STRUCTURAL ANALYSIS OF PTEN TUMOR SUPPRESSOR ENZYME: UNLOCKING THE CLOSED DOORS

TAMIM HUSSEIN
DR. JOSEPH REBEHMED
MAY 2, 2024
CAPSTONE PROJECT

Table of Contents

Abstract

This study presents a comprehensive structural study of PTEN (Phosphatase and tensin homolog), a critical tumor suppressor enzyme known for its dual phosphatase activity. The study aims to elucidate the molecular architecture of PTEN, focusing on its phosphatase domain, and to understand the enzyme's catalytic site and the corresponding enzymatic activity. Utilizing advanced structural biology and bioinformatics tools, we have identified key residues involved in PTEN's enzymatic activity, including Cys124, Arg130, and His93, Lys125, Lys128, and Val45, which are crucial for substrate binding and catalysis. This study provide insights into residues that could have role in the hydrolysis (dephosphorylation) reaction caried out by PTEN. Our findings reveal the structural similarities between PTEN and other proteins, such as Tensin and auxilin, suggesting an evolutionary conservation of these structural features. This paper was able to provide a docking structure that represent PTEN protein bound to the IP3 ligand and validate the structure through different tools. Despite these advances, several aspects of PTEN's structure and function remain unresolved, including the exact nature of its dimeric structure and the roles of its various hotspot residues. This study underscores the complexity of PTEN's regulation and the need for continued research to fully understand its role in cellular processes and disease, particularly cancer. The implications of our findings for the development of targeted therapies for cancer are discussed, highlighting the potential of PTEN as a therapeutic target.

Introduction

Cancers are genetic disorders that are driven by mutations on the level of germ-line genes or somatic cells (Liu Y. et al., 2015). Treating cancer cells is achieved by targeting two main genes: Oncogenes and Tumor Suppressor Genes (Weinberg R. et. al, 1991). In the field of targeted therapies against specific proteins involved in cancer, several advancements were made in the gain-of-function mutations, in Oncogenes (Liu Y. et al., 2015). However, challenges arise in targeting the loss-of-function mutations in the tumor suppressor genes using standard approaches of pharmacology (Liu Y. et al., 2015). Recent advancements in screening and computational resources have paved the way to getting more insights on the location of mutations, mode of action, and resistance to cancer therapy of tumor suppressor genes and their product proteins (Lee J et. Al, 2024).

Several tumor suppressor genes were candidates for research and structural analysis since the identification of tumor suppressor genes. Lung cancer studies identified PTEN, TP53, and FHIT genes as frequently mutated proteins in this cancer type (Lee J et. Al, 2024). Also, Breast cancer studies identified BRCA1 and BRCA2 genes, PTEN gene, and P53 whose mutation is a master driver force for tumor growth and mutagenesis (Liu Y. et al., 2015). In fact, it is thought that about 80% of mutations linked to cancer that have been found are in tumor suppressor genes (Liu Y. et al., 2015).

These tumor suppressor genes are not separate entities. On the contrary, they share common pathways and are part of a complex network of interactions that can influence the development and progression of tumors (L.H. et. al, 2010). P-53 which is a tumor suppressor protein that has received huge attention in cancer, as several human tumors have shown some mutation or modification in p53 function and activity (see Agarwal *et al.* 1998; Braithwaite *et al.* 2005 as cited in L.H. et al., 2010). Tumor suppressor protein p53 is involved in the regulation of cellular responses to stress by mediating toxicant-induced renal cell death. Additionally, it activates the cell cycle inhibitor protein p21, which prevents cell death (L.H. et al., 2010). In his article, Pessoa (2019), mentions some interactions among tumor suppressor genes, the PTEN and CDKN2A Interactions: A strong association

was noted for the co-occurrence of PTEN and CDKN2A deletions, suggesting that these two tumor suppressor genes often alter together. This co-occurrence was more frequently detected in gliomas, indicating its importance in the development and progression of these tumors (Pessoa A. et al., 2019).

PTEN, is a phosphatidylinositol-3,4,5-triphosphate (PIP3) phospholipid phosphatase (Dempsey R., 2021). It is a tumor suppressor gene that produces a 403-amino acid polypeptide, a dual lipid/protein phosphatase, which is structurally complex protein and consists of several domains and substructures that contribute to its function (Masson, G. R., 2020). The primary components of PTEN's structure include: the DUSP domain (15–185), which is crucial for PTEN's phosphatase activity containing the protein's active site having three core loops (TI, WPD, and P loops) that are essential for its function, including an arginine loop that plays a key role in membrane binding and the C2 Domain (192–353), which is involved in the calcium-independent phosphatase activity of PTEN (Masson, G. R., 2020). The C2 domain contains calcium-binding region (CBR) loops that are important for its function (Masson, G. R., 2020). It is also involved in the regulation of PTEN's activity, including its interaction with the membrane and its allosteric regulation (Masson, G. R., 2020). Besides, PTEN also have 3 disordered loops that are important for protein interaction with membrane and with other proteins, these loops are located between the DUSP and the C2 domains (Masson, G. R., 2020). The protein terminates with a in a 50 amino acid disordered C-terminal tail (Dempsey R., 2021 & Masson, G. R., 2020). Upon having its natural substrate (PIP3) embedded in the plasma membrane, PTEN poses those crucial membrane binding properties (Dempsey R., 2021).

This phosphatase and tensin homolog deleted in chromosome 10 (PTEN) is thought to be necessary for tumor suppressor properties because of its ability to negatively regulate Phosphoinositide 3-kinase (PI3K) and Akt/mTOR signaling (Pessoa A. et al., 2019). As a result, it plays a significant role in the control of events such as cell cycle arrest, apoptosis, angiogenesis, adhesion and invasion, and DNA damage repair (Pessoa A. et al., 2019). PTEN is essential for controlling cell growth and metabolic pathways in both healthy and pathological conditions (Dempsey R., 2021). It catalyzes the conversion of phospholipid phosphatidylinositol-3,4,5-trophosphate (PIP3) to PIP2 (Dempsey R., 2021). As one of the main tumor suppressor genes, PTEN frequently has loss-of-function mutations in a variety of cancer types (Dempsey R., 2021). PTEN acts as a brake on the

AKT/mTOR cell proliferation/antiapoptotic signaling pathway, which is activated in response to elevated PIP3 levels (Dempsey R., 2021 & Lee J et. al., 2024). In addition, PTEN acts as regulator of PI3K, which catalyzes the reverse reaction of PTEN, PIP2 to PIP3 (Costa A. et al., 2015).

Along with p53, sporadic somatic mutations in PTEN render this tumor suppressor gene one of the two most commonly mutated in human malignancies, particularly in solid tumors (Isabel E. et al., 2011). In fact, many studies have found out that PTEN hamartoma tumor-related syndromes (PHTSs), also known as inherited hamartoma, Cowden syndrome, and associated conditions, are caused by germline mutations in the PTEN gene (Orloff S., 2008 & Waite K.A., 2002 as cited by Isabel E. et al., 2011). Besides its clear involvement in PHTS, germline PTEN mutations have been found in patients with neurodevelopmental disorders, such as autism spectrum disorders (ASDs) and developmental delay (DD) (Isabel E. et al., 2011). As studies have shown that lipid phosphatase activity of PTEN is necessary for its tumor suppressor function, insights on the specific mutations in the active site might have direct implications in tumor suppressor activity reactivation and cancer treatment (Han S., 2000). Even though several studies have described many loss-of-function mutations leading to impaired phosphatase activity on the level of catalytic site, few have discussed the structural implications of the observed mutations on the active site as well as their natural occurrence in evolutionary related proteins.

Although phosphatase-independent functions of PTEN have also been described to contribute to its tumor suppressive activity as mentioned by Leslie R. et al. (2011), buffering PI3K signaling in virtue of its catalytic activity is assumed to be its key role. Thus, the precise dissection of PTEN catalytic properties is highly relevant for human disease (Isabel E. et al., 2011). The analysis of active site of crystal structure of PTEN (Lee O., 1999), reveals its basic pocket structure, which is predicted to be as is to accommodate the negative charge of PIP3 phosphate atoms. It was also evident the presence of the phosphatase signature motif ([123] HCKAGKGR [130]) of the P-loop, located in the bottom of the active site pocket, along with the WPD loop and the TI loop that were previously described. Studies working with PTEN identified several mutations associated with the active site and the P-loop in particular, many of which gave a truncated non-functional protein with no catalytic activity. This gave insights into specific locations within the active site that are tolerant to mutations and other locations where mutations are non-tolerated. This study presents a structural analysis of the PTEN's catalytic

site, investigating how mutations and variants in key binding residues, affect the interaction with its substrate, PIP3, and subsequent enzymatic activity.

