IN SILICO ANALYSIS OF ANTI-CERVICAL CANCER DRUG OFF-TARGET EFFECTS ON DIVERSE PROTEIN ISOFORMS FOR ENHANCED THERAPEUTIC STRATEGIES

A. Iqbal1, F. Ali2, S. Choudhary3, A. Qayyum3, F. Arshad1, S. Ashraf1, M. Aziz2, A.U. Shakil1, S.A. Sehgal1, M. Hussain1, M. Sajid1

1Faculty of Life Sciences, Department of Biotechnology, University of Okara, Okara, Punjab, Pakistan
2Department of Bioinformatics, Islamia University Bahawalpur, Bahawalpur, Punjab, Pakistan
*Corresponding author: infobiotec@uo.edu.pk

Received 25 September 2023; Accepted 20 November 2023

Background. Cervical cancer is a serious medical condition that affects hundreds of thousands of individuals worldwide annually. The selection and analysis of suitable gene targets in the early stages of drug design are crucial for combating this disease. However, overlooking the presence of various protein isoforms may result in unwanted therapeutic or harmful side effects.

Objective. This study aimed to provide a computational analysis of the interactions between cervical cancer drugs and their targets, influenced by alternative splicing.

Methods. Using open-access databases, we targeted 45 FDA-approved cervical cancer drugs that target various genes having more than two distinct protein-coding isoforms. To check the conservation of binding pocket in isoforms of the genes, multiple sequence analysis was performed. To better understand the associations between proteins and FDA-approved drugs at the isoform level, we conducted molecular docking analysis.

Results. The study reveals that many drugs lack potential targets at the isoform level. Further examination of various isoforms of the same gene revealed distinct ligand-binding pocket configurations, including differences in size, shape, electrostatic characteristics, and structure.

Conclusions. This study highlights the potential risks of focusing solely on the canonical isoform, and ignoring the impact of cervical cancer drugs on- and off-target effects at the isoform level. These findings emphasize the importance of considering interactions between drugs and their targets at the isoform level to promote effective treatment outcomes.

Keywords: cervical cancer; isoforms; molecular docking; interaction analysis; bioinformatics approaches.

Introduction

In developing countries, cervical cancer is the main cause of cancer-related deaths and loss of life loss [1]. Several years earlier than the median age at which breast, lung, and ovarian cancers are diagnosed, cervical cancer is commonly diagnosed in the fifth decade of life [2]. Ninety percent of the 270,000 cervical cancer fatalities in 2015 occurred in low- and middle-income countries (LMIC), where mortality was 18 times higher than in developed nations [3]. Nearly all cervical cancers are caused by high-risk subtypes of the human papillomavirus (HPV), and screening and vaccination programs are effective preventive measures against HPV [4]. The two most prevalent histological subtypes (squamous cell carcinoma, and adenocarcinoma) account for 70% and 25% of all cervical malignancies, respectively [5, 6]. The major decrease in cervical cancer mortality has been attributed to the development and implementation of screening programs [7]. Cervical cancer has a poor prognosis following metastasis or recurrence, with a 5-year overall survival (OS) rate of approximately 17% [8]. To improve the efficacy of cervical cancer treatment, it is crucial to identify novel therapeutic targets and survival-associated biomarkers.

Major innovations in large-scale multi-omics research provide a unique perspective for systems biology analyses of the emergence and spread of cancers. HPV contributes to the development of cervical cancer, which is considered to be a virus-driven malignancy. Early HPV infection may simply be a result of external causes, such as changes in the genome, eventually causing cervical epithelial cells to become malignant (e.g., gene fusion, non-coding RNAs, copy number variation, DNA methylation, and somatic DNA mutations) [9–13]. Transcriptomic and epigenetic modifications have been the focus of several prospective studies. However, Alternative splicing (AS) in cancer post-transcriptional protein isoforms has not yet been thoroughly studied.
In eukaryotes, a remarkable biological process known as alternative splicing, which promotes proteomic diversity, allows a single gene to express several protein isoforms. In humans, where more than 94% of genes are alternatively spliced, the occurrence and properties of alternative splicing are highly diverse [14–16]. This method enables cancer cells to generate abnormal proteins with altered functional domains, which promotes carcinogenesis [17–19]. In malignancies, these domain changes can lead to complicated remodeling and protein-protein interactions. Some essential oncogenic splicing variations can control tumor epithelial-to-mesenchymal transition and biological processes in cancer stem cells [20]. Gene expression is properly controlled in a context-specific manner, even if gene isoforms appear to have different and sometimes even opposing functions.

