Oncological drug discovery: AI meets structure-based computational research

Gorostiola Gonzalez, M.; Janssen, A.P.A.; IJzerman, A.P.; Heitman, L.H.; Westen, G.J.P. van

Oncological drug discovery: AI meets structure-based computational research

Marina Gorostiola González [a,b], Antonius P.A. Janssen [b,c], Adriaan P. IJzerman [a], Laura H. Heitman [a,b], Gerard J.P. van Westen [a,*]

[a] Division of Drug Discovery and Safety, Leiden Academic Centre for Drug Research, Leiden University, the Netherlands
[b] Oncode Institute, Utrecht, the Netherlands
[c] Molecular Physiology, Leiden Institute of Chemistry, Leiden University, the Netherlands

The integration of machine learning and structure-based methods has proven valuable in the past as a way to prioritize targets and compounds in early drug discovery. In oncological research, these methods can be highly beneficial in addressing the diversity of neoplastic diseases portrayed by the different hallmarks of cancer. Here, we review six use case scenarios for integrated computational methods, namely driver prediction, computational mutagenesis, (off)-target prediction, binding site prediction, virtual screening, and allosteric modulation analysis. We address the heterogeneity of integration approaches and individual methods, while acknowledging their current limitations and highlighting their potential to bring drugs for personalized oncological therapies to the market faster.

Keywords: Cancer; Artificial intelligence; Machine learning; Structure-based drug design; Hallmarks of cancer

Introduction

In recent years, the scientific community has seen the increased use of computational approaches to accelerate the discovery of relevant targets and to prioritize small molecules in all disease areas. These approaches include data-driven artificial intelligence (AI)/machine learning (ML),[1,2] as well as structure-based (SB) methods, such as those based on docking and molecular dynamics (MD).[3] Moreover, advances in computing power and in experimental structure elucidation have made it possible to integrate these two types of methods, for example to use ML-based scoring functions to rank the accuracy of docking results,[4] or to use structure-derived data (such as interaction fingerprints or MD trajectories) as input for bioactivity prediction models.[5,6] These advances have emerged from the joint efforts of the computational drug discovery community and are generally applicable to the subfield of oncological drug discovery, which shares most of the challenges and characteristics of broader drug discovery. Nevertheless, oncological drug discovery also entails its own unique traits that reflect the complexity and diversity of neoplastic diseases, as summarized in Box 1.[7,8] Understanding this diversity is a key aspect in the development of personalized anticancer treatments, which are increasingly being deployed in clinical practice.[9,10] The (computational) drug discovery field is gradually moving towards cancer-specific applications and/or demonstrating the applicability of therapies against cancer-related targets.

Here, we review the efforts made to integrate AI/ML and SB methods within computational drug discovery strategies that are specifically applied to, or can potentially impact, the field of cancer research (Table 1). The reviewed articles cover different parts of the oncology drug discovery pipeline, and we focus on six computational use case scenarios and four integration

Abbreviations: AI, Artificial intelligence; DHODH, dihydroorotate dehydrogenase; FEP, Free energy perturbation; MD, Molecular dynamics; ML, Machine learning; QSAR, Quantitative structure–activity relationship; RF, Random forest; SB, Structure-based; TCGA, The Cancer Genome Atlas; VS, Virtual screening.