Methods

Evolutionary study tools and materials

We used the NCBI BLASTP tool to search the SwissProt database against the query sequence of PTEN_HUMAN protein obtained from UniProtKb database under the UniProt ID, P60484. The specified database for the algorithm to search through was UniProtKb/SwissProt database. The search was set to search for the top 250 hits at an E-value no more than 0.001 to ensure accuracy and eliminate similarity by chance. BLASTP works by breaking the query sequence into 3-letter words, where each word's score is calculated based on amino acid values from the BLOSUM62 matrix. Words with scores of 12 or more (indicating highly conserved amino acids) form the initial search set. Then, synonyms (similar words) are added to the search set if their scores exceed the threshold value. After that, BLASTP scans the database, identifying protein sequences containing two or more words/synonyms from the search set, short matches serve as seeds for extended alignments in both directions. Raw scores are converted into bit scores, considering scoring matrices and database size. The obtained results were cleaned by removing sequences larger than 600 amino acids and less than 200 to ensure accurate covering of the protein query. Multiple Sequence Alignment (MSA) was carried out in MUSCLE tool. MUSCLE uses the iterative method, a dynamic programming approach, to find the best alignment between each pair of sequences. It starts by producing a low-quality alignment progressively and gradually improves it by iterative realignment until no more improvements in the alignment scores. This is done by first finding the two sequences with the highest pairwise similarity and align them using standard pairwise dynamic programming alignment. Then the algorithm finds the sequence that is most similar to a profile of the alignment of the first two, and align it to the first two by profile sequence alignment. This is repeated several times until all sequences have been included in the multiple alignment. The next step is to remove sequence x1and realign it to a profile of the other aligned sequences x2…xN by profile-sequence alignment. The last step

involves repeating the previous realignment step a fixed number of times, or until the alignment score converges.

Following the alignment by MUSCLE, we took the Phylogenetic tree generated as a result of the MSA job. The tree (taken is used) in ITOL tool to be visualized and to annotate specific characteristics in regard to PTEN_HUMAN protein. The MSA sequences were downloaded as FASTA (aligned) and put in WebLogo for visualization purposes where this tool provides a graph displaying the relative frequency of residues present in each position of the protein to use it for conservation analysis. The same strategy of sequence alignment cleaning and downloading was used with PTEN proteins in different species to study the conservation of PTEN in different species.

Processing of the ligand and receptor (Pre-Docking step)

The receptor protein chosen had the PDB id: 1D5R. The choice of this PDB structure was based on the idea that it is a monomeric structure having the catalytic site and only missing part of the C-terminal specifically the region between residue 281 and 313, also it had the TLA ligand bound to the catalytic site and this protein structure was used by several literature papers. This study uses the recommendations of the paper by Samia Aci-Sèche et al. (2023) of docking preparation and working. First, the receptor protein was transferred to chimera X to remove the ligand TLA which was interfering with the active site. Also, the Alfa-Fold prediction tool was used to predict the missing structure from the C-terminus. Alfa-Fold tool iteratively searches the Alfa-fold database for structures, then it assigns base on similarity the best model present. Alpha-fold structure details are present in figure 1.

AlphaFold prediction matching modified_Protein.pdb						
Chain	**UniProt Id**	**UniProt Name**	**RMSD**	**Length**	**Seen**	**% Id**
A	A0A024QYR5	A0A024QYR5_PANPA	1.59	338	307	100

Figure1 from Chimera X, represents the details of the fetched structure by Alpha-Fold used during the preparation of docking receptor protein.

This protein was transferred to AutoDock tools to be further processed. The processing of this protein included: adding polar hydrogens to ensure the accuracy and clarity of the active site, water molecules were removed, and

Kolman charges were added. The method for calculating Kolman charges involves solving a linear equation for the partial charges, which is derived from the constraints and restraints applied to the molecule. These constraints and restraints can include total charge, electrostatic moments, electrostatic potential, or equality of certain charges resulting from symmetry. The calculation process involves the use of matrices and vectors to find the optimal point charges that minimize the quantity of the restraints or constraints (Sigfridsson, E. & Ryde, U., 1998). Finally, the ligand position in the receptor was approximated using the grid box service. The ligand (IP3) used was taken from the instance coordinates of the structural homologous structure VSP phosphatase (PDB id: 3V0H). BLAST was performed on the structure of the modified protein using Chimera X. The matrix used was BLOSSUM62, the cutoff significance was 1e (-3), and the query was performed on the structures of the PDB database. The results are present in the supplementary files and documents under BLASTChimeraX.csv. The hit 37, A, B: Voltage-sensor containing phosphatase, having 339 amino acid residues, is a structure obtained by X-RAY DIFFRACTION with a resolution of 1.89Å. The BLAST score of this structure was 183, making it one of the most similar protein structures to our protein. The ligand associated with this structure is the IP3, which was then used for the docking process. The ligand was prepared to docking by adding charges. IP3 has three phosphate groups each of which depends on protonation state which might bear a charge 0, -1 or -2. Therefore, IP3 might exist in several ionization states with different levels and arrangements of protonation on the phosphates. NMR measurements gave an estimate of the PI (3,4,5) P3 overall charge of -5.05 at physiological pH, which was used here. The receptor and ligand were saved as ". pdbqt" extension.

Docking and Visualization Tools and methodology

Docking was performed via the Autodock vina tool, a tool which takes as input the ligand and the receptor proteins. In the vina directory a file called "output.pdqt" was created, in this file the output ligand atomic details will be written. Also, a file called "log.txt" file was created where the log of the docking process was stored in it. The last file was called "configuration.txt" file which had the actions shown in figure 2 to be used in the docking command later.

```
receptor = modified_Protein.pdbqt
ligand = I3P.pdbqt

center_x = 35.6
center_y = 88.8
center_z = 27.8

size_x = 18
size_y = 18
size_z = 22

exhaustiveness = 8
```

Figure 2

The receptor keyword specifies the input receptor protein, ligand keyword takes the input ligand where both ligand an receptor should be within the directory of the executable vina program. The following center_x, center_y, and center_z keywords represent the position of the ligand within the receptor. The size_x, size_y, and size_z keywords represent the grid size in Å (angstrom) of the potential position of the ligand in the active site. Finally, the exhaustiveness parameter is a value that increases accuracy in the prediction of the location of the ligand by consuming more CPU power to try different possibilities and positions. AutoDock Vina is a program designed for molecular docking and virtual screening, aiming to improve upon the speed and accuracy of previous versions like AutoDock 4. It achieves notable gains in accuracy and speed, providing a speed increase of about two folds over AutoDock 4 and improving binding mode predictions' accuracy. A new scoring function, effective optimization, and multithreading capabilities on multi-core machines are some of the factors contributing to this improvement. The general functional form of the conformation-dependent part of the scoring function AutoDock Vina is designed to work with is The scoring function in AutoDock Vina is designed to work with a conformation-dependent part, denoted as (c), which is a sum of intermolecular and intramolecular contributions. The AutoDock Vina tool is characterized with its minimal rectangular parallelepiped, which is used to determine the search space of the ligand. This rectangular structure is increased in size with each iteration to have a whole evaluation of the active possible sites of the ligand. It is decided on the 22.5Å to be the final increase in size and is followed by random assigning of 5Å in all directions of the ligand search space. Finally, RMSD is used to assess the structures.

To further validate our results, another docking tool, HDOCK was used to perform docking where the same protein prepared for AutoDock was used along with the same ligand. This docking tool computes models through a

comprehensive workflow that integrates several steps (Yan Y. et al., 2017). It starts from the input structure where it does sequence similarity search against the PDB sequence database to find homologous sequences for both receptor and ligand molecules (Yan Y. et al., 2017). This step uses the HHSuite package for proteins. If a complex template is found through the sequence similarity search, HDOCK performs template-based docking. Otherwise, it proceeds with free docking using the input structures. HDOCK uses HDOCKlite, a hierarchical FFT-based docking program, to globally sample putative binding orientations (Yan Y. et al., 2017). This step incorporates the binding site information provided by the user. Finally, the models are made available for the user. The first trials of docking using HDOCK resulted in 103 models of which only 5 made it to the correct binding catalytic site. To ensure the tool was working correctly the processed experimental structure of the protein PTEN (PDB ID: 1D5R) was provided without the experimental ligand bound. Then the ligand was bound as separate file and docking was performed. The result was perfect match between the experimental and docked structure. To further ensure the docking tool gave us accurate results, the binding site residues were specified and the docking was repeated. The docking trials used until this point had the option Template-free docking only option selected. This option by default, when not selected, allows HDOCK to conduct the hybrid protocol of template-based modelling and free docking. The last trial we tend to deselect this option, while still providing the active site, to see whether the algorithm choice of ligand positioning would still support the results previously obtained without using PDB database templates.