Aberrant protein isoforms that cause diseases have the potential to be effective drug targets in addition to serving as significant biomarkers [21, 22]. In this study, we examined the effectiveness of FDA-approved drugs against various cervical cancer-related gene isoforms. Using structural analysis and clinical data on the expression of these genes, we curated the drug interaction data for various isoforms of different genes implicated in cervical cancer, and evaluated their effectiveness against protein isoforms. In this study, we primarily focused on cervical cancer and examined whether the drugs were effective against target gene isoforms.

**Methods**

**Collection of genes and their protein isoforms**

We identified genes associated with cervical cancer using the COSMIC database [23], an online resource for somatically acquired mutations reported in human cancers. More than 30 genes may contribute to cervical cancer (Supplementary File 1). Based on the number of patient samples, the top five genes out of 30 were selected and used for further analysis. The Ensemble genome database [24] was used to curate the gene isoforms and protein sequences of these genes. Using the COSMIC Mutation ID, mutations were identified in the genes and matched with the variants of each gene isoform using the Ensemble database.

**Curation of drugs-target interaction data**

Using the Drug Gene Interaction Database (DGIdb) [25], we curated the FDA Approved drugs for our genes. Using this database, more than 40 drugs that received FDA approval were identified. These drugs were retrieved from the Drug Bank [26] and chEBI [27].

**Sequence analysis of gene isoforms**

To check the conservation of binding pocket in isoforms of the genes, Binding Pockets of the canonical proteins were predicted through the COACH Server (https://zhanggroup.org/COACH/). We identified domains from the EMBL-EBI InterPro database [28] and aligned them with the sequences of the canonical protein and its protein isoforms. Using the Bioconductor program msa, we curated the drug interaction data for various cervical cancer and examined whether protein-coding isoforms are upregulated or downregulated in cervical cancer. The expression of gene isoforms was examined in normal patient samples using GTEx and in tumor samples using TCGA, both of which were drawn from 307 Cervical Cancer Samples available in the UCSC Xena database. We also visualized the exon structure of the gene isoforms to better understand the pattern of alternative splicing in various isoforms of the genes.

**Gene isoforms expression in normal and tumors samples**

We examined the clinical data offered by UCSC Xena [29] for patients with cervical cancer, which is an online resource for analyzing multi-omics, clinical, and phenotypic data. We used UCSC Xena to compare TCGA tumor samples with normal GTEx samples to evaluate whether protein-coding isoforms are upregulated or downregulated in cervical cancer. The expression of gene isoforms was examined in normal patient samples using GTEx and in tumor samples using TCGA, both of which were drawn from 307 Cervical Cancer Samples available in the UCSC Xena database. We also visualized the exon structure of the gene isoforms to better understand the pattern of alternative splicing in various isoforms of the genes.

**Structure prediction of protein isoforms and ligand docking**

To better understand the associations between proteins and their ligands (drugs), we predicted the 3D structures of protein isoforms using a number of tools for structural level study of the different isoforms of the proteins. Protein isoform structures were predicted using the structure prediction tools trRosetta [30], Robetta [31], Swiss-Model [32], and I-TASSER [33]. Furthermore, the ERRAT quality factor and the favored, allowed, and disabled regions in the Ramachandran plot were used to evaluate the predicted structures. After evaluation, we used SiteMap53 [34] to determine the drug target region in those protein isoforms’
3D structures. The predicted 3D structures of the protein isoforms were prepared for docking analysis using Chimera 1.15 rc. We used Pyrex software to perform ligand-protein docking analysis, and considered a number of drugs that have already been approved for such proteins to check the effectiveness of these drugs against various protein isoforms that are affected by disease. Poses of the protein-ligand complexes were captured to further analyze the pocket sizes, shapes, and electrostatic surfaces of the docked protein isoforms.