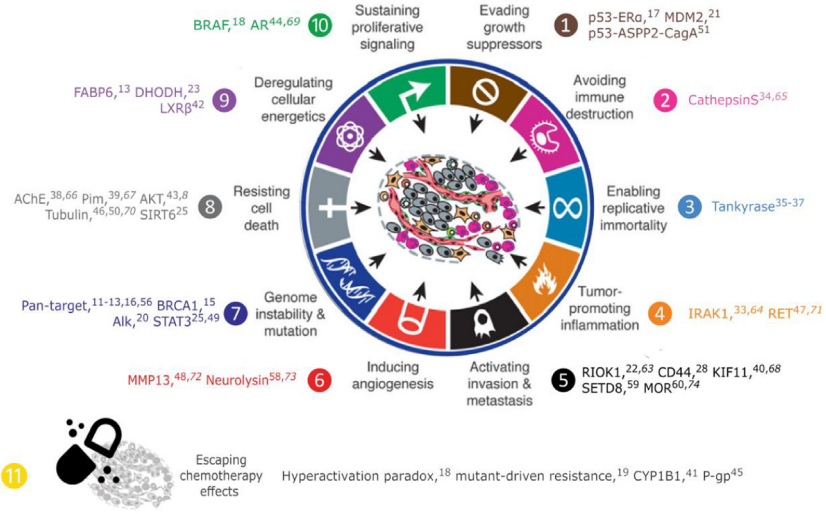
* *Corresponding author.van Westen, G.J.P. (gerard@lacdr.leidenuniv.nl)*

1359-6446/
https://doi.org/10.1016/j.drudis.2022.03.005

Box 1. Targeting the hallmarks of cancer

In their description of the hallmarks of cancer, Hanahan and Weinberg (2000)7 defined six underlying traits that are common to tumorigenesis. In the light of new evidence, these were later complemented by two additional emerging hallmarks and two enabling characteristics.8 These hallmarks paved the way to understanding the complexity and diversity of neoplastic diseases. Understanding this diversity is a key aspect for the development of personalized anti-cancer treatments.

A combination of artificial intelligence (AI) and structure-based methods can be used to address cancer drug discovery research in a more holistic way, tackling all the hallmarks of cancer. In this review, we provide an overview of the biological relevance of the drug discovery targets in cancer and their relevance to the hallmarks and characteristics of cancer (numbered 1 to 10 in the box figure). An eleventh 'hallmark', the ability of cancer cells to escape chemotherapy effects, is added here and is a key aspect to consider in oncology drug discovery strategies.



Supporting references are cited for the target of each hallmark; references cited in italic text provide additional support on the connection of a certain target to a hallmark. Figure adapted from Hanahan and Weinberg.[8]

methods (Fig. 1). In the following sections, we cover each of these use scenarios: computational mutagenesis, (off)-target prediction, binding site prediction, virtual screening (VS), and allosteric modulation analysis. The four ML–SB integration methods that we discuss are: (A) the use of structural data as input for ML models, (B) ML-based scoring functions for SB applications, (C) ML as a tool to analyze MD simulations, and (D) sequential or parallel pipelines in which SB and ML methods are used independently but complementarily. The biological impact of these computational strategies in cancer research is exemplified by the link between the targets addressed in the reviewed publications and each of the ten defined hallmarks of cancer, as well as an additional eleventh 'hallmark' that is of high relevance in oncological drug discovery, namely chemotherapy-escaping capabilities (Box 1). The heterogeneity of the use cases and methods (Table 1) goes hand-in-hand with that of the molecular targets in the reviewed publications, and illustrates the diverse potential of the combined use of AI and SB methods in oncological drug discovery.

Driver prediction

One of the main use case scenarios for, most frequently ML-based, computational cancer research is the prediction of gene and mutation drivers that should be prioritized as targets for anti-cancer therapies. These approaches are pan-target by definition and are usually pan-cancer; they are not focused on specific targets or cancer types. They often start from multi-omics datasets obtained from cancer patients, such as the data on somatic mutations,[11,12] copy number variations,[12] epigenetics,[12] or RNAseq[13] available from The Cancer Genome Atlas (TCGA), and their applicability depends on the availability of such data types. The work of Bailey *et al.*[11] provides an extensive overview of the wide array of tools that are available for driver prediction and, more importantly, of the importance of combining different tools to maximize predictive performance. The approach described by Bailey *et al.*[11] joined SB and ML methods in parallel, but they are more frequently incorporated sequentially.[12,13] Knijnenburg *et al.*[12] and Liñares-Blanco *et al.*[13] created classification models (logistic regression and random forest (RF), respectively) trained on omics data to predict cancer-related outcomes, such as homologous recombination deficiency and tumorigenic phenotype. In both cases, feature importance was used to prioritize genes for further SB analysis. Knijnenburg *et al.*[12] performed *in silico* mutagenesis studies for each detected variant that had potential to affect protein stability. Some of the substitutions were also analyzed with MD, and appeared to alter protein dynamics even if they were not predicted to alter protein stability. Conversely, Liñares-Blanco *et al.*[13] used ML-derived information to perform a drug repurposing VS approach, in which FDA-approved anti-cancer drugs were docked into the available crystal structures of the computationally prioritized proteins, such as FABP6.

