**MINI PROJECT REPORT**

**Integrated In-Silico Analysis of BRCA1:**

**Protein Domain Mapping and**

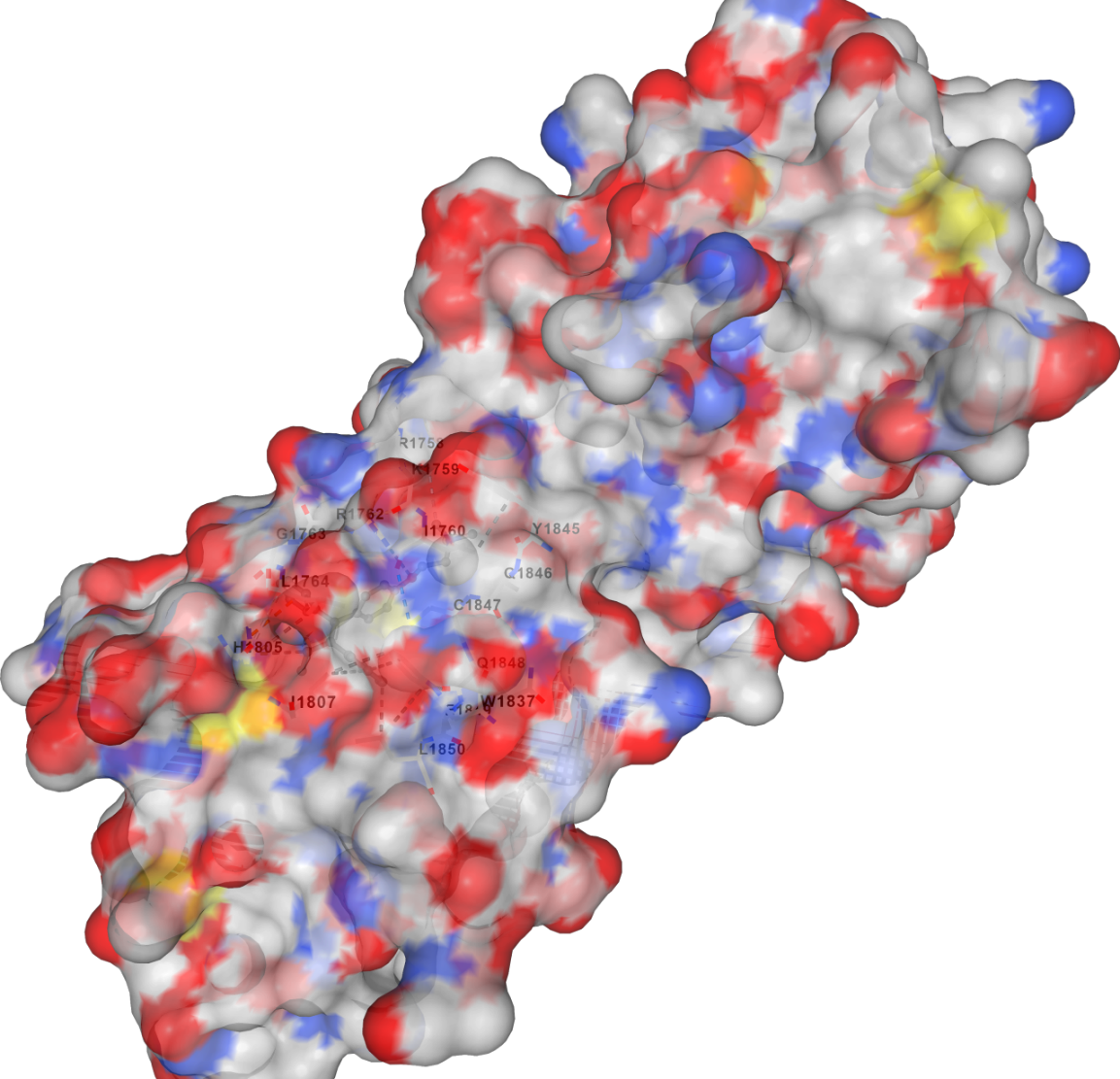
**Ligand Docking with Tamoxifen**

Done by

MANJUBASINI K

DEPARTMENT OF INDUSTRIAL BIOTECHNOLOGY

GOVERNMENT COLLEGE OF TECHNOLOGY,COIMBATORE

****

**1.INTODUCTION:**

The **BRCA1 (Breast Cancer type 1 susceptibility protein)** is a tumor suppressor gene that plays a critical role in the maintenance of genomic stability. It is involved in multiple essential cellular processes such as **DNA repair, transcriptional regulation, and cell cycle checkpoint control**. Germline mutations in the BRCA1 gene are strongly associated with an increased risk of developing **hereditary breast and ovarian cancers**.Testing for inherited harmful changes in **BRCA1**  may be done using a blood or saliva sample. That is because when a mutation is inherited, it is found in every cell in the body, including blood cells and cells that are present in saliva. Sometimes people with cancer find out that they have a harmful change in **BRCA1**  **[when their tumor is tested to see if they are a candidate for treatment](https://www.cancer.gov/about-cancer/treatment/types/biomarker-testing-cancer-treatment)** with a particular [targeted therapy](https://www.cancer.gov/Common/PopUps/popDefinition.aspx?id=CDR0000270742&version=Patient&language=en). Because harmful BRCA changes found in the tumor may have been inherited or may have arisen later in someone’s lifetime, someone with such a change in their tumor should consider getting tested to find out if the change was inherited.

The BRCA1 protein contains several important **structural domains**, such as the **RING finger domain**, **coiled-coil domain**, and **BRCT (BRCA1 C-Terminus) domain**, which contribute to its tumor suppressor function. Studying these conserved domains can provide insights into its molecular mechanisms and potential disruption in cancer.

In recent years, **bioinformatics tools** have been widely used to study protein structure, function, and interactions. Along with **domain identification** using databases like **Pfam** and **InterPro**, **NCBI BLAST** enables comparison of BRCA1 across different species to study evolutionary conservation.