**Interaction analysis**

The Discovery Studio 2021 Client was used to examine the protein-ligand complexes. We examined how the drug, which has high binding affinity for canonical proteins, interacts with different protein isoforms. Furthermore, we examined the interactions between the hydrophobic and hydrogen sites in different docked protein isoforms.

**Results**

**Drugs Target Genes have multiple isoforms**

More than 30 genes linked to cervical cancer were identified to have missense mutations (Supplementary File 1). Five genes were selected for further analysis considering the number of patient samples. We identified FDA-approved drug interactions to analyze interactions between drugs and their target protein isoforms. We retrieved more than 145 entries belonging to five distinct Cervical Cancer genes.

A partial list of the summary tables is presented in Table 1. We found that the bulk of the candidate genes had two or more transcribed spliced variants and protein isoforms (Fig. 1).

Our findings demonstrate that the majority of cancer drug target genes undergo splicing and produce many gene isoforms that may be functionally distinct and react with drugs in different ways, highlighting the significance of obtaining protein isoforms and alternative splicing in drug development.

![Graph showing frequency of transcript variants and protein-coding isoforms](image)

**Table 1: FDA Approved Drugs against target genes and protein-coding isoforms**

<table>
<thead>
<tr>
<th>Genes</th>
<th>Abbreviations</th>
<th>FDA approved drugs</th>
<th>Number of transcript</th>
<th>Number of protein isoforms</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>KRAS</td>
<td>Kirsten rat sarcoma virus</td>
<td>Cetuximab, AZD-4785, Selumetinib, CC-223, AZD-8835, PD-0325901, Trametinib, Ridaforolimus</td>
<td>14</td>
<td>9</td>
<td>[35–38]</td>
</tr>
<tr>
<td>SMAD4</td>
<td>Mothers against decapentaplegic homolog 4</td>
<td>Lysine, Sapanisertib, Fluorouracil, Alectinib, Crizotinib, Cetuximab, Gemcitabine, Irinotecan, Carboplatin, Placlitaxel</td>
<td>10</td>
<td>4</td>
<td>[39–42]</td>
</tr>
<tr>
<td>PIK3CA</td>
<td>Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha</td>
<td>Trastuzumab, Tensirolimus, Serabelisib, Taselisib, CC-223, INK-1117, Alpelisib, Buparlisib, Capivasertib</td>
<td>16</td>
<td>10</td>
<td>[43, 44]</td>
</tr>
<tr>
<td>ERBB3</td>
<td>Human epidermal growth factor receptor 3</td>
<td>Pertuzumab, Trastuzumab, MM-121, AV-203, AMG-888, Patritumab, Duligotuzumab, Saptinib, MM-111</td>
<td>33</td>
<td>12</td>
<td>[45–47]</td>
</tr>
<tr>
<td>FBXW7</td>
<td>F-Box and WD Repeat Domain Containing 7</td>
<td>Tensirolimus, Sirolimus, Regorafenib, Vorinostat Belinostat, Entinostat, Docetaxel, Vorinostat, AR-42</td>
<td>24</td>
<td>8</td>
<td>[35, 48, 49]</td>
</tr>
</tbody>
</table>
Differences in binding pockets among protein isoforms

Using several sequence alignments, we identified the precise interaction residues in the drug-binding region of each protein isoform. We performed multiple sequence alignments between the Pfam functional domains, canonical proteins, protein isoform sequences, and predicted protein-binding pockets. Here, we describe the sequence alignment plots of a few genes.

Cellular functions essential for cancer development, such as cell growth, proliferation, motility, survival, and metabolism, are regulated by Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha (PIK3CA) [50]. PIK3CA has four protein isoforms (PIK3CA-201, PIK3CA-203, PIK3CA-204, and PIK3CA-205). Protein isoforms PIK3CA-203 & 204 have 21 and 118 residues, respectively. Comparison with the predicted pocket binding of the canonical protein. It was found that these residues lack residues in the binding pocket, which indicates that these residues completely lack the predicted pocket binding (Fig. 2). The canonical proteins and protein isoforms, PIK3CA-201 and 205, were found to have identical sequences in the predicted binding pocket. However, we found variations in the C-terminal regions and domain PF00454 of the protein isoforms PIK3CA-201 and 205 (Fig. 2). We examined the C-terminal region of the canonical protein, PIK3CA-201 & 205, and the Pfam domain PF00454 to further explain this variation. In previous studies, we have found that the C-terminal region is necessary for catalysis. This has been suggested to be a crucial PI3Ks regulating component [51]. The Pfam domain is a domain of the p100α catalytic subunit of PIK3CA. However, in USP13-PIK3CA, the entire C-terminal region was replaced by USP13, which affected its catalytic activity. Since PIK3CA-201 and PIK3CA-205 have the same upstream regions, the fusion proteins produced by the two protein isoforms should ideally have the same structure. Additionally, we aligned two other USP13-PIK3CA protein sequences in the FusionGDB database to support this claim, and all sequences had overlapping interference residues with the predicted pocket binding (Supplementary File 2). These sequence-level data indicate that the drug may target all