It is important to note that the input data, features, and outcome variables that are selected for cancer driver prediction are

TABLE 1

Overview of the reviewed literature categorized by use case scenario.

Reference	Target/ligand dataset	Hallmark of cancer	Artificial intelligence method(s)	Structure-based method(s)	Integration approach
Driver prediction					
Bailey *et al.*[11]	Pan-target/TCGA-MC3 set	7 [11]	Various	Various	Ⓓ
Knijnenburg *et al.*[12]	Pan-target/TCGA-MC3 set	7 [12]	Logistic regression classifier	FoldX, MD	Ⓓ
Liñares-Blanco *et al.*[13]	Pan-target (FABP6)/TCGA	7 9 [13]	RF and generalized linear classifiers	Docking	Ⓓ
Computational mutagenesis					
Masso *et al.*[15]	BRCA1/ClinVar	7 [15]	RF classifier	Structure-derived features	Ⓐ
Pandurangan and Blundell[16]	Pan-target/ProTherm benchmark	7 [16]	ML ensemble classifier	Structure-derived features	Ⓐ
Chitrala *et al.*[17]	P53-ERα/NA	1 [17]	One-layer NN	Protein–protein docking	Ⓐ
Babbitt *et al.*[18]	BRAF/FDA	10 11 [8,18]	Seven stacked classifiers	MD	Ⓒ
Aldeghi *et al.*[19]	Abl/Platinum database, in-house set	11 [19]	Extremely randomized regression trees	Free energy perturbation (FEP)	Ⓓ
Patil *et al.*[20]	Kinome (Alk)/UniProt, literature	7	SVM, RF, NeuralNet, LR	MD (metadynamics)	Ⓓ
(Off)-target prediction					
Pande *et al.*[21]	Pan-target (MDM2)/literature	1 [21]	CoMFA/CoMSIA PLS regressor, DT, RF, KNN, MLP, SVM classifiers	Docking, MD	Ⓐ
Lim *et al.*[22]	Pan-target (RIOK1, PDE3)/ChEMBL, DrugBank, literature datasets, TCGA-CCLE	5 [63]	ElasticNet, SVR regressors	Ligand binding space search in genome, docking	Ⓐ
Zhi *et al.*[23]	DHODH/STRING, KEGG, ChEMBL, ZINC	9 [23]	Multi-GNN	Docking, MD	Ⓓ
Binding site prediction					
Kawaguchi *et al.*[28]	CD44/NA (pre-trained)	5 [28]	Bayesian active learning	Protein–protein docking	Ⓑ
Taherzadeh *et al.*[29]	Pan-target/BioLip	1 (protein–protein binding)	RF classifier, DBSCAN	Structure-derived features	Ⓐ
Virtual screening					
Che *et al.*[33]	IRAK1/ChEMBL, DUD-E	4 [64]	SVM classifier	Docking	Ⓑ
Yang *et al.*[34]	Cathepsin S/PDBbind, CSAR, GC3/4, ChEMBL	2 [65]	XGBoost regressor	Similarity-based docking	Ⓑ
Berishvili *et al.*[35–37]	Pan-target, Tankyrase/ZINC	3 [37]	DNN	Docking, MD, FEP	Ⓑ
Adeshina *et al.*[38]	Pan-target (AChE)/ChEMBL, DUD-E	8 [66]	XGBoost classifier	Docking	Ⓑ
Kalaki and Asadollahi-Baboli[39]	Pim/in-house dataset	8 [67]	PCA, PLS classifier	Docking	Ⓐ
Li *et al.*[40]	KIF11/KEGG BRITE, DrugBank, STITCH	5 [68]	Bayesian Additive Regression Trees	Bow-pharmacological space (protein–ligand interactions)	Ⓐ
Raju *et al.*[41]	CYP1B1/ChEMBL, PubChem, literature, DUD-E, Maybridge, ChemBridge, Natural compound library	11 [41]	SVM, RF, ANN classifiers	Docking, MD	Ⓓ
Chen *et al.*[42]	LXRβ/ChEMBL, Binding DB, in-house library, GSMTL	9 [42]	SVM, Naïve Bayes classifiers	Docking, MD	Ⓓ
Halder and Cordeiro[43]	AKT/ChEMBL, Asinex library	8 [8]	LDA, XGBoost and other classifiers	MD	Ⓓ
Azhagiya Singam *et al.*[44]	AR/Tox21, CompTox	10 [69]	SVM classifiers	Docking	Ⓓ
Kadioglu and Efferth[45]	P-gp/ChEMBL	11 [45]	RF classifier	Docking	Ⓓ
Guo *et al.*[46]	Tubulin/ChEMBL	8 [70]	Naïve Bayes classifiers	Docking, MD	Ⓓ
Burggraaff *et al.*[47]	RET/ChEMBL, ZINC	2 [71]	RF classifiers	(Induced-fit) docking, metadynamics	Ⓓ
Chen *et al.*[48]	MMP13/Traditional Chinese Medicine Database	6 [72]	RF, gradient boosting, AdaBoost, deep learning	MD	Ⓓ