Chimera X tool was used to visualize the molecules of PTEN generated by docking servers. This tool is an atomic building tool that takes is able to interpret low resolution structures through the integration of machine-learning structure predictions, provide likelihood-based fitting in maps, and compute per-residue scores to identify modeling errors. Chimera X is capable of generating 3D models of biological structures with unprecedented detail due to the advancements in cryoelectronic (cryo-ET) data along with structure prediction tools and other techniques. Different properties were used to do the structural analysis including: Hydrogen bonding, contacts, and clashes. The default parameters were used to find the hydrogen bonding network of the ligand and the residues of the active site. Hydrogen bonds are found based on this “new” crystal survey provided by the Cambridge Structural Database where they extracted the 3D distribution of the hydrogen bonds around the hydrogen bonding groups and calculated the propensity of that group to form hydrogen bond, which allows this technique to differ from other techniques that usually consider only the individual geometric parameters (Mills E., 1996). Contacts, which are all kinds of direct interactions: polar and nonpolar, favorable and unfavorable (including clashes). Clashes, on the other hand, are unfavorable interactions where atoms are too close together. Clashes

and contacts depend on the atomic VDW (van der Waals) radii which are used in sphere and ball displays. VDW radii also compute surface areas and molecular surfaces, as well as identify collisions and contacts. A change in an atom's atom type or number of attached hydrogens may also have an impact on its VDW radius. The values are kept as the radius atom attribute. In Clashes and contacts, the pairs of atoms are compared against this formula:

$$\textbf{VDW overlap} \geq [\, \boldsymbol{cutoff} \,] \text{ Å}$$

The ***VDW overlap*** between two atoms is defined as the sum of their VDW radii minus the distance between their centers. An overlap of 0 means the VDW surfaces are just touching. A positive overlap is how far the VDW surfaces interpenetrate, whereas a negative overlap is the distance of separation between the VDW surfaces. A larger positive *cutoff* restricts the results to more severe clashes, whereas a negative *cutoff* can also identify pairs of atoms that are farther apart. In Clashes only, after subtracting [*allowance*] for H-bonding, that is before using the previous formula the VDW radii overlapping value undergoes:

$$\text{VDW overlap} - \text{an } allowance > 0$$

reflects the observation that atoms sharing a hydrogen bond can come closer to each other than would be expected from their VDW radii. The *allowance* is only subtracted for pairs comprised of a donor (or donor-borne hydrogen) and an acceptor (Li J et al., 1998). The allowance value by default is 0.4 and is usually between 0.2 and 0.6.

In this study we mutated several residues by the mutate residue tool present in Chimera X. The tool provides several rotamers variations including: Richardson (common atom), Dunbrack, Dynameomics, and Richardson (mode), each is associated with a set of amino acids that are defined and have specific characteristics. In this study we used the Richardson (common atom) rotamer determinator, this tool allows the substitution of the selected amino acid by any of the 20 standard amino acids present. This rotamer library is capable of covering 94.5% of examples in the highest quality protein data with 153 rotamers and can make a significant contribution to improving the accuracy of new structures (Lovell C., 2000). Lovell C. et al. (2000) states that the accuracy of this rotamer library is that it employes more comprehensive methodologies and high-resolution data, resulting in improved rotamer libraries that are less prone to perpetuating inaccuracies. Depending on each residue properties such as, bonding and interactions, along with the number of chi angles, its corresponding rotamer properties are predicted.

Binding free energy calculations

The ligand binding free energy, ΔG, estimation was done in PRODIGY tool. This tool takes as input a PDB file, protein chain, and the ligand chain and generates a table of binding energies. This tool was previously used for protein-protein binding energy prediction of the interface. However, Anna V. et al. (2019), presented the new technique called PRODIGY-LIG, which allows protein-ligand binding energy calculation. It demonstrates that the $\Delta G_{binding}$ can be described by the number and type of interfacial residue–residue contacts in combination with properties of the non-interacting surface (Kastritis et al., 2014 as cited in Anna V. et al., 2019). This tool is trained to be able to generate the $\Delta G_{noelect}$ by only making use of structural terms (i.e. Acs, atomic contacts).

The other tool utilized was PISA, which allows the determination of binding energies by examining the surface area change between non-bound and bound/buried surface area. The VEP tool uses a comprehensive database of genomic variation and annotation data to predict the functional consequences of variants. This includes predicting whether a variant is likely to be damaging, neutral, or uncertain (Herrero, J. et al., 2016).

The tool integrates data from multiple sources, including Ensembl's own databases and external databases, to provide a comprehensive view of the potential effects of variants. The defined parameters used in this tool are HSDC, indicates residues that contain the across-interface hydrogen bond, salt bridge or disulfide bond atoms, ASA, indicates the solvent-accessible surface area of the corresponding residue, in $Å^2$. The next parameters are BSA, indicates Buried Surface Area, $Å^2$, and the ΔG Solvation energy effect, kcal/mol. The solvation energy gain of the interface is calculated as difference in solvation energies of all residues between dissociated and associated (interfacing) structures. Therefore, positive solvation energy ΔG of a residue makes a negative contribution to the solvation energy gain of the interface, which corresponds to hydrophobic effect. Solvation energy estimates in PISA do not include the effect of satisfied hydrogen bonds and salt bridges across the interface. The effect of hydrogen bonds (-0.44 kcal/mol per bond), salt bridges (additional -0.15 kcal/mol per salt bridge) and disulphide bonds (-4 kcal/mol per bond) is calculated separately.

HotSpot Wizard is another tool used in this study. The methodology for generating results in the HotSpot Wizard tool involves a multi-phase process aimed at identifying hot spots in a protein and designing smart

libraries for protein engineering. First, the annotation of the Protein. The protein is annotated using several prediction tools and databases to explore the mutational landscape and find promising mutagenesis targets. Then, the identification of Mutagenesis Hot Spots. Here, four protein engineering strategies are employed to identify suitable hot spots for improving desired protein properties:

FUNC: Analysis of functional hot spots.

FLEX: Analysis of stability hot spots/structural flexibility approach.

CONS: Analysis of stability hot spots / sequence consensus approach.

CORREL: Analysis of correlated hot spots.

Each strategy is applied to identify hot spots that could be targeted for mutagenesis to enhance protein stability, catalytic activity, substrate specificity, and enantioselectivity. Finally, HotSpot Wizard provides a way to prioritize amino acids at the randomized positions.

Results and Discussion

Mutations in the protein sequence can alter the structure conferred by the sequence of the wildtype, leading to changes in the protein's folding, stability, interactions with other molecules, and overall function. This can occur even if the conformation of the mutant protein is highly similar to that of the wildtype, with alterations and distortions far beyond the mutation site. This study is based on three levels: the evolutionary relationships and variants present in the previous ancestors, the structural significance of the changes in amino acids and their effects on properties of protein and substrate recognition, and the functional significance with binding IP3 and dephosphorylation activity.

Evolution of PTEN protein

The results of BLASTp tool were 37 sequences from the SwissProt database of proteins sharing similarity with the PTEN_Human protein. However, more than 13 sequences shared less than 30% percent identity, of which are some PTEN proteins from different species. The query coverage was greater than 50% for 70% of the proteins obtained as visualized by figure in supplementary documents. The E-value associated with the results was less than 1e -30 for the majority of the resulting sequences, lowering the similarity by chance hypothesis.

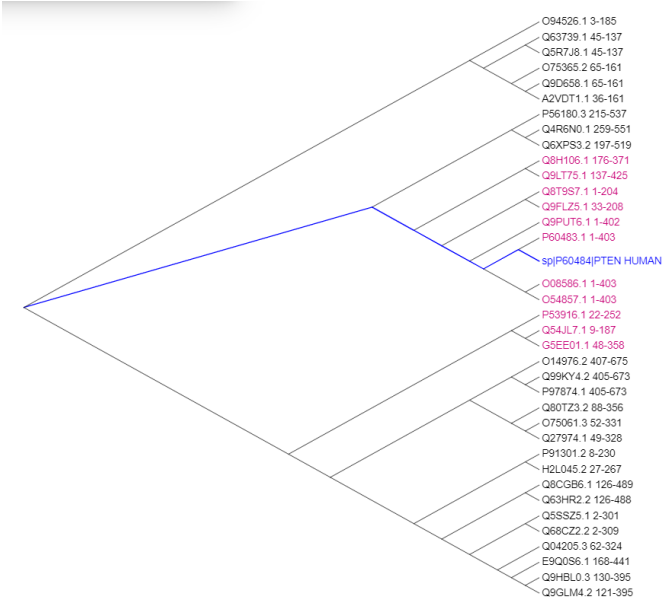


Figure 3 generated using ITOL tool, represents the evolutionary tree of PTEN_HUMAN. It represents in blue the PTEN_HUMAN ID associated with its evolutionary path. It represents, in red, the PTEN proteins for other

Four sequences shared complete coverage, percent identity of 100%, and e-value of 0.0 showing identicality to the PTEN_HUMAN protein. These sequences correspond to PTEN_CANLF, PTEN_MOUSE, PTEN_RAT, and PTEN_FROG. However, a limitation to these results is the low number of sequences present in the SwisProt database that are in similarity with the protein. This limitation is due to the criteria followed by this paper to use only Swissport reviewed protein molecules.

The results of BLASTp can be divided into two categories: PTEN proteins of various species and the other proteins of phosphatases, sharing common features with the PTEN protein due to evolutional and functional relevance. Figure 3 by ITOL below represents the phylogenic tree of PTEN_Human protein. This figure shows the divergent evolution of PTEN protein where it represents the divergence, from a common ancestor, of various Phosphatase families such as, Auxilin, Tensin, Tyr/Ser phosphatases and the PTEN family. In order, to look more deeply into the sequence level and assess the conservation of residues, Multiple Sequence Alignment was performed, where the results are represented as the most frequent amino acid/s per position in the sequence

as represented in figure 4. This figure represents the non-conserved nature among the sequence of PTEN and its related proteins. However, this is not the case for the P-loop motif of the active site where the residues appeared to be conserved throughout the phosphatase families. Figure also shows the basic nature of the P-loop with the basic residues shown in red the P-loop contains histidine, lysine, and arginine. The WPD loop and TI loop also appear to be conserved showing key residues that are conserved along the species. The conserved D92, W111, and L112 residues of the WPD loop and the Tyrosine and Isoleucine repetitions of the TI loop. These conserved residues become more observable in the PTEN protein Pubsum tool representation in figure 5.

