Article

Predicting Anticancer Drug Resistance Mediated by Mutations

Yu-Feng Lin [1],[†], Jia-Jun Liu [2],[†], Yu-Jen Chang [2], Chin-Sheng Yu [3], Wei Yi [1], Hsien-Yuan Lane [4],[5],[6]
and Chih-Hao Lu [2],[4],[7],[*]

[1] Department of Medical Laboratory Science and Biotechnology, Asia University, Taichung 41354, Taiwan;
  yflin@asia.edu.tw (Y.-F.L.); yiwyiw0918@gmail.com (W.Y.)
[2] The Ph.D. Program of Biotechnology and Biomedical Industry, China Medical University,
  Taichung 40402, Taiwan; u107305606@cmu.edu.tw (J.-J.L.); u107311201@cmu.edu.tw (Y.-J.C.)
[3] Department of Information Engineering and Computer Science, Feng Chia University,
  Taichung 40201, Taiwan; yucs@fcu.edu.tw
[4] Graduate Institute of Biomedical Sciences, China Medical University, Taichung 40402, Taiwan;
  hylane@gmail.com
[5] Department of Psychiatry, China Medical University Hospital, Taichung 40402, Taiwan
[6] Brain Disease Research Center, China Medical University Hospital, Taichung 40402, Taiwan
[7] Department of Medical Laboratory Science and Biotechnology, China Medical University,
  Taichung 40402, Taiwan
[*] Correspondence: chlu@mail.cmu.edu.tw
[†] These authors contributed equally to this work and share first authorship.

Abstract: Cancer drug resistance presents a challenge for precision medicine. Drug-resistant mutations are always emerging. In this study, we explored the relationship between drug-resistant mutations and drug resistance from the perspective of protein structure. By combining data from previously identified drug-resistant mutations and information of protein structure and function, we used machine learning-based methods to build models to predict cancer drug resistance mutations. The performance of our combined model achieved an accuracy of 86%, a Matthews correlation coefficient score of 0.57, and an F1 score of 0.66. We have constructed a fast, reliable method that predicts and investigates cancer drug resistance in a protein structure. Nonetheless, more information is needed concerning drug resistance and, in particular, clarification is needed about the relationships between the drug and the drug resistance mutations in proteins. Highly accurate predictions regarding drug resistance mutations can be helpful for developing new strategies with personalized cancer treatments. Our novel concept, which combines protein structure information, has the potential to elucidate physiological mechanisms of cancer drug resistance.

Keywords: cancer drug; drug resistance; single amino acid variation; protein structure; machine learning; feature selection; personalized therapeutics

1. Introduction

One of the greatest challenges of this century is precision medicine, highlighted by the search for personalized cancer medicine. The expectation is that by analyzing tumors at the molecular level, scientists can design treatment that specifically adapts to specific molecular subgroups of tumors and even individual patient characteristics, greatly improving the therapeutic outcomes. In recent decades, targeted cancer therapy has been associated with significant improvements in survival rates, so has become one of the standard strategies for cancer treatment [1]. However, a problem with targeted drug therapy is the emergence of cancer drug resistance [2], as with chemotherapy [3,4]. Many studies have investigated the mechanisms of resistance to chemotherapy and their solutions [5–14]. The mechanisms of how cancer cells acquire targeted drug resistance are not clarified. Some patients develop resistance to targeted drug therapy post-treatment, possibly due to the occurrence of drug-resistant mutations [15]. For example, gefitinib and

afatinib, epidermal growth factor receptor tyrosine kinase inhibitors (EGFR-TKIs) that have been developed as targeted therapies for use in patients with non-small cell lung cancer (NSCLC) who have EGFR-activating mutations, significantly improve survival rates in this patient population [16–18]. However, a recent case study has reported that a patient with NSCLC and the EGFR L858R mutation developed acquired resistance to gefitinib [19]. Gene sequencing of the blood revealed an EGFR kinase domain duplication (EGFR-KDD mutation) that has not previously been reported as an acquired mutation due to EGFR-TKI resistance in NSCLC [19]. The gradual development of acquired resistance within tumors is characterized by subpopulations of cells that may acquire or already possess the mutations that enable them to escape the effects of targeted therapies [20–22]. Adaptive responses of tumor cells to treatment promote a variety of mechanisms that lead to drug resistance, including changes in the tumor microenvironment, repair of DNA damage, changes in drug targets, alterations of signaling pathways, and changes in the pharmacology of cells [4,23]. Many experimental studies have revealed intrinsic or acquired resistance mechanisms that validate the use of strategic changes to treatment strategies, such as combination therapy, to overcome resistance to tumor treatment [24,25]. Notably, searching for successful treatment strategies requires multiple experimental conditions, cell lines and different time series modeling techniques, all of which are expensive and time-consuming to conduct with traditional hypothesis-driven experimental methods [26–28].

In recent years, the public release of data from large-scale drug screening programs [29–32] has helped to promote the development of precision oncology. These data have been applied by computational methods to identify putative drug response biomarkers [29,33] and have been used to develop predictive models to predict drug sensitivity [34]. The development of computational strategies for predicting drug responses is vital for guiding drug discovery and reducing the required amount of experimental work. Various machine learning models have been proposed that predict drug responses and enable the discovery of drug response biomarkers, such as the random forest algorithm [34–37], support vector machine (SVM) algorithms [38], neural networks [39] and Bayesian multitask multiple kernel learning [40,41]. Many methods have been developed to predict drug responses using analytical data based on signaling networks, cellular dynamics, and high-throughput data [42–46].

Evidence from the study of genomics has produced vital information. As proteins are responsible for coordinating biological processes of cells, understanding and analyzing the role of proteins in organisms is necessary for correct assessments of disease status [47]. Proteomes are much more complex than their corresponding genomes, as proteomes involve mechanisms such as alternative splicing and post-translational modifications. Correct biological functions of proteins depend upon them coiling and folding correctly into three-dimensional (3D) structures; gene mutations can lead to structural changes that influence protein functions [48,49]. In many types of gene mutations, most disease-related single amino acid variations (SAVs) occur in structurally or functionally important positions. SAV refers to an amino acid substitution caused by a genetic polymorphism. In extreme cases, nonsynonymous encoding variants alter protein sequences and thus change the entire protein structure or function. The unique physical and chemical properties of each kind of amino acid mean that the occurrence of mutations at different positions in the sequence affects protein conformation and function to varying degrees.

In this study, we explored the relationship between resistance mutations and cancer drug resistance from the perspective of protein structures and we built a reliable prediction model to distinguish which mutation(s) might be responsible for drug resistance. The data used to build prediction models were obtained from the Catalogue of Somatic Mutations in Cancer (COSMIC) database, the world's largest expert-curated database of somatic mutations in human cancers [50]. These mutations are recorded in the literature as drug resistance mutations and include both intrinsic resistance (before treatment) and acquired resistance (after treatment). The prediction models were constructed by machine learning-based methods, including genetic algorithms and SVMs, to predict drug resistance

mutations involving protein structures. The performance of our combined model demonstrated an accuracy of 86%, an MCC (Matthews correlation coefficient) of 0.57 and an F1 score of 0.66, indicating that mutations may lead to drug resistance by altering protein structures and their positions in 3D space.

2. Results

2.1. Performance Evaluation of the Training Set

This study employed the machine learning method to construct several different drug resistance SAV prediction (DRSP) models. Each model was subjected to the five-fold cross-validation technique to evaluate predictive performance. All prediction models were optimized using the MCC as the fitness function. Table 1 presents comparisons of predictive performances from the different prediction models. In our experiment, the prediction of the simulated complexes ($DRSP_s$) outperformed the prediction of the crystal complexes ($DRSP_c$). The training set from the simulated and crystal complexes was divided into subgroups based on the spatial distance between the SAVs and the drug. According to the study dataset, approximately 60% of drug-resistant SAVs were located within an 8 Å distance from the drug; this 8 Å space surrounding the drug was regarded as the drug-binding pocket. We therefore used a cut-off of 8 Å, then divided the model according to SAV location; within the 8 Å space surrounding the drug ($DRSP^i$, where "i" = interior), or beyond that space ($DRSP^e$, where "e" = exterior). The interior models ($DRSP_c^i$ and $DRSP_s^i$) had higher sensitivity than the exterior models ($DRSP_c^e$ and $DRSP_s^e$), although the exterior models exhibited higher specificity than the interior models. Thus, we speculated that combining the $DRSP^e$ and $DRSP^i$ models would yield superior performances for both prediction models (simulated and crystal complexes). The combined interior and exterior modeled complexes yielded an accuracy of 85%, an MCC of 0.56 and F1 score of 0.65. The performance was slightly better with the combined interior and exterior crystal complexes, with an accuracy of 86%, an MCC of 0.57 and F1 score of 0.66, which may be due to the SAV positions, which are capable of directly or indirectly influencing drug binding to the protein. The superior predictive performance based on the 8 Å distance suggests that spatial distance is a primary consideration for calculations into direct and indirect factors affecting drug resistance. Indeed, spatial distance appears to be the primary contributor in the impact of resistance SAVs.

Table 1. Comparison of predictive performances from different prediction models. All predictions were optimized using the Matthews correlation coefficient (MCC) as the fitness function.

