A | W | CA | 1.702 |
| 25A | P | | | | |
| 26A | P | 24A | H | CA | 1.360 |
| 27A | L | 25A | L | CA | 0.587 |
| 28A | R | 26A | P | CA | 0.191 |
| 29A | N | 27A | A | CA | 0.213 |
| 30A | E | 28A | E | CA | 0.136 |
| 31A | F | 29A | L | CA | 0.394 |
| 32A | K | 30A | Q | CA | 0.514 |
| 33A | Y | 31A | L | CA | 0.436 |
| 34A | F | 32A | F | CA | 0.397 |
| 35A | Q | 33A | K | CA | 0.554 |
| 36A | R | 34A | R | CA | 0.894 |
| 37A | M | 35A | A | CA | 0.996 |
| 38A | T | 36A | T | CA | 0.301 |
| 39A | T | | | | |
| 40A | T | 37A | L | CA | 4.099 |
| 41A | S | 38A | G | CA | 4.507 |
| 42A | S | | | | |
| 43A | V | | | | |
| 44A | E | | | | |
| 45A | G | | | | |
| 46A | K | | | | |
| 47A | Q | | | | |
| 48A | N | 39A | K | CA | 1.331 |

# Appendix part III

| | | | | | |
|---|---|---|---|---|---|
| 49A | L | 40A | P | CA | 0.573 |
| 50A | V | 41A | I | CA | 0.524 |
| 51A | I | 42A | V | CA | 0.575 |
| 52A | M | 43A | M | CA | 0.466 |
| 53A | G | 44A | G | CA | 0.538 |
| 54A | R | 45A | R | CA | 0.515 |
| 55A | K | 46A | N | CA | 0.811 |
| 56A | T | 47A | T | CA | 0.641 |
| 57A | W | 48A | F | CA | 0.579 |
| 58A | F | 49A | E | CA | 0.950 |
| 59A | S | 50A | S | CA | 0.857 |
| 60A | I | 51A | I | CA | 0.996 |
| 61A | P | 52A | G | CA | 3.036 |
| 62A | E | | | | |
| 63A | K | | | | |
| 64A | N | | | | |
| 65A | R | 53A | R | CA | 2.463 |
| 66A | P | 54A | P | CA | 1.389 |
| 67A | L | 55A | L | CA | 0.606 |
| 68A | K | 56A | P | CA | 0.367 |
| 69A | D | 57A | G | CA | 0.681 |
| 70A | R | 58A | R | CA | 0.617 |
| 71A | I | 59A | L | CA | 0.641 |
| 72A | N | 60A | N | CA | 0.667 |
| 73A | I | 61A | I | CA | 0.566 |
| 74A | V | 62A | V | CA | 0.551 |
| 75A | L | 63A | L | CA | 0.350 |
| 76A | S | 64A | S | CA | 0.218 |
| 77A | R | 65A | R | CA | 0.439 |
| 78A | E | 66A | Q | CA | 0.865 |
| | | 67A | T | | |
| 79A | L | 68A | D | CA | 3.843 |
| 80A | K | 69A | Y | CA | 4.314 |
| 81A | E | 70A | Q | CA | 3.547 |
| 82A | P | | | | |
| 83A | P | 71A | P | CA | 1.291 |
| 84A | R | 72A | E | CA | 0.916 |
| 85A | G | 73A | G | CA | 1.404 |
| 86A | A | 74A | V | CA | 0.731 |
| 87A | H | | | | |
| 88A | F | 75A | T | CA | 1.400 |
| 89A | L | 76A | V | CA | 0.980 |
| 90A | A | 77A | V | CA | 0.601 |
| 91A | K | 78A | A | CA | 0.380 |
| 92A | S | 79A | T | CA | 0.945 |
| 93A | L | 80A | L | CA | 0.916 |
| 94A | D | 81A | E | CA | 1.335 |
| 95A | D | 82A | D | CA | 1.401 |
| 96A | A | 83A | A | CA | 1.005 |
| 97A | L | 84A | V | CA | 0.970 |
| 98A | R | 85A | V | CA | 2.717 |
| 99A | L | 86A | A | CA | 3.095 |
| 100A | I | 87A | A | CA | 2.828 |

# Appendix part III

| | | | | | |
|---|---|---|---|---|---|
| 101A | E | | | | |
| 102A | Q | 88A | G | CA | 5.152 |
| 103A | P | | | | |
| 104A | D | | | | |
| 105A | L | | | | |
| 106A | A | 89A | D | CA | 3.140 |
| 107A | S | | | | |
| 108A | K | | | | |
| 109A | V | 90A | V | CA | 4.024 |
| 110A | D | 91A | E | CA | 4.468 |
| 111A | M | 92A | E | CA | 2.934 |
| 112A | V | 93A | L | CA | 1.230 |
| 113A | W | 94A | M | CA | 0.654 |
| 114A | I | 95A | I | CA | 0.758 |
| 115A | V | 96A | I | CA | 0.716 |
| 116A | G | 97A | G | CA | 1.313 |
| 117A | G | 98A | G | CA | 0.580 |
| 118A | S | 99A | A | CA | 0.443 |
| 119A | S | 100A | T | CA | 0.245 |
| 120A | V | 101A | I | CA | 0.558 |
| 121A | Y | 102A | Y | CA | 0.600 |
| 122A | Q | 103A | N | CA | 0.706 |
| 123A | E | 104A | Q | CA | 0.963 |
| 124A | A | 105A | C | CA | 1.178 |
| 125A | M | 106A | L | CA | 0.782 |
| 126A | N | 107A | A | CA | 2.882 |
| 127A | Q | | | | |
| 128A | P | 108A | A | CA | 3.684 |
| 129A | G | | | | |
| 130A | H | 109A | A | CA | 5.546 |
| 131A | L | 110A | D | CA | 1.282 |
| 132A | R | 111A | R | CA | 0.572 |
| 133A | L | 112A | L | CA | 0.366 |
| 134A | F | 113A | Y | CA | 0.303 |
| 135A | V | 114A | L | CA | 0.336 |
| 136A | T | 115A | T | CA | 0.170 |
| 137A | R | 116A | H | CA | 0.319 |
| 138A | I | 117A | I | CA | 0.273 |
| 139A | M | 118A | E | CA | 1.167 |
| 140A | Q | 119A | L | CA | 1.691 |
| 141A | E | 120A | T | CA | 0.858 |
| 142A | F | 121A | T | CA | 1.121 |
| 143A | E | 122A | E | CA | 1.037 |
| 144A | S | 123A | G | CA | 1.017 |
| 145A | D | 124A | D | CA | 1.043 |
| 146A | T | 125A | T | CA | 1.343 |
| 147A | F | 126A | W | CA | 0.904 |
| 148A | F | 127A | F | CA | 0.686 |
| 149A | P | 128A | P | CA | 0.581 |
| 150A | E | 129A | D | CA | 0.642 |
| 151A | I | 130A | Y | CA | 1.629 |
| | | 131A | E | | |
| 152A | D | 132A | Q | CA | 2.518 |

# Appendix part III

| | | | | | |
|---|---|---|---|---|---|
| 153A | L | | | | |
| 154A | G | 133A | Y | CA | 4.330 |
| 155A | K | 134A | N | CA | 2.183 |
| 156A | Y | 135A | W | CA | 1.915 |
| 157A | K | 136A | Q | CA | 1.878 |
| 158A | L | 137A | E | CA | 2.272 |
| 159A | L | 138A | I | CA | 2.300 |
| 160A | P | | | | |
| 161A | E | 139A | E | CA | 2.440 |
| 162A | Y | | | | |
| 163A | P | | | | |
| 164A | G | | | | |
| 165A | V | | | | |
| 166A | L | | | | |
| 167A | S | 140A | H | CA | 3.423 |
| 168A | E | 141A | E | CA | 3.705 |
| 169A | V | 142A | S | CA | 2.154 |
| 170A | Q | 143A | Y | CA | 1.039 |
| 171A | E | 144A | A | CA | 3.026 |
| 172A | E | 145A | A | CA | 3.721 |
| | | 146A | D | | |
| | | 147A | D | | |
| | | 148A | K | | |
| 173A | K | 149A | N | CA | 3.631 |
| 174A | G | 150A | P | CA | 2.899 |
| 175A | I | 151A | H | CA | 0.833 |
| 176A | K | 152A | N | CA | 0.552 |
| 177A | Y | 153A | Y | CA | 0.228 |
| 178A | K | 154A | R | CA | 0.746 |
| 179A | F | 155A | F | CA | 0.272 |
| 180A | E | 156A | S | CA | 0.244 |
| 181A | V | 157A | L | CA | 0.801 |
| 182A | Y | 158A | L | CA | 0.895 |
| 183A | E | 159A | E | CA | 1.048 |
| 184A | K | 160A | R | CA | 1.303 |
| 185A | K | 161A | V | CA | 5.317 |
| 186A | D | | | | |

# Appendix part III

**b) Core identification (core regions):**

Details of the quality of superposition between **3K45** and the other fixed molecule**s.** As shown, all the local deviation between CA atoms of every matching residues in both structures is less than 2 Å indicating quite good fit between the structures

**Table 09.**Residues used for superposition between 3K45 and 1U70

| Residue number (Fixed) | Residue name (Fixed) | Residue number (Rotated) | Residue name (Rotated) | Atom name | Deviation (Å) |
|---|---|---|---|---|---|
| 1A | V | 1A | V | CA | 0.597 |
| 2A | R | 2A | R | CA | 0.536 |
| 3A | P | 3A | P | CA | 0.457 |
| 4A | L | 4A | L | CA | 0.512 |
| 5A | N | 5A | N | CA | 0.159 |
| 6A | C | 6A | C | CA | 0.112 |
| 7A | I | 7A | I | CA | 0.130 |
| 8A | V | 8A | V | CA | 0.188 |
| 9A | A | 9A | A | CA | 0.257 |
| 10A | V | 10A | V | CA | 0.329 |
| 11A | S | 11A | S | CA | 0.111 |
| 12A | Q | 12A | Q | CA | 0.166 |
| 13A | N | 13A | N | CA | 0.061 |
| 14A | M | 14A | M | CA | 0.151 |
| 15A | G | 15A | G | CA | 0.332 |
| 16A | I | 16A | I | CA | 0.463 |
| 17A | G | 17A | G | CA | 1.343 |
| 18A | K | 18A | K | CA | 0.524 |
| 19A | N | 19A | N | CA | 0.564 |
| 20A | G | 20A | G | CA | 1.795 |
| 21A | D | 21A | D | CA | 0.981 |
| 22A | L | 22A | R | CA | 0.425 |
| 23A | P | 23A | P | CA | 0.662 |
| 24A | W | 24A | W | CA | 0.613 |
| 25A | P | 25A | P | CA | 0.744 |
| 26A | P | 26A | P | CA | 0.528 |
| 27A | L | 27A | L | CA | 0.363 |
| 28A | R | 28A | R | CA | 0.483 |
| 29A | N | 29A | N | CA | 0.175 |
| 30A | E | 30A | E | CA | 0.403 |
| 31A | F | 31A | F | CA | 0.377 |
| 32A | K | 32A | K | CA | 0.210 |
| 33A | Y | 33A | Y | CA | 0.255 |
| 34A | F | 34A | F | CA | 0.365 |
| 35A | Q | 35A | Q | CA | 0.326 |
| 36A | R | 36A | R | CA | 0.355 |
| 37A | M | 37A | M | CA | 0.457 |
| 38A | T | 38A | T | CA | 0.153 |
| 39A | T | 39A | T | CA | 0.274 |
| 40A | T | 40A | T | CA | 0.165 |
| 41A | S | 41A | S | CA | 0.170 |

| 42A | S | 42A | S | CA | 0.557 |
|---|---|---|---|---|---|
| 43A | V | 43A | V | CA | 0.303 |
| 44A | E | 44A | E | CA | 0.276 |
| 45A | G | 45A | G | CA | 0.261 |
| 46A | K | 46A | K | CA | 0.365 |
| 47A | Q | 47A | Q | CA | 0.271 |
| 48A | N | 48A | N | CA | 0.232 |
| 49A | L | 49A | L | CA | 0.171 |
| 50A | V | 50A | V | CA | 0.061 |
| 51A | I | 51A | I | CA | 0.136 |
| 52A | M | 52A | M | CA | 0.035 |
| 53A | G | 53A | G | CA | 0.137 |
| 54A | R | 54A | R | CA | 0.216 |
| 55A | K | 55A | K | CA | 0.210 |
| 56A | T | 56A | T | CA | 0.299 |
| 57A | W | 57A | W | CA | 0.162 |
| 58A | F | 58A | F | CA | 0.394 |
| 59A | S | 59A | S | CA | 0.476 |
| 60A | I | 60A | I | CA | 0.460 |
| 61A | P | 61A | P | CA | 0.457 |
| 62A | E | 62A | E | CA | 0.225 |
| 63A | K | 63A | K | CA | 0.450 |
| 64A | N | 64A | N | CA | 0.291 |
| 65A | R | 65A | R | CA | 0.233 |
| 66A | P | 66A | P | CA | 0.365 |
| 67A | L | 67A | L | CA | 0.423 |
| 68A | K | 68A | K | CA | 0.414 |
| 69A | D | 69A | D | CA | 0.354 |
| 70A | R | 70A | R | CA | 0.119 |
| 71A | I | 71A | I | CA | 0.154 |
| 72A | N | 72A | N | CA | 0.146 |
| 73A | I | 73A | I | CA | 0.091 |
| 74A | V | 74A | V | CA | 0.224 |
| 75A | L | 75A | L | CA | 0.238 |
| 76A | S | 76A | S | CA | 0.261 |
| 77A | R | 77A | R | CA | 0.449 |
| 78A | E | 78A | E | CA | 0.729 |
| 79A | L | 79A | L | CA | 0.337 |
| 80A | K | 80A | K | CA | 0.429 |
| 81A | E | 81A | E | CA | 0.804 |
| 82A | P | 82A | P | CA | 0.755 |
| 83A | P | 83A | P | CA | 0.666 |
| 84A | R | 84A | R | CA | 0.386 |
| 85A | G | 85A | G | CA | 0.156 |
| 86A | A | 86A | A | CA | 0.588 |
| 87A | H | 87A | H | CA | 0.238 |
| 88A | F | 88A | F | CA | 0.060 |
| 89A | L | 89A | L | CA | 0.194 |
| 90A | A | 90A | A | CA | 0.207 |
| 91A | K | 91A | K | CA | 0.251 |
| 92A | S | 92A | S | CA | 0.339 |
| 93A | L | 93A | L | CA | 0.063 |
| 94A | D | 94A | D | CA | 0.432 |

# Appendix part III

| 95A | D | 95A | D | CA | 0.558 |
|---|---|---|---|---|---|
| 96A | A | 96A | A | CA | 0.181 |
| 97A | L | 97A | L | CA | 0.344 |
| 98A | R | 98A | R | CA | 0.111 |
| 99A | L | 99A | L | CA | 0.222 |
| 100A | I | 100A | I | CA | 0.260 |
| 101A | E | 101A | E | CA | 0.270 |
| 102A | Q | 102A | Q | CA | 0.231 |
| 103A | P | 103A | P | CA | 0.138 |
| 104A | D | 104A | E | CA | 0.359 |
| 105A | L | 105A | L | CA | 0.765 |
| 106A | A | 106A | A | CA | 1.826 |
| 107A | S | 107A | S | CA | 1.763 |
| 108A | K | 108A | K | CA | 0.734 |
| 109A | V | 109A | V | CA | 0.319 |
| 110A | D | 110A | D | CA | 0.389 |
| 111A | M | 111A | M | CA | 0.100 |
| 112A | V | 112A | V | CA | 0.117 |
| 113A | W | 113A | W | CA | 0.031 |
| 114A | I | 114A | I | CA | 0.274 |
| 115A | V | 115A | V | CA | 0.284 |
| 116A | G | 116A | G | CA | 0.736 |
| 117A | G | 117A | G | CA | 1.555 |
| 118A | S | 118A | S | CA | 0.971 |
| 119A | S | 119A | S | CA | 0.975 |
| 120A | V | 120A | V | CA | 0.730 |
| 121A | Y | 121A | Y | CA | 0.481 |
| 122A | Q | 122A | Q | CA | 0.740 |
| 123A | E | 123A | E | CA | 0.694 |
| 124A | A | 124A | A | CA | 0.502 |
| 125A | M | 125A | M | CA | 0.875 |
| 126A | N | 126A | N | CA | 0.557 |
| 127A | Q | 127A | Q | CA | 1.454 |
| 128A | P | 128A | P | CA | 0.735 |
| 129A | G | 129A | G | CA | 0.761 |
| 130A | H | 130A | H | CA | 0.806 |
| 131A | L | 131A | L | CA | 0.542 |
| 132A | R | 132A | R | CA | 0.283 |
| 133A | L | 133A | L | CA | 0.315 |
| 134A | F | 134A | F | CA | 0.137 |
| 135A | V | 135A | V | CA | 0.189 |
| 136A | T | 136A | T | CA | 0.263 |
| 137A | R | 137A | R | CA | 0.309 |
| 138A | I | 138A | I | CA | 0.173 |
| 139A | M | 139A | M | CA | 0.121 |
| 140A | Q | 140A | Q | CA | 0.072 |
| 141A | E | 141A | E | CA | 0.185 |
| 142A | F | 142A | F | CA | 0.395 |
| 143A | E | 143A | E | CA | 0.619 |
| 144A | S | 144A | S | CA | 0.422 |
| 145A | D | 145A | D | CA | 0.556 |
| 146A | T | 146A | T | CA | 0.991 |
| 147A | F | 147A | F | CA | 0.792 |

| 148A | F | 148A | F | CA | 0.353 |
|---|---|---|---|---|---|
| 149A | P | 149A | P | CA | 0.340 |
| 150A | E | 150A | E | CA | 0.618 |
| 151A | I | 151A | I | CA | 0.409 |
| 152A | D | 152A | D | CA | 0.434 |
| 153A | L | 153A | L | CA | 0.306 |
| 154A | G | 154A | G | CA | 0.226 |
| 155A | K | 155A | K | CA | 0.425 |
| 156A | Y | 156A | Y | CA | 0.692 |
| 157A | K | 157A | K | CA | 0.294 |
| 158A | L | 158A | L | CA | 0.514 |
| 159A | L | 159A | L | CA | 0.396 |
| 160A | P | 160A | P | CA | 0.218 |
| 161A | E | 161A | E | CA | 0.397 |
| 162A | Y | 162A | Y | CA | 0.360 |
| 163A | P | 163A | P | CA | 0.285 |
| 164A | G | 164A | G | CA | 0.333 |
| 165A | V | 165A | V | CA | 0.300 |
| 166A | L | 166A | L | CA | 0.172 |
| 167A | S | 167A | S | CA | 0.201 |
| 168A | E | 168A | E | CA | 0.035 |
| 169A | V | 169A | V | CA | 0.098 |
| 170A | Q | 170A | Q | CA | 0.228 |
| 171A | E | 171A | E | CA | 0.242 |
| 172A | E | 172A | E | CA | 0.363 |
| 173A | K | 173A | K | CA | 0.333 |
| 174A | G | 174A | G | CA | 0.184 |
| 175A | I | 175A | I | CA | 0.193 |
| 176A | K | 176A | K | CA | 0.186 |
| 177A | Y | 177A | Y | CA | 0.263 |
| 178A | K | 178A | K | CA | 0.148 |
| 179A | F | 179A | F | CA | 0.141 |
| 180A | E | 180A | E | CA | 0.123 |
| 181A | V | 181A | V | CA | 0.205 |
| 182A | Y | 182A | Y | CA | 0.179 |
| 183A | E | 183A | E | CA | 0.238 |
| 184A | K | 184A | K | CA | 0.466 |
| 185A | K | 185A | K | CA | 0.630 |

