Article

Integrating In Silico and In Vitro Approaches to Identify Natural Peptides with Selective Cytotoxicity against Cancer Cells

Hui-Ju Kao [1,2], Tzu-Han Weng [3], Chia-Hung Chen [1,2], Yu-Chi Chen [1,2], Yu-Hsiang Chi [4], Kai-Yao Huang [1,2,5,6,*] and Shun-Long Weng [5,7,8,*]

1   Department of Medical Research, Hsinchu MacKay Memorial Hospital, Hsinchu City 300, Taiwan
2   Department of Medical Research, Hsinchu Municipal MacKay Children's Hospital, Hsinchu City 300, Taiwan
3   Department of Dermatology, MacKay Memorial Hospital, Taipei City 104, Taiwan
4   National Center for High-Performance Computing, Hsinchu City 300, Taiwan
5   Department of Medicine, MacKay Medical College, New Taipei City 252, Taiwan
6   Institute of Biomedical Sciences, MacKay Medical College, New Taipei City 252, Taiwan
7   Department of Obstetrics and Gynecology, Hsinchu MacKay Memorial Hospital, Hsinchu City 300, Taiwan
8   Department of Obstetrics and Gynecology, Hsinchu Municipal MacKay Children's Hospital,
     Hsinchu City 300, Taiwan
*   Correspondence: kyh.l765@mmh.org.tw (K.-Y.H.); 4467@mmh.org.tw (S.-L.W.)

Abstract: Anticancer peptides (ACPs) are bioactive compounds known for their selective cytotoxicity against tumor cells via various mechanisms. Recent studies have demonstrated that in silico machine learning methods are effective in predicting peptides with anticancer activity. In this study, we collected and analyzed over a thousand experimentally verified ACPs, specifically targeting peptides derived from natural sources. We developed a precise prediction model based on their sequence and structural features, and the model's evaluation results suggest its strong predictive ability for anticancer activity. To enhance reliability, we integrated the results of this model with those from other available methods. In total, we identified 176 potential ACPs, some of which were synthesized and further evaluated using the MTT colorimetric assay. All of these putative ACPs exhibited significant anticancer effects and selective cytotoxicity against specific tumor cells. In summary, we present a strategy for identifying and characterizing natural peptides with selective cytotoxicity against cancer cells, which could serve as novel therapeutic agents. Our prediction model can effectively screen new molecules for potential anticancer activity, and the results from in vitro experiments provide compelling evidence of the candidates' anticancer effects and selective cytotoxicity.

Keywords: anticancer peptide; antitumor peptide; anticancer activity; selective cytotoxicity; in silico analysis; in vitro experiments; machine learning

1. Introduction

Cancer remains a significant global health burden in the 21st century, ranking among the leading causes of death alongside cardiovascular diseases in many countries [1]. Annually, tens of millions of individuals receive cancer diagnoses worldwide, and nearly half of them succumb to the disease [2]. While surgical resection is a traditional and effective therapy for many cancer types, various additional treatments have been developed to reduce cancer cell growth and progression. These include radiation therapy, chemotherapy, immunotherapy, hormone therapy, targeted therapy, and other approaches [3]. Among them, chemotherapy, commonly known as "chemo", is the most prevalent treatment that uses drugs to slow cancer growth or eliminate cancer cells [4,5]. Multiple chemotherapy drugs are in clinical use today due to the tendency of cancer cells to grow and divide faster than healthy cells, making them ideal targets for these drugs. However, chemotherapy often damages non-cancerous cells, causing side effects like fatigue, hair loss, lung tissue damage, cardiac and renal problems, peripheral neuropathy, and infertility [6].

Anticancer peptides (ACPs), typically consisting of 10–50 amino acids, are bioactive molecules that induce cytotoxicity against cancer cells by disrupting and penetrating cell or organelle membranes. Their selective cytotoxicity has been observed in various cancers, positioning them as potential novel antineoplastic agents [7]. One notable difference between tumor and healthy cells lies in the membrane's electrical properties. Tumor cells secrete large amounts of lactate through glucose and glutamine metabolism, which results in a negatively charged surface [8,9]. ACPs leverage this difference to disrupt cancer cell membranes via electrostatic interactions with their anionic components, thereby selectively lysing cancer cells [10–12]. Compared to antibodies and small molecules, ACPs are increasingly viewed as effective and safer alternatives to chemotherapy, offering high selectivity, penetration, and easy modification. With cancer remaining a leading cause of death globally, ACPs are attracting attention for their clinical potential as a new class of antineoplastic drugs.

In recent years, in silico approaches have been employed to identify peptides with cytotoxicity against cancer cells. In 2013, Tyagi et al. introduced the AntiCP [13] model, which discriminates between ACPs and non-ACPs based on limited available data. Hajisharifi et al. later improved ACP prediction by combining Chou's pseudo-amino acid composition with other sequence features [14]. Vijayakumar and Lakshmi developed a novel feature encoding method that identifies apoptotic domains in a peptide, enhancing sensitivity in ACP detection [15]. Chen et al. created the iACP web tool, using a feature selection algorithm to identify key features for ACP prediction [16]. Li and Wang investigated the correlation between anticancer activity and amino acid sequence properties, including amino acid composition, average chemical shift, and reduced amino acid composition [17].

Other researchers have developed models using genetic algorithms (GAs), SMOTE (Synthetic Minority Oversampling Technique), and support vector machines (SVMs) to enhance ACP prediction [18,19]. Wei et al. created ACPred-FL [20] using the minimum redundancy maximum relevance (mRMR) method to select informative features, significantly improving predictive performance. ACPred [21] used SVMs coupled with amino acid and amphiphilic pseudo-amino acid compositions, revealing that hydrophobic residues in the α-helix and cysteine residues in the β-sheet structures correlate with anticancer activity. mACPpred [22] was created based on selected physicochemical and compositional properties. Recently, AntiCP 2.0 [23], a refined version of the original model, was trained using amino acid composition and an ETree classifier, achieving state-of-the-art accuracy.

Despite numerous methods for ACP identification, comprehensive evaluation via in vitro or in vivo cytotoxicity assays remains limited. Therefore, a robust analysis platform is needed that provides high-accuracy prediction and experimental validation. In this study, we aim to develop a strategy for identifying natural peptides with cytotoxicity against cancer cells through a combined computational and experimental approach.

2. Results

The workflow, outlined in Figure 1, includes several key steps: data collection and preprocessing, analysis of ACP features, construction of a novel prediction model, evaluation of the model's performance, identification of natural candidate ACPs through multiple predictive tools, and validation of ACPs' selective cytotoxicity against specific cancer cells. The details of each step are described below.

2.1. Data Collection and Preprocessing of Peptides with Anticancer Activity

As summarized in Table 1, a total of 1462 experimentally verified ACP sequences were gathered from the literature [14,20,22,24] and public tools and databases including ACPred [21], ACPred-FL [20], AntiCP [13], AntiCP2 [23], APD3 [25], CAMP [26], CancerPPD [27], DADP [28], dbAMP [29], DRAMP [30], EnACP [24], mACPpred [22], and SATPdb [31]. An additional 2875 non-ACP sequences were sourced from existing tools for predicting ACPs, such as ACPred [21], ACPred-FL [20], AntiCP [13], AntiCP2 [23], EnACP [24], and mACPpred [22].
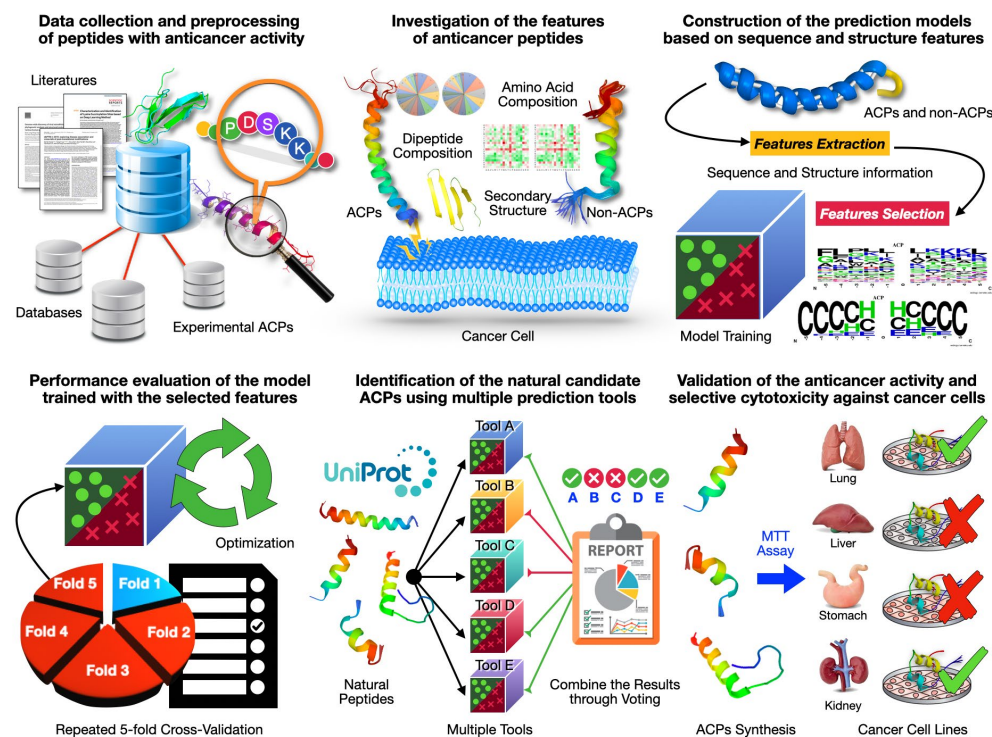
Figure 1. Flowchart depicting process for identifying natural peptides with selective cytotoxicity against cancer cells.