Additionally, **Computer-Aided Drug Design (CADD)** has become a powerful tool to simulate **ligand–protein interactions**. **Tamoxifen**, a selective estrogen receptor modulator, is commonly used in the treatment of hormone-positive breast cancers and has shown binding interactions with BRCA-related proteins.

This project aims to conduct an **integrated in-silico study** of the BRCA1 protein using domain mapping, sequence alignment, and docking simulation. Tools such as **UniProt, Pfam, BLAST, AutoDock Vina, and PyMOL** will be utilized to analyze the protein structurally and functionally. Additionally, **R programming** will be applied for basic data visualization to represent the findings clearly.

**2.METHODOLOGY :**

**2.1 PROTEIN SEQUENCE RETRIEVAL**

The protein sequence of the **human BRCA1 gene** was retrieved from the **UniProt database**, which is a curated resource for protein sequences and functional information.

**UniProt Entry Name:** BRCA1\_HUMAN

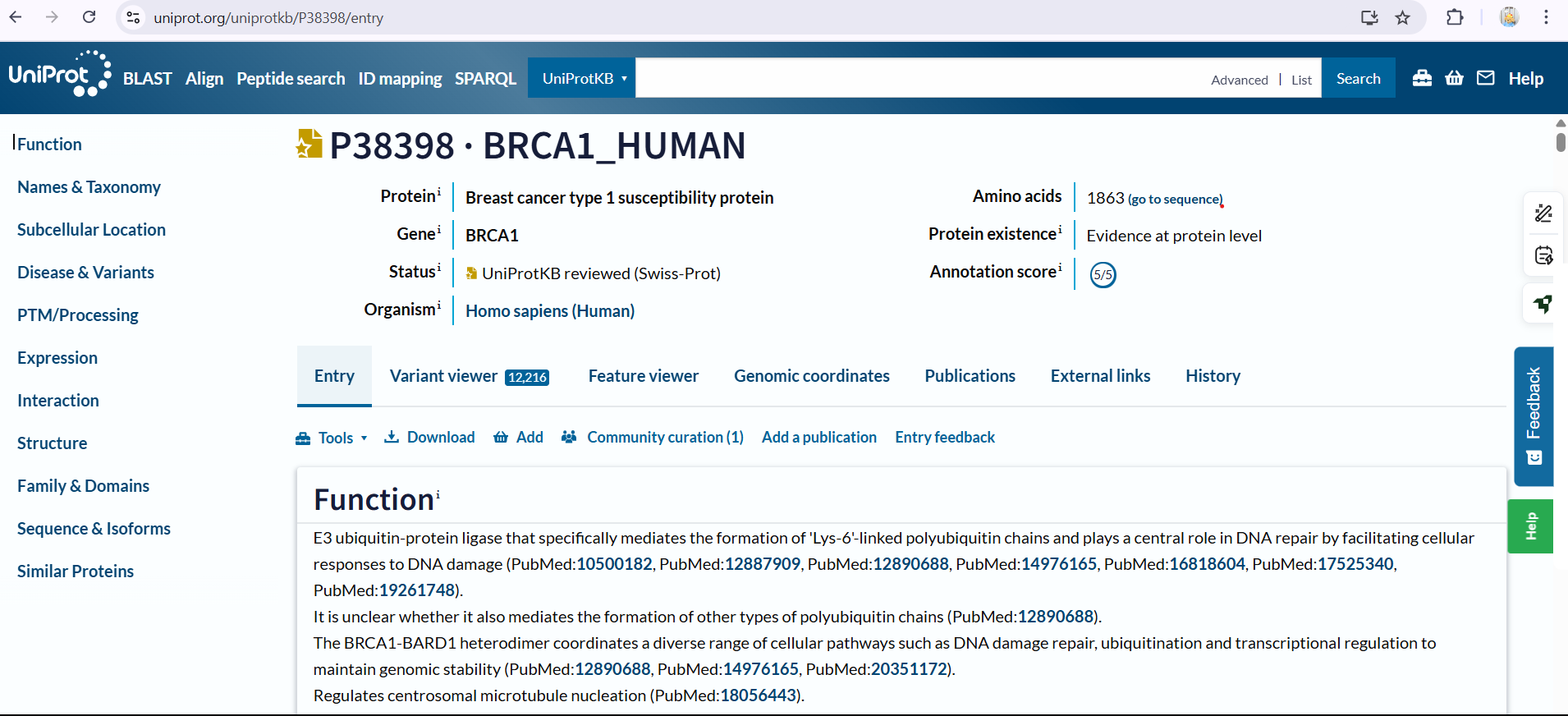
**Accession Number:** **P38398**

**Sequence Length:** 1,863 amino acids

**FASTA format** was selected and downloaded for further analysis.

The sequence serves as the foundation for all downstream bioinformatics and structural studies in this project, including domain identification, BLAST comparison, and molecular docking.

**Fig. 1: UniProt Page for BRCA1 (P38398)**



**FASTA FORMAT SEQUENCES:**

>sp|P38398|BRCA1\_HUMAN Breast cancer type 1 susceptibility protein OS=Homo sapiens OX=9606 GN=BRCA1 PE=1 SV=2

MDLSALRVEEVQNVINAMQKILECPICLELIKEPVSTKCDHIFCKFCMLKLLNQKKGPSQ

CPLCKNDITKRSLQESTRFSQLVEELLKIICAFQLDTGLEYANSYNFAKKENNSPEHLKD

EVSIIQSMGYRNRAKRLLQSEPENPSLQETSLSVQLSNLGTVRTLRTKQRIQPQKTSVYI

ELGSDSSEDTVNKATYCSVGDQELLQITPQGTRDEISLDSAKKAACEFSETDVTNTEHHQ

PSNNDLNTTEKRAAERHPEKYQGSSVSNLHVEPCGTNTHASSLQHENSSLLLTKDRMNVE

KAEFCNKSKQPGLARSQHNRWAGSKETCNDRRTPSTEKKVDLNADPLCERKEWNKQKLPC

SENPRDTEDVPWITLNSSIQKVNEWFSRSDELLGSDDSHDGESESNAKVADVLDVLNEVD

EYSGSSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKTYRKKASLPNLSHVTEN

LIIGAFVTEPQIIQERPLTNKLKRKRRPTSGLHPEDFIKKADLAVQKTPEMINQGTNQTE

QNGQVMNITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTKAEPISSSISNMELELNI

HNSKAPKKNRLRRKSSTRHIHALELVVSRNLSPPNCTELQIDSCSSSEEIKKKKYNQMPV

RHSRNLQLMEGKEPATGAKKSNKPNEQTSKRHDSDTFPELKLTNAPGSFTKCSNTSELKE

FVNPSLPREEKEEKLETVKVSNNAEDPKDLMLSGERVLQTERSVESSSISLVPGTDYGTQ

ESISLLEVSTLGKAKTEPNKCVSQCAAFENPKGLIHGCSKDNRNDTEGFKYPLGHEVNHS

RETSIEMEESELDAQYLQNTFKVSKRQSFAPFSNPGNAEEECATFSAHSGSLKKQSPKVT

FECEQKEENQGKNESNIKPVQTVNITAGFPVVGQKDKPVDNAKCSIKGGSRFCLSSQFRG

NETGLITPNKHGLLQNPYRIPPLFPIKSFVKTKCKKNLLEENFEEHSMSPEREMGNENIP

STVSTISRNNIRENVFKEASSSNINEVGSSTNEVGSSINEIGSSDENIQAELGRNRGPKL

NAMLRLGVLQPEVYKQSLPGSNCKHPEIKKQEYEEVVQTVNTDFSPYLISDNLEQPMGSS

HASQVCSETPDDLLDDGEIKEDTSFAENDIKESSAVFSKSVQKGELSRSPSPFTHTHLAQ

GYRRGAKKLESSEENLSSEDEELPCFQHLLFGKVNNIPSQSTRHSTVATECLSKNTEENL

LSLKNSLNDCSNQVILAKASQEHHLSEETKCSASLFSSQCSELEDLTANTNTQDPFLIGS

SKQMRHQSESQGVGLSDKELVSDDEERGTGLEENNQEEQSMDSNLGEAASGCESETSVSE

DCSGLSSQSDILTTQQRDTMQHNLIKLQQEMAELEAVLEQHGSQPSNSYPSIISDSSALE

DLRNPEQSTSEKAVLTSQKSSEYPISQNPEGLSADKFEVSADSSTSKNKEPGVERSSPSK

CPSLDDRWYMHSCSGSLQNRNYPSQEELIKVVDVEEQQLEESGPHDLTETSYLPRQDLEG

TPYLESGISLFSDDPESDPSEDRAPESARVGNIPSSTSALKVPQLKVAESAQSPAAAHTT