# Appendix part III

**Table 10.** Residues used for superposition between 3K45 and 1DR1

| Residue number (Fixed) | Residue name (Fixed) | Residue number (Rotated) | Residue name (Rotated) | Atom name | Deviation (Å) |
|---|---|---|---|---|---|
| 3A | P | 3A | S | CA | 0.844 |
| 4A | L | 4A | L | CA | 0.792 |
| 5A | N | 5A | N | CA | 0.568 |
| 6A | C | 6A | S | CA | 0.640 |
| 7A | I | 7A | I | CA | 0.263 |
| 8A | V | 8A | V | CA | 0.186 |
| 9A | A | 9A | A | CA | 0.263 |
| 10A | V | 10A | V | CA | 0.420 |
| 11A | S | 11A | C | CA | 0.309 |
| 12A | Q | 12A | Q | CA | 0.874 |
| 13A | N | 13A | N | CA | 1.066 |
| 14A | M | 14A | M | CA | 0.863 |
| 15A | G | 15A | G | CA | 0.905 |
| 16A | I | 16A | I | CA | 0.759 |
| 17A | G | 17A | G | CA | 1.181 |
| 18A | K | 18A | K | CA | 1.196 |
| 19A | N | 19A | D | CA | 1.529 |
| 20A | G | 20A | G | CA | 1.540 |
| 21A | D | 21A | N | CA | 1.114 |
| 22A | L | 22A | L | CA | 0.814 |
| 23A | P | 23A | P | CA | 0.868 |
| 24A | W | 24A | W | CA | 0.790 |
| 25A | P | 25A | P | CA | 0.798 |
| 26A | P | 26A | P | CA | 1.058 |
| 27A | L | 27A | L | CA | 0.921 |
| 28A | R | 28A | R | CA | 1.313 |
| 29A | N | 29A | N | CA | 0.817 |
| 30A | E | 30A | E | CA | 0.489 |
| 31A | F | 31A | Y | CA | 0.437 |
| 32A | K | 32A | K | CA | 0.122 |
| 33A | Y | 33A | Y | CA | 0.161 |
| 34A | F | 34A | F | CA | 0.153 |
| 35A | Q | 35A | Q | CA | 0.155 |
| 36A | R | 36A | R | CA | 0.175 |
| 37A | M | 37A | M | CA | 0.164 |
| 38A | T | 38A | T | CA | 0.098 |
| 39A | T | 39A | S | CA | 0.178 |
| 40A | T | 40A | T | CA | 0.215 |
| 41A | S | 41A | S | CA | 0.217 |
| 42A | S | 42A | H | CA | 0.788 |
| 43A | V | 43A | V | CA | 0.536 |
| 44A | E | 44A | E | CA | 0.884 |
| 45A | G | 45A | G | CA | 0.624 |
| 46A | K | 46A | K | CA | 0.279 |
| 47A | Q | 47A | Q | CA | 0.145 |
| 48A | N | 48A | N | CA | 0.385 |
| 49A | L | 49A | A | CA | 0.424 |
| 50A | V | 50A | V | CA | 0.125 |
| 51A | I | 51A | I | CA | 0.190 |

# Appendix part III

| 52A | M | 52A | M | CA | 0.237 |
|---|---|---|---|---|---|
| 53A | G | 53A | G | CA | 0.229 |
| 54A | R | 54A | K | CA | 0.569 |
| 55A | K | 55A | K | CA | 0.594 |
| 56A | T | 56A | T | CA | 0.342 |
| 57A | W | 57A | W | CA | 0.188 |
| 58A | F | 58A | F | CA | 0.543 |
| 59A | S | 59A | S | CA | 0.498 |
| 60A | I | 60A | I | CA | 0.432 |
| 61A | P | 61A | P | CA | 0.687 |
| 62A | E | 62A | E | CA | 0.803 |
| 63A | K | 63A | K | CA | 1.365 |
| 64A | N | 64A | N | CA | 1.009 |
| 65A | R | 65A | R | CA | 0.490 |
| 66A | P | 66A | P | CA | 0.252 |
| 67A | L | 67A | L | CA | 0.346 |
| 68A | K | 68A | K | CA | 0.150 |
| 69A | D | 69A | D | CA | 0.346 |
| 70A | R | 70A | R | CA | 0.280 |
| 71A | I | 71A | I | CA | 0.428 |
| 72A | N | 72A | N | CA | 0.429 |
| 73A | I | 73A | I | CA | 0.358 |
| 74A | V | 74A | V | CA | 0.447 |
| 75A | L | 75A | L | CA | 0.370 |
| 76A | S | 76A | S | CA | 0.635 |
| 77A | R | 77A | R | CA | 0.848 |
| 78A | E | 78A | E | CA | 0.982 |
| 79A | L | 79A | L | CA | 0.881 |
| 80A | K | 80A | K | CA | 0.640 |
| 81A | E | 81A | E | CA | 0.664 |
| 82A | P | 82A | A | CA | 0.574 |
| 83A | P | 83A | P | CA | 0.698 |
| 84A | R | 84A | K | CA | 0.732 |
| 85A | G | 85A | G | CA | 0.795 |
| 86A | A | 86A | A | CA | 0.548 |
| 87A | H | 87A | H | CA | 0.550 |
| 88A | F | 88A | Y | CA | 0.488 |
| 89A | L | 89A | L | CA | 0.472 |
| 90A | A | 90A | S | CA | 0.303 |
| 91A | K | 91A | K | CA | 0.776 |
| 92A | S | 92A | S | CA | 0.790 |
| 93A | L | 93A | L | CA | 0.726 |
| 94A | D | 94A | D | CA | 0.467 |
| 95A | D | 95A | D | CA | 0.292 |
| 96A | A | 96A | A | CA | 0.230 |
| 97A | L | 97A | L | CA | 0.482 |
| 98A | R | 98A | A | CA | 0.606 |
| 99A | L | 99A | L | CA | 0.549 |
| 100A | I | 100A | L | CA | 0.373 |
| 101A | E | 101A | D | CA | 0.853 |
| 102A | Q | 102A | S | CA | 1.257 |
| 104A | D | 104A | E | CA | 1.674 |
| 105A | L | 105A | L | CA | 1.366 |

# Appendix part III

| 107A | S | 107A | S | CA | 1.970 |
|---|---|---|---|---|---|
| 108A | K | 108A | K | CA | 0.459 |
| 109A | V | 109A | V | CA | 0.111 |
| 110A | D | 110A | D | CA | 0.271 |
| 111A | M | 111A | M | CA | 0.289 |
| 112A | V | 112A | V | CA | 0.221 |
| 113A | W | 113A | W | CA | 0.237 |
| 114A | I | 114A | I | CA | 0.453 |
| 115A | V | 115A | V | CA | 0.446 |
| 116A | G | 116A | G | CA | 1.128 |
| 117A | G | 117A | G | CA | 0.884 |
| 118A | S | 118A | T | CA | 1.305 |
| 119A | S | 119A | A | CA | 1.523 |
| 120A | V | 120A | V | CA | 0.813 |
| 121A | Y | 121A | Y | CA | 0.835 |
| 122A | Q | 122A | K | CA | 1.392 |
| 123A | E | 123A | A | CA | 1.232 |
| 124A | A | 124A | A | CA | 1.172 |
| 125A | M | 125A | M | CA | 1.405 |
| 126A | N | 126A | E | CA | 1.746 |
| 127A | Q | 127A | K | CA | 1.230 |
| 128A | P | 128A | P | CA | 1.191 |
| 129A | G | 129A | I | CA | 1.288 |
| 130A | H | 130A | N | CA | 0.880 |
| 131A | L | 131A | H | CA | 1.381 |
| 132A | R | 132A | R | CA | 1.163 |
| 133A | L | 133A | L | CA | 0.724 |
| 134A | F | 134A | F | CA | 0.423 |
| 135A | V | 135A | V | CA | 0.239 |
| 136A | T | 136A | T | CA | 0.226 |
| 137A | R | 137A | R | CA | 0.347 |
| 138A | I | 138A | I | CA | 0.426 |
| 139A | M | 139A | L | CA | 0.601 |
| 140A | Q | 140A | H | CA | 0.593 |
| 141A | E | 141A | E | CA | 0.950 |
| 142A | F | 142A | F | CA | 1.077 |
| 143A | E | 143A | E | CA | 1.016 |
| 144A | S | 144A | S | CA | 1.161 |
| 145A | D | 145A | D | CA | 0.867 |
| 146A | T | 146A | T | CA | 1.616 |
| 147A | F | 147A | F | CA | 1.158 |
| 148A | F | 148A | F | CA | 0.760 |
| 149A | P | 149A | P | CA | 1.082 |
| 150A | E | 150A | E | CA | 1.325 |
| 151A | I | 151A | I | CA | 1.901 |
| 153A | L | 153A | Y | CA | 1.534 |
| 154A | G | 154A | K | CA | 1.056 |
| 155A | K | 155A | D | CA | 1.076 |
| 156A | Y | 156A | F | CA | 0.562 |
| 157A | K | 157A | K | CA | 0.706 |
| 158A | L | 158A | L | CA | 0.450 |
| 159A | L | 159A | L | CA | 0.829 |
| 160A | P | 160A | T | CA | 1.159 |

# Appendix part III

| 161A | E | 161A | E | CA | 1.221 |
|---|---|---|---|---|---|
| 162A | Y | 162A | Y | CA | 1.271 |
| 163A | P | 163A | P | CA | 1.421 |
| 164A | G | 164A | G | CA | 1.572 |
| 165A | V | 165A | V | CA | 1.189 |
| 166A | L | 166A | P | CA | 1.215 |
| 167A | S | 167A | A | CA | 1.261 |
| 168A | E | 168A | D | CA | 1.415 |
| 169A | V | 169A | I | CA | 0.990 |
| 170A | Q | 170A | Q | CA | 0.866 |
| 171A | E | 171A | E | CA | 0.842 |
| 172A | E | 172A | E | CA | 1.038 |
| 173A | K | 173A | D | CA | 0.566 |
| 174A | G | 174A | G | CA | 0.536 |
| 175A | I | 175A | I | CA | 0.777 |
| 176A | K | 176A | Q | CA | 0.484 |
| 177A | Y | 177A | Y | CA | 0.582 |
| 178A | K | 178A | K | CA | 1.026 |
| 179A | F | 179A | F | CA | 0.716 |
| 180A | E | 180A | E | CA | 0.628 |
| 181A | V | 181A | V | CA | 0.460 |
| 182A | Y | 182A | Y | CA | 0.289 |
| 183A | E | 183A | Q | CA | 0.611 |
| 184A | K | 184A | K | CA | 0.564 |
| 185A | K | 185A | S | CA | 1.014 |
| 186A | D | 186A | V | CA | 1.438 |

# Appendix part III

**Table 11.** Residues used for superposition between 3K45 and 4H95

| Residue number (Fixed) | Residue name (Fixed) | Residue number (Rotated) | Residue name (Rotated) | Atom name | Deviation (Å) |
|---|---|---|---|---|---|
| 1A | V | | | | |
| 4A | L | 6A | V | CA | 1.159 |
| 5A | N | 7A | A | CA | 0.702 |
| 6A | C | 8A | I | CA | 0.368 |
| 7A | I | 9A | I | CA | 0.262 |
| 8A | V | 10A | V | CA | 0.213 |
| 9A | A | 11A | A | CA | 0.307 |
| 10A | V | 12A | A | CA | 0.273 |
| 11A | S | 13A | L | CA | 0.597 |
| 12A | Q | 14A | K | CA | 1.557 |
| | | 15A | P | | |
| 14A | M | 17A | L | CA | 0.961 |
| 15A | G | 18A | G | CA | 0.299 |
| 16A | I | 19A | I | CA | 0.583 |
| 17A | G | 20A | G | CA | 1.277 |
| 18A | K | 21A | Y | CA | 0.972 |
| 19A | N | 22A | K | CA | 1.078 |
| 20A | G | 23A | G | CA | 1.138 |
| 21A | D | 24A | K | CA | 0.942 |
| 22A | L | 25A | M | CA | 0.565 |
| 23A | P | 26A | P | CA | 1.925 |
| 25A | P | | | | |
| 26A | P | 28A | R | CA | 0.988 |
| 27A | L | 29A | L | CA | 0.791 |
| 28A | R | 30A | R | CA | 1.353 |
| 29A | N | 31A | K | CA | 0.909 |
| 30A | E | 32A | E | CA | 0.696 |
| 31A | F | 33A | I | CA | 0.393 |
| 32A | K | 34A | R | CA | 0.457 |
| 33A | Y | 35A | Y | CA | 0.544 |
| 34A | F | 36A | F | CA | 0.462 |
| 35A | Q | 37A | K | CA | 0.415 |
| 36A | R | 38A | D | CA | 0.704 |
| 37A | M | 39A | V | CA | 0.634 |
| 38A | T | 40A | T | CA | 0.467 |
| 39A | T | 41A | T | CA | 0.709 |
| 40A | T | 42A | R | CA | 0.677 |
| 41A | S | 43A | T | CA | 1.165 |
| 42A | S | 44A | T | CA | 1.405 |
| 43A | V | 45A | K | CA | 1.622 |
| 46A | K | 48A | T | CA | 0.970 |
| 47A | Q | 49A | R | CA | 0.729 |
| 48A | N | 50A | N | CA | 0.550 |
| 49A | L | 51A | A | CA | 0.526 |
| 50A | V | 52A | V | CA | 0.331 |
| 51A | I | 53A | I | CA | 0.357 |
| 52A | M | 54A | M | CA | 0.255 |
| 53A | G | 55A | G | CA | 0.273 |
| 54A | R | 56A | R | CA | 0.294 |

# Appendix part III

| | | | | | |
|---|---|---|---|---|---|
| 55A | K | 57A | K | CA | 0.392 |
| 56A | T | 58A | T | CA | 0.414 |
| 57A | W | 59A | W | CA | 0.432 |
| 58A | F | 60A | E | CA | 0.572 |
| 59A | S | 61A | S | CA | 0.456 |
| 60A | I | 62A | I | CA | 0.469 |
| 61A | P | 63A | P | CA | 0.209 |
| 62A | E | 64A | Q | CA | 0.817 |
| 63A | K | 65A | K | CA | 0.680 |
| 64A | N | 66A | F | CA | 0.306 |
| 65A | R | 67A | R | CA | 0.519 |
| 66A | P | 68A | P | CA | 0.572 |
| 67A | L | 69A | L | CA | 0.707 |
| 68A | K | 70A | P | CA | 0.867 |
| 69A | D | 71A | D | CA | 0.315 |
| 70A | R | 72A | R | CA | 0.354 |
| 71A | I | 73A | L | CA | 0.469 |
| 72A | N | 74A | N | CA | 0.450 |
| 73A | I | 75A | I | CA | 0.534 |
| 74A | V | 76A | I | CA | 0.326 |
| 75A | L | 77A | L | CA | 0.325 |
| 76A | S | 78A | S | CA | 0.938 |
| 77A | R | 79A | R | CA | 1.874 |
| | | 83A | N | | |
| 87A | H | 90A | I | CA | 1.180 |
| 88A | F | 91A | I | CA | 0.369 |
| 89A | L | 92A | H | CA | 0.119 |
| 90A | A | 93A | A | CA | 0.067 |
| 91A | K | 94A | S | CA | 1.437 |
| 95A | D | 98A | S | CA | 1.493 |
| 96A | A | 99A | S | CA | 1.697 |
| 101A | E | | | | |
| 102A | Q | | | | |
| 103A | P | | | | |
| 104A | D | | | | |
| 105A | L | | | | |
| 108A | K | | | | |
| 109A | V | 106A | V | CA | 1.363 |
| 110A | D | 107A | E | CA | 1.005 |
| 111A | M | 108A | R | CA | 0.706 |
| 112A | V | 109A | V | CA | 0.607 |
| 113A | W | 110A | F | CA | 0.495 |
| 114A | I | 111A | I | CA | 0.484 |
| 115A | V | 112A | I | CA | 0.355 |
| 116A | G | 113A | G | CA | 0.349 |
| 117A | G | 114A | G | CA | 0.420 |
| 118A | S | 115A | A | CA | 0.932 |
| 119A | S | 116A | E | CA | 1.079 |
| 120A | V | 117A | I | CA | 0.971 |
| 121A | Y | 118A | Y | CA | 1.138 |
| 122A | Q | 119A | N | CA | 1.786 |
| 124A | A | 121A | L | CA | 1.391 |
| 125A | M | 122A | I | CA | 1.129 |