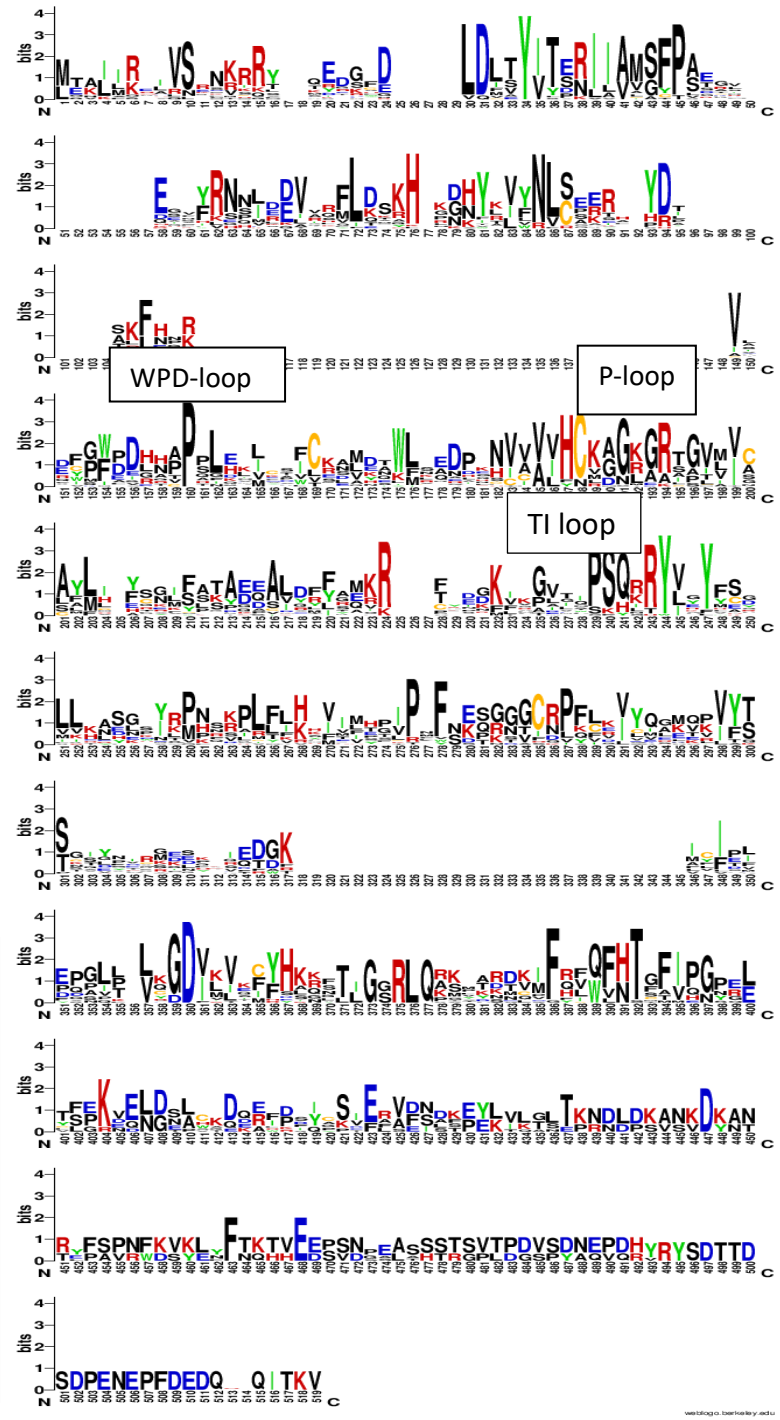


Figure 4, shows WebLogo where residues appear based on their respective frequency of occurrence in the MSA of the PTEN_HUMAN related proteins. The P-loop, WPD loop, and TI loop are highlighted.

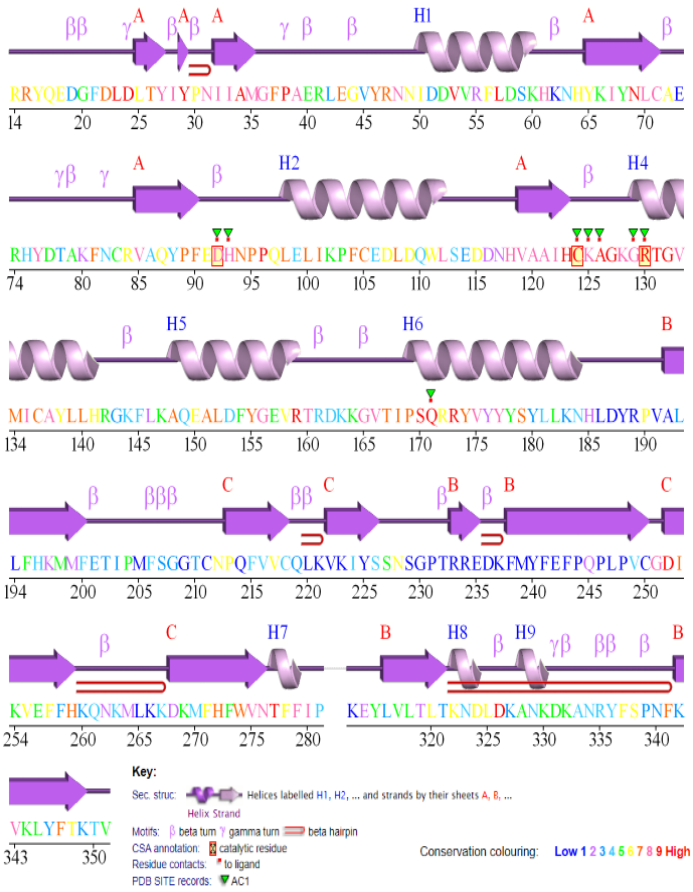


Figure 5 by PUBsum tool output of the structute having PDB ID: 1d5r, showing the different secondary structure in the PTEN protein. It also represents the residues present in the active site and those bound to the ligand. This figure presents by color scheme the conservation of each residue in the sequence of PTEN.

This figure illustrates importance of specific residues that are specific to the PTEN protein especially in the active catalytic site. The color scale present illustrates the conservation of the corresponding residue throughout the similar structures of PDB database. These results build on previous studies and suggest that that while there may not be significant sequence conservation between auxilin, tensin, and PTEN proteins overall, the conservation observed in the active site, catalytic site, and specifically the P-loop region is indicative of shared functional characteristics related to their enzymatic activities. Furthermore, this study proposes that the lack of conservation in other regions may be attributed to differences in their cellular localization, membrane-binding mechanisms, and possibly other structural or functional requirements specific to each protein. The conservation in active site among Auxilin, Tensin, and PTEN proteins supports the notion that these proteins share a common enzymatic activity. In the case of PTEN, it is a phosphatase enzyme responsible for removing phosphate groups from phosphoinositide substrates. The conservation in these regions suggests that auxilin and tensin may also possess enzymatic activities related to phosphate modification or similar biochemical reactions. This is supported by the fact that P-loop is a conserved motif in various nucleotide binding proteins including GTPases and kinases. In PTEN, the P-loop is

essential for binding to the phosphate group of the substrate. The conservation of the P-loop in Auxilin, Tensin, and PTEN suggests a shared requirement for nucleotide binding or interaction with phosphate-containing molecules. This supports the hypothesis that these proteins may share functional similarities related to nucleotide or phosphate binding. The lack of overall sequence conservation outside of the active site region might be attributed to differences in cellular localization, membrane-binding mechanisms, and other structural or functional requirements specific to each protein. For example, Auxilin is primarily involved in vesicle uncoating in the cytoplasm, whereas PTEN acts at the cell membrane to regulate phosphoinositide signaling. These differences in localization and function may drive evolutionary divergence in non-conserved regions while maintaining key functional elements. Further experimental validation, such as structural comparisons, biochemical assays, and functional studies, could provide additional evidence to confirm the proposed functional similarities and evolutionary divergence among Auxilin, Tensin, and PTEN proteins.