Models	Accuracy	Sensitivity	Specificity	MCC	Precision	F1 Score
$DRSP_c$	0.8377	0.5338	0.9224	0.4936	0.6574	0.5892
$DRSP_s$	0.8508	0.5188	0.9434	0.5241	0.7188	0.6026
$DRSP_c^e$	0.8886	0.5385	0.9378	0.4803	0.5490	0.5437
$DRSP_c^i$	0.7819	0.7284	0.8224	0.5536	0.7564	0.7421
$DRSP_s^e$	0.8886	0.5577	0.9351	0.4888	0.5472	0.5524
$DRSP_s^i$	0.7660	0.6914	0.8224	0.5196	0.7467	0.7179
$DRSP_c^e + DRSP_c^i$	0.8557	0.6541	0.9119	0.5724	0.6744	0.6641
$DRSP_s^e + DRSP_s^i$	0.8508	0.6391	0.9099	0.5567	0.6641	0.6513

2.2. Performance Evaluation of the Testing Set

The testing set contained three resistance SAVs located in three proteins: L505H in the BRAF protein kinase (a participant in MAP kinase/ERK signaling) [51]; V215E in the dual specificity mitogen-activated protein kinase 2 (MAP2K2, a downstream kinase from BRAF in the MAPK pathway) [34]; and G2032R in the c-ros oncogene 1 (ROS1, a receptor tyrosine kinase that acts as an oncogene driver of non-small cell lung cancer [NSCLC]) [52] (Table 2). The molecular docking method was used to simulate the BRAF-vemurafenib, MAP2K2-PD0325901, and ROS1-crizotinib structural complexes. Different prediction models were used to perform drug-resistant SAV predictions, according to the distance between the

drug-resistant SAV and the drug in the simulated complexes. Table 2 presents distances between drug-resistant SAVs and drugs, the prediction models, and the predicted results. The $DRSP_c^i$ model correctly predicted three drug-resistant SAV mutations; L505H in BRAF, V215E in MAP2K2 and G2032R in ROS1. However, L505H in BRAF and V215E in MAP2K2 were not correctly predicted by the $DRSP_s^e$ model. Specificity values of the testing sets predicted by the combined models from the crystal and simulated complexes were 0.84 and 0.93, respectively. The predicted results from the testing sets indicate that the crystal complex performs better than the simulated complex prediction model in the identification of drug-resistant SAVs. In contrast, the simulated complex prediction model performs better in detecting non-drug-resistant SAVs.

Table 2. The predicted results from the testing sets.

Protein	Drug-Resistant SAV	Distance [1]	Model	Predicted Result
BRAF	L505H	5.41	$DRSP_c^i$	TP
			$DRSP_s^i$	FN
MAP2K2	V215E	4.27	$DRSP_c^i$	TP
			$DRSP_s^i$	FN
ROS1	G2032R	3.30	$DRSP_c^i$	TP
			$DRSP_s^i$	TP

[1] The distance between the drug-resistant SAV and the docked drug.

2.3. Case Study: L505 in BRAF and V215E in MAP2K2

The *BRAF* gene is an oncogene that occurs in 8% of human cancers [53], while the *BRAF* mutation usually occurs in melanoma and colorectal cancer [54,55]. The substitution of the valine residue by glutamic acid at amino acid position 600 of the BRAF protein is the most frequently observed *BRAF* mutation (occurring in approximately 90% of patients) [53,56]. Previous studies have documented how the use of a MAP2K2 inhibitor for *BRAF*-mutated melanoma triggers the PI3K/AKT pathway and leads to drug treatment resistance, although acquired *MAP2K2* mutations have also been associated with the drug resistance [57,58]. BRAF inhibitor therapy can provide short-term curative effects, but it also induces resistance and often disease progression [59]. Several unique drug-resistant SAVs have been reported in the COSMIC database. Five L505 mutations in the BRAF protein have been linked to vemurafenib; one of these L505 mutations is mutated to histidine [60,61]. Moreover, a V215E mutation in MAP2K2 is reportedly associated with PD0325901 drug resistance [62].

Our prediction model analyzes the interaction characteristics of drug-protein complexes. However, not all drug-protein complexes have a crystal structure for analysis. For those complexes lacking a crystal structure, we would use the docking model to simulate the complex before analyzing whether the SAVs are drug-resistant or not. Our docking model has proven reliable for our follow-up analysis, because the positions of each drug in the simulated complexes overlap the drug positions in the crystal complexes (Figures 1a, 2a and 3a). We were also able to simulate each mutated amino acid so that we could investigate changes in the structures after substitution (Figures 1b, 2b and 3b). Simulated mutated amino acids were generated by PS^2, a protein structure prediction server [63]. The BRAF leucine residue 505 is 5.41 Å away from the docking space for vemurafenib, while the MAP2K2 valine residue 215 is 4.27 Å away from PD0325901. Both drug-resistant SAVs are positioned within the distance of 8 Å. Thus, the $DRSP_c^i$ would be applied. This model incorporates several critical features (see Table S1): the drug-resistant SAVs are found with low levels of five elements (W-H-neu, W-Z-low, W-Z-med, W-P-low, and W-P-neu) within the surrounding amino acid residues. This finding implies that fewer neutral amino acids are associated with drug-resistant SAVs. BRAF residue L505 is positioned very near the functional amino acid R506, which exhibits a distinct conformation for

molecular binding [64]. We speculated that this unique pattern can be used to predict drug resistance and highlight drug-protein interactions.
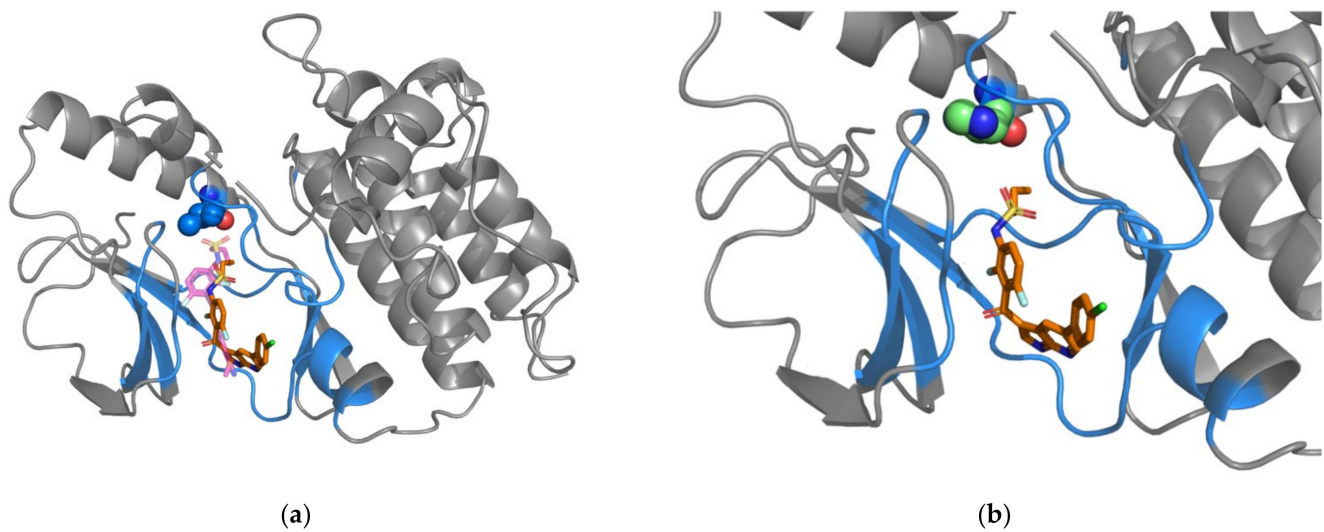

Figure 1. Simulated and crystal structures of the BRAF-drug complex. (**a**) The simulated structure of the BRAF-vemurafenib complex. Docked vemurafenib is indicated by the orange-colored stick. The magenta-colored stick represents the pyrazolopyridine inhibitor, which is located in the crystal structure of the BRAF-pyrazolopyridine inhibitor complex. The gray-colored cartoon structures of BRAF (PDBID: 3TV6 [65]) were drawn using PyMOL software. Residues found within 8 Å from vemurafenib are represented by blue coloring. The drug-resistant SAV (L505) is represented as spheres in blue. (**b**) Simulation of the amino acid mutated to histidine (H505) is shown in the color green.

Figure 2. Simulated and crystal structures of the MAP2K2-drug complex. (**a**) The simulated structure of the MAP2K2-PD0325901 complex. Docked PD0325901 is indicated by the yellow-colored stick. The blue-colored stick indicates the PD184352-like inhibitor. The gray-colored cartoon structures of MAP2K2 (PDBID: 1S9I [66]) were drawn using PyMOL software. Residues found within 8 Å from PD0325901 are represented by pink coloring. The drug-resistant SAV (V215) is represented as spheres in pink. (**b**) Simulation of mutated glutamic acid (E215) is shown in the color green.

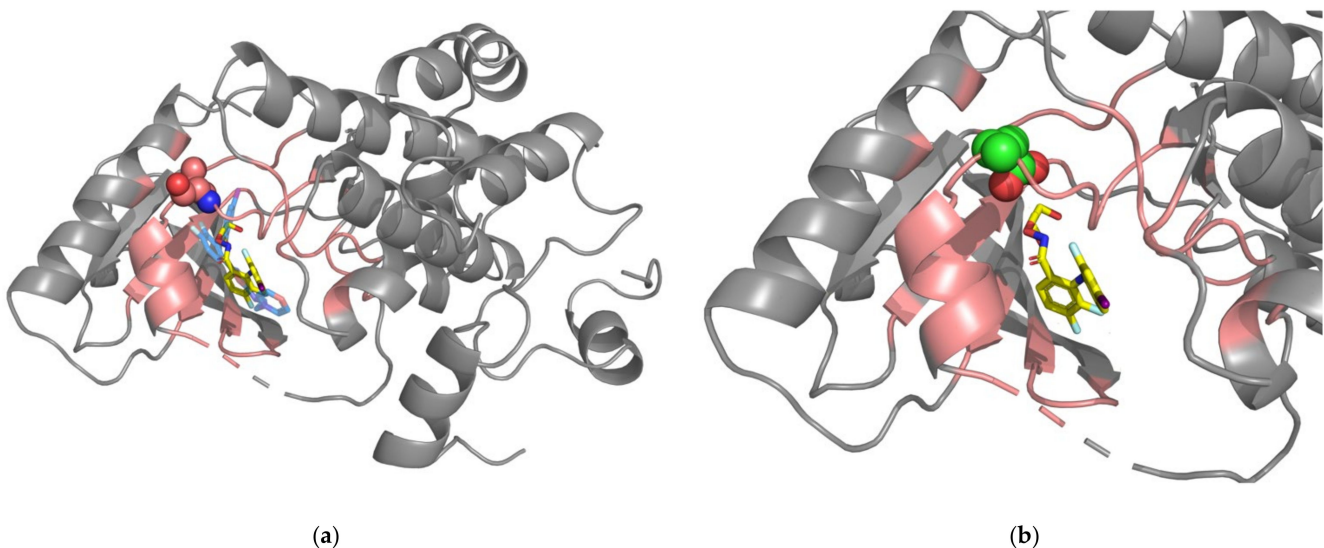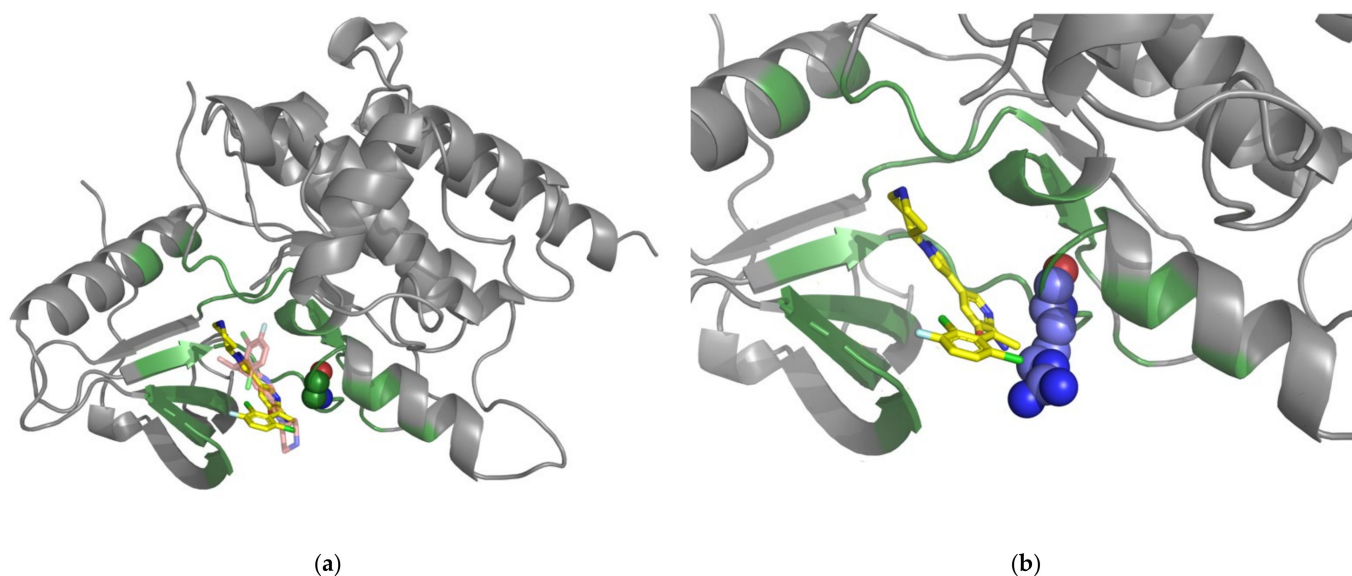(**a**)   (**b**)