# Appendix part III

| | | | | | |
|---|---|---|---|---|---|
| **126A** | **N** | **123A** | **N** | **CA** | **1.760** |
| **127A** | **Q** | **124A** | **N** | **CA** | **1.109** |
| **128A** | **P** | **125A** | **S** | **CA** | **1.762** |
| **131A** | **L** | **128A** | **S** | **CA** | **1.301** |
| **132A** | **R** | **129A** | **H** | **CA** | **0.552** |
| **133A** | **L** | **130A** | **L** | **CA** | **0.517** |
| **134A** | **F** | **131A** | **L** | **CA** | **0.321** |
| **135A** | **V** | **132A** | **I** | **CA** | **0.276** |
| **136A** | **T** | **133A** | **T** | **CA** | **0.451** |
| **137A** | **R** | **134A** | **E** | **CA** | **0.692** |
| **138A** | **I** | **135A** | **I** | **CA** | **0.943** |
| **139A** | **M** | **136A** | **E** | **CA** | **0.459** |
| | | **138A** | **P** | | |
| | | **139A** | **S** | | |
| | | **140A** | **P** | | |
| | | **142A** | **S** | | |
| **144A** | **S** | **145A** | **M** | **CA** | **1.399** |
| **145A** | **D** | **146A** | **D** | **CA** | **1.390** |
| **146A** | **T** | **147A** | **T** | **CA** | **0.746** |
| **147A** | **F** | **148A** | **F** | **CA** | **0.642** |
| **148A** | **F** | **149A** | **L** | **CA** | **0.986** |
| **149A** | **P** | **150A** | **K** | **CA** | **1.511** |
| **154A** | **G** | | | | |
| **156A** | **Y** | **156A** | **W** | **CA** | **1.104** |
| **157A** | **K** | **157A** | **T** | **CA** | **1.154** |
| **158A** | **L** | **158A** | **K** | **CA** | **0.491** |
| **159A** | **L** | **159A** | **Q** | **CA** | **0.493** |
| **160A** | **P** | **160A** | **P** | **CA** | **0.746** |
| **161A** | **E** | **161A** | **K** | **CA** | **1.788** |
| | | **162A** | **S** | | |
| | | **163A** | **E** | | |
| **162A** | **Y** | **164A** | **L** | **CA** | **1.444** |
| **163A** | **P** | **165A** | **Q** | **CA** | **1.911** |
| | | **166A** | **K** | | |
| | | **167A** | **F** | | |
| | | **168A** | **V** | | |
| | | **169A** | **G** | | |
| | | **171A** | **T** | | |
| | | **172A** | **V** | | |
| **167A** | **S** | | | | |
| **168A** | **E** | **175A** | **D** | **CA** | **1.261** |
| **169A** | **V** | **176A** | **D** | **CA** | **1.003** |
| **170A** | **Q** | **177A** | **I** | **CA** | **0.725** |
| **171A** | **E** | **178A** | **K** | **CA** | **1.039** |
| **172A** | **E** | **179A** | **E** | **CA** | **1.419** |
| **175A** | **I** | **182A** | **F** | **CA** | **1.481** |
| **176A** | **K** | **183A** | **T** | **CA** | **0.750** |
| **177A** | **Y** | **184A** | **Y** | **CA** | **0.663** |
| **178A** | **K** | **185A** | **N** | **CA** | **0.665** |
| **179A** | **F** | **186A** | **Y** | **CA** | **0.631** |
| **180A** | **E** | **187A** | **T** | **CA** | **0.697** |
| **181A** | **V** | **188A** | **L** | **CA** | **0.626** |
| **182A** | **Y** | **189A** | **W** | **CA** | **0.409** |

| | | | | | |
|---|---|---|---|---|---|
| **183A** | **E** | **190A** | **T** | **CA** | **0.912** |
| **184A** | **K** | **191A** | **R** | **CA** | **1.588** |
| **186A** | **D** | | | | |

# Appendix part III

**Table 12.** Residues used for superposition between 3K45 and 4GH8

| Residue number (Fixed) | Residue name (Fixed) | Residue number (Rotated) | Residue name (Rotated) | Atom name | Deviation (Å) |
|---|---|---|---|---|---|
| 1A | V | | | | |
| 2A | R | | | | |
| 3A | P | 1A | M | CA | 0.684 |
| 4A | L | 2A | I | CA | 0.533 |
| 5A | N | 3A | S | CA | 0.947 |
| 6A | C | 4A | L | CA | 0.422 |
| 7A | I | 5A | I | CA | 0.190 |
| 8A | V | 6A | A | CA | 0.162 |
| 9A | A | 7A | A | CA | 0.398 |
| 10A | V | 8A | L | CA | 0.218 |
| 11A | S | 9A | A | CA | 1.070 |
| 12A | Q | 10A | V | CA | 1.739 |
| 13A | N | 11A | D | CA | 1.022 |
| 14A | M | 12A | R | CA | 0.843 |
| 15A | G | 13A | V | CA | 0.374 |
| 16A | I | 14A | I | CA | 0.395 |
| 17A | G | 15A | G | CA | 0.315 |
| 18A | K | 16A | M | CA | 0.826 |
| 19A | N | 17A | E | CA | 1.246 |
| 20A | G | 18A | N | CA | 0.847 |
| 21A | D | 19A | A | CA | 0.284 |
| 22A | L | 20A | M | CA | 0.372 |
| 23A | P | 21A | P | CA | 1.485 |
| 24A | W | 22A | W | CA | 0.902 |
| 25A | P | 23A | P | CA | 0.984 |
| 26A | P | 24A | P | CA | 0.905 |
| 27A | L | 25A | L | CA | 0.694 |
| 28A | R | 26A | P | CA | 1.221 |
| 29A | N | 27A | A | CA | 1.389 |
| 30A | E | 28A | D | CA | 0.855 |
| 31A | F | 29A | L | CA | 0.619 |
| 32A | K | 30A | A | CA | 0.877 |
| 33A | Y | 31A | W | CA | 0.816 |
| 34A | F | 32A | F | CA | 0.337 |
| 35A | Q | 33A | K | CA | 0.843 |
| 36A | R | 34A | R | CA | 1.055 |
| 37A | M | 35A | N | CA | 0.935 |
| 38A | T | 36A | T | CA | 0.474 |
| 39A | T | | | | |
| 42A | S | | | | |
| 43A | V | | | | |
| 44A | E | | | | |
| 45A | G | | | | |
| 46A | K | | | | |
| 47A | Q | | | | |
| 48A | N | 39A | K | CA | 1.020 |
| 49A | L | 40A | P | CA | 0.560 |
| 50A | V | 41A | V | CA | 0.421 |
| 51A | I | 42A | I | CA | 0.381 |

| 52A | M | 43A | M | CA | 0.182 |
|---|---|---|---|---|---|
| 53A | G | 44A | G | CA | 0.222 |
| 54A | R | 45A | R | CA | 0.381 |
| 55A | K | 46A | H | CA | 0.622 |
| 56A | T | 47A | T | CA | 0.437 |
| 57A | W | 48A | W | CA | 0.278 |
| 58A | F | 49A | E | CA | 1.008 |
| 59A | S | 50A | S | CA | 0.776 |
| 60A | I | 51A | I | CA | 0.665 |
| 61A | P | 52A | P | CA | 0.984 |
| 62A | E | 53A | E | CA | 1.482 |
| 64A | N | 55A | N | CA | 1.259 |
| 65A | R | 56A | R | CA | 0.745 |
| 66A | P | 57A | P | CA | 0.602 |
| 67A | L | 58A | L | CA | 0.325 |
| 68A | K | 59A | P | CA | 0.425 |
| 69A | D | 60A | G | CA | 0.652 |
| 70A | R | 61A | R | CA | 0.404 |
| 71A | I | 62A | K | CA | 0.504 |
| 72A | N | 63A | N | CA | 0.425 |
| 73A | I | 64A | I | CA | 0.315 |
| 74A | V | 65A | I | CA | 0.298 |
| 75A | L | 66A | L | CA | 0.303 |
| 76A | S | 67A | S | CA | 0.431 |
| 77A | R | 68A | S | CA | 0.360 |
| 78A | E | 69A | Q | CA | 1.410 |
| 79A | L | | | | |
| 80A | K | 70A | P | CA | 1.456 |
| 81A | E | | | | |
| | | 74A | D | | |
| 86A | A | 76A | V | CA | 1.094 |
| 87A | H | | | | |
| 88A | F | 77A | T | CA | 1.286 |
| 89A | L | 78A | W | CA | 1.056 |
| 90A | A | 79A | V | CA | 0.873 |
| 91A | K | 80A | K | CA | 0.512 |
| 92A | S | 81A | S | CA | 0.901 |
| 93A | L | 82A | V | CA | 1.208 |
| 94A | D | 83A | D | CA | 1.324 |
| 95A | D | 84A | E | CA | 1.274 |
| 96A | A | 85A | A | CA | 1.306 |
| 97A | L | 86A | I | CA | 0.922 |
| 102A | Q | | | | |
| 103A | P | | | | |
| 104A | D | | | | |
| 105A | L | | | | |
| 107A | S | | | | |
| 108A | K | | | | |
| 112A | V | 95A | I | CA | 1.188 |
| 113A | W | 96A | M | CA | 0.470 |
| 114A | I | 97A | V | CA | 0.398 |
| 115A | V | 98A | I | CA | 0.329 |
| 116A | G | 99A | G | CA | 0.610 |

| | | | | | |
|---|---|---|---|---|---|
| 117A | G | 100A | G | CA | 0.456 |
| 118A | S | 101A | G | CA | 0.427 |
| 119A | S | 102A | R | CA | 0.411 |
| 120A | V | 103A | V | CA | 0.371 |
| 121A | Y | 104A | Y | CA | 0.335 |
| 122A | Q | 105A | E | CA | 0.249 |
| 123A | E | 106A | Q | CA | 0.173 |
| 124A | A | 107A | F | CA | 0.154 |
| 125A | M | 108A | L | CA | 0.949 |
| 127A | Q | | | | |
| 128A | P | | | | |
| 131A | L | 112A | Q | CA | 1.794 |
| 132A | R | 113A | K | CA | 0.819 |
| 133A | L | 114A | L | CA | 0.706 |
| 134A | F | 115A | Y | CA | 0.290 |
| 135A | V | 116A | L | CA | 0.614 |
| 136A | T | 117A | T | CA | 0.503 |
| 137A | R | 118A | H | CA | 0.698 |
| 138A | I | 119A | I | CA | 0.808 |
| 141A | E | 122A | E | CA | 1.457 |
| 142A | F | 123A | V | CA | 1.505 |
| 143A | E | 124A | E | CA | 1.189 |
| 144A | S | 125A | G | CA | 0.995 |
| 145A | D | 126A | D | CA | 0.989 |
| 146A | T | 127A | T | CA | 0.936 |
| 147A | F | 128A | H | CA | 0.620 |
| 148A | F | 129A | F | CA | 0.592 |
| 149A | P | 130A | P | CA | 0.715 |
| 150A | E | 131A | D | CA | 1.077 |
| 151A | I | 132A | Y | CA | 1.459 |
| 152A | D | 133A | E | CA | 0.284 |
| 153A | L | 134A | P | CA | 0.505 |
| 154A | G | 135A | D | CA | 0.792 |
| 155A | K | 136A | D | CA | 0.778 |
| 156A | Y | 137A | W | CA | 1.106 |
| 157A | K | 138A | E | CA | 1.378 |
| 158A | L | 139A | S | CA | 1.250 |
| 159A | L | 140A | V | CA | 1.167 |
| 160A | P | | | | |
| 162A | Y | | | | |
| 163A | P | | | | |
| 164A | G | | | | |
| 165A | V | | | | |
| 166A | L | | | | |
| 169A | V | 144A | F | CA | 1.348 |
| 170A | Q | 145A | H | CA | 1.192 |
| | | 148A | D | | |
| | | 149A | A | | |
| | | 150A | Q | | |
| | | 151A | N | | |
| 173A | K | 152A | S | CA | 1.503 |
| 174A | G | | | | |
| 175A | I | 153A | H | CA | 0.884 |

| | | | | | |
|---|---|---|---|---|---|
| **176A** | **K** | **154A** | **S** | **CA** | **0.851** |
| **177A** | **Y** | **155A** | **Y** | **CA** | **0.596** |
| **178A** | **K** | **156A** | **C** | **CA** | **0.718** |
| **179A** | **F** | **157A** | **F** | **CA** | **0.571** |
| **180A** | **E** | **158A** | **E** | **CA** | **0.569** |
| **181A** | **V** | **159A** | **I** | **CA** | **0.727** |
| **182A** | **Y** | **160A** | **L** | **CA** | **1.058** |
| **183A** | **E** | **161A** | **E** | **CA** | **1.273** |
| **184A** | **K** | **162A** | **R** | **CA** | **1.614** |
| **185A** | **K** | | | | |
| **186A** | **D** | | | | |

# Appendix part III

**Table 13.** Residues used for superposition between 3K45 and 1ZDR

| Residue number (Fixed) | Residue name (Fixed) | Residue number (Rotated) | Residue name (Rotated) | Atom name | Deviation (Å) |
|---|---|---|---|---|---|
| 1A | V | | | | |
| 2A | R | | | | |
| 3A | P | 1A | M | CA | 1.446 |
| 4A | L | 2A | I | CA | 0.991 |
| 5A | N | 3A | S | CA | 0.760 |
| 6A | C | 4A | H | CA | 0.363 |
| 7A | I | 5A | I | CA | 0.188 |
| 8A | V | 6A | V | CA | 0.154 |
| 9A | A | 7A | A | CA | 0.274 |
| 10A | V | 8A | M | CA | 0.554 |
| 11A | S | 9A | D | CA | 0.410 |
| 12A | Q | 10A | E | CA | 0.428 |
| 13A | N | 11A | N | CA | 0.817 |
| 14A | M | 12A | R | CA | 0.830 |
| 15A | G | 13A | V | CA | 0.384 |
| 16A | I | 14A | I | CA | 0.371 |
| 17A | G | 15A | G | CA | 0.792 |
| 18A | K | 16A | K | CA | 0.394 |
| 19A | N | 17A | D | CA | 0.569 |
| 20A | G | 18A | N | CA | 0.671 |
| 21A | D | 19A | R | CA | 0.810 |
| 22A | L | 20A | L | CA | 1.065 |
| 24A | W | 22A | W | CA | 1.471 |
| 25A | P | | | | |
| 26A | P | 23A | H | CA | 1.968 |
| 27A | L | 24A | L | CA | 1.276 |
| 28A | R | 25A | P | CA | 1.484 |
| 29A | N | 26A | A | CA | 1.526 |
| 30A | E | 27A | D | CA | 0.796 |
| 31A | F | 28A | L | CA | 0.263 |
| 32A | K | 29A | A | CA | 0.416 |
| 33A | Y | 30A | Y | CA | 0.052 |
| 34A | F | 31A | F | CA | 0.141 |
| 35A | Q | 32A | K | CA | 0.338 |
| 36A | R | 33A | R | CA | 0.465 |
| 37A | M | 34A | V | CA | 0.509 |
| 38A | T | 35A | T | CA | 0.771 |
| 39A | T | 36A | M | CA | 0.973 |
| 41A | S | | | | |
| 42A | S | | | | |
| 43A | V | | | | |
| 44A | E | | | | |
| 45A | G | | | | |
| 46A | K | | | | |
| 47A | Q | | | | |
| 48A | N | 38A | H | CA | 1.340 |
| 49A | L | 39A | A | CA | 0.705 |
| 50A | V | 40A | I | CA | 0.796 |
| 51A | I | 41A | V | CA | 0.574 |

# Appendix part III

| | | | | | |
|---|---|---|---|---|---|
| 52A | M | 42A | M | CA | 0.349 |
| 53A | G | 43A | G | CA | 0.380 |
| 54A | R | 44A | R | CA | 0.707 |
| 55A | K | 45A | K | CA | 0.714 |
| 56A | T | 46A | T | CA | 0.236 |
| 57A | W | 47A | F | CA | 0.467 |
| 58A | F | 48A | E | CA | 1.041 |
| 59A | S | 49A | A | CA | 0.458 |
| 60A | I | 50A | I | CA | 0.804 |
| 61A | P | | | | |
| 63A | K | | | | |
| 64A | N | | | | |
| 66A | P | 53A | P | CA | 1.364 |
| 67A | L | 54A | L | CA | 1.256 |
| 68A | K | 55A | P | CA | 1.234 |
| 69A | D | 56A | G | CA | 1.284 |
| 70A | R | 57A | R | CA | 1.415 |
| 71A | I | 58A | D | CA | 0.890 |
| 72A | N | 59A | N | CA | 0.804 |
| 73A | I | 60A | V | CA | 0.730 |
| 74A | V | 61A | V | CA | 0.435 |
| 75A | L | 62A | V | CA | 0.727 |
| 76A | S | 63A | T | CA | 0.557 |
| 77A | R | 64A | G | CA | 1.083 |
| 78A | E | 65A | N | CA | 1.897 |
| 83A | P | 70A | P | CA | 1.266 |
| 84A | R | 71A | E | CA | 1.973 |
| 86A | A | 73A | C | CA | 1.420 |
| 87A | H | | | | |
| 88A | F | 74A | L | CA | 0.742 |
| 89A | L | 75A | V | CA | 0.556 |
| 90A | A | 76A | L | CA | 0.424 |
| 91A | K | 77A | H | CA | 1.213 |
| 95A | D | 81A | E | CA | 1.817 |
| 98A | R | 84A | Q | CA | 1.373 |
| 99A | L | 85A | W | CA | 1.891 |
| 103A | P | | | | |
| 104A | D | | | | |
| 107A | S | | | | |
| 108A | K | | | | |
| 109A | V | | | | |
| 111A | M | 92A | E | CA | 1.644 |
| 112A | V | 93A | V | CA | 0.467 |
| 113A | W | 94A | F | CA | 0.506 |
| 114A | I | 95A | I | CA | 0.386 |
| 115A | V | 96A | I | CA | 0.495 |
| 116A | G | 97A | G | CA | 0.293 |
| 117A | G | 98A | G | CA | 1.242 |
| 118A | S | 99A | A | CA | 0.698 |
| 119A | S | 100A | E | CA | 0.520 |
| 120A | V | 101A | L | CA | 0.714 |
| 121A | Y | 102A | F | CA | 0.531 |
| 122A | Q | 103A | R | CA | 0.503 |

# Appendix part III

| | | | | | |
|---|---|---|---|---|---|
| 123A | E | 104A | A | CA | 0.754 |
| 124A | A | 105A | T | CA | 1.066 |
| 125A | M | 106A | M | CA | 0.924 |
| 127A | Q | | | | |
| 128A | P | | | | |
| 131A | L | 110A | D | CA | 1.254 |
| 132A | R | 111A | R | CA | 0.414 |
| 133A | L | 112A | L | CA | 0.494 |
| 134A | F | 113A | Y | CA | 0.175 |
| 135A | V | 114A | V | CA | 0.135 |
| 136A | T | 115A | T | CA | 0.144 |
| 137A | R | 116A | K | CA | 0.340 |
| 138A | I | 117A | I | CA | 0.520 |
| 139A | M | 118A | F | CA | 0.551 |
| 140A | Q | 119A | A | CA | 0.295 |
| 141A | E | 120A | S | CA | 0.440 |
| 142A | F | 121A | F | CA | 0.593 |
| 143A | E | 122A | P | CA | 0.942 |
| 144A | S | 123A | G | CA | 0.318 |
| 145A | D | 124A | D | CA | 0.141 |
| 146A | T | 125A | T | CA | 0.799 |
| 147A | F | 126A | F | CA | 0.731 |
| 148A | F | 127A | Y | CA | 0.599 |
| 149A | P | 128A | P | CA | 0.210 |
| 150A | E | 129A | P | CA | 0.330 |
| 151A | I | 130A | I | CA | 0.501 |
| 152A | D | 131A | S | CA | 0.430 |
| 153A | L | 132A | D | CA | 0.948 |
| 154A | G | 133A | D | CA | 0.808 |
| 155A | K | 134A | E | CA | 0.023 |
| 156A | Y | 135A | W | CA | 0.477 |
| 157A | K | 136A | E | CA | 1.138 |
| 158A | L | 137A | I | CA | 0.873 |
| 159A | L | 138A | V | CA | 0.833 |
| 160A | P | | | | |
| 162A | Y | | | | |
| 163A | P | | | | |
| 164A | G | | | | |
| 167A | S | | | | |
| 171A | E | 145A | K | CA | 1.427 |
| | | 147A | E | | |
| | | 148A | K | | |
| 175A | I | 151A | Y | CA | 1.121 |
| 176A | K | 152A | E | CA | 1.290 |
| 177A | Y | 153A | H | CA | 0.774 |
| 178A | K | 154A | A | CA | 0.590 |
| 179A | F | 155A | F | CA | 0.320 |
| 180A | E | 156A | I | CA | 0.226 |
| 181A | V | 157A | I | CA | 0.316 |
| 182A | Y | 158A | Y | CA | 0.443 |
| 183A | E | 159A | E | CA | 0.780 |
| 184A | K | 160A | R | CA | 1.737 |
| 186A | D | | | | |