DTAGYNAMEESVSREKPELTASTERVNKRMSMVVSGLTPEEFMLVYKFARKHHITLTNLI

TEETTHVVMKTDAEFVCERTLKYFLGIAGGKWVVSYFWVTQSIKERKMLNEHDFEVRGDV

VNGRNHQGPKRARESQDRKIFRGLEICCYGPFTNMPTDQLEWMVQLCGASVVKELSSFTL

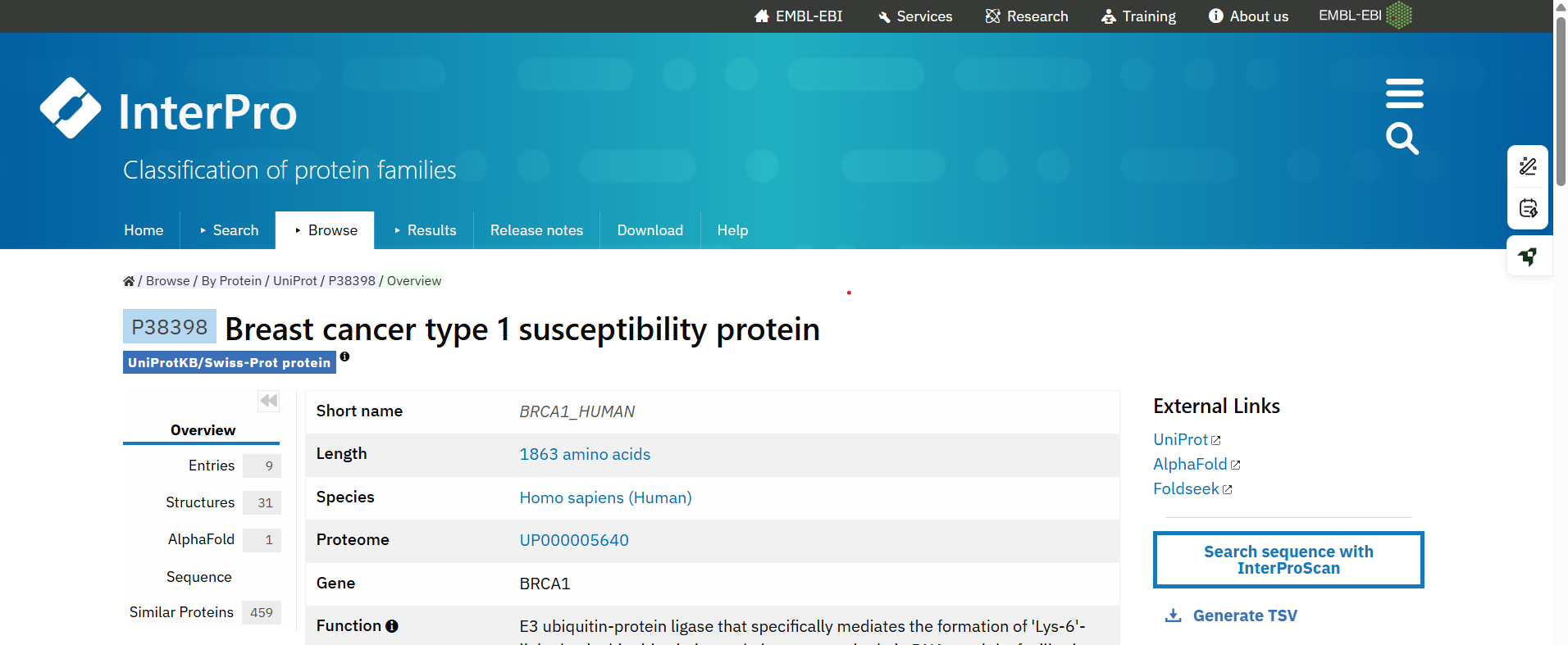
GTGVHPIVVVQPDAWTEDNGFHAIGQMCEAPVVTREWVLDSVALYQCQELDTYLIPQIPH

SHY

**2.2 DOMAIN ANALYSIS**

Domains are distinct functional and/or structural units in a protein. Usually they are responsible for a particular function or interaction, contributing to the overall role of a protein. Domains may exist in a variety of biological contexts, where similar domains can be found in proteins with different functions.

To identify the functional domains present in the BRCA1 protein, two well-established databases were used: **Pfam** and **InterPro**. These tools provide information about conserved domain families, domain architecture, and functional annotations.



#### ****Pfam Search:****

The UniProt accession number **P38398** was used to search in Pfam.

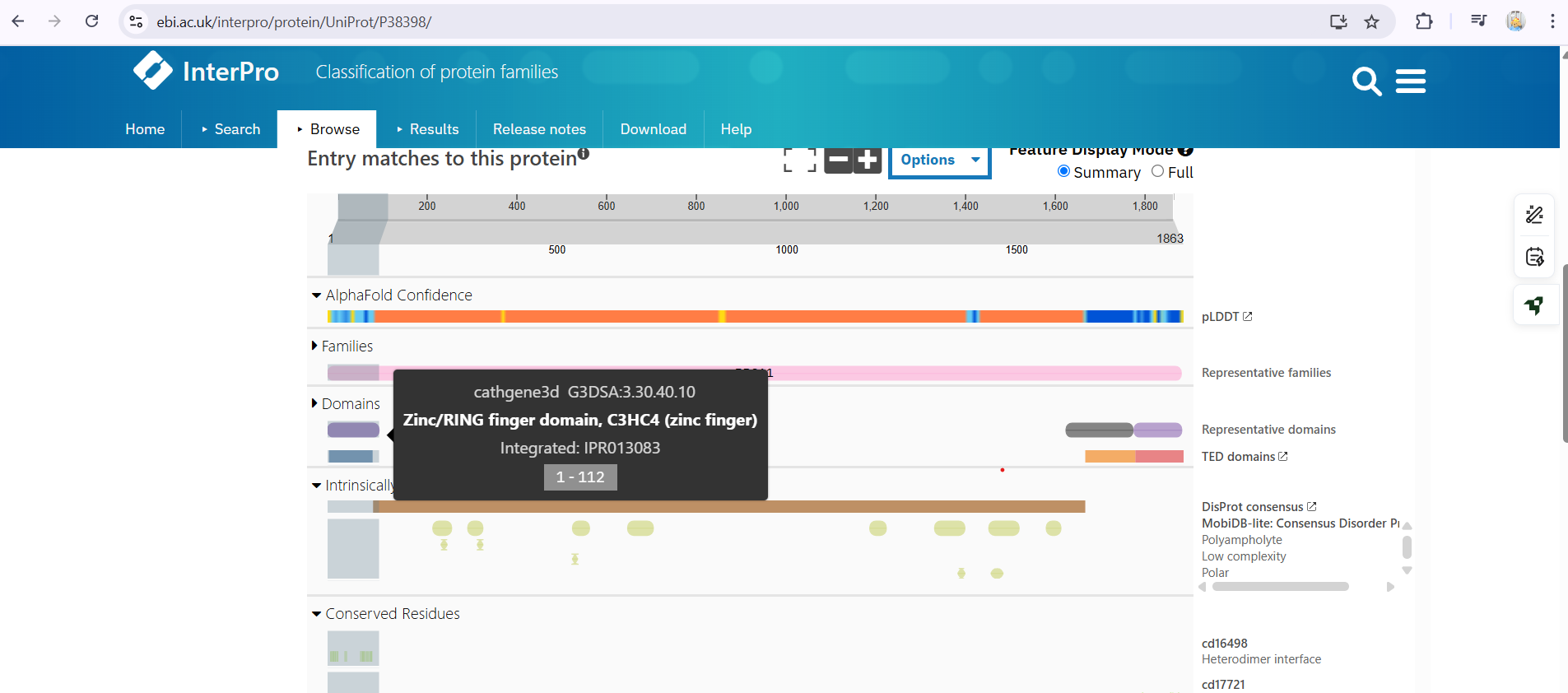
The domain architecture of BRCA1 was visualized, showing conserved domains such as:

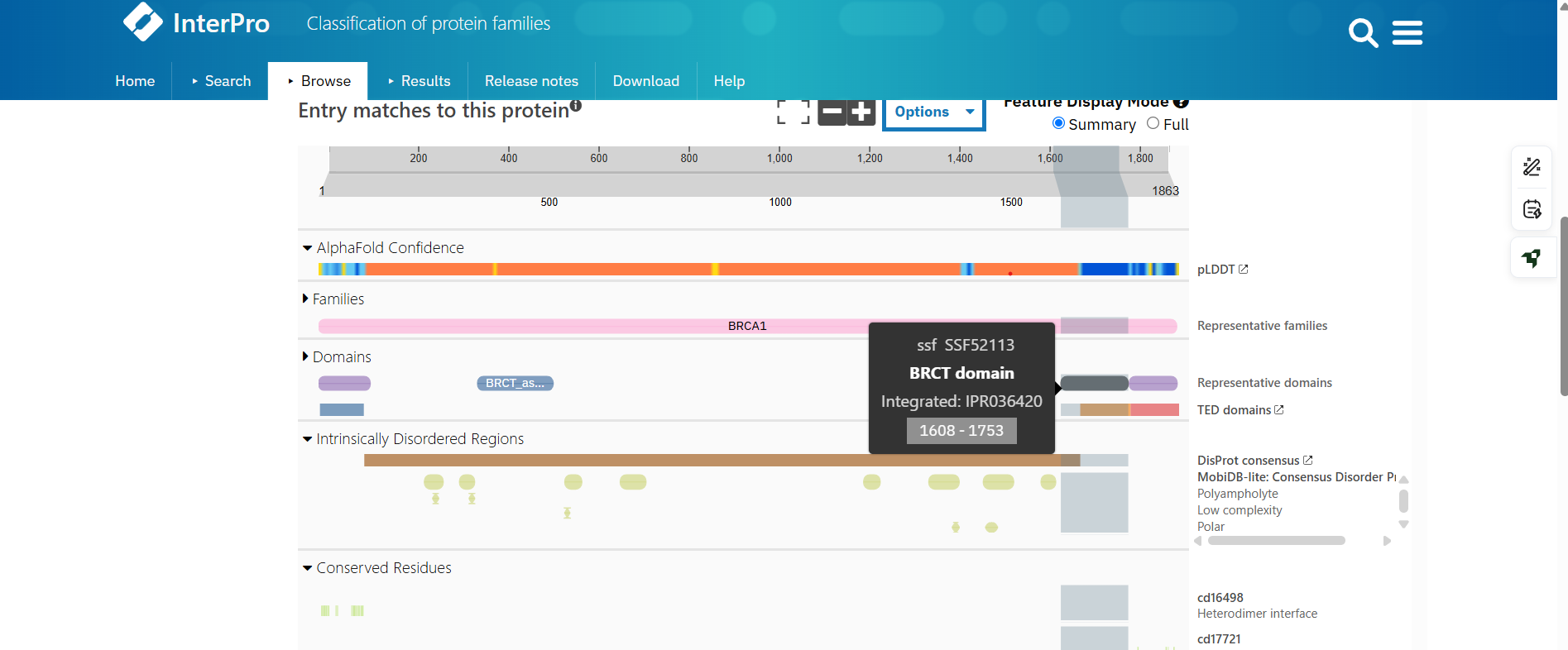
**RING finger domain (Pfam: PF00097)** at the N-terminal

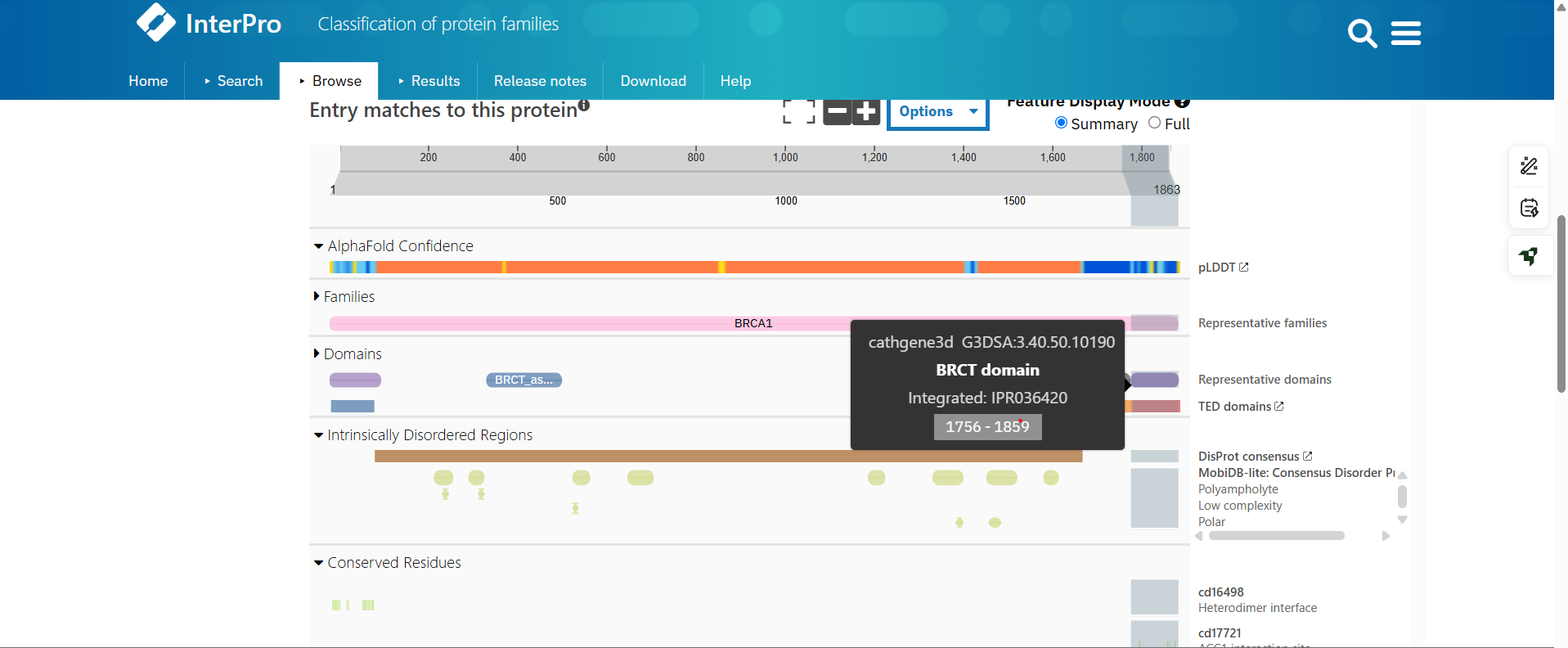
**BRCT domains (Pfam: PF00533)** at the C-terminal

The domains play crucial roles in **DNA repair, protein–protein interactions**, and **ubiquitin ligase activity**.

