IDENTIFICATION AND CLASSIFICATION OF SMALL-MOLECULE BINDING SITES IN THE BREAST CANCER PROTEOME

AGAM KANODIA, 131518

KAMAKSHI BHARDWAJ, 131506

MAY-2017

Submitted in partial fulfilment of the requirements for the award of the degree of Bachelor of Technology in Bioinformatics

SUPERVISOR: Dr. JAYASHREE RAMANA

DEPARTMENT OF BIOTECHNOLOGY AND BIOINFORMATICS

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY

WAKNAGHAT, SOLAN-173234, HIMACHAL PRADESH

Table of Contents

CERTIFICATE

This is to certify that the project work entitled "**Identification and Classification of Small-Molecule Binding Sites in the Breast Cancer Proteome"** is a bonafide work carried out by **Agam Kanodia (131518)** with **Kamakshi Bhardwaj (131506)** during period from August 2016 to April 2017 under the supervision of **Dr. Jayashree Ramana**, Assistant Professor, Department of Biotechnology and Bioinformatics, Jaypee University of Information Technology, Waknaghat, towards partial fulfilment of the requirements for the award of the degree of Bachelor of Technology and submitted to **Department of Biotechnology and Bioinformatics**, Jaypee University of Information Technology, Waknaghat.

Date:

_____

Dr. Jayashree Ramana (Supervisor)

Assistant Professor (Senior Grade)

Department of Biotechnology and Bioinformatics

Jaypee University of Information Technology

DECLARATION

We hereby declare that the work presented in this report entitled "Identification and Classification of Small-Molecule Binding Sites in the Breast Cancer Proteome " in partial fulfilment of the requirements for the award of the degree of Bachelor of Technology in Bioinformatics submitted in the Department of Biotechnology and Bioinformatics, Jaypee University of Information Technology, Waknaghat, Solan-173234, Himachal Pradesh is an authentic record of our own work carried out over a period from August 2016 to April 2017 under the supervision of Dr Jayashree Ramana, Assistant Professor (Senior Grade), Department of Biotechnology and Bioinformatics. The matter embodied in the report has not been submitted for the award of any other degree or diploma.

Agam Kanodia 131518

Kamakshi Bhardwaj 131506

ACKNOWLEDGEMENT

This thesis represents a part of our dream which would not be possible without the contribution of many others, who directly or indirectly contributed in the development of this work.

We would like to express our sincere gratitude and indebtedness to our guide Dr. Jayashree Ramana for her exemplary guidance, valuable feedback and constant encouragement throughout the duration of the project. We are highly grateful to her for providing us this opportunity.

We would hereby express our sincere thanks to all the faculties of BI department, Dr. Chittranjan Rout, Dr. Jayashree Ramana, Dr. Tiratha Raj Singh and Dr. Raghu M. Yennamalli whose teachings gave us conceptual understanding and clarity which ultimately made our work easier. We greatly acknowledge Dr. R.S. Chauhan, Professor and Head of Department for valuable assistance.

Credit also goes to all other friends and family as their encouragement kept us in good stead. Their encouragement and continuous support has given us strength and confidence to complete the initial part of the project without any difficulty.

Signature of Student
Name: Agam Kanodia
Date:

Signature of Student
Name: Kamakshi Bhardwaj
Date:

ABSTRACT

We have performed the RNA-Seq data analysis on the dataset containing the tumor and normal breast samples and identified deregulated genes. Further, to find the structures of the protein products of these genes we searched the Protein Data Bank and then we scanned the surface of these protein structures and identified binding sites. We then classified the binding sites obtained into enzyme active sites, protein-protein interaction sites, or other sites that may lie outside of functional sites on the basis of their surface characteristics and on the basis of the information collected from various databases like CSA, UniProt and BRENDA. Our study focuses on the high throughput screening of breast cancer specific binding sites and their categorisation for possible evaluation to evaluate and pave an ease towards novel drug target ident ification which may prove to be clinically- relevant protein targets.

Signature of Student	Signature of Student	Signature of Supervisor
Name: Agam Kanodia	Name: Kamakshi Bhardwaj	Name: Dr. Jayashree Ramana
Date:	Date:	Date:

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1 – INTRODUCTION

1.1.OBJECTIVE

The objective of the project is to identify small-molecule binding sites on proteins corresponding to the de-regulated genes involved in the Breast cancer and to broadly classify them into three categories namely: Enzymatic active sites (ENZ), Protein-protein interaction (PPI) and Others (OTH).

1.2. CANCER

Cancer is an abnormal growth of cell likely to affect other parts of the body. It takes place due to factors such as mutations or abnormal transformations in the genes that helps in regulation of growing cells thereby keeping them fit. A tumor can be of two categories: benign and malignant. Benign tumors are not cancerous. These cells are almost normal in appearance, do not attack any nearby tissues nor extend to the entire body. While malignant tumors are considered to be cancerous. If not treated in time can spread in the other parts of the body as well [1] [2].

1.3. BREAST CANCER

Breast cancer is malignant tumor that usually starts off in the inner lining of milk ducts or the lobules that supply them with milk. *Lobular carcinoma* is a breast cancer that starts off in the lobules, while the one that develops from the ducts is known as *ductal carcinoma.* It is the most common invasive cancer in female's worldwide accounting for 16% of all female cancers and 22.9% of invasive cancers in women [3]. Breast cancer signs and symptoms often appear when the tumour grows large enough to be felt as a lump in the breast or when the cancer spreads to surrounding tissues and organs. The most common symptom of ductal carcinoma is a firm or hard lump that feels very different from the rest of the breast. Lobular carcinoma often does not form a lump. It feels more like the tissue in the breast is getting thicker or harder [4]. According to National Cancer Institute, 232340, breast cancers are reported in the USA each year, as well as about 39,620 deaths is caused

by the disease. The stats reported in **Figure 1** shows the incidence of breast cancer and mortality rate in three countries- India, U.S and China. Now talking about the genetic mutations that can lead to breast cancer, the most significant ones are the BRAC1 and BRAC2 gene mutations. These mutations account for up to 90% of the total genetic influence with a risk of breast cancer of 60-80% in those affected. Other significant mutations include p53, PTEN, STK11, ATM, BRIP1 gene mutations.

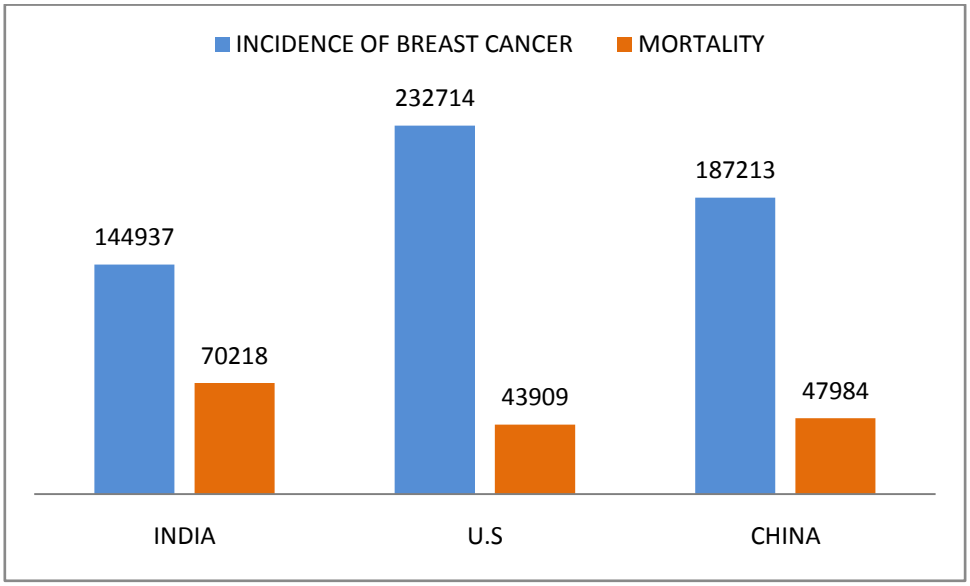


Figure 1: Statistics for breast cancer in India, U.S. and China for the Year 2014(Adopted from WHO, International Agency for Research on Cancer

1.4.Next generation RNA-Seq

Next generation RNA-sequencing is an emerging technology that is rapidly outcompeting microarrays as the technology of choice for whole transcriptome studies. However, the analysis of the RNA data generated by NGS often poses a significant hurdle for many biologists as the data is very huge. RNA-Seq extends the possibilities of transcriptome studies to the analysis of previously unidentified genes and of splice variants. RNA-Seq analysis is a powerful method for discovering, profiling, and quantifying RNA transcripts. RNA-Seq is used to analyse the continually changing cellular transcriptome [5]. Specifically, RNA-Seq facilitates

the ability to look at alternative gene spliced transcripts, post-transcriptional modifications, gene fusion, mutations/SNPs and changes in gene expression [6]. Cancer being a genetic disease is driven by heritable or somatic mutations therefore NGS technology particularly RNA-sequencing will have a remarkable effect on the detection, management and treatment of disease.

1.5. BINDING SITES

Binding sites are the regions on a protein or a piece of DNA or RNA to which ligands may form a chemical bond.In particular, the protein-binding site is the most fascinating and important mechanistic arbiter of protein function. The binding pocket is in the state of a three-dimensional cleft that is made out of amino acids from various residues of the essential amino acid sequence. Amino acids forming the pocket are usually not adjacent to each other in the primary structure, but form the active site as a result of folding in creating the tertiary structure.Identification of binding sites is accomplished by analysing the 3-D structure of a protein. These structures can be used to scan for binding pockets to develop lead compounds for drugdiscovery.In addition to detecting binding sites, algorithms have been developed to score these binding sitesbased on whether they can accommodate a small molecule [7] [8].

In this study, we analyzed RNA-Seq data of Breast cancer taking 6 tumor samples and 5 normal samples from the GEO datasets and identified genes that are deregulated by comparing the data between cancer and normal samples. Protein products of these genes were scanned for binding sites that possess shape and physicochemical properties that can accommodate small-molecule probes or therapeutic agents (druggable).The functional relevance of binding sites is explored by classifying them into known enzyme active sites(ENZ), protein-protein interaction sites(PPI), or other sites(OTH) that may lie outside of functional sites [9].

CHAPTER 2 – DATASETS AND METHODS

2.1. METHODOLOGY

The methodology adopted is described below:

1. Download RNA-Seq NGS data of Breast cancer from the NCBI GEO database which satisfy following criteria:
   a. Include both normal and tumor samples with each type having a count of more than one.
2. For the downloaded dataset:
   a. Convert data to .fastq format from .sra format using SRAToolkit.
   b. Perform the pre-processing of RNA-Seq reads which involves the following steps:
      i. Quality checking using FastQC tool to get a quick impression of the quality of data.
      ii. Filtering using filteR tool in order to filter out the lower quality data, adapters, PCR primers and other artefacts which would otherwise hinder in further analysis.
   c. Perform Mapping of RNA-Seq reads using splice aware aligner TopHat, which involves alignment of the filtered reads obtained against the UCSC *H. sapiens* reference genome (build hg19) giving you the output file in SAM format.
   d. Perform Transcriptome assembly and differential expression analysis on the output files obtained from TopHat using Cufflinks tool to identify differentially expressed transcripts and genes giving you the output file containing differential gene expression values across the given sample.
3. From the output file obtained in the last step take the gene symbols and map them to their respective UniProt ids using UniProt Retrieve/Id mapping tool and also download the fasta sequences of each of them.
4. Perform CD-HIT on each of the sequence files obtained to generate non-redundant entries.
5. Perform BLASTp on CD-HIT output files to validate the structures of our query if available or not. Download the blast output file.

6. From the downloaded files, extract the hits with 98% sequence identity and then obtain the PDB IDs for these entries using UniProt retrieve/ID mapping tool.
7. Obtain the resolution for each of the PDB ID obtained in the last step.
8. Remove the duplicates from the data, keeping the entries having the highest resolution among duplicates.
9. Sort the above data on the basis of resolution, keeping only the PDB IDs having resolution between 0 and1.6.
10. Retrieve the structures of the PDB IDs obtained in the above step using RCSB downloader.
11. Put each of the structure in the DoGSiteScorer to obtain the binding site information which includes D-score and position of each binding pocket and the residues involved in it.
12. Functional data is collected for each of the PDB IDs from databases: BRENDA, CSA, and UniProt.
13. Binding site classification is done on the basis of the surface characteristics and the functional data. Each binding site identified by DoGSiteScorer was visually inspected and manually annotated to determine its functional role in the protein and is classified into three categories: Enzymatic binding site, Binding site involved in Protein-Protein interaction and Others.
    a. It is said to be an Enzymatic binding site if any of the catalytic residue is present inside the binding pocket or there is information available in BRENDA database or CSA database for the corresponding PDB IDs.
    b. If the binding site was at a protein-protein interaction (PPI) interface on the original structure, the binding site was labeled ‘PPI’.
    c. Otherwise, if the binding site was neither enzymatic nor part of the interaction interface, it was labeled ‘Other’ (OTH).

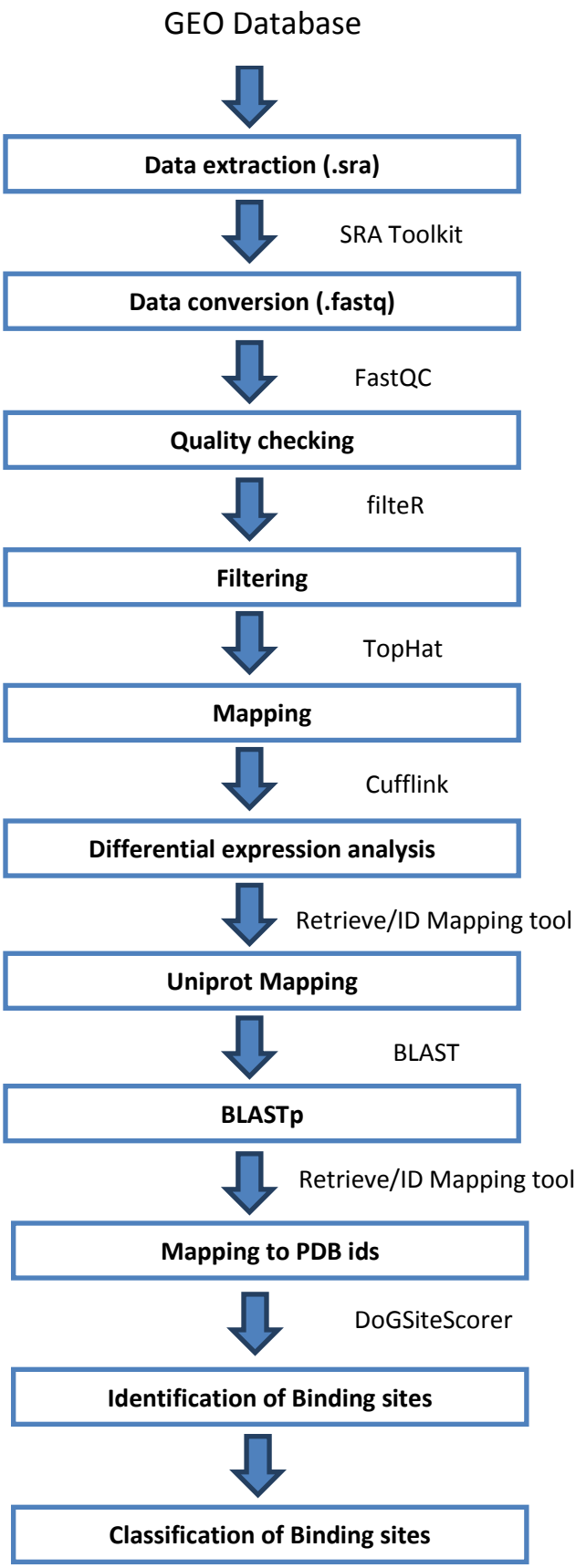


Figure 2: Methodology

2.2. DATABASES USED

1. SRA:Sequence Read Archive database provides a public repository for DNA sequencing data, especially the short fragmented reads generated as an output from the high throughput sequencing techniques. These reads information comes from various sequencing platforms, such as Illumina Genome Analyzer, Roche 454 GS System, HelicosHeliscope, Complete Genomics, and Pacific Biosciences SMRTApplied Biosystems SOLiD System and many more [10].

2.GEO:It is a public repository that archives and freely distributes next-generation sequencing (NGS), microarray and other forms of high-throughput genomic data sets. Approximately 90% of the data in GEOare gene expression studies which catalogues broad range of biological data including evolution,ecology,disease, development, immunity, metabolism,toxicology, and more [11].

3. UniProt: is a freely available database, containing functional information, protein sequences and its biological annotation derived from either research literature or genome sequencing projects [12].

4. PDB: is a freely accessible repository having information majorly about 3-D structures of proteins, nucleotides and large biological molecules [13].

5.BRENDA: is a comprehensive enzyme information system representing a comprehensive compendium of enzymes identified till date.It is a resource that includesbiochemical and molecularinformation on enzymes that have been categorised by the IUBMB. Every categorised enzyme is characterized with respect to its catalysed bio-chemical reactions [14].

6. CSA:is a database detailing enzyme catalytic residues and active sites in enzymes of 3-D structure. Database consists of organisation of catalytic amino acid which contains only those amino acid thought to be directly involved in nearly aspect of the reaction which is catalysed by an enzyme [15].

2.3. TOOLS AND SOFTWARE USED

1. SRA Toolkit: It is software provided by the SRA database that allows to access data present in the SRA database and converts it into files from SRA format to many formats including fastq format.

2. FastQC: It is a quality control tool which is used to perform quality checks on raw sequence data obtained from highthroughput sequencing (HTS) techniques [16].

3. filteR: It is used to filter out the lower quality reads determined by performing quality check using FastQC, which would otherwise hinder the protocol.

4. TopHat: It is a fast splice mapper for the RNA-Seq reads. RNA-Seq reads aligned to mammalian sized genomes by using ultrahigh-throughput short read aligner Bowtie, which further analyse the mapping reads to identify spliced junctions between set of exons [17].

5. Cufflink: It is an assembler which assembles transcripts and estimates the abundances on the basis of reads. Also, it tests for differential expression (DE) and regulation or dysregulation in RNA-Seq sample datasets [18].

6. UniProt Retrieve/ID mapping tool: This tool is used to retrieve the corresponding UniProt entries to a list of identifiers submitted, or to map them from or to an external database.

7. CD-HIT: It is a globally used program for comparing and clusteringnucleotide/protein sequences besides, it also reduces the computational and manual efforts required to perform analysis and aids in deciphering the data structure and correct the bias within a dataset [19].

8. BLAST: Basic Local Alignment Search Tool (BLAST) compares nucleotide or protein sequences against sequence databases and on the basis of statistical significance calculated it can be used to not only decipher functional and evolutionary relationship between sequences but also in identification of members of various gene families [20].

9. RCSB downloader: It is a tool usually used to download FASTA sequence files, ligand data files for one or many PDB entries and experimental and coordinate data files.

10. DoGSiteScorer: It is an automated analysis tool for pocket detection which can also be used for protein druggability assessment and analysis[21].

2.4. DATASETS

SRA data of invasive carcinoma of the breast were downloaded from the NCBI GEO database. The downloaded RNA-Seq NGS data (Illumina Genome AnalyzerIix platform) has 6 Breast cancer samples and 5 normal breast samples as listed in the Table 1.

Breast Cancer samples(Source: Breast invasive ductal carcinoma)	Normal breast samples(Source:Adult normal breast)
GSM715426	GSM984203
GSM715427	GSM984204
GSM715428	GSM984205
GSM715429	GSM984206
GSM715430	GSM984207
GSM715431	

Table1: Breast cancer GEO datasets included in the study.

CHAPTER 3 – RESULTS AND DISCUSSION

3.1. RNA-Seq ANALYSIS

We did RNA-Seq analysis on the dataset containing 5 normal breast samples and 6 tumorous samples. In the analysis process we checked the quality of the reads using FastQC, filtered out the poor quality reads using filteR, filtered reads obtained were aligned against the UCSC *H. sapiens* reference genome using TopHat and then weidentified differentially expressed transcripts and genes using Cufflinks tool giving us the results shown in Table 2 and Figure 3.

Total number of genes	22,403
Number of up-regulated genes	8,716
Number of down-regulated genes	9,025
Number of non-regulated genes	4,662

Table 2: Output of the differential expression analysis for RNA-Seq data.

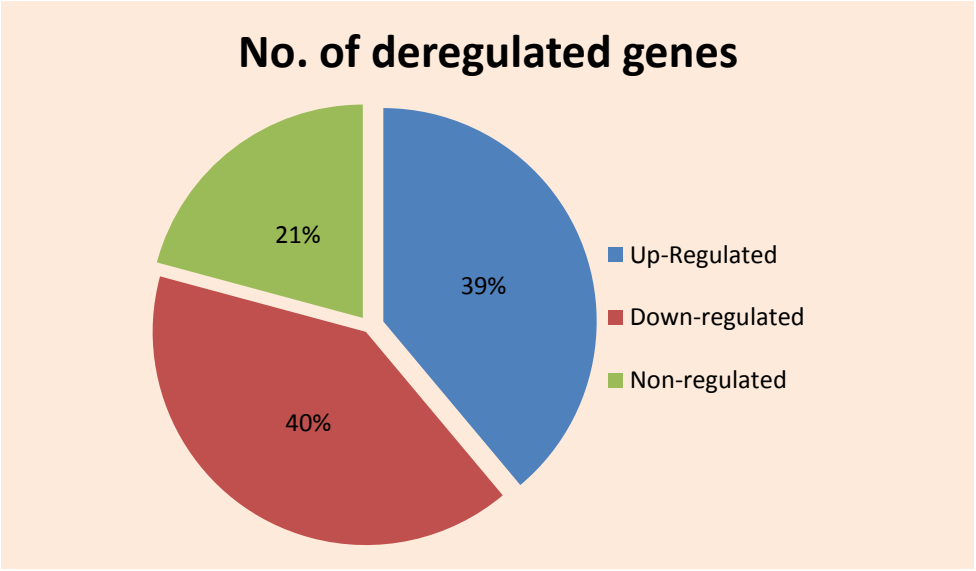


Figure 3: A pie-chart showing total number of differentially regulated genes: up-regulated, down-regulated and non-regulated.

3.2. UniProt MAPPING FROM GENE SYMBOLS TO UniProtKB AC/ID

We mapped the gene symbols of the deregulated genes to their respective UniProtKB AC/ID using UniProt Retrieve/Id Mapping tool and also downloaded the fasta sequences of the mapped ids.

- Out of 8716 up-regulated genes 8165 were mapped to 8266 reviewed Swiss-Prot IDs as shown in Figure 4.

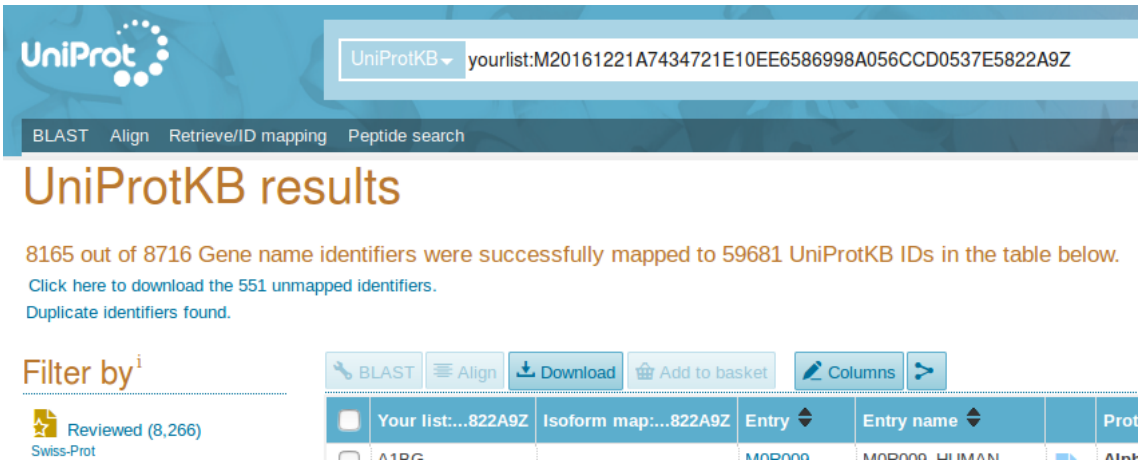


Figure 4: Uniprot mapping from gene symbol to UniProtKB AC/ID output1

- Out of 9025 Down-regulated genes 8085 genes were successfully mapped to 8196 reviewed Uniprot IDs as shown in Figure 5.

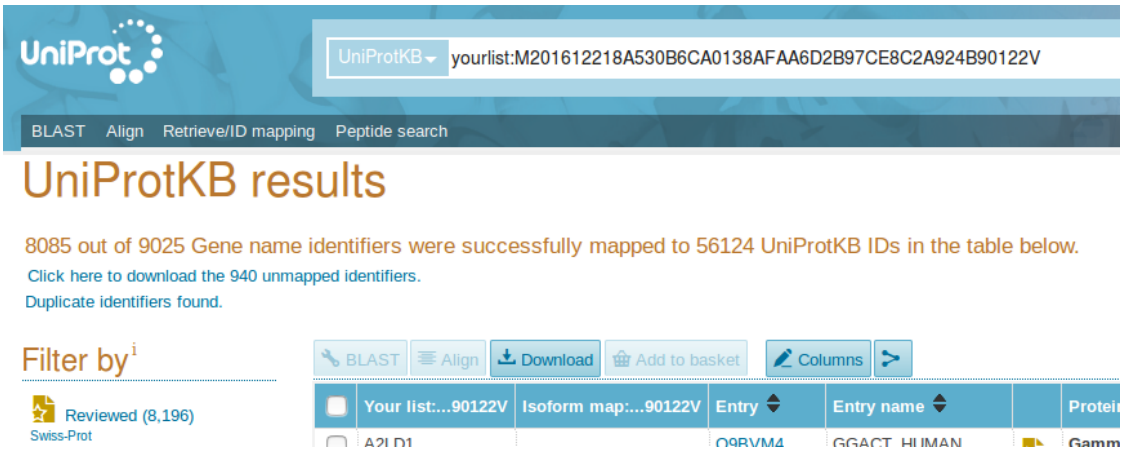


Figure 5: Uniprot mapping from gene symbol to UniProtKB AC/ID output2

3.3. CD-HIT

CD-Hit outputs two files for each Down-regulated and up-regulated genes after filtering out the duplicate entries in the data. Output files contain 8085 down-regulated and 8165 up-regulated sequences.

3.4. BLAST

After getting the CD-HIT output files we input those files in the BLASTp to carry out protein blast in order to validate if structures of the proteins are available or not. We get the blast output files containing 103381 hits for the up-regulated genes and 82621 hits for the down-regulated genes respectively.

3.5. DATA SORTING

We sorted the blast output data on the basis of Sequence Identity,keeping only the entries having Sequence identity ≥ 98%. Out of 103381 entries in the up-regulated file and 82621 entries in the down-regulated file, we are left with 4148 entries and— 3952 entries having sequence identity ≥ 98% respectively.

3.6. UniProt MAPPING FROM UniProtKB AC/ID TO PDB IDs

From the above BLAST output files containing entries with Sequence identity ≥ 98%, we mapped the query UniProtKB AC/ID to their respective PDB IDs using UniProt Retrieve/ID mapping tool.

- 1166 out of 1188 identifiers from UniProtKB Ac/ID were successfully mapped to 9372 PDB IDs for the protein products corresponding to up-regulated genes as shown in Figure 6.

Figure 6: Uniprot mapping from UniProtKB Ac/ID to PDB IDs output1

- 1134 out of 1145 identifiers from UniProtKB Ac/ID were successfully mapped to 8935 PDB IDs for the protein products corresponding to down-regulated genes as shown in Figure 7.