# Appendix part III

**Table 14.** Residues used for superposition between 3K45 and 3JW3

| Residue number (Fixed) | Residue name (Fixed) | Residue number (Rotated) | Residue name (Rotated) | Atom name | Deviation (Å) |
|---|---|---|---|---|---|
| | | -2A | H | | |
| | | -1A | H | | |
| | | 0A | H | | |
| 2A | R | | | | |
| 3A | P | 2A | R | CA | 1.056 |
| 4A | L | 3A | V | CA | 0.262 |
| 5A | N | 4A | S | CA | 0.675 |
| 6A | C | 5A | F | CA | 0.699 |
| 7A | I | 6A | M | CA | 0.228 |
| 8A | V | 7A | V | CA | 0.445 |
| 9A | A | 8A | A | CA | 0.164 |
| 10A | V | 9A | M | CA | 0.463 |
| 11A | S | 10A | D | CA | 0.173 |
| 12A | Q | 11A | E | CA | 0.316 |
| 13A | N | 12A | N | CA | 0.832 |
| 14A | M | 13A | R | CA | 0.750 |
| 15A | G | 14A | V | CA | 0.449 |
| 16A | I | 15A | I | CA | 0.430 |
| 17A | G | 16A | G | CA | 0.737 |
| 18A | K | 17A | K | CA | 0.594 |
| 19A | N | 18A | D | CA | 0.627 |
| 20A | G | 19A | N | CA | 0.691 |
| 21A | D | 20A | N | CA | 0.636 |
| 22A | L | 21A | L | CA | 1.094 |
| 23A | P | 22A | P | CA | 1.815 |
| 24A | W | 23A | W | CA | 1.753 |
| 25A | P | | | | |
| 26A | P | 24A | R | CA | 1.245 |
| 27A | L | 25A | L | CA | 0.129 |
| 28A | R | 26A | P | CA | 0.412 |
| 29A | N | 27A | S | CA | 0.464 |
| 30A | E | 28A | E | CA | 0.447 |
| 31A | F | 29A | L | CA | 0.511 |
| 32A | K | 30A | Q | CA | 0.463 |
| 33A | Y | 31A | Y | CA | 0.187 |
| 34A | F | 32A | V | CA | 0.441 |
| 35A | Q | 33A | K | CA | 0.441 |
| 36A | R | 34A | K | CA | 0.345 |
| 37A | M | 35A | T | CA | 0.425 |
| 38A | T | 36A | T | CA | 0.796 |
| 39A | T | | | | |
| 42A | S | | | | |
| 43A | V | | | | |
| 44A | E | | | | |
| 45A | G | | | | |
| 46A | K | | | | |
| 47A | Q | | | | |
| 48A | N | 39A | H | CA | 1.230 |
| 49A | L | 40A | P | CA | 0.811 |

| 50A | V | 41A | L | CA | 0.800 |
|---|---|---|---|---|---|
| 51A | I | 42A | I | CA | 0.842 |
| 52A | M | 43A | M | CA | 0.629 |
| 53A | G | 44A | G | CA | 0.592 |
| 54A | R | 45A | R | CA | 0.540 |
| 55A | K | 46A | K | CA | 0.893 |
| 56A | T | 47A | N | CA | 0.949 |
| 57A | W | 48A | Y | CA | 0.934 |
| 58A | F | 49A | E | CA | 1.052 |
| 59A | S | 50A | A | CA | 1.036 |
| 60A | I | 51A | I | CA | 1.177 |
| 61A | P | | | | |
| 63A | K | | | | |
| 64A | N | | | | |
| 66A | P | 54A | P | CA | 1.396 |
| 67A | L | 55A | L | CA | 1.318 |
| 68A | K | 56A | P | CA | 0.968 |
| 69A | D | 57A | G | CA | 0.334 |
| 70A | R | 58A | R | CA | 0.569 |
| 71A | I | 59A | R | CA | 0.776 |
| 72A | N | 60A | N | CA | 0.888 |
| 73A | I | 61A | I | CA | 0.902 |
| 74A | V | 62A | I | CA | 0.709 |
| 75A | L | 63A | V | CA | 0.642 |
| 76A | S | 64A | T | CA | 0.866 |
| 77A | R | 65A | R | CA | 0.591 |
| 78A | E | 66A | N | CA | 1.020 |
| | | 67A | E | | |
| 82A | P | | | | |
| 83A | P | 71A | V | CA | 1.359 |
| 84A | R | 72A | E | CA | 1.980 |
| 86A | A | 74A | C | CA | 0.739 |
| 87A | H | | | | |
| 88A | F | 75A | E | CA | 1.397 |
| 89A | L | 76A | V | CA | 0.739 |
| 90A | A | 77A | A | CA | 0.705 |
| 91A | K | 78A | H | CA | 0.705 |
| 92A | S | 79A | S | CA | 0.770 |
| 93A | L | 80A | V | CA | 1.674 |
| 94A | D | 81A | E | CA | 1.627 |
| 95A | D | 82A | E | CA | 0.979 |
| 96A | A | 83A | V | CA | 0.730 |
| 97A | L | 84A | F | CA | 0.959 |
| 100A | I | 87A | C | CA | 1.126 |
| 101A | E | | | | |
| 103A | P | | | | |
| 104A | D | | | | |
| 105A | L | | | | |
| 107A | S | | | | |
| 108A | K | | | | |
| 112A | V | 93A | I | CA | 0.815 |
| 113A | W | 94A | F | CA | 0.678 |
| 114A | I | 95A | I | CA | 0.700 |

# Appendix part III

| 115A | V | 96A | I | CA | 0.818 |
|---|---|---|---|---|---|
| 116A | G | 97A | G | CA | 0.939 |
| 117A | G | 98A | G | CA | 0.717 |
| 118A | S | 99A | A | CA | 0.739 |
| 119A | S | 100A | Q | CA | 0.624 |
| 120A | V | 101A | I | CA | 0.723 |
| 121A | Y | 102A | Y | CA | 0.857 |
| 122A | Q | 103A | D | CA | 1.020 |
| 123A | E | 104A | L | CA | 1.171 |
| 124A | A | 105A | F | CA | 0.950 |
| 125A | M | 106A | L | CA | 0.822 |
| 127A | Q | | | | |
| 128A | P | | | | |
| 131A | L | 110A | D | CA | 1.419 |
| 132A | R | 111A | K | CA | 0.784 |
| 133A | L | 112A | L | CA | 1.061 |
| 134A | F | 113A | Y | CA | 0.257 |
| 135A | V | 114A | I | CA | 0.218 |
| 136A | T | 115A | T | CA | 0.241 |
| 137A | R | 116A | K | CA | 0.391 |
| 138A | I | 117A | I | CA | 0.590 |
| 139A | M | 118A | H | CA | 1.051 |
| 140A | Q | 119A | H | CA | 1.054 |
| 141A | E | 120A | A | CA | 0.726 |
| 142A | F | 121A | F | CA | 0.540 |
| 143A | E | 122A | E | CA | 0.794 |
| 144A | S | 123A | G | CA | 1.006 |
| 145A | D | 124A | D | CA | 0.135 |
| 146A | T | 125A | T | CA | 1.049 |
| 147A | F | 126A | F | CA | 0.802 |
| 148A | F | 127A | F | CA | 0.492 |
| 149A | P | 128A | P | CA | 0.853 |
| 150A | E | 129A | E | CA | 1.117 |
| 154A | G | 133A | T | CA | 0.931 |
| 157A | K | 136A | K | CA | 1.835 |
| 158A | L | 137A | E | CA | 1.539 |
| 159A | L | 138A | V | CA | 1.597 |
| 160A | P | | | | |
| 162A | Y | | | | |
| 163A | P | | | | |
| 164A | G | | | | |
| 165A | V | | | | |
| 166A | L | | | | |
| 169A | V | 142A | K | CA | 1.992 |
| | | 144A | L | | |
| 171A | E | 145A | T | CA | 1.766 |
| | | 147A | E | | |
| | | 148A | K | | |
| | | 149A | N | | |
| 173A | K | 150A | P | CA | 1.542 |
| 174A | G | | | | |
| 175A | I | 151A | Y | CA | 0.410 |
| 176A | K | 152A | T | CA | 0.338 |

| | | | | | |
|---|---|---|---|---|---|
| **177A** | **Y** | **153A** | **Y** | **CA** | **0.338** |
| **178A** | **K** | **154A** | **Y** | **CA** | **0.515** |
| **179A** | **F** | **155A** | **Y** | **CA** | **0.534** |
| **180A** | **E** | **156A** | **H** | **CA** | **0.601** |
| **181A** | **V** | **157A** | **V** | **CA** | **0.755** |
| **182A** | **Y** | **158A** | **Y** | **CA** | **1.247** |
| **183A** | **E** | **159A** | **E** | **CA** | **1.764** |
| **184A** | **K** | **160A** | **K** | **CA** | **1.956** |

# Appendix part III

**Table 15.** Residues used for superposition between 3K45 and 2QK8

| Residue number (Fixed) | Residue name (Fixed) | Residue number (Rotated) | Residue name (Rotated) | Atom name | Deviation (Å) |
|---|---|---|---|---|---|
| 1A | V | | | | |
| 2A | R | 1A | M | CA | 0.279 |
| 3A | P | 2A | I | CA | 0.724 |
| 4A | L | 3A | V | CA | 0.589 |
| 5A | N | 4A | S | CA | 1.011 |
| 6A | C | 5A | F | CA | 0.875 |
| 7A | I | 6A | M | CA | 0.637 |
| 8A | V | 7A | V | CA | 0.520 |
| 9A | A | 8A | A | CA | 0.345 |
| 10A | V | 9A | M | CA | 0.787 |
| 11A | S | 10A | D | CA | 0.459 |
| 12A | Q | 11A | E | CA | 0.412 |
| 13A | N | 12A | N | CA | 0.548 |
| 14A | M | 13A | R | CA | 0.645 |
| 15A | G | 14A | V | CA | 0.550 |
| 16A | I | 15A | I | CA | 0.418 |
| 17A | G | 16A | G | CA | 0.588 |
| 18A | K | 17A | K | CA | 0.502 |
| 19A | N | 18A | D | CA | 0.287 |
| 20A | G | 19A | N | CA | 0.261 |
| 21A | D | 20A | N | CA | 1.000 |
| 22A | L | 21A | L | CA | 0.891 |
| 23A | P | 22A | P | CA | 1.295 |
| 24A | W | 23A | W | CA | 1.664 |
| 25A | P | | | | |
| 26A | P | 24A | R | CA | 1.070 |
| 27A | L | 25A | L | CA | 0.336 |
| 28A | R | 26A | P | CA | 0.202 |
| 29A | N | 27A | S | CA | 0.259 |
| 30A | E | 28A | E | CA | 0.373 |
| 31A | F | 29A | L | CA | 0.453 |
| 32A | K | 30A | Q | CA | 0.351 |
| 33A | Y | 31A | Y | CA | 0.448 |
| 34A | F | 32A | V | CA | 0.438 |
| 35A | Q | 33A | K | CA | 0.413 |
| 36A | R | 34A | K | CA | 0.393 |
| 37A | M | 35A | T | CA | 0.647 |
| 38A | T | 36A | T | CA | 0.485 |
| 39A | T | | | | |
| 42A | S | | | | |
| 43A | V | | | | |
| 44A | E | | | | |
| 45A | G | | | | |
| 46A | K | | | | |
| 47A | Q | | | | |
| 48A | N | 39A | H | CA | 1.359 |
| 49A | L | 40A | P | CA | 1.038 |
| 50A | V | 41A | L | CA | 1.025 |
| 51A | I | 42A | I | CA | 0.785 |

# Appendix part III

| | | | | | |
|---|---|---|---|---|---|
| 52A | M | 43A | M | CA | 0.635 |
| 53A | G | 44A | G | CA | 0.528 |
| 54A | R | 45A | R | CA | 0.972 |
| 55A | K | 46A | K | CA | 0.961 |
| 56A | T | 47A | N | CA | 0.845 |
| 57A | W | 48A | Y | CA | 1.152 |
| 58A | F | 49A | E | CA | 1.696 |
| 59A | S | 50A | A | CA | 1.502 |
| 60A | I | 51A | I | CA | 1.174 |
| 62A | E | | | | |
| 63A | K | | | | |
| 64A | N | | | | |
| 66A | P | 54A | P | CA | 1.418 |
| 67A | L | 55A | L | CA | 0.664 |
| 68A | K | 56A | P | CA | 0.183 |
| 69A | D | 57A | G | CA | 0.757 |
| 70A | R | 58A | R | CA | 0.673 |
| 71A | I | 59A | R | CA | 0.990 |
| 72A | N | 60A | N | CA | 1.162 |
| 73A | I | 61A | I | CA | 0.811 |
| 74A | V | 62A | I | CA | 0.808 |
| 75A | L | 63A | V | CA | 0.613 |
| 76A | S | 64A | T | CA | 0.982 |
| 77A | R | 65A | R | CA | 1.398 |
| 78A | E | 66A | N | CA | 1.715 |
| | | 67A | E | | |
| 82A | P | | | | |
| 83A | P | 71A | V | CA | 1.227 |
| 84A | R | 72A | E | CA | 1.495 |
| 85A | G | 73A | G | CA | 1.618 |
| 86A | A | 74A | C | CA | 0.732 |
| 87A | H | | | | |
| 88A | F | 75A | E | CA | 1.774 |
| 89A | L | 76A | V | CA | 1.018 |
| 90A | A | 77A | A | CA | 0.773 |
| 91A | K | 78A | H | CA | 0.633 |
| 92A | S | 79A | S | CA | 1.328 |
| 93A | L | 80A | V | CA | 1.793 |
| 95A | D | 82A | E | CA | 1.677 |
| 96A | A | 83A | V | CA | 0.843 |
| 97A | L | 84A | F | CA | 1.380 |
| 100A | I | 87A | C | CA | 1.268 |
| 101A | E | | | | |
| 103A | P | | | | |
| 104A | D | | | | |
| 105A | L | | | | |
| 107A | S | | | | |
| 108A | K | | | | |
| 112A | V | 93A | I | CA | 1.391 |
| 113A | W | 94A | F | CA | 0.886 |
| 114A | I | 95A | I | CA | 0.910 |
| 115A | V | 96A | F | CA | 0.562 |
| 116A | G | 97A | G | CA | 1.074 |

# Appendix part III

| | | | | | |
|---|---|---|---|---|---|
| 117A | G | 98A | G | CA | 1.054 |
| 118A | S | 99A | A | CA | 0.892 |
| 119A | S | 100A | Q | CA | 0.601 |
| 120A | V | 101A | I | CA | 0.581 |
| 121A | Y | 102A | Y | CA | 0.982 |
| 122A | Q | 103A | D | CA | 1.442 |
| 123A | E | 104A | L | CA | 1.481 |
| 124A | A | 105A | F | CA | 1.286 |
| 125A | M | 106A | L | CA | 1.606 |
| 126A | N | | | | |
| 127A | Q | 107A | P | CA | 1.513 |
| 128A | P | | | | |
| 131A | L | 110A | D | CA | 1.642 |
| 132A | R | 111A | K | CA | 0.968 |
| 133A | L | 112A | L | CA | 0.983 |
| 134A | F | 113A | Y | CA | 0.369 |
| 135A | V | 114A | I | CA | 0.276 |
| 136A | T | 115A | T | CA | 0.409 |
| 137A | R | 116A | K | CA | 0.517 |
| 138A | I | 117A | I | CA | 0.445 |
| 139A | M | 118A | H | CA | 1.127 |
| 140A | Q | 119A | H | CA | 1.328 |
| 141A | E | 120A | A | CA | 1.187 |
| 142A | F | 121A | F | CA | 1.326 |
| 143A | E | 122A | E | CA | 1.613 |
| 145A | D | 124A | D | CA | 0.486 |
| 146A | T | 125A | T | CA | 0.814 |
| 147A | F | 126A | F | CA | 0.640 |
| 148A | F | 127A | F | CA | 0.408 |
| 149A | P | 128A | P | CA | 1.237 |
| 150A | E | 129A | E | CA | 1.339 |
| 152A | D | 131A | D | CA | 1.158 |
| 153A | L | 132A | M | CA | 1.798 |
| 154A | G | 133A | T | CA | 1.384 |
| 157A | K | 136A | K | CA | 1.675 |
| 158A | L | 137A | E | CA | 1.800 |
| 159A | L | 138A | V | CA | 1.695 |
| 160A | P | | | | |
| 162A | Y | | | | |
| 163A | P | | | | |
| 164A | G | | | | |
| 165A | V | | | | |
| 166A | L | | | | |
| | | 144A | L | | |
| | | 147A | E | | |
| | | 149A | N | | |
| 175A | I | 151A | Y | CA | 0.954 |
| 176A | K | 152A | T | CA | 0.970 |
| 177A | Y | 153A | Y | CA | 0.489 |
| 178A | K | 154A | Y | CA | 0.873 |
| 179A | F | 155A | Y | CA | 0.443 |
| 180A | E | 156A | H | CA | 0.452 |
| 181A | V | 157A | V | CA | 0.470 |

| | | | | | |
|---|---|---|---|---|---|
| **182A** | **Y** | **158A** | **Y** | **CA** | **1.375** |
| **183A** | **E** | **159A** | **E** | **CA** | **1.493** |
| **184A** | **K** | **160A** | **K** | **CA** | **1.490** |