**Figure 2: Domain architecture of BRCA1 as predicted by Pfam (P38398)**  
The BRCA1 protein contains a RING domain and two BRCT domains that are highly conserved.







| **Domain Name** | **Start–End Position** | **Function** |
| --- | --- | --- |
| RING finger | | 1-112 | Ubiquitin ligase activity |
| BRCT 1 | | 1608-1753 | Protein–protein interaction in DNA repair |
| BRCT 2 | | 1756–1859 | Protein–protein interaction |

**2.3 Sequence Comparison (BLAST Analysis)**

### ****Sequence Comparison Using BLASTP****

To assess the evolutionary conservation of BRCA1, the protein sequence was subjected to a **BLASTP (Protein–Protein BLAST)** search using the **NCBI BLAST tool**.

The FASTA sequence retrieved from **UniProt (P38398)** was used as the query input.

The **non-redundant protein sequences (nr)** database was selected.

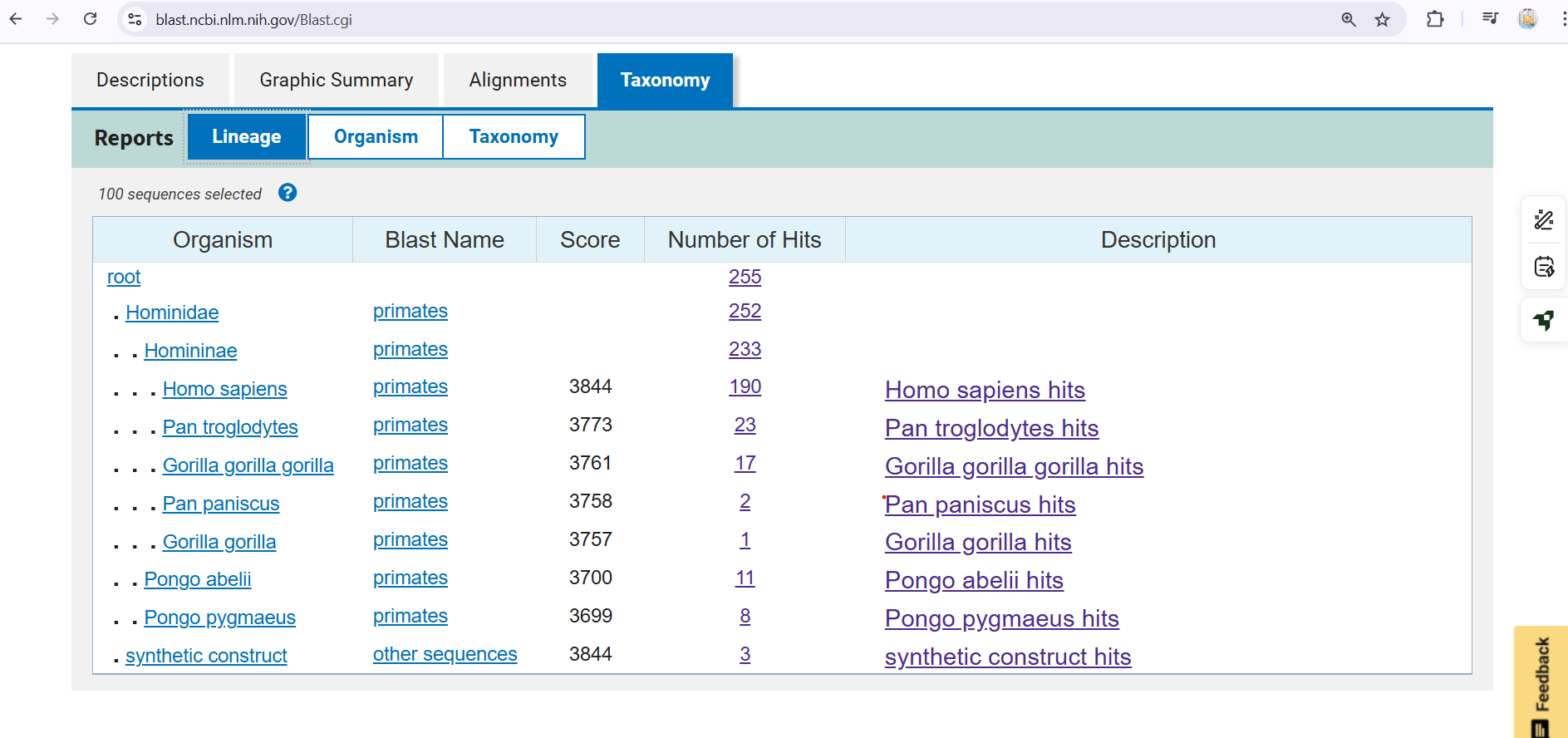
**Default parameters** were used, and the search was performed against the **Homo sapiens** BRCA1 protein.

### Purpose:

To **compare the BRCA1 protein sequence in humans** with other organisms, and to identify:

* **How conserved (similar)** the protein is across species
* **Which parts of the protein (e.g., domains)** are **evolutionarily important**
* Potential **model organisms** for future cancer studies (e.g., mouse, zebrafish)

**Figure 3: Taxonomic distribution of BLASTP hits for BRCA1 protein**  
Most homologous sequences were found in primates, including Homo sapiens, Pan troglodytes (chimpanzee), and Gorilla species, indicating a high level of evolutionary conservation within the Hominidae family.



The BRCA1 protein sequence was analyzed using **BLASTP** against the non-redundant protein database. The taxonomy view of the results showed a high number of homologous sequences within the **Hominidae** family.