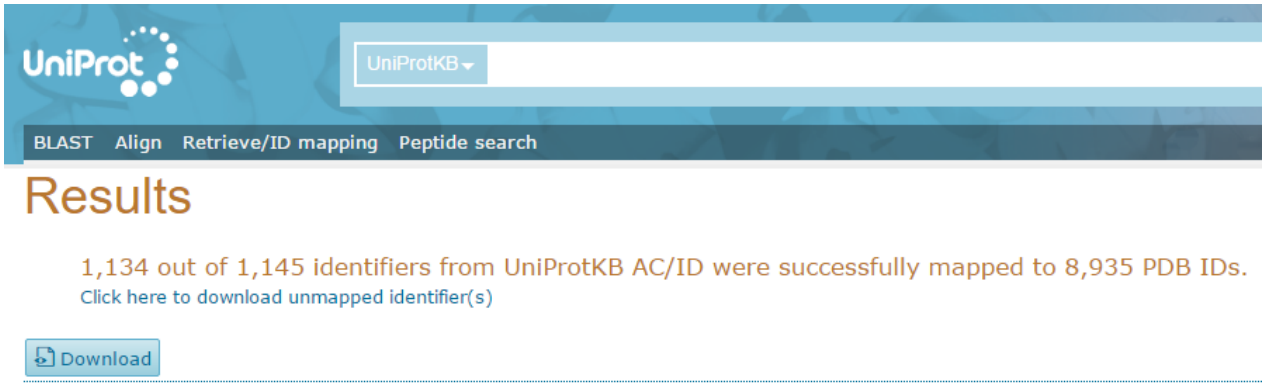


Figure 7: Uniprot mapping from UniProtKB Ac/ID to PDB IDs output2

3.7. Sorting data on the basis of Resolution

After mapping the UniProtKB AC/ID to their respective PDB IDs, we extracted the resolution of each of the PDB ID manually and then we:

a. Removed the duplicates from the data, keeping the entries having the highest resolution among duplicates.
b. And then sorted the data obtained after removing the duplicates on the basis of resolution, keeping only the PDB IDs having resolution between 0 and1.6.

After doing all the sorting we were left with 87 PDB entries corresponding to the up-regulated genes and 38 PDB entries corresponding to down-regulated genes as represented in Table 3 and Table 4 respectively.

SERIAL NUMBER	PDB entries corresponding to Up-regulated genes		
	UniProtKB	PDB ID	Resolution(Å)
1	P07108	2FJ9	1.6
2	P35219	2W2J	1.6
3	P48059	3F6Q	1.6
4	P51800	2PFI	1.6
5	P60508	1Y4M	1.6
6	Q13395	2HA8	1.6
7	Q14318	2AWG	1.6
8	Q16832	2WUH	1.6
9	Q3YBR2	2WZO	1.6
10	Q49AH0	2W50	1.6
11	Q5JTV8	4TVS	1.6
12	Q6ZW76	4NJ8	1.6
13	Q96FJ0	2ZNV	1.6
14	Q96ST3	1PO4	1.6
15	Q9NQW7	3CTZ	1.6
16	Q9NQX3	1JLJ	1.6
17	Q9ULT8	3DKM	1.6
18	O00584	3T0O	1.5
19	P36551	2AEX	1.5
20	Q5T0W9	5LZK	1.5
21	P48147	3DDU	1.5
22	Q9UMZ3	4IKC	1.5
23	Q9Y3Q0	3FEE	1.5
24	A1Z1Q3	4IQY	1.5
25	P26651	4J8S	1.5
26	P61020	2HEI	1.5
27	P63272	3H7H	1.5
28	Q5T2D3	4BOU	1.5
29	Q96EQ8	5DKA	1.5
30	Q9H1B7	2B9G	1.5
31	Q9H4G4	1SMB	1.5
32	Q9HC52	3I91	1.5
33	Q9UBP6	3CKK	1.5
34	Q9Y530	4J5S	1.5
35	O43763	3A03	1.5
36	Q9Y6X5	4LQY	1.5
37	Q8N157	4ESR	1.5
38	Q00872	2YXM	1.5
39	Q8WUD1	2A5J	1.5
40	O00151	2PKT	1.5
41	O95390	5E4G	1.5
42	P05165	2JKU	1.5
43	P07203	2F8A	1.5

44	P10768	3FCX	1.5
45	P27482	1GGZ	1.5
46	P29590	4WJN	1.5
47	P57105	2JIN	1.5
48	Q14554	4I6X	1.5
49	Q68DC2	4NL9	1.5
50	Q96HC4	2UZC	1.5
51	Q9BPZ3	3KUT	1.5
52	Q9HCJ0	3KTP	1.5
53	Q9NRW4	1WRM	1.5
54	P51157	3E5H	1.4
55	O60870	2CKK	1.4
56	Q8N1Q8	4AE7	1.4
57	Q9Y6W3	2QFE	1.4
58	Q96BD5	2PUY	1.4
59	O43897	3EDI	1.4
60	O75112	4YDP	1.4
61	O95218	3G9Y	1.4
62	O95721	4WY4	1.4
63	P56962	4WY4	1.4
64	Q06481	5JBT	1.4
65	Q8IYD1	3KUJ	1.4
66	Q9H0S4	3BER	1.4
67	Q9HB21	1EAZ	1.4
68	Q9HB90	3LLU	1.4
69	Q9NXS2	3PB7	1.4
70	Q9Y237	3UI5	1.4
71	Q16674	1I1J	1.3
72	Q6PHR2	4WZX	1.3
73	O43598	4P5E	1.3
74	Q9BTC0	4L7X	1.3
75	O94875	4IGZ	1.3
76	P30533	2FCW	1.2
77	O75747	2WWE	1.2
78	P51151	1WMS	1.2
79	Q96AZ1	4QPN	1.2
80	P51687	1MJ4	1.2
81	Q9BVM4	3JUB	1.2
82	O95881	1SEN	1.1
83	Q6P1N9	2XIO	1.1
84	Q86YW5	2FRG	1.1
85	Q9Y2N7	4WN5	1.1
86	Q9NRD5	2GZV	1.1
87	Q9BQ65	4H7W	1.1

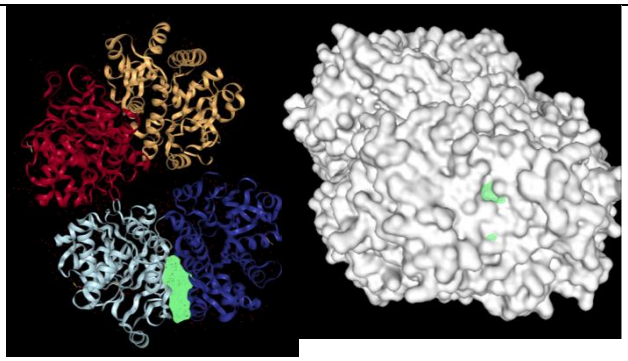
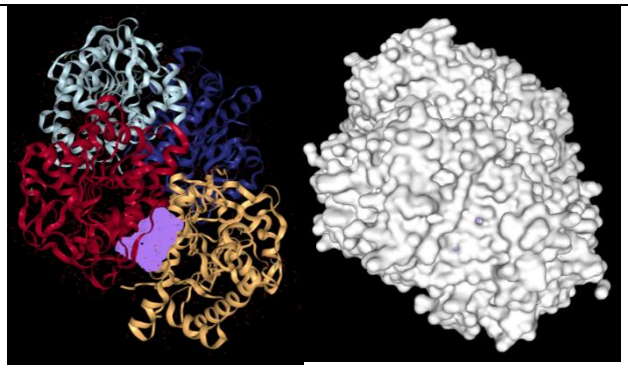
Table 3: UniProtKb AC/ID, PDB ID and Resolution of the PDB ids having resolution between 0 and 1.6 corresponding to the up-regulated genes.

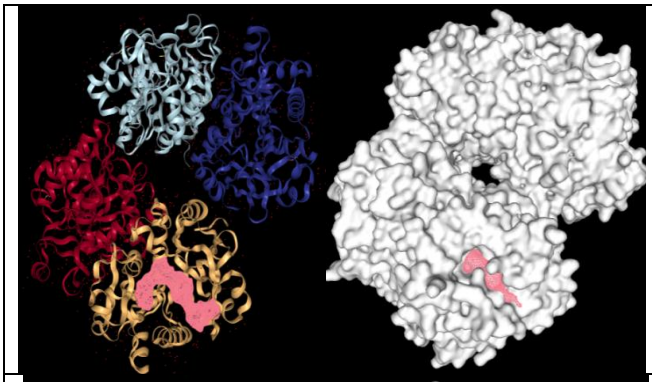
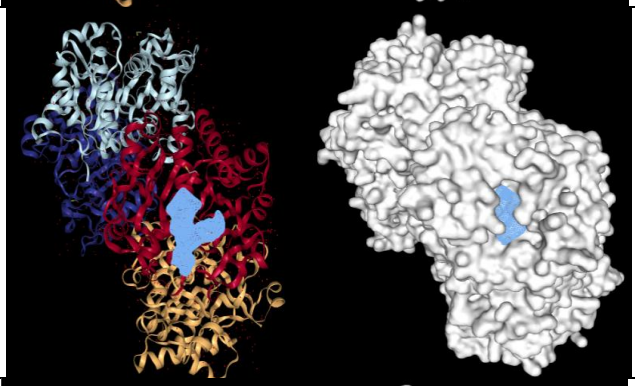
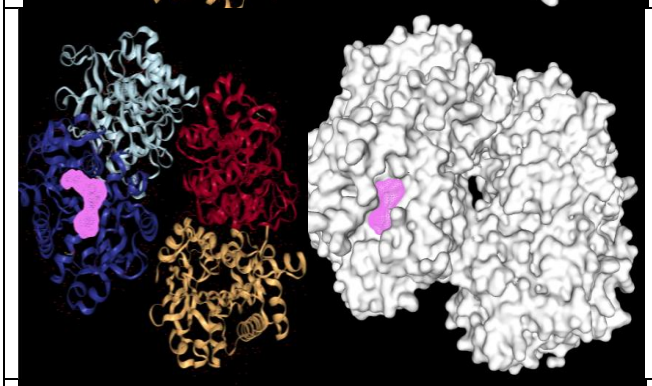
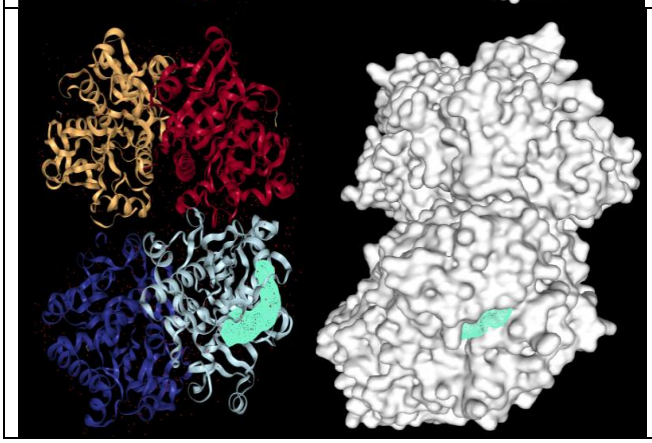
SERIAL NUMBER	PDB entries corresponding to Down-regulated genes		
	UniProtKB	PDB ID	Resolution(Å)
1	P30043	1HE5	1.5
2	P12111	1KNT	1.6
3	P22749	1L9L	0.9
4	Q92820	1L9X	1.6
5	P53674	1OKI	1.4
6	Q9Y692	1OQJ	1.5
7	P42126	1SG4	1.3
8	Q9Y547	1TVG	1.6
9	Q6UWP2	1XG5	1.5
10	Q9UIJ7	1ZD8	1.4
11	P36551	2AEX	1.5
12	Q9BS40	2BO9	1.6
13	O95684	2D68	1.6
14	Q8WWI1	2EAQ	1.4
15	Q9Y600	2JIS	1.6
16	Q9GZV9	2P39	1.5
17	Q99704	2V76	1.6
18	P13497	3EDG	1.2
19	Q14847	3I35	1.4
20	O60307	3KHF	1.2
21	Q7L1T6	3LF5	1.2
22	P07910	3LN4	1.2
23	O43169	3NER	1.4
24	Q15744	3T92	1.5
25	Q9ULY5	3WH3	1.3
26	A6NK44	3ZW5	1.6
27	P22001	4BGC	1.2
28	Q96LW7	4DWN	1.5
29	P35754	4RQR	1.0
30	Q96P48	4X1V	1.5
31	O75976	5AQ0	0.9
32	P51530	5EAY	1.5
33	Q9H6S1	5EP6	1.4
34	O75791	5GJH	1.2
35	Q96FZ2	5KO9	1.5
36	P61923	5MC7	1.6
37	Q8WWN9	5MR1	1.2
38	O75608	5SYM	1.5

Table 4: UniProtKb AC/ID, PDB ID and Resolution of the PDB IDs having resolution between 0 and 1.6 corresponding to the down-regulated genes.

3.8. DoGSiteScorer

We downloaded structure of each of the above mentioned PDB IDs and manually put them in the DoGSiteScorer to obtain the binding site information which includes D-score and position of each binding pocket and the residues involved in it. We only kept the entry for which the D-score ≥ 0.7. All the binding pockets having 0.7≥ D-score>0.8 are normal binding pockets that are not likely druggable and the binding sites with D-score ≥ 0.8 are considered as druggable binding sites. We then characterized the binding sites into enzymatic binding sites (ENZ), protein-protein interaction binding sites (PPI) and others (OTH) on the basis of amino acids involved, surface characteristics of the pockets and the information available in various other databases like CSA and BRENDA as shown in the tables below.

PDB id: 1l9x	Resolution: 1.6	CSA: Available
	Dscore: 0.83 Binding site type: Druggable Amino acid residues: A(2),R(0),N(0),D(0),C(0), E(4),Q(2),G(0),H(0),I(4), L(0),K(2),M(0),F(2),P(2),S(4),T(0),W(0),Y(8), V(2) Binding site category: ENZ	
	Dscore: 0.82 Binding site type: Druggable Amino acid residues: A(3),R(1),N(0),D(0),C(0),E(5),Q(3),G(1),H(0), I(4), L(0),K(2),M(0),F(3), P(3), S(5),T(0),W(1),Y(8),V(3) Binding site category: ENZ	

	Dscore: 0.82 Binding site type: Druggable Amino acid residues: A(1),R(1),N(0),D(0),B(0),C(2), E(2),Q(3),Z(0),G(2),H(2),I(2), L(8),K(2),M(1),F(2),P(1),S(3), T(0),W(1),Y(3),V(0) Binding site category: OTH(ENZ+PPI)
	Dscore: 0.79 Binding site type: Normal Amino acid residues: A(1),R(0),N(0),D(0),B(0),C(1), E(1),Q(1),Z(0),G(2),H(2),I(2), L(2),K(2),M(1),F(2),P(1),S(2), T(0),W(1),Y(2),V(0) Binding site category: OTH(ENZ/PPI)
	Dscore: 0.79 Binding site type: Normal Amino acid residues: A(0),R(0),N(0),D(1),B(0),C(2), E(0),Q(3),Z(0),G(2),H(2),I(0), L(7),K(1),M(0),F(1),P(0),S(3), T(0),W(1),Y(0),V(1) Binding site category:OTH(ENZ/PPI)
	Dscore: 0.79 Binding site type: Normal Amino acid residues: A(1),R(0),N(0),D(0),B(0),C(1), E(2),Q(2),Z(0),G(2),H(2),I(3), L(4),K(3),M(1),F(2),P(1),S(2), T(0),W(2),Y(2),V(2) Binding site category: ENZ

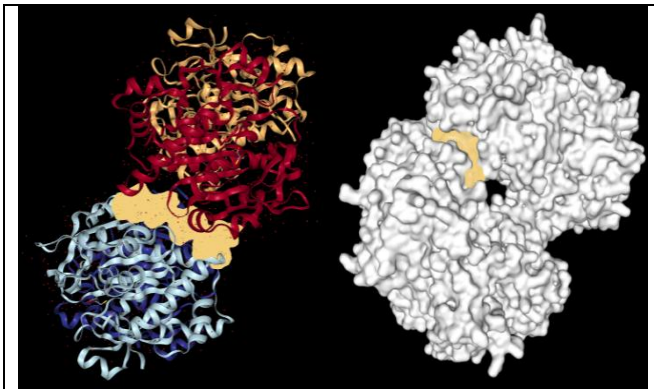
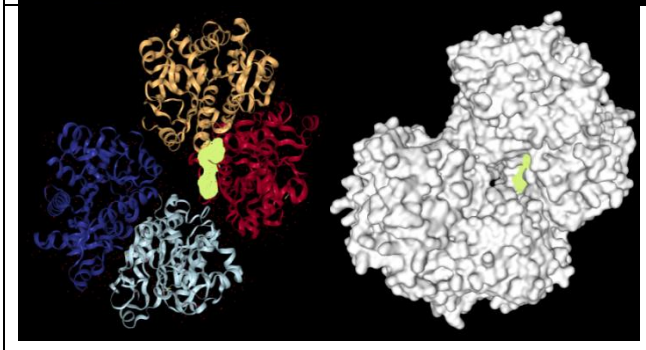
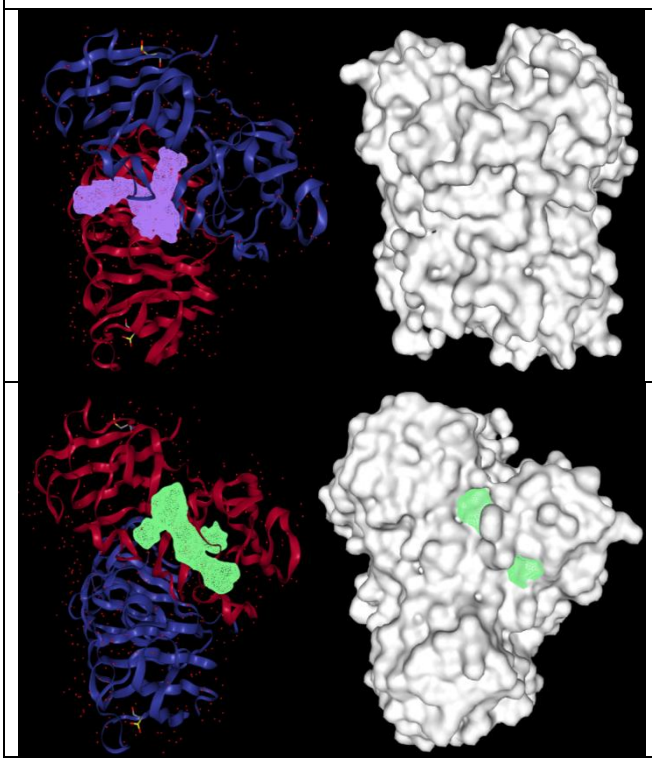
	Dscore: 0.78 Binding site type: Normal Amino acid residues: A(0),R(2),N(4),D(4),B(0),C(0), E(0),Q(3),Z(0),G(7),H(2),I(0), L(2),K(8),M(0),F(5),P(2),S(2), T(2),W(0),Y(7),V(2) Binding site category: OTH(ENZ/PPI)
	Dscore: 0.74 Binding site type: Normal Amino acid residues: A(4),R(2),N(2),D(0),B(0),C(0), E(1),Q(2),Z(0),G(1),H(1),I(0), L(1),K(4),M(1),F(4),P(2),S(0), T(0),W(1),Y(1),V(2) Binding site category: ENZ

Table 5: Structural information for the protein Gamma-glutamyl hydrolase.

PDB id:1oki	Resolution: 1.4 Å	CSA:N/A
	Dscore: 0.84 Binding site type: Druggable Amino acid residues: A(0),R(8),N(2),D(0),B(0),C(0), E(4),Q(2),Z(0),G(4),H(0),I(2), L(1),K(0),M(0),F(0),P(0),S(6), T(2),W(1),Y(10),V(2) Binding site category: ENZ	
	Dscore: 0.82 Binding site type: Druggable Amino acid residues: A(1),R(1),N(0),D(1),B(0),C(0), E(4),Q(2),Z(0),G(3),H(0),I(0), L(0),K(0),M(2),F(2),P(1),S(3), T(0),W(3),Y(2),V(1) Binding site category: OTH	

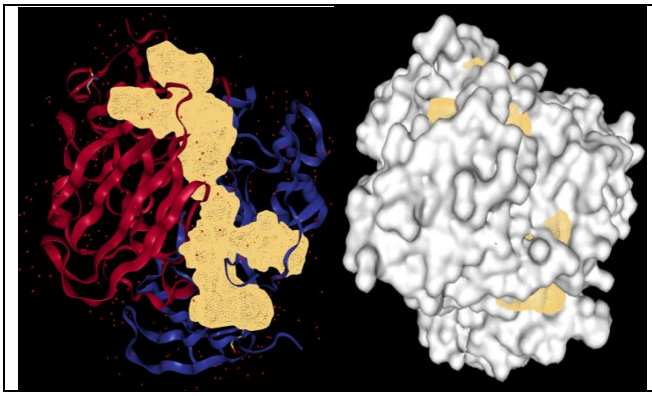
	Dscore: 0.8 Binding site type: Druggable Amino acid residues: A(6),R(6),N(2),D(8),B(0),C(0), E(4),Q(5),Z(0),G(4),H(3),I(2), L(2),K(6),M(2),F(0),P(8),S(4), T(2),W(3),Y(6),V(4) Binding site category: ENZ

Table 6:Structural information for the protein: Beta-crystallin B1

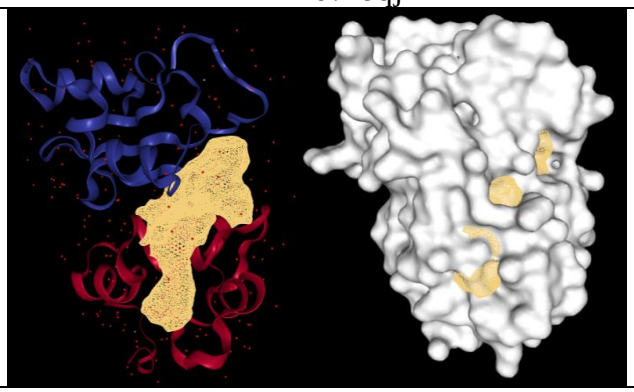
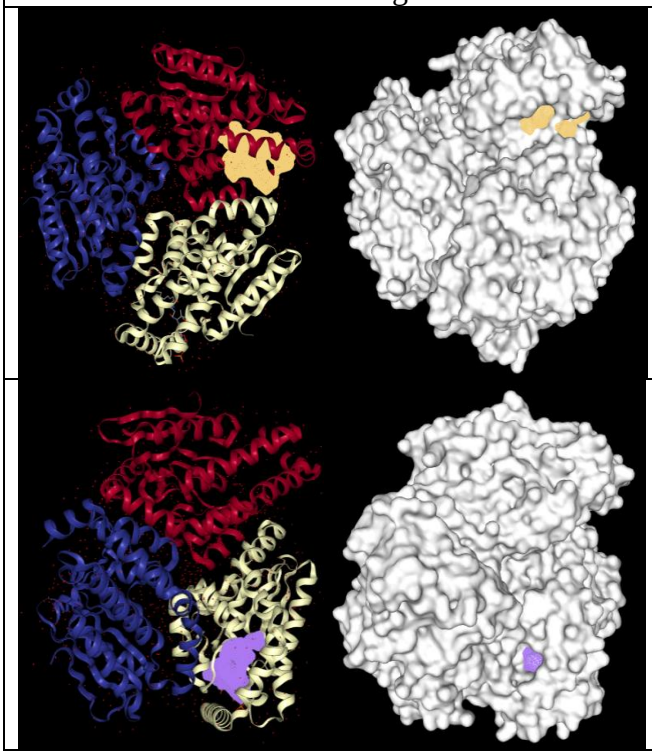
PDB id:1oqj	Resolution: 1.55 Å	CSA:N/A
	Dscore: 0.8 Binding site type: Druggable Amino acid residues: A(2),R(2),N(1),D(3),B(0),C(1), E(2),Q(1),Z(0),G(4),H(1),I(4), L(2),K(3),M(1),F(2),P(1),S(2), T(1),W(1),Y(2),V(0) Binding site category: ENZ	

Table7:Structural information for the proteinGlucocorticoid modulatory element-binding protein 1

PDB id: 1sg4	Resolution: 1.3 Å	CSA: Available
	Dscore: 0.83 Binding site type: Druggable Amino acid residues: A(3),R(0),N(2),D(2),B(0),C(0), E(0),Q(1),Z(0),G(2),H(0),I(3), L(12),K(0),M(1),F(3),P(2),S(1), T(1),W(3),Y(1),V(3) Binding site category: OTH	
	Dscore: 0.83 Binding site type: Druggable Amino acid residues: A(3),R(0),N(2),D(2),B(0),C(0), E(0),Q(1),Z(0),G(2),H(0),I(3), L(9),K(0),M(2),F(3),P(2),S(1), T(1),W(3),Y(1),V(3) Binding site category: OTH	

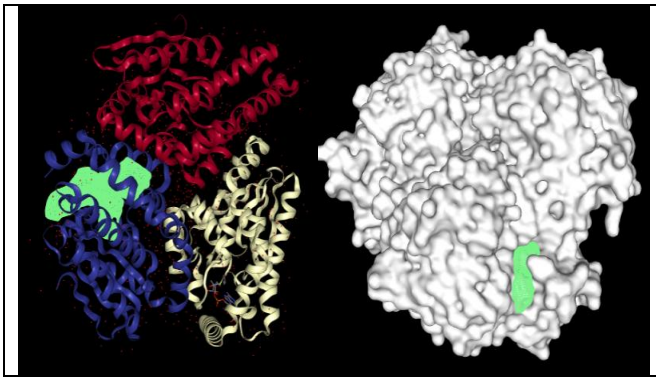
	Dscore: 0.83 Binding site type: Druggable Amino acid residues: A(4),R(0),N(2),D(1),B(0),C(0), E(0),Q(1),Z(0),G(3),H(1),I(4), L(9),K(0),M(3),F(3),P(2),S(1), T(0),W(2),Y(0),V(1) Binding site category: PPI

Table 8: Structural information for the protein Enoyl-CoA delta isomerase 1, mitochondrial

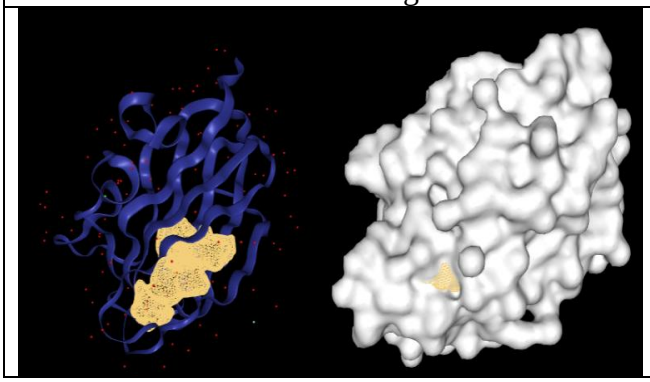
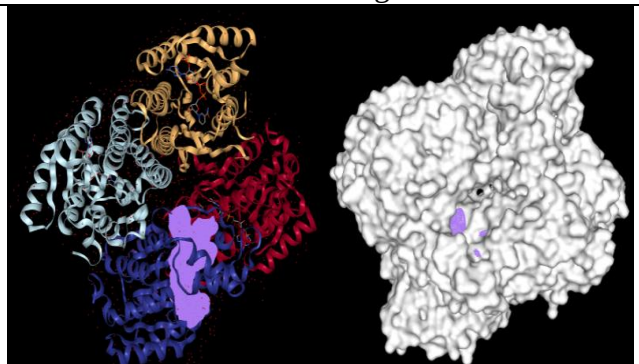
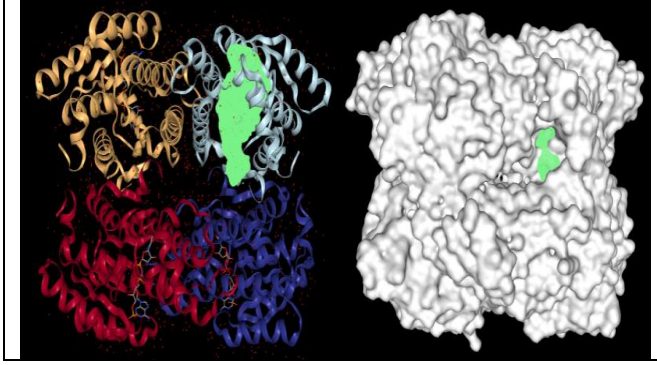
PDB id:1tvg	Resolution: 1.6 Å	CSA:N\A
	Dscore: 0.79 Binding site type: Normal Amino acid residues: A(0),R(0),N(1),D(0),B(0),C(0), E(4),Q(2),Z(0),G(0),H(1),I(6), L(3),K(1),M(0),F(4),P(0),S(2), T(1),W(1),Y(1),V(4) Binding site category: OTH	

Table 9: Structural information for the protein Intraflagellar transport protein 25 homolog