Figure 3. Simulated and crystal structures of the ROS1-drug complex. (**a**) The simulated structure of the ROS1-crizotinib structure complex. Docked crizotinib is indicated by the yellow-colored stick. The transparent orange-colored stick indicates crizotinib. The gray-colored cartoon structures of ROS1 (PDBID: 3ZBF [67]) were drawn using PyMOL software. Residues found within 8 Å from crizotinib are represented by green coloring. The drug-resistant SAV (G2032) is represented as spheres. (**b**) Simulation of mutated arginine (R2032) is shown in the color purple.

2.4. Case Study: The ROS1-G2032R Mutation

The rearrangement of the *ROS1* gene has been reported in several different tumors, including NSCLC [68]. Almost all patients with NSCLC receiving first- or second-generation tyrosine kinase inhibitors develop treatment resistance. The secondary kinase-domain mutation ROS1-G2032R has been identified in patients refractory to treatment with crizotinib, an ALK/ROS1/MET inhibitor [69,70]. The G2032R mutation also significantly reduces the cellular potency of lorlatinib [71]. Alteration of the glycine residue 2032 to arginine in the ROS1 structure is responsible for blocking the drug binding [72,73]. In our predicted system, the ROS1 glycine 2032 is 3.3 Å away from the docking space for crizotinib (within the 8 Å distance surrounding crizotinib), so it would be appropriate to apply the $DRSP_c^i$ model. The amino acid residues surrounding drug-resistant SAVs are found with neutral and low polarizability properties. Glycine is simply an amino acid with a neutral charge, while mutated arginine is a large-volume amino acid with an electrically charged side chain. These massive changes in the microenvironment represent an obstacle for drug binding. Our prediction system predicted correctly in this case.

3. Discussion

A critical problem that has emerged with cancer-targeted therapies is persistent drug resistance, which is linked to various factors including increased drug efflux, epigenetics, inhibition of apoptosis, and drug inactivation [74,75]. These factors lead to escape pathways and adaptive mechanisms that can even include inactivation of the drug. Another challenge in drug resistance is the heterogeneity of tumors [76]. Genetic heterogeneity influences drug responses and causes drug resistance [77,78]. Although several large-scale tumor genomic projects exist, such as the 1000 Genomes Project [79], COSMIC [50] and the TCGA database [80], no models have systematically investigated the relationship between drug resistance and genetic mutations. Clarifying this relationship would benefit the drug development process, by serving as a reference during drug design and modifications. Our DRSP model attempts to provide a comprehensive high-throughput system for the full appreciation of drug resistance. Our approach offers a spatial structure that incorporates microenvironmental properties and interprets how drugs interact with protein residues.

This system is advantageous for not only determining which SAVs are associated with drug resistance, but it also assists with drug development.

When we divided our prediction system into two subgroups, based on the 8 Å distance between docked drug and protein, we found that combining the interior (within the 8 Å distance) and exterior (beyond the 8 Å distance) models resulted in superior performances for both prediction models (simulated and crystal complexes). We suggest that this phenomenon may be related to the location of the SAV, indicating that it is closely associated with the drug interaction. The binding pocket is a dynamic, active site that hosts drug-protein interactions. Within this space, not all of the mutations are capable of influencing drug resistance; each amino residue within the 8 Å distance has a distinct function, such as arginine and histidine, which are more likely to be located in the ligand binding site [81,82]. More research is needed to explore these discrepant functions.

Our drug prediction systems analyzed selected features from the energy arising from interactions between drug molecules and SAVs, microenvironmental properties surrounding SAVs, protein structural characteristics and sequence conservation profiles of the SAVs (Figure 4). The four models share three features: the weighted contact number of aromatic amino acid, W-E-aro (FWY); the weighted contact number of sulfur amino acid, W-E-sul (CM), and an average entropy value containing 11 residues, ETP-avg11. Phenylalanine, tyrosine and tryptophan play a key role in the stability of the folding structure, while cysteine and methionine are important for structural maintenance; both molecules are hydrophobic. ETP-avg11 values may assist with evolutionary conservation.

Figure 4. Selected features that were applied in the four drug prediction models.

Other features were selected for distances within 8 Å in the simulated and crystal complexes: the weighted contact number of the medium-volume amino acid, W-V-med (NVEQIL); the weighted contact number of the amino acid with neutral polarity, W-P-neu (PATGS); a BLOSUM62 substitution matrix, SSI-b62; and an average value over 7 residues, ETP-avg7. Around 40% of drug-resistant SAVs are located beyond the drug-binding pocket, but also lead to drug resistance. We also identified several features that are located beyond the 8 Å distance from the ligand: van der Waals forces, DKE-vdw; the weighted contact number of oxygen atoms, WCN-o; the weighted contact number of the amino acid with neutral polarity, W-H-neu (GASCTPHY); the weighted contact number of the small-volume amino acid, W-V-sma (GASCTPD); the weighted contact number of the amino acid with low polarizability, W-Z-low (GASDT); the weighted contact number of the acidic amino acid, W-E-aci (DE); a secondary structure, SSE; an H-bond acceptor of the backbone H-bond energy, EHB-acc; and an average value over 5 residues, ETP-avg5. Although further study is needed for most of these selected features to determine the mechanisms of drug resistance, our results indicate that it is possible to reliably predict drug-resistant mutations.

The greatest obstacle facing any attempt to reliably predict drug-resistant mutations is the identification of those SAVs capable of influencing drug resistance. The scant data in the literature makes it challenging to link with certainty any SAVs with drug resistance. Similarly, although several reliable prediction tools such as SIFT and PolyPhen can determine the pathogenic degree of variants, their rarity makes them difficult to define and their potential for drug resistance is uncertain. As illustrated by the TCGA records, only 132 BRAF missense variants (48 of which are predicted to be deleterious, i.e., associated with tumorigenesis, but are not necessarily drug-resistant) have been identified in 731 cancer patients across 26 tumor types; 82.76% of those patients have mutations in

valine residue 600. The frequencies of other mutations amount to less than 2%, and most mutations are only detected once. Interestingly, both our crystal and simulated models predicted three mutations (F516L, M517V, and G596S) as capable of causing drug resistance (Figure 5a). Although the evidence is insufficient at this time, further research is warranted with these variants. The TCGA records have revealed the MAP2K2 protein in association with 51 missense variants in 57 patients across 16 tumor types; all of these variants are rare, as each appears only twice at the most. Moreover, 24 of these missense variants are deleterious. Interestingly, both of our prediction systems identified one of the 24 missense variants—mutation of G83 to serine was predicted as a drug-resistant mutation (Figure 5b). G83 is in a glycine-rich conservative loop, a catalytic region for binding and positioning ATP [83]. Although this mutation is not directly involved in drug interactions, it might alter drug binding and subsequently lead to drug resistance. TCGA data show that the ROS1 protein is found in 475 missense variants (40 of which have been identified by SIFT and PolyPhen as deleterious) in 384 patients across 29 cancer types; the frequencies of these variants are below 1%. One of those 40 missense variants (S2088 mutated to phenylalanine) was predicted to be a drug-resistant mutation by both of our systems (Figure 5c). Although most cancer-related mutations are rare, differences between individual patterns are important for precise medicine for tailoring targeting therapy.