Docking tools results and chosen model

The results of the docking tool for Alfa-Fold modified protein and the Experimental structure missing some parts on the C-terminal, are represented in Tables 1 and 2. The criteria followed to choose the closest model to the natural occurring structure complex is same as described by bender and his colleagues (Bender J. et al., 2020). Model 1 of the Table 2 shows RMSD upper and lower bounds tending to 0. This value represents the lower limit and the upper limit of the RMSD range. It indicates the minimum and maximum

AutoDock Vina-Alfafold modified

model name	affinity (kcal/mol)	distance from best mode RMSD lower bound	RMSD upper bound	strict hydrogen bonds	contacts	clashes
1	-5.0	0.000	0.000	3	-	0
2	-4.7	6.990	10.348	0	-	0
3	-4.6	2.292	5.663	6	-	0
4	-4.5	5.533	8.520	1	-	0
5	-4.4	1.948	6.241	6	-	0
6	-4.4	5.419	9.012	2	-	0
7	-4.3	1.545	4.365	4	-	0

Table 1, represents the results of AutoDock vina tool for the protein generated by Alfa-fold program of the Chimera X tool. The colomns were dictated by the docking tool. H.bonds and clashes were calculated by Chimera X

AutoDock Vina-Experimental Structure

model name	affinity (kcal/mol)	distance from best mode RMSD lower bound	RMSD upper bound	strict hydrogen bonds	contacts	clashes
1	-6.7	0.000	0.000	14	84	1
2	-5.5	1.310	3.836	9	71	1
3	-5.5	1.908	6.590	7	51	1
4	-5.3	1.986	5.540	3	56	0
5	-4.9	7.808	10.080	1	23	0
6	-4.6	6.728	9.230	0	19	0
7	-4.6	6.415	8.900	3	29	0
8	-4.5	6.126	8.100	0	23	0
9	-4.4	2.158	5.430	9	66	0

Table 2, represents the docking results of AutoDock Vina, where the experimental structure missing some parts from the C-terminus is used. The colomns were generated by the ViewDockx from Chimera X tool and they were dictated by the docking tool, AutoDock vina.

deviation in the spatial arrangement of the ligand between the predicted mode and the best mode, respectively. While RMSD was a contributing factor the low energy and docking score associated with this model (-6.7) as compared to the other models makes it good candidate of being the complex structure. Other parameters were used to assess the quality of the structure

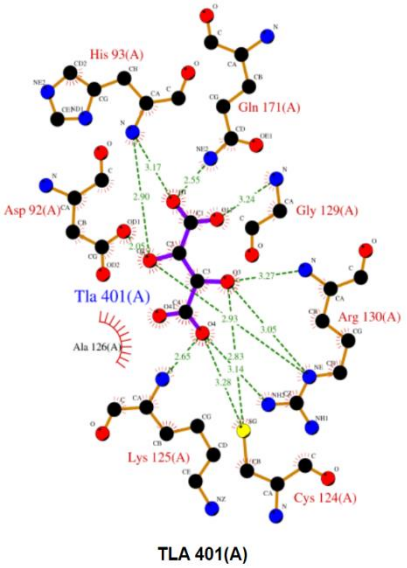
which are mainly the high number of hydrogen bonds and low number of clashes normally associated with a ligand in its active site. The clashes present could be due to flexibility within the active site and different conformations taken by PTEN. This is supported by Masson, G. et al. (2021), where the flexibility of the active

site allows the side chains to adopt different chi angles. This might lead to change in the position of atoms toward the ligand eliminating this clash. The models generated with Alfa-fold showed less docking score even for same RMSD, distance from the best model as compared to the experimental structure. While filling the missing structure from experimentally obtained structure using Alpha-fold is common practice to ensure accurate and valid results, this study have demonstrated that Alpha-fold would fail to predict accurately the active site resulting in deviations from the experimental structure. In this study, we tend to use the experimental structure even though it has some missing residues from the C-terminus. For this reason, we hypothesize that the location of the missing parts far away from the active site helped us do this act of not using the Alfa-fold generated model. Also, the Alfa-fold modified protein was seen to substitute the missing part with mainly loops which is also compatible with the theory that experimentally derived crystalized structures may not be able to detect loops and turns loosely defined or in low complexity regions. This is also corelated to the fact that PTEN exist in aqueous solutions and interact with membranes making it harder to detect such flexible structures.

Results of HDOCK upon using hybrid template-positive service were based on the PDB structure 5BUG as indicated by the docking software. The resulting models details of the docking especially model 1 referring to model 1 on the AutoDock vina tool presented in Table 2 along with models 2 and 5 corresponding to docking models of AutoDock vina 2 and 3 present in Table 2, are all available in the supplementary material under model 1, 2, and 5 files respectively. The resulting models of the template-free query had 2 models of the top 5 present in the active site which correspond to models 1 and 2, respectively, of the template-positive job. Thus, those results support the AutoDock vina results where the lowest energy structure and the predicted as top model will be selected for further investigation in this paper.

Structural Analysis of P-loop residues and associated mutations

This importance of P-loop in the active site illustrated by the conservation of its residues as illustrated by the figures (4 &5) along with the association between cancer progression and the mutations in this loop calls for analysis of these residues. The experimental structure of PTEN (PDB id: 5BZZ) having TLA (L(+)-Tartaric acid) as ligand, where the interactions of this latter with the active site residues are shown in figure 6 from PDBsum database.



TI loop, and H93 & D92 of the

Figure 6 from PDB sum shows the interactions of TLA with the surrounding residues and constituent atoms. Residues 125-9 of P-loop, Q171 of the TI loop, and H93 & D92 of the WPD loop.

Figure 6 suggest important residues for the function and structure of the PTEN protein. However, the presence of the natural substrate in the active site would drive more suggestions and implications. In fact, it has been hard to model PTEN in complex with its ligand IP3 due to the complex environment in which these two molecules interact. Here, I refer to the naturaly buried nature of the IP3 in the plasma membrane and the binding of the PTEN protein to the plasma membrane to bind this substrate. This made it hard for researchers to crystalize the structure of this complex due to flexible and aqueous environment of where it exists due to cell membranes. Docking is the tool used by this study to investigate the complex assumed to be the closest to what a PTEN-IP3 complex looks like. In this study we refer to this complex as PTEN complex for simplicity. In order to get insights into the binding energy of the complex, Pissa tool was employed which results are presented in the Table 3. These results of ΔG (kcal/mol), indicates the solvation free energy gain upon formation of the interface, in kcal/M. The value -0.9 shows

PTEN protein (receptor)			IP3 (ligand)			Interface area, Å2	ΔG kcal/mol	ΔG P-value	N(HB)
N(at)	N(res)	surfaceÅ2	N(at)	N(res)	surfaceÅ2				
33	13	15715	24	1	482	294.6	-0.9	0.485	12

Table 3, from Pisa, correspond to the interface of the PTEN complex. N(at) and N(res) correspond to interfacing atoms and residues in the corresponding structure, respectively. Surface Å2 is total solvent accessible surface area. N(HB) number of potential hydrogen bonds

that the process of bringing the proteins together to form the interface results in a decrease in the overall solvation energy. In other words, the proteins are more energetically favorable when interacting at the interface compared to when they are in their isolated states. This is also supported by the ΔG P-loop value which is a statistical measure used to assess the specificity of a protein-protein interface based on the observed solvation free energy gain at that interface. The value of 0.485 < 0.5 indicates suggests that the observed interface is energetically favorable and more stable than expected. The higher hydrophobicity of the interface indicates stronger interaction between the proteins, potentially resulting in a more stable complex. It states that the interactions in the interface are site specific and not by random assignment. The other tool employed to study interactions between the ligand and its receptor protein is the PRODIGY tool aimed on the prediction of affinity in protein-small ligand complexes (Anna V. et al., 2019). The results of PRODIGY are present in Table 4.

Complex	ΔG_{noelec} (kcal/mol)	CC	CO	CN	CX	OO	OX	NO	NN	NX	XX
PTEN complex	-6.93	531	1444	187	253	378	95	443	0	88	3

Table 4, results of PRODIGY, ΔGnoelec parameter represents the 'no electrostatic protocol', the following parameters represent atoms interactions C: carbon, O: oxygen, N: nitrogen, X: other atoms and polar hydrogens

The next step after assessing the stability of the complex is to analyze the structural factors contributing to this stability on the level of the catalytic site residues. HotSpot Wizard 3.0 tool provides the ability to analyze the structural features of the PTEN complex. The results of HotSpot Wizard were on 4 levels. This tool provided insight on functional, stability (structural), stability (sequence consensus), and correlated hotspots. The functional hotspots which refer to the highly mutable residues located in the catalytic pocket. This tool identified the catalytic site residue Valine (V45) as highly mutable with a mutability score of 8, which also appears as a stability hotspot sharing hydrogen bonds with its environment. The results cross with the results of Figure 4 where it indicated that the valine is usually substituted with another non-polar residue mainly, isoleucine, leucine, and methionine without causing loss of function besides, most of its mutations are non-damaging as indicated by the results of the HotSpot Wizard. This could indicate a role of this residue in substrate binding, catalysis, or other essential interactions that define the protein's role in biological processes.

This requires further analysis as the determination of functional hotspots might drive research forward into establishing targeted drugs and treatments. The stability hotspots generated by HotSpot Wizard are mainly present in around the active site and present in both domains, the phosphatase and the N-terminal domain. The correlated spots results of the HotSpot Wizard tool were indicative of double role of the PTEN as protein and lipid phosphatase, where the results show contact between residues of the active sites and those present considerably far from active site. This would indicate regulatory role done by those residues to maintain cellular homeostasis and proper physiological function of PTEN. The majority of those interactions take place through the histidine residue (H93) as shown in figure 7.