Pdb id: 1xg5	Resolution: 1.53 Å	CSA: Available
	Dscore: 0.81 Binding site type: Druggable Amino acid residues: A(6),R(4),N(5),D(1),B(0),C(3), E(1),Q(0),Z(0),G(6),H(1),I(3), L(6),K(2),M(2),F(1),P(2),S(5), T(4),W(0),Y(2),V(6) Binding site category: ENZ	
	Dscore: 0.81 Binding site type: Druggable Amino acid residues: A(5),R(3),N(5),D(1),B(0),C(3), E(1),Q(0),Z(0),G(6),H(2),I(3), L(6),K(2),M(1),F(1),P(2),S(6), T(5),W(0),Y(2),V(5) Binding site category: ENZ	

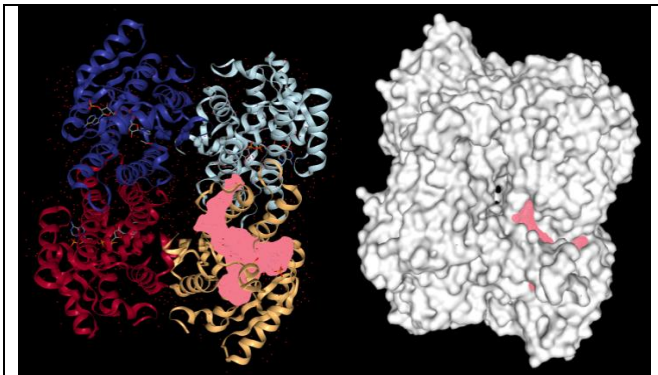
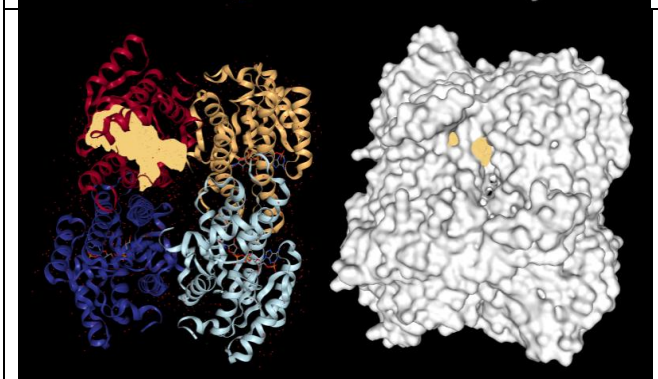
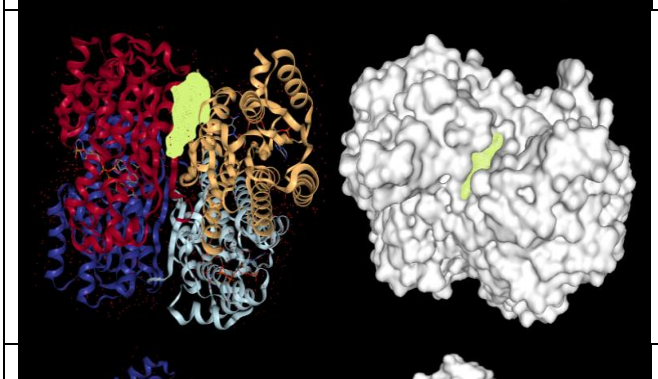
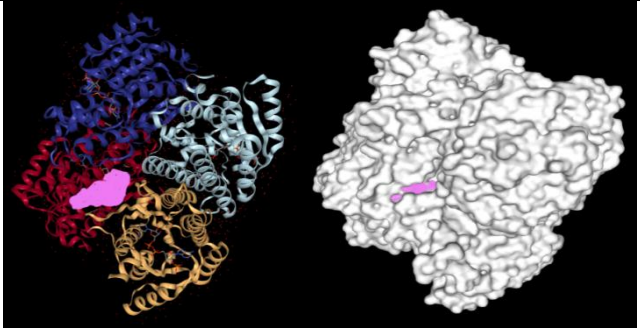
	Dscore: 0.81 Binding site type: Druggable Amino acid residues: A(5),R(3),N(5),D(2),B(0),C(3), E(1),Q(2),Z(0),G(7),H(1),I(2), L(7),K(2),M(1),F(1),P(2),S(4), T(3),W(0),Y(1),V(6) Binding site category: ENZ
	Dscore: 0.8 Binding site type: Druggable Amino acid residues: A(5),R(4),N(4),D(1),B(0),C(3), E(1),Q(0),Z(0),G(6),H(2),I(2), L(6),K(2),M(1),F(1),P(1),S(6), T(4),W(0),Y(2),V(5) Binding site category: ENZ
	Dscore: 0.8 Binding site type: Druggable Amino acid residues: A(2),R(2),N(0),D(2),B(0),C(0), E(0),Q(2),Z(0),G(1),H(0),I(7), L(0),K(0),M(3),F(0),P(4),S(2), T(2),W(0),Y(2),V(0) Binding site category: OTH
	Dscore: 0.79 Binding site type: Normal Amino acid residues: A(1),R(4),N(0),D(1),B(0),C(1), E(3),Q(1),Z(0),G(1),H(2),I(2), L(2),K(0),M(2),F(0),P(1),S(1), T(4),W(0),Y(0),V(0) Binding site category: ENZ

Table 10: Structural information for the proteinDehydrogenase/reductase SDR family member 11

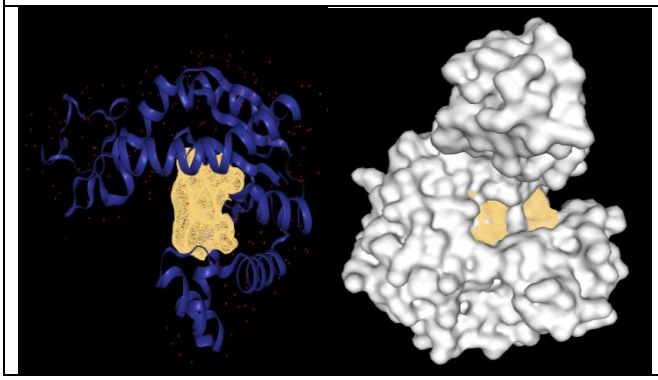
PDB id: 1zd8	Resolution: 1.48 Å	CSA: N/A
	Dscore: 0.82 Binding site type: Druggable Amino acid residues: A(0),R(0),N(1),D(0),B(0),C(0), E(4),Q(2),Z(0),G(0),H(1),I(6), L(3),K(1),M(0),F(4),P(0),S(2), T(1),W(1),Y(1),V(4) Binding site category: PPI	

Table 11: Structural information for the protein GTP: AMP phosphotransferase AK3, mitochondrial

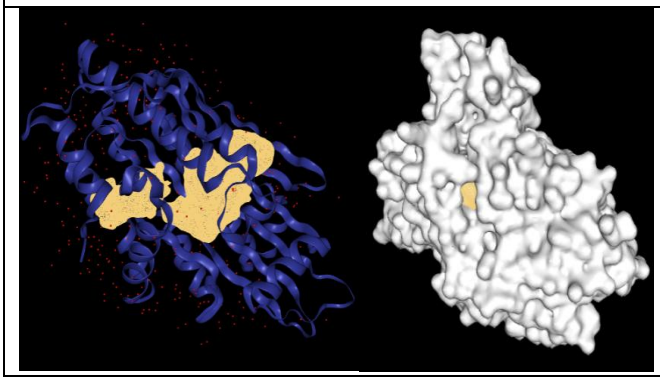
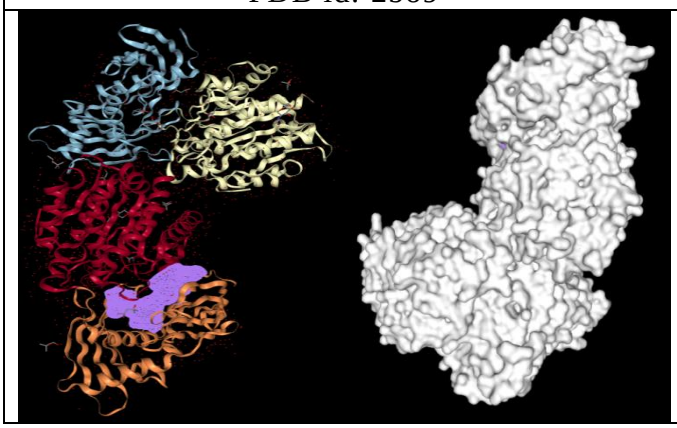
PDB id: 2aex	Resolution: 1.58 Å	CSA: N/A
	Dscore: 0.78 Binding site type: Normal Amino acid residues: A(0),R(6),N(2),D(4),B(0),C(2), E(0),Q(1),Z(0),G(8),H(1),I(5), L(3),K(1),M(1),F(5),P(2),S(8), T(3),W(2),Y(3),V(3) Binding site category: ENZ	

Table 12: Structural information for the protein Oxygen-dependent coproporphyrinogen-III oxidase, mitochondrial

PDB id: 2bo9	Resolution: 1.6 Å	CSA: Available
	Dscore: 0.83 Binding site type: Druggable Amino acid residues: A(1),R(0),N(2),D(0),B(0),C(1), E(3),Q(1),Z(0),G(3),H(1),I(3), L(4),K(3),M(1),F(2),P(2),S(2), T(4),W(2),Y(3),V(6) Binding site category: OTH	

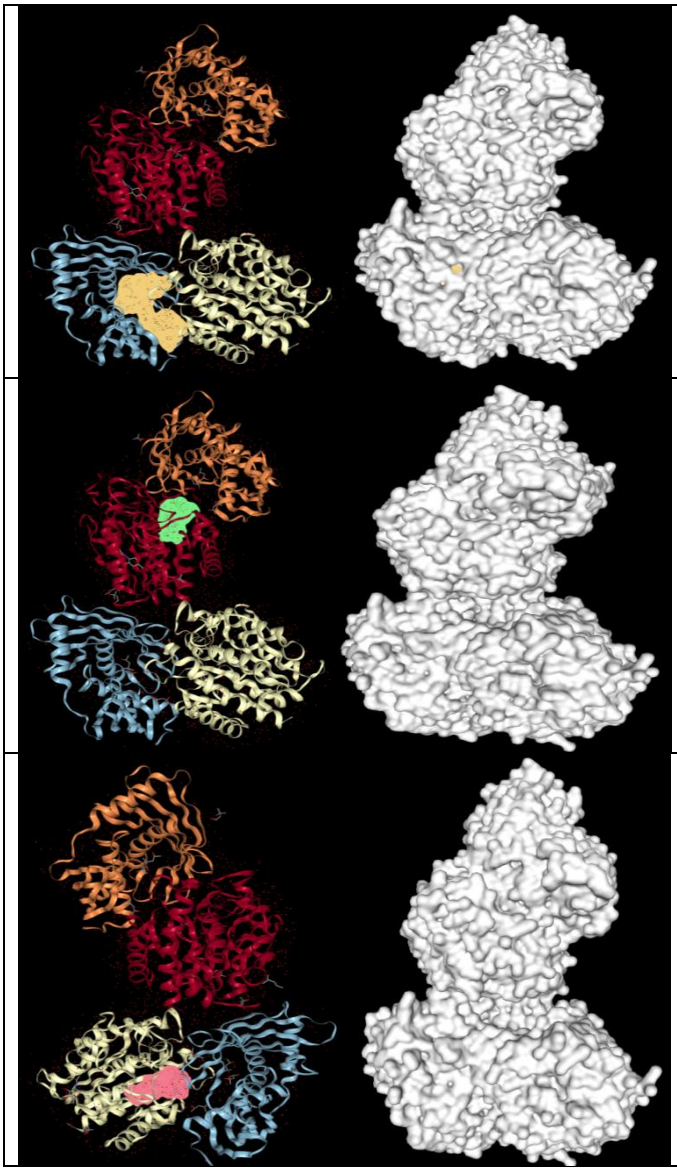
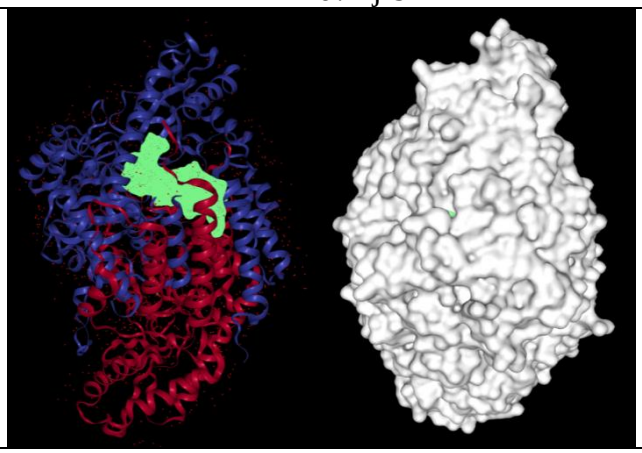
	Dscore: 0.8 Binding site type: Druggable Amino acid residues: A(1),R(0),N(3),D(0),B(0),C(1), E(4),Q(2),Z(0),G(3),H(2),I(5), L(4),K(4),M(1),F(2),P(2),S(2), T(5),W(2),Y(5),V(6) Binding site category: OTH
	Dscore: 0.72 Binding site type: Normal Amino acid residues: A(1),R(2),N(1),D(2),B(0),C(0), E(1),Q(2),Z(0),G(1),H(3),I(2), L(1),K(0),M(1),F(1),P(0),S(4), T(2),W(1),Y(2),V(1) Binding site category: ENZ
	Dscore: 0.71 Binding site type: Normal Amino acid residues: A(1),R(2),N(1),D(1),B(0),C(0), E(1),Q(2),Z(0),G(1),H(3),I(2), L(1),K(0),M(1),F(1),P(0),S(4), T(2),W(1),Y(2),V(1) Binding site category: ENZ

Table13: Structural information for the protein Latexin

PDB id: 2jis	Resolution: 1.6 Å	CSA: Available
	Dscore: 0.83 Binding site type: Druggable Amino acid residues: A(7),R(3),N(2),D(1),B(0),C(0), E(2),Q(2),Z(0),G(5),H(2),I(3), L(11),K(2),M(1),F(2),P(3),S(3), T(2),W(0),Y(0),V(4) Binding site category: ENZ	

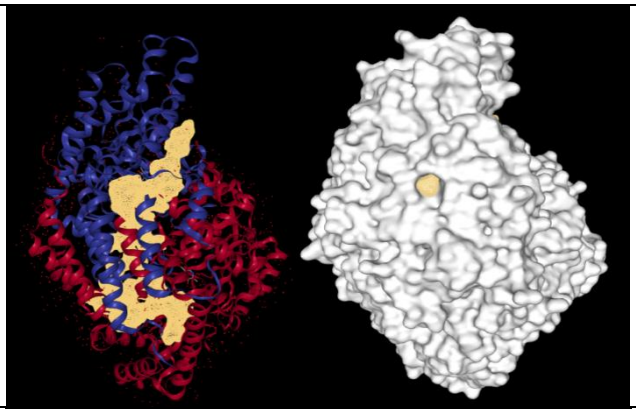
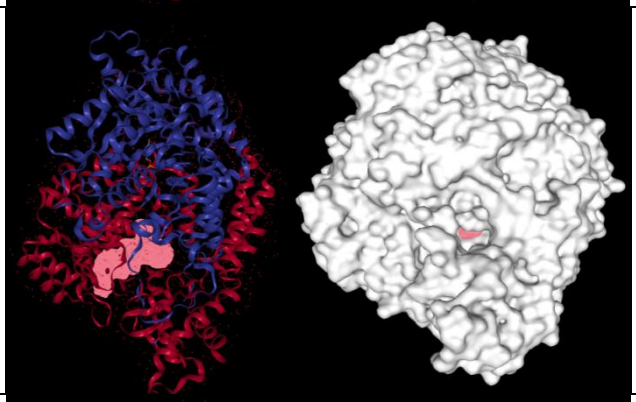
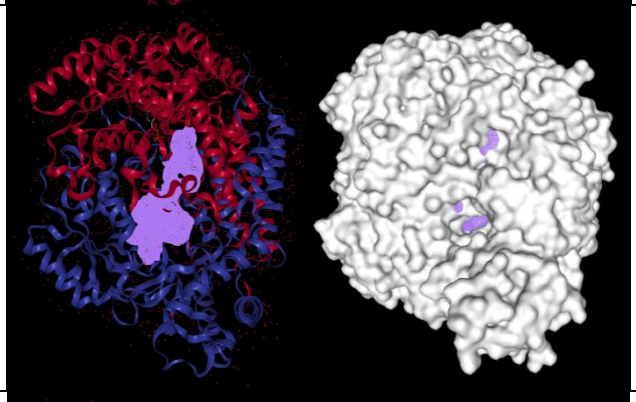
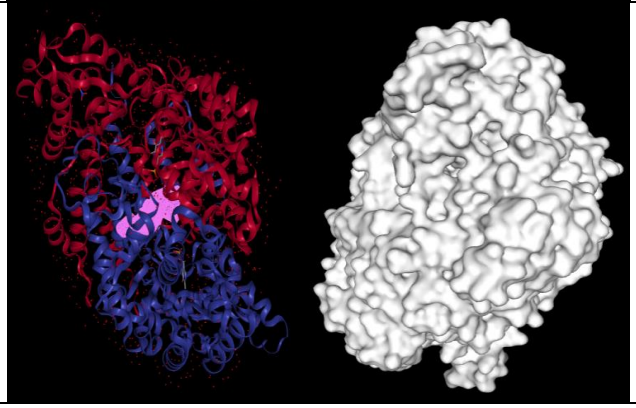
	Dscore: 0.81 Binding site type: Druggable Amino acid residues: A(5),R(0),N(2),D(2),B(0),C(2), E(3),Q(1),Z(0),G(5),H(2),I(3), L(2),K(2),M(1),F(2),P(1),S(7), T(4),W(1),Y(1),V(0) Binding site category: ENZ
	Dscore: 0.81 Binding site type: Druggable Amino acid residues: A(10),R(8),N(3),D(3),B(0),C(4), E(4),Q(8),Z(0),G(9),H(3),I(5), L(16),K(12),M(2),F(5),P(6),S(6), T(7),W(2),Y(7),V(11) Binding site category: ENZ
	Dscore: 0.8 Binding site type: Druggable Amino acid residues: A(2),R(2),N(2),D(4),B(0),C(2), E(4),Q(2),Z(0),G(8),H(2),I(4), L(4),K(3),M(0),F(2),P(1),S(5), T(5),W(1),Y(2),V(3) Binding site category: ENZ
	Dscore: 0.79 Binding site type: Normal Amino acid residues: A(0),R(4),N(0),D(2),B(0),C(3), E(6),Q(2),Z(0),G(2),H(0),I(0), L(0),K(0),M(0),F(2),P(2),S(0), T(0),W(0),Y(0),V(1) Binding site category: ENZ

Table 14: Structural information for the protein Cysteine sulfinic acid decarboxylase

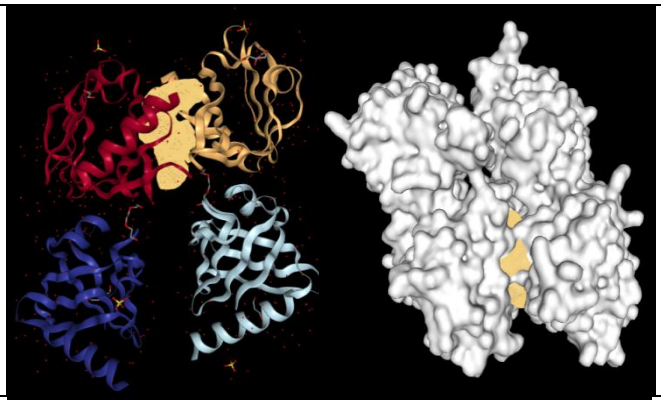
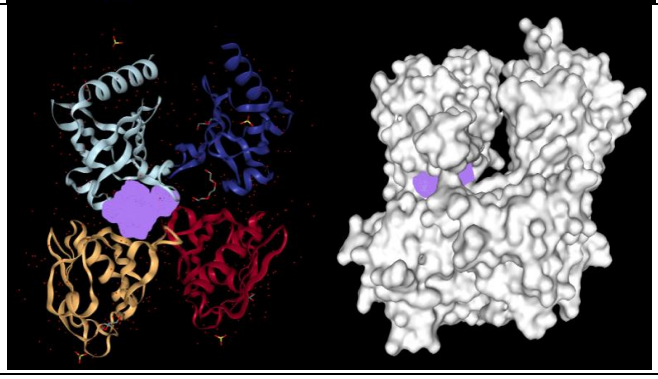
PDB id: 2v76	Resolution: 1.6 Å	CSA: N/A
	Dscore: 0.81 Binding site type: Druggable Amino acid residues: A(2),R(4),N(0),D(0),B(0),C(0), E(1),Q(4),Z(0),G(1),H(0),I(1), L(7),K(0),M(0),F(0),P(5),S(3), T(2),W(0),Y(0),V(1) Binding site category: OTH	
	Dscore: 0.72 Binding site type: Normal Amino acid residues: A(1),R(1),N(0),D(0),B(0),C(1), E(2),Q(1),Z(0),G(2),H(1),I(0), L(5),K(0),M(0),F(1),P(1),S(1), T(1),W(1),Y(1),V(3) Binding site category: OTH	

Table 15: Structural information for the proteinDocking protein 1

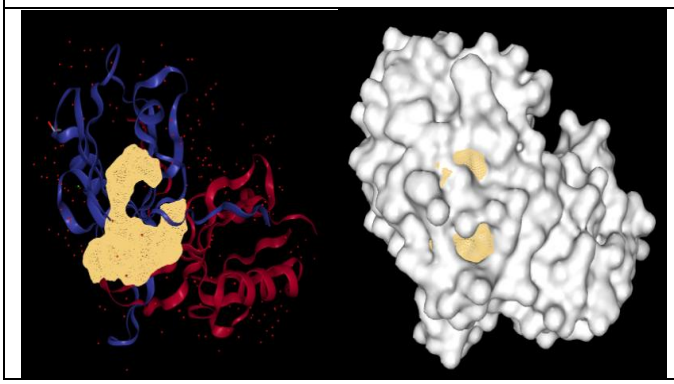
PDB id: 3khf	Resolution: 1.2 Å	CSA: N/A
	Dscore: 0.82 Binding site type: Druggable Amino acid residues: A(2),R(4),N(1),D(3),B(0),C(0), E(0),Q(1),Z(0),G(3),H(2),I(2), L(5),K(0),M(1),F(0),P(2),S(2), T(3),W(0),Y(1),V(2) Binding site category: ENZ	

Table 16:Structural information for the protein Microtubule-associated serine/threonine-protein kinase 3

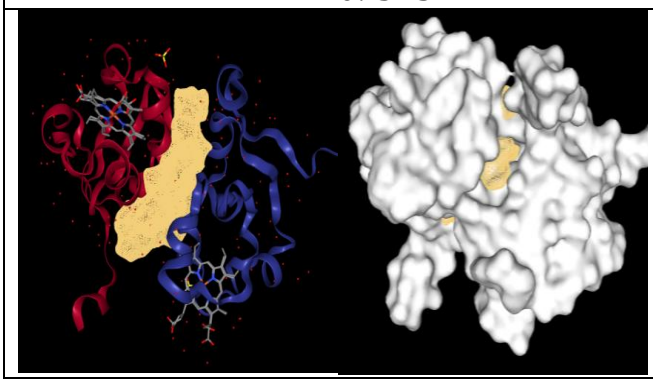
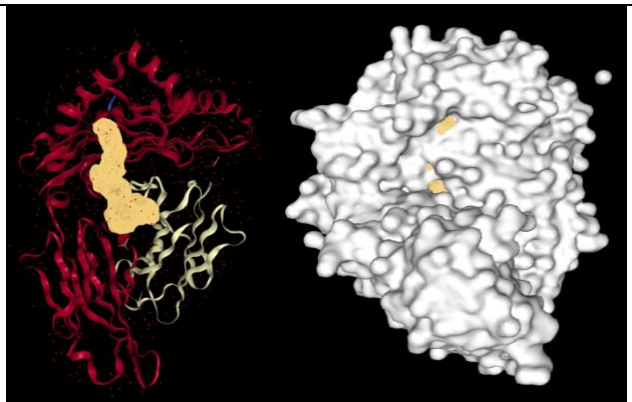
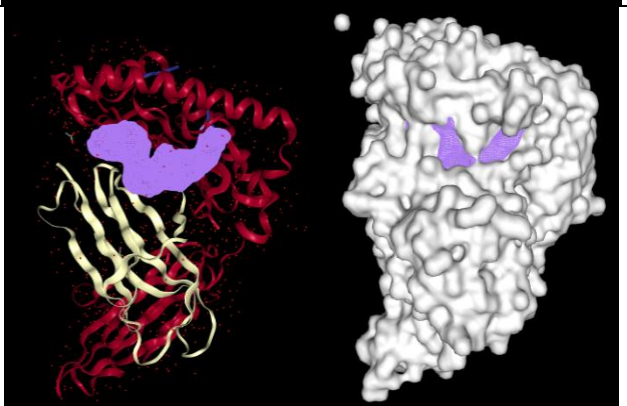
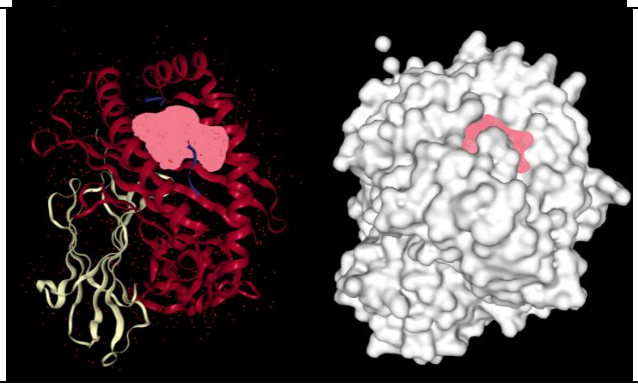
PDB id: 3lf5	Resolution: 1.25 Å	CSA: N/A
	Dscore: 0.77 Binding site type: Normal Amino acid residues: A(5),R(2),N(2),D(4),B(0),C(1), E(1),Q(5),Z(0),G(4),H(0),I(1), L(4),K(7),M(4),F(0),P(0),S(2), T(1),W(0),Y(0),V(1) Binding site category: ENZ	

Table 17: Structural information for the protein Cytochrome b5 reductase 4

PDB id: 3ln4	Resolution: 1.296 Å	CSA: N/A
	Dscore: 0.84 Binding site type: Druggable Amino acid residues: A(3),R(4),N(2),D(5),B(0),C(0), E(1),Q(2),Z(0),G(3),H(1),I(0), L(6),K(2),M(2),F(4),P(1),S(6), T(2),W(0),Y(4),V(1) Binding site category: ENZ	
	Dscore: 0.81 Binding site type: Druggable Amino acid residues: A(1),R(5),N(0),D(4),B(0),C(0), E(0),Q(1),Z(0),G(0),H(3),I(2), L(2),K(1),M(1),F(4),P(2),S(6), T(3),W(0),Y(1),V(1) Binding site category: ENZ	
	Dscore: 0.8 Binding site type: Druggable Amino acid residues: A(1),R(2),N(2),D(1),B(0),C(0), E(2),Q(1),Z(0),G(1),H(2),I(1), L(3),K(0),M(1),F(0),P(2),S(2), T(3),W(2),Y(2),V(4) Binding site category: OTH(ENZ/PPI)	

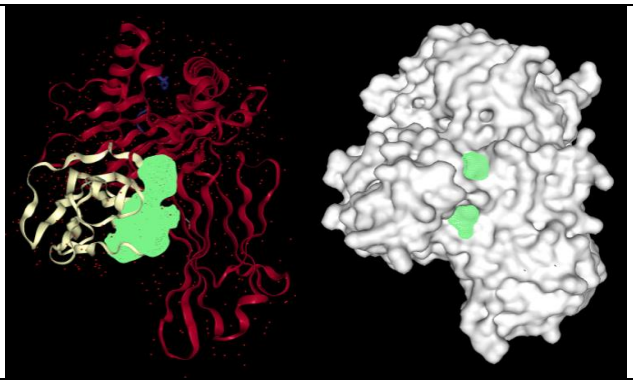
	Dscore: 0.78 Binding site type: Normal Amino acid residues: A(3),R(3),N(0),D(3),B(0),C(0), E(2),Q(1),Z(0),G(1),H(2),I(1), L(1),K(4),M(1),F(1),P(1),S(2), T(2),W(3),Y(2),V(4) Binding site category: OTH(ENZ/PPI)

Table 18: Structural information for the protein Heterogeneous nuclear ribonucleoproteins C1/C2