# Appendix part III

**Table 16.** Residues used for superposition between 3K45 and 2ZZA

| Residue number (Fixed) | Residue name (Fixed) | Residue number (Rotated) | Residue name (Rotated) | Atom name | Deviation (Å) |
|---|---|---|---|---|---|
| 1A | V | | | | |
| 2A | R | 1A | V | CA | 0.574 |
| 3A | P | 2A | I | CA | 0.517 |
| 4A | L | 3A | V | CA | 0.991 |
| 5A | N | 4A | S | CA | 1.286 |
| 6A | C | 5A | M | CA | 0.564 |
| 7A | I | 6A | I | CA | 0.513 |
| 8A | V | 7A | A | CA | 0.135 |
| 9A | A | 8A | A | CA | 0.152 |
| 10A | V | 9A | L | CA | 0.219 |
| 11A | S | 10A | A | CA | 0.618 |
| 12A | Q | 11A | N | CA | 1.417 |
| 13A | N | 12A | N | CA | 1.414 |
| 14A | M | 13A | R | CA | 0.918 |
| 15A | G | 14A | V | CA | 0.424 |
| 16A | I | 15A | I | CA | 0.616 |
| 17A | G | 16A | G | CA | 0.848 |
| 18A | K | 17A | L | CA | 0.836 |
| 19A | N | 18A | D | CA | 1.222 |
| 20A | G | 19A | N | CA | 1.016 |
| 21A | D | 20A | K | CA | 0.707 |
| 22A | L | 21A | M | CA | 0.974 |
| 23A | P | 22A | P | CA | 1.659 |
| 24A | W | 23A | W | CA | 1.702 |
| 25A | P | | | | |
| 26A | P | 24A | H | CA | 1.360 |
| 27A | L | 25A | L | CA | 0.587 |
| 28A | R | 26A | P | CA | 0.191 |
| 29A | N | 27A | A | CA | 0.213 |
| 30A | E | 28A | E | CA | 0.136 |
| 31A | F | 29A | L | CA | 0.394 |
| 32A | K | 30A | Q | CA | 0.514 |
| 33A | Y | 31A | L | CA | 0.436 |
| 34A | F | 32A | F | CA | 0.397 |
| 35A | Q | 33A | K | CA | 0.554 |
| 36A | R | 34A | R | CA | 0.894 |
| 37A | M | 35A | A | CA | 0.996 |
| 38A | T | 36A | T | CA | 0.301 |
| 39A | T | | | | |
| 42A | S | | | | |
| 43A | V | | | | |
| 44A | E | | | | |
| 45A | G | | | | |
| 46A | K | | | | |
| 47A | Q | | | | |
| 48A | N | 39A | K | CA | 1.331 |
| 49A | L | 40A | P | CA | 0.573 |
| 50A | V | 41A | I | CA | 0.524 |
| 51A | I | 42A | V | CA | 0.575 |

# Appendix part III

| | | | | | |
|---|---|---|---|---|---|
| 52A | M | 43A | M | CA | 0.466 |
| 53A | G | 44A | G | CA | 0.538 |
| 54A | R | 45A | R | CA | 0.515 |
| 55A | K | 46A | N | CA | 0.811 |
| 56A | T | 47A | T | CA | 0.641 |
| 57A | W | 48A | F | CA | 0.579 |
| 58A | F | 49A | E | CA | 0.950 |
| 59A | S | 50A | S | CA | 0.857 |
| 60A | I | 51A | I | CA | 0.996 |
| 62A | E | | | | |
| 63A | K | | | | |
| 64A | N | | | | |
| 66A | P | 54A | P | CA | 1.389 |
| 67A | L | 55A | L | CA | 0.606 |
| 68A | K | 56A | P | CA | 0.367 |
| 69A | D | 57A | G | CA | 0.681 |
| 70A | R | 58A | R | CA | 0.617 |
| 71A | I | 59A | L | CA | 0.641 |
| 72A | N | 60A | N | CA | 0.667 |
| 73A | I | 61A | I | CA | 0.566 |
| 74A | V | 62A | V | CA | 0.551 |
| 75A | L | 63A | L | CA | 0.350 |
| 76A | S | 64A | S | CA | 0.218 |
| 77A | R | 65A | R | CA | 0.439 |
| 78A | E | 66A | Q | CA | 0.865 |
| | | 67A | T | | |
| 82A | P | | | | |
| 83A | P | 71A | P | CA | 1.291 |
| 84A | R | 72A | E | CA | 0.916 |
| 85A | G | 73A | G | CA | 1.404 |
| 86A | A | 74A | V | CA | 0.731 |
| 87A | H | | | | |
| 88A | F | 75A | T | CA | 1.400 |
| 89A | L | 76A | V | CA | 0.980 |
| 90A | A | 77A | V | CA | 0.601 |
| 91A | K | 78A | A | CA | 0.380 |
| 92A | S | 79A | T | CA | 0.945 |
| 93A | L | 80A | L | CA | 0.916 |
| 94A | D | 81A | E | CA | 1.335 |
| 95A | D | 82A | D | CA | 1.401 |
| 96A | A | 83A | A | CA | 1.005 |
| 97A | L | 84A | V | CA | 0.970 |
| 101A | E | | | | |
| 103A | P | | | | |
| 104A | D | | | | |
| 105A | L | | | | |
| 107A | S | | | | |
| 108A | K | | | | |
| 112A | V | 93A | L | CA | 1.230 |
| 113A | W | 94A | M | CA | 0.654 |
| 114A | I | 95A | I | CA | 0.758 |
| 115A | V | 96A | I | CA | 0.716 |
| 116A | G | 97A | G | CA | 1.313 |

| | | | | | |
|---|---|---|---|---|---|
| 117A | G | 98A | G | CA | 0.580 |
| 118A | S | 99A | A | CA | 0.443 |
| 119A | S | 100A | T | CA | 0.245 |
| 120A | V | 101A | I | CA | 0.558 |
| 121A | Y | 102A | Y | CA | 0.600 |
| 122A | Q | 103A | N | CA | 0.706 |
| 123A | E | 104A | Q | CA | 0.963 |
| 124A | A | 105A | C | CA | 1.178 |
| 125A | M | 106A | L | CA | 0.782 |
| 127A | Q | | | | |
| 129A | G | | | | |
| 131A | L | 110A | D | CA | 1.282 |
| 132A | R | 111A | R | CA | 0.572 |
| 133A | L | 112A | L | CA | 0.366 |
| 134A | F | 113A | Y | CA | 0.303 |
| 135A | V | 114A | L | CA | 0.336 |
| 136A | T | 115A | T | CA | 0.170 |
| 137A | R | 116A | H | CA | 0.319 |
| 138A | I | 117A | I | CA | 0.273 |
| 139A | M | 118A | E | CA | 1.167 |
| 140A | Q | 119A | L | CA | 1.691 |
| 141A | E | 120A | T | CA | 0.858 |
| 142A | F | 121A | T | CA | 1.121 |
| 143A | E | 122A | E | CA | 1.037 |
| 144A | S | 123A | G | CA | 1.017 |
| 145A | D | 124A | D | CA | 1.043 |
| 146A | T | 125A | T | CA | 1.343 |
| 147A | F | 126A | W | CA | 0.904 |
| 148A | F | 127A | F | CA | 0.686 |
| 149A | P | 128A | P | CA | 0.581 |
| 150A | E | 129A | D | CA | 0.642 |
| 151A | I | 130A | Y | CA | 1.629 |
| | | 131A | E | | |
| 153A | L | | | | |
| 156A | Y | 135A | W | CA | 1.915 |
| 157A | K | 136A | Q | CA | 1.878 |
| 160A | P | | | | |
| 162A | Y | | | | |
| 163A | P | | | | |
| 164A | G | | | | |
| 165A | V | | | | |
| 166A | L | | | | |
| 170A | Q | 143A | Y | CA | 1.039 |
| | | 146A | D | | |
| | | 147A | D | | |
| | | 148A | K | | |
| 175A | I | 151A | H | CA | 0.833 |
| 176A | K | 152A | N | CA | 0.552 |
| 177A | Y | 153A | Y | CA | 0.228 |
| 178A | K | 154A | R | CA | 0.746 |
| 179A | F | 155A | F | CA | 0.272 |
| 180A | E | 156A | S | CA | 0.244 |
| 181A | V | 157A | L | CA | 0.801 |

| | | | | | |
|---|---|---|---|---|---|
| **182A** | **Y** | **158A** | **L** | **CA** | **0.895** |
| **183A** | **E** | **159A** | **E** | **CA** | **1.048** |
| **184A** | **K** | **160A** | **R** | **CA** | **1.303** |
| **186A** | **D** | | | | |

# Appendix part III

**c) Core extension:**

the following tables represent the results of the superposition has been undertaken using the core regions and by extending the core in the case having a new regions contain better distance deviation , and the deletion of the absent residues corresponding to the fixed molecule.

**Table 17.** Residues used for superposition between 3K45 and 1U70

| Residue number (Fixed) | Residue name (Fixed) | Residue number (Rotated) | Residue name (Rotated) | Atom name | Deviation (Å) |
|---|---|---|---|---|---|
| 4A | L | 4A | L | CA | 0.467 |
| 5A | N | 5A | N | CA | 0.129 |
| 6A | C | 6A | C | CA | 0.110 |
| 7A | I | 7A | I | CA | 0.125 |
| 8A | V | 8A | V | CA | 0.235 |
| 9A | A | 9A | A | CA | 0.221 |
| 10A | V | 10A | V | CA | 0.256 |
| 11A | S | 11A | S | CA | 0.078 |
| 12A | Q | 12A | Q | CA | 0.132 |
| 14A | M | 14A | M | CA | 0.243 |
| 15A | G | 15A | G | CA | 0.360 |
| 16A | I | 16A | I | CA | 0.442 |
| 17A | G | 17A | G | CA | 1.275 |
| 18A | K | 18A | K | CA | 0.456 |
| 19A | N | 19A | N | CA | 0.493 |
| 20A | G | 20A | G | CA | 1.737 |
| 21A | D | 21A | D | CA | 0.952 |
| 22A | L | 22A | R | CA | 0.458 |
| 26A | P | 26A | P | CA | 0.556 |
| 27A | L | 27A | L | CA | 0.341 |
| 28A | R | 28A | R | CA | 0.459 |
| 29A | N | 29A | N | CA | 0.158 |
| 30A | E | 30A | E | CA | 0.364 |
| 31A | F | 31A | F | CA | 0.324 |
| 32A | K | 32A | K | CA | 0.171 |
| 33A | Y | 33A | Y | CA | 0.211 |
| 34A | F | 34A | F | CA | 0.321 |
| 35A | Q | 35A | Q | CA | 0.297 |
| 36A | R | 36A | R | CA | 0.315 |
| 37A | M | 37A | M | CA | 0.427 |
| 38A | T | 38A | T | CA | 0.112 |
| 48A | N | 48A | N | CA | 0.209 |
| 49A | L | 49A | L | CA | 0.247 |
| 50A | V | 50A | V | CA | 0.070 |
| 51A | I | 51A | I | CA | 0.134 |
| 52A | M | 52A | M | CA | 0.136 |
| 53A | G | 53A | G | CA | 0.247 |
| 54A | R | 54A | R | CA | 0.253 |
| 55A | K | 55A | K | CA | 0.269 |

| 56A | T | 56A | T | CA | 0.218 |
|---|---|---|---|---|---|
| 57A | W | 57A | W | CA | 0.069 |
| 58A | F | 58A | F | CA | 0.362 |
| 59A | S | 59A | S | CA | 0.450 |
| 60A | I | 60A | I | CA | 0.393 |
| 66A | P | 66A | P | CA | 0.289 |
| 67A | L | 67A | L | CA | 0.343 |
| 68A | K | 68A | K | CA | 0.385 |
| 69A | D | 69A | D | CA | 0.357 |
| 70A | R | 70A | R | CA | 0.171 |
| 71A | I | 71A | I | CA | 0.241 |
| 72A | N | 72A | N | CA | 0.245 |
| 73A | I | 73A | I | CA | 0.153 |
| 74A | V | 74A | V | CA | 0.203 |
| 75A | L | 75A | L | CA | 0.207 |
| 76A | S | 76A | S | CA | 0.362 |
| 77A | R | 77A | R | CA | 0.544 |
| 88A | F | 88A | F | CA | 0.200 |
| 89A | L | 89A | L | CA | 0.271 |
| 90A | A | 90A | A | CA | 0.246 |
| 91A | K | 91A | K | CA | 0.211 |
| 112A | V | 112A | V | CA | 0.025 |
| 113A | W | 113A | W | CA | 0.078 |
| 114A | I | 114A | I | CA | 0.320 |
| 115A | V | 115A | V | CA | 0.341 |
| 116A | G | 116A | G | CA | 0.651 |
| 117A | G | 117A | G | CA | 1.584 |
| 118A | S | 118A | S | CA | 1.019 |
| 119A | S | 119A | S | CA | 1.005 |
| 120A | V | 120A | V | CA | 0.743 |
| 121A | Y | 121A | Y | CA | 0.495 |
| 122A | Q | 122A | Q | CA | 0.760 |
| 131A | L | 131A | L | CA | 0.540 |
| 132A | R | 132A | R | CA | 0.317 |
| 133A | L | 133A | L | CA | 0.316 |
| 134A | F | 134A | F | CA | 0.163 |
| 135A | V | 135A | V | CA | 0.166 |
| 136A | T | 136A | T | CA | 0.195 |
| 137A | R | 137A | R | CA | 0.280 |
| 138A | I | 138A | I | CA | 0.161 |
| 145A | D | 145A | D | CA | 0.433 |
| 146A | T | 146A | T | CA | 0.875 |
| 147A | F | 147A | F | CA | 0.686 |
| 148A | F | 148A | F | CA | 0.353 |
| 149A | P | 149A | P | CA | 0.404 |
| 175A | I | 175A | I | CA | 0.203 |
| 176A | K | 176A | K | CA | 0.172 |
| 177A | Y | 177A | Y | CA | 0.279 |
| 178A | K | 178A | K | CA | 0.108 |
| 179A | F | 179A | F | CA | 0.098 |
| 180A | E | 180A | E | CA | 0.086 |
| 181A | V | 181A | V | CA | 0.224 |
| 182A | Y | 182A | Y | CA | 0.204 |

| | | | | | |
|---|---|---|---|---|---|
| **183A** | **E** | **183A** | **E** | **CA** | **0.243** |
| **184A** | **K** | **184A** | **K** | **CA** | **0.459** |

# Appendix part III

**Table 18.** Residues used for superposition between 3K45 and 1DR1

| Residue number (Fixed) | Residue name (Fixed) | Residue number (Rotated) | Residue name (Rotated) | Atom name | Deviation (Å) |
|---|---|---|---|---|---|
| 4A | L | 4A | L | CA | 0.705 |
| 5A | N | 5A | N | CA | 0.490 |
| 6A | C | 6A | S | CA | 0.584 |
| 7A | I | 7A | I | CA | 0.160 |
| 8A | V | 8A | V | CA | 0.141 |
| 9A | A | 9A | A | CA | 0.163 |
| 10A | V | 10A | V | CA | 0.357 |
| 11A | S | 11A | C | CA | 0.197 |
| 12A | Q | 12A | Q | CA | 0.907 |
| 14A | M | 14A | M | CA | 0.865 |
| 15A | G | 15A | G | CA | 0.901 |
| 16A | I | 16A | I | CA | 0.679 |
| 17A | G | 17A | G | CA | 1.051 |
| 18A | K | 18A | K | CA | 1.050 |
| 19A | N | 19A | D | CA | 1.366 |
| 20A | G | 20A | G | CA | 1.412 |
| 21A | D | 21A | N | CA | 0.959 |
| 22A | L | 22A | L | CA | 0.661 |
| 26A | P | 26A | P | CA | 0.934 |
| 27A | L | 27A | L | CA | 0.822 |
| 28A | R | 28A | R | CA | 1.236 |
| 29A | N | 29A | N | CA | 0.720 |
| 30A | E | 30A | E | CA | 0.412 |
| 31A | F | 31A | Y | CA | 0.468 |
| 32A | K | 32A | K | CA | 0.228 |
| 33A | Y | 33A | Y | CA | 0.139 |
| 34A | F | 34A | F | CA | 0.168 |
| 35A | Q | 35A | Q | CA | 0.291 |
| 36A | R | 36A | R | CA | 0.310 |
| 37A | M | 37A | M | CA | 0.282 |
| 38A | T | 38A | T | CA | 0.219 |
| 48A | N | 48A | N | CA | 0.500 |
| 49A | L | 49A | A | CA | 0.532 |
| 50A | V | 50A | V | CA | 0.210 |
| 51A | I | 51A | I | CA | 0.209 |
| 52A | M | 52A | M | CA | 0.279 |
| 53A | G | 53A | G | CA | 0.282 |
| 54A | R | 54A | K | CA | 0.652 |
| 55A | K | 55A | K | CA | 0.612 |
| 56A | T | 56A | T | CA | 0.279 |
| 57A | W | 57A | W | CA | 0.299 |
| 58A | F | 58A | F | CA | 0.644 |
| 59A | S | 59A | S | CA | 0.494 |
| 60A | I | 60A | I | CA | 0.490 |
| 66A | P | 66A | P | CA | 0.430 |
| 67A | L | 67A | L | CA | 0.500 |
| 68A | K | 68A | K | CA | 0.325 |
| 69A | D | 69A | D | CA | 0.470 |

# Appendix part III

| 70A | R | 70A | R | CA | 0.404 |
|---|---|---|---|---|---|
| 71A | I | 71A | I | CA | 0.588 |
| 72A | N | 72A | N | CA | 0.583 |
| 73A | I | 73A | I | CA | 0.458 |
| 74A | V | 74A | V | CA | 0.499 |
| 75A | L | 75A | L | CA | 0.372 |
| 76A | S | 76A | S | CA | 0.643 |
| 77A | R | 77A | R | CA | 0.823 |
| 88A | F | 88A | Y | CA | 0.614 |
| 89A | L | 89A | L | CA | 0.549 |
| 90A | A | 90A | S | CA | 0.238 |
| 91A | K | 91A | K | CA | 0.643 |
| 112A | V | 112A | V | CA | 0.313 |
| 113A | W | 113A | W | CA | 0.118 |
| 114A | I | 114A | I | CA | 0.365 |
| 115A | V | 115A | V | CA | 0.471 |
| 116A | G | 116A | G | CA | 1.143 |
| 117A | G | 117A | G | CA | 0.924 |
| 118A | S | 118A | T | CA | 1.357 |
| 119A | S | 119A | A | CA | 1.539 |
| 120A | V | 120A | V | CA | 0.785 |
| 121A | Y | 121A | Y | CA | 0.822 |
| 122A | Q | 122A | K | CA | 1.382 |
| 131A | L | 131A | H | CA | 1.335 |
| 132A | R | 132A | R | CA | 1.083 |
| 133A | L | 133A | L | CA | 0.703 |
| 134A | F | 134A | F | CA | 0.319 |
| 135A | V | 135A | V | CA | 0.161 |
| 136A | T | 136A | T | CA | 0.238 |
| 137A | R | 137A | R | CA | 0.425 |
| 138A | I | 138A | I | CA | 0.406 |
| 145A | D | 145A | D | CA | 0.752 |
| 146A | T | 146A | T | CA | 1.538 |
| 147A | F | 147A | F | CA | 1.123 |
| 148A | F | 148A | F | CA | 0.782 |
| 149A | P | 149A | P | CA | 1.144 |
| 175A | I | 175A | I | CA | 0.746 |
| 176A | K | 176A | Q | CA | 0.551 |
| 177A | Y | 177A | Y | CA | 0.660 |
| 178A | K | 178A | K | CA | 1.094 |
| 179A | F | 179A | F | CA | 0.736 |
| 180A | E | 180A | E | CA | 0.633 |
| 181A | V | 181A | V | CA | 0.401 |
| 182A | Y | 182A | Y | CA | 0.219 |
| 183A | E | 183A | Q | CA | 0.501 |
| 184A | K | 184A | K | CA | 0.528 |