(continued on next page)

TABLE 1 (*CONTINUED*)

Reference	Target/ligand dataset	Hallmark of cancer	Artificial intelligence method(s)	Structure-based method(s)	Integration approach
Chen *et al.*[49]	STAT3/literature set, ZINC	7 [49]	Nine regressors, 3D QSAR	Docking, MD	Ⓓ
Guo *et al.*[50]	Tubulin/ChemDiv	8 [70]	Discovery studio prediction models	Docking, MD	Ⓓ
Junaid *et al.*[51]	P53-ASPP2-CagA/rationally designed	1 [51]	ML module in MOE	MD	Ⓓ
Allosteric modulation analysis					
Lu *et al.*[25]	SIRT6, STAT3/PDB, commercial	8, 7 [25]	SVM	Geometric binding site predictor	Ⓐ
Song *et al.*[56]	Pan-target/PDB	7 [56]	RF, neural networks	Structure-derived features	Ⓐ
Uyar *et al.*[58]	Neurolysin/PDB	6 [73]	ElasticNet, PCA, LDA	MD	Ⓒ
Chen *et al.*[59]	SETD8/cBioPortal	5 [59]	Markov state model, tICA, clustering	MD	Ⓒ
Hu *et al.*[60]	MOR/rationally designed	5 [74]	Markov state model, tICA	MD	Ⓒ

ANN, Artificial neural network; CoMFA, Comparative molecular field analysis; CoMSIA, Comparative molecular similarity indices analysis; DHODH, Dihydroorotate dehydrogenase; DNN, Deep neural network; DT, Decision Tree; GNN, graph neural network; GSMTL, Guangdong Small Molecule Tangible Library; KEGG, Kyoto Encyclopedia of Genes and Genomes; KNN, K-nearest neighbor; LDA, Linear discriminant analysis; LR, Linear regression; MD, Molecular dynamics; MDM2, Mouse double minute 2; ML, Machine learning; MLP, multi-layer perceptron; MOR, μ opioid receptor; NN, Neural network; PCA, Principal components analysis; PDB, Protein Database; PLS, Partial least squares; QSAR, Quantitative structure–activity relationship; RF, Random forest; SVM, Support vector machine; SVR, Support vector regression; TCGA-CCLE, The Cancer Genome Atlas Cancer Cell Line Encyclopedia; tICA, Time-structure based independent component analysis.
[a]See Box 1.
[b]See Fig. 1.