(**a**)　　　　　　　(**b**)　　　　　　　(**c**)
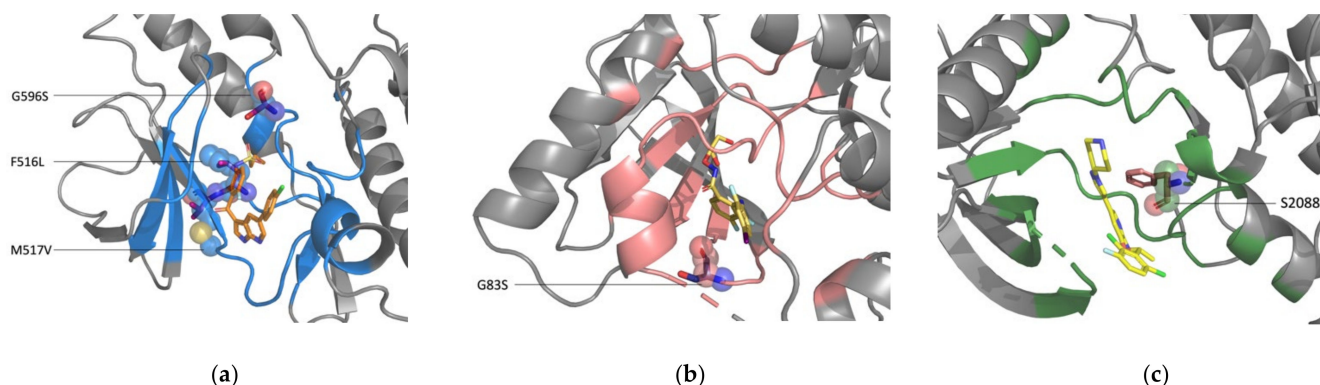
Figure 5. Simulated structures of the protein-drug complexes with mutation alterations. Wild-type amino acids are represented as spheres and mutated-type amino acids as sticks. (**a**) The BRAF-vemurafenib complex with the F516L, M517V, and G596S mutations. (**b**) The MAP2K2-PD0325901 structure complex with the G83S mutation. (**c**) The ROS1-crizotinib structure complex with the S2088F mutation.

Although the crystal drug-protein complex might more realistically reflect the binding condition, the crystallization structure has some limitations. Not all of the protein has the drug-protein co-crystal complex. Thus, we built the crystal complex and simulated two prediction systems to confirm the influence of SAVs on drug resistance. Our modeled systems can also analyze whether a particular SAV influences drug resistance without the crystal complex. Before performing any prediction, we would use the docking method to create the protein-drug complex. Although the conformation of the protein structure might differ slightly due to binding with different compounds, the binding site would be the identical region. A comparison of our stimulated and crystal complexes reveals consistent drug binding sites. Our stimulated complex is therefore reliable.

The primary impediment facing our study research is the small pool of data available for analysis. Although many large-scale cancer genome projects exist, scant information is available regarding drug resistance. Another challenge is how to confirm which SAVs would likely impact drug resistance. Understanding which SAVs influence the effects of a drug is invaluable for research and drug development in the cancer field. Our system provides characteristics of SAVs that help to elucidate their functions. The features used by our prediction systems may serve as valuable factors for further study.

4. Materials and Methods

4.1. Dataset Preparation

All SAV-related data were obtained from The Cancer Genome Atlas (TCGA) [80] and COSMIC databases [50]. The TCGA project has collected and sequenced gene mutations across many different cancers, while COSMIC is the largest available global database that includes somatic mutations from human cancers. To date, 582 unique drug-resistance mutations located in 22 genes have been annotated in COSMIC and 2,531 missense mutations located in these genes have been recorded in the TCGA. The impacts of these mutations can be appropriately interpreted by methodologies such as those used in studies predicting changes in amino acids that affect protein function, as with the SIFT (Sorting Intolerant from Tolerant) program [84] and by methodology that predicts the impact of protein sequence variants, such as PolyPhen [85]. For our study, non-drug-resistant SAVs were collected which were identified as deleterious by SIFT and PolyPhen. All SAVs were mapped to the identified protein structure obtained from the Protein Data Bank (PDB) via the Universal Protein Resource (UniProt). SAVs that lacked structure, were duplicated, or were ambiguous were filtered out, leaving a total of 136 drug-resistant and 589 non-drug-resistant SAVs in our dataset, which were classified into training and testing sets. The training set contained 11 protein-drug crystal complexes and included 133 drug-resistant and 477 non-drug-resistant SAVs. PDB alphanumeric identifiers (IDs) and the numbers of drug-resistant and non-drug-resistant SAVs are listed for each training set protein-drug complex in Table 3.

Table 3. Numbers of drug-resistant and non-drug-resistant SAVs for each protein and drug in the training set.

Protein	Drug	PDB ID	Drug-Resistant [1]	Non-Drug-Resistant [2]
ABL1	Imatinib	1OPJ	31	36
ALK	Alectinib	3AOX	24	50
BTK	Ibrutinib	5P9I	4	36
EGFR	Osimertinib	4ZAU	15	54
ESR1	Raloxifene	1ERR	6	23
FLT3	Quizartinib	4RT7	5	48
KIT	Imatinib	1T46	21	51
MAP2K1	PD0325901	3VVH	2	31
PDGFRA	Sunitinib	6JOK	1	65
SMO	Vismodegib	5L7I	17	42
MET	Crizotinib	2WGJ	7	41
TOTAL			133	477

[1] Numbers of drug-resistant SAVs; [2] Numbers of non-drug-resistant SAVs.

The testing set was used for the performance evaluation, which contained three drug-resistant SAVs (BRAF, MAP2K2, and ROS1) located in three proteins (Table 4). The BRAF-vemurafenib, MAP2K2-PD0325901, and ROS1-crizotinib complexes were simulated by the molecular docking method.

Table 4. Numbers of drug-resistant and deleterious SAVs for each protein and drug in the testing set.

Protein	Drug	PDB ID	Drug-Resistant [1]	Non-Drug-Resistant [2]
BRAF	Vemurafenib	3TV6	1	48
MAP2K2	PD0325901	1S9I	1	24
ROS1	Crizotinib	3ZBF	1	40
TOTAL			3	112

[1] Numbers of drug-resistant SAVs; [2] Numbers of non-drug-resistant SAVs.

4.2. Construction of Prediction Systems

The machine learning method was used to build two DRSP systems. The first system contained two prediction models, $DRSP_c$ and $DRSP_s$, constructed according to the features generated by the crystal and simulated protein-drug complexes of the training set, respectively. The simulated protein-drug complexes of the training set were obtained by removing non-protein molecules from the protein structures and docking the drug into the protein structure. The second prediction system divided the training set into two subgroups, according to whether the minimum distance between drug and protein was greater or less than 8 Å in the protein-drug crystal complexes. Figure 6 shows the numbers of drug-resistant and non-drug-resistant SAV distributions for each distance interval (2 Å). A total of 60.09% drug-resistant SAVs were within 8 Å, so were regarded as located within the drug-binding pocket of proteins. These SAVs are considered to be capable of directly affecting the binding of the drug. The remaining 39.91% of drug-resistant SAVs were beyond the drug-binding pocket of proteins and were regarded as indirectly affecting drug binding activity. The training set was split into two subgroups (e and i), and four prediction models were built: $DRSP_c^e$ and $DRSP_c^i$ for the crystal protein-drug complexes and $DRSP_s^e$ and $DRSP_s^i$ for the simulated protein-drug complexes.
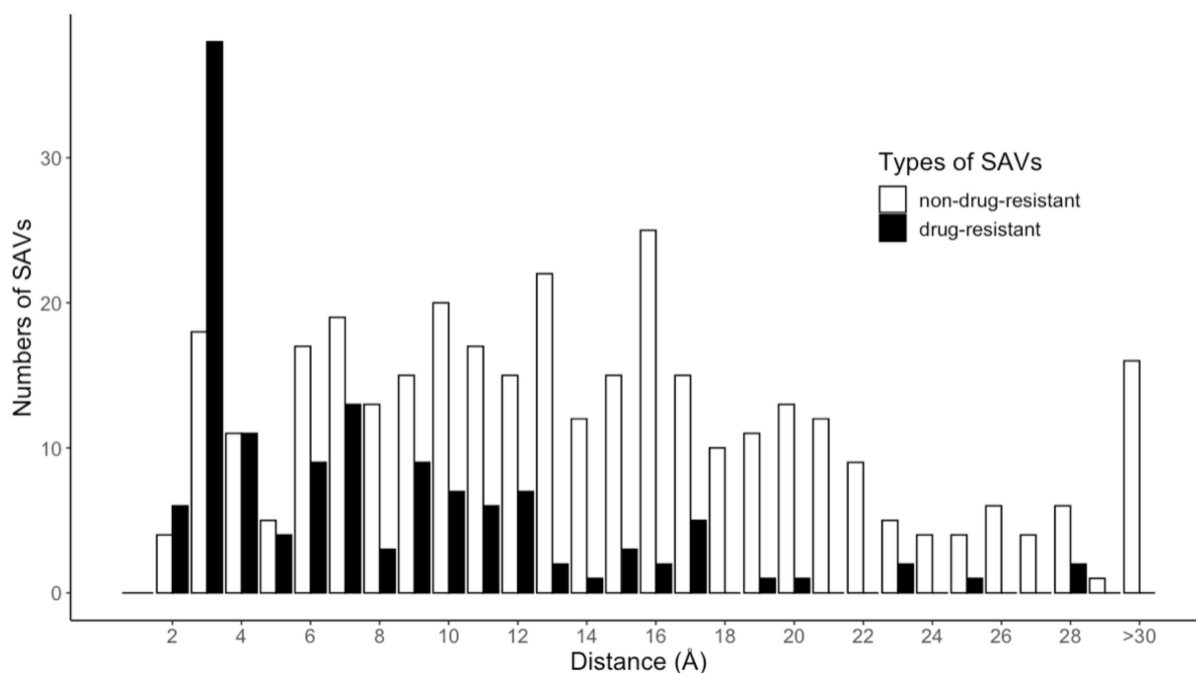
Figure 6. Distributions of drug-resistant and non-drug-resistant SAVs in the training set.

4.3. Machine Learning Method

Each prediction model was an SVM classifier module. The SVM method is a supervised learning model that uses statistical risk minimization to estimate the hyperplane of a classification. This model is widely used for classifying protein structure or function in computational biology [86–91]. All SVM calculations were performed using LIBSVM (version 3.24) [92,93], incorporating the radial basis function (RBF) kernel. The RBF kernel is the most generalized form of kernelization and is widely used because of its similarity to the Gaussian distribution. Optimized classification was identified by using hyperparameter tuning techniques for the given datasets. The parameters (penalty and gamma values of the RBF kernel) were trained by exponentially increasing the grid search from 2^{-15} to 2^{15}, incorporating best values of informative measures with a 5-fold cross-validation during model training.