Table 1. Data statistics for the training and testing datasets.

Dataset	Number of ACPs	Number of Non-ACPs
Raw data	1462	2875
Length 10–50 aa	1344	2361
Training dataset	804	1494
Testing dataset	41,489 natural peptides	

To prevent overfitting, where the model becomes too well-suited to the training data, redundant peptide sequences were removed from both positive and negative datasets. Only sequences between 10 and 50 residues long were retained. The remaining dataset comprised 804 ACP sequences and 1494 non-ACP sequences, resulting in a 1:2 ratio of positive to negative sequences. At the time of analysis, this was the most comprehensive data available.

To identify potential ACPs derived from natural sources, peptides ranging from 10 to 50 amino acids in length without an anticancer activity annotation were extracted from the Universal Protein Knowledgebase (UniProtKB) [32], resulting in 41,489 sequences used as the testing dataset.

2.2. Investigation of Sequence and Structural Features of Anticancer Peptides

Several studies have shown that sequence-based features are highly effective for predicting protein functions. In this research, amino acid composition (AAC) [33], dipeptides composition (DPC) [34], and k-spaced amino acid pairs (CKSAAPs) [35] were employed to differentiate ACPs from non-ACPs. After preprocessing, the occurrence frequencies of the 20 amino acids were calculated to identify consensus motifs in ACP sequences. Figure 2 compares the composition of essential amino acids between ACP and non-ACP sequences, revealing that aliphatic residues glycine (G) and leucine (L) are enriched in ACPs. Aromatic amino acids phenylalanine (F) and tryptophan (W) are also more frequent in ACPs compared to non-ACPs. Notably, lysine (K), a basic amino acid, shows the most statistically

significant difference in frequency. This indicates that ACPs primarily interact with cancer cells through electrostatic interactions with anionic phospholipids in the plasma membrane. This is a key mechanism through which ACPs disrupt membrane integrity, leading to the leakage of cellular contents [7,36].
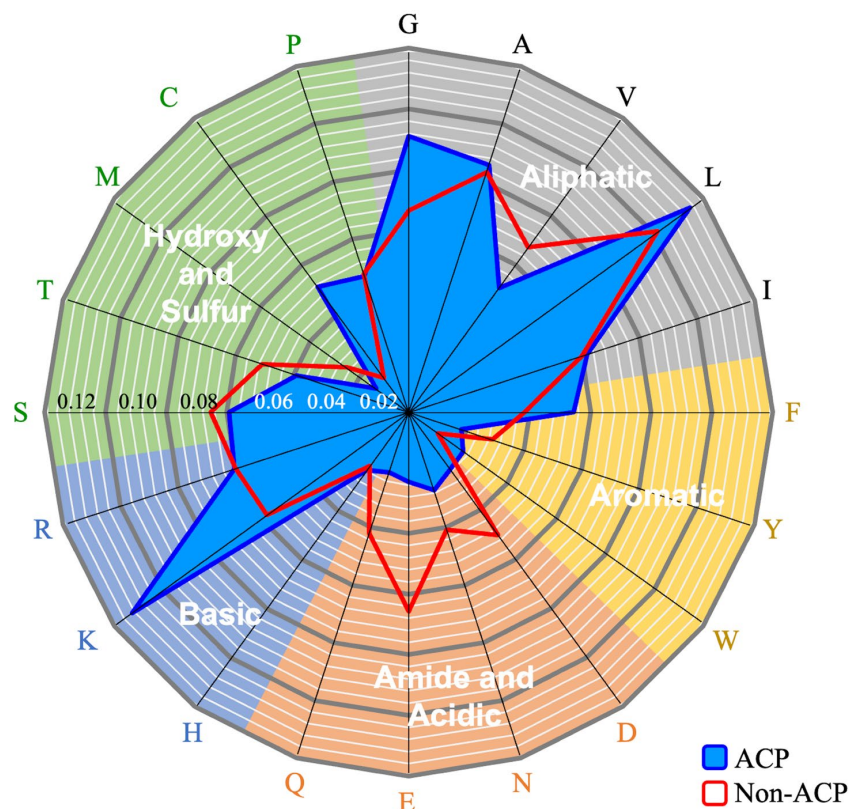

Figure 2. Comparison of amino acid compositions between ACPs and non-ACPs. The uppercase letters around the image represent various amino acids. The blue lines represent the amino acid compositions of ACPs, while the red lines represent those of non-ACPs.

Additionally, cysteine (C), an amino acid with both polar and hydrophobic properties, plays a critical role in protein structure and stability and is more frequently found in ACPs. Studies have shown that cysteine-rich, cationic peptides with antimicrobial activity also exhibit cytostatic effects against cancer cells [37–40]. Many of these ACPs, including defensins and bacteriocins, have been reported to show low cytotoxic and hemolytic effects on normal cells [41]. Conversely, amide and acidic amino acids like asparagine (N), glutamine (Q), aspartic acid (D), and glutamic acid (E) are polar and negatively charged at physiological pH. The results show that these amino acids are less prevalent in ACPs.

Analyzing amino acid pairs helps estimate the significance of different combinations and their characteristics. For each peptide sequence, the composition of amino acid pairs was measured at k-spaced intervals of zero, one, two, and three residues. Figure 3 shows the frequency differences of 400 k-spaced amino acid pairs between ACPs and non-ACPs using 20 × 20 matrices, highlighting enriched and suppressed pairs in red and green, respectively. At zero spacing (k = 0), the pairs correspond to dipeptides and are enriched with aliphatic (G, A, V, I, L) and basic (K, R, H) amino acids, such as AK, RR, GG, GL, GK, LA, LL, LK, KA, KI, KL, and KK. When k = 1, pairs like AxK, GxL, IxK, LxK, KxA, KxL, and FxK are significantly different between ACPs and non-ACPs. At k = 2, pairs such as AxxK, GxxC, LxxL, LxxK, KxxA, KxxL, and KxxK show marked differences. At k = 3, the pairs AxxxA, CxxxC, GxxxK, LxxxA, LxxxL, KxxxK, and FxxxL are enriched in ACPs. The presence of sulfur-containing cysteine in various combinations across all k-spacings indicates the importance of these pairs for distinguishing ACPs.
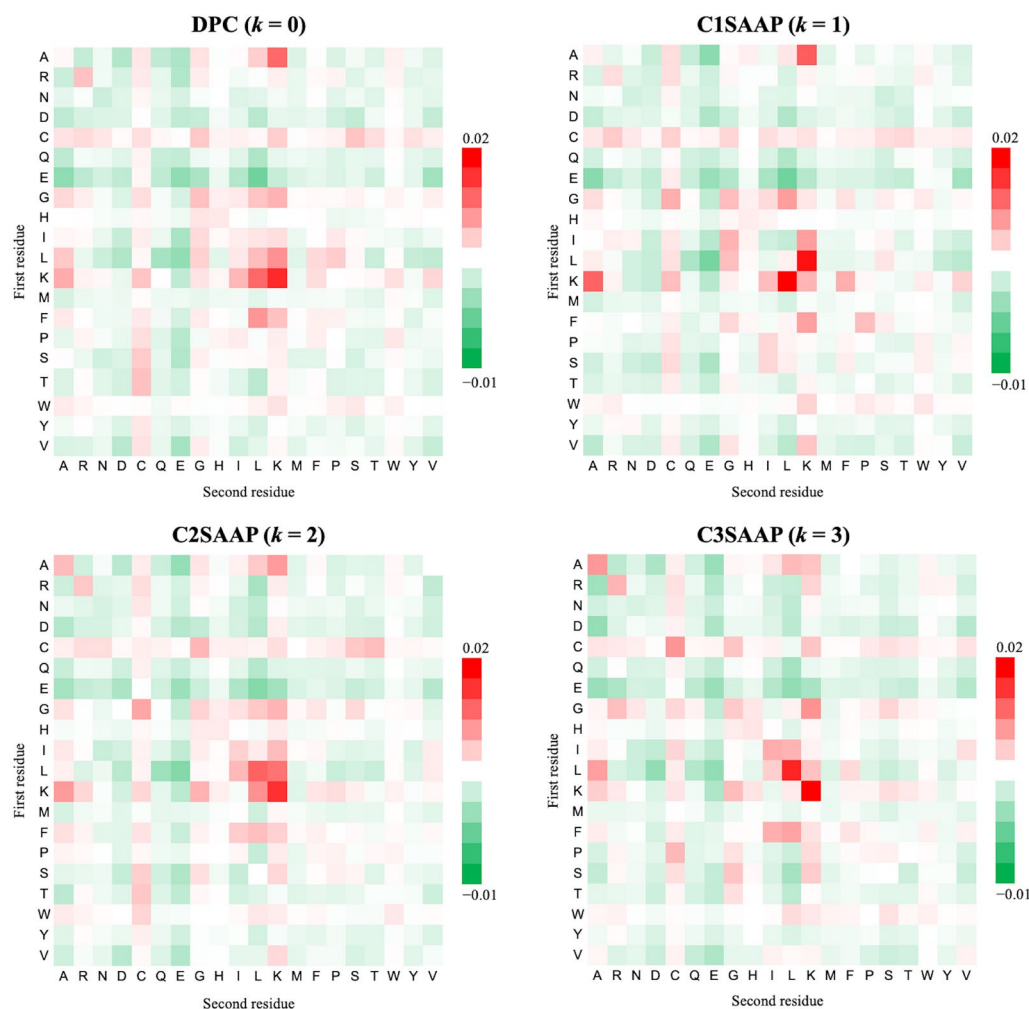
Figure 3. Comparison of frequencies of occurrence of 20 × 20 amino acid pairs separated by k residues between ACPs and non-ACPs.