In order to investigate the mutations effects, it is important that the environment of the catalytic sites residues is known. A tool called PLIP is able to provide a summary of the interactions of the ligand at the interface in the active site of PTEN. Specifically, this tool discusses the hydrophobic and hydrogen interactions of the ligand with the corresponding residues. Taken as a whole the 3 loops of the active site create positively charged substrate pocket compared with other DUSPs, allowing PTEN to accommodate the bulky, negatively charged PI(3,4,5)P3 substrate with cysteine-124 sitting at the bottom of the pocket ready for catalysis (Masson, G., 2020). This is evident in the results of PLIP tool where it shows that the basic

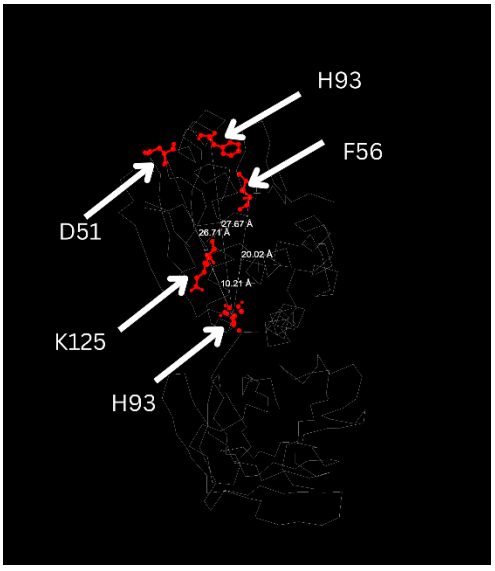


Figure 7 from HotSpot Wizard, shows the corelated hotspots with the Histidine (H93)

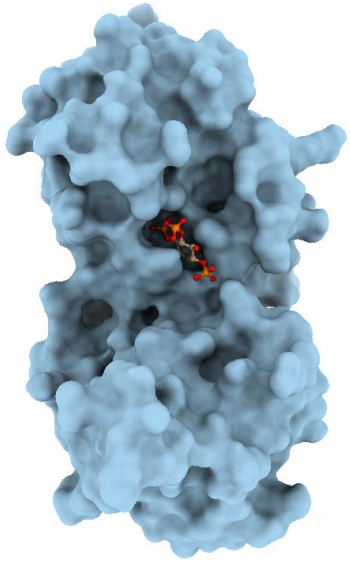


Figure8, shows the ligand (IP3) buried in the active site

sidechains of the basic amino acids at the side chain shares salt bridges with the ligand. This shows is shown in figure 8 illustrating the pocket like structure of the PTEN accommodating the ligand IP3. The Table 5, illustrates the hydrogen bonding network of the ligand in its active site from PLIP tool. Table 6, illustrate the respective parameters, structure name, HSDC, ASA, BSA, and ΔG of solvation change. These results show the complex network of interactions that stabilize IP3 in the catalytic site and proposes some candidates involved in the dephosphorylation (hydrolysis) reaction. The list of candidates include: ASP 92, THR 131, THR 167, and CYS 124.

Index	Residue	AA	Distance H-A	Distance D-A	Donor Angle	Protein donor?	Side chain	Donor Atom	Acceptor Atom
1	125A	LYS	2.94	3.82	150.34	✓	×	937 [Nam]	2600 [O2]
2	126A	ALA	2.55	3.38	142.00	✓	×	946 [Nam]	2601 [O3]
3	128A	LYS	2.11	2.81	127.33	×	✓	2584 [O3]	963 [N3]
4	128A	LYS	3.02	3.83	139.95	✓	×	955 [Nam]	2599 [O3]
5	129A	GLY	2.63	3.15	112.82	✓	×	964 [Nam]	2584 [O3]
6	130A	ARG	1.89	2.87	176.88	✓	×	968 [Nam]	2599 [O3]
7	171A	GLN	2.33	3.18	143.82	✓	✓	1303 [Nam]	2585 [O3]

Table 5, describes the hydrogen bond network of the active site residues with the ligand IP3 along with the associated atom numbers of the PTEN complex

A:CYS 124		4.47	4.47	||||||||||	0.18
[SG]					
A:LYS 125		39.90	10.38	|||	0.17
[CB]					
[CD]					
A:ALA 126		47.09	22.03	|||||	0.31
[N]	H				
[CB]					
A:GLY 127		1.35	0.00		0.00
A:LYS 128		87.13	49.15	||||||	0.43
[N]	H				
[CB]					
[CD]					
[CE]					
[NZ]					
A:GLY 129		10.92	10.92	||||||||||	0.06
[N]	H				
[CA]					
A:ARG 130		8.03	7.59	|||||||||	-1.00
[N]	H				
[CG]					
[NE]					
[NH2]	H				

Table 6, from PIZA, shows the following columns structure name, HSDC, ASA, BSA, and ΔG

Mutations in the catalytic site have been discussed in previous papers, however structural features associated with this distortion remained not understood. Isabel E., et al. (2011) discusses the mutations of PTEN and provides the outcome of those mutations on the protein activity. The paper mentions several mutations causing loss of function PTEN protein. The list included: H123Y, C124S, A126D, G127N, K125R, K128T, G129D, R130A, R130K, T131I, and T131L. The paper also mentions the residue D92 which had very low tolerance to mutation since its substitution to any residue yielded an inactive enzyme (Isabel E., et al., 2011). Masson, G. R. et al. (2020) reveal the role played by Cysteine 124 residue where this study mentions that the catalytic cysteine

in the P loop forms a covalent cysteinyl phosphoenzyme intermediate. This study suggests that the phosphate involved in this intermediate structure is hydrolyzed (removed) by the action of ASP 92, giving an important role to the ASP 92 residue. Han, S. Y. et al. (2000) and Nelen MR et al. (1999) add another mutation causing loss-of-function, which is C124R and H123R, where these mutations appeared in Cowden disease. This disease is seen to be greatly associated with a mutation in the PTEN protein (Nelen MR, 1999). Based on the previous findings of several studies which give indication about residues that are tolerant to mutation and other that are intolerant to mutation such as cysteine 124 and Arginine 130 (Masson, G. R. et al., 2020). Taking it all together, the conserved HC*XX*G*XX*R motif of the P-loop have residues that

Figure 9 from CANSAR database, presents the frequency of mutations observed in catalytic site residues

are crucial for the function of the PTEN along with some residues from TI and WPD loops. CANSAR database shows the frequency of mutation for each residue in PTEN with respect to cancer cases recorded as shown in figure 9. The greatest mutations were seen in the Arginine 130 residue with 154 mutations far away from other residues. Cysteine 136 and Aspartate 92 fall in the second and third places with 23 and 18 mutations each respectively. Figure 10, illustrate the interaction network of the key residues of the P-loop with the ligand IP3. C124 and R130 are

conserved residues within the P-loop, which through their interactions with the ligand are able to stabilize it especially through their hydrogen bonds. In the figure 10, no clashes were detected which indicate stability within the protein structure. Based on these structural features associated with each residue, having mutation in those regions would remove the network of bonds and electrostatic interactions leading to unstable complex. Moreover, as is evident in Figure 10-A, the substrate is mainly attached to the buried cysteine which plays a role in its recognition or else the absence of cysteine with its properties of being hydrophilic, having specific size of

the atom, and ability to participate in high energy intermediates will indeed cause protein being non-functional. The mutants C124S and R130K were investigated in this study by using the rotamers tool in Chimera X by Richardson and colleagues.

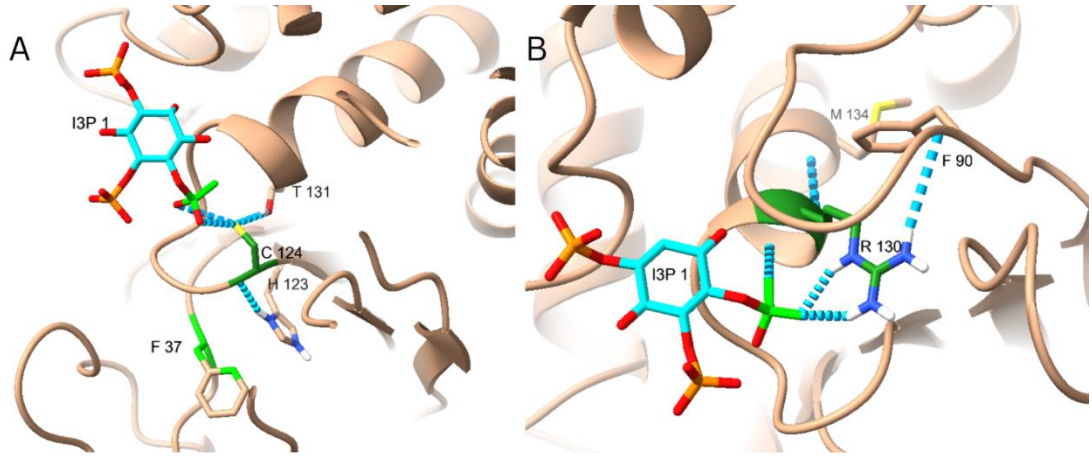


Figure 10, represents the structural representation of key binding residues A- cysteine 124 and B- Arginine 130. The residues are colored in Forest Green, the ligand IP3 is colored in cyan. The pseudobonds representt the hydrogen bonds associated with each residue, and the atoms/residues colored in lime represent those that have contacts with the corresponding residue.