4.4. Feature Selection

The genetic algorithm (GA) [94–96] was used to select critical features and optimize classification performance. The basic GA procedures were as follows: N solutions (S_i, $i = 1, \ldots, N$) were randomly generated in the initial population, with each solution S_i represented as a set of m-dimensional feature vectors (f_j^i, $j = 1, \ldots, m$) indicating the binary representations of m features. If $f_j^i = 1$, the jth feature was retained; if $f_j^i = 0$, the feature j^{th} was eliminated for feeding into the SVM. For each generation of τ, the three basic mechanisms driving the evolutionary processes were performed, consisting of the selection, mutation and crossover processes. The selection operators were defined as $\alpha^\tau = \max\left\{S_1^\tau, \ldots, S_{N/2}^\tau, \alpha^{\tau-1}\right\}$ and $\beta^\tau = \max\left\{S_{\frac{N}{2}+1}^\tau, \ldots, S_N^\tau, \beta^{\tau-1}\right\}$. The solutions α^τ and β^τ had the best fitness values in each half of the N solutions and $\alpha^{\tau-1}$ and $\beta^{\tau-1}$ in previous generations, respectively. For the special case of $\tau = 0$, the fitness values of α^0 and β^0 were defined as 0. A new solution in next-generation $\tau + 1$, $S_i^{\tau+1}$, was considered equal to α^τ if i was an odd numerical value, while $S_i^{\tau+1}$ was equal to β^τ if i was an even numerical value.

Four informative measures (Equations (1)–(4)) calculated from the 5-fold cross-validation were used as the fitness functions in the selection process. They consisted of accuracy (Acc), the MCC, and the F1 score (F1), as well as the summation of sensitivity and weighted specificity (Hybrid) values, and were calculated as follows:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{2}$$

$$F1 = \frac{2 \times Precision \times Sensitivity}{Precision + Sensitivity} \tag{3}$$

$$Hybrid = Sensitivity + \delta \times Specificity \tag{4}$$

where $Precision = \frac{TP}{TP+FP}$, $Sensitivity = \frac{TP}{TP+FN}$, and $Specificity = \frac{TN}{TN+FP}$, TP represents true-positives, TN represents true-negatives, FP represents false-positives, FN represents false-negatives and δ is the ratio of the number of positives to negatives.

After inputting the selection operators, two types of mutations were applied to the N solution S_is. In the case of $i = 1, \ldots, N/2$, every b bit of the vectors was subject to mutation: $b = \sim b$, if the mutation rate was less than 0.1. In the case of $i = \frac{N}{2} + 1, \ldots, N$, we randomly chose a bit from each vector and subjected the bits to mutation without any mutation thresholds. The one-point crossover operations were carried out between S_{2p-1} and S_{2p}, where $p = 1, \ldots, N/2$ and proceeded as follows: the feature vectors from r to m of S_{2p-1} and S_{2p} were swapped if the crossover rate was less than 0.5, where r was randomly selected from 1 to m.

In this work, the processes of feature selections and model training were repeated for a total of 8 runs. For each run, the number of generations was 150 and there were 80 solutions ($N = 80$) for each generation. Finally, the best feature set and predictive performances were selected with their optimized MCC values. The prediction system is shown schematically in Figure 7.
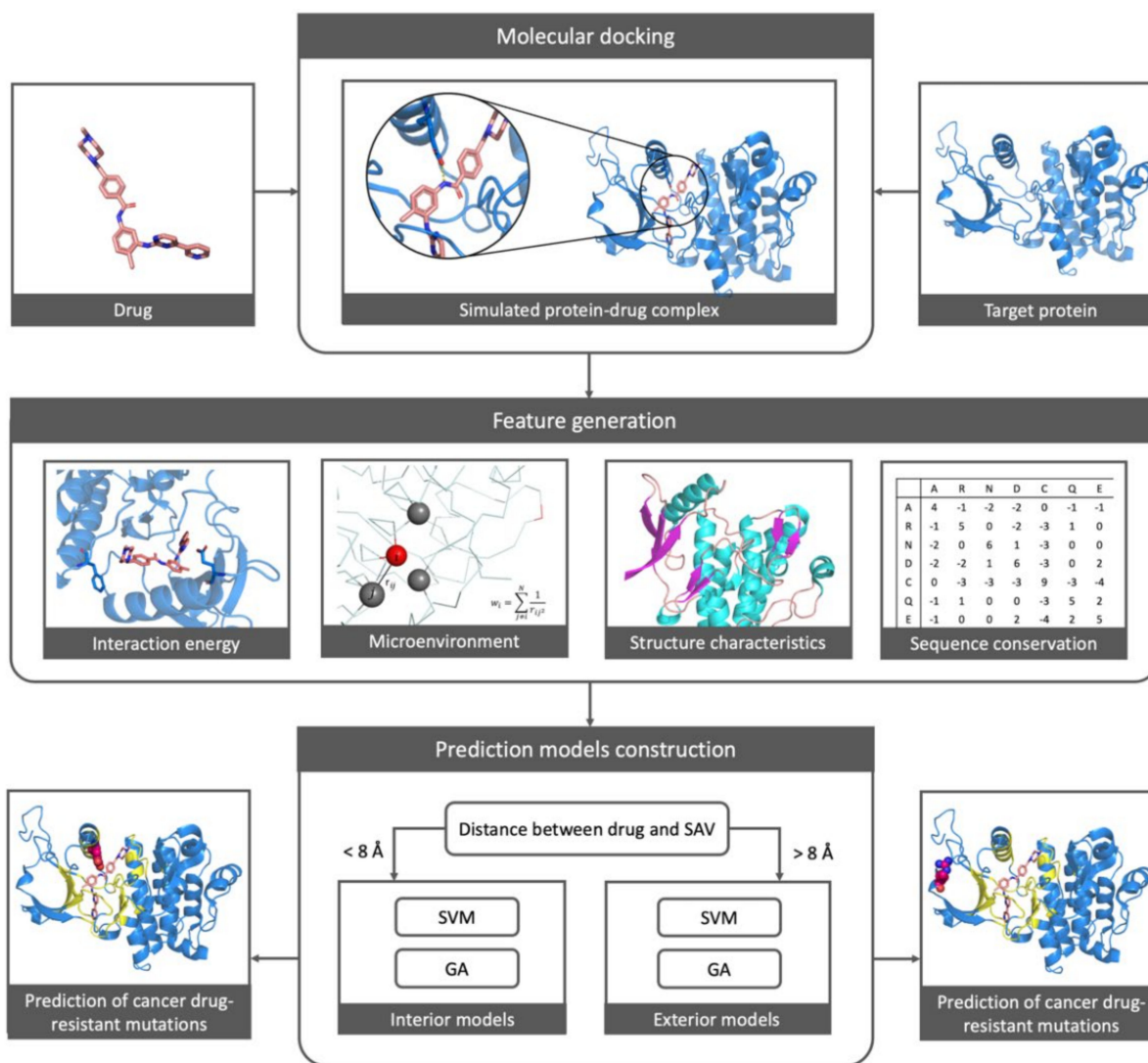
Figure 7. The workflow diagram represents the study's prediction system for cancer drug resistance.

4.5. Generation of Feature Sets

In this study, we used 45 features to describe SAV characteristics, which were classified into four categories: the interaction energy between the drug and the SAV; the microenvironmental properties surrounding the SAV; the protein structural characteristics of the SAV; and the sequence conservation of the SAV (Table S1). Three types of the interaction energies between drugs and SAVs (electrostatic, hydrogen-bonding, and van der Waals forces) were calculated from the crystal or simulated protein-drug complexes using iGEM-DOCK software [97], a widely used molecular docking program that provides accurate predictions of protein-molecule interactions. Using iGEMDOCK with an accurate set of parameters ensured that the drugs were docked into specific proteins and the simulated protein-drug complexes were also obtained.

The weighted contact number (WCN) model [98] was used to describe our 26 microenvironment-associated features of SAVs. The local packing density profile of this WCN model is highly correlated with the sequence conservation profile [99]. The WCN value of atom *i* was calculated by $\text{WCN}_i = \sum_{j \neq i}^{N} \frac{1}{r_{ij}^2}$, where r_{ij} was the distance between atom *i* and atom *j*, while *N* was the number of calculated atoms. In this work, atom *i* was defined as the Cα atom of SAV, and the different microenvironment properties were represented by calculated different atom types or the source of atom *j*. The atom type of *j* could be Cα atoms, nitrogen atoms or oxygen atoms of an amino acid, representing the residue-,

nitrogen- and oxygen-packing densities of a SAV, respectively. The packing density of the SAVs were then divided into different classifications representing the microenvironmental properties where the SAV was located. According to the physicochemical properties of the residues, we used the following classification schemes [100] to refer to the different microenvironments surrounding the amino acid residues: H for polar (RKEDQN), neutral (GASTPHY), and hydrophobic (CVLIMFW); V for small (GASCTPD), medium (NVEQIL), and large (MHKFRYW); Z for low (GASDT), medium (CPNVEQIL), and high polarizability (KMHFRYW); P for low (LIFWCMVY), neutral (PATGS), and high polarity (HQRKNED); F for charged (DEHKR), polar (CGNQSTY), and nonpolar (AFILMPVW); and E for acidic (DE), basic (HKR), aromatic (FWY), amide (NQ), small hydroxyl (ST), sulfur-containing (CM), aliphatic 1 (AGP), and aliphatic 2 (ILV).

Five structure-associated features were derived from the PDB and DSSP databases [101,102]. Firstly, the B-factor value of the SAV $C\alpha$ atom was identified; this value represents the diminished intensity of scattered X-rays after atoms are displaced from their mean positions in a crystal structure. This displacement may be the result of temperature-dependent atomic vibrations, or because of static disorder in a crystal lattice. Our model also used critical DSSP information regarding solvent accessibility and energy derived from the acceptor and donor backbone hydrogen bonds. The last structural descriptors represent the secondary structure elements of SAV defined by DSSP; 1 for the α-helix (H, G and I), -1 for the β-sheet (B and E) and 0 for loop (T, S, and others).

In the last category of characteristics, 11 sequence-associated features were used in this study. The three commonly used substitution indices were used; the BLOSUM62 [103,104], PAM250 [105], and the position-specific scoring matrix (PSSM), which was derived from PSI-BLAST [106]. The evolutional entropy values derived from PSI-BLAST were used to denote a sliding window containing several amino acids on either side of the SAV. This window of amino acids was used to calculate average entropy values. Window lengths of 1, 3, 5, 7, 9, 11, 13 and 15 are centered on the SAV, representing sequence conservations from near to far-ranging.