Protein folding, which involves the number, spatial arrangement, and connectivity of secondary structure elements, plays a crucial role in biological functions [42]. Using the PEP2D tool [43], the secondary structure elements composition (SSEC) was predicted for each peptide. Figure 4 compares the secondary structure compositions between ACPs and non-ACPs, revealing that ACPs are composed of 56.9% random coils, 31.9% alpha-helices, and 11.2% beta-strands. The results show a significant difference in coil structures between ACPs and non-ACPs, with ACPs showing fewer helices and more beta-strands. Furthermore, the SSEC of the first and last 10 residues were analyzed separately. The C-terminus of ACPs contains more beta-strands and fewer helices and coils compared to non-ACPs, while no significant difference was found at the N-terminus. Studies investigating ACP structure and biological activity [44–48] suggest that the alpha-helical structure plays a crucial role in the anticancer effects and selective cytotoxicity of ACPs against cancer cells [49–51].

Finally, the amino acid and structural element compositions at the N- and C-terminus of ACPs and non-ACPs were compared using the TwoSampleLogo version 1.21 [52]. Figure 5 shows position-specific AAC and SSEC for the first and last five amino acids. Positively charged amino acids lysine (K) and arginine (R) are enriched at the C-terminus of ACPs, while nonpolar residues like phenylalanine (F), leucine (L), tryptophan (W), and proline (P) are particularly abundant at the N-terminus. Previous studies [13,53] suggest that the positively charged C-terminus plays a significant role in affecting tumor growth and progression. In addition, position-specific SSEC analysis indicates a greater prevalence of C-terminal beta-strands in ACPs compared to non-ACPs.
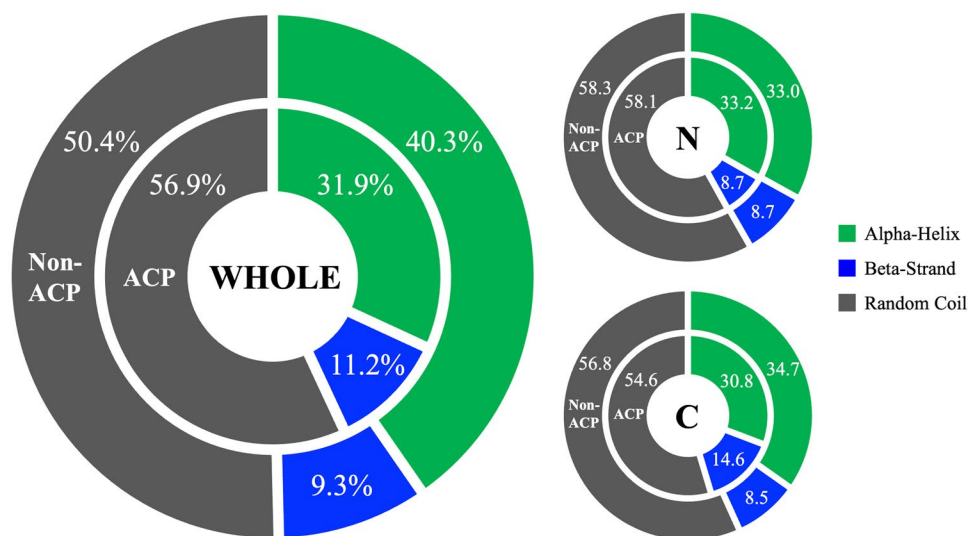
Figure 4. Comparison of compositions of secondary structure elements between ACPs and non-ACPs.
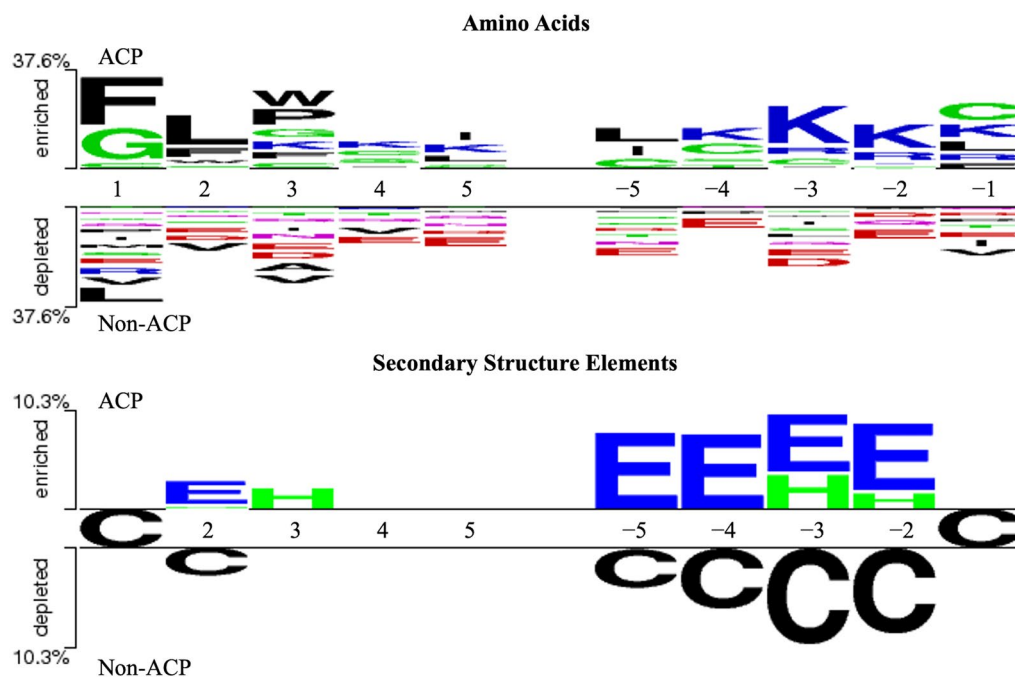

Figure 5. A comparison of the composition of amino acids and secondary structure elements at the N- and C-terminal regions between ACPs and non-ACPs. The uppercase letters in the upper part of figure represent the amino acids in the peptide sequences, with blue indicating positively charged amino acids and red indicating negatively charged amino acids. In the lower part, the uppercase letters represent the secondary structure of the peptide sequences, where H stands for alpha-helix, E stands for beta-sheet, and C stands for random coil.

Overall, the amino acid compositions and conformations of the C-terminal region appear to play a critical role in determining a peptide's ability to suppress cancer cells.

2.3. Construction of Prediction Models Based on Sequence and Structural Features

To evaluate the discrimination capability of the investigated features for distinguishing ACPs from non-ACPs, we trained models using each feature subset and validated them through five repetitions of five-fold cross-validation. Each peptide sequence was encoded using different feature encoding methods, including AAC, DPC, C1SAAP, C2SAAP,

C3SAAP, SSEC, N-AAC, N-SSEC, C-AAC, and C-SSEC. The LIBSVM tool [54] was used to build the SVM prediction models. Table 2 presents the results, where the AAC model achieved satisfactory results, with a sensitivity of 89.00%, specificity of 89.48%, accuracy of 89.31%, and a Matthews correlation coefficient (MCC) of 0.77 in distinguishing ACPs from non-ACPs. The models based on CKSAAP features demonstrated exceptional performance, with the C1SAAP model delivering the best results, showing a sensitivity of 90.00%, specificity of 90.09%, accuracy of 90.06%, and an MCC of 0.79.

Table 2. Results from five-fold cross-validation experiments for models trained with single feature sets.

Model	Sensitivity (%)	Specificity (%)	Accuracy (%)	MCC *
AAC	89.00 ± 0.0038	89.48 ± 0.0030	89.31 ± 0.0022	0.77 ± 0.0046
DPC	88.71 ± 0.0030	88.96 ± 0.0025	88.87 ± 0.0008	0.76 ± 0.0014
C1SAAP	90.00 ± 0.0017	90.09 ± 0.0035	90.06 ± 0.0022	0.79 ± 0.0042
C2SAAP	89.70 ± 0.0046	89.53 ± 0.0044	89.59 ± 0.0033	0.78 ± 0.0068
C3SAAP	89.10 ± 0.0014	90.58 ± 0.0025	90.06 ± 0.0016	0.79 ± 0.0033
SSEC	62.29 ± 0.0140	64.08 ± 0.0167	63.46 ± 0.0070	0.25 ± 0.0083
N-AAC	73.26 ± 0.0052	74.67 ± 0.0055	74.18 ± 0.0036	0.46 ± 0.0065
N-SSEC	50.92 ± 0.1771	50.67 ± 0.1782	50.76 ± 0.0539	0.02 ± 0.0027
C-AAC	69.18 ± 0.0077	70.08 ± 0.0048	69.77 ± 0.0032	0.38 ± 0.0069
C-SSEC	55.50 ± 0.0997	54.42 ± 0.1230	54.80 ± 0.0468	0.10 ± 0.0344

* MCC: Matthews correlation coefficient. The values represent the mean and standard deviation of all measurements.