![Figure 2](image-url)

Figure 2: Sequence alignments of predicted pocket binding residues of PIK3CA isoforms. The predicted binding pocket residues; aligned Pfam domains; and PIK3CA-201, PIK3CA-203, PIK3CA-204, and PIK3CA-205 are shown from top to bottom. Each line included the sequence logo of the consensus sequences at the top. Residues in the sequence that coincided with anticipated binding residues are shown in blue. Purple color suggests that this residue is conserved in approximately 50% of all sequences. Similar amino acids are shown under the pink shading.
USP13–PIK3CA fusion protein splice-variant protein isoforms. As a result, splice-variation within the PIK3CA gene does not influence binding to its targets in protein isoforms PIK3CA–201 and 205, whereas it may affect PIK3CA–203 and 204, which do not have a predicted binding pocket.

In 285 cervical cancers, PIK3CA 16 targetable oncogenic mutations were found to be the most common oncogenic mutation, with KRAS mutation coming in second. However, despite significant efforts, cancers with KRAS mutations remain challenging to treat because of the plasticity of tumor cells and acquisition of additional mutations. Multiple sequence alignments of KRAS protein isoforms (KRAS–201, 202, 205, 203, 204, 210, and 214) indicated that isoforms 201, 202, 205, and 203 have binding residues and are thus likely targets of drugs, while isoforms 204 and 207 lack binding pockets and are not predicted to be targets of KRAS-targeting drugs, as shown in Fig. 3. Further investigation revealed variations in KRAS isoforms 202, 205, 203, and 204 at the C terminal compared to KRAS isoform 201. These findings suggest that further efforts are required to specifically target KRAS protein isoforms.

**High gene isoforms expression in tumor tissues**

Using clinical information from the UCSC Xena, which is accessible through several projects (TCGA, GTEx, and TARGET), we were able to determine the expression of gene isoforms. We observed the expression of PIK3CA and KRAS gene isoforms in TCGA samples of cervical and breast cancers, as shown in Fig. 4A. The expression of the gene isoforms was nearly the same in both the cancer types. The gene isoform (PIK3CA–204/ENST00000477735.1) is not expressed in tumor and normal samples, and is thus ignored. The gene isoform (PIK3CA–203/ENST00000468036.1) was highly expressed in TCGA tumor samples, in contrast to normal GTEx samples. Although we previously found that gene isoform–203 does not have a predicted binding pocket, we observed that tumor cells expressed it. This should be included in future studies to examine the on- and off-target effects of these drugs.

Using transcriptome expression data from TCGA repository, it was possible to compare the expression of KRAS gene isoforms

\[
\text{KRAS–202/ENST00000311936.7, } \\
\text{KRAS–203/ENST00000556131.1, } \\
\text{and } \text{KRAS–204/ENST00000557334.5) in cervical and breast samples (Fig. 4B). Compared to normal samples, tumor samples were shown to have higher levels of KRAS–203 expression. Sequence analysis of FBXW7, ERBB3, and SMAD4 is shown in Supplementary File 3. Future studies analyzing the on- and off-target effects of drugs should consider these gene isoforms, as they are expressed in tumors.}
\]

**Figure 3:** Sequence alignments of the predicted pocket-binding residues of various KRAS protein isoforms. Using Bioconductor software msa, Cluster Omega was used to align the binding residues with the protein isoform sequences. The predicted binding pocket residues, aligned Pfam domains, and KRAS protein isoforms are shown from top to bottom. Each line included the sequence logo of the consensus sequences at the top. Residues in the sequence that coincided with anticipated binding residues are shown in blue. Purple color suggests that this residue is conserved in approximately 50% of all sequences. Similar amino acids are shown under the pink shading.