5. Conclusions

In this project, we developed a drug-resistant prediction system that provides SAV environmental properties comprising the interaction energy, structure characteristics, and microenvironmental components. Although drug resistance is a more complicated mechanism involving more than a single mutation, an in-depth evaluation of the SAVs can determine how a particular SAV is associated with drug resistance and its value for drug development. Moreover, although targeted therapy is the preferred strategy for cancer treatment, it needs to consider molecular heterogeneity for the cancer being treated and the particular mutation pattern. Our prediction system may also be applied to different mutation patterns. The information gleaned can be used to generate individually-specific treatment strategies for true precision medicine.

References

1. Sawyers, C. Targeted cancer therapy. *Nature* **2004**, *432*, 294–297. [CrossRef]
2. Camidge, D.R.; Pao, W.; Sequist, L.V. Acquired resistance to TKIs in solid tumours: Learning from lung cancer. *Nat. Rev. Clin. Oncol.* **2014**, *11*, 473–481. [CrossRef]
3. Tukagoshi, S. Cancer chemotherapy; past, present and future—From the aspect of fundamental studies. *Gan Kagaku Ryoho* **2003**, *30*, 1398–1403.
4. Asano, T. Drug Resistance in Cancer Therapy and the Role of Epigenetics. *J. Nippon Med. Sch.* **2020**, *87*, 244–251. [CrossRef]
5. Hinds, M.; Deisseroth, K.; Mayes, J.; Altschuler, E.; Jansen, R.; Ledley, F.D.; Zwelling, L.A. Identification of a point mutation in the topoisomerase II gene from a human leukemia cell line containing an amsacrine-resistant form of topoisomerase II. *Cancer Res.* **1991**, *51*, 4729–4731.
6. Jhaveri, M.S.; Morrow, C.S. Methylation-mediated regulation of the glutathione S-transferase P1 gene in human breast cancer cells. *Gene* **1998**, *210*, 1–7. [CrossRef]
7. Cabral, F.R.; Brady, R.C.; Schibler, M.J. A mechanism of cellular resistance to drugs that interfere with microtubule assembly. *Ann. N. Y. Acad. Sci.* **1986**, *466*, 745–756. [CrossRef]
8. Moscow, J.A.; Cowan, K.H. Multidrug resistance. *J. Natl. Cancer Inst.* **1988**, *80*, 14–20. [CrossRef]
9. Zhang, Y.; Yang, S.H.; Guo, X.L. New insights into Vinca alkaloids resistance mechanism and circumvention in lung cancer. *Biomed. Pharmacother.* **2017**, *96*, 659–666. [CrossRef]
10. Tsuruo, T. Mechanisms of multidrug resistance and implications for therapy. *Jpn. J. Cancer Res.* **1988**, *79*, 285–296. [CrossRef]
11. Yao, Y.; Zhou, Y.; Liu, L.; Xu, Y.; Chen, Q.; Wang, Y.; Wu, S.; Deng, Y.; Zhang, J.; Shao, A. Nanoparticle-Based Drug Delivery in Cancer Therapy and Its Role in Overcoming Drug Resistance. *Front. Mol. Biosci.* **2020**, *7*, 193. [CrossRef]
12. Ashrafizadeh, M.; Mirzaei, S.; Gholami, M.H.; Hashemi, F.; Zabolian, A.; Raei, M.; Hushmandi, K.; Zarrabi, A.; Voelcker, N.H.; Aref, A.R.; et al. Hyaluronic acid-based nanoplatforms for Doxorubicin: A review of stimuli-responsive carriers, co-delivery and resistance suppression. *Carbohydr. Polym.* **2021**, *272*, 118491. [CrossRef]
13. Mirzaei, S.; Gholami, M.H.; Hashemi, F.; Zabolian, A.; Farahani, M.V.; Hushmandi, K.; Zarrabi, A.; Goldman, A.; Ashrafizadeh, M.; Orive, G. Advances in understanding the role of P-gp in doxorubicin resistance: Molecular pathways, therapeutic strategies, and prospects. *Drug Discov. Today* **2021**. [CrossRef]
14. Maleki Dana, P.; Sadoughi, F.; Asemi, Z.; Yousefi, B. The role of polyphenols in overcoming cancer drug resistance: A comprehensive review. *Cell Mol. Biol. Lett.* **2022**, *27*, 1–26. [CrossRef]
15. Villanueva, J.; Vultur, A.; Herlyn, M. Resistance to BRAF inhibitors: Unraveling mechanisms and future treatment options. *Cancer Res.* **2011**, *71*, 7137–7140. [CrossRef]
16. Mok, T.S.; Wu, Y.L.; Thongprasert, S.; Yang, C.H.; Chu, D.T.; Saijo, N.; Sunpaweravong, P.; Han, B.; Margono, B.; Ichinose, Y.; et al. Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. *N. Engl. J. Med.* **2009**, *361*, 947–957. [CrossRef]
17. Rosell, R.; Carcereny, E.; Gervais, R.; Vergnenegre, A.; Massuti, B.; Felip, E.; Palmero, R.; Garcia-Gomez, R.; Pallares, C.; Sanchez, J.M.; et al. Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): A multicentre, open-label, randomised phase 3 trial. *Lancet Oncol.* **2012**, *13*, 239–246. [CrossRef]
18. Lee, S.M.; Khan, I.; Upadhyay, S.; Lewanski, C.; Falk, S.; Skailes, G.; Marshall, E.; Woll, P.J.; Hatton, M.; Lal, R.; et al. First-line erlotinib in patients with advanced non-small-cell lung cancer unsuitable for chemotherapy (TOPICAL): A double-blind, placebo-controlled, phase 3 trial. *Lancet Oncol.* **2012**, *13*, 1161–1170. [CrossRef]
19. He, C.; Wang, Y. Role of the EGFR-KDD mutation as a possible mechanism of acquired resistance of non-small cell lung cancer to EGFR tyrosine kinase inhibitors: A case report. *Mol. Clin. Oncol.* **2022**, *16*, 30. [CrossRef]
20. Jackman, D.; Pao, W.; Riely, G.J.; Engelman, J.A.; Kris, M.G.; Janne, P.A.; Lynch, T.; Johnson, B.E.; Miller, V.A. Clinical definition of acquired resistance to epidermal growth factor receptor tyrosine kinase inhibitors in non-small-cell lung cancer. *J. Clin. Oncol.* **2010**, *28*, 357–360. [CrossRef]
21. West, H.; Oxnard, G.R.; Doebele, R.C. Acquired resistance to targeted therapies in advanced non-small cell lung cancer: New strategies and new agents. *Am. Soc. Clin. Oncol. Educ. Book* **2013**, *33*, e272–e278. [CrossRef]
22. Neel, D.S.; Bivona, T.G. Resistance is futile: Overcoming resistance to targeted therapies in lung adenocarcinoma. *NPJ Precis. Oncol.* **2017**, *1*, 1–6. [CrossRef]
23. Gottesman, M.M.; Lavi, O.; Hall, M.D.; Gillet, J.P. Toward a Better Understanding of the Complexity of Cancer Drug Resistance. *Annu. Rev. Pharmacol. Toxicol.* **2016**, *56*, 85–102. [CrossRef]
24. Obenauf, A.C.; Zou, Y.; Ji, A.L.; Vanharanta, S.; Shu, W.; Shi, H.; Kong, X.; Bosenberg, M.C.; Wiesner, T.; Rosen, N.; et al. Therapy-induced tumour secretomes promote resistance and tumour progression. *Nature* **2015**, *520*, 368–372. [CrossRef]

25. Straussman, R.; Morikawa, T.; Shee, K.; Barzily-Rokni, M.; Qian, Z.R.; Du, J.; Davis, A.; Mongare, M.M.; Gould, J.; Frederick, D.T.; et al. Tumour micro-environment elicits innate resistance to RAF inhibitors through HGF secretion. *Nature* **2012**, *487*, 500–504. [CrossRef]

26. Nayar, U.; Cohen, O.; Kapstad, C.; Cuoco, M.S.; Waks, A.G.; Wander, S.A.; Painter, C.; Freeman, S.; Persky, N.S.; Marini, L.; et al. Acquired HER2 mutations in ER(+) metastatic breast cancer confer resistance to estrogen receptor-directed therapies. *Nat. Genet.* **2019**, *51*, 207–216. [CrossRef]

27. Mao, X.G.; Wang, C.; Liu, D.Y.; Zhang, X.; Wang, L.; Yan, M.; Zhang, W.; Zhu, J.; Li, Z.C.; Mi, C.; et al. Hypoxia upregulates HIG2 expression and contributes to bevacizumab resistance in glioblastoma. *Oncotarget* **2016**, *7*, 47808–47820. [CrossRef]

28. Li, L.; Nan, F.; Guo, Q.; Guan, D.; Zhou, C. Resistance to bevacizumab in ovarian cancer SKOV3 xenograft due to EphB4 overexpression. *J. Cancer Res. Ther.* **2019**, *15*, 1282–1287. [CrossRef]

29. Barretina, J.; Caponigro, G.; Stransky, N.; Venkatesan, K.; Margolin, A.A.; Kim, S.; Wilson, C.J.; Lehar, J.; Kryukov, G.V.; Sonkin, D.; et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **2012**, *483*, 603–607. [CrossRef]

30. Yang, W.; Soares, J.; Greninger, P.; Edelman, E.J.; Lightfoot, H.; Forbes, S.; Bindal, N.; Beare, D.; Smith, J.A.; Thompson, I.R.; et al. Genomics of Drug Sensitivity in Cancer (GDSC): A resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.* **2013**, *41*, D955–D961. [CrossRef]

31. Basu, A.; Bodycombe, N.E.; Cheah, J.H.; Price, E.V.; Liu, K.; Schaefer, G.I.; Ebright, R.Y.; Stewart, M.L.; Ito, D.; Wang, S.; et al. An interactive resource to identify cancer genetic and lineage dependencies targeted by small molecules. *Cell* **2013**, *154*, 1151–1161. [CrossRef]

32. Seashore-Ludlow, B.; Rees, M.G.; Cheah, J.H.; Cokol, M.; Price, E.V.; Coletti, M.E.; Jones, V.; Bodycombe, N.E.; Soule, C.K.; Gould, J.; et al. Harnessing Connectivity in a Large-Scale Small-Molecule Sensitivity Dataset. *Cancer Discov.* **2015**, *5*, 1210–1223. [CrossRef]