Unfortunately, the model trained using SSEC features could not effectively distinguish ACPs from non-ACPs, resulting in suboptimal performance, with a sensitivity of 62.29%, specificity of 64.08%, accuracy of 63.46%, and an MCC of 0.25. Additionally, the models trained on N- or C-terminal amino acid or secondary structure element compositions also yielded subpar sensitivity values (all below 70%) except for the N-AAC model, which performed slightly better.

These findings suggest that sequence-based features are valuable for characterizing peptides with anticancer activity. However, secondary structure elements provide limited predictive power, possibly because they were approximated as substitutions.

2.4. Performance Evaluation of Model Trained by Hybrid Feature Sets

Based on previous results, models trained using sequence-based features demonstrated efficient performance in classification on the training dataset. However, according to prior research [22,23,55], models that incorporate hybrid feature sets generally achieve higher average accuracy than those utilizing individual features. Consequently, to enhance predictive capability, these features were combined both additively and in a more integrated manner and applied to the SVM classifier.

As depicted in Table 3, the models that integrated sequence and structural characteristics showed improved performance. The model utilizing a combination of AAC and DPC features achieved a sensitivity of 91.02%, a specificity of 90.12%, an accuracy of 90.44%, and an MCC of 0.80. Remarkably, the model combining AAC, DPC, and CKSAAP delivered the best overall performance, with a sensitivity of 91.17%, a specificity of 90.83%, an accuracy of 90.95%, and an MCC of 0.81. Although combining multiple sequence-based features enhanced classification performance, models that integrated both sequence and structural features still showed less satisfactory results. Specifically, the model combining AAC, DPC, and SSEC yielded slightly reduced predictive performance, with a sensitivity of 84.3%, specificity of 85.07%, accuracy of 84.8%, and an MCC of 0.68. Similar results were observed when adding CKSAAP and SSEC to the combination.

Table 3. Results from five-fold cross-validation experiments for models trained with hybrid feature sets.

Model	Sensitivity (%)	Specificity (%)	Accuracy (%)	MCC *
AAC + DPC	91.02 ± 0.0025	90.12 ± 0.0047	90.44 ± 0.0037	0.80 ± 0.0074
AAC + DPC + SSEC	84.30 ± 0.0069	85.07 ± 0.0028	84.80 ± 0.0036	0.68 ± 0.0080
AAC + DPC + CKSAAP	91.17 ± 0.0037	90.83 ± 0.0032	90.95 ± 0.0021	0.81 ± 0.0043
AAC + DPC + CKSAAP + SSEC	86.52 ± 0.0047	87.07 ± 0.0033	86.88 ± 0.0015	0.72 ± 0.0030

* MCC: Matthews correlation coefficient. The values represent the mean and standard deviation of all measurements.

Five-fold cross-validation was used, and Figure 6 illustrates the comparison of receiver operating characteristic (ROC) curves between the SVM models trained using all feature sets. The area under the ROC curve (AUC) for each model was measured. In summary, the model trained by combining sequence-based features such as AAC, DPC, and CKSAAP significantly enhances the predictive performance for distinguishing between ACPs and non-ACPs.

Model	AUC
AAC	94.9%
DPC	94.8%
C1SAAP	95.4%
SSC	67.1%
AAC+DPC	96.0%
AAC+DPC+SSC	91.0%
AAC+DPC+CKSAAP	96.4%
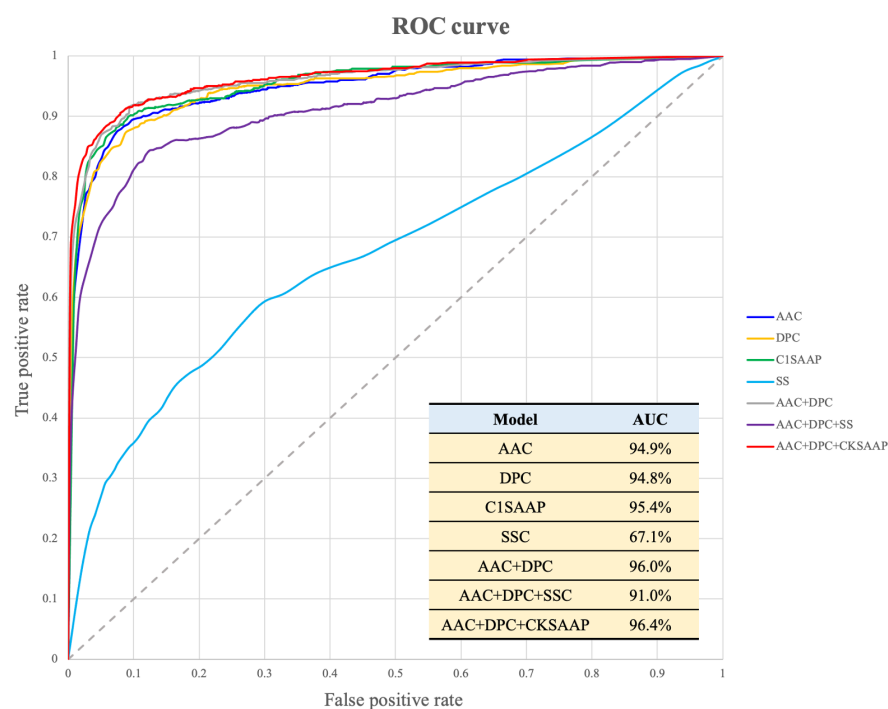
Figure 6. The ROC curves of models trained by sequence and structure-based features based on the results of five-fold cross-validation experiments.

2.5. Identification of Natural ACPs by Integrating Multiple Tools

Antimicrobial peptides have been discovered across a broad range of life forms; however, only a small number of peptides with anticancer activity have been identified and validated through biological experiments. We propose a strategy for identifying anticancer peptides derived from the natural environment, which may offer cancer treatment with fewer complications. To achieve more precise identification, various approaches were used to predict anticancer activity from the natural peptide dataset sourced from UniProtKB [32]. These tools include ACPred [21], ACPred-FL [20], AntiCP [13], AntiCP2 [23], iACP [16], mACPpred [22], and our proposed model.

As outlined in the methods section, a peptide was considered a candidate if it was predicted as a positive case by all the aforementioned tools. In total, 176 natural peptides out of 41,489 were considered potential ACP candidates (Supplementary Table S1). Table 4

lists the top 20 candidates with the highest probability, with many of these peptides originating from plants, notably from species such as Oldenlandia affinis (OLDAF), Chassalia parviflora (CHAPA), Psychotria brachyceras (PSYBR), and Psychotria leiocarpa (PSYLE) in the Rubiaceae family, as well as Viola odorata (VIOOD), Viola hederacea (VIOHE), Viola inconspicua (VIOIN), Melicytus dentatus (MELDN), and Melicytus chathamicus (MELCT) in the Violaceae family. Additional candidates were identified in amphibians from the Viperidae family, such as Crotalus durissus ruruima (CRODR), Crotalus viridis (CROW), and Crotalus durissus terrificus (CRODU), and in species of tree frogs like Phyllomedusa trinitatis (PHYTB) and Ranoidea caerulea (RANCA), as well as in bees and bacteria.

Furthermore, functional enrichment analysis was performed to identify the biological themes present in the candidate ACPs. As shown in Figure 7, the results indicated that these candidates were significantly enriched in Gene Ontology (GO) [56] terms related to defense responses to bacteria and fungi, the killing of cells from other organisms, cell cytolysis, degranulation, and hemolysis within the biological process (BP) category. The significantly enriched GO terms in the cellular component (CC) category included extracellular region and membrane, and for molecular function (MF), the enriched terms involved toxin and hormone activity. These findings suggest that these peptides may play a crucial role in regulating defense responses against various pathogens, likely aiding in the fight against harmful bacteria and cancer cells.
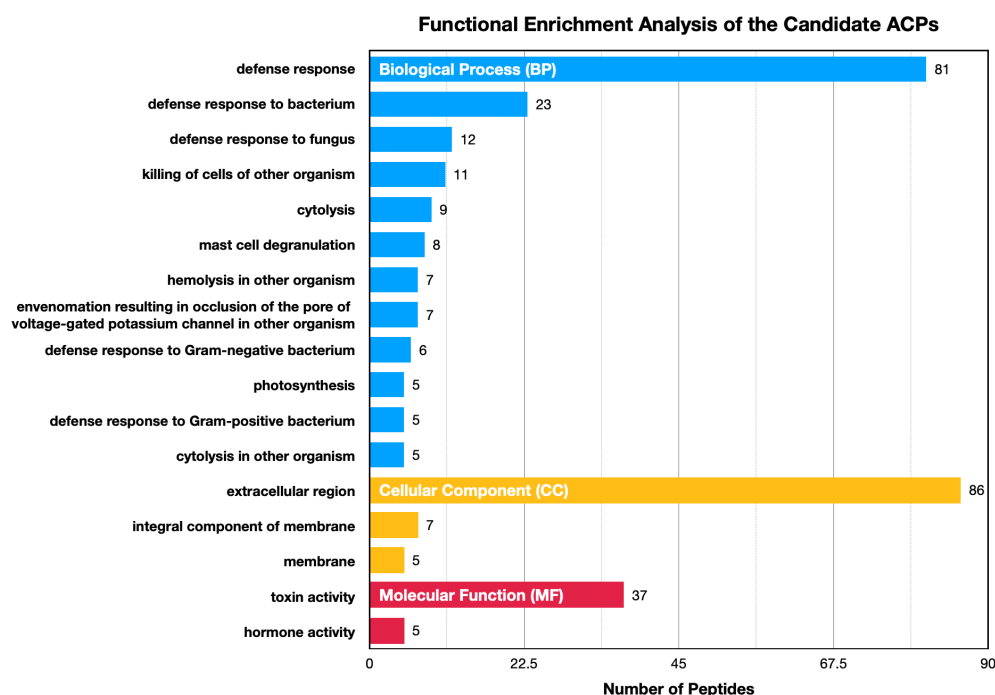

Figure 7. Functional enrichment analysis for candidate ACPs highlighting significant GO terms.