# Appendix part III

**Table 19.** Residues used for superposition between 3K45 and 4H95

| Residue number (Fixed) | Residue name (Fixed) | Residue number (Rotated) | Residue name (Rotated) | Atom name | Deviation (Å) |
|---|---|---|---|---|---|
| 4A | L | 6A | V | CA | 1.361 |
| 5A | N | 7A | A | CA | 0.469 |
| 6A | C | 8A | I | CA | 0.218 |
| 7A | I | 9A | I | CA | 0.215 |
| 8A | V | 10A | V | CA | 0.237 |
| 9A | A | 11A | A | CA | 0.229 |
| 10A | V | 12A | A | CA | 0.381 |
| 11A | S | 13A | L | CA | 0.288 |
| 12A | Q | 14A | K | CA | 1.369 |
| 14A | M | 17A | L | CA | 0.899 |
| 15A | G | 18A | G | CA | 0.219 |
| 16A | I | 19A | I | CA | 0.565 |
| 17A | G | 20A | G | CA | 1.098 |
| 18A | K | 21A | Y | CA | 0.795 |
| 19A | N | 22A | K | CA | 0.866 |
| 20A | G | 23A | G | CA | 0.891 |
| 21A | D | 24A | K | CA | 0.844 |
| 22A | L | 25A | M | CA | 0.645 |
| 26A | P | 28A | R | CA | 0.800 |
| 27A | L | 29A | L | CA | 0.562 |
| 28A | R | 30A | R | CA | 1.297 |
| 29A | N | 31A | K | CA | 0.612 |
| 30A | E | 32A | E | CA | 0.453 |
| 31A | F | 33A | I | CA | 0.448 |
| 32A | K | 34A | R | CA | 0.402 |
| 33A | Y | 35A | Y | CA | 0.451 |
| 34A | F | 36A | F | CA | 0.474 |
| 35A | Q | 37A | K | CA | 0.480 |
| 36A | R | 38A | D | CA | 0.625 |
| 37A | M | 39A | V | CA | 0.611 |
| 38A | T | 40A | T | CA | 0.519 |
| 48A | N | 50A | N | CA | 0.149 |
| 49A | L | 51A | A | CA | 0.185 |
| 50A | V | 52A | V | CA | 0.360 |
| 51A | I | 53A | I | CA | 0.317 |
| 52A | M | 54A | M | CA | 0.151 |
| 53A | G | 55A | G | CA | 0.197 |
| 54A | R | 56A | R | CA | 0.429 |
| 55A | K | 57A | K | CA | 0.559 |
| 56A | T | 58A | T | CA | 0.553 |
| 57A | W | 59A | W | CA | 0.495 |
| 58A | F | 60A | E | CA | 0.571 |
| 59A | S | 61A | S | CA | 0.625 |
| 60A | I | 62A | I | CA | 0.591 |
| 66A | P | 68A | P | CA | 0.494 |
| 67A | L | 69A | L | CA | 0.797 |
| 68A | K | 70A | P | CA | 0.811 |
| 69A | D | 71A | D | CA | 0.346 |
| 70A | R | 72A | R | CA | 0.252 |

# Appendix part III

| | | | | | |
|---|---|---|---|---|---|
| **71A** | **I** | **73A** | **L** | **CA** | **0.148** |
| **72A** | **N** | **74A** | **N** | **CA** | **0.099** |
| **73A** | **I** | **75A** | **I** | **CA** | **0.387** |
| **74A** | **V** | **76A** | **I** | **CA** | **0.211** |
| **75A** | **L** | **77A** | **L** | **CA** | **0.162** |
| **76A** | **S** | **78A** | **S** | **CA** | **0.774** |
| **77A** | **R** | **79A** | **R** | **CA** | **1.828** |
| **88A** | **F** | **91A** | **I** | **CA** | **0.572** |
| **89A** | **L** | **92A** | **H** | **CA** | **0.496** |
| **90A** | **A** | **93A** | **A** | **CA** | **0.566** |
| **91A** | **K** | **94A** | **S** | **CA** | **1.632** |
| **112A** | **V** | **109A** | **V** | **CA** | **0.475** |
| **113A** | **W** | **110A** | **F** | **CA** | **0.495** |
| **114A** | **I** | **111A** | **I** | **CA** | **0.393** |
| **115A** | **V** | **112A** | **I** | **CA** | **0.345** |
| **116A** | **G** | **113A** | **G** | **CA** | **0.368** |
| **117A** | **G** | **114A** | **G** | **CA** | **0.424** |
| **118A** | **S** | **115A** | **A** | **CA** | **0.880** |
| **119A** | **S** | **116A** | **E** | **CA** | **0.881** |
| **120A** | **V** | **117A** | **I** | **CA** | **0.769** |
| **121A** | **Y** | **118A** | **Y** | **CA** | **1.040** |
| **122A** | **Q** | **119A** | **N** | **CA** | **1.594** |
| **131A** | **L** | **128A** | **S** | **CA** | **1.356** |
| **132A** | **R** | **129A** | **H** | **CA** | **0.372** |
| **133A** | **L** | **130A** | **L** | **CA** | **0.407** |
| **134A** | **F** | **131A** | **L** | **CA** | **0.147** |
| **135A** | **V** | **132A** | **I** | **CA** | **0.137** |
| **136A** | **T** | **133A** | **T** | **CA** | **0.238** |
| **137A** | **R** | **134A** | **E** | **CA** | **0.463** |
| **138A** | **I** | **135A** | **I** | **CA** | **0.675** |
| **145A** | **D** | **146A** | **D** | **CA** | **1.010** |
| **146A** | **T** | **147A** | **T** | **CA** | **0.672** |
| **147A** | **F** | **148A** | **F** | **CA** | **0.611** |
| **148A** | **F** | **149A** | **L** | **CA** | **0.989** |
| **149A** | **P** | **150A** | **K** | **CA** | **1.457** |
| **175A** | **I** | **182A** | **F** | **CA** | **1.572** |
| **176A** | **K** | **183A** | **T** | **CA** | **0.668** |
| **177A** | **Y** | **184A** | **Y** | **CA** | **0.304** |
| **178A** | **K** | **185A** | **N** | **CA** | **0.312** |
| **179A** | **F** | **186A** | **Y** | **CA** | **0.417** |
| **180A** | **E** | **187A** | **T** | **CA** | **0.469** |
| **181A** | **V** | **188A** | **L** | **CA** | **0.386** |
| **182A** | **Y** | **189A** | **W** | **CA** | **0.188** |
| **183A** | **E** | **190A** | **T** | **CA** | **0.776** |
| **184A** | **K** | **191A** | **R** | **CA** | **1.521** |

# Appendix part III

**Table 20.** Residues used for superposition between 3K45 and 4GH8

| Residue number (Fixed) | Residue name (Fixed) | Residue number (Rotated) | Residue name (Rotated) | Atom name | Deviation (Å) |
|---|---|---|---|---|---|
| 4A | L | 2A | I | CA | 0.452 |
| 5A | N | 3A | S | CA | 0.786 |
| 6A | C | 4A | L | CA | 0.308 |
| 7A | I | 5A | I | CA | 0.131 |
| 8A | V | 6A | A | CA | 0.181 |
| 9A | A | 7A | A | CA | 0.402 |
| 10A | V | 8A | L | CA | 0.154 |
| 11A | S | 9A | A | CA | 0.963 |
| 12A | Q | 10A | V | CA | 1.683 |
| 14A | M | 12A | R | CA | 0.834 |
| 15A | G | 13A | V | CA | 0.361 |
| 16A | I | 14A | I | CA | 0.336 |
| 17A | G | 15A | G | CA | 0.213 |
| 18A | K | 16A | M | CA | 0.834 |
| 19A | N | 17A | E | CA | 1.217 |
| 20A | G | 18A | N | CA | 0.750 |
| 21A | D | 19A | A | CA | 0.192 |
| 22A | L | 20A | M | CA | 0.261 |
| 26A | P | 24A | P | CA | 0.707 |
| 27A | L | 25A | L | CA | 0.582 |
| 28A | R | 26A | P | CA | 1.127 |
| 29A | N | 27A | A | CA | 1.328 |
| 30A | E | 28A | D | CA | 0.773 |
| 31A | F | 29A | L | CA | 0.429 |
| 32A | K | 30A | A | CA | 0.729 |
| 33A | Y | 31A | W | CA | 0.708 |
| 34A | F | 32A | F | CA | 0.208 |
| 35A | Q | 33A | K | CA | 0.629 |
| 36A | R | 34A | R | CA | 0.868 |
| 37A | M | 35A | N | CA | 0.782 |
| 38A | T | 36A | T | CA | 0.379 |
| 48A | N | 39A | K | CA | 0.948 |
| 49A | L | 40A | P | CA | 0.397 |
| 50A | V | 41A | V | CA | 0.259 |
| 51A | I | 42A | I | CA | 0.282 |
| 52A | M | 43A | M | CA | 0.183 |
| 53A | G | 44A | G | CA | 0.185 |
| 54A | R | 45A | R | CA | 0.446 |
| 55A | K | 46A | H | CA | 0.631 |
| 56A | T | 47A | T | CA | 0.382 |
| 57A | W | 48A | W | CA | 0.393 |
| 58A | F | 49A | E | CA | 1.119 |
| 59A | S | 50A | S | CA | 0.852 |
| 60A | I | 51A | I | CA | 0.751 |
| 66A | P | 57A | P | CA | 0.868 |
| 67A | L | 58A | L | CA | 0.464 |
| 68A | K | 59A | P | CA | 0.200 |
| 69A | D | 60A | G | CA | 0.424 |
| 70A | R | 61A | R | CA | 0.371 |

# Appendix part III

| | | | | | |
|---|---|---|---|---|---|
| **71A** | **I** | **62A** | **K** | **CA** | **0.444** |
| **72A** | **N** | **63A** | **N** | **CA** | **0.604** |
| **73A** | **I** | **64A** | **I** | **CA** | **0.399** |
| **74A** | **V** | **65A** | **I** | **CA** | **0.309** |
| **75A** | **L** | **66A** | **L** | **CA** | **0.378** |
| **76A** | **S** | **67A** | **S** | **CA** | **0.547** |
| **77A** | **R** | **68A** | **S** | **CA** | **0.482** |
| **88A** | **F** | **77A** | **T** | **CA** | **1.546** |
| **89A** | **L** | **78A** | **W** | **CA** | **1.274** |
| **90A** | **A** | **79A** | **V** | **CA** | **1.083** |
| **91A** | **K** | **80A** | **K** | **CA** | **0.688** |
| **112A** | **V** | **95A** | **I** | **CA** | **0.955** |
| **113A** | **W** | **96A** | **M** | **CA** | **0.284** |
| **114A** | **I** | **97A** | **V** | **CA** | **0.262** |
| **115A** | **V** | **98A** | **I** | **CA** | **0.222** |
| **116A** | **G** | **99A** | **G** | **CA** | **0.537** |
| **117A** | **G** | **100A** | **G** | **CA** | **0.425** |
| **118A** | **S** | **101A** | **G** | **CA** | **0.428** |
| **119A** | **S** | **102A** | **R** | **CA** | **0.384** |
| **120A** | **V** | **103A** | **V** | **CA** | **0.394** |
| **121A** | **Y** | **104A** | **Y** | **CA** | **0.350** |
| **122A** | **Q** | **105A** | **E** | **CA** | **0.211** |
| **131A** | **L** | **112A** | **Q** | **CA** | **1.716** |
| **132A** | **R** | **113A** | **K** | **CA** | **0.769** |
| **133A** | **L** | **114A** | **L** | **CA** | **0.726** |
| **134A** | **F** | **115A** | **Y** | **CA** | **0.346** |
| **135A** | **V** | **116A** | **L** | **CA** | **0.635** |
| **136A** | **T** | **117A** | **T** | **CA** | **0.511** |
| **137A** | **R** | **118A** | **H** | **CA** | **0.670** |
| **138A** | **I** | **119A** | **I** | **CA** | **0.735** |
| **145A** | **D** | **126A** | **D** | **CA** | **1.086** |
| **146A** | **T** | **127A** | **T** | **CA** | **0.926** |
| **147A** | **F** | **128A** | **H** | **CA** | **0.634** |
| **148A** | **F** | **129A** | **F** | **CA** | **0.635** |
| **149A** | **P** | **130A** | **P** | **CA** | **0.815** |
| **175A** | **I** | **153A** | **H** | **CA** | **0.746** |
| **176A** | **K** | **154A** | **S** | **CA** | **0.713** |
| **177A** | **Y** | **155A** | **Y** | **CA** | **0.508** |
| **178A** | **K** | **156A** | **C** | **CA** | **0.787** |
| **179A** | **F** | **157A** | **F** | **CA** | **0.618** |
| **180A** | **E** | **158A** | **E** | **CA** | **0.601** |
| **181A** | **V** | **159A** | **I** | **CA** | **0.785** |
| **182A** | **Y** | **160A** | **L** | **CA** | **1.088** |
| **183A** | **E** | **161A** | **E** | **CA** | **1.272** |
| **184A** | **K** | **162A** | **R** | **CA** | **1.516** |

# Appendix part III

**Table 21.** Residues used for superposition between 3K45 and 1ZDR

| Residue number (Fixed) | Residue name (Fixed) | Residue number (Rotated) | Residue name (Rotated) | Atom name | Deviation (Å) |
|---|---|---|---|---|---|
| 4A | L | 2A | I | CA | 0.808 |
| 5A | N | 3A | S | CA | 0.903 |
| 6A | C | 4A | H | CA | 0.444 |
| 7A | I | 5A | I | CA | 0.241 |
| 8A | V | 6A | V | CA | 0.239 |
| 9A | A | 7A | A | CA | 0.451 |
| 10A | V | 8A | M | CA | 0.798 |
| 11A | S | 9A | D | CA | 0.344 |
| 12A | Q | 10A | E | CA | 0.255 |
| 14A | M | 12A | R | CA | 1.011 |
| 15A | G | 13A | V | CA | 0.702 |
| 16A | I | 14A | I | CA | 0.589 |
| 17A | G | 15A | G | CA | 0.980 |
| 18A | K | 16A | K | CA | 0.665 |
| 19A | N | 17A | D | CA | 0.903 |
| 20A | G | 18A | N | CA | 0.796 |
| 21A | D | 19A | R | CA | 0.858 |
| 22A | L | 20A | L | CA | 1.213 |
| 26A | P | 23A | H | CA | 1.575 |
| 27A | L | 24A | L | CA | 0.997 |
| 28A | R | 25A | P | CA | 1.236 |
| 29A | N | 26A | A | CA | 1.358 |
| 30A | E | 27A | D | CA | 0.840 |
| 31A | F | 28A | L | CA | 0.035 |
| 32A | K | 29A | A | CA | 0.143 |
| 33A | Y | 30A | Y | CA | 0.182 |
| 34A | F | 31A | F | CA | 0.252 |
| 35A | Q | 32A | K | CA | 0.182 |
| 36A | R | 33A | R | CA | 0.321 |
| 37A | M | 34A | V | CA | 0.453 |
| 38A | T | 35A | T | CA | 0.658 |
| 48A | N | 38A | H | CA | 1.061 |
| 49A | L | 39A | A | CA | 0.472 |
| 50A | V | 40A | I | CA | 0.594 |
| 51A | I | 41A | V | CA | 0.405 |
| 52A | M | 42A | M | CA | 0.248 |
| 53A | G | 43A | G | CA | 0.454 |
| 54A | R | 44A | R | CA | 0.558 |
| 55A | K | 45A | K | CA | 0.494 |
| 56A | T | 46A | T | CA | 0.342 |
| 57A | W | 47A | F | CA | 0.452 |
| 58A | F | 48A | E | CA | 0.888 |
| 59A | S | 49A | A | CA | 0.399 |
| 60A | I | 50A | I | CA | 1.001 |
| 66A | P | 53A | P | CA | 1.174 |
| 67A | L | 54A | L | CA | 0.992 |
| 68A | K | 55A | P | CA | 1.011 |
| 69A | D | 56A | G | CA | 0.842 |