33. Garnett, M.J.; Edelman, E.J.; Heidorn, S.J.; Greenman, C.D.; Dastur, A.; Lau, K.W.; Greninger, P.; Thompson, I.R.; Luo, X.; Soares, J.; et al. Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* **2012**, *483*, 570–575. [CrossRef]

34. Iorio, F.; Knijnenburg, T.A.; Vis, D.J.; Bignell, G.R.; Menden, M.P.; Schubert, M.; Aben, N.; Goncalves, E.; Barthorpe, S.; Lightfoot, H.; et al. A Landscape of Pharmacogenomic Interactions in Cancer. *Cell* **2016**, *166*, 740–754. [CrossRef]

35. Cortes-Ciriano, I.; van Westen, G.J.; Bouvier, G.; Nilges, M.; Overington, J.P.; Bender, A.; Malliavin, T.E. Improved large-scale prediction of growth inhibition patterns using the NCI60 cancer cell line panel. *Bioinformatics* **2016**, *32*, 85–95. [CrossRef]

36. Naulaerts, S.; Dang, C.C.; Ballester, P.J. Precision and recall oncology: Combining multiple gene mutations for improved identification of drug-sensitive tumours. *Oncotarget* **2017**, *8*, 97025–97040. [CrossRef]

37. Gayvert, K.M.; Aly, O.; Platt, J.; Bosenberg, M.W.; Stern, D.F.; Elemento, O. A Computational Approach for Identifying Synergistic Drug Combinations. *PLoS Comput. Biol.* **2017**, *13*, e1005308. [CrossRef]

38. Huang, C.; Mezencev, R.; McDonald, J.F.; Vannberg, F. Open source machine-learning algorithms for the prediction of optimal cancer drug therapies. *PLoS ONE* **2017**, *12*, e0186906. [CrossRef]

39. Menden, M.P.; Iorio, F.; Garnett, M.; McDermott, U.; Benes, C.H.; Ballester, P.J.; Saez-Rodriguez, J. Machine learning prediction of cancer cell sensitivity to drugs based on genomic and chemical properties. *PLoS ONE* **2013**, *8*, e61318. [CrossRef]

40. Costello, J.C.; Heiser, L.M.; Georgii, E.; Gonen, M.; Menden, M.P.; Wang, N.J.; Bansal, M.; Ammad-ud-din, M.; Hintsanen, P.; Khan, S.A.; et al. A community effort to assess and improve drug sensitivity prediction algorithms. *Nat. Biotechnol.* **2014**, *32*, 1202–1212. [CrossRef]

41. Gonen, M.; Margolin, A.A. Drug susceptibility prediction against a panel of drugs using kernelized Bayesian multitask learning. *Bioinformatics* **2014**, *30*, i556–i563. [CrossRef]

42. Komarova, N.L.; Wodarz, D. Drug resistance in cancer: Principles of emergence and prevention. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 9714–9719. [CrossRef]

43. Faratian, D.; Goltsov, A.; Lebedeva, G.; Sorokin, A.; Moodie, S.; Mullen, P.; Kay, C.; Um, I.H.; Langdon, S.; Goryanin, I.; et al. Systems biology reveals new strategies for personalizing cancer medicine and confirms the role of PTEN in resistance to trastuzumab. *Cancer Res.* **2009**, *69*, 6713–6720. [CrossRef]

44. Tomasetti, C.; Levy, D. An elementary approach to modeling drug resistance in cancer. *Math. Biosci. Eng.* **2010**, *7*, 905–918. [CrossRef]

45. Sun, X.; Bao, J.; Nelson, K.C.; Li, K.C.; Kulik, G.; Zhou, X. Systems modeling of anti-apoptotic pathways in prostate cancer: Psychological stress triggers a synergism pattern switch in drug combination therapy. *PLoS Comput. Biol.* **2013**, *9*, e1003358. [CrossRef]

46. Choi, J.; Park, S.; Ahn, J. RefDNN: A reference drug based neural network for more accurate prediction of anticancer drug resistance. *Sci. Rep.* **2020**, *10*, 1861. [CrossRef]

47. Jimenez, C.R.; Zhang, H.; Kinsinger, C.R.; Nice, E.C. The cancer proteomic landscape and the HUPO Cancer Proteome Project. *Clin. Proteom.* **2018**, *15*, 4. [CrossRef]

48. Kato, S.; Han, S.Y.; Liu, W.; Otsuka, K.; Shibata, H.; Kanamaru, R.; Ishioka, C. Understanding the function-structure and function-mutation relationships of p53 tumor suppressor protein by high-resolution missense mutation analysis. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 8424–8429. [CrossRef]

49. Soskine, M.; Tawfik, D.S. Mutational effects and the evolution of new protein functions. *Nat. Rev. Genet.* **2010**, *11*, 572–582. [CrossRef]

50. Tate, J.G.; Bamford, S.; Jubb, H.C.; Sondka, Z.; Beare, D.M.; Bindal, N.; Boutselakis, H.; Cole, C.G.; Creatore, C.; Dawson, E.; et al. COSMIC: The Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res.* **2019**, *47*, D941–D947. [CrossRef]

51. Wagenaar, T.R.; Ma, L.; Roscoe, B.; Park, S.M.; Bolon, D.N.; Green, M.R. Resistance to vemurafenib resulting from a novel mutation in the BRAFV600E kinase domain. *Pigment. Cell Melanoma Res.* **2014**, *27*, 124–133. [CrossRef]

52. Gou, W.; Zhou, X.; Liu, Z.; Wang, L.; Shen, J.; Xu, X.; Li, Z.; Zhai, X.; Zuo, D.; Wu, Y. CD74-ROS1 G2032R mutation transcriptionally up-regulates Twist1 in non-small cell lung cancer cells leading to increased migration, invasion, and resistance to crizotinib. *Cancer Lett.* **2018**, *422*, 19–28. [CrossRef]

53. Pratilas, C.A.; Xing, F.; Solit, D.B. Targeting oncogenic BRAF in human cancer. *Curr. Top. Microbiol. Immunol.* **2012**, *355*, 83–98. [CrossRef]

54. Long, G.V.; Stroyakovskiy, D.; Gogas, H.; Levchenko, E.; de Braud, F.; Larkin, J.; Garbe, C.; Jouary, T.; Hauschild, A.; Grob, J.J.; et al. Combined BRAF and MEK inhibition versus BRAF inhibition alone in melanoma. *N. Engl. J. Med.* **2014**, *371*, 1877–1888. [CrossRef]

55. Barras, D. BRAF Mutation in Colorectal Cancer: An Update. *Biomark Cancer* **2015**, *7* (Suppl. S1), 9–12. [CrossRef]

56. Brose, M.S.; Volpe, P.; Feldman, M.; Kumar, M.; Rishi, I.; Gerrero, R.; Einhorn, E.; Herlyn, M.; Minna, J.; Nicholson, A.; et al. BRAF and RAS mutations in human lung cancer and melanoma. *Cancer Res.* **2002**, *62*, 6997–7000.

57. Van Allen, E.M.; Wagle, N.; Sucker, A.; Treacy, D.J.; Johannessen, C.M.; Goetz, E.M.; Place, C.S.; Taylor-Weiner, A.; Whittaker, S.; Kryukov, G.V.; et al. The genetic landscape of clinical resistance to RAF inhibition in metastatic melanoma. *Cancer Discov.* **2014**, *4*, 94–109. [CrossRef]

58. Wagle, N.; Van Allen, E.M.; Treacy, D.J.; Frederick, D.T.; Cooper, Z.A.; Taylor-Weiner, A.; Rosenberg, M.; Goetz, E.M.; Sullivan, R.J.; Farlow, D.N.; et al. MAP kinase pathway alterations in BRAF-mutant melanoma patients with acquired resistance to combined RAF/MEK inhibition. *Cancer Discov.* **2014**, *4*, 61–68. [CrossRef]

59. Luebker, S.A.; Koepsell, S.A. Diverse Mechanisms of BRAF Inhibitor Resistance in Melanoma Identified in Clinical and Preclinical Studies. *Front. Oncol.* **2019**, *9*, 268. [CrossRef]

60. Hoogstraat, M.; Gadellaa-van Hooijdonk, C.G.; Ubink, I.; Besselink, N.J.; Pieterse, M.; Veldhuis, W.; van Stralen, M.; Meijer, E.F.; Willems, S.M.; Hadders, M.A.; et al. Detailed imaging and genetic analysis reveal a secondary BRAF(L505H) resistance mutation and extensive intrapatient heterogeneity in metastatic BRAF mutant melanoma patients treated with vemurafenib. *Pigment. Cell Melanoma Res.* **2015**, *28*, 318–323. [CrossRef]

61. Rizos, H.; Menzies, A.M.; Pupo, G.M.; Carlino, M.S.; Fung, C.; Hyman, J.; Haydu, L.E.; Mijatov, B.; Becker, T.M.; Boyd, S.C.; et al. BRAF inhibitor resistance mechanisms in metastatic melanoma: Spectrum and clinical impact. *Clin. Cancer Res.* **2014**, *20*, 1965–1977. [CrossRef]

62. Hatzivassiliou, G.; Liu, B.; O'Brien, C.; Spoerke, J.M.; Hoeflich, K.P.; Haverty, P.M.; Soriano, R.; Forrest, W.F.; Heldens, S.; Chen, H.; et al. ERK inhibition overcomes acquired resistance to MEK inhibitors. *Mol. Cancer Ther.* **2012**, *11*, 1143–1154. [CrossRef]

63. Chen, C.C.; Hwang, J.K.; Yang, J.M. (PS)2-v2: Template-based protein structure prediction server. *BMC Bioinform.* **2009**, *10*, 366. [CrossRef]

64. Karoulia, Z.; Wu, Y.; Ahmed, T.A.; Xin, Q.; Bollard, J.; Krepler, C.; Wu, X.; Zhang, C.; Bollag, G.; Herlyn, M.; et al. An Integrated Model of RAF Inhibitor Action Predicts Inhibitor Activity against Oncogenic BRAF Signaling. *Cancer Cell* **2016**, *30*, 485–498. [CrossRef]

65. Wenglowsky, S.; Ren, L.; Ahrendt, K.A.; Laird, E.R.; Aliagas, I.; Alicke, B.; Buckmelter, A.J.; Choo, E.F.; Dinkel, V.; Feng, B.; et al. Pyrazolopyridine Inhibitors of B-Raf(V600E). Part 1: The Development of Selective, Orally Bioavailable, and Efficacious Inhibitors. *ACS Med. Chem. Lett.* **2011**, *2*, 342–347. [CrossRef]