**Figure 4A:** The expression of PIK3CA isoforms was nearly the same in both the cancer types. The gene isoform (PIK3CA–204/ENST00000477735.1) is not expressed in tumor and normal samples, and is thus ignored. The gene isoform (PIK3CA–203/ENST00000468036.1) was highly expressed in TCGA tumor samples, in contrast to normal GTEx samples. Although we previously found that gene isoform–203 does not have a predicted binding pocket, we observed that tumor cells expressed it. This should be included in future studies to examine the on- and off-target effects of these drugs.

**Figure 4B:** Using transcriptome expression data from TCGA repository, it was possible to compare the expression of KRAS gene isoforms

\[
\text{KRAS–202/ENST00000311936.7, } \\
\text{KRAS–203/ENST00000556131.1, } \\
\text{and } \text{KRAS–204/ENST00000557334.5) in cervical and breast samples (Fig. 4B). Compared to normal samples, tumor samples were shown to have higher levels of KRAS–203 expression. Sequence analysis of FBXW7, ERBB3, and SMAD4 is shown in Supplementary File 3. Future studies analyzing the on- and off-target effects of drugs should consider these gene isoforms, as they are expressed in tumors.}
\]
Interactions of drugs on structural level

Although we have shown changes in binding pockets across gene isoforms at the sequence level, structural-level research is the only way to gain more solid proof that drugs bind to their target protein isoforms in distinct ways. We studied the KRAS gene, which has seven distinct protein isoforms, together with known drugs that target them, to understand how a certain drug molecule interacts with several isoforms of a protein.

The three-dimensional (3D) structure of each protein isoform was predicted using various databases. The best predicted structures were projected to have ERRAT scores greater than 94. The structures with poor ERRAT values were further improved.

Next, using Pyrex, we conducted docking analysis while considering the selection of drugs that have been identified to target this disease protein target. After analyzing the docked positions, we observed that, although some drugs bind similarly to protein isoforms, others bind extremely differently. For instance, protein isoforms KRAS-201, 202, 205, and 214 had good Kras-201, 202, 205, and 214 scores, respectively. These six protein isoforms had strong binding affinities for Trametinib, although KRAS-202 had a low binding affinity. All protein isoforms showed good binding affinity with ridoforolimus. While the remaining drugs showed good binding affinities with these protein isoforms, certain protein isoforms displayed lower affinities.

In the case of PIK3CA, the protein isoforms PIK3CA-203 and 204 showed low binding affinity for approved FDA Drugs, as these protein isoforms have short sequences and do not have a predicted binding pocket. All the other protein isoforms of KRAS (KRAS-201, 202, 205, 210, 213, and 214) have high binding affinities. AZD-4785 had good KRAS-201, 202, 205, and 214 scores, respectively. These six protein isoforms had strong binding affinities for Temsirolimus, although KRAS-202 had a low binding affinity. All protein isoforms showed good binding affinity with Temsirolimus. While the remaining drugs showed good binding affinities with these protein isoforms, certain protein isoforms displayed lower affinities.

In the case of PIK3CA, the protein isoforms PIK3CA-203 and 204 showed low binding affinity with FDA Approved drugs (Table 2). This finding supports our previous finding that these protein isoforms have very small sequences and do not have a predicted binding pocket. All the other protein isoforms of KRAS (KRAS-201, 202, 205, 210, 213, and 214) have high binding affinities. AZD-4785 had good KRAS-201, 202, 205, and 214 scores, respectively. These six protein isoforms had strong binding affinities for Temsirolimus, although KRAS-202 had a low binding affinity. All protein isoforms showed good binding affinity with Temsirolimus. While the remaining drugs showed good binding affinities with these protein isoforms, certain protein isoforms displayed lower affinities.
To explain how different pocket sizes, shapes, and electrostatic potential surfaces may create an illusion similar to the binding mode even when the scores are the same in some instances. Here, we examined the temsirolimus binding mode in all four protein isoforms and discovered that while the binding scores were similar, the binding patterns varied greatly, as shown in (Fig. 5). The molecular docking results for FBXW7, ERBB3, and SMAD4 are shown in Supplementary File 4. These results led us to hypothesize that, despite the identity of the ligand-binding residues, the binding pocket structures change in size, form, and dynamic properties, resulting in different binding patterns for a single drug in several protein isoforms with various binding affinity values.