Table 4. Top 20 potential natural ACP candidates ranked by probability.

Entry Name	Sequence	ACPred	ACPred-FL	AntiCP	AntiCP2	iACP	mACPpred	Our Model
KAB4_OLDAF	GLPVCGETCVGGTCNTPGCTCSWPVCTRD	98.60%	98.11%	72.58%	96.00%	99.73%	98.17%	99.60%
CYO22_VIOOD	GLPICGETCVGGTCNTPGCTCSWPVCTRN	99.50%	95.12%	71.77%	95.00%	99.90%	98.42%	99.70%
THN2_VISAL	KSCCPNTTGRNIYNTCRFGGGSREVCASLSGCKIISASTCPSYPDK	99.50%	99.22%	70.56%	94.00%	99.52%	96.51%	96.60%
CYH3_VIOHE	GLPVCGETCFGGTCNTPGCICDPWPVCTRN	98.70%	92.89%	71.77%	95.00%	99.80%	98.80%	99.50%
MYX_CRODR	YKQCHKKGGHCFPKEKICIPPSSDFGKMDCRWRWKCCKKGSG	99.60%	99.22%	70.56%	91.00%	99.83%	94.71%	95.10%
KAB10_OLDAF	GLPTCGETCFGGTCNTPGCSCSSWPICTRD	99.40%	98.11%	70.56%	90.00%	99.93%	98.48%	99.60%
PROTO_POLPI	ILGTILGLLKSL	97.80%	99.22%	95.16%	54.00%	88.85%	98.35%	99.50%
CYO23_VIOOD	GLPTCGETCFGGTCNTPGCTCDSSWPICTHN	99.70%	98.11%	70.56%	87.00%	99.93%	98.47%	99.60%
CYPLE_PSYLE	SVTPIVCGETCFGGTCNTPGCSCSWPICTK	99.90%	99.22%	68.95%	87.00%	99.97%	96.86%	99.70%
CYPLD_PSYBR	GLPVCGESCFGGTCNTPGCSCTWPVCTRD	98.10%	95.12%	72.18%	87.00%	98.28%	98.01%	99.50%
ATOX_PHYTB	LTWKIPTRFCGVT	91.90%	99.22%	83.47%	50.00%	96.44%	96.64%	91.10%
CR12_RANCA	GLLGVLGSVAKHVLPHVVPVIAEHL	99.30%	98.11%	70.16%	84.00%	99.79%	98.62%	86.10%
KAB14_OLDAF	GLPVCGESCFGGTCNTPGCACDPWPVCTRD	88.70%	83.42%	71.77%	86.00%	99.76%	97.94%	99.00%
CYPLC_PSYLE	GDLPVCGETCFGGTCNTPGCVCAWPVCTR	95.70%	98.11%	68.15%	83.00%	99.25%	98.21%	99.40%
CYPLB_PSYLE	GDLPICGETCFGGTCNTPGCVCAWPVCNR	95.10%	98.11%	67.74%	83.00%	99.43%	97.87%	99.50%
GRAB_GRASX	IGGIISFFKRLF	100.00%	99.22%	85.08%	69.00%	82.43%	96.20%	98.90%
CIRF_CHAPA	AIPCGESCVWIPCISAAIGCSCKNKVCYR	99.60%	82.67%	75.81%	89.00%	99.80%	98.54%	99.50%
CYVNA_VIOIN	GIPVCGETCTLGTCYTAGCSCSWPVCTRN	99.60%	98.11%	71.37%	82.00%	99.80%	98.12%	99.50%
PNG1_PANCL	LNWGAILKHIIK	99.90%	99.22%	81.85%	58.00%	99.20%	98.22%	99.00%
PSMA3_STAAN	MEFVAKLFKFFKDLLGKFLGNN	98.60%	97.98%	75.81%	82.00%	96.15%	87.97%	85.70%

2.6. Validating the Selective Cytotoxicity of ACPs against Specific Cancer Cells

Although the anticancer activity of some peptides has been confirmed through in vitro experiments, previous studies often evaluated the cytotoxic effects in only a few cell lines, without exploring the underlying mechanisms of action in depth. To validate the anticancer activity and selective cytotoxicity of the predicted ACPs, we synthesized and tested ten putative ACPs, labeled 1 to 10, for their effects on inhibiting cell proliferation using over 30 cancer cell lines. This evaluation employed the MTT colorimetric assay and included a diverse array of cancers such as skin, lung, colon, liver, breast, stomach, endometrial, ovarian, hypopharyngeal, lymphoma, pancreatic, fibrosarcoma, prostate, brain, oral cavity, and bone cancer, as summarized in Table 5. Additionally, two peptides, numbered 11 and 12, which were highly ranked in our model but not in others, were synthesized for comparative analysis.

Table 5. List of putative ACPs selected for validation experiments.

Peptide	UniProt ID	Length	Sequence
ACP1	KAB4_OLDAF	29	GLPVCGETCVGGTCNTPGCTCSWPVCTRD
ACP2	CIRF_CHAPA	29	AIPCGESCVWIPCISAAIGCSCKNKVCYR
ACP3	PSMA3_STAAN	22	MEFVAKLFKFFKDLLGKFLGNN
ACP4	CYMEK_MELDN	31	GSIPCGESCVWIPCISSVVGCACKNKVCYKN
ACP5	CYVNA_VIOIN	29	GIPVCGETCTLGTCYTAGCSCSWPVCTRN
ACP6	CIRB_CHAPA	31	GVIPCGESCVFIPCISTLLGCSCKNKVCYRN
ACP7	THN2_VISAL	46	KSCCPNTTGRNIYNTCRFGGGSREVCASLSGCKIISASTCPSYPDK
ACP8	MYX_CRODR	42	YKQCHKKGGHCFPKEKICIPPSSDFGKMDCRWRWKCCKKGSG
ACP9	CR12_RANCA	25	GLLGVLGSVAKHVLPHVVPVIAEHL
ACP10	CYPLE_PSYLE	30	SVTPIVCGETCFGGTCNTPGCSCSWPICTK
ACP11	UT114_PEA	15	EQQQQQQPQNRRFRE
ACP12	TL11_SPIOL	22	FKGGGPYGQGVTRGQDLSGKDF

As depicted in Figure 8 and Supplementary Table S2, the results confirmed the anticancer effects and selective cytotoxicity of these putative peptides against human cancer cells. The half-maximal inhibitory concentration (IC50) was calculated to measure the inhibitory capacity. Peptides 1 to 5 demonstrated broad cytotoxicity against various cancer cell lines even at concentrations below 50 μM, impacting cells such as A431, H1299, A549, HT29, HepG2, HEC-1-A, FaDu, HL-60, Daudi, Panc-1, and DU145. Peptides 6 and 7 showed notable anticancer effects, particularly on liver and breast cancers, and also affected endometrial and hypopharyngeal cancer cells. Peptide 8 was crucial in selectively targeting SKOV-3 ovarian carcinoma, HT1080 fibrosarcoma, DU145 prostate cancer, and DBTRG brain tumor cells. Peptide 9 was particularly effective against liver cancer cells, especially HepG2 and Mahlavu, while Peptide 10 showed pronounced cytotoxicity in BT474 cells. Interestingly, the DBTRG cell line was relatively resistant to these ACPs, except for Peptide 11, which exhibited specific cytotoxicity against brain tumor cells at a concentration of 32.12 μM. Peptide 12 also demonstrated anticancer activity not only against the triple-negative breast cancer cell line MDA-MB-231 but also in the Burkitt lymphoma-derived Daudi cell line.