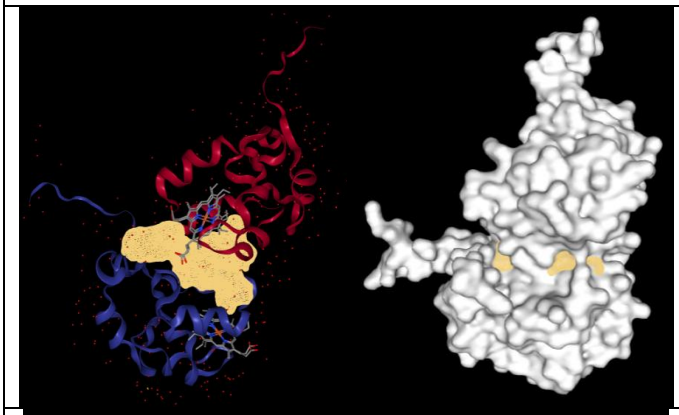
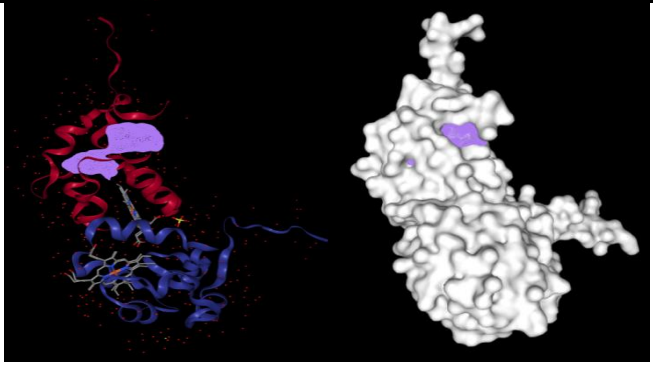
PDB id: 3ner	Resolution: 1.45 Å	CSA: N/A
	Dscore: 0.77 Binding site type: Normal Amino acid residues: A(1),R(2),N(0),D(4),B(0),C(0), E(1),Q(6),Z(0),G(5),H(3),I(2), L(2),K(1),M(0),F(1),P(1),S(2), T(0),W(0),Y(3),V(3) Binding site category: ENZ	
	Dscore: 0.73 Binding site type: Normal Amino acid residues: A(3),R(1),N(1),D(2),B(0),C(0), E(1),Q(0),Z(0),G(2),H(1),I(1), L(4),K(1),M(0),F(0),P(0),S(1), T(0),W(1),Y(0),V(3) Binding site category: PPI	

Table 19: Structural information for the protein Cytochrome b5 type B

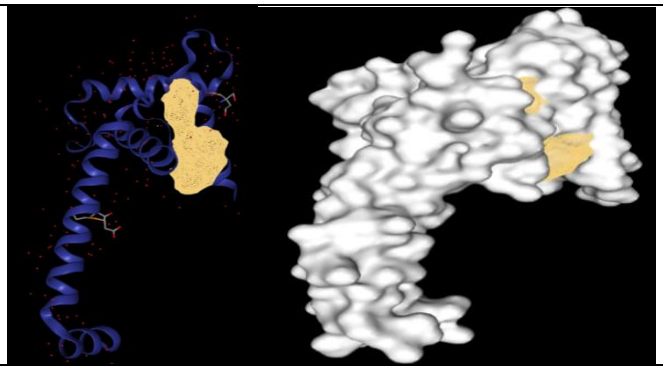
PDB id: 3t92	Resolution: 1.5 Å	CSA: N/A
	Dscore: 0.74 Binding site type: Normal Amino acid residues: A(1),R(2),N(0),D(0),B(0),C(2), E(4),Q(1),Z(0),G(0),H(2),I(2), L(0),K(1),M(0),F(0),P(1),S(0), T(0),W(0),Y(1),V(2) Binding site category: OTH	

Table 20: Structural information for the protein CCAAT/enhancer-binding protein epsilon-histone acetyltransferase

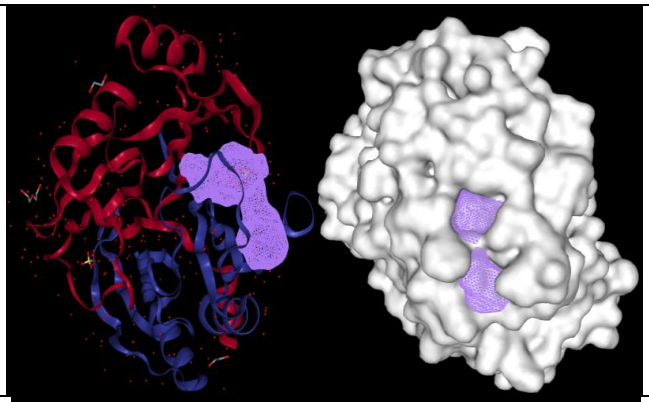
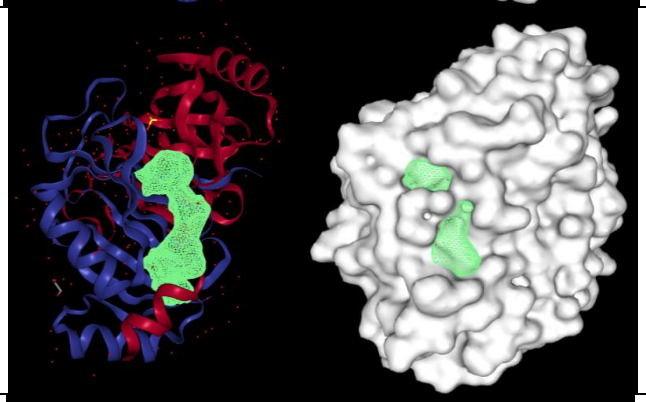
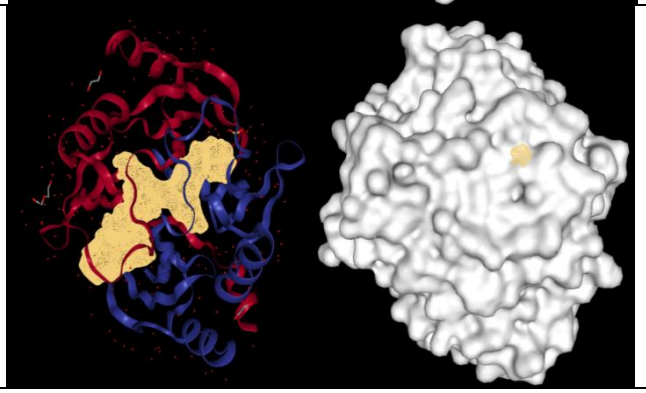
PDB id: 3zw5	Resolution: 1.6 Å	CSA: N/A
	Dscore: 0.84 Binding site type: Druggable Amino acid residues: A(2),R(1),N(1),D(1),B(0),C(1), E(2),Q(2),Z(0),G(3),H(0),I(2), L(2),K(2),M(4),F(2),P(0),S(1), T(2),W(0),Y(0),V(1) Binding site category: PPI	
	Dscore: 0.83 Binding site type: Druggable Amino acid residues: A(2),R(2),N(3),D(4),B(0),C(1), E(0),Q(5),Z(0),G(3),H(5),I(4), L(3),K(5),M(2),F(2),P(5),S(2), T(3),W(0),Y(1),V(7) Binding site category: OTH(ENZ/PPI)	
	Dscore: 0.8 Binding site type: Druggable Amino acid residues: A(0),R(2),N(1),D(2),B(0),C(1), E(0),Q(4),Z(0),G(0),H(2),I(5), L(2),K(4),M(0),F(2),P(1),S(0), T(1),W(0),Y(1),V(1) Binding site category: ENZ	

Table 21:Structural information for the protein Glyoxalase domain-containing protein 5

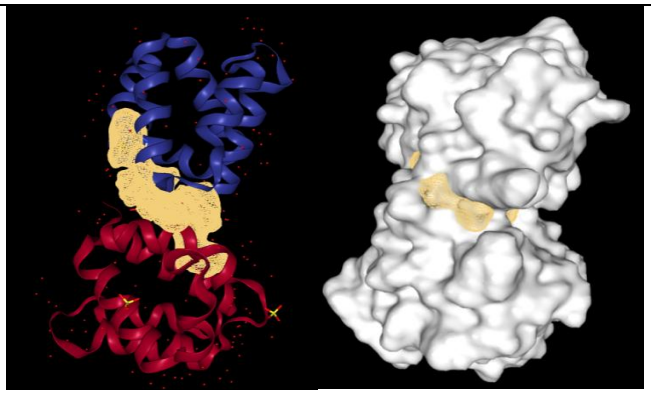
PDB id: 4dwn	Resolution: 1.581 Å	CSA: N/A
	Dscore: 0.83 Binding site type: Druggable Amino acid residues: A(1),R(4),N(0),D(4),B(0),C(1), E(4),Q(1),Z(0),G(2),H(1),I(0), L(2),K(0),M(0),F(0),P(1),S(1), T(0),W(0),Y(0),V(0) Binding site category: OTH(ENZ/PPI)	

Table 22: Structural information for the protein Caspase recruitment domain-containing protein 19

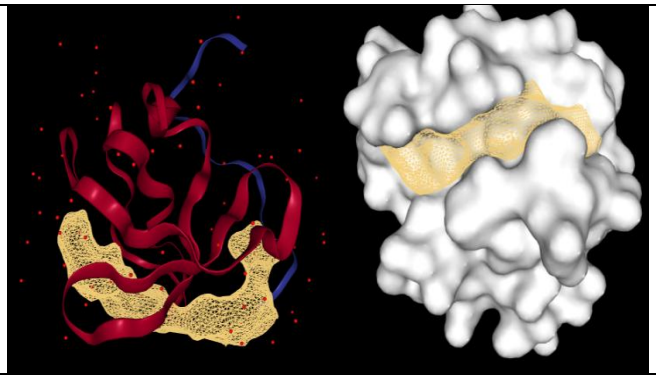
PDB id: 4x1v	Resolution: 1.58 Å	CSA: N/A
	Dscore: 0.79 Binding site type: Normal Amino acid residues: A(0),R(2),N(1),D(2),B(0),C(0), E(0),Q(4),Z(0),G(1),H(1),I(1), L(5),K(2),M(0),F(1),P(1),S(1), T(0),W(1),Y(0),V(1) Binding site category: PPI	

Table 23: Structural information for the protein Arf-GAP with Rho-GAP domain, ANK repeat and PH domain-containing protein 1

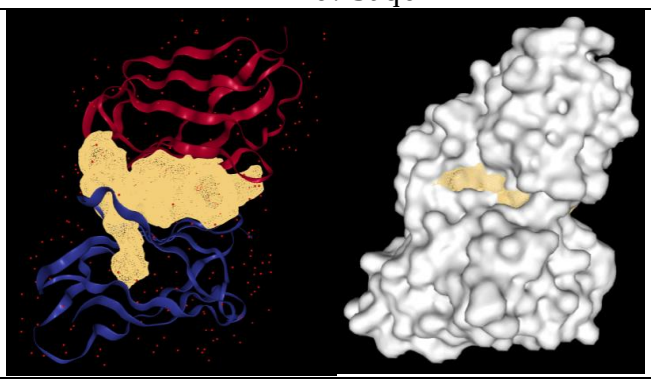
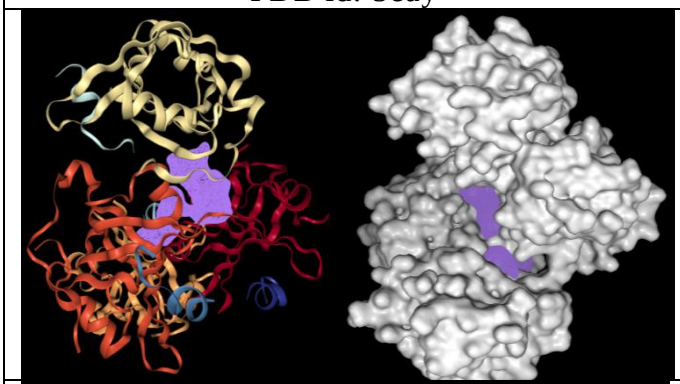
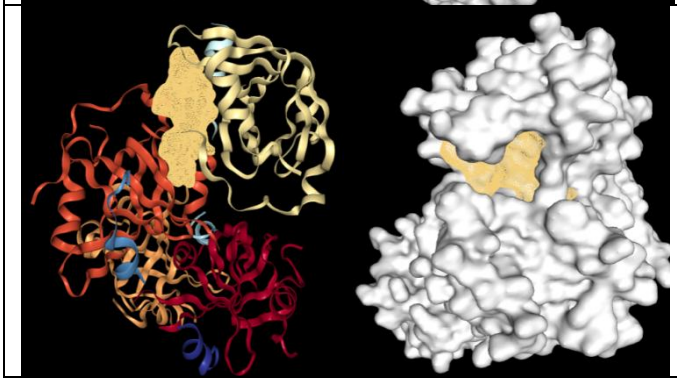
PDB id: 5aq0	Resolution: 0.95 Å	CSA: N/A
	Dscore: 0.81 Binding site type: Druggable Amino acid residues: A(2),R(2),N(6),D(1),B(0),C(0), E(0),Q(2),Z(0),G(6),H(3),I(5), L(5),K(1),M(1),F(5),P(3),S(2), T(7),W(0),Y(3),V(5) Binding site category: ENZ	

Table 24: Structural information for the protein Carboxypeptidase D

PDB id: 5eay	Resolution: 1.55 Å	CSA: N/A
	Dscore: 0.86 Binding site type: Druggable Amino acid residues: A(1),R(4),N(3),D(1),B(0),C(0), E(0),Q(5),Z(0),G(1),H(0),I(1), L(3),K(0),M(2),F(1),P(2),S(4), T(1),W(0),Y(2),V(2) Binding site category: OTH(ENZ/PPI)	
	Dscore: 0.86 Binding site type: Druggable Amino acid residues: A(0),R(1),N(3),D(0),B(0),C(0), E(2),Q(4),Z(0),G(1),H(1),I(2), L(5),K(1),M(2),F(0),P(4),S(0), T(2),W(0),Y(0),V(0) Binding site category: PPI	

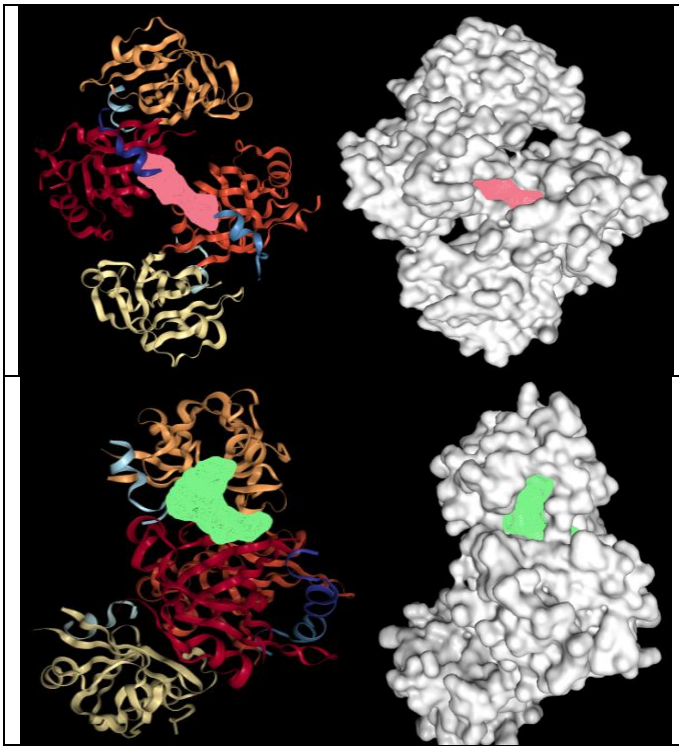
	Dscore: 0.81 Binding site type: Druggable Amino acid residues: A(2),R(2),N(0),D(0),B(0),C(0), E(2),Q(1),Z(0),G(0),H(0),I(2), L(6),K(0),M(4),F(0),P(0),S(0), T(2),W(0),Y(2),V(0) Binding site category: OTH
	Dscore: 0.77 Binding site type: Normal Amino acid residues: A(1),R(4),N(2),D(0),B(0),C(0), E(0),Q(2),Z(0),G(1),H(0),I(1), L(2),K(0),M(2),F(1),P(2),S(5), T(0),W(0),Y(2),V(2) Binding site category: PPI

Table 25:Structural information for the protein DNA replication ATP-dependent helicase/nuclease DNA2

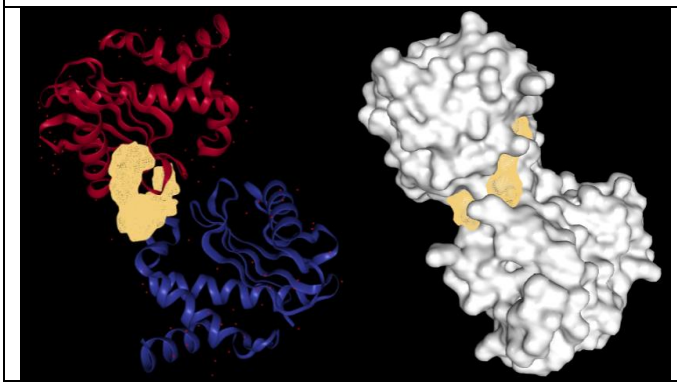
PDB id: 5mc7	Resolution: 1.6 Å	CSA: N/A
	Dscore: 0.78 Binding site type: Normal Amino acid residues: A(1),R(3),N(3),D(4),B(0),C(0), E(0),Q(3),Z(0),G(1),H(1),I(2), L(3),K(3),M(0),F(1),P(0),S(2), T(0),W(0),Y(1),V(0) Binding site category: OTH(ENZ/PPI)	

Table 26: Structural information for the protein Coatomer subunit zeta-1

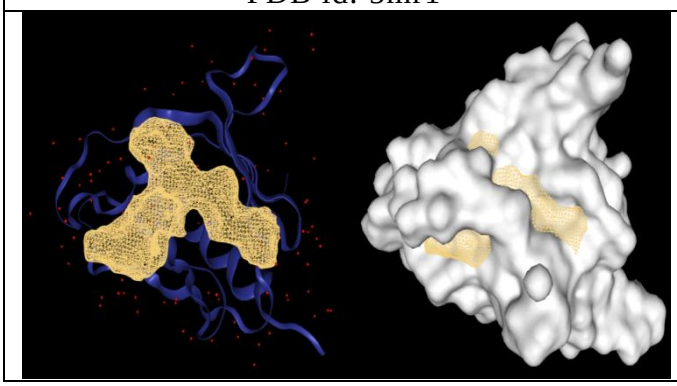
PDB id: 5mr1	Resolution: 1.2 Å	CSA: N/A
	Dscore: 0.75 Binding site type: Normal Amino acid residues: A(0),R(0),N(1),D(0),B(0),C(0), E(0),Q(1),Z(0),G(2),H(0),I(1), L(0),K(6),M(0),F(5),P(0),S(1), T(1),W(2),Y(1),V(1) Binding site category: PPI	

Table 27: Structural information for the protein Interactor protein for cytohesin exchange factors 1

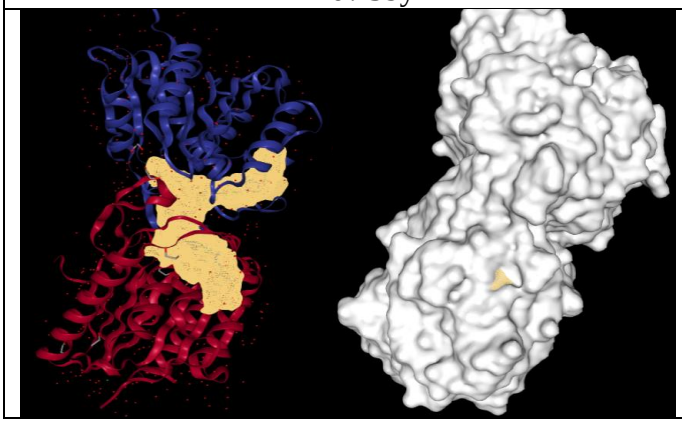
PDB id: 5sym	Resolution: 1.55 Å	CSA: N/A
	Dscore: 0.81 Binding site type: Druggable Amino acid residues: A(2),R(2),N(2),D(5),B(0),C(0), E(4),Q(2),Z(0),G(5),H(3),I(6), L(14),K(2),M(3),F(6),P(6),S(6), T(2),W(3),Y(0),V(5) Binding site category: ENZ	

Table 28: Structural information for the protein Acyl-protein thioesterase 1

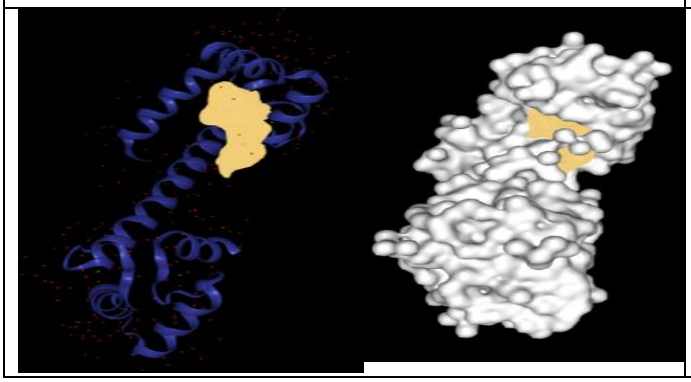
PDB id : 1ggz	Resolution: 1.5 Å	CSA: N/A
	Dscore: 0.79 Binding site type: Normal Amino acid residues: A(4),R(0),N(0),D(0),C(0), E(0),Q(2),G(0),H(0),I(4), L(0),K(2),M(0),F(2),P(2), S(0),T(0),W(1),Y(8),V(2) Binding site category: PPI	

Table 29:Structural information for the protein Calmodulin-like protein 3

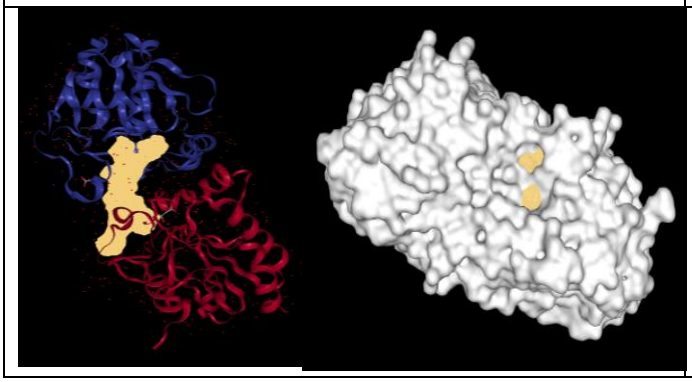
PDB id: 2f8a	Resolution: 1.5 Å	CSA: Available
	Dscore: 0.81 Binding site type: Druggable Amino acid residue: A(1),R(1),N(0),D(0),B(0),C(2), E(2),Q(3),Z(0),G(2),H(2),I(2), L(8),K(2),M(1),F(2),P(1),S(3), T(0),W(1),Y(3),V(0) Binding site category: ENZ	

Table 30:Structural information for the protein Glutathione peroxidase 13

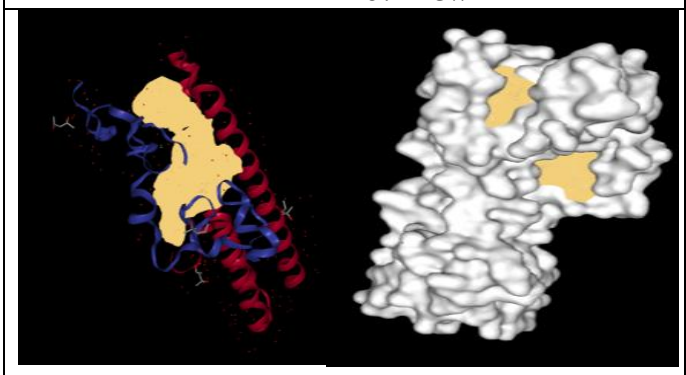
PDB id: 2fcw	Resolution: 1.26 Å	CSA: N/A
	Dscore: 0.8 Binding site type: Druggable Amino acid residue: A(1),R(0),N(0),D(0),B(0),C(1), E(1),Q(1),Z(0),G(2),H(2),I(2), L(2),K(2),M(1),F(2),P(1),S(2), T(0),W(1),Y(2),V(0) Binding site category: OTH	

Table 31: Structural information for the protein Alpha-2-macroglobulin receptor-associated protein

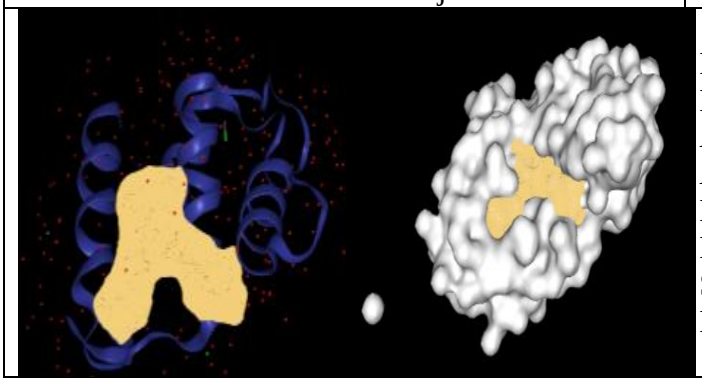
PDB id: 2fj9	Resolution: 1.6 Å	CSA: N/A
	Dscore: 0.76 Binding site type: Normal Amino acid residues: A(3),R(1),N(0),D(0),C(0), E(5),Q(3),G(1),H(0),I(4), L(0),K(2),M(0),F(3),P(3), S(5),T(0),W(1),Y(8),V(3) Binding site category: PPI	

Table 32: Structural information for the protein Acyl-CoA-binding protein

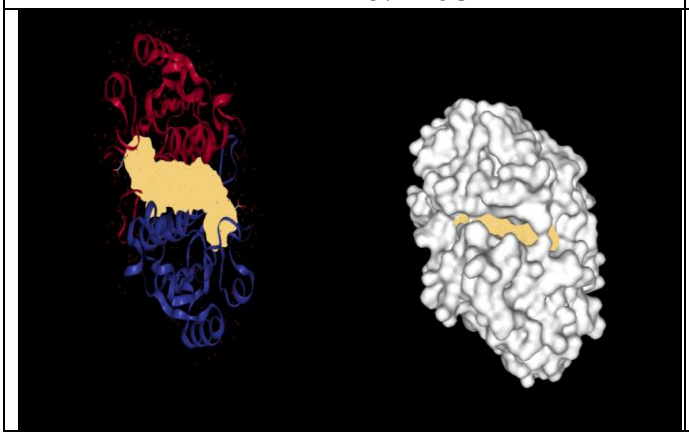
PDB id: 2ha8	Resolution: 1.6 Å	CSA: N/A
	Dscore : 0.8 Binding site type: Druggable Amino acid residue: A(0),R(0),N(0),D(1),B(0),C(2), E(0),Q(3),Z(0),G(2),H(2),I(0), L(7),K(1),M(0),F(1),P(0),S(3), T(0),W(1),Y(0),V(1) Binding site category: OTH(ENZ/PPI)	

Table 33: Structural information for the protein Probable methyltransferase TARBP1

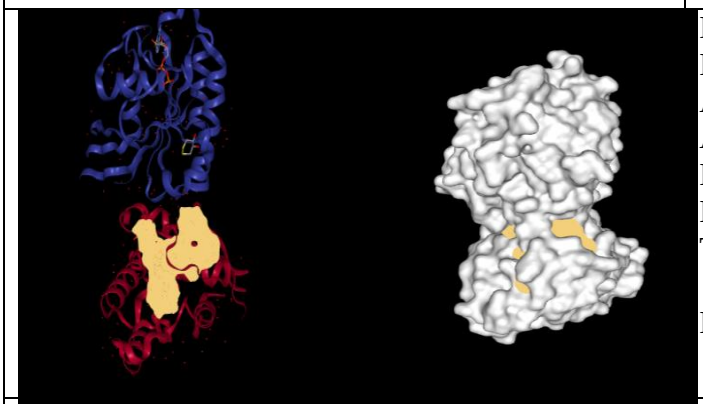
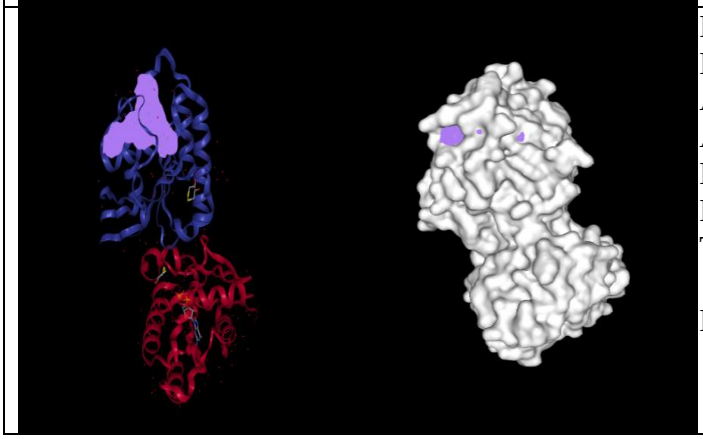
PDB id: 2hei	Resolution: 1.55 Å	CSA: Available
	Dscore: 0.81 Binding site type: Druggable Amino acid residues: A(1),R(0),N(0),D(0),B(0),C(1), E(2),Q(2),Z(0),G(2),H(2),I(3), L(4),K(3),M(1),F(2),P(1),S(2), T(0),W(2),Y(2),V(2) Binding site category: OTH	
	Dscore: 0.73 Binding site type:Normal Amino acid residues: A(0),R(2),N(4),D(4),B(0),C(0), E(0),Q(3),Z(0),G(7),H(2),I(0), L(2),K(8),M(0),F(5),P(2),S(2), T(2),W(0),Y(7),V(2) Binding site category: ENZ	