Tables 7 & 8 represent the rotamers along with some properties associated with Serine and Lysine substitutes. These tables show the possible conformations of the corresponding residues in the protein. With respect to Ser 124, the results

SER 124 (C124S) rotamers			
Chi 1	Prevalence	H-bonds	Clashes
62.0	0.489007	0	2
-65.0	0.290717	0	2
-177.0	0.220277	0	2

Table 7, shows the rotamers associated with C124S mutant, serine rotamers

LYS 130 (R130K) rotamers						
Chi 1	Chi 2	Chi 3	Chi 4	Prevalence	H-Bonds	Clashes
-67.0	180	180	180	0.2479	1	6
-177	180	180	180	0.1646	1	13
-62	-68	180	180	0.0782	2	9
-67	180	180	-65	0.0569	1	5
-177	180	180	65	0.0497	0	15
-67	180	180	65	0.0426	1	6
-67	180	-68	180	0.0406	0	1
-67	180	68	180	0.0386	2	1
-177	180	180	-65	0.0376	1	14
-177	68	180	180	0.0325	0	20

Table 8, showing the top 10 hits based on the prevalence factor of the R130K mutant. Rotamers with the adjacent chi angles, 4 chi angles are present in Lysine residue, the prevalence score, h.bonds and clashes.

suggest that the greatest possibility is the one with chi 1 angle 62, based on the prevalence score. However, the clashes and hydrogen bonds indicate unstable complex resulting from this substitution. Similarly the most probable structure for Lysine is the one having chi 1, 2 , 3, & 4 being -67.0 180 180 180, respectively. This change results in 6 clashes and 1 hydrogen bond, while arginine had no clashes and additional 6 hydrogen bonds. Lysine residues at positions 125 and 128 are essential residues in the active site due to the stability they provide to the kigand. This observation is prominent in figure 12, showing the interactions of K125 and K128. The ligand is colored in cyan, however, only small part can be seen due to the green color all over the ligand signifying large number of contacts. K128 act as a backbone for the ligand anchoring it in place. These two residues undergo acetylation as a PTM to regulate the activity of PTEN protein (Han, S. Y., 2000). Acetylation of Lysine in the active site causes neutralizing of the basic charge of the lysine, which can lead to changes in protein structure and/or interactions (Kinga K. et al., 2009). This indeed affects the PTEN complex due to the fact that what maintains the integraty of this complex is the positive charge of the active site accommodating IP3. Being the most common residue to be acetylated, Lysine, this enforces the reason for conservation of these 2 together which makes the regulation by acetylation more certain due to the presence of 2 Lysine residues.

Functional Insights on the hydrolyzing residues

Phosphoinositide phosphatase activity of PTEN is critical for its tumor suppressor function (Han Y. et al., 2000). This importance of catalytic function of PTEN requires the understanding of the mechanism of dephosphorylation of inositol ligands. While some studies found D92 as a possible candidate to dephosphorylate the IP3 ligand such as this study by Masson R., et al. (2020) which states that D92 hydrolyses the covalent cysteinyl phosphoenzyme intermediate. A previous study has supported this assumption where it explains that the loss-of-function properties of the D92A mutation is in accordance with the importance of this residue in PTEN phosphatase catalysis, although the study explains that the role of ASP 92 as the general acid in the hydrolysis reaction remains unclear. Figure 11-A illustrates the interactions that D92 is involved in, as seen D92 has hydrogen bonds with N94 and K125, also it does contacts with IP3 ligand.

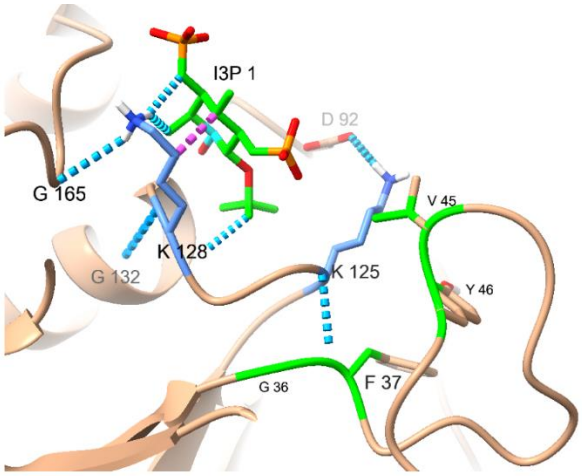


Figure 12, shows the interactions of the Lysine residues in the active site(K125 and K128). The Lysine residues are colored in sky blue and the ligand is colored with cyan. The lime greem color correspond to contacts and the blue dotted pseudobonds refer to hydrogen bonds

Based on figure 10, it was realized that both residues (R130 and C124) share interactions with the same phosphate atom suggesting potential hydrolysis takes place on this specific site, which is also supported by having this P3 the most buried among the 3 phosphates of IP3. These result support the previous work. However, an observation was also realized in this study. Threonine 131 which also a residue buried in the active site. As seen in figure 11-B, this residue has a hydrophilic side chain which is able to initiate nucleophilic attacks making it suitable for such reactions as hydrolysis. The interaction seen between cysteine and Threonine 131 is significant as the orientation of oxygen atom, interaction residues, and the theory of cysteinyl phosphoenzyme intermediate might have some implications. This might require further investigations to validate.

Conclusion

This structural study of PTEN has revealed fascinating insights into the molecular architecture of this critical tumor suppressor protein enzyme. PTEN's structure, characterized by its phosphatase domain and C2 lipid-binding domain, underscores its dual function in regulating both lipid and protein phosphorylation pathways. This study was able to give insights on key residues that might be involved in the catalysis of the ligands/substrates of PTEN namely, Asp92 and Thr131. Furthermore, the identification of key residues involved

in PTEN's enzymatic activity, stabilization forces, and regulation including Cys124, Arg130, and His93, Lys125, Lys128, and Val45, provides valuable insights into the enzyme's catalytic mechanism and modulation.

Despite these advances, several aspects of PTEN's structure and function remain elusive. The exact nature of PTEN's dimeric structure, the roles of its translational variants, and the mechanisms of action governing its enzymatic activity and regulation are areas that warrant further investigation. Understanding these aspects is crucial for knowing the full spectrum of PTEN's functions and its role in disease, particularly in cancer.

Moreover, the structural similarities between PTEN and other proteins, such as Tensin proteins and auxilin, suggest a broader evolutionary conservation of these structural features especially in catalytic residues despite early divergent evolution. This finding opens avenues for comparative studies to gain insights into the functional and structural adaptations of these proteins across different species.

In conclusion, this structural study of PTEN has significantly advanced our understanding of this enzyme's role in cellular processes and its importance in tumor suppression. However, the complexity of PTEN's regulation and the outstanding questions about its structure and function underscore the need for continued research. Future studies will undoubtedly shed more light on the intricate mechanisms underlying PTEN's activity and its potential as a therapeutic target in cancer treatment.

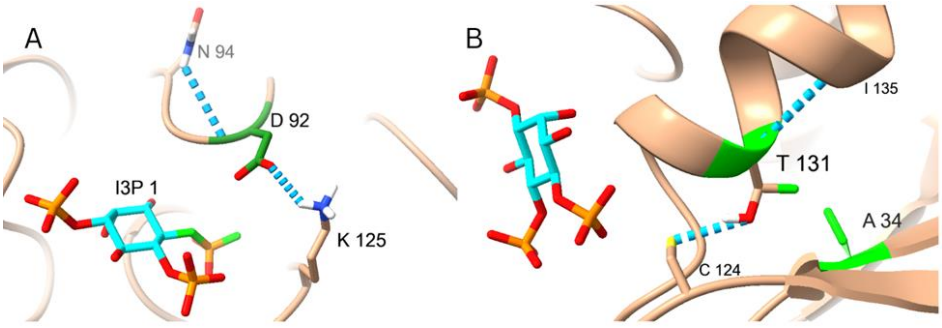


Figure 11, shows two residues having hydrophilic side chains, A-Aspartate 92 and B-Threonine 131. Both contribute to interactions within the protein. In 10-A, D92 is colored with forest green while in 10-B, T131 is colored with Beige. The green colored atoms/bonds are the contacts, while the pseudobonds colored in blue represent h.bonds

References

1. Robert A. Weinberg, Tumor Suppressor Genes.*Science***254**,1138-1146(1991).DOI:10.1126/science.1659741

2. Liu, Y., Hu, X., Han, C., Wang, L., Zhang, X., He, X., & Lu, X. (2015). Targeting tumor suppressor genes for cancer therapy. *BioEssays : news and reviews in molecular, cellular and developmental biology*, *37*(12), 1277–1286. https://doi.org/10.1002/bies.201500093.