# Appendix part III

| | | | | | |
|---|---|---|---|---|---|
| **70A** | **R** | **57A** | **R** | **CA** | **1.049** |
| **71A** | **I** | **58A** | **D** | **CA** | **0.587** |
| **72A** | **N** | **59A** | **N** | **CA** | **0.457** |
| **73A** | **I** | **60A** | **V** | **CA** | **0.563** |
| **74A** | **V** | **61A** | **V** | **CA** | **0.574** |
| **75A** | **L** | **62A** | **V** | **CA** | **0.901** |
| **76A** | **S** | **63A** | **T** | **CA** | **0.681** |
| **88A** | **F** | **74A** | **L** | **CA** | **0.574** |
| **89A** | **L** | **75A** | **V** | **CA** | **0.519** |
| **90A** | **A** | **76A** | **L** | **CA** | **0.551** |
| **91A** | **K** | **77A** | **H** | **CA** | **1.185** |
| **112A** | **V** | **93A** | **V** | **CA** | **0.508** |
| **113A** | **W** | **94A** | **F** | **CA** | **0.459** |
| **114A** | **I** | **95A** | **I** | **CA** | **0.445** |
| **115A** | **V** | **96A** | **I** | **CA** | **0.472** |
| **116A** | **G** | **97A** | **G** | **CA** | **0.246** |
| **117A** | **G** | **98A** | **G** | **CA** | **1.261** |
| **118A** | **S** | **99A** | **A** | **CA** | **0.575** |
| **119A** | **S** | **100A** | **E** | **CA** | **0.494** |
| **120A** | **V** | **101A** | **L** | **CA** | **0.768** |
| **121A** | **Y** | **102A** | **F** | **CA** | **0.576** |
| **122A** | **Q** | **103A** | **R** | **CA** | **0.454** |
| **131A** | **L** | **110A** | **D** | **CA** | **1.216** |
| **132A** | **R** | **111A** | **R** | **CA** | **0.379** |
| **133A** | **L** | **112A** | **L** | **CA** | **0.366** |
| **134A** | **F** | **113A** | **Y** | **CA** | **0.148** |
| **135A** | **V** | **114A** | **V** | **CA** | **0.159** |
| **136A** | **T** | **115A** | **T** | **CA** | **0.257** |
| **137A** | **R** | **116A** | **K** | **CA** | **0.479** |
| **138A** | **I** | **117A** | **I** | **CA** | **0.354** |
| **145A** | **D** | **124A** | **D** | **CA** | **0.320** |
| **146A** | **T** | **125A** | **T** | **CA** | **1.071** |
| **147A** | **F** | **126A** | **F** | **CA** | **0.934** |
| **148A** | **F** | **127A** | **Y** | **CA** | **0.692** |
| **149A** | **P** | **128A** | **P** | **CA** | **0.403** |
| **175A** | **I** | **151A** | **Y** | **CA** | **0.826** |
| **176A** | **K** | **152A** | **E** | **CA** | **0.967** |
| **177A** | **Y** | **153A** | **H** | **CA** | **0.511** |
| **178A** | **K** | **154A** | **A** | **CA** | **0.454** |
| **179A** | **F** | **155A** | **F** | **CA** | **0.412** |
| **180A** | **E** | **156A** | **I** | **CA** | **0.264** |
| **181A** | **V** | **157A** | **I** | **CA** | **0.215** |
| **182A** | **Y** | **158A** | **Y** | **CA** | **0.336** |
| **183A** | **E** | **159A** | **E** | **CA** | **0.731** |
| **184A** | **K** | **160A** | **R** | **CA** | **1.839** |

# Appendix part III

**Table 22.** Residues used for superposition between 3K45 and 3JW3

| Residue number (Fixed) | Residue name (Fixed) | Residue number (Rotated) | Residue name (Rotated) | Atom name | Deviation (Å) |
|---|---|---|---|---|---|
| 4A | L | 3A | V | CA | 0.373 |
| 5A | N | 4A | S | CA | 0.654 |
| 6A | C | 5A | F | CA | 0.565 |
| 7A | I | 6A | M | CA | 0.361 |
| 8A | V | 7A | V | CA | 0.477 |
| 9A | A | 8A | A | CA | 0.336 |
| 10A | V | 9A | M | CA | 0.515 |
| 11A | S | 10A | D | CA | 0.207 |
| 12A | Q | 11A | E | CA | 0.481 |
| 14A | M | 13A | R | CA | 0.969 |
| 15A | G | 14A | V | CA | 0.647 |
| 16A | I | 15A | I | CA | 0.577 |
| 17A | G | 16A | G | CA | 0.838 |
| 18A | K | 17A | K | CA | 0.725 |
| 19A | N | 18A | D | CA | 0.885 |
| 20A | G | 19A | N | CA | 0.918 |
| 21A | D | 20A | N | CA | 0.680 |
| 22A | L | 21A | L | CA | 1.225 |
| 26A | P | 24A | R | CA | 1.172 |
| 27A | L | 25A | L | CA | 0.255 |
| 28A | R | 26A | P | CA | 0.365 |
| 29A | N | 27A | S | CA | 0.476 |
| 30A | E | 28A | E | CA | 0.532 |
| 31A | F | 29A | L | CA | 0.578 |
| 32A | K | 30A | Q | CA | 0.425 |
| 33A | Y | 31A | Y | CA | 0.243 |
| 34A | F | 32A | V | CA | 0.597 |
| 35A | Q | 33A | K | CA | 0.614 |
| 36A | R | 34A | K | CA | 0.503 |
| 37A | M | 35A | T | CA | 0.520 |
| 38A | T | 36A | T | CA | 0.938 |
| 48A | N | 39A | H | CA | 1.084 |
| 49A | L | 40A | P | CA | 0.619 |
| 50A | V | 41A | L | CA | 0.642 |
| 51A | I | 42A | I | CA | 0.633 |
| 52A | M | 43A | M | CA | 0.398 |
| 53A | G | 44A | G | CA | 0.298 |
| 54A | R | 45A | R | CA | 0.211 |
| 55A | K | 46A | K | CA | 0.598 |
| 56A | T | 47A | N | CA | 0.716 |
| 57A | W | 48A | Y | CA | 0.668 |
| 58A | F | 49A | E | CA | 0.742 |
| 59A | S | 50A | A | CA | 0.766 |
| 60A | I | 51A | I | CA | 0.961 |
| 66A | P | 54A | P | CA | 1.148 |
| 67A | L | 55A | L | CA | 1.343 |
| 68A | K | 56A | P | CA | 0.975 |
| 69A | D | 57A | G | CA | 0.311 |
| 70A | R | 58A | R | CA | 0.515 |

# Appendix part III

| 71A | I | 59A | R | CA | 0.536 |
|---|---|---|---|---|---|
| 72A | N | 60A | N | CA | 0.715 |
| 73A | I | 61A | I | CA | 0.717 |
| 74A | V | 62A | I | CA | 0.456 |
| 75A | L | 63A | V | CA | 0.363 |
| 76A | S | 64A | T | CA | 0.499 |
| 77A | R | 65A | R | CA | 0.166 |
| 88A | F | 75A | E | CA | 1.397 |
| 89A | L | 76A | V | CA | 0.723 |
| 90A | A | 77A | A | CA | 0.764 |
| 91A | K | 78A | H | CA | 0.727 |
| 112A | V | 93A | I | CA | 0.671 |
| 113A | W | 94A | F | CA | 0.510 |
| 114A | I | 95A | I | CA | 0.536 |
| 115A | V | 96A | I | CA | 0.651 |
| 116A | G | 97A | G | CA | 0.794 |
| 117A | G | 98A | G | CA | 0.496 |
| 118A | S | 99A | A | CA | 0.672 |
| 119A | S | 100A | Q | CA | 0.682 |
| 120A | V | 101A | I | CA | 0.834 |
| 121A | Y | 102A | Y | CA | 0.971 |
| 122A | Q | 103A | D | CA | 1.252 |
| 131A | L | 110A | D | CA | 1.351 |
| 132A | R | 111A | K | CA | 0.756 |
| 133A | L | 112A | L | CA | 1.005 |
| 134A | F | 113A | Y | CA | 0.095 |
| 135A | V | 114A | I | CA | 0.205 |
| 136A | T | 115A | T | CA | 0.245 |
| 137A | R | 116A | K | CA | 0.413 |
| 138A | I | 117A | I | CA | 0.455 |
| 145A | D | 124A | D | CA | 0.411 |
| 146A | T | 125A | T | CA | 1.291 |
| 147A | F | 126A | F | CA | 1.027 |
| 148A | F | 127A | F | CA | 0.672 |
| 149A | P | 128A | P | CA | 1.052 |
| 175A | I | 151A | Y | CA | 0.472 |
| 176A | K | 152A | T | CA | 0.271 |
| 177A | Y | 153A | Y | CA | 0.214 |
| 178A | K | 154A | Y | CA | 0.548 |
| 179A | F | 155A | Y | CA | 0.374 |
| 180A | E | 156A | H | CA | 0.393 |
| 181A | V | 157A | V | CA | 0.529 |
| 182A | Y | 158A | Y | CA | 1.095 |
| 183A | E | 159A | E | CA | 1.609 |
| 184A | K | 160A | K | CA | 1.727 |

# Appendix part III

**Table 23.** Residues used for superposition between 3K45 and 2QK8

| Residue number (Fixed) | Residue name (Fixed) | Residue number (Rotated) | Residue name (Rotated) | Atom name | Deviation (Å) |
|---|---|---|---|---|---|
| 4A | L | 3A | V | CA | 0.471 |
| 5A | N | 4A | S | CA | 0.874 |
| 6A | C | 5A | F | CA | 0.717 |
| 7A | I | 6A | M | CA | 0.598 |
| 8A | V | 7A | V | CA | 0.363 |
| 9A | A | 8A | A | CA | 0.440 |
| 10A | V | 9A | M | CA | 1.023 |
| 11A | S | 10A | D | CA | 0.526 |
| 12A | Q | 11A | E | CA | 0.479 |
| 14A | M | 13A | R | CA | 0.881 |
| 15A | G | 14A | V | CA | 0.812 |
| 16A | I | 15A | I | CA | 0.695 |
| 17A | G | 16A | G | CA | 0.864 |
| 18A | K | 17A | K | CA | 0.841 |
| 19A | N | 18A | D | CA | 0.650 |
| 20A | G | 19A | N | CA | 0.529 |
| 21A | D | 20A | N | CA | 1.307 |
| 22A | L | 21A | L | CA | 1.149 |
| 26A | P | 24A | R | CA | 0.838 |
| 27A | L | 25A | L | CA | 0.541 |
| 28A | R | 26A | P | CA | 0.437 |
| 29A | N | 27A | S | CA | 0.436 |
| 30A | E | 28A | E | CA | 0.529 |
| 31A | F | 29A | L | CA | 0.583 |
| 32A | K | 30A | Q | CA | 0.198 |
| 33A | Y | 31A | Y | CA | 0.405 |
| 34A | F | 32A | V | CA | 0.558 |
| 35A | Q | 33A | K | CA | 0.556 |
| 36A | R | 34A | K | CA | 0.424 |
| 37A | M | 35A | T | CA | 0.685 |
| 38A | T | 36A | T | CA | 0.578 |
| 48A | N | 39A | H | CA | 1.209 |
| 49A | L | 40A | P | CA | 0.861 |
| 50A | V | 41A | L | CA | 0.873 |
| 51A | I | 42A | I | CA | 0.605 |
| 52A | M | 43A | M | CA | 0.459 |
| 53A | G | 44A | G | CA | 0.429 |
| 54A | R | 45A | R | CA | 0.796 |
| 55A | K | 46A | K | CA | 0.705 |
| 56A | T | 47A | N | CA | 0.588 |
| 57A | W | 48A | Y | CA | 0.933 |
| 58A | F | 49A | E | CA | 1.418 |
| 59A | S | 50A | A | CA | 1.267 |
| 60A | I | 51A | I | CA | 0.952 |
| 66A | P | 54A | P | CA | 1.149 |
| 67A | L | 55A | L | CA | 0.607 |
| 68A | K | 56A | P | CA | 0.190 |
| 69A | D | 57A | G | CA | 0.598 |
| 70A | R | 58A | R | CA | 0.473 |

# Appendix part III

| | | | | | |
|---|---|---|---|---|---|
| **71A** | **I** | **59A** | **R** | **CA** | **0.771** |
| **72A** | **N** | **60A** | **N** | **CA** | **0.972** |
| **73A** | **I** | **61A** | **I** | **CA** | **0.646** |
| **74A** | **V** | **62A** | **I** | **CA** | **0.590** |
| **75A** | **L** | **63A** | **V** | **CA** | **0.488** |
| **76A** | **S** | **64A** | **T** | **CA** | **0.868** |
| **77A** | **R** | **65A** | **R** | **CA** | **1.445** |
| **88A** | **F** | **75A** | **E** | **CA** | **1.621** |
| **89A** | **L** | **76A** | **V** | **CA** | **0.876** |
| **90A** | **A** | **77A** | **A** | **CA** | **0.840** |
| **91A** | **K** | **78A** | **H** | **CA** | **0.831** |
| **112A** | **V** | **93A** | **I** | **CA** | **1.256** |
| **113A** | **W** | **94A** | **F** | **CA** | **0.738** |
| **114A** | **I** | **95A** | **I** | **CA** | **0.740** |
| **115A** | **V** | **96A** | **F** | **CA** | **0.452** |
| **116A** | **G** | **97A** | **G** | **CA** | **0.914** |
| **117A** | **G** | **98A** | **G** | **CA** | **0.913** |
| **118A** | **S** | **99A** | **A** | **CA** | **0.709** |
| **119A** | **S** | **100A** | **Q** | **CA** | **0.455** |
| **120A** | **V** | **101A** | **I** | **CA** | **0.595** |
| **121A** | **Y** | **102A** | **Y** | **CA** | **0.915** |
| **122A** | **Q** | **103A** | **D** | **CA** | **1.429** |
| **131A** | **L** | **110A** | **D** | **CA** | **1.604** |
| **132A** | **R** | **111A** | **K** | **CA** | **0.901** |
| **133A** | **L** | **112A** | **L** | **CA** | **0.940** |
| **134A** | **F** | **113A** | **Y** | **CA** | **0.288** |
| **135A** | **V** | **114A** | **I** | **CA** | **0.091** |
| **136A** | **T** | **115A** | **T** | **CA** | **0.258** |
| **137A** | **R** | **116A** | **K** | **CA** | **0.570** |
| **138A** | **I** | **117A** | **I** | **CA** | **0.483** |
| **145A** | **D** | **124A** | **D** | **CA** | **0.256** |
| **146A** | **T** | **125A** | **T** | **CA** | **1.107** |
| **147A** | **F** | **126A** | **F** | **CA** | **0.908** |
| **148A** | **F** | **127A** | **F** | **CA** | **0.617** |
| **149A** | **P** | **128A** | **P** | **CA** | **1.372** |
| **175A** | **I** | **151A** | **Y** | **CA** | **0.829** |
| **176A** | **K** | **152A** | **T** | **CA** | **0.688** |
| **177A** | **Y** | **153A** | **Y** | **CA** | **0.319** |
| **178A** | **K** | **154A** | **Y** | **CA** | **0.662** |
| **179A** | **F** | **155A** | **Y** | **CA** | **0.658** |
| **180A** | **E** | **156A** | **H** | **CA** | **0.640** |
| **181A** | **V** | **157A** | **V** | **CA** | **0.616** |
| **182A** | **Y** | **158A** | **Y** | **CA** | **1.390** |
| **183A** | **E** | **159A** | **E** | **CA** | **1.502** |
| **184A** | **K** | **160A** | **K** | **CA** | **1.511** |

# Appendix part III

**Table 24.** Residues used for superposition between 3K45 and 2ZZA

| Residue number (Fixed) | Residue name (Fixed) | Residue number (Rotated) | Residue name (Rotated) | Atom name | Deviation (Å) |
|---|---|---|---|---|---|
| 4A | L | 3A | V | CA | 0.775 |
| 5A | N | 4A | S | CA | 1.056 |
| 6A | C | 5A | M | CA | 0.373 |
| 7A | I | 6A | I | CA | 0.357 |
| 8A | V | 7A | A | CA | 0.103 |
| 9A | A | 8A | A | CA | 0.221 |
| 10A | V | 9A | L | CA | 0.308 |
| 11A | S | 10A | A | CA | 0.675 |
| 12A | Q | 11A | N | CA | 1.437 |
| 14A | M | 13A | R | CA | 1.019 |
| 15A | G | 14A | V | CA | 0.507 |
| 16A | I | 15A | I | CA | 0.697 |
| 17A | G | 16A | G | CA | 0.910 |
| 18A | K | 17A | L | CA | 0.894 |
| 19A | N | 18A | D | CA | 1.276 |
| 20A | G | 19A | N | CA | 1.074 |
| 21A | D | 20A | K | CA | 0.759 |
| 22A | L | 21A | M | CA | 1.024 |
| 26A | P | 24A | H | CA | 1.345 |
| 27A | L | 25A | L | CA | 0.578 |
| 28A | R | 26A | P | CA | 0.241 |
| 29A | N | 27A | A | CA | 0.195 |
| 30A | E | 28A | E | CA | 0.141 |
| 31A | F | 29A | L | CA | 0.314 |
| 32A | K | 30A | Q | CA | 0.418 |
| 33A | Y | 31A | L | CA | 0.373 |
| 34A | F | 32A | F | CA | 0.299 |
| 35A | Q | 33A | K | CA | 0.359 |
| 36A | R | 34A | R | CA | 0.729 |
| 37A | M | 35A | A | CA | 0.826 |
| 38A | T | 36A | T | CA | 0.217 |
| 48A | N | 39A | K | CA | 1.235 |
| 49A | L | 40A | P | CA | 0.356 |
| 50A | V | 41A | I | CA | 0.357 |
| 51A | I | 42A | V | CA | 0.393 |
| 52A | M | 43A | M | CA | 0.277 |
| 53A | G | 44A | G | CA | 0.359 |
| 54A | R | 45A | R | CA | 0.358 |
| 55A | K | 46A | N | CA | 0.694 |
| 56A | T | 47A | T | CA | 0.522 |
| 57A | W | 48A | F | CA | 0.460 |
| 58A | F | 49A | E | CA | 0.849 |
| 59A | S | 50A | S | CA | 0.775 |
| 60A | I | 51A | I | CA | 0.877 |
| 66A | P | 54A | P | CA | 1.334 |
| 67A | L | 55A | L | CA | 0.745 |
| 68A | K | 56A | P | CA | 0.425 |
| 69A | D | 57A | G | CA | 0.675 |
| 70A | R | 58A | R | CA | 0.613 |

# Appendix part III

| | | | | | |
|---|---|---|---|---|---|
| **71A** | **I** | **59A** | **L** | **CA** | **0.536** |
| **72A** | **N** | **60A** | **N** | **CA** | **0.673** |
| **73A** | **I** | **61A** | **I** | **CA** | **0.503** |
| **74A** | **V** | **62A** | **V** | **CA** | **0.405** |
| **75A** | **L** | **63A** | **L** | **CA** | **0.173** |
| **76A** | **S** | **64A** | **S** | **CA** | **0.162** |
| **77A** | **R** | **65A** | **R** | **CA** | **0.610** |
| **88A** | **F** | **75A** | **T** | **CA** | **1.517** |
| **89A** | **L** | **76A** | **V** | **CA** | **1.052** |
| **90A** | **A** | **77A** | **V** | **CA** | **0.799** |
| **91A** | **K** | **78A** | **A** | **CA** | **0.603** |
| **112A** | **V** | **93A** | **L** | **CA** | **0.959** |
| **113A** | **W** | **94A** | **M** | **CA** | **0.421** |
| **114A** | **I** | **95A** | **I** | **CA** | **0.556** |
| **115A** | **V** | **96A** | **I** | **CA** | **0.547** |
| **116A** | **G** | **97A** | **G** | **CA** | **1.193** |
| **117A** | **G** | **98A** | **G** | **CA** | **0.458** |
| **118A** | **S** | **99A** | **A** | **CA** | **0.324** |
| **119A** | **S** | **100A** | **T** | **CA** | **0.133** |
| **120A** | **V** | **101A** | **I** | **CA** | **0.382** |
| **121A** | **Y** | **102A** | **Y** | **CA** | **0.410** |
| **122A** | **Q** | **103A** | **N** | **CA** | **0.509** |
| **131A** | **L** | **110A** | **D** | **CA** | **1.271** |
| **132A** | **R** | **111A** | **R** | **CA** | **0.454** |
| **133A** | **L** | **112A** | **L** | **CA** | **0.412** |
| **134A** | **F** | **113A** | **Y** | **CA** | **0.393** |
| **135A** | **V** | **114A** | **L** | **CA** | **0.427** |
| **136A** | **T** | **115A** | **T** | **CA** | **0.164** |
| **137A** | **R** | **116A** | **H** | **CA** | **0.355** |
| **138A** | **I** | **117A** | **I** | **CA** | **0.318** |
| **145A** | **D** | **124A** | **D** | **CA** | **1.133** |
| **146A** | **T** | **125A** | **T** | **CA** | **1.451** |
| **147A** | **F** | **126A** | **W** | **CA** | **1.018** |
| **148A** | **F** | **127A** | **F** | **CA** | **0.823** |
| **149A** | **P** | **128A** | **P** | **CA** | **0.737** |
| **175A** | **I** | **151A** | **H** | **CA** | **0.792** |
| **176A** | **K** | **152A** | **N** | **CA** | **0.532** |
| **177A** | **Y** | **153A** | **Y** | **CA** | **0.214** |
| **178A** | **K** | **154A** | **R** | **CA** | **0.651** |
| **179A** | **F** | **155A** | **F** | **CA** | **0.197** |
| **180A** | **E** | **156A** | **S** | **CA** | **0.291** |
| **181A** | **V** | **157A** | **L** | **CA** | **0.971** |
| **182A** | **Y** | **158A** | **L** | **CA** | **1.062** |
| **183A** | **E** | **159A** | **E** | **CA** | **1.179** |
| **184A** | **K** | **160A** | **R** | **CA** | **1.277** |