**190 hits** were observed in Homo sapiens

Additional hits were found in related primates such as:

Pan troglodytes (chimpanzee): 23 hits

Gorilla gorilla: 17 hits

Pongo abelii (Sumatran orangutan): 11 hits

The **high BLAST scores (3773–3844)** and **extensive query coverage** across these species suggest strong evolutionary conservation of BRCA1 among primates.

**2.4 Ligand–Protein Docking Simulation**

### ****2.4.1 Objective****

The aim of this study is to predict the potential binding interaction between **Tamoxifen**, a known breast cancer therapeutic agent, and the **BRCT domain of BRCA1**, a tumor suppressor protein involved in DNA repair mechanisms. Molecular docking helps estimate the binding affinity and preferred orientation of the ligand when bound to the target protein.

### ****2.4.2 Tools and Resources****

**CB-Dock2** Web Tool (AutoDock Vina-based):  
https://cadd.labshare.cn/cb-dock2/

**Protein structure (1JNX)**: Retrieved from RCSB PDB

**Ligand structure (Tamoxifen)**: Retrieved from PubChem

Software performs cavity detection, ligand preparation, and docking simulation automatically.

### ****2.4.3 Methodology****

#### **Ligand Preparation**

#### The 3D structure of Tamoxifen was downloaded from PubChem and converted to .pdb format.The structure was directly uploaded to CB-Dock2, which performed internal optimization.

#### **Protein Preparation**

#### The BRCT domain of **BRCA1** (PDB ID: 1JNX) was downloaded from the RCSB Protein Data Bank in **legacy** .pdb **format**.The structure was uploaded to CB-Dock2 without further modification.

#### **Docking Procedure**

Both the receptor (BRCA1) and ligand (Tamoxifen) were uploaded.CB-Dock2 detected potential binding cavities and performed docking using AutoDock Vina.Five binding poses were generated and ranked by **binding energy (kcal/mol)**.

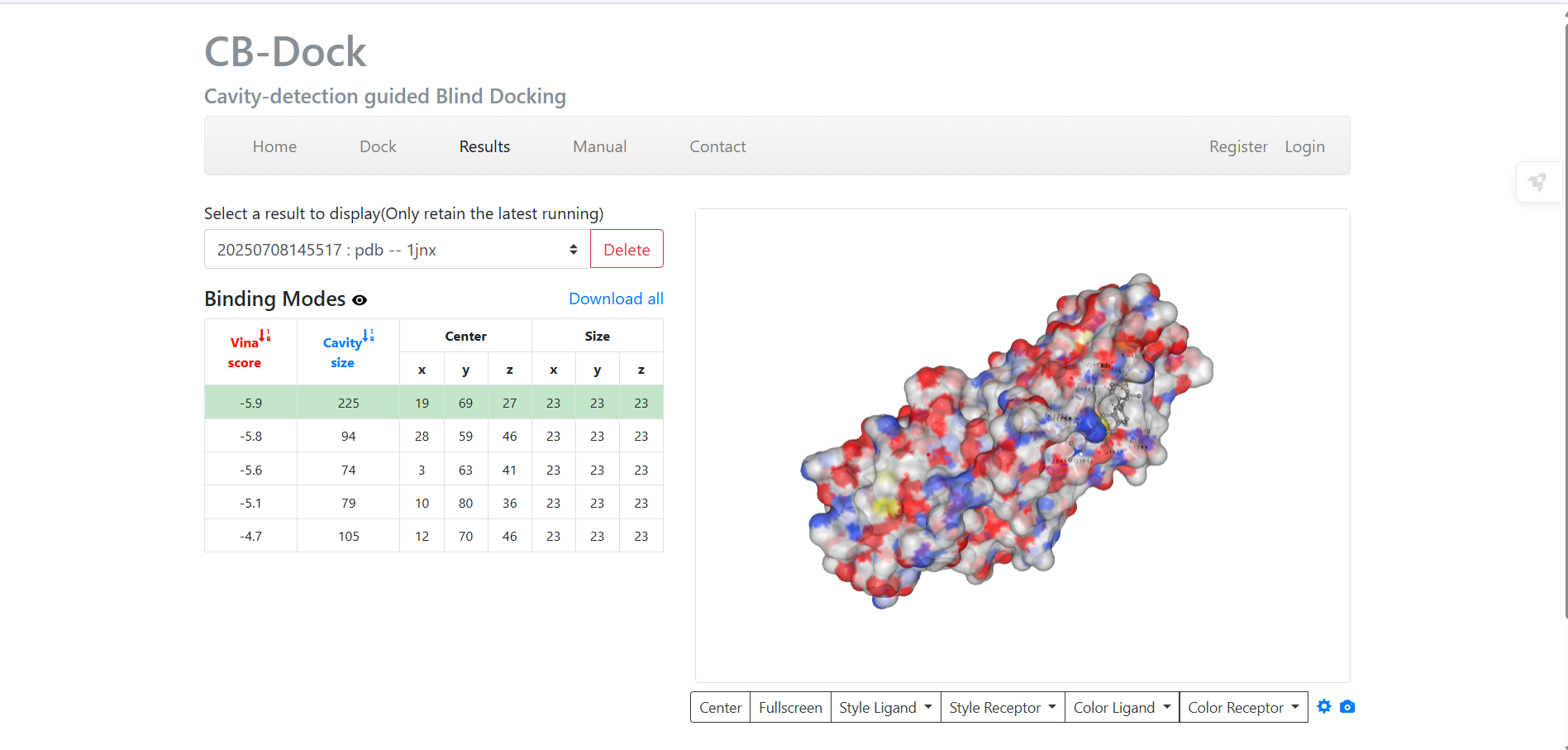
2.4.4 Results and Analysis

|  |  |  |  |
| --- | --- | --- | --- |
| Pose | Binding energy(kcal/mol) | Cavity size | Center(x,y,z) |
| 1 | -5.9 | 225 | (19,27,23) |
| 2 | -5.8 | 94 | (28,56,23) |
| 3 | -5.6 | 74 | (3,60,23) |
| 4 | -5.1 | 109 | (30,46,23) |
| 5 | -4.7 | 105 | (12,46,23) |

The **best docking pose** showed a **binding energy of –5.9 kcal/mol**, indicating moderate affinity.Tamoxifen was observed to bind within a pocket in the BRCT domain.The cavity had a size of 225 Å³, centered at coordinates **(19, 27, 23)**.