Interaction analysis of the target protein isoforms was performed to determine the type and number of interactions between the docked temsirolimus and PIK3CA protein isoforms. When a complex has a significant number of hydrogen bonds together with a small number of salt bridges, hydrophobic contacts, and \( \pi-\pi \) interactions, it is said to be strong. To determine the number of interactions generated by each molecule, we tested each docked drug differently (Fig. 6). According to the interaction study, complexes with strong binding affinities produced the most hydrogen bonds (Table 4).

PIK3CA-Canonical and protein isoforms 201 and 205 were shown to have strong interactions, whereas the docked PIK3CA-203 complex was found to have weak interactions.

### Table 2: Binding Affinity Values of the KRAS-Canonical protein and its protein isoforms

<table>
<thead>
<tr>
<th>Drugs</th>
<th>KRAS-Canonical</th>
<th>KRAS-201</th>
<th>KRAS-202</th>
<th>KRAS-205</th>
<th>KRAS-214</th>
<th>KRAS-213</th>
<th>KRAS-210</th>
<th>KRAS-203</th>
<th>KRAS-204</th>
</tr>
</thead>
<tbody>
<tr>
<td>AZD-4785</td>
<td>-6.9</td>
<td>-7.3</td>
<td>-7.2</td>
<td>-7.1</td>
<td>-7.2</td>
<td>-6.4</td>
<td>-7.9</td>
<td>-5.7</td>
<td>-4</td>
</tr>
<tr>
<td>AZD-8835</td>
<td>-8.3</td>
<td>-8.2</td>
<td>-8.2</td>
<td>-8.7</td>
<td>-7.9</td>
<td>-7.7</td>
<td>-8.5</td>
<td>-6.3</td>
<td>-4.9</td>
</tr>
<tr>
<td>CC-223</td>
<td>-7.2</td>
<td>-7.7</td>
<td>-7.5</td>
<td>-7.5</td>
<td>-7.7</td>
<td>-7.3</td>
<td>-7.7</td>
<td>-6.1</td>
<td>-4.5</td>
</tr>
<tr>
<td>PD-0325901</td>
<td>-7.7</td>
<td>-7.3</td>
<td>-6.9</td>
<td>-6.9</td>
<td>-7.4</td>
<td>-6.3</td>
<td>-6.8</td>
<td>-5.4</td>
<td>-4.3</td>
</tr>
<tr>
<td>Ridaforolimus</td>
<td>-10.1</td>
<td>-10.2</td>
<td>-9.7</td>
<td>-9.6</td>
<td>-10.6</td>
<td>-10.7</td>
<td>-9.6</td>
<td>-8.8</td>
<td>-5.3</td>
</tr>
<tr>
<td>Selumetinib</td>
<td>-7.2</td>
<td>-7.2</td>
<td>-7.6</td>
<td>-7.7</td>
<td>-7.1</td>
<td>-6.6</td>
<td>-7.4</td>
<td>-5.9</td>
<td>-4.8</td>
</tr>
<tr>
<td>Trametinib</td>
<td>-8.2</td>
<td>-9</td>
<td>-7.9</td>
<td>-9.6</td>
<td>-9</td>
<td>-8.6</td>
<td>-8.9</td>
<td>-6.9</td>
<td>-5.5</td>
</tr>
</tbody>
</table>