# References

# References


- **Abali EE, Skacel NE, Celikkaya H, Hsieh YC. (2008)** Chapter 9 regulation of human dihydrofolate reductase activity and expression. Vitam. Horm. 79: 267-92.
- **Aryal, S. (2021).** NMR Spectroscopy- Definition, Principle, Steps, Parts, Uses. Microbe Notes. Retrieved from https://microbenotes.com/nuclear-magnetic-resonance-nmr-spectroscopy/#principle-of-nuclear-magnetic-resonance-nmr-spectroscopy
- **Beierlein, J. M., Karri, N. G., & Anderson, A. C. (2010)**. Targeted mutations of Bacillus anthracis dihydrofolate reductase condense complex structure−activity relationships. Journal of medicinal chemistry, 53(20), 7327–7336. https://doi.org/10.1021/jm100727t.
- **Biochemistry Glossary**: Peptide Bond Formation. Retrieved from https://www.drawittoknowit.com/course/biochemistry/glossary/biochemical-pathway/peptide-bond-formation
- **Chakraborty S, Venkatramani R, Rao BJ, Asgeirsson B, Dandekar AM. (2013)** Protein structure quality assessment based on the distance profiles of consecutive backbone Cα atoms. F1000Res. Oct 10;2:211. doi: 10.12688/f1000research.2-211.v3. PMID: 24555103; PMCID: PMC3892923.
- **CNX** Open Stack Microbio 07 04 proteinstr.jpg - Wikimedia Commons. (2016). Retrieved from:https://commons.wikimedia.org/wiki/File:OSC_Microbio_07_04_proteinstr.jpg
- **Cody V., Luft J. R., Pangborn W. (2005)** Understanding the role of Leu22 variants in methotrexate resistance: comparison of wild-type and Leu22Arg variant mouse and human dihydrofolate reductase ternary crystal complexes with methotrexate and NADPH. Acta Cryst. D61, 147-155. Available from: https://doi.org/10.1107/s0907444904030422
- **Cody, V., Luft, J. R., & Pangborn, W. (2005).** Understanding the role of Leu22 variants in methotrexate resistance: comparison of wild-type and Leu22Arg variant mouse and human dihydrofolate reductase ternary crystal complexes with methotrexate and NADPH. Acta crystallographica. Section D, Biological crystallography, 61(Pt 2), 147–155. https://doi.org/10.1107/S0907444904030422.
- **Courtesy of Malvern-PANalytical B.V.**,What is X-Ray Powder Diffraction analysis-XRPD?. Retrieved from http://www.xrpd.eu/
- **Cowman AF, Lew AM (November 1989).** "Antifolate drug selection results in duplication and rearrangement of chromosome 7 in Plasmodium chabaudi". Molecular and Cellular Biology. 9 (11): 5182–8. doi:10.1128/mcb.9.11.5182. PMC 363670. PMID 2601715
- **Darnell, S. (2020).** Why Structure Prediction Matters. DNASTAR. Retrieved from https://www.dnastar.com/blog/structural-biology/why-structure-prediction-matters/
- **Davies, J.F. *et. al.* (1990).** Crystal structures of recombinant human dihydrofolate reductase complexed with folate and 5-deazafolate. Biochemistry Oct 9;29(40):9467-9479. Available from: https://doi.org/10.1021/bi00492a021
- **DeJarnette, C., Luna-Tapia, A., Estredge, L. R., & Palmer, G. E. (2020)**. Dihydrofolate Reductase Is a Valid Target for Antifungal Development in the Human Pathogen *Candida albicans*. *mSphere*, *5*(3), e00374-20. https://doi.org/10.1128/mSphere.00374-20.
- **Deléage, G., Gouy, M., & Brevern, A. (2021).** Bioinformatique (3rd ed., p. 47). france: [Dunod].
- **Deng, H., Jia, Y., & Zhang, Y. (2018).** Protein structure prediction. *International* journal of modern physics. B, 32(18), 1840009. https://doi.org/10.1142/S021797921840009X

# References


- **Drenth J (1994)**, principles of x-ray crystallography. springer-verlag, New york
- **Eldon L. Ulrich; Hideo Akutsu; et al** ; Nucleic Acids Research 36, D402-D408 **(2008)** doi: 10.1093/nar/gkm957
- **Futterman, S. (1957)** Enzymatic reduction of folic acid and dihydrofolic acid to tetrahydrofolic acid. J. Biol. Chem. 228: 1031-8.
- **Gangjee, A., Li, W., Lin, L., Zeng, Y., Ihnat, M., Warnke, L. A., Green, D. W., Cody, V., Pace, J., & Queener, S. F.** (**2009**). Design, synthesis, and X-ray crystal structures of 2,4-diaminofuro[2,3-d]pyrimidines as multireceptor tyrosine kinase and dihydrofolate reductase inhibitors. Bioorganic & medicinal chemistry, 17(20), 7324–7336. https://doi.org/10.1016/j.bmc.2009.08.044.
- **García-Álvarez, L., Busto, J., Peregrina, J., Avenoza, A., & Oteo, J. (2016).** Applications of 1H Nuclear Magnetic Resonance Spectroscopy in Clinical Microbiology. Applications Of Molecular Spectroscopy To Current Research In The Chemical And Biological Sciences,pg 284. doi: 10.5772/64450
- **Golovin, A., Dimitropoulos, D., Oldfield, T., Rachedi, A., Henrick, K. (2005).** MSDsite: a database search and retrieval system for the analysis and viewing of bound ligands and active sites. Proteins, 58(1):190-199. Available from https://doi.org/10.1002/prot.20288
- **Goodsell, D. (2002).** Dihydrofolate Reductase. RCSB Protein Data Bank. doi: 10.2210/rcsb_pdb/mom_2002_10
- **Goodsell, D. S. (2002).** Dihydrofolate reductase. RCSB Protein Data Bank. https://doi.org/10.2210/rcsb_pdb/mom_2002_10
- **H., Berman; Henrick, K.; Nakamura, H.; Markley, J. L. (January 2007).** "The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data". Nucleic Acids Res. 35 (Database issue): D301–D303. doi:10.1093/nar/gkl971. PMC 1669775. PMID 17142228
- **Haddad Y, Adam V, Heger Z (2020)** Ten quick tips for homology modeling of high-resolution protein 3D structures. PLoS Comput Biol 16(4): e1007449. https://doi.org/10.1371/journal.pcbi.1007449
- **Hartl, F. U., & Hayer-Hartl, M.** Converging concepts of protein folding in vitro and in vivo. Nature Structural and Molecular Biology 16, 574–581 **(2009)** doi:10.1038/nsmb.1591.
- **Héctor Zamora , C. (2021).** NMR Spectroscopy Principles, Interpreting an NMR Spectrum and Common Problems. Technology Network. Retrieved from https://www.technologynetworks.com/analysis/articles/nmr-spectroscopy-principles-interpreting-an-nmr-spectrum-and-common-problems-355891#D9

  http://www.sbg.bio.ic.ac.uk/ezmol/EzMol_Data/images/student/pdf/protein-structure.pdf

  https://leeds.primo.exlibrisgroup.com/permalink/44LEE_INST/5rdkl9/alma991006873019705181

  https://leeds.primo.exlibrisgroup.com/permalink/44LEE_INST/5rdkl9/alma991006873019705181
- **J.S. Richardson , D.C. Richardson, (1992)** Protein Science - Vol. 1, Issue 3 "The Protein Tourist: DOUBLY-WOUND PARALLEL ALPHA/BETA PROTEINS, OR NUCLEOTIDE-BINDING DOMAINS",

# References


- **Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., & Madden, T. (2008).** NCBI BLAST: a better web interface. Nucleic Acids Research, 36(Web Server), W5-W9. doi: 10.1093/nar/gkn201
- **K.Sumathi, P.Ananthalakshmi, M.N.A. Md Roshan** and **K.Sekar ,** 3d-ss (3Dimensional Structural Superposition) , **2005**, http://cluster.physics.iisc.ernet.in/3dss/.
- khan, T. (2020). How to Search Protein Data Bank (PDB) | Sci Chores. Retrieved from https://www.scichores.com/2020/08/how-to-search-protein-data-bank-pdb-sci.html
- **Kinjo, A., Suzuki, H., Yamashita, R., Ikegawa, Y., Kudou, T., & Igarashi, R. et al. (2011).** Protein Data Bank Japan (PDBj): maintaining a structural data archive and resource description framework format. Nucleic Acids Research, 40(D1), D453-D460. doi: 10.1093/nar/gkr811
- **Kumar, S., Demo, G., Koca, J., & Wimmerova, M. (2012).** In Silico Engineering of Proteins That Recognize Small Molecules. Protein Engineering. doi: 10.5772/28001.
- **Lam, A. J., Rao, L., Anazawa, Y., Okada, K., Chiba, K., Niwa, S., Gennerich, A., Nowakowski, D. W., & McKenney, R. J. (2020).** A highly conserved 310-helix within the Kinesin Motor Domain is critical for kinesin function and human health. https://doi.org/10.1101/2020.09.22.308320
- **Larry Li., Biologydictionary.net Editors. (2014).** Organ. Retrieved from https://biologydictionary.net/tertiary-structure/
- **Lee, J., Freddolino, P., & Zhang, Y. (2017).** Ab Initio Protein Structure Prediction. From Protein Structure To Function With Bioinformatics, 3-35. doi: 10.1007/978-94-024-1069-3_1
- **Lever, G. (2015).** Large-Scale Quantum-Mechanical Enzymology. Springer Theses. doi: 10.1007/978-3-319-19351-9
- **M. Dorn et al. Márcio Dorna,⁎, Mariel Barbachan e Silva b, Luciana S. Buriol a, Luis C. Lamba** / Computational Biology and Chemistry 53 **(2014)** 251–276
- **Manske, M.** (**2001**). Beta sheet - Wikipedia, from https://en.wikipedia.org/wiki/Beta_sheet
- **Marcella Martos, Meredith Turnbough. (2012, April 27).** What Are Proteins?. ASU - Ask A Biologist. Retrieved from https://askabiologist.asu.edu/venom/what-are-proteins
- **McTigue, M. A., Davies, J. F., 2nd, Kaufman, B. T., & Kraut, J. (1992).** Crystal structure of chicken liver dihydrofolate reductase complexed with NADP+ and biopterin. Biochemistry, 31(32), 7264–7273. https://doi.org/10.1021/bi00147a009.
- **MedlinePlus** [Internet]. Bethesda (MD): National Library of Medicine (US); [updated 2020 Jun 24]. Noonan syndrome; [updated 2020 Jun 18; reviewed **2018** Jun 01; cited 2020 Jul 1]; [about 5 p.]. Available from: https://medlineplus.gov/genetics/condition/noonan-syndrome/.
- **Mutharasappan, N., Ravi Rao, G., Mariadasse, R., Poopandi, S., Mathimaran, A., Dhamodharan, P., Sundarraj, R., Jeyaraj Pandian, C. and Jeyaraman, J., 2020.** Experimental and Computational Methods to Determine Protein Structure and Stability. Frontiers in Protein Structure, Function, and Dynamics, pp.23-55.
- **OpenStax College**, Biology. OpenStax College. 30 May **2013**. <http://cnx.org/content/col11448/latest/>.

# References


- **Osborne MJ, Schnell J, Benkovic SJ, Dyson HJ, Wright PE** (August **2001**). "Backbone dynamics in dihydrofolate reductase complexes: role of loop flexibility in the catalytic mechanism". Biochemistry. 40 (33): 9846–59. doi:10.1021/bi010621k. PMID 11502178
- **Pan, A., Pranavathiyani, G., & Chakraborty, S. (2021).** Computational Modeling of Protein Three-Dimensional Structure: Methods and Resources. Molecular Docking For Computer-Aided Drug Design, 155-178. doi: 10.1016/b978-0-12-822312-3.00023-0
- **PDB.** (**2007**). Sucrose porin 1a0s.png [Image]. Retrieved from https://upload.wikimedia.org/wikipedia/commons/thumb/f/fb/Sucrose_porin_1a0s.png/800px-Sucrose_porin_1a0s.png
- **Penhallurick, R. W., & Ichiye, T. (2021)**. Pressure Adaptations in Deep-Sea Moritella Dihydrofolate Reductases: Compressibility versus Stability. Biology, 10(11), 1211. https://doi.org/10.3390/biology10111211
- **Pundir, S., Martin, M. J., O'Donovan, C., & UniProt Consortium (2016).** UniProt Tools. Current protocols in bioinformatics, 53, 1.29.1–1.29.15. https://doi.org/10.1002/0471250953.bi0129s53
- **R. B. Russell & G. J. Barton**, Multiple Protein sequence alignment from tertiary structure comparision, PROTEINS: Struct. Funct. Genet.,14, 309-323, **1992.**
- **Rachedi, A. (1994)** Three-dimensional structural studies on dihydrofolate reductase, Chapter 8 Ph.D Thesis, University of Leed, UK. Availabe from
- **Rachedi, A. (2011).** Sequence Structure and Function Server, **2011**, https://bioinformaticstools.org/ssfs/
- **Raimondi, M. V., Randazzo, O., La Franca, M., Barone, G., Vignoni, E., Rossi, D., & Collina, S. (2019).** DHFR Inhibitors: Reading the Past for Discovering Novel Anticancer Agents. Molecules (Basel, Switzerland), 24(6), 1140. https://doi.org/10.3390/molecules24061140
- **Reddy Michael K,. (2020,** October 20). amino acid. Encyclopedia Britannica. https://www.britannica.com/science/amino-acid
- **Rehman I, Farooq M, Botelho S.** Biochemistry, Secondary Protein Structure. [Updated **2021** Dec 15]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2022 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK470235/
- **Rives A. *et al.* (2021)** Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences, Proc. Natl. Acad. Sci., 118 (15) :e2016239118. Available from https://doi.org/10.1073/pnas.2016239118
- **Russell, p. (2010).** Amino Acid Structure. Retrieved 28 April 2022, from https://www.mun.ca/biology/scarr/iGen3_06-01.html
- **Sliwoski, G., Kothiwale, S., Meiler, J., & Lowe, E. (2013).** Computational Methods in Drug Discovery. Pharmacological Reviews, 66(1), 334-395. doi: 10.1124/pr.112.007336
- **Smith SL, Patrick P, Stone D, Phillips AW, Burchall JJ. J.** Biol. Chem Porcine liver dihydrofolate reductase. Purification, properties, and amino acid sequence.. 254, 11475-84, **(1979).**
- **Smyth, M. S., & Martin, J. H. (2000).** x ray crystallography. Molecular pathology : MP, 53(1), 8–14. https://doi.org/10.1136/mp.53.1.8

# References


- **Srinivasan B., Tonddast-Navaei S., Roy A., Zhou H., Skolnick J.** Chemical Space of Escherichia coli Dihydrofolate Reductase Inhibitors: New Approaches for Discovering Novel Drugs for Old Bugs. Med. Res. Rev. **2019**;39:684–705. doi: 10.1002/med.21538. Structural Bioinformatics Group: protein-structure
- **Srinivasan, B., Tonddast-Navaei, S., Roy, A., Zhou, H., & Skolnick, J. (2019)**. Chemical space of Escherichia coli dihydrofolate reductase inhibitors: New approaches for discovering novel drugs for old bugs. *Medicinal research reviews*, *39*(2), 684–705. https://doi.org/10.1002/med.21538.
- **Stryer L, Berg JM, Tymoczko JL (2002).** Biochemistry (5th ed.). San Francisco: W.H. Freeman. ISBN 0-7167-4955-6.open access
- **Sumathi, K., Ananthalakshmi, P., Roshan, M. N., & Sekar, K. (2006).** https://doi.org/10.1093/nar/gkl036
- **Swaminathan, J. (2008).** File:PDB 1dn3 EBI.jpg [Image]. Retrieved from https://commons.wikimedia.org/wiki/File:PDB_1dn3_EBI.jpg
- **The UniProt Consortium**, UniProt: the universal protein knowledgebase in **2021**, Nucleic Acids Research, Volume 49, Issue D1, 8 January 2021, Pages D480–D489, https://doi.org/10.1093/nar/gkaa1100
- **Totrov, M.M.** and **Abagyan, R.A**. Efficient Parallelization of The Energy, Surface and Derivative Calculations For internal Coordinate Mechanics. J. Comput. Chem., 15, 1105–1112, **1994**.
- **Uday,H.,** Quaternary Structure of protein vector illustration. Protein structure and folding. [Image], retrived from https://www.shutterstock.com/fr/image-vector/quaternary-structure-protein-vector-illustration-folding-1961746360
- **Velankar, S., Best, C., Beuth, B., Boutselakis, C., Cobley, N., & Sousa Da Silva, A. et al. (2009).** PDBe: Protein Data Bank in Europe. Nucleic Acids Research, 38(Database), D308-D317. doi: 10.1093/nar/gkp916
- **Wang, H. (2015).** Cryo-electron microscopy for structural biology: current status and future perspectives. Science China Life Sciences, 58(8), 750-756. doi: 10.1007/s11427-015-4851-2
- **Warwick, T. (2018).** What is Cryo-EM? | Blue Scientific. Retrieved from https://blue-scientific.com/cryo-em/