Additionally, we have generated dose–response curves for the combination of each ACP and the cancer cell line that exhibited the most significant effect, as shown in Supplementary Figure S1. The anticancer mechanisms and effects of ACPs can differ significantly across various cancer cell lines due to the unique biological characteristics and microenvironments of these cells. Many studies have demonstrated that ACPs can exert a range of anticancer activities, such as inhibiting cell migration, suppressing angiogenesis, displaying antioxidant properties, halting cell proliferation, inducing apoptosis, and exerting cytotoxic effects [57,58]. This diversity in action mechanisms results in ACPs having selective efficacy against different types of cancer cells.
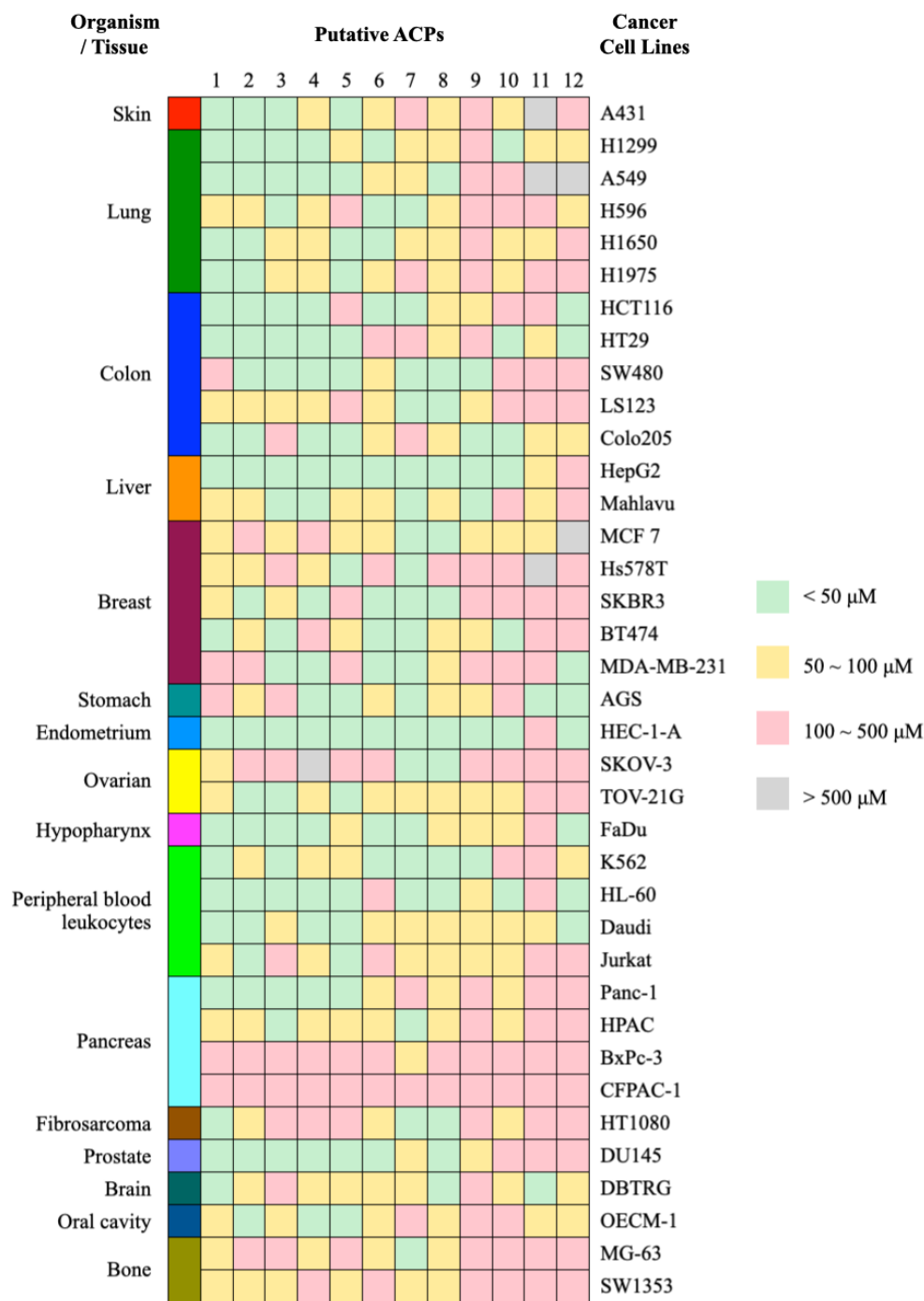
Figure 8. A validation of the anticancer activity of the putative ACPs against various cancer cell lines using the MTT colorimetric assay.

For instance, certain ACPs may show higher selectivity for cancer cells with a highly negatively charged cell membrane due to stronger electrostatic interactions between the peptides and the cancer cell membrane, leading to the targeted disruption of cancer cell membranes and the subsequent induction of cell death through necrosis or apoptosis [58]. Furthermore, the anticancer efficacy of ACPs is influenced by their amino acid composition, structural properties, hydrophobicity, and amphipathic nature, which enhance their interactions with cancer cell membranes [59].

In summary, the functional diversity and adaptability of ACPs result in varying levels of effectiveness in different cancer treatments, underscoring their potential for targeted

cancer therapies. However, further research is necessary to elucidate the mechanisms driving these reactions. Importantly, we have shown that our model can more accurately predict peptides that possess anticancer activity.

3. Discussion

In this study, we merged mathematical modeling with in vitro experiments across various cancer cell lines to investigate the anticancer activity and selective cytotoxicity of peptides. This approach provided a strategic framework for researchers to identify and characterize anticancer peptides derived from natural sources. We assembled the largest collection of experimentally validated ACPs in our training dataset compared to previous studies. The analysis of sequence features offered insights into the putative functions of ACPs, particularly through the comparison of k-spaced amino acid pairs between ACPs and non-ACPs. The cross-validation results from the training dataset confirmed the effective discrimination capability of the investigated features, with some models achieving accuracies over 90%.

Furthermore, by integrating the predictions from our model with other bioinformatics tools, we enhanced the accuracy and reliability of identifying potential ACPs. The functional enrichment analysis suggested that most of the predicted ACPs play a crucial role in targeting and destroying harmful cells. As noted, the anticancer effects and selective cytotoxicity of these peptides were substantiated through in vitro experiments involving numerous human cancer cell lines. Although the mechanisms behind the anticancer properties of these peptides are not yet fully understood, our findings underscore their potential to exhibit cytotoxicity against cancer cells, emphasizing the need for further exploration to fully elucidate their therapeutic capabilities.

4. Materials and Methods

4.1. Redundancy Removal

To minimize overestimation, the CD-HIT V4.8.1 [60] was employed to decrease redundancy in the ACP sequences within the training dataset by applying a cutoff of 80% similarity. Additionally, to better mimic real-world conditions, homologous sequences in the non-ACP dataset were also filtered out when their identity exceeded 50%. This approach ensures a more accurate and reliable dataset for subsequent analyses.

4.2. Feature Investigation

Sequence-based features, including amino acid composition (AAC) [33], dipeptide composition (DPC) [34], and k-spaced amino acid pairs (CKSAAPs) [35], are widely utilized in analyzing protein functions and developing prediction models. AAC specifically quantifies the frequency of occurrence of each of the 20 standard amino acids in a protein sequence, which is essential for feature encoding. This process can be described as follows:

$$f(i) = \frac{\sum x_i}{L} (1 \leq i \leq 20)$$

Given a peptide, where i represents each type of amino acid, x_i stands the number of occurrences of each amino acid, and L is the full length of the considered peptide.

The CKSAAP method estimates the frequencies of occurrence of 20×20 types of amino acid pairs in a peptide. These pairs are defined by their separation through a specific number of other amino acids, known as the gap. The formulation can be expressed as follows:

$$f(i,j) = \frac{\sum x_{i,j}}{L - k} (1 \leq i, j \leq 20)$$

where i,j represents each amino acid pair, $x_{i,j}$ stands the number of occurrences of each amino acid pair separated by k amino acids, and L is the full length of the considered peptide; $k = 0, 1, 2,$ and 3 were considered as features to be applied in the prediction of anticancer activity.

Additionally, the composition of secondary structure elements (SSECs) was considered a primary feature for investigation in this study. The secondary structures of all peptides were predicted using the PEP2D web tool [43]. This tool uniquely predicts the secondary structure of peptides based on their amino acid sequences using a method tailored specifically for peptides rather than proteins.

4.3. Construction of Prediction Models

The support vector machine (SVM) is a supervised machine learning algorithm widely applied to various biological classification problems. In this study, the SVM algorithm implemented in LIBSVM [54] was used as the classifier. LIBSVM is a publicly available SVM tool that employs the radial basis function (RBF) as its kernel function. The flexibility of the decision boundary, or hyperplane, is determined by two parameters: gamma (γ) and cost (C). We utilized LIBSVM to construct classification models using feature vectors based on both sequence and structural characteristics.

4.4. Performance Evaluation

In this study, we used the training data to build the ACP prediction model employing LIBSVM. To evaluate the model's performance, we conducted five repetitions of a five-fold cross-validation procedure. We utilized the following measures to estimate the predictive performance of the model, encompassing *TP* (true positive), *FN* (false negative), *TN* (true negative), and *FP* (false positive):

$$Sensitivity\ (Sn) = \frac{TP}{TP + FN}$$

$$Specificity\ (Sp) = \frac{TN}{TN + FP}$$

$$Accuracy\ (Acc) = \frac{TP + TN}{TP + FP + TN + FN}$$

$$Matthews\ Correlation\ Coefficient\ (MCC) = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

4.5. Identification of Candidate ACPs Using Multiple Prediction Tools

Novel natural ACP candidates were predicted using the proposed model alongside six other existing models, including ACPred [21], ACPred-FL [20], AntiCP [13], AntiCP2 [23], iACP [16], and mACPpred [22]. Peptide sequences extracted from natural species were input into these prediction tools in FASTA format. By integrating the results from all seven prediction tools, we were able to obtain multiple decision-making outcomes. A majority voting method was adopted to synthesize these results into a final decision. A peptide was nominated as a candidate only if it received unanimous approval from all seven tools. If candidates obtained an equal number of votes, they were ranked based on the scores from the proposed model.