Table 34: Structural information for the protein Ras-related protein Rab-5B

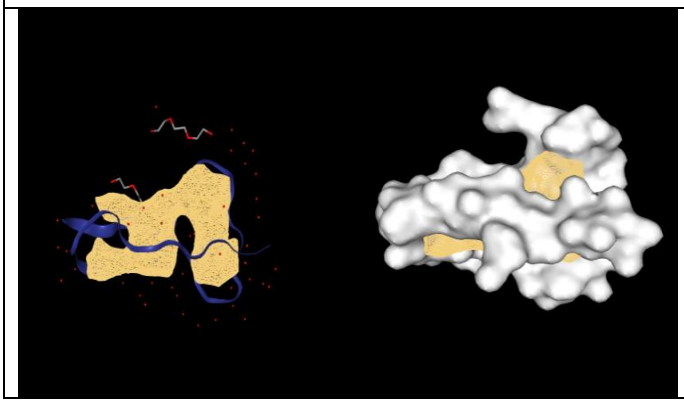
PDB id: 2jku	Resolution: 1.5 Å	CSA: N/A
	Dscore: 0.81 Binding site type: Druggable Amino acid residues: A(4),R(2),N(2),D(0),B(0),C(0), E(1),Q(2),Z(0),G(1),H(1),I(0), L(1),K(4),M(1),F(4),P(2),S(0), T(0),W(1),Y(1),V(2) Binding site category: OTH	

Table 35: Structural information for the protein Propionyl-CoA carboxylase alpha chain, mitochondrial

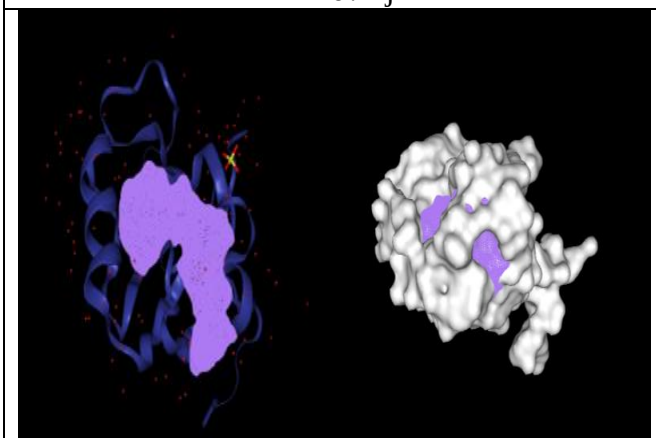
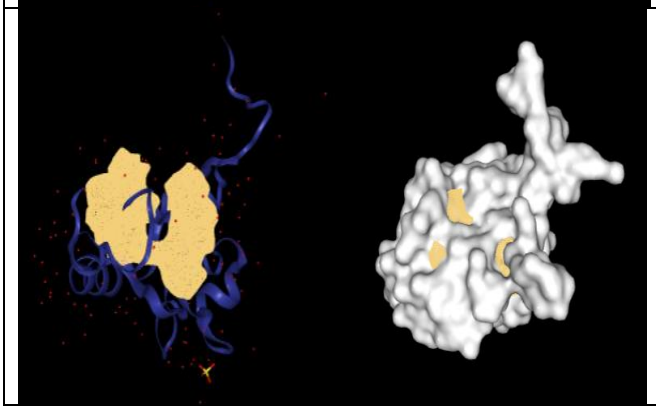
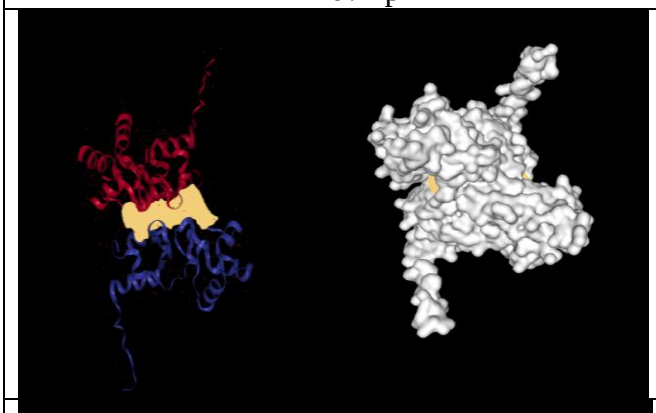
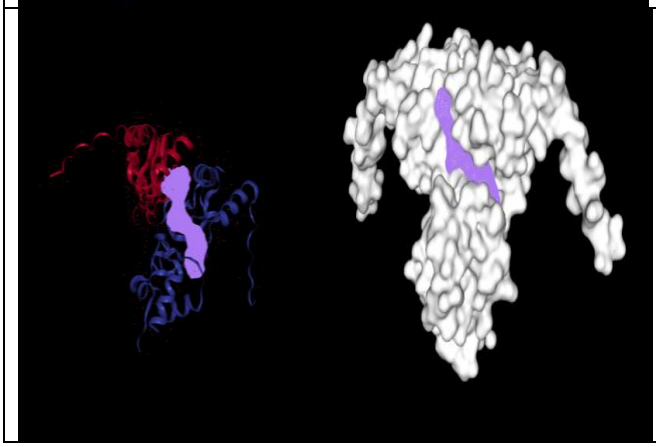
PDB id: 2jin	Resolution: 1.5 Å	CSA: N/A
	Dscore:0.77 Binding site type: Normal Amino acid residues: A(0),R(8),N(2),D(0),B(0),C(0), E(4),Q(2),Z(0),G(4),H(0),I(2), L(1),K(0),M(0),F(0),P(0),S(6), T(2),W(1),Y(10),V(2) Binding site category: OTH	
	Dscore: 0.7 Binding site type: Normal Amino acid residues: A(1),R(1),N(0),D(1),B(0),C(0), E(4),Q(2),Z(0),G(3),H(0),I(0), L(0),K(0),M(2),F(2),P(1),S(3), T(0),W(3),Y(2),V(1) Binding site category: OTH	

Table 36: Structural information for the protein Synaptojanin-2-binding protein

PDB id: 2pfi	Resolution: 1.6 Å	CSA: N/A
	Dscore: 0.81 Binding site type: Druggable Amino acid residues: A(6),R(6),N(2),D(8),B(0),C(0), E(4),Q(5),Z(0),G(4),H(3),I(2), L(2),K(6),M(2),F(0),P(8),S(4), T(2),W(3),Y(6),V(4) Binding site category:ENZ	
	Dscore: 0.85 Binding site type: Druggable Amino acid residues: A(2),R(0),N(1),D(0),B(0),C(0), E(2),Q(1),Z(0),G(4),H(0),I(4), L(2),K(3),M(1),F(2),P(1),S(2), T(1),W(1),Y(2),V(0) Binding site category: PPI	

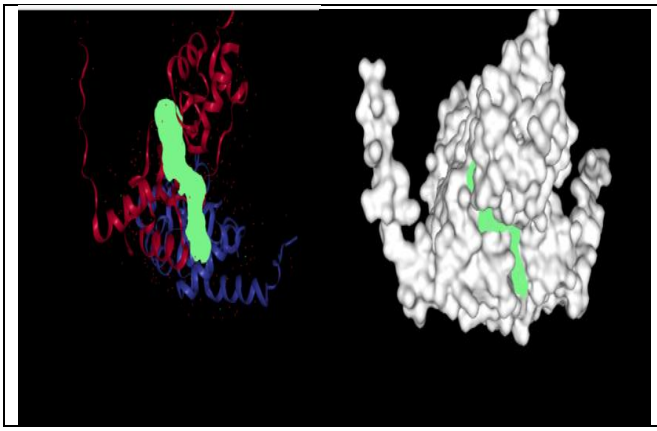
	Dscore: 0.85 Binding site type: Druggable Amino acid residues: A(3),R(0),N(2),D(2),B(0),C(0), E(0),Q(1),Z(0),G(2),H(0),I(3), L(12),K(0),M(1),F(3),P(2),S(1), T(1),W(3),Y(1),V(3) Binding site category: PPI

Table 37: Structural information for the protein Chloride channel protein ClC-Ka

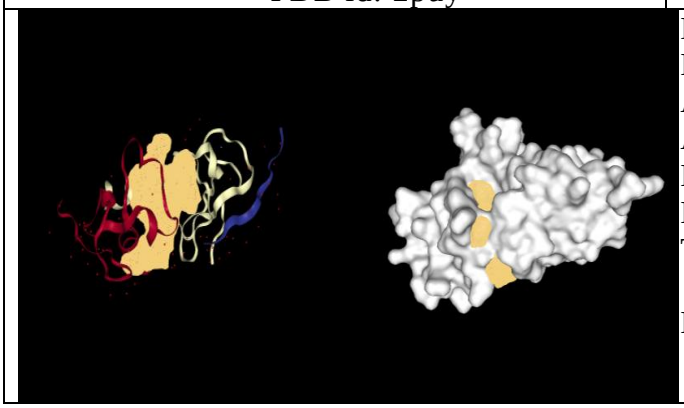
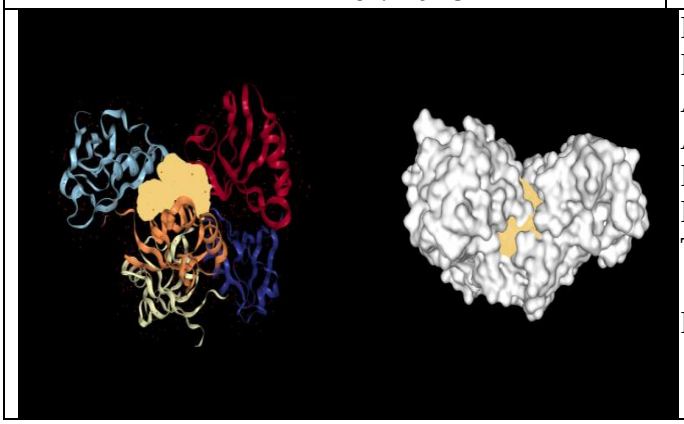
PDB id: 2puy	Resolution: 1.43 Å	CSA: N/A
	Dscore: 0.81 Binding site type: Druggable Amino acid residues: A(3),R(0),N(2),D(2),B(0),C(0), E(0),Q(1),Z(0),G(2),H(0),I(3), L(9),K(0),M(2),F(3),P(2),S(1), T(1),W(3),Y(1),V(3) Binding site category: PPI	

Table 38: Structural information for the protein PHD finger protein 21A

PDB id : 2uzc	Resolution : 1.5 Å	CSA: N/A
	Dscore: 0.81 Binding site type: Druggable Amino acid residues: A(4),R(0),N(2),D(1),B(0),C(0), E(0),Q(1),Z(0),G(3),H(1),I(4), L(9),K(0),M(3),F(3),P(2),S(1), T(0),W(2),Y(0),V(1) Binding site category: PPI	

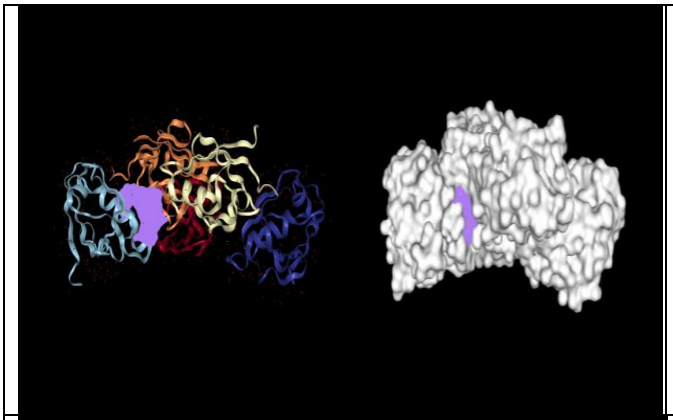
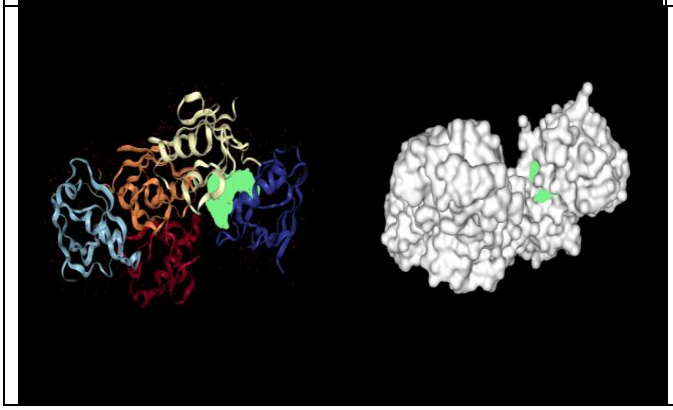
	Dscore: 0.74 Binding site type: Normal Amino acid residues: A(0),R(0),N(1),D(0),B(0),C(0), E(4),Q(2),Z(0),G(0),H(1),I(6), L(3),K(1),M(0),F(4),P(0),S(2), T(1),W(1),Y(1),V(4) Binding site category: PPI
	Dscore: 0.76 Binding site type: Normal Amino acid residues: A(1),R(2),N(1),D(2),B(0),C(0), E(1),Q(2),Z(0),G(1),H(3),I(2), L(1),K(0),M(1),F(1),P(0),S(4), T(2),W(1),Y(2),V(1) Binding site category: OTH

Table 39: Structural information for the protein PDZ and LIM domain protein 5

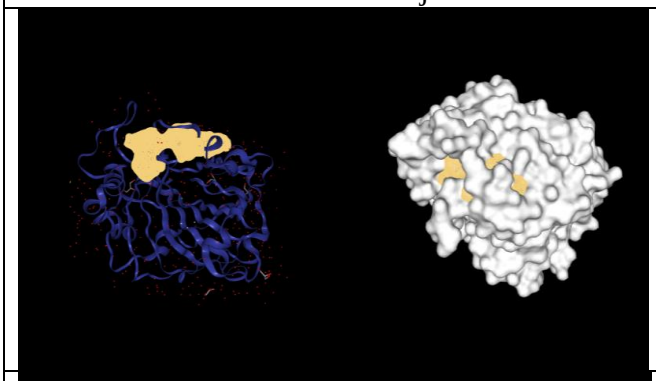
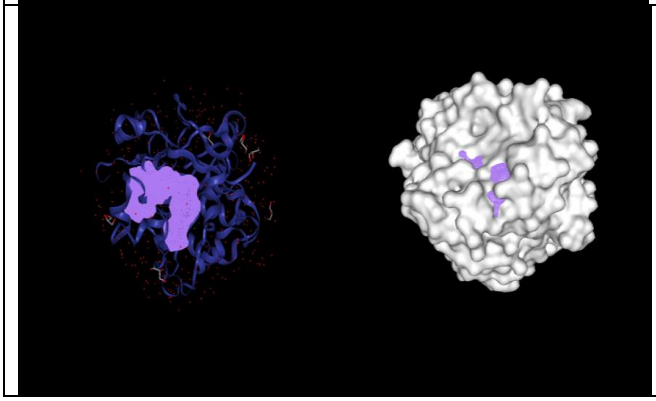
PDB id: 2w2j	Resolution: 1.6 Å	CSA: Available
	Dscore: 0.82 Binding site type: Druggable Amino acid residues: A(6),R(4),N(5),D(1),B(0),C(3), E(1),Q(0),Z(0),G(6),H(1),I(3), L(6),K(2),M(2),F(1),P(2),S(5), T(4),W(0),Y(2),V(6) Binding site category: ENZ	
	Dscore: 0.88 Binding site type: Druggable Amino acid residues: A(3),R(4),N(2),D(5),B(0),C(0), E(1),Q(2),Z(0),G(3),H(1),I(0), L(6),K(2),M(2),F(4),P(1),S(6), T(2),W(0),Y(4),V(1) Binding site category: ENZ	

Table 40: Structural information for the protein Carbonic anhydrase-related protein

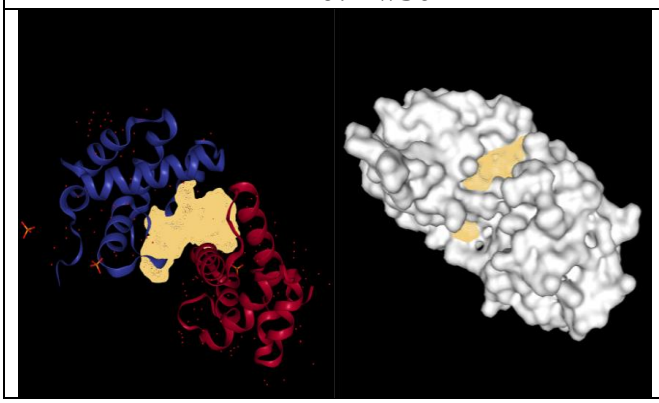
PDB id: 2w50	Resolution: 1.6 Å	CSA: N/A
	Dscore: 0.79 Binding site type: Normal Amino acid residues: A(5),R(3),N(5),D(8),B(0),C(3), E(1),Q(0),Z(0),G(6),H(5),I(3), L(6),K(2),M(1),F(1),P(2),S(6), T(5),W(0),Y(2),V(5) Binding site category: ENZ	

Table 41: Structural information for the protein Cerebral dopamine neurotrophic factor

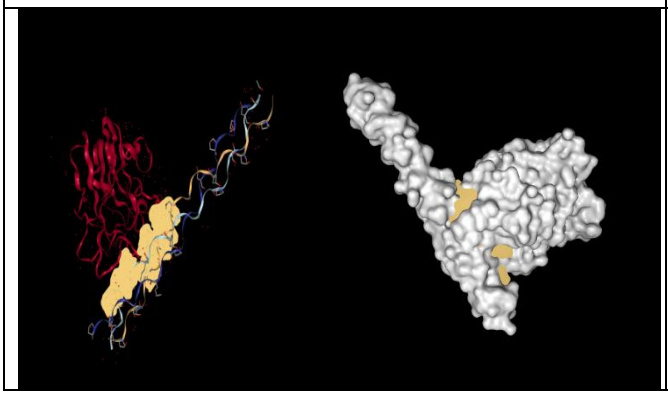
PDB id: 2wuh	Resolution: 1.6 Å	CSA: N/A
	Dscore: 0.8 Binding site type: Druggable Amino acid residues: A(5),R(4),N(4),D(4),B(0),C(3), E(1),Q(0),Z(0),G(6),H(5),I(2), L(6),K(2),M(1),F(1),P(1),S(6), T(4),W(0),Y(2),V(5) Binding site category:ENZ	

Table 42: Structural information for the protein Discoidin domain-containing receptor protein – tyrosine kinase

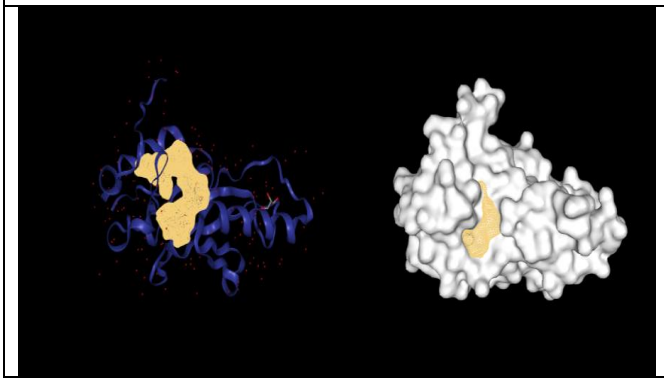
PDB id: 2wzo	Resolution: 1.6 Å	CSA: N/A
	Dscore: 0.75 Binding site type: normal Amino acid residues: A(1),R(5),N(0),D(8),B(0),C(6), E(4),Q(1),Z(0),G(0),H(6),I(2), L(0),K(1),M(0),F(0),P(1),S(0), T(0),W(0),Y(1),V(2) Binding site category: ENZ	

Table 43: Structural information for the protein Transforming growth factor beta regulator 1

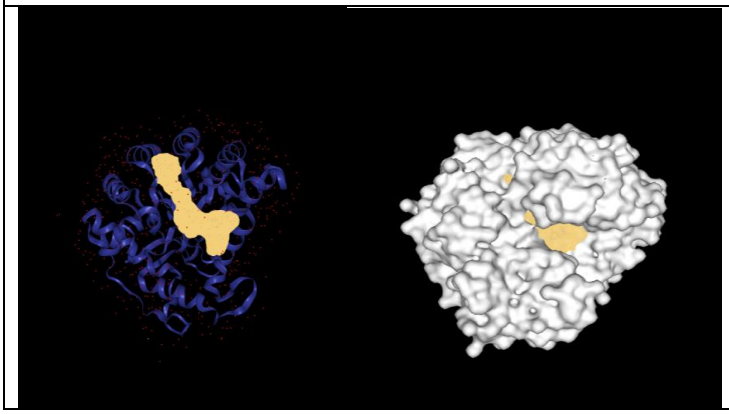
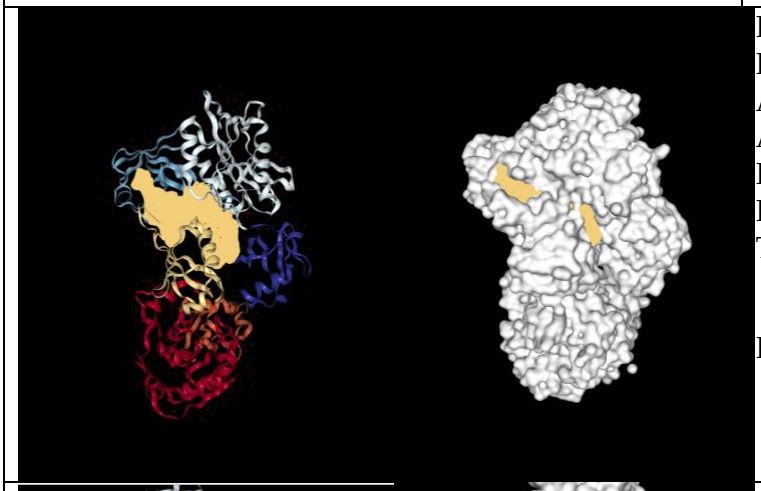
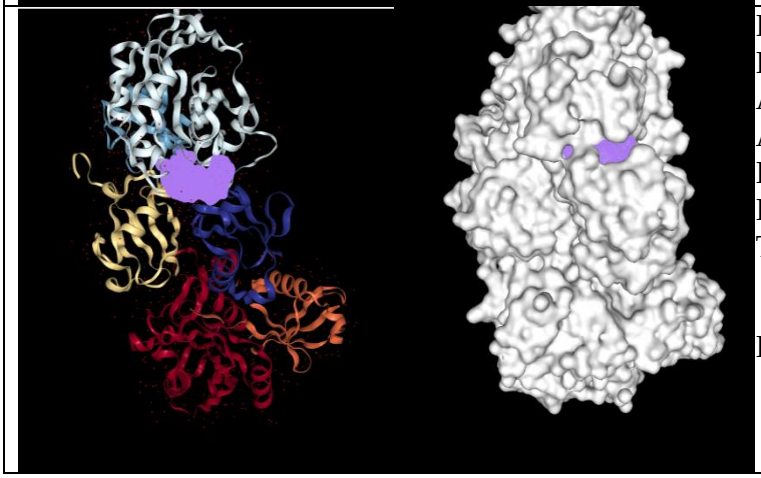
PDB id: 2xio	Resolution: 1.19 Å	CSA: N/A
	Dscore: 0.82 Binding site type: Druggable Amino acid residues: A(0),R(0),N(1),D(0),B(0),C(0), E(4),Q(2),Z(0),G(0),H(1),I(6), L(3),K(1),M(0),F(4),P(0),S(2), T(1),W(1),Y(1),V(4) Binding site category: OTH	

Table 44: Structural information for the protein Putative deoxyribonuclease TATDN1

PDB id: 2znv	Resolution: 1.6 Å	CSA: N/A
	Dscore: 0.8 Binding site type: Druggable Amino acid residues: A(2),R(2),N(0),D(2),B(0),C(0), E(0),Q(2),Z(0),G(1),H(0),I(7), L(0),K(0),M(3),F(0),P(4),S(2), T(2),W(0),Y(2),V(0) Binding site category: PPI	
	Dscore: 0.81 Binding site type: Druggable Amino acid residues: A(5),R(0),N(5),D(0),B(0),C(0), E(1),Q(2),Z(0),G(7),H(1),I(2), L(7),K(2),M(1),F(1),P(2),S(4), T(3),W(0),Y(1),V(6) Binding site category: OTH	

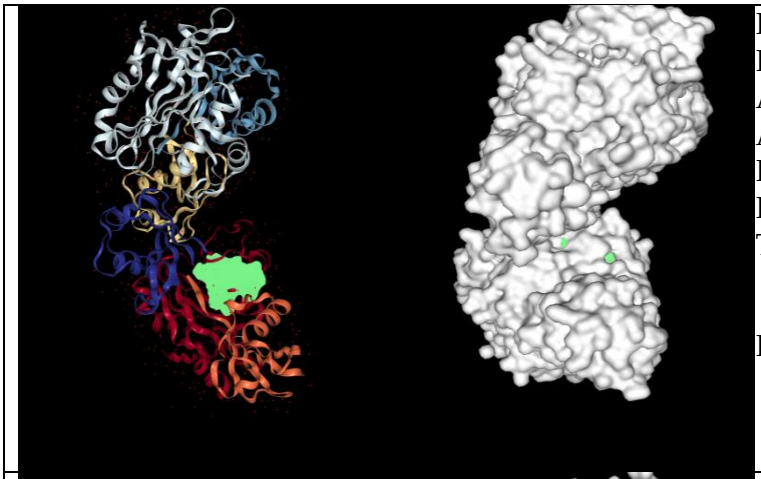
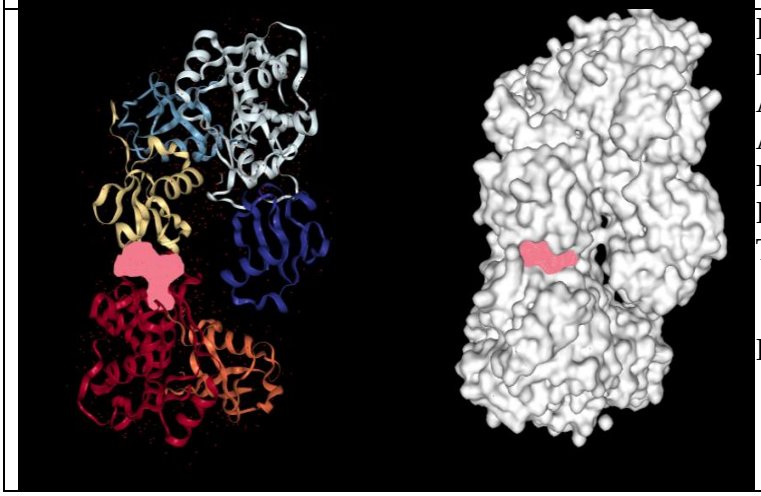
	Dscore: 0.76 Binding site type: Normal Amino acid residues: A(1),R(4),N(0),D(1),B(0),C(1), E(3),Q(1),Z(0),G(1),H(2),I(2), L(2),K(0),M(2),F(0),P(1),S(1), T(4),W(0),Y(0),V(0) Binding site category: OTH
	Dscore: 0.72 Binding site type: Normal Amino acid residues: A(0),R(0),N(2),D(1),B(0),C(0), E(0),Q(1),Z(0),G(8),H(0),I(5), L(3),K(1),M(1),F(5),P(2),S(8), T(3),W(2),Y(3),V(3) Binding site category: PPI

Table 45: Structural information for the protein AMSH-like

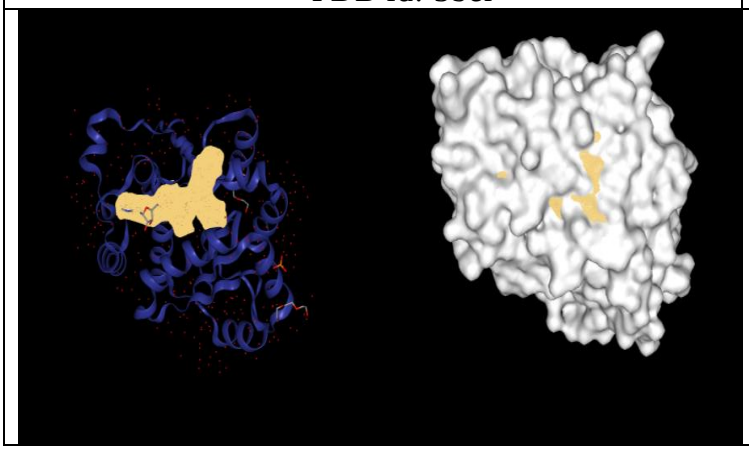
PDB id: 3ber	Resolution: 1.4 Å	CSA: N/A
	Dscore: 0.76 Binding site type: Normal Amino acid residues: A(1),R(4),N(0),D(6),B(0),C(5), E(4),Q(1),Z(0),G(0),H(8),I(2), L(0),K(1),M(0),F(0),P(1),S(0), T(0),W(0),Y(1),V(2) Binding site category: ENZ	

Table 46: Structural information for the protein Probable ATP-dependent RNA helicase DDX47.