66. Ohren, J.F.; Chen, H.; Pavlovsky, A.; Whitehead, C.; Zhang, E.; Kuffa, P.; Yan, C.; McConnell, P.; Spessard, C.; Banotai, C.; et al. Structures of human MAP kinase kinase 1 (MEK1) and MEK2 describe novel noncompetitive kinase inhibition. *Nat. Struct. Mol. Biol.* **2004**, *11*, 1192–1197. [CrossRef]

67. Awad, M.M.; Katayama, R.; McTigue, M.; Liu, W.; Deng, Y.L.; Brooun, A.; Friboulet, L.; Huang, D.; Falk, M.D.; Timofeevski, S.; et al. Acquired resistance to crizotinib from a mutation in CD74-ROS1. *N. Engl. J. Med.* **2013**, *368*, 2395–2401. [CrossRef]

68. Davies, K.D.; Doebele, R.C. Molecular pathways: ROS1 fusion proteins in cancer. *Clin. Cancer Res.* **2013**, *19*, 4040–4045. [CrossRef]

69. Gainor, J.F.; Tseng, D.; Yoda, S.; Dagogo-Jack, I.; Friboulet, L.; Lin, J.J.; Hubbeling, H.G.; Dardaei, L.; Farago, A.F.; Schultz, K.R.; et al. Patterns of Metastatic Spread and Mechanisms of Resistance to Crizotinib in ROS1-Positive Non-Small-Cell Lung Cancer. *JCO Precis. Oncol.* **2017**, *1*, 1–13. [CrossRef]

70. Song, A.; Kim, T.M.; Kim, D.W.; Kim, S.; Keam, B.; Lee, S.H.; Heo, D.S. Molecular Changes Associated with Acquired Resistance to Crizotinib in ROS1-Rearranged Non-Small Cell Lung Cancer. *Clin. Cancer Res.* **2015**, *21*, 2379–2387. [CrossRef]

71. Lin, J.J.; Shaw, A.T. Recent Advances in Targeting ROS1 in Lung Cancer. *J. Thorac. Oncol.* **2017**, *12*, 1611–1625. [CrossRef]

72. Drilon, A.; Jenkins, C.; Iyer, S.; Schoenfeld, A.; Keddy, C.; Davare, M.A. ROS1-dependent cancers—Biology, diagnostics and therapeutics. *Nat. Rev. Clin. Oncol.* **2021**, *18*, 35–55. [CrossRef]

73. Lee, J.; Sun, J.M.; Lee, S.H.; Ahn, J.S.; Park, K.; Choi, Y.; Ahn, M.J. Efficacy and Safety of Lorlatinib in Korean Non-Small-Cell Lung Cancer Patients With ALK or ROS1 Rearrangement Whose Disease Failed to Respond to a Previous Tyrosine Kinase Inhibitor. *Clin. Lung Cancer* **2019**, *20*, 215–221. [CrossRef]

74. Housman, G.; Byler, S.; Heerboth, S.; Lapinska, K.; Longacre, M.; Snyder, N.; Sarkar, S. Drug resistance in cancer: An overview. *Cancers* **2014**, *6*, 1769–1792. [CrossRef]
75. Mansoori, B.; Mohammadi, A.; Davudian, S.; Shirjang, S.; Baradaran, B. The Different Mechanisms of Cancer Drug Resistance: A Brief Review. *Adv. Pharm. Bull.* **2017**, *7*, 339–348. [CrossRef]
76. Konieczkowski, D.J.; Johannessen, C.M.; Garraway, L.A. A Convergence-Based Framework for Cancer Drug Resistance. *Cancer Cell* **2018**, *33*, 801–815. [CrossRef]
77. Wilmott, J.S.; Tembe, V.; Howle, J.R.; Sharma, R.; Thompson, J.F.; Rizos, H.; Lo, R.S.; Kefford, R.F.; Scolyer, R.A.; Long, G.V. Intratumoral molecular heterogeneity in a BRAF-mutant, BRAF inhibitor-resistant melanoma: A case illustrating the challenges for personalized medicine. *Mol. Cancer Ther.* **2012**, *11*, 2704–2708. [CrossRef]
78. Burrell, R.A.; McGranahan, N.; Bartek, J.; Swanton, C. The causes and consequences of genetic heterogeneity in cancer evolution. *Nature* **2013**, *501*, 338–345. [CrossRef]
79. Genomes Project, C.; Auton, A.; Brooks, L.D.; Durbin, R.M.; Garrison, E.P.; Kang, H.M.; Korbel, J.O.; Marchini, J.L.; McCarthy, S.; McVean, G.A.; et al. A global reference for human genetic variation. *Nature* **2015**, *526*, 68–74. [CrossRef]
80. Cancer Genome Atlas Research, N.; Weinstein, J.N.; Collisson, E.A.; Mills, G.B.; Shaw, K.R.; Ozenberger, B.A.; Ellrott, K.; Shmulevich, I.; Sander, C.; Stuart, J.M. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.* **2013**, *45*, 1113–1120. [CrossRef]
81. Stank, A.; Kokh, D.B.; Fuller, J.C.; Wade, R.C. Protein Binding Pocket Dynamics. *Acc. Chem. Res.* **2016**, *49*, 809–815. [CrossRef]
82. Bianchi, V.; Gherardini, P.F.; Helmer-Citterich, M.; Ausiello, G. Identification of binding pockets in protein structures using a knowledge-based potential derived from local structural similarities. *BMC Bioinform.* **2012**, *13* (Suppl. S4), S17. [CrossRef]
83. Wu, P.K.; Park, J.I. MEK1/2 Inhibitors: Molecular Activity and Resistance Mechanisms. *Semin. Oncol.* **2015**, *42*, 849–862. [CrossRef]
84. Ng, P.C.; Henikoff, S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* **2003**, *31*, 3812–3814. [CrossRef]
85. Adzhubei, I.A.; Schmidt, S.; Peshkin, L.; Ramensky, V.E.; Gerasimova, A.; Bork, P.; Kondrashov, A.S.; Sunyaev, S.R. A method and server for predicting damaging missense mutations. *Nat. Methods* **2010**, *7*, 248–249. [CrossRef]
86. Hua, S.; Sun, Z. A novel method of protein secondary structure prediction with high segment overlap measure: Support vector machine approach. *J. Mol. Biol.* **2001**, *308*, 397–407. [CrossRef]
87. Yu, C.S.; Wang, J.Y.; Yang, J.M.; Lyu, P.C.; Lin, C.J.; Hwang, J.K. Fine-grained protein fold assignment by support vector machines using generalized npeptide coding schemes and jury voting from multiple-parameter sets. *Proteins* **2003**, *50*, 531–536. [CrossRef]
88. Yu, C.S.; Lin, C.J.; Hwang, J.K. Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. *Protein Sci.* **2004**, *13*, 1402–1406. [CrossRef]
89. Chen, Y.C.; Lin, Y.S.; Lin, C.J.; Hwang, J.K. Prediction of the bonding states of cysteines using the support vector machines based on multiple feature vectors and cysteine state sequences. *Proteins* **2004**, *55*, 1036–1042. [CrossRef]
90. Lei, Z.; Dai, Y. An SVM-based system for predicting protein subnuclear localizations. *BMC Bioinform.* **2005**, *6*, 291. [CrossRef]
91. Ward, J.J.; McGuffin, L.J.; Buxton, B.F.; Jones, D.T. Secondary structure prediction with support vector machines. *Bioinformatics* **2003**, *19*, 1650–1655.
92. Chang, C.C.; Lin, C.J. LIBSVM: A Library for Support Vector Machines. *ACM Trans. Intell. Syst. Technol.* **2011**, *2*, 1–27. [CrossRef]
93. Lin, C.-J. Formulations of support vector machines: A note from an optimization point of view. *Neural Comput.* **2001**, *13*, 307–317.
94. Lu, C.H.; Chen, Y.C.; Yu, C.S.; Hwang, J.K. Predicting disulfide connectivity patterns. *Proteins* **2007**, *67*, 262–270. [CrossRef]
95. Yu, C.S.; Lu, C.H. Identification of antifreeze proteins and their functional residues by support vector machine and genetic algorithms based on n-peptide compositions. *PLoS ONE* **2011**, *6*, e20445. [CrossRef]
96. Liu, J.J.; Yu, C.S.; Wu, H.W.; Chang, Y.J.; Lin, C.P.; Lu, C.H. The structure-based cancer-related single amino acid variation prediction. *Sci. Rep.* **2021**, *11*, 13599. [CrossRef]
97. Yang, J.-M.; Chen, C.-C. GEMDOCK: A generic evolutionary method for molecular docking. *Proteins Struct. Funct. Bioinform.* **2004**, *55*, 288–304. [CrossRef]
98. Lin, C.P.; Huang, S.W.; Lai, Y.L.; Yen, S.C.; Shih, C.H.; Lu, C.H.; Huang, C.C.; Hwang, J.K. Deriving protein dynamical properties from weighted protein contact number. *Proteins* **2008**, *72*, 929–935. [CrossRef]
99. Shih, C.H.; Chang, C.M.; Lin, Y.S.; Lo, W.C.; Hwang, J.K. Evolutionary information hidden in a single protein structure. *Proteins* **2012**, *80*, 1647–1657. [CrossRef]
100. Yu, C.S.; Chen, Y.C.; Lu, C.H.; Hwang, J.K. Prediction of protein subcellular localization. *Proteins* **2006**, *64*, 643–651. [CrossRef]
101. Cheng, J.; Randall, A.Z.; Sweredoski, M.J.; Baldi, P. SCRATCH: A protein structure and structural feature prediction server. *Nucleic Acids Res.* **2005**, *33*, W72–W76. [CrossRef]
102. Kabsch, W.; Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, *22*, 2577–2637. [CrossRef]
103. Choi, Y.; Sims, G.E.; Murphy, S.; Miller, J.R.; Chan, A.P. Predicting the functional effect of amino acid substitutions and indels. *PLoS ONE* **2012**, *7*, e46688. [CrossRef]
104. Henikoff, S.; Henikoff, J.G. Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 10915–10919.

105. Jones, D.T.; Taylor, W.R.; Thornton, J.M. The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* **1992**, *8*, 275–282.

106. Altschul, S.F.; Madden, T.L.; Schaffer, A.A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D.J. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25*, 3389–3402. [CrossRef]