4.6. Functional Enrichment Analysis

Gene Ontology (GO) [56] is a comprehensive resource that describes the functions of gene products across all living species in three independent categories: cellular component (CC), molecular function (MF), and biological process (BP). To provide a functional interpretation for the identified peptides, GO annotations for each peptide were obtained from the Universal Protein Knowledgebase (UniProtKB) [32].

4.7. Evaluation of Anticancer Activity and Selective Cytotoxicity

Cancer cell lines were purchased from BCRC (Hsinchu, R.O.C) and their culture conditions were created according to BCRC's suggestion. The culture medium used for the cancer cell lines included 10% fetal bovine serum (Gibco, Grand Island, NY, USA) and

1% penicillin/streptomycin (Gibco) in 5% CO_2 at 37 °C. Cells were seeded into 96-well tissue culture plates at a concentration of 1×10^4 cells per 200 μL per well and allowed to settle overnight. The cells were then treated with serial dilutions of various ACPs. After 48 h of incubation, cell viability for each line was assessed using the MTT colorimetric assay (Sigma-Aldrich, St. Louis, MO, USA). The peptide preparation and dilution in our anticancer assays are as follows: ACPs (10 mg) were dissolved in 100 μL of DMSO to generate 100 mg/mL stock solutions. These stock solutions were then diluted 100 times with a culture medium to obtain a 1 mg/mL solution. Finally, serial dilutions were prepared in a culture medium containing 1% DMSO to achieve the desired ACP concentrations for the MTT assay. Cell viability was expressed as a percentage of the untreated control, and the inhibitory concentration at which 50% of the cells survived (IC50) was determined from the dose–response curve.

5. Conclusions

The findings of this study significantly advance the development of cancer therapies by identifying potential anticancer peptides (ACPs) through both computational predictions and experimental validations. These ACPs exhibit selective cytotoxicity towards cancer cells, presenting a promising alternative to traditional treatments like chemotherapy and surgery, which often come with severe side effects. The discovery of these natural peptides with anticancer properties opens the door to novel, targeted therapies that are both effective and have fewer side effects.

Moreover, our study demonstrates the effectiveness of integrating computational and experimental approaches, which pave the way for more efficient discovery and validation processes for ACPs in future research. This dual approach not only increases the likelihood of discovering potent therapeutic agents but also positions ACPs as excellent candidates for the development of new anticancer drugs. By leveraging both computational predictions and empirical validations, we enhance the potential to identify effective treatments that could play a crucial role in future cancer therapies.

In conclusion, the strategy outlined in this study, which integrates in silico and in vitro approaches, offers a comprehensive and reliable platform for the identification and characterization of natural anticancer peptides. This integrated strategy not only facilitates the discovery of potent therapeutic agents but also establishes a robust framework for translating these promising peptides into effective cancer treatments.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Mahase, E. Cancer overtakes CVD to become leading cause of death in high income countries. *BMJ* **2019**, *366*, l5368. [CrossRef] [PubMed]
2. Ma, X.; Yu, H. Global burden of cancer. *Yale J. Biol. Med.* **2006**, *79*, 85–94. [PubMed]
3. Tohme, S.; Simmons, R.L.; Tsung, A. Surgery for Cancer: A Trigger for Metastases. *Cancer Res.* **2017**, *77*, 1548–1552. [CrossRef] [PubMed]
4. Lamson, D.W.; Brignall, M.S. Antioxidants in cancer therapy; their actions and interactions with oncologic therapies. *Altern. Med. Rev.* **1999**, *4*, 304–329. [PubMed]
5. Potmesil, M. Camptothecins: From bench research to hospital wards. *Cancer Res.* **1994**, *54*, 1431–1439. [PubMed]
6. Coates, A.; Abraham, S.; Kaye, S.B.; Sowerbutts, T.; Frewin, C.; Fox, R.M.; Tattersall, M.H. On the receiving end--patient perception of the side-effects of cancer chemotherapy. *Eur. J. Cancer Clin. Oncol.* **1983**, *19*, 203–208. [CrossRef] [PubMed]
7. Gaspar, D.; Veiga, A.S.; Castanho, M.A. From antimicrobial to anticancer peptides. A review. *Front. Microbiol.* **2013**, *4*, 294. [CrossRef] [PubMed]
8. Perez-Tomas, R.; Perez-Guillen, I. Lactate in the Tumor Microenvironment: An Essential Molecule in Cancer Progression and Treatment. *Cancers* **2020**, *12*, 3244. [CrossRef] [PubMed]
9. Schweizer, F. Cationic amphiphilic peptides with cancer-selective toxicity. *Eur. J. Pharmacol.* **2009**, *625*, 190–194. [CrossRef] [PubMed]
10. Rodrigues, E.G.; Dobroff, A.S.; Taborda, C.P.; Travassos, L.R. Antifungal and antitumor models of bioactive protective peptides. *An. Acad. Bras. Ciências* **2009**, *81*, 503–520. [CrossRef] [PubMed]
11. Droin, N.; Hendra, J.B.; Ducoroy, P.; Solary, E. Human defensins as cancer biomarkers and antitumour molecules. *J. Proteom.* **2009**, *72*, 918–927. [CrossRef] [PubMed]
12. Harris, F.; Dennison, S.R.; Singh, J.; Phoenix, D.A. On the selectivity and efficacy of defense peptides with respect to cancer cells. *Med. Res. Rev.* **2013**, *33*, 190–234. [CrossRef] [PubMed]
13. Tyagi, A.; Kapoor, P.; Kumar, R.; Chaudhary, K.; Gautam, A.; Raghava, G.P. In silico models for designing and discovering novel anticancer peptides. *Sci. Rep.* **2013**, *3*, 2984. [CrossRef] [PubMed]
14. Hajisharifi, Z.; Piryaiee, M.; Mohammad Beigi, M.; Behbahani, M.; Mohabatkar, H. Predicting anticancer peptides with Chou's pseudo amino acid composition and investigating their mutagenicity via Ames test. *J. Theor. Biol.* **2014**, *341*, 34–40. [CrossRef] [PubMed]
15. Vijayakumar, S.; Ptv, L. ACPP: A Web Server for Prediction and Design of Anti-cancer Peptides. *Int. J. Pept. Res. Ther.* **2015**, *21*, 99–106. [CrossRef]
16. Chen, W.; Ding, H.; Feng, P.; Lin, H.; Chou, K.C. iACP: A sequence-based tool for identifying anticancer peptides. *Oncotarget* **2016**, *7*, 16895–16909. [CrossRef] [PubMed]
17. Li, F.M.; Wang, X.Q. Identifying anticancer peptides by using improved hybrid compositions. *Sci. Rep.* **2016**, *6*, 33910. [CrossRef] [PubMed]
18. Akbar, S.; Hayat, M.; Iqbal, M.; Jan, M.A. iACP-GAEnsC: Evolutionary genetic algorithm based ensemble classification of anticancer peptides by utilizing hybrid feature space. *Artif. Intell. Med.* **2017**, *79*, 62–70. [PubMed]
19. Kabir, M.; Arif, M.; Ahmad, S.; Ali, Z.; Swati, Z.N.K.; Yu, D.-J. Intelligent computational method for discrimination of anticancer peptides by incorporating sequential and evolutionary profiles information. *Chemom. Intell. Lab. Syst.* **2018**, *182*, 158–165. [CrossRef]
20. Wei, L.; Zhou, C.; Chen, H.; Song, J.; Su, R. ACPred-FL: A sequence-based predictor using effective feature representation to improve the prediction of anti-cancer peptides. *Bioinformatics* **2018**, *34*, 4007–4016. [CrossRef] [PubMed]
21. Schaduangrat, N.; Nantasenamat, C.; Prachayasittikul, V.; Shoombuatong, W. ACPred: A Computational Tool for the Prediction and Analysis of Anticancer Peptides. *Molecules* **2019**, *24*, 1973. [CrossRef]
22. Boopathi, V.; Subramaniyam, S.; Malik, A.; Lee, G.; Manavalan, B.; Yang, D.C. mACPpred: A Support Vector Machine-Based Meta-Predictor for Identification of Anticancer Peptides. *Int. J. Mol. Sci.* **2019**, *20*, 1964. [CrossRef] [PubMed]
23. Agrawal, P.; Bhagat, D.; Mahalwal, M.; Sharma, N.; Raghava, G.P.S. AntiCP 2.0: An updated model for predicting anticancer peptides. *Brief. Bioinform.* **2021**, *22*, bbaa153. [CrossRef]
24. Ge, R.; Feng, G.; Jing, X.; Zhang, R.; Wang, P.; Wu, Q. EnACP: An Ensemble Learning Model for Identification of Anticancer Peptides. *Front. Genet.* **2020**, *11*, 760. [CrossRef] [PubMed]
25. Wang, G.; Li, X.; Wang, Z. APD3: The antimicrobial peptide database as a tool for research and education. *Nucleic Acids Res.* **2016**, *44*, D1087–D1093. [CrossRef] [PubMed]
26. Thomas, S.; Karnik, S.; Barai, R.S.; Jayaraman, V.K.; Idicula-Thomas, S. CAMP: A useful resource for research on antimicrobial peptides. *Nucleic Acids Res.* **2010**, *38*, D774–D780. [CrossRef]
27. Tyagi, A.; Tuknait, A.; Anand, P.; Gupta, S.; Sharma, M.; Mathur, D.; Joshi, A.; Singh, S.; Gautam, A.; Raghava, G.P. CancerPPD: A database of anticancer peptides and proteins. *Nucleic Acids Res.* **2015**, *43*, D837–D843. [CrossRef] [PubMed]