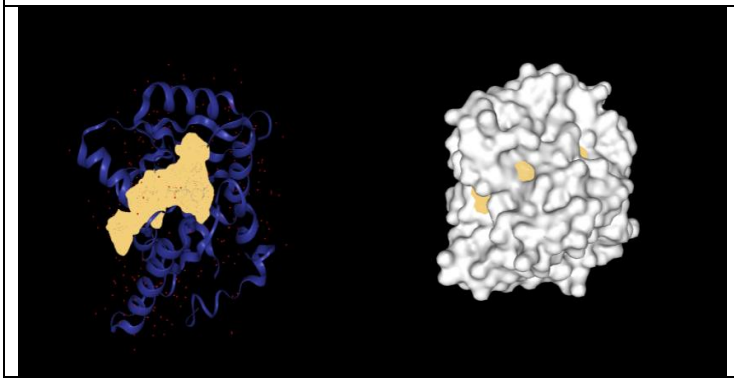
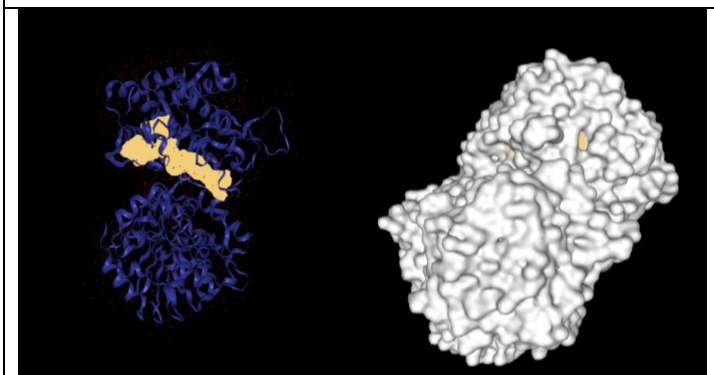
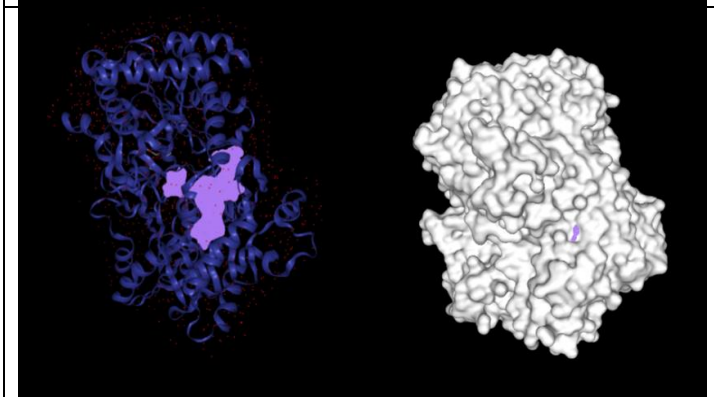
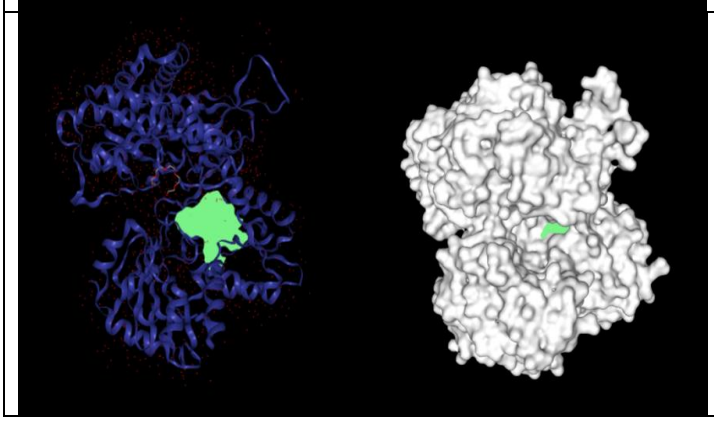
PDB id: 3ckk	Resolution:1.55 Å	CSA: N/A
	Dscore: 0.79 Binding site type: Normal Amino acid residues: A(3),R(3),N(2),D(3),B(0),C(8), E(2),Q(1),Z(0),G(1),H(5),I(1), L(1),K(4),M(1),F(1),P(1),S(2), T(2),W(3),Y(2),V(4) Binding site category: ENZ	

Table 47: Structural information for the protein tRNA (guanine-N(7)-)-methyltransferase

PDB id: 3ctz	Resolution: 1.6 Å	CSA: Available
	Dscore: 0.8 Binding site type: Druggable Amino acid residues: A(1),R(5),N(2),D(4),B(0),C(3), E(2),Q(1),Z(0),G(1),H(8),I(1), L(3),K(0),M(1),F(0),P(2),S(2), T(3),W(2),Y(2),V(4) Binding site category: ENZ	
	Dscore: 0.77 Binding site type: Normal Amino acid residues: A(1),R(4),N(2),D(6),B(0),C(8), E(1),Q(6),Z(0),G(5),H(3),I(2), L(2),K(1),M(0),F(1),P(1),S(2), T(0),W(0),Y(3),V(3) Binding site category: ENZ	
	Dscore: 0.79 Binding site type: Normal Amino acid residues: A(3),R(3),N(2),D(3),B(0),C(5), E(2),Q(1),Z(0),G(1),H(8),I(1), L(1),K(4),M(1),F(1),P(1),S(2), T(2),W(3),Y(2),V(4) Binding site category:ENZ	

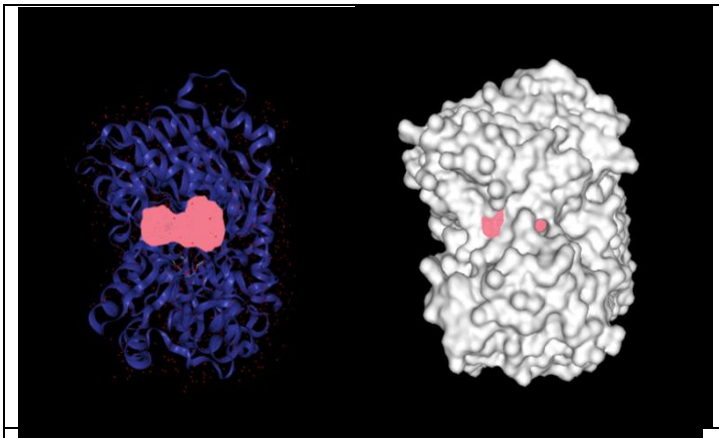
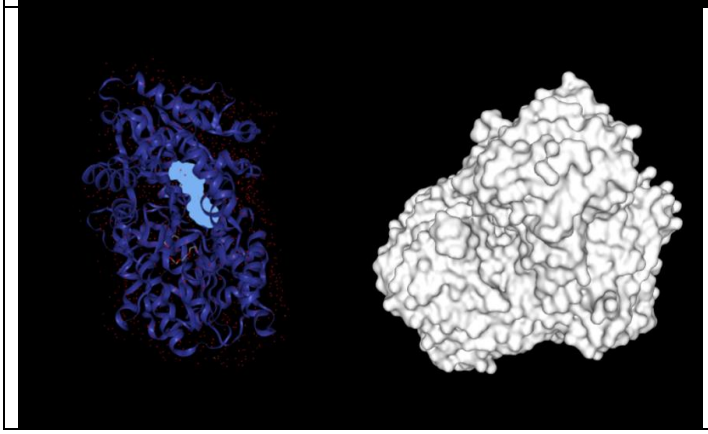
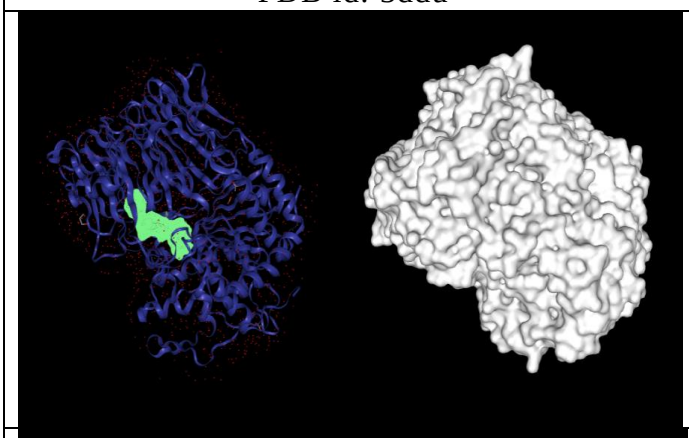
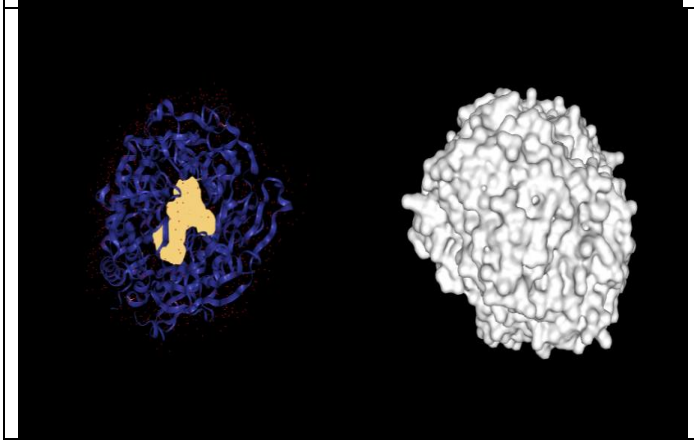
	Dscore: 0.82 Binding site type: Druggable Amino acid residues: A(1),R(5),N(0),D(4),B(0),C(0), E(0),Q(1),Z(0),G(0),H(3),I(2), L(2),K(1),M(1),F(4),P(2),S(6), T(3),W(0),Y(1),V(1) Binding site category: OTH
	Dscore: 0.76 Binding site type: Normal Amino acid residues: A(0),R(2),N(1),D(6),B(0),C(3), E(0),Q(1),Z(0),G(2),H(10),I(1), L(0),K(6),M(0),F(5),P(0),S(1), T(1),W(2),Y(1),V(1) Binding site category: ENZ

Table 48: Structural information for the protein Xaa-Pro aminopeptidase 1

PDB id: 3ddu	Resolution: 1.56 Å	CSA: Available
	Dscore: 0.81 Binding site type: Druggable Amino acid residues: A(2),R(2),N(2),D(5),B(0),C(5), E(4),Q(2),Z(0),G(5),H(7),I(6), L(14),K(2),M(3),F(6),P(6),S(6), T(2),W(3),Y(1),V(5) Binding site category: ENZ	
	Dscore: 0.79 Binding site type: Normal Amino acid residues: A(1),R(3),N(3),D(4),B(0),C(3), E(0),Q(3),Z(0),G(1),H(9),I(2), L(3),K(3),M(0),F(1),P(0),S(2), T(0),W(0),Y(1),V(0) Binding site category: ENZ	

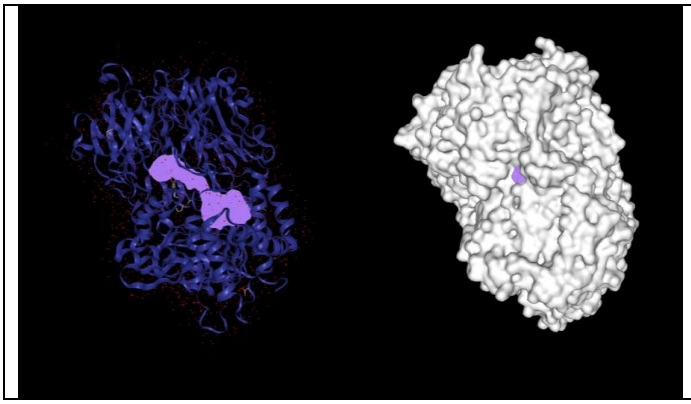
	Dscore: 0.77 Binding site type: Normal Amino acid residues: A(1),R(4),N(2),D(5),B(0),C(5), E(0),Q(2),Z(0),G(1),H(7),I(1), L(2),K(0),M(2),F(1),P(2),S(5), T(0),W(0),Y(2),V(2) Binding site category: ENZ

Table 49: Structural information for the protein Prolyl endopeptidase

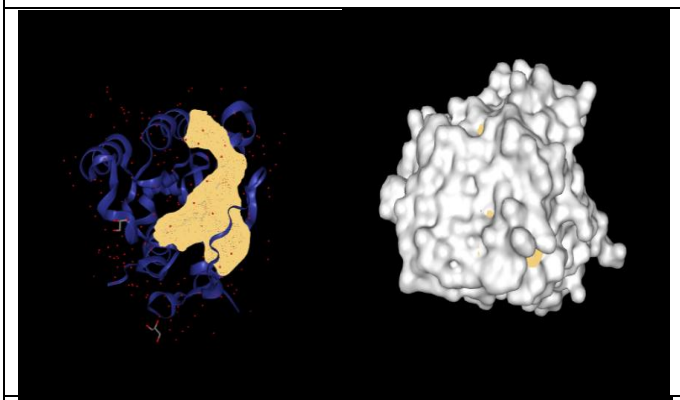
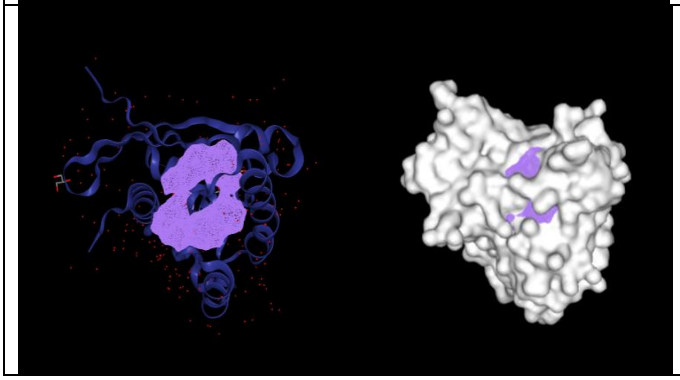
PDB id: 3e5h	Resolution: 1.499 Å	CSA: Available
	Dscore: 0.76 Binding site type: Normal Amino acid residues: A(0),R(4),N(0),D(2),B(0),C(3), E(6),Q(2),Z(0),G(2),H(0),I(0), L(0),K(0),M(0),F(2),P(2),S(0), T(0),W(0),Y(0),V(1) Binding site category: OTH	
	Dscore: 0.72 Binding site type: Normal Amino acid residues: A(3),R(1),N(1),D(2),B(0),C(0), E(1),Q(0),Z(0),G(2),H(1),I(1), L(4),K(1),M(0),F(0),P(0),S(1), T(0),W(1),Y(0),V(3) Binding site category: OTH	

Table 50: Structural information for the protein Ras-related protein Rab-28

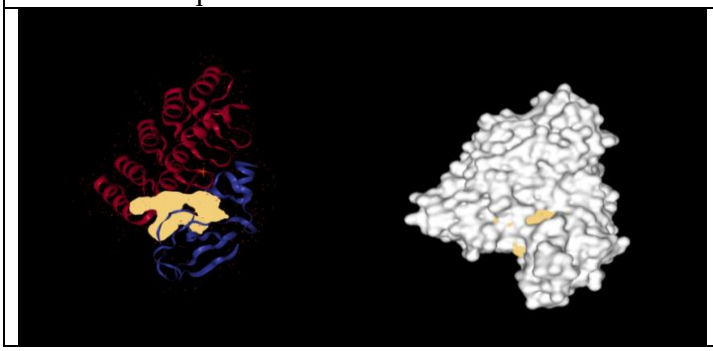
PDB id: 3f6q	Resolution: 1.6 Å	CSA: N/A
	Dscore: 0.86 Binding site type: Druggable Amino acid residues: A(1),R(4),N(3),D(1),B(0),C(0), E(0),Q(5),Z(0),G(1),H(0),I(1), L(3),K(3),M(2),F(1),P(2),S(4), T(1),W(3),Y(5),V(2) Binding site category: OTH	

Table 51: Structural information for the protein LIM and senescent cell antigen-like-containing domain protein 1

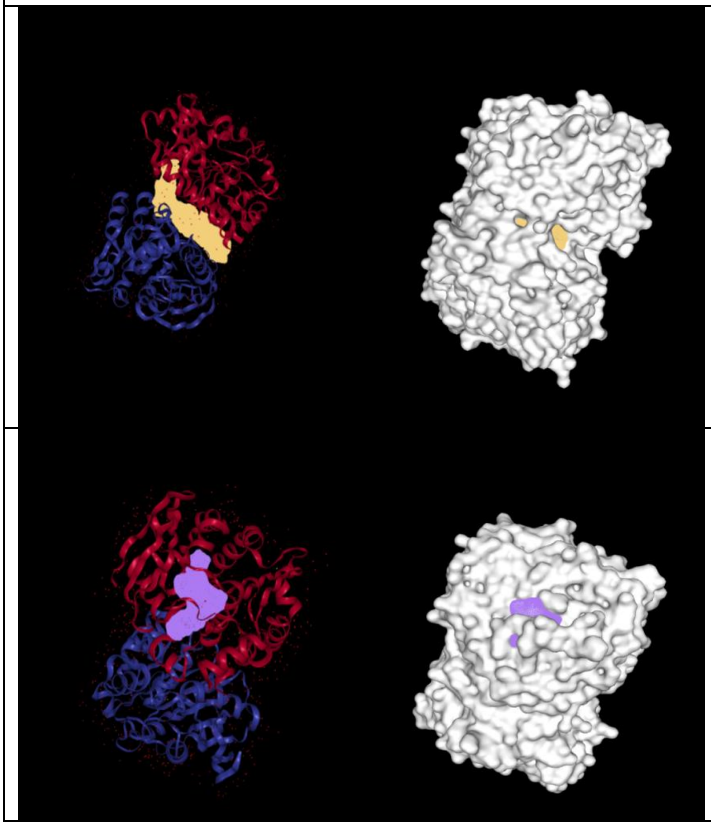
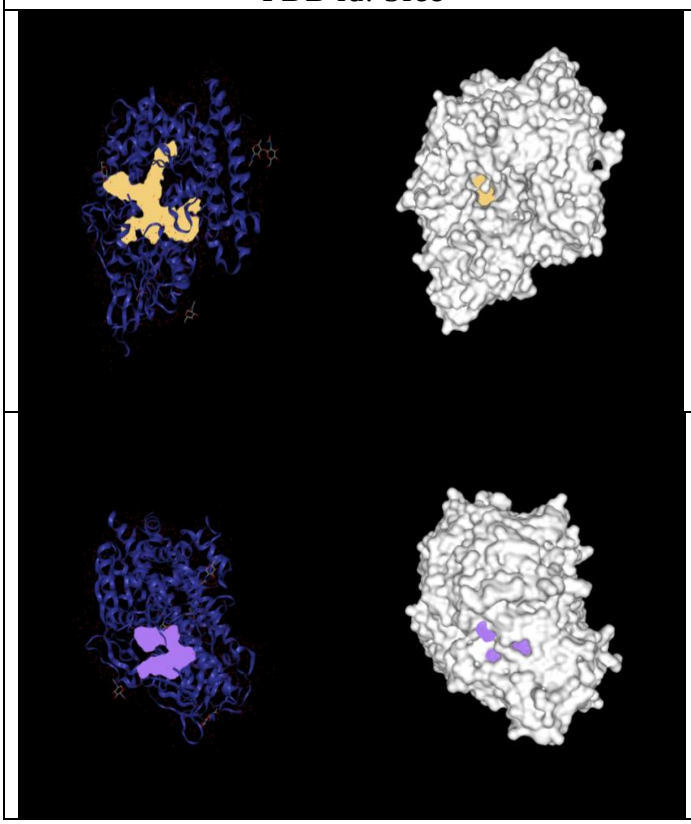
PDB id: 3fcx	Resolution: 1.5 Å	CSA: Available
	Dscore: 0.79 Binding site type: Normal Amino acid residues: A(1),R(4),N(2),D(6),B(0),C(2), E(4),Q(1),Z(0),G(0),H(8),I(2), L(0),K(1),M(0),F(0),P(1),S(0), T(0),W(0),Y(1),V(2) Binding site category: ENZ	
	Dscore: 0.77 Binding site type: Normal Amino acid residues: A(1),R(2),N(0),D(4),B(0),C(0), E(1),Q(6),Z(0),G(5),H(3),I(2), L(2),K(1),M(0),F(1),P(1),S(2), T(0),W(0),Y(3),V(3) Binding site category: OTH	

Table 52: Structural information for the protein S-formylglutathione hydrolase

PDB id: 3fee	Resolution: 1.56 Å	CSA: Available
	Dscore: 0.8 Binding site type: Druggable Amino acid residues: A(10),R(8),N(3),D(3),B(0),C(4), E(4),Q(8),Z(0),G(9),H(8),I(5), L(16),K(12),M(2),F(5),P(6),S(6), T(7),W(2),Y(7),V(11) Binding site category:ENZ	
	Dscore: 0.8 Binding site type: Druggable Amino acid residues: A(2),R(2),N(2),D(4),B(0),C(2), E(4),Q(2),Z(0),G(8),H(2),I(4), L(4),K(3),M(0),F(2),P(1),S(5), T(5),W(1),Y(2),V(3) Binding site category: ENZ	

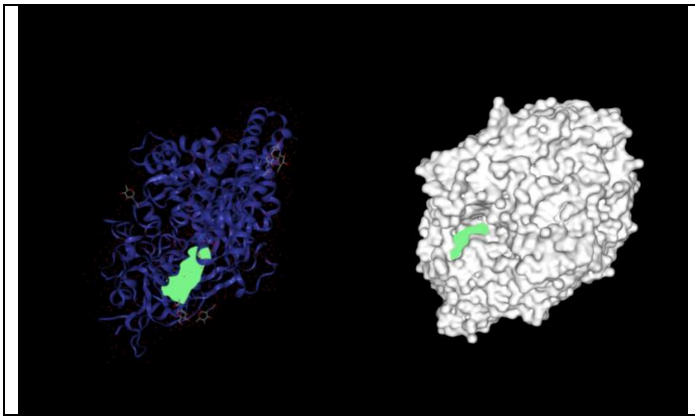
	Dscore: 0.71 Binding site type: Normal Amino acid residues: A(1),R(0),N(1),D(0),B(0),C(0), E(1),Q(2),Z(0),G(1),H(0),I(2), L(1),K(0),M(1),F(1),P(0),S(4), T(2),W(1),Y(2),V(1) Binding site category: PPI

Table 53: Structural information for the protein N-acetylated-alpha-linked acidic dipeptidase 2

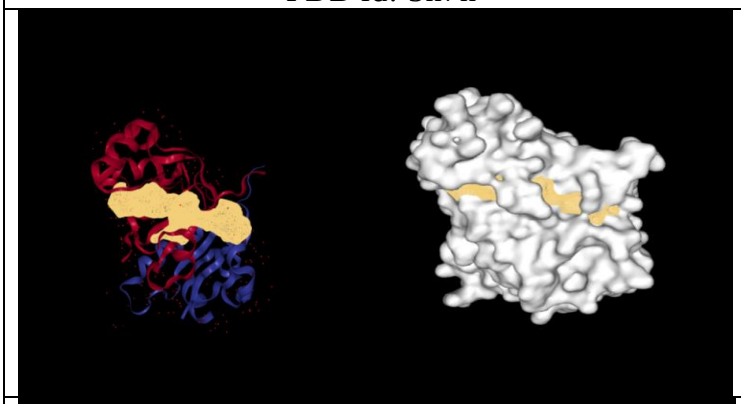
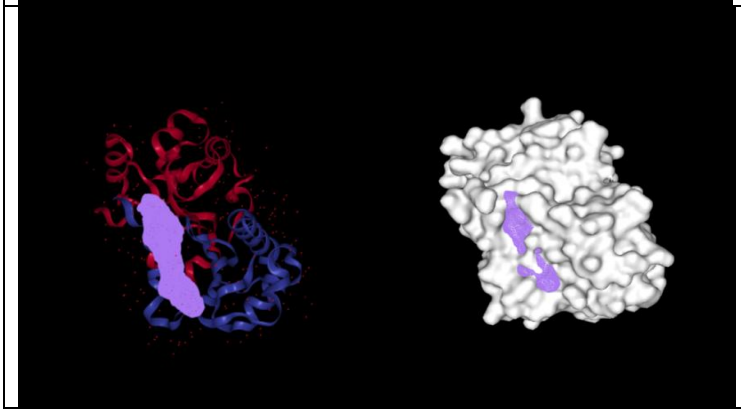
PDB id: 3h7h	Resolution: 1.55 Å	CSA: N/A
	Dscore: 0.85 Binding site type: Druggable Amino acid residues: A(0),R(1),N(0),D(0),B(0),C(0), E(2),Q(4),Z(0),G(1),H(1),I(2), L(5),K(1),M(2),F(0),P(4),S(0), T(2),W(0),Y(0),V(0) Binding site category: PPI	
	Dscore: 0.77 Binding site type: Normal Amino acid residues: A(1),R(2),N(0),D(0),B(0),C(0), E(4),Q(1),Z(0),G(0),H(0),I(2), L(0),K(1),M(0),F(0),P(1),S(0), T(0),W(0),Y(1),V(2) Binding site category: PPI	

Table 54: Structural information for the protein Transcription elongation factor SPT4

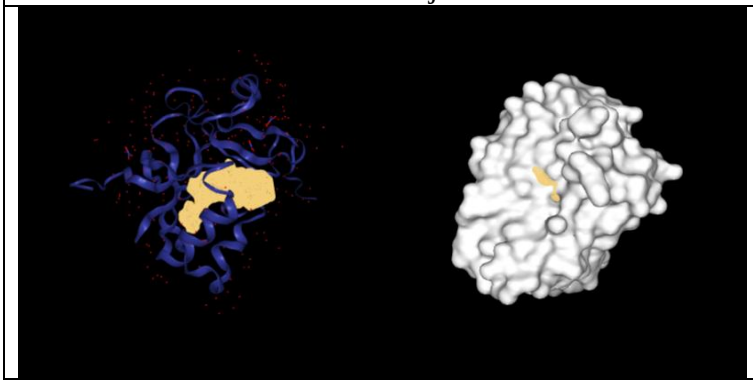
PDB id: 3jub	Resolution: 1.2 Å	CSA: N/A
	Dscore: 0.8 Binding site type: Druggable Amino acid residues: A(2),R(2),N(3),D(4),B(0),C(8), E(0),Q(5),Z(0),G(3),H(5),I(4), L(3),K(5),M(2),F(2),P(5),S(2), T(3),W(0),Y(1),V(7) Binding site category: ENZ	

Table 55: Structural information for the protein Gamma-glutamylaminecyclotransferase