### Table 3: Binding affinity values of the PIK3CA-Canonical, PIK3CA-201, 205, 203, and 204

<table>
<thead>
<tr>
<th>Drugs</th>
<th>PIK3CA-Canonical</th>
<th>PIK3CA-201</th>
<th>PIK3CA-205</th>
<th>PIK3CA-203</th>
<th>PIK3CA-204</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC-223</td>
<td>-8.6</td>
<td>-8.3</td>
<td>-8</td>
<td>-6.4</td>
<td>-6.4</td>
</tr>
<tr>
<td>ALPELISIB</td>
<td>-8.9</td>
<td>-8.8</td>
<td>-9.5</td>
<td>-7.5</td>
<td>-7.7</td>
</tr>
<tr>
<td>BUPARLISIB</td>
<td>-8.6</td>
<td>-8.2</td>
<td>-8.1</td>
<td>-6.2</td>
<td>-6.4</td>
</tr>
<tr>
<td>CAPIVASERTIB</td>
<td>-9.5</td>
<td>-9.6</td>
<td>-8.9</td>
<td>-6.6</td>
<td>-6.6</td>
</tr>
<tr>
<td>INK-1117</td>
<td>-9</td>
<td>-9</td>
<td>-9</td>
<td>-6.8</td>
<td>-6.8</td>
</tr>
<tr>
<td>SERABELISIB</td>
<td>-8.9</td>
<td>-9.1</td>
<td>-9</td>
<td>-6.8</td>
<td>-6.8</td>
</tr>
<tr>
<td>TASELISIB</td>
<td>-9.7</td>
<td>-9.7</td>
<td>-8.2</td>
<td>-7.2</td>
<td>-6.9</td>
</tr>
<tr>
<td>TEMSIROLIMUS</td>
<td>-9.5</td>
<td>-11.6</td>
<td>-10.2</td>
<td>-9</td>
<td>-9</td>
</tr>
<tr>
<td>TRASTUZUMAB</td>
<td>-10.5</td>
<td>-9.6</td>
<td>-9.6</td>
<td>-6.8</td>
<td>-7.7</td>
</tr>
</tbody>
</table>

### Table 4: Hydrogen and Hydrophobic interactions of the docked protein isoforms with drugs

<table>
<thead>
<tr>
<th>Protein</th>
<th>Hydrogen Interactions</th>
<th>Hydrophobic Interactions</th>
</tr>
</thead>
<tbody>
<tr>
<td>PIK3CA-Canonical</td>
<td>GLU, ASN, ASP, ASP, TYR</td>
<td>THR</td>
</tr>
<tr>
<td>PIK3CA-201</td>
<td>ASP, ASN, LYS, SER</td>
<td>LYS, ASP, ASN, PRO, GLN</td>
</tr>
<tr>
<td>PIK3CA-205</td>
<td>ARG, ASP, ASP, LYS, PHE</td>
<td>GLU, TYR, LYS</td>
</tr>
<tr>
<td>PIK3CA-203</td>
<td>SER, THR</td>
<td>ARG, GLU</td>
</tr>
</tbody>
</table>
Figure 5: Ligand-binding pocket of PIK3CA protein isoforms: (A) Canonical Protein, (B) PIK3CA-201, (C) PIK3CA-205, and (D) PIK3CA-202 with the drug temsirolimus.

Figure 6: Ligplot analysis of the interactions between PIK3CA protein isoforms and temsirolimus. Hydrophobic interactions between amino acid residues are indicated by red arcs, whereas hydrogen bonds are represented by dashed green lines with specified bond lengths.
Discussion

Although current target prediction methods have shown the accuracy of genomic, chemical, and pharmacological data in drug target interaction prediction, these methods frequently concentrate only on the canonical protein while disregarding the on- or even off-target isoform-level interactions that are linked to the action of the chemical [53]. Previous studies have related cancer-specific aberrant splicing to drug resistance. However, little is known about the therapeutic effects of this drug on specific tissues and its side effects on other tissues. Gene isoforms produced by alternative splicing can be expressed at different levels and exhibit various, perhaps conflicting, activities in various tissues and/or organs [54, 55]. We postulated that various protein isoforms formed by alternative splicing might develop into candidates for off-target or non-target drug interactions due to the presence or lack of target binding sequences in different alternative splicing of genes specifically involved in cervical cancer. Our findings show that most small-molecule therapeutic targets have a variety of protein isoforms. Therefore, it is feasible that most pharmacologically targeting gene isoforms have functional differences and show isoform-level changes in their interactions with the drug.