28. Novkovic, M.; Simunic, J.; Bojovic, V.; Tossi, A.; Juretic, D. DADP: The database of anuran defense peptides. *Bioinformatics* **2012**, *28*, 1406–1407. [CrossRef]

29. Jhong, J.H.; Chi, Y.H.; Li, W.C.; Lin, T.H.; Huang, K.Y.; Lee, T.Y. dbAMP: An integrated resource for exploring antimicrobial peptides with functional activities and physicochemical properties on transcriptome and proteome data. *Nucleic Acids Res.* **2019**, *47*, D285–D297. [CrossRef] [PubMed]

30. Kang, X.; Dong, F.; Shi, C.; Liu, S.; Sun, J.; Chen, J.; Li, H.; Xu, H.; Lao, X.; Zheng, H. DRAMP 2.0, an updated data repository of antimicrobial peptides. *Sci. Data* **2019**, *6*, 148. [CrossRef] [PubMed]

31. Singh, S.; Chaudhary, K.; Dhanda, S.K.; Bhalla, S.; Usmani, S.S.; Gautam, A.; Tuknait, A.; Agrawal, P.; Mathur, D.; Raghava, G.P. SATPdb: A database of structurally annotated therapeutic peptides. *Nucleic Acids Res.* **2016**, *44*, D1119–D1126. [CrossRef] [PubMed]

32. UniProt, C. UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Res.* **2021**, *49*, D480–D489.

33. Sahu, S.S.; Panda, G. A novel feature representation method based on Chou's pseudo amino acid composition for protein structural class prediction. *Comput. Biol. Chem.* **2010**, *34*, 320–327. [CrossRef] [PubMed]

34. Park, K.J.; Kanehisa, M. Prediction of protein subcellular locations by support vector machines using compositions of amino acids and amino acid pairs. *Bioinformatics* **2003**, *19*, 1656–1663. [CrossRef] [PubMed]

35. Chen, Y.Z.; Tang, Y.R.; Sheng, Z.Y.; Zhang, Z. Prediction of mucin-type O-glycosylation sites in mammalian proteins using the composition of k-spaced amino acid pairs. *BMC Bioinform.* **2008**, *9*, 101. [CrossRef] [PubMed]

36. Seo, M.D.; Won, H.S.; Kim, J.H.; Mishig-Ochir, T.; Lee, B.J. Antimicrobial peptides for therapeutic applications: A review. *Molecules* **2012**, *17*, 12276–12286. [CrossRef] [PubMed]

37. Baindara, P.; Gautam, A.; Raghava, G.P.S.; Korpole, S. Anticancer properties of a defensin like class IId bacteriocin Laterosporulin10. *Sci. Rep.* **2017**, *7*, 46541. [CrossRef] [PubMed]

38. Ghosh, S.K.; McCormick, T.S.; Weinberg, A. Human Beta Defensins and Cancer: Contradictions and Common Ground. *Front. Oncol.* **2019**, *9*, 341. [CrossRef] [PubMed]

39. Oelkrug, C.; Hartke, M.; Schubert, A. Mode of action of anticancer peptides (ACPs) from amphibian origin. *Anticancer Res.* **2015**, *35*, 635–643.

40. Baindara, P.; Kapoor, A.; Korpole, S.; Grover, V. Cysteine-rich low molecular weight antimicrobial peptides from Brevibacillus and related genera for biotechnological applications. *World J. Microbiol. Biotechnol.* **2017**, *33*, 124. [CrossRef] [PubMed]

41. Wu, D.; Gao, Y.; Qi, Y.; Chen, L.; Ma, Y.; Li, Y. Peptide-based cancer therapy: Opportunity and challenge. *Cancer Lett.* **2014**, *351*, 13–22. [CrossRef]

42. Martin, A.C.; Orengo, C.A.; Hutchinson, E.G.; Jones, S.; Karmirantzou, M.; Laskowski, R.A.; Mitchell, J.B.; Taroni, C.; Thornton, J.M. Protein folds and functions. *Structure* **1998**, *6*, 875–884. [CrossRef] [PubMed]

43. Singh, H.; Singh, S.; Singh Raghava, G.P. Peptide Secondary Structure Prediction using Evolutionary Information. *bioRxiv* **2019**, 558791. [CrossRef]

44. Shai, Y. Mechanism of the binding, insertion and destabilization of phospholipid bilayer membranes by alpha-helical antimicrobial and cell non-selective membrane-lytic peptides. *Biochim. Biophys. Acta* **1999**, *1462*, 55–70. [CrossRef] [PubMed]

45. Papo, N.; Shahar, M.; Eisenbach, L.; Shai, Y. A novel lytic peptide composed of DL-amino acids selectively kills cancer cells in culture and in mice. *J. Biol. Chem.* **2003**, *278*, 21018–21023. [CrossRef] [PubMed]

46. Papo, N.; Shai, Y. New lytic peptides based on the D,L-amphipathic helix motif preferentially kill tumor cells compared to normal cells. *Biochemistry* **2003**, *42*, 9346–9354. [CrossRef] [PubMed]

47. Mai, J.C.; Mi, Z.; Kim, S.H.; Ng, B.; Robbins, P.D. A proapoptotic peptide for the treatment of solid tumors. *Cancer Res.* **2001**, *61*, 7709–7712. [PubMed]

48. Papo, N.; Oren, Z.; Pag, U.; Sahl, H.G.; Shai, Y. The consequence of sequence alteration of an amphipathic alpha-helical antimicrobial peptide and its diastereomers. *J. Biol. Chem.* **2002**, *277*, 33913–33921. [CrossRef] [PubMed]

49. Dennison, S.R.; Whittaker, M.; Harris, F.; Phoenix, D.A. Anticancer alpha-helical peptides and structure/function relationships underpinning their interactions with tumour cell membranes. *Curr. Protein Pept. Sci.* **2006**, *7*, 487–499. [CrossRef] [PubMed]

50. Huang, Y.; Feng, Q.; Yan, Q.; Hao, X.; Chen, Y. Alpha-helical cationic anticancer peptides: A promising candidate for novel anticancer drugs. *Mini Rev. Med. Chem.* **2015**, *15*, 73–81. [CrossRef] [PubMed]

51. Chiangjong, W.; Chutipongtanate, S.; Hongeng, S. Anticancer peptide: Physicochemical property, functional aspect and trend in clinical application (Review). *Int. J. Oncol.* **2020**, *57*, 678–696. [CrossRef] [PubMed]

52. Crooks, G.E.; Hon, G.; Chandonia, J.M.; Brenner, S.E. WebLogo: A sequence logo generator. *Genome Res.* **2004**, *14*, 1188–1190. [CrossRef] [PubMed]

53. Manavalan, B.; Basith, S.; Shin, T.H.; Choi, S.; Kim, M.O.; Lee, G. MLACP: Machine-learning-based prediction of anticancer peptides. *Oncotarget* **2017**, *8*, 77121–77136. [CrossRef] [PubMed]

54. Chang, C.C.; Lin, C.J. LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.* **2011**, *2*, 27:1–27:27. [CrossRef]

55. Huang, K.Y.; Tseng, Y.J.; Kao, H.J.; Chen, C.H.; Yang, H.H.; Weng, S.L. Identification of subtypes of anticancer peptides based on sequential features and physicochemical properties. *Sci. Rep.* **2021**, *11*, 13594. [PubMed]

56. The Gene Ontology, C. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* **2019**, *47*, D330–D338.

57.   Ghadiri, N.; Javidan, M.; Sheikhi, S.; Tastan, O.; Parodi, A.; Liao, Z.; Tayybi Azar, M.; Ganjalikhani-Hakemi, M. Bioactive peptides: An alternative therapeutic approach for cancer management. *Front. Immunol.* **2024**, *15*, 1310443. [CrossRef] [PubMed]

58.   Jafari, A.; Babajani, A.; Sarrami Forooshani, R.; Yazdani, M.; Rezaei-Tavirani, M. Clinical Applications and Anticancer Effects of Antimicrobial Peptides: From Bench to Bedside. *Front. Oncol.* **2022**, *12*, 819563. [CrossRef] [PubMed]

59.   Ghaly, G.; Tallima, H.; Dabbish, E.; Badr ElDin, N.; Abd El-Rahman, M.K.; Ibrahim, M.A.A.; Shoeib, T. Anti-Cancer Peptides: Status and Future Prospects. *Molecules* **2023**, *28*, 1148. [CrossRef] [PubMed]

60.   Li, W.; Godzik, A. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **2006**, *22*, 1658–1659. [CrossRef] [PubMed]