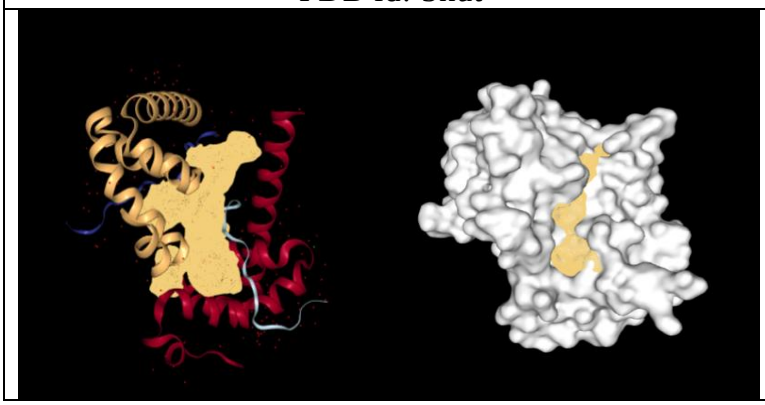
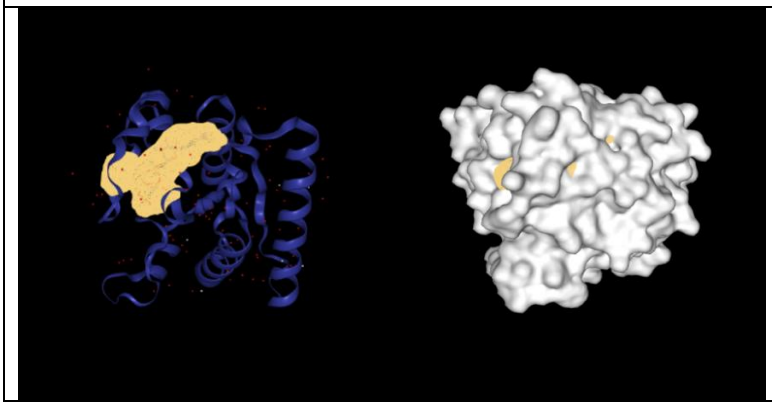
PDB id: 3kut	Resolution: 1.5 Å	CSA: N/A
	Dscore: 0.81 Binding site type: Druggable Amino acid residues: A(2),R(2),N(0),D(0),B(0),C(0), E(2),Q(1),Z(0),G(0),H(0),I(2), L(6),K(0),M(4),F(0),P(0),S(0), T(2),W(0),Y(2),V(0) Binding site category: PPI	

Table 56: Structural information for the protein Polyadenylate-binding protein-interacting protein 2

PDB id: 3llu	Resolution: 1.4 Å	CSA: N/A
	Dscore: 0.81 Binding site type: Druggable Amino acid residues: A(2),R(2),N(2),D(0),B(0),C(0), E(4),Q(2),Z(0),G(5),H(0),I(6), L(14),K(2),M(3),F(6),P(6),S(6), T(2),W(3),Y(0),V(5) Binding site category: OTH	

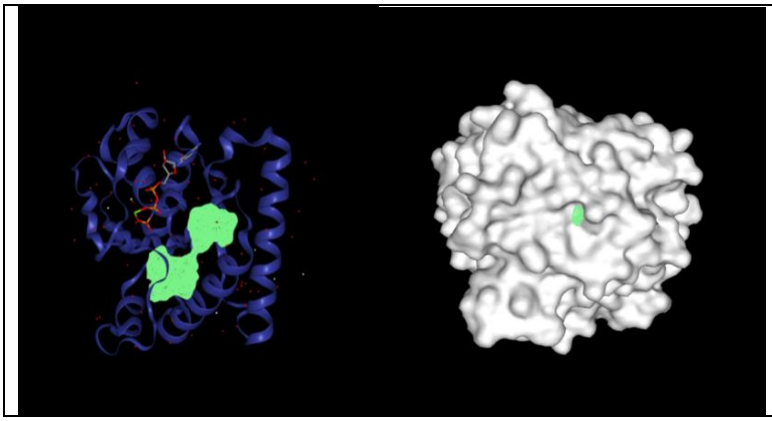
	Dscore: 0.72 Binding site type: Normal Amino acid residues: A(0),R(5),N(1),D(6),B(0),C(6), E(0),Q(1),Z(0),G(2),H(8),I(1), L(0),K(6),M(0),F(5),P(0),S(1), T(1),W(2),Y(1),V(1) Binding site category: ENZ

Table 57: Structural information for the protein Ras-related GTP-binding protein C

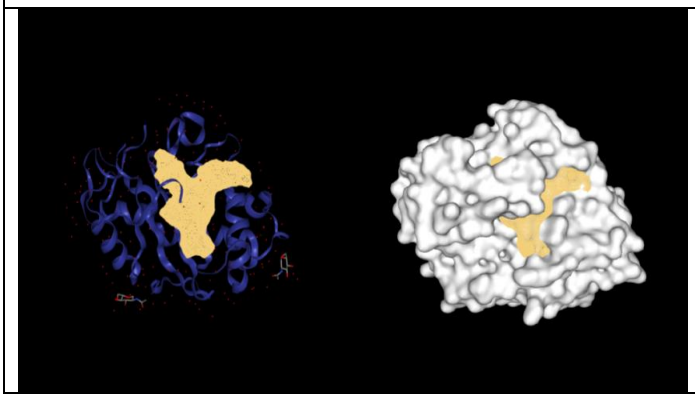
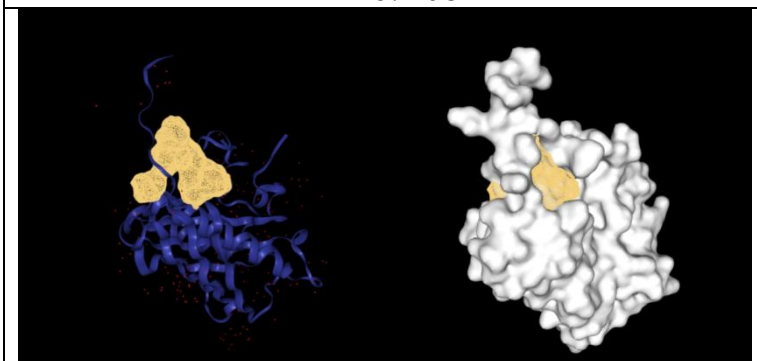
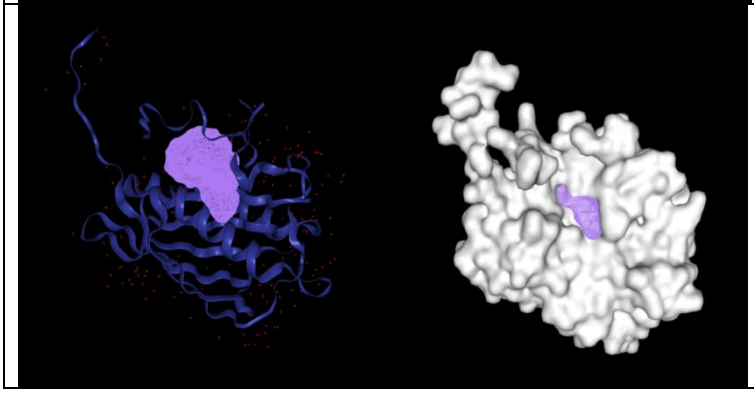
PDB id: 3t0o	Resolution: 1.59 Å	CSA: Available
	Dscore: 0.72 Binding site type: Normal Amino acid residues: A(1),R(6),N(3),D(4),B(0),C(8), E(0),Q(3),Z(0),G(1),H(10),I(2), L(3),K(3),M(0),F(1),P(0),S(2), T(0),W(2),Y(1),V(0) Binding site category: ENZ	

Table 58: Structural information for the protein Ribonuclease T2

PDB id: 4ae7	Resolution: 1.45 Å	CSA: N/A
	Dscore: 0.74 Binding site type: Normal Amino acid residues: A(1),R(2),N(1),D(2),B(0),C(0), E(1),Q(2),Z(0),G(1),H(0),I(2), L(1),K(0),M(1),F(1),P(0),S(4), T(2),W(1),Y(2),V(1) Binding site category: OTH	
	Dscore: 0.72 Binding site type: Normal Amino acid residues: A(1),R(2),N(0),D(4),B(0),C(3), E(1),Q(2),Z(0),G(1),H(6),I(2), L(1),K(0),M(2),F(1),P(0),S(4), T(2),W(1),Y(2),V(2) Binding site category: ENZ	

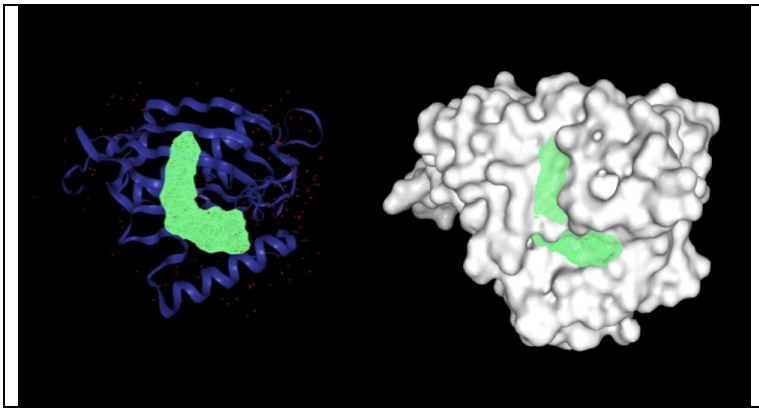
	Dscore: 0.71 Binding site type: Normal Amino acid residues: A(0),R(1),N(1),D(1),B(0),C(0), E(1),Q(0),Z(0),G(1),H(0),I(2), L(1),K(0),M(1),F(1),P(0),S(4), T(2),W(1),Y(2),V(3) Binding site category: PPI

Table 59: Structural information for the protein Acyl-coenzyme Athioesterase THEM5

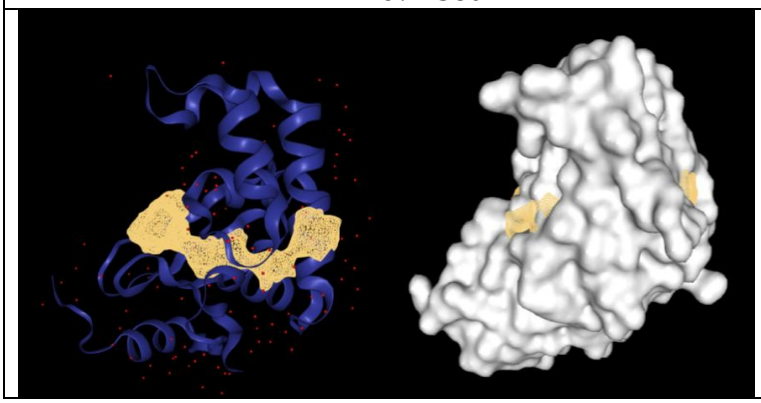
PDB id: 4bou	Resolution: 1.55 Å	CSA: N/A
	Dscore: 0.79 Binding site type: Normal Amino acid residues: A(1),R(3),N(3),D(4),B(0),C(5), E(0),Q(3),Z(0),G(1),H(10),I(2), L(3),K(3),M(0),F(1),P(0),S(2), T(0),W(2),Y(1),V(0) Binding site category: ENZ	

Table 60: Structural information for the protein UbiquitinylHydrolase 1

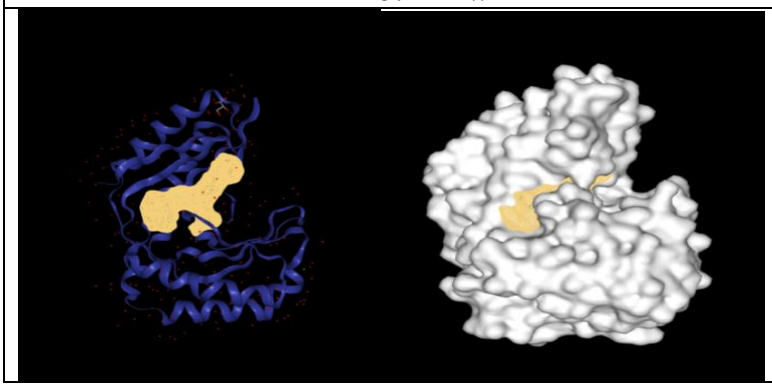
PDB id: 4h7w	Resolution: 1.1 Å	CSA: N/A
	Dscore: 0.74 Binding site type: Normal Amino acid residues: A(0),R(0),N(1),D(1),B(0),C(1), E(0),Q(1),Z(0),G(2),H(2),I(1), L(0),K(6),M(0),F(5),P(0),S(1), T(1),W(2),Y(1),V(1) Binding site category: OTH	

Table 61: Structural information for the protein snRNA phosphodiesterase

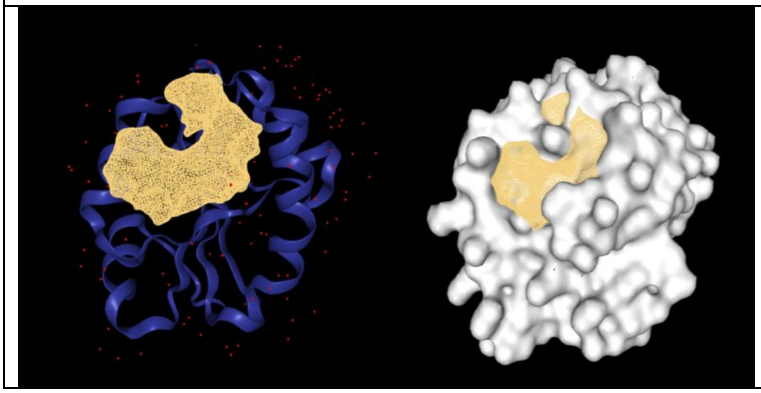
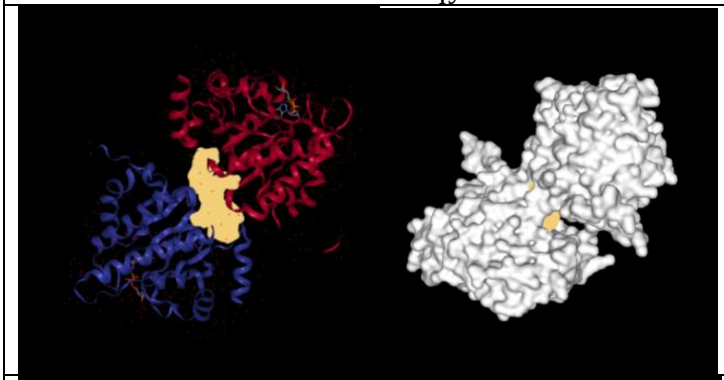
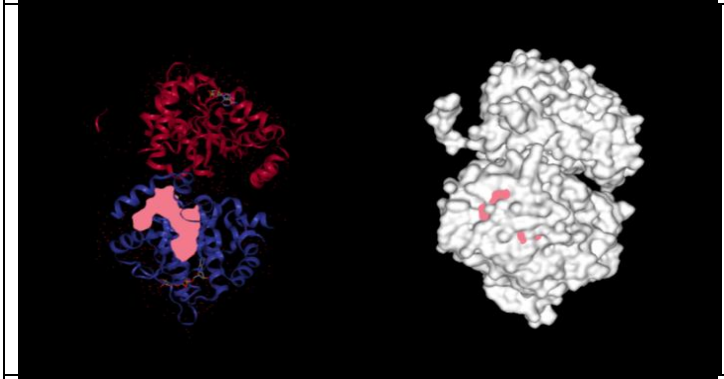
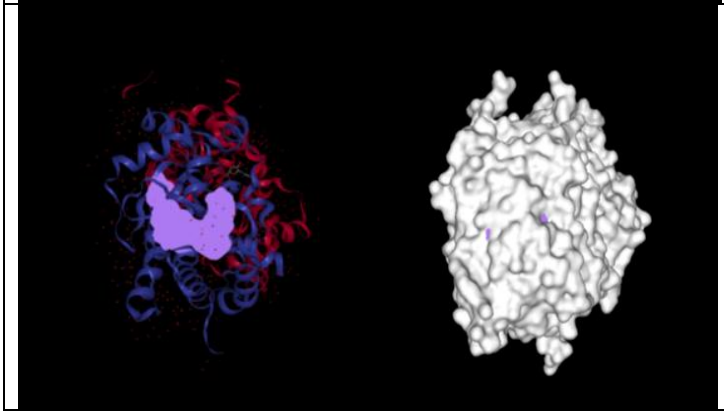
PDB id: 4i6x	Resolution: 1.5 Å	CSA: N/A
	Dscore: 0.78 Binding site type: Normal Amino acid residues: A(7),R(0),N(1),D(0),B(0),C(0), E(0),Q(4),Z(0),G(1),H(0),I(10), L(5),K(2),M(0),F(1),P(1),S(1), T(0),W(1),Y(0),V(1) Binding site category: PPI	

Table 62: Structural information for the protein disulphide- isomerase

PDB id: 4iqy	Resolution: 1.55 Å	CSA: N/A
	Dscore: 0.83 Binding site type: Druggable Amino acid residues: A(1),R(4),N(3),D(1),B(0),C(0), E(0),Q(5),Z(0),G(1),H(0),I(1), L(3),K(0),M(2),F(1),P(2),S(4), T(1),W(0),Y(2),V(2) Binding site category: OTH	
	Dscore: 0.82 Binding site type: Druggable Amino acid residues: A(2),R(2),N(6),D(1),B(0),C(0), E(0),Q(2),Z(0),G(6),H(3),I(5), L(5),K(1),M(1),F(5),P(3),S(2), T(7),W(0),Y(3),V(5) Binding site category: OTH	
	Dscore: 0.8 Binding site type: Druggable Amino acid residues: A(2),R(2),N(0),D(5),B(0),C(4), E(2),Q(1),Z(0),G(0),H(6),I(2), L(6),K(0),M(4),F(0),P(0),S(0), T(2),W(0),Y(2),V(0) Binding site category: ENZ	

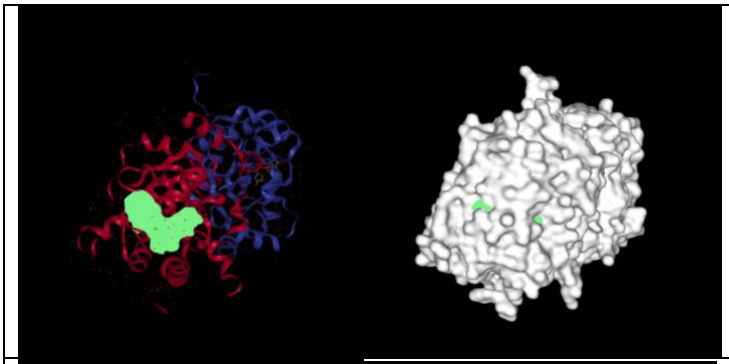
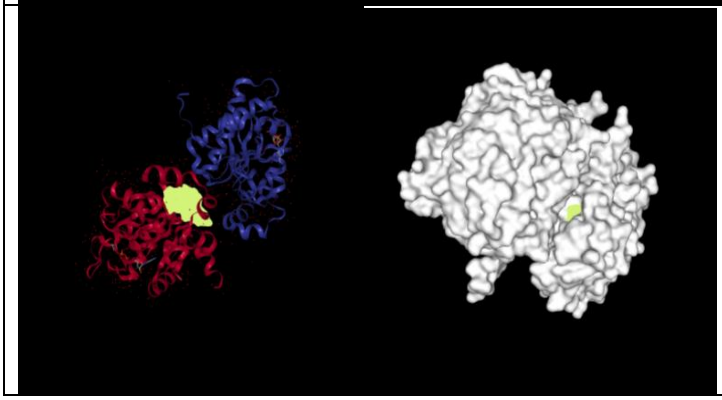
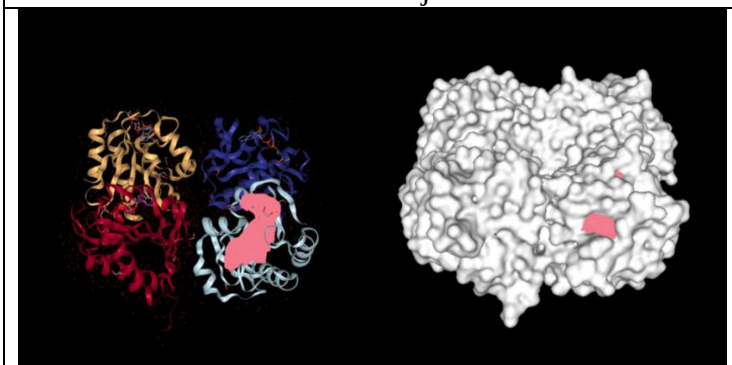
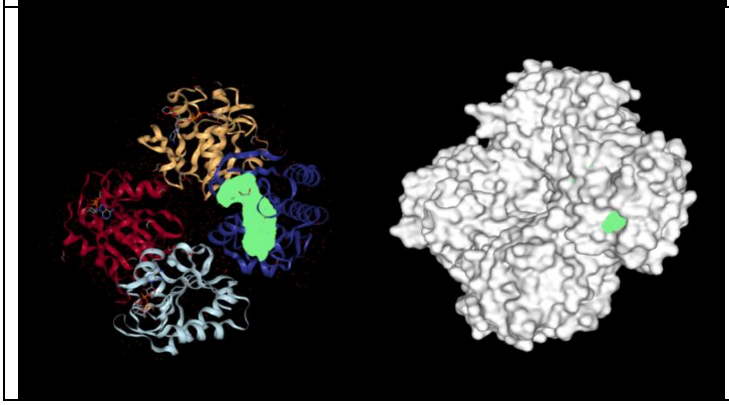
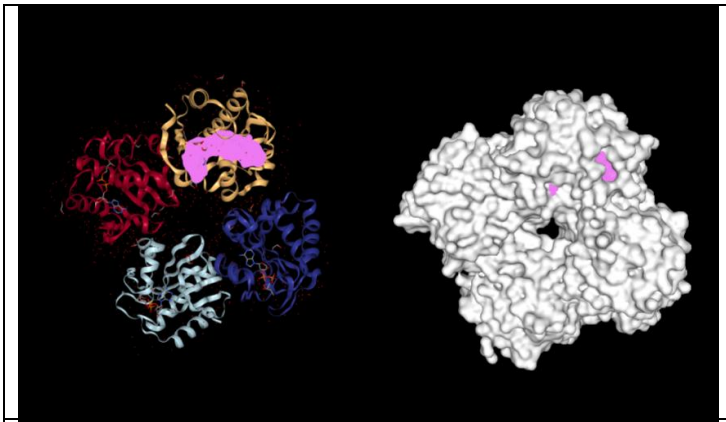
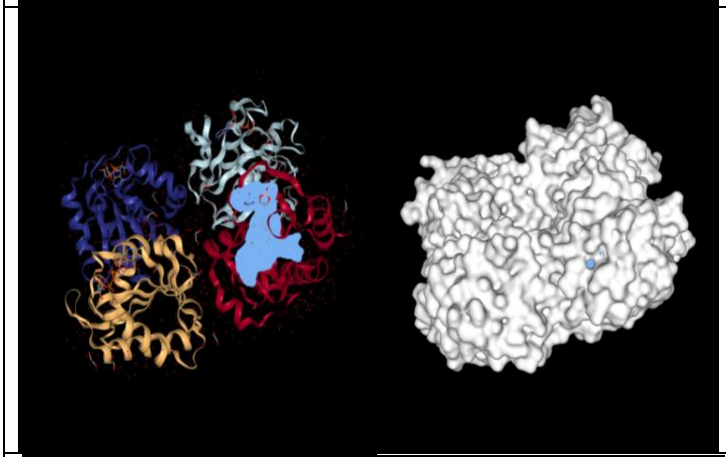
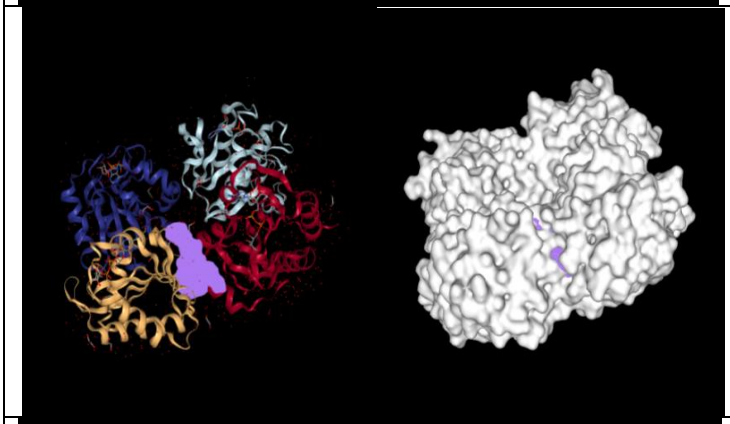
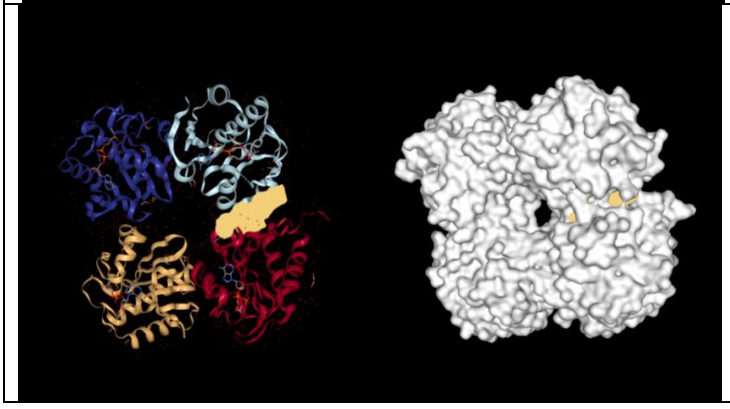
	Dscore: 0.74 Binding site type: Normal Amino acid residues: A(0),R(0),N(1),D(0),B(0),C(0), E(0),Q(1),Z(0),G(2),H(0),I(1), L(0),K(6),M(0),F(5),P(0),S(1), T(1),W(2),Y(1),V(1) Binding site category: OTH
	Dscore: 0.72 Binding site type: Normal Amino acid residues: A(1),R(1),N(0),D(0),B(0),C(1), E(2),Q(1),Z(0),G(2),H(1),I(0), L(5),K(0),M(0),F(1),P(1),S(1), T(1),W(1),Y(1),V(3) Binding site category: OTH

Table 63: Structural information for the protein o – acetyl – ADP – ribose deacetylase MACROD2

PDB id: 4j5s	Resolution: 1.55 Å	CSA: N/A
	Dscore: 0.84 Binding site type: Druggable Amino acid residues: A(0),R(8),N(2),D(0),B(0),C(0), E(4),Q(2),Z(0),G(4),H(0),I(2), L(1),K(0),M(0),F(0),P(0),S(6), T(2),W(1),Y(10),V(2) Binding site category: OTH	
	Dscore: 0.83 Binding site type: Druggable Amino acid residues: A(1),R(1),N(0),D(1),B(0),C(0), E(4),Q(2),Z(0),G(3),H(0),I(0), L(0),K(0),M(2),F(2),P(1),S(3), T(0),W(3),Y(2),V(1) Binding site category: OTH	

	Dscore: 0.83 Binding site type: Druggable Amino acid residues: A(7),R(3),N(2),D(1),B(0),C(0), E(2),Q(2),Z(0),G(5),H(0),I(3), L(11),K(2),M(1),F(2),P(3),S(3), T(2),W(0),Y(0),V(4) Binding site category: OTH
	Dscore: 0.82 Binding site type: Druggable Amino acid residues: A(2),R(4),N(1),D(6),B(0),C(5), E(0),Q(1),Z(0),G(3),H(9),I(2), L(5),K(0),M(1),F(0),P(2),S(2), T(3),W(0),Y(1),V(2) Binding site category: ENZ
	Dscore: 0.75 Binding site type: Normal Amino acid residues: A(1),R(2),N(1),D(4),B(0),C(5), E(1),Q(2),Z(0),G(1),H(7),I(2), L(1),K(0),M(1),F(1),P(0),S(4), T(2),W(1),Y(2),V(1) Binding site category:ENZ
	Dscore: 0.74 Binding site type: Normal Amino acid residues: A(4),R(2),N(2),D(0),B(0),C(0), E(1),Q(2),Z(0),G(1),H(1),I(0), L(1),K(4),M(1),F(4),P(2),S(0), T(0),W(1),Y(1),V(2) Binding site category: OTH