2.4.5 Visualization

**Figure 4**: Binding pose of Tamoxifen docked with BRCA1 (PDB: 1JNX) shown using CB-Dock2. The receptor is represented in surface mode and the ligand (Tamoxifen) is shown in yellow.



### ****2.4.6 Interpretation****

The docking results suggest that **Tamoxifen** can form a stable interaction with the **BRCT domain of BRCA1**, although the binding energy is lower compared to classical Tamoxifen targets like estrogen receptors. This may indicate off-target or secondary interactions of Tamoxifen with BRCA1-related signaling pathways in breast cancer treatment.

## ****3. Discussion****

The primary objective of this mini-project was to explore the interaction of **Tamoxifen**, a selective estrogen receptor modulator (SERM), with the **BRCT domain of BRCA1**, a critical DNA repair protein. This computational study was carried out using **Pfam domain analysis**, **sequence similarity via BLAST**, and **molecular docking using AutoDock Vina (via CB-Dock2)**.

### ****3.1 Interpretation of Domain and Homology Analysis****

The **Pfam scan** confirmed the presence of the **BRCT domain**, a highly conserved region that interacts with phosphorylated proteins to regulate DNA damage response and cell cycle control.  
The **BLAST search** results showed that BRCA1 is highly conserved among **Homo sapiens** and other **primates**, indicating its essential role in maintaining genomic stability. The absence of top hits in rodents like Mus musculus suggests evolutionary divergence in this region, especially within its BRCT domain.

### ****3.2 Docking Significance****

Tamoxifen is primarily known to target estrogen receptors in breast cancer treatment. However, the docking results suggest a **possible moderate binding affinity (–5.9 kcal/mol)** to BRCA1. This opens the possibility that Tamoxifen could have **secondary molecular effects**, perhaps influencing BRCA1-related signaling or DNA repair functions indirectly.

While the binding energy is not as strong as typical enzyme-inhibitor complexes, it still represents a **biologically meaningful interaction** worth exploring in wet-lab settings.

### ****3.3 Relevance to Cancer Research****

BRCA1 mutations are well known to contribute to **hereditary breast and ovarian cancers**. Understanding how drugs like Tamoxifen may bind to or modulate BRCA1 can be valuable for:

Exploring off-target drug effects,

Designing combination therapies, or

Developing new small molecules targeting BRCT domains.

This project demonstrates the **importance of in silico docking** as a first step in evaluating drug–protein interactions for therapeutic applications.

## ****4. Conclusion****

This mini-project explored the molecular interaction between **Tamoxifen**, a widely used anti-breast cancer drug, and **BRCA1**, a tumor suppressor protein involved in DNA repair. Using a combination of **Pfam domain identification**, **sequence similarity analysis (BLAST)**, and **molecular docking (CB-Dock2 with AutoDock Vina)**, the following conclusions were drawn:

**Pfam analysis** revealed the presence of the **BRCT domain** in BRCA1, which is essential for its interaction with DNA repair proteins.

**BLAST results** confirmed a high level of conservation of BRCA1 among human and primate sequences, emphasizing its evolutionary importance.

**Docking studies** demonstrated that Tamoxifen binds moderately to a predicted surface pocket within the BRCT domain of BRCA1, with a binding energy of **–5.9 kcal/mol**.

Although BRCA1 is not the primary target of Tamoxifen, the observed interaction may suggest secondary effects or off-target binding potential. The study underscores the usefulness of **in silico tools** in predicting drug–protein interactions and sets a foundation for future **experimental validation**.

## ****5. References****

**1.Pfam Protein Families Database**  
Mistry, J., et al. (2021). Pfam: The protein families database in 2021. Nucleic Acids Research, 49(D1), D412–D419.  
Retrieved from: https://pfam.xfam.org

**2.BLAST (Basic Local Alignment Search Tool)**  
Altschul, S. F., et al. (1990). Basic local alignment search tool. Journal of Molecular Biology, 215(3), 403–410.  
Tool accessed at: [https://blast.ncbi.nlm.nih.gov/](https://blast.ncbi.nlm.nih.gov/" \t "_new)

**3.UniProt Database**  
The UniProt Consortium. (2023). UniProt: the universal protein knowledgebase in 2023. Nucleic Acids Research, 51(D1), D523–D531.  
Retrieved from: [https://www.uniprot.org/](https://www.uniprot.org/" \t "_new)

**4.RCSB Protein Data Bank**  
Burley, S. K., et al. (2021). RCSB Protein Data Bank: powerful new tools for exploring 3D structures of biological macromolecules. Nucleic Acids Research, 49(D1), D437–D451.  
Structure of BRCA1 accessed from: https://www.rcsb.org/structure/1JNX

**5.CB-Dock2: Protein–Ligand Docking Server**  
Liu, Y., et al. (2022). CB-Dock2: Improved protein–ligand blind docking by integrating cavity detection, docking, and homologous template fitting. Nucleic Acids Research, 50(W1), W159–W164.  
Tool accessed at: https://cadd.labshare.cn/cb-dock2/

**6.PubChem Compound Database (Tamoxifen)**  
Kim, S., et al. (2021). PubChem in 2021: new data content and improved web interfaces. Nucleic Acids Research, 49(D1), D1388–D1395.  
Compound accessed at: [https://pubchem.ncbi.nlm.nih.gov/compound/Tamoxifen](https://pubchem.ncbi.nlm.nih.gov/compound/Tamoxifen" \t "_new)