We found that KRAS-203 is highly expressed in tumor samples. Sequence alignment and data analysis of the gene expression patterns in the TCGA and GTEx datasets uncovered significant data, such as medicines that skip alternative gene isoforms that are also expressed in cancer but perhaps are not targeted, while the drugs that might possibly target alternative protein isoforms that are variously expressed across many normal tissues and are involved in the process of cancer development. Furthermore, the ability of the same medication to bind to several structurally related protein isoforms with various affinities was verified using a drug docking study and structural analysis of KRAS and PIK3CA proteins. These findings are basically two processes in which both could possibly lead to far-off impacts that could result in drug resistance.

In comparison to the canonical isoform, we observed low KRAS isoform expression in TCGA samples. We observed, via structural docking, that various medicines can interact with all protein isoforms in various ways. It remains unknown whether the secondary protein isoforms behave similarly to or differently from the downregulated primary isoform, carcinogenic, or overexpressed. In contrast, the different protein isoforms, with the exception of KRAS-204, which was not expressed in normal or tumor samples, showed variable and greater expression in healthy tissues than in tumor tissues. These protein isoforms can act as tumor suppressors or regulators, counteracting the functions of carcinogenic protein isoforms. The immediate inhibition of these protein isoforms may be undesirable under these conditions. Although the precise roles of these protein isoforms are still unknown, separating sites from non-targets at the splice level is a crucial step in the early stages of drug discovery.

Owing to restrictions on the availability of data, we were challenged by several limitations in the current study. The first is the lack of mapping of gene isoforms between public online databases and older studies. For example, differences in exon numbers are frequently reported between these two sources. Public databases such as Ensemble do not contain many gene isoforms previously described in the literature. This makes it extremely challenging to annotate these gene and protein isoforms both structurally and functionally. Therefore, the major aspects of our study are that the overexpression of protein isoforms that are more advantageous for the development of cancer should be suppressed, and the main aims for suppression should be the gene isoforms that are upregulated in cancer. This is obviously a restriction because these two hypotheses might be incorrect; however, we do not currently have any better methods for evaluating the roles of these unidentified protein isoforms. Furthermore, the inclusion of the actual gene-level expression of these gene isoforms will strengthen this claim. To our knowledge, there is currently no comprehensive database that includes the expression of all the protein isoforms on a complete proteome scale. In our opinion, the importance of identifying pharmacological targets at the protein isoform level should be emphasized. However, our results add to those of a recent study that identified mean mRNA expression across tissues and variance in expression across tissues as the two key characteristics that separate effective medications from ineffective ones [56].

Conclusions

This study highlights the potential risks of focusing solely on the canonical isoform, and ignoring the impact of cervical cancer drugs on- and off-target effects at the isoform level. Identifying additional cancer biomarkers at the isoform level and connecting them to treatment sensitivity using
computational methods is crucial. Our findings indicate that the protein isoforms have distinct binding pocket confirmations, which indicate the potential variations in drug binding and efficacy. Some isoforms completely lack the binding pocket, which highlights the importance of considering the drugs effectiveness across the isoforms. Some isoforms were found to be upregulated in tumor samples, suggesting them as potential therapeutic targets. Molecular docking analysis revealed that protein isoforms have varying binding affinity with FDA approved drugs, which is essential to predict the possibility of designing protein isoform-level medication. Sufficient structural and functional knowledge of these isoforms is necessary to achieve this goal.

**Interests disclosure**

The authors have no conflicts of interest to declare that are relevant to the content of this article.

**Funding**

None.

**References**


Це дослідження підкреслює потенційні ризики зосередження виключно на канонічній ізоформі й ігнорування цільових мішеней на ранніх стадіях розробки ліків є вкрай важливим для боротьби з цією хворобою. Однак якщо залишити наявність різних ізоформ білка поза увагою, то можуть виникнути небажані терапевтичні або шкідливі побічні ефекти.

**КОНCLUSION.**

Це дослідження підкреслює потенційні ризики зосередження виключно на канонічній ізоформі й ігнорування цільових мішеней на ранніх стадіях розробки ліків є вкрай важливим для боротьби з цією хворобою. Однак якщо залишити наявність різних ізоформ білка поза увагою, то можуть виникнути небажані терапевтичні або шкідливі побічні ефекти.

**Ключові слова:** рак шийки матки; ізоформи; молекулярний докінг; аналіз взаємодії; біоінформатика.