3. Pessôa, I. A., Amorim, C. K., Ferreira, W. A. S., Sagica, F., Brito, J. R., Othman, M., Meyer, B., Liehr, T., & de Oliveira, E. H. C. (2019). Detection and Correlation of Single and Concomitant *TP53, PTEN,* and *CDKN2A* Alterations in Gliomas. *International journal of molecular sciences, 20*(11), 2658. https://doi.org/10.3390/ijms20112658

4. Dempsey, D. R., Viennet, T., Iwase, R., Park, E., Henriquez, S., Chen, Z., Jeliazkov, J. R., Palanski, B. A., Phan, K. L., Coote, P., Gray, J. J., Eck, M. J., Gabelli, S. B., Arthanari, H., & Cole, P. A. (2021). The structural basis of PTEN regulation by multi-site phosphorylation. *Nature structural & molecular biology*, *28*(10), 858–868. https://doi.org/10.1038/s41594-021-00668-5

5. Masson, G. R., & Williams, R. L. (2020). Structural Mechanisms of PTEN Regulation. *Cold Spring Harbor perspectives in medicine*, *10*(3), a036152. https://doi.org/10.1101/cshperspect.a036152

6. Isabel Rodríguez-Escudero, María D. Oliver, Amparo Andrés-Pons, María Molina, Víctor J. Cid, Rafael Pulido, A comprehensive functional analysis of PTEN mutations: implications in tumor- and autism-related syndromes, *Human Molecular Genetics*, Volume 20, Issue 21, 1 November 2011, Pages 4132–4142, https://doi.org/10.1093/hmg/ddr337.

7. Orloff M.S., Eng C.. Genetic and phenotypic heterogeneity in the PTEN hamartoma tumour syndrome, Oncogene, 2008, vol. 27 (pg. 5387-5397).

8. Waite K.A., Eng C.. Protean PTEN: form and function, Am. J. Hum. Genet., 2002, vol. 70 (pg. 829-844).

9. Leslie N.R., Foti M.. Non-genomic loss of PTEN function in cancer: not in my genes, Trends Pharmacol. Sci., 2011, vol. 32 (pg. 131-140)

10. Lee J.O., Yang H., Georgescu M.M., Di Cristofano A., Maehama T., Shi Y., Dixon J.E., Pandolfi P., Pavletich N.P.. Crystal structure of the PTEN tumor suppressor: implications for its phosphoinositide phosphatase activity and membrane association, Cell, 1999, vol. 99 (pg. 323-334).

11. Mills, J. E., & Dean, P. M. (1996). Three-dimensional hydrogen-bond geometry and probability information from a crystal survey. *Journal of computer-aided molecular design*, *10*(6), 607–622. https://doi.org/10.1007/BF00134183

12. Han, S. Y., Kato, H., Kato, S., Suzuki, T., Shibata, H., Ishii, S., Shiiba, K., Matsuno, S., Kanamaru, R., & Ishioka, C. (2000). Functional evaluation of PTEN missense mutations using in vitro phosphoinositide phosphatase assay. *Cancer research*, *60*(12), 3147–3151.

13. Debbie J. Marsh, Jennifer B. Kum, Kathryn L. Lunetta, Michael J. Bennett, Robert J. Gorlin, S. Faisal Ahmed, Joann Bodurtha, Carol Crowe, Mary A. Curtis, Majed Dasouki, Teresa Dunn, Howard Feit, Michael T. Geraghty, John M. Graham, Shirley V. Hodgson, Alasdair Hunter, Bruce R. Korf, David Manchester, Susan Miesfeldt, Victoria A. Murday, Katherine L. Nathanson, Melissa Parisi, Barbara Pober, Corrado Romano, John L. Tolmie, Richard Trembath, Robin M. Winter, Elaine H. Zackai, Roberto T. Zori, Liang-Ping Weng, Patricia L. M. Dahia, Charis Eng, *PTEN* Mutation Spectrum and Genotype-Phenotype Correlations in Bannayan-Riley-Ruvalcaba Syndrome Suggest a Single Entity With Cowden Syndrome, *Human Molecular Genetics*, Volume 8, Issue 8, August 1999, Pages 1461–1472, https://doi.org/10.1093/hmg/8.8.1461

14. Yan Y, Tao H, He J, Huang S-Y.* The HDOCK server for integrated protein-protein docking. Nature Protocols, 2020; doi: https://doi.org/10.1038/s41596-020-0312-x.

15. Yan Y, Zhang D, Zhou P, Li B, Huang S-Y. HDOCK: a web server for protein-protein and protein-DNA/RNA docking based on a hybrid strategy. Nucleic Acids Res. 2017;45(W1):W365-W373.

16. Yan Y, Wen Z, Wang X, Huang S-Y. Addressing recent docking challenges: A hybrid strategy to integrate template-based and free protein-protein docking. Proteins 2017;85:497-512.

17. Huang S-Y, Zou X. A knowledge-based scoring function for protein-RNA interactions derived from a statistical mechanics-based iterative method. Nucleic Acids Res. 2014;42:e55.

18. Huang S-Y, Zou X. An iterative knowledge-based scoring function for protein-protein recognition. Proteins 2008;72:557-579.

19. Lovell, S.C., Word, J.M., Richardson, J.S. and Richardson, D.C. (2000), The penultimate rotamer library‡. Proteins, 40: 389-408. https://doi.org/10.1002/1097-0134(20000815)40:3<389::AID-PROT50>3.0.CO;2-2

20. Li AJ, Nussinov R. A set of van der Waals and coulombic radii of protein atoms for molecular and solvent-accessible surface calculation, packing evaluation, and docking. Proteins. 1998 Jul 1;32(1):111-27. PMID: 9672047.

21. Sigfridsson, E. and Ryde, U. (1998), Comparison of methods for deriving atomic charges from the electrostatic potential and moments. J. Comput. Chem., 19: 377-395. https://doi.org/10.1002/(SICI)1096-987X(199803)19:4<377::AID-JCC1>3.0.CO;2-P

22. Samia Aci-Sèche, Stéphane Bourg, Pascal Bonnet, Joseph Rebehmed, Alexandre G. de Brevern, Julien Diharce, A perspective on the sharing of docking data. Data in Brief, Volume 49, 2023,109386,ISSN 2352-3409, https://doi.org/10.1016/j.dib.2023.109386.

23. Yan, Y., Zhang, D., Zhou, P., Li, B., & Huang, S. Y. (2017). HDOCK: a web server for protein-protein and protein-DNA/RNA docking based on a hybrid strategy. *Nucleic acids research*, *45*(W1), W365–W373. https://doi.org/10.1093/nar/gkx407

24. Bender, B.J., Gahbauer, S., Luttens, A. *et al.* A practical guide to large-scale docking. *Nat Protoc* **16**, 4799–4832 (2021). https://doi.org/10.1038/s41596-021-00597-z

25. Masson, G. R., & Williams, R. L. (2020). Structural Mechanisms of PTEN Regulation. *Cold Spring Harbor perspectives in medicine*, *10*(3), a036152. https://doi.org/10.1101/cshperspect.a036152

26. Anna Vangone, Joerg Schaarschmidt, Panagiotis Koukos, Cunliang Geng, Nevia Citro, Mikael E Trellet, Li C Xue, Alexandre M J J Bonvin, Large-scale prediction of binding affinity in protein–small ligand complexes: the PRODIGY-LIG web server, *Bioinformatics*, Volume 35, Issue 9, May 2019, Pages 1585–1587, https://doi.org/10.1093/bioinformatics/bty816

27. Han, S. Y., Kato, H., Kato, S., Suzuki, T., Shibata, H., Ishii, S., Shiiba, K., Matsuno, S., Kanamaru, R., & Ishioka, C. (2000). Functional evaluation of PTEN missense mutations using in vitro phosphoinositide phosphatase assay. *Cancer research*, *60*(12), 3147–3151.

28. Nelen MR, Kremer H, Konings IB, Schoute F, van Essen AJ, Koch R, Woods CG, Fryns JP, Hamel B, Hoefsloot LH, Peeters EA, Padberg GW. Novel PTEN mutations in patients with Cowden disease: absence of clear genotype-phenotype correlations. Eur J Hum Genet. 1999 Apr;7(3):267-73. doi: 10.1038/sj.ejhg.5200289. PMID: 10234502.

29. Han, S. Y., Kato, H., Kato, S., Suzuki, T., Shibata, H., Ishii, S., Shiiba, K., Matsuno, S., Kanamaru, R., & Ishioka, C. (2000). Functional evaluation of PTEN missense mutations using in vitro phosphoinositide phosphatase assay. *Cancer research*, *60*(12), 3147–3151.

30. Kinga Kamieniarz, Robert Schneider,Tools to Tackle Protein Acetylation,Chemistry & Biology,Volume 16, Issue 10,2009,Pages 1027-1029,ISSN 1074-5521, https://doi.org/10.1016/j.chembiol.2009.10.002.

31. Vangone A., Schaarschmidt J., Koukos P., Geng C., Citro N., Trellet M.E., Xue L., Bonvin A.M.J.J. "Large-scale prediction of binding affinity in protein-small ligand complexes: the PRODIGY-LIG web server", Bioinformatics, bty816, https://doi.org/10.1093/bioinformatics/bty816

32. Kastritis P.L. et al. (2014) Proteins feel more than they see: fine-tuning of binding affinity by properties of the non-interacting surface. J. Mol. Biol., 426, 2632–2652.

33. Herrero, J., Muffato, M., Beal, K., Fitzgerald, S., Gordon, L., Pignatelli, M., Vilella, A. J., Searle, S. M., Amode, R., Brent, S., Spooner, W., Kulesha, E., Yates, A., & Flicek, P. (2016). Ensembl comparative genomics resources. Database : the journal of biological databases and curation, 2016, bav096. https://doi.org/10.1093/database/bav096

34. Trott, O., & Olson, A. J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry*, *31*(2), 455–461. https://doi.org/10.1002/jcc.21334