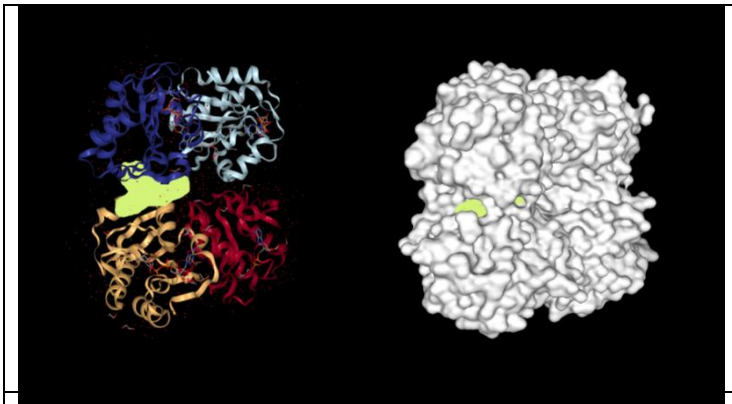
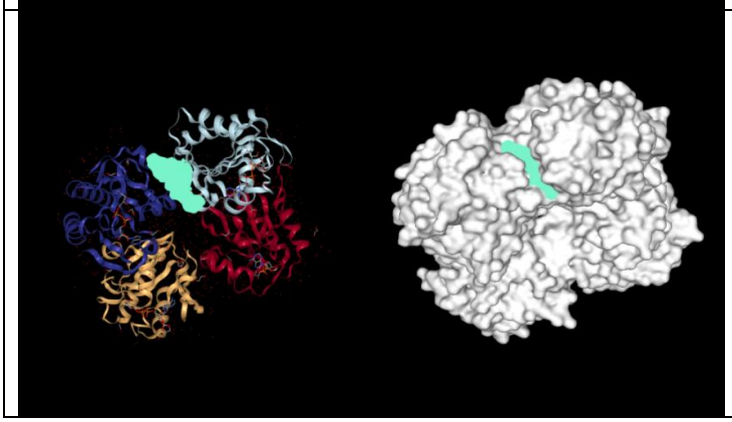
	Dscore: 0.72 Binding site type: Normal Amino acid residues: A(1),R(1),N(3),D(0),B(0),C(0), E(0),Q(3),Z(0),G(1),H(0),I(2), L(3),K(3),M(0),F(1),P(0),S(2), T(0),W(0),Y(1),V(0) Binding site category: OTH
	Dscore: 0.7 Binding site type: Normal Amino acid residues: A(8),R(1),N(0),D(1),B(0),C(0), E(1),Q(6),Z(0),G(5),H(0),I(7), L(2),K(1),M(0),F(1),P(9),S(2), T(5),W(0),Y(3),V(3) Binding site category: PPI

Table 64: Structural information for the protein o - acetyl –ADP - deacetylase 1

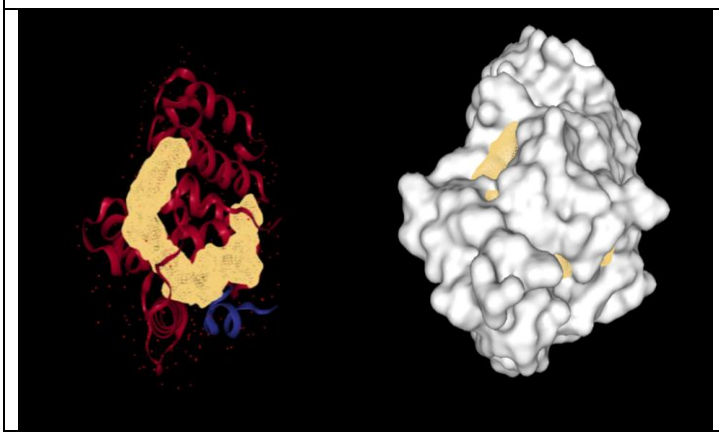
PDB id: 4j8s	Resolution: 1.55 Å	CSA: N/A
	Dscore: 0.82 Binding site type: Druggable Amino acid residues: A(2),R(1),N(1),D(1),B(0),C(1), E(2),Q(2),Z(0),G(3),H(1),I(2), L(2),K(2),M(4),F(2),P(0),S(1), T(2),W(0),Y(0),V(1) Binding site category: OTH	

Table 65: Structural information for the protein mRNA decay activator protein ZFP36

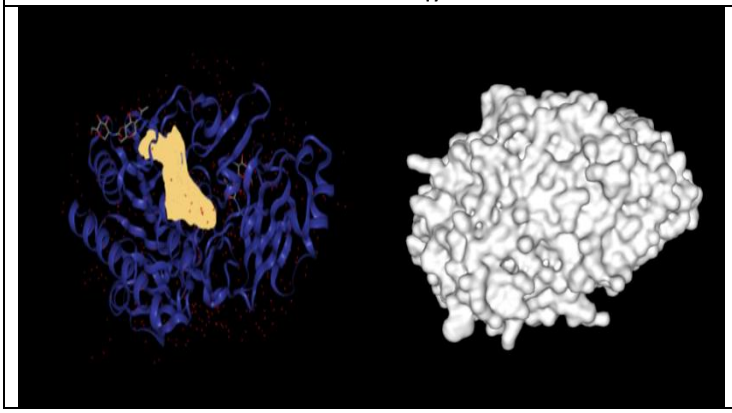
PDB id: 4lqy	Resolution: 1.54 Å	CSA: N/A
	Dscore: 0.85 Binding site type: Druggable Amino acid residues: A(4),R(4),N(1),D(7),B(0),C(5), E(2),Q(2),Z(0),G(3),H(7),I(2), L(4),K(2),M(4),F(2),P(0),S(1), T(2),W(0),Y(3),V(1) Binding site category: ENZ	

Table 66: Structural information for the protein Bis(5'- adenosyl)-triphophatase ENPP4

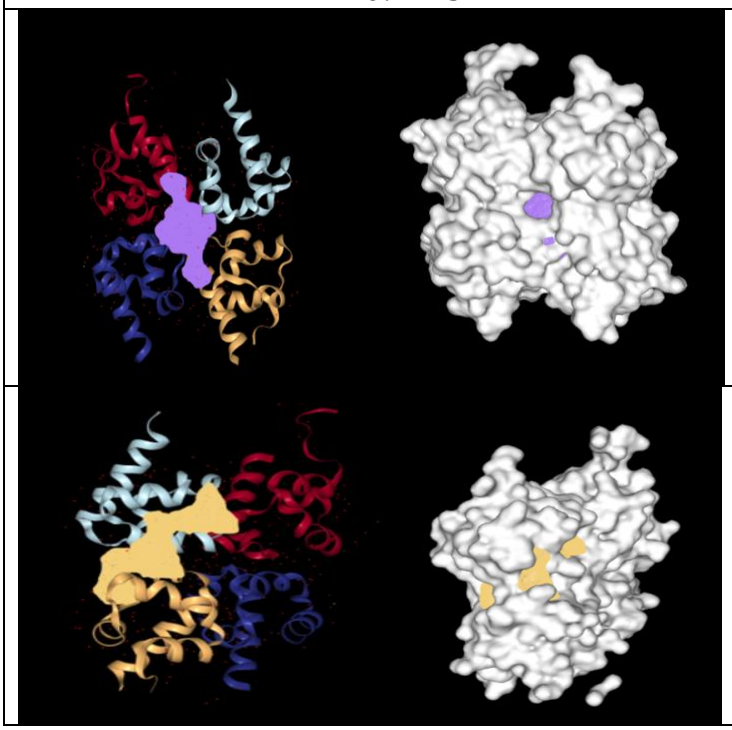
PDB id: 4nl9	Resolution: 1.5 Å	CSA: N/A
	Dscore: 0.86 Binding site type: Druggable Amino acid residues: A(2),R(2),N(3),D(1),B(0),C(1), E(2),Q(2),Z(0),G(3),H(0),I(2), L(4),K(2),M(4),F(2),P(0),S(1), T(2),W(0),Y(0),V(1) Binding site category: OTH	
	Dscore: 0.8 Binding site type: Druggable Amino acid residues: A(6),R(1),N(1),D(1),B(0),C(1), E(2),Q(2),Z(0),G(3),H(0),I(8), L(2),K(1),M(4),F(0),P(0),S(1), T(4),W(0),Y(0),V(1) Binding site category: PPI	

Table 67: Structural information for the protein Ankyrin repeat and SAM domain – containing protein 6

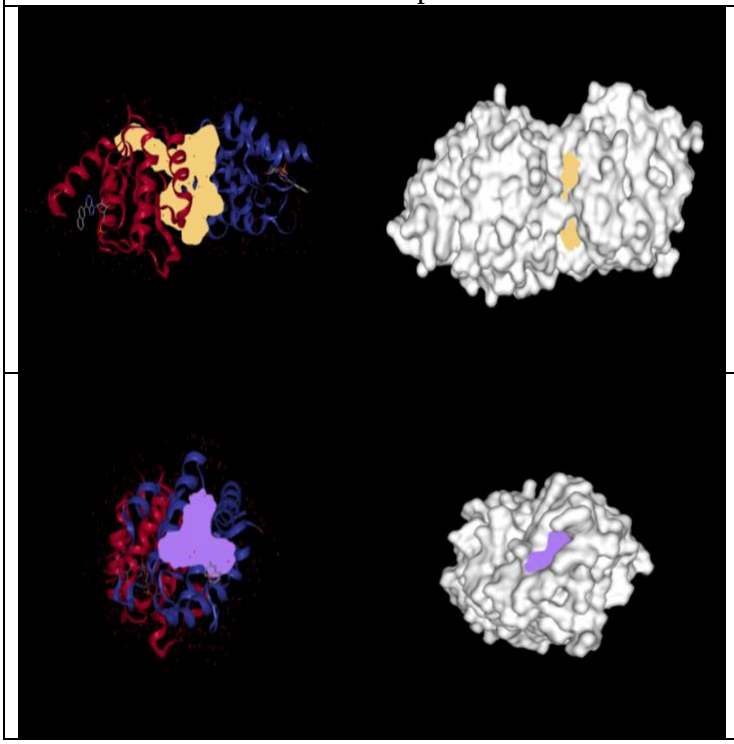
PDB id: 4p5e	Resolution: 1.35 Å	CSA: N/A
	Dscore: 0.8 Binding site type: Druggable Amino acid residues: A(2),R(2),N(2),D(1),B(0),C(1), E(4),Q(2),Z(0),G(3),H(0),I(2), L(2),K(2),M(0),F(2),P(0),S(1), T(2),W(0),Y(0),V(0) Binding site category: OTH	
	Dscore: 0.78 Binding site type: Normal Amino acid residues: A(1),R(10),N(1),D(12),B(0),C(8), E(2),Q(2),Z(0),G(3),H(8),I(2), L(2),K(2),M(4),F(2),P(0),S(0), T(2),W(0),Y(2),V(0) Binding site category: ENZ	

Table 68: Structural information for the protein 2'- deoxynucleoside 5'-phosphate N-hydrolase 1

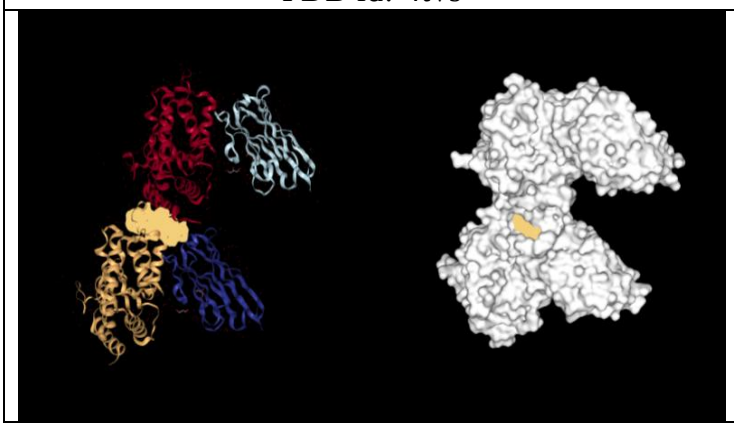
PDB id: 4tvs	Resolution: 1.6 Å	CSA: N/A
	Dscore: 0.81 Binding site type: Druggable Amino acid residues: A(10),R(2),N(1),D(2),B(0),C(1), E(0),Q(4),Z(0),G(0),H(0),I(5), L(2),K(4),M(0),F(2),P(1),S(0), T(9),W(0),Y(1),V(1) Binding site category: PPI	

Table 69: Structural information for the protein Torsin-1A-interacting protein 1

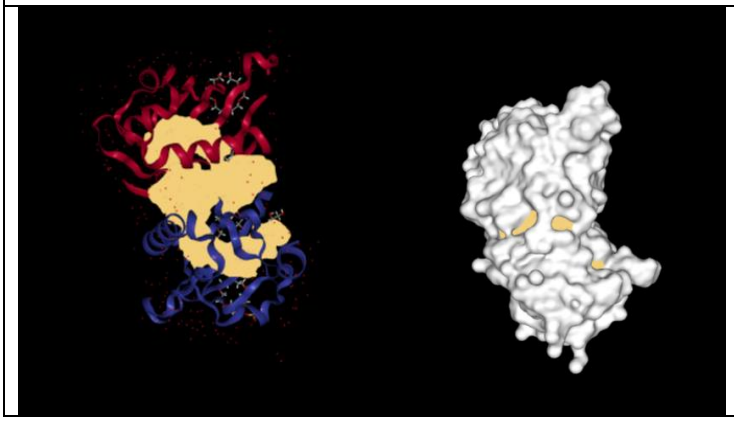
PDB id: 4wn5	Resolution: 1.15 Å	CSA: N/A
	Dscore: 0.81 Binding site type: Druggable Amino acid residues: A(2),R(2),N(3),D(1),B(0),C(1), E(0),Q(5),Z(0),G(3),H(0),I(4), L(3),K(5),M(2),F(2),P(5),S(2), T(3),W(0),Y(1),V(7) Binding site category: OTH	

Table 70: Structural information for the protein Hypoxia-inducible factor 3-alpha

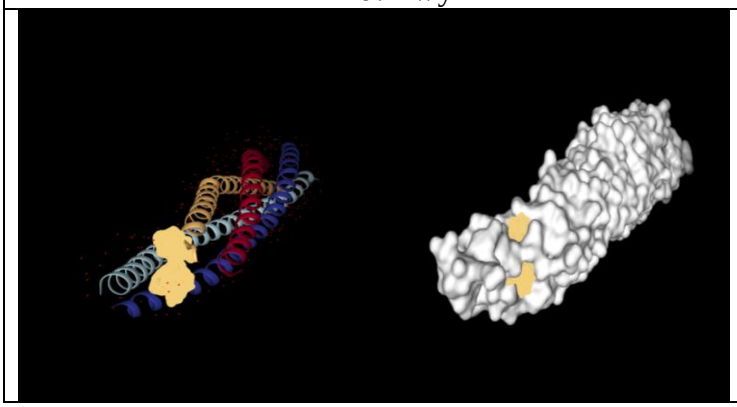
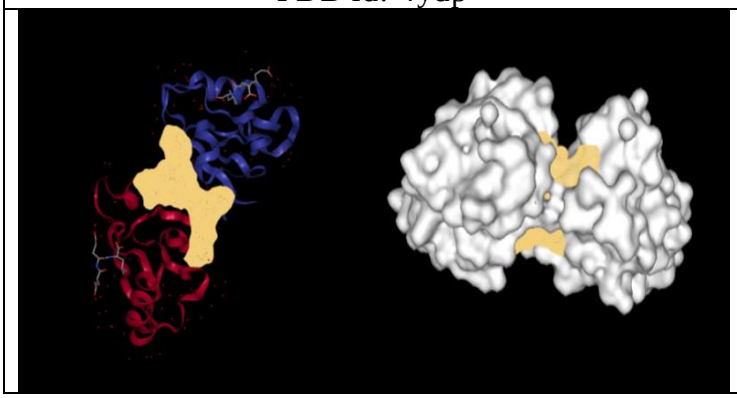
PDB id: 4wy4	Resolution: 1.4 Å	CSA: N/A
	Dscore: 0.74 Binding site type: Normal Amino acid residues: A(10),R(0),N(0),D(0),B(0),C(0), E(4),Q(1),Z(0),G(0),H(0),I(2), L(0),K(1),M(0),F(0),P(1),S(0), T(5),W(0),Y(1),V(2) Binding site category: PPI	

Table 71: Structural information for the protein Vesicle-associated membrane protein 8

PDB id: 4ydp	Resolution: 1.4 Å	CSA: N/A
	Dscore: 0.77 Binding site type: Normal Amino acid residues: A(1),R(1),N(0),D(0),B(0),C(1), E(4),Q(1),Z(0),G(0),H(1),I(2), L(0),K(3),M(0),F(0),P(1),S(0), T(0),W(0),Y(1),V(4) Binding site category: OTH	

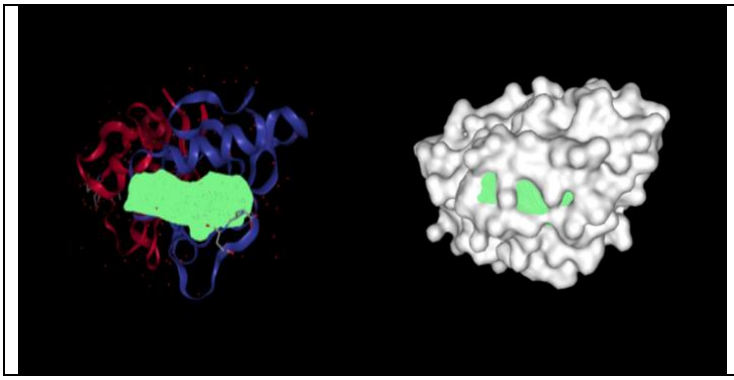
	Dscore: 0.7 Binding site type: Normal Amino acid residues: A(1),R(3),N(0),D(8),B(0),C(4), E(4),Q(1),Z(0),G(0),H(5),I(0), L(0),K(1),M(0),F(0),P(1),S(0), T(0),W(0),Y(1),V(1) Binding site category: ENZ

Table 72: Structural information for the protein LIM domain-binding protein 3

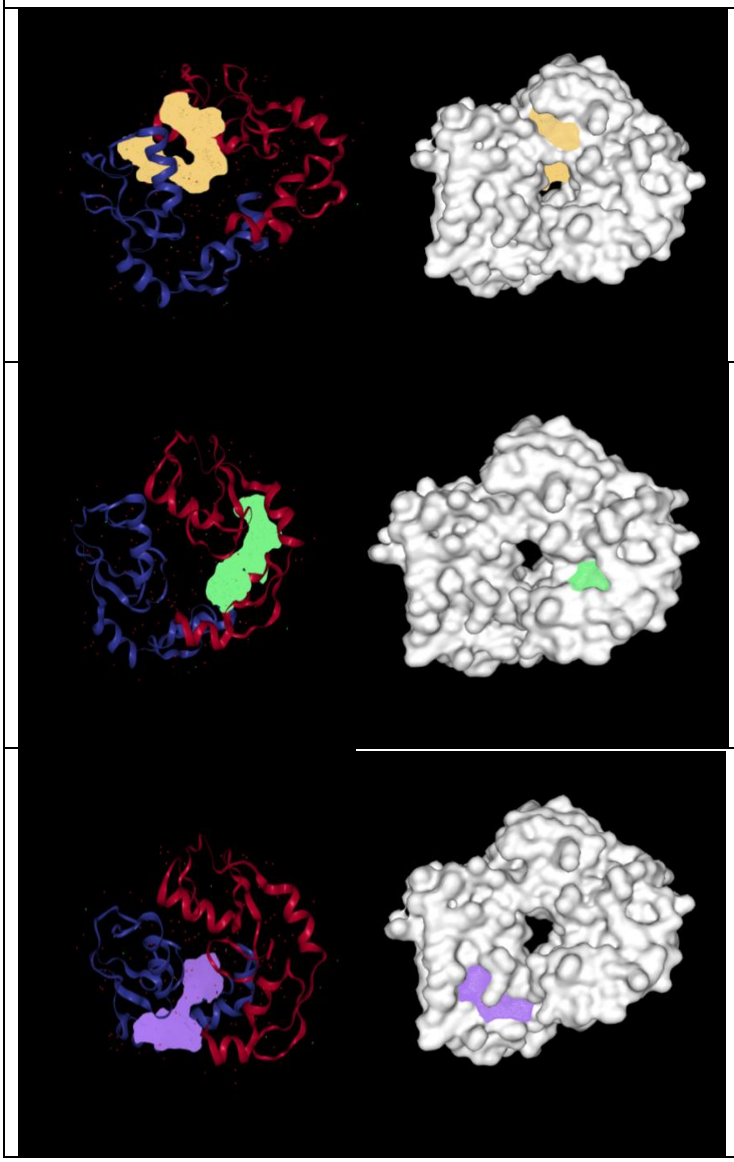
PDB id:5dka	Resolution: 1.55 Å	CSA: N/A
	Dscore: 0.77 Binding site type: Normal Amino acid residues: A(4),R(2),N(0),D(0),B(0),C(0), E(4),Q(1),Z(0),G(0),H(0),I(2), L(0),K(1),M(0),F(0),P(1),S(0), T(0),W(0),Y(1),V(2) Binding site category: OTH	
	Dscore: 0.77 Binding site type: Normal Amino acid residues: A(1),R(2),N(0),D(4),B(0),C(7), E(1),Q(6),Z(0),G(5),H(8),I(2), L(2),K(1),M(0),F(1),P(1),S(2), T(0),W(0),Y(3),V(3) Binding site category: ENZ	
	Dscore: 0.76 Binding site type: Normal Amino acid residues: A(0),R(1),N(0),D(0),B(0),C(0), E(2),Q(1),Z(0),G(2),H(0),I(0), L(5),K(0),M(0),F(1),P(1),S(1), T(1),W(1),Y(2),V(3) Binding site category: OTH	

Table 73: Structural information for the protein E3-ubiquitin-protein ligase RNF125

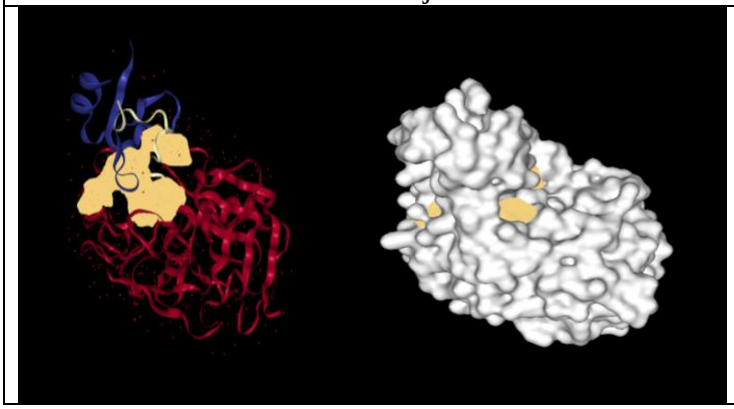
PDB id: 5jbt	Resolution: 1.4 Å	CSA: N/A
	Dscore: 0.73 Binding site type: Normal Amino acid residues: A(1),R(4),N(0),D(7),B(0),C(5), E(2),Q(1),Z(0),G(2),H(10),I(0), L(0),K(0),M(0),F(2),P(1),S(1), T(1),W(3),Y(1),V(3) Binding site category: ENZ	

Table 74: Structural information for the protein trypsin-3

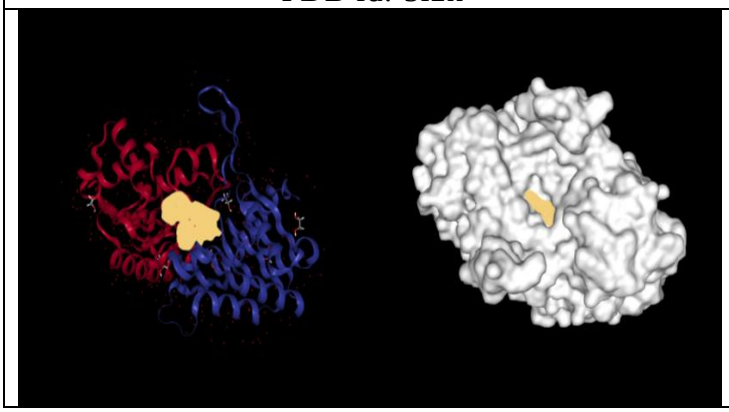
PDB id: 5lzk	Resolution: 1.575Å	CSA: N/A
	Dscore: 0.74 Binding site type: Normal Amino acid residues: A(10),R(0),N(1),D(1),B(0),C(0), E(1),Q(2),Z(0),G(1),H(0),I(2), L(1),K(0),M(1),F(1),P(0),S(4), T(4),W(1),Y(2),V(1) Binding site category: PPI	

Table 75: Structural information for the protein FAM83B

3.9. Distribution of Binding pockets

Total number of Proteins (Seq similarity ≥98%)	Proteins with structures (0≤Resolution≤1.6)	Proteins with binding sites	Proteins with druggable binding sites	No.of druggable binding sites	Binding site type		
					ENZ	PPI	OTH
4970	125	71	45	76	61	28	55

Table 76: Distribution of Protein Structures and Binding Sites (Druggable/Normal) among Breast cancer

CHAPTER 4 – CONCLUSION

The sequencing of genome has provided various opportunities and access to next generation resources for the development of cancer therapies. Our current study focuses on the high throughput screening of breast cancer specific binding sites and their categorisation for possible evaluation to evaluate and pave an ease towards novel drug target identification.While Biological methods like CRIPSR/Cas9 and siRNA convenient tools explore the role of potential drug targets and chemical tools provide a global view of approach to interrogate new targets. Our study revealed various properties associated with binding sites which might result in candidate biomarker screening. 4970 proteins (≥98 identity) considered for analysis; out of which 125 proteins having structures with resolution in range 0-1.6 Å, followed by 71 proteins in binding site category, 45 proteins with 76 druggable binding sites. Binding sites were further classified into Enzymatic (61), protein-protein interactions (28) and other (55). Our analysis can be carried further by the breast cancer scientific community performing molecular screening and experimental variation on various cell lines and help in customized or personalised therapy.

CHAPTER 5 – REFERENCES

1. What Is Cancer? National Cancer Institute. 2017. Available from: https://www.cancer.gov/about cancer/ understanding/ what-is-cancer.
2. Cancer World Health Organization. 2017. Available from: http://www.who.int/mediacentre/factsheets/fs297/en/
3. Breast Cancer Signs and Symptoms. Cancer.org. 2017. Available from: https://www.cancer.org/cancer/breast-cancer/about/breast-cancer-signs-and-symptoms.html
4. Symptoms of breast cancer - Canadian Cancer Society. www.cancer.ca. 2017. Available from: http://www.cancer.ca/en/cancer-information/cancer-type/breast/signs-and-symptoms/?region=bc
5. Wang, Zhong, Mark Gerstein, and Michael Snyder. "RNA-Seq: A Revolutionary Tool For Transcriptomics". Nat Rev Genet 10.1 2009; 57-63.
6. Nagalakshmi U, Waern K, Snyder M. RNA-Seq: a method for comprehensive transcriptome analysis. Current Protocols in Molecular Biology. 2010 Jan 13:4-11.
7. Labute P, Santavy M. Locating binding sites in protein structures. Journal of Chemical Computing Group. 2007.
8. Binding site. En.wikipedia.org. 2017. Available from: https:// en.wikipedia.org/ wiki/Binding_site.
9. Xu D, Jalal SI, Sledge GW, Meroueh SO. Small-molecule binding sites to explore protein–protein interactions in the cancer proteome. Molecular BioSystems. 2016;12(10):3067-87.
10. Kodama Y, Shumway M, Leinonen R. The Sequence Read Archive: explosive growth of sequencing data. Nucleic acids research. 2012 Jan 1;40(D1):D54-6.
11. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic acids research. 2002 Jan 1;30(1):207-10.
12. UniProt Consortium. The universal protein resource (UniProt). Nucleic acids research. 2008 Jan 1;36(suppl 1):D190-5.
13. Bernstein FC, Koetzle TF, Williams GJ, Meyer EF, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. The protein data bank. European Journal of Biochemistry. 1977 Nov 1;80(2):319-24.
14. Schomburg I, Chang A, Ebeling C, Gremse M, Heldt C, Huhn G, Schomburg D. BRENDA, the enzyme database: updates and major new developments. Nucleic acids research. 2004 Jan 1;32(suppl 1):D431-3.
15. Porter CT, Bartlett GJ, Thornton JM. The Catalytic Site Atlas: a resource of catalytic sites and residues identified in enzymes using structural data. Nucleic acids research. 2004 Jan 1;32(suppl 1):D129-33.
16. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. Bioinformatics. 2011 Mar 15;27(6):863-4.
17. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics. 2009 May 1;25(9):1105-11.

18. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature protocols. 2012 Mar 1;7(3):562-78.
19. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 2006 Jul 1;22(13):1658-9.
20. Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL. NCBI BLAST: a better web interface. Nucleic acids research. 2008 Jul 1;36(suppl 2):W5-9.
21. Volkamer A, Kuhn D, Rippmann F, Rarey M. DoGSiteScorer: a web server for automatic binding site prediction, analysis and druggability assessment. Bioinformatics. 2012 Aug 1;28(15):2074-5.