not homogeneous. Key aspects such as the tumor microenvironment or metastasis are often neglected. With regard to cancer patient data, most of the publications use TCGA, which provides high quality and standardized data. However, the TCGA data has been frozen since 2016, highlighting the need for updated cancer patient databases, such as the Genomic Data Commons.[14] Overall, the use of sequential pipelines, sometimes including experimental assays, could help to account for the differential effect of the different types of available data on tumorigenesis.

Computational mutagenesis

Knowing the effect of specific point mutations on protein function and 'drugability' is key to the development of personalized anti-cancer therapies, as well as to decision-making in the clinic. *In vitro* mutagenesis studies are time- and cost-expensive, thus *in silico* computational studies are a good starting point that can be used to prioritize mutants for experimental analysis.

Most of the computational mutagenesis approaches that are reviewed here use structural data to train ML classifiers.[15–18] These structural data might originate directly from a crystal structure,[15,16] in combination with docking studies,[17] or from MD.[18] The approaches developed by Masso *et al.*[15] and by Pandurangan *et al.*[16] extract features from a geometrical representation derived from wildtype and mutant crystal structures and from homology models. Those features were used in classification models to predict variant clinical significance and protein stability, respectively. Protein–protein interaction stability can also be predicted from protein–protein docking-derived features, such as those used by Chitrala *et al.*[17] to predict the p53–ERα interaction stability for wildtype p53 and three breast-cancer-related p53 polymorphisms. Moreover, computational mutagenesis studies are used to predict the effect of mutations on ligand-binding dynamics. Babbitt *et al.*[18] studied the hyperactivating effect of BRAF V600E-targeting inhibitors in wildtype cells using MD. Here, differences in rapid dynamics in bound and unbound functional states for each amino acid were modelled in stacked classification models to detect conserved dynamic function. The models showed that the V600E mutation greatly alters dynamics, leading to lower predictive performance.

The performance of the classification models used for mutagenesis prediction varies highly depending on the amount of experimental mutagenesis data available for training and validation.[15–18] Hence, some authors have compared the performance of SB methods alone to that of ML models for these tasks.[19,20] For example, Aldeghi *et al.*[19] benchmarked the performance of free energy perturbation (FEP), ML, and Monte Carlo methods in predicting the change in inhibitor affinity for Abl kinase variants. The classifier trained on a pan-target dataset was not able to generalize on the test set, but when trained on a reduced Abl-specific dataset, the performance of the classifier was comparable to those of FEP and Monte Carlo methods. Nevertheless, computational time was drastically less when ML was used. Similarly, Patil *et al.*[20] created an MD protocol to determine the activation status of any kinase variant. This information is critical for the prioritization of kinase inhibitors that target the active or the inactive conformation, and hence for the prevention of unwanted side effects. For this reason, we have selected Alk kinase as a case study. Here, the long-term dynamics of the active and the inactive conformations were explored using the metadynamics method. Changes in RMSD (the root-mean-square deviation of atomic positions) and hydrogen bond occupation data were used to calculate a score for the wildtype and the mutant, and a final score was compared to a defined threshold. This approach outperformed a kinome-wide ML model and other common impact prediction tools, such as SIFT and Polyphen.

The computational mutagenesis approaches that are reviewed here are able to capture differences in protein stability and conformation,[16,20] protein–protein interactions,[17] ligand-binding affinity and dynamics,[18,19] and clinical significance.[15] Nevertheless, their applicability is often limited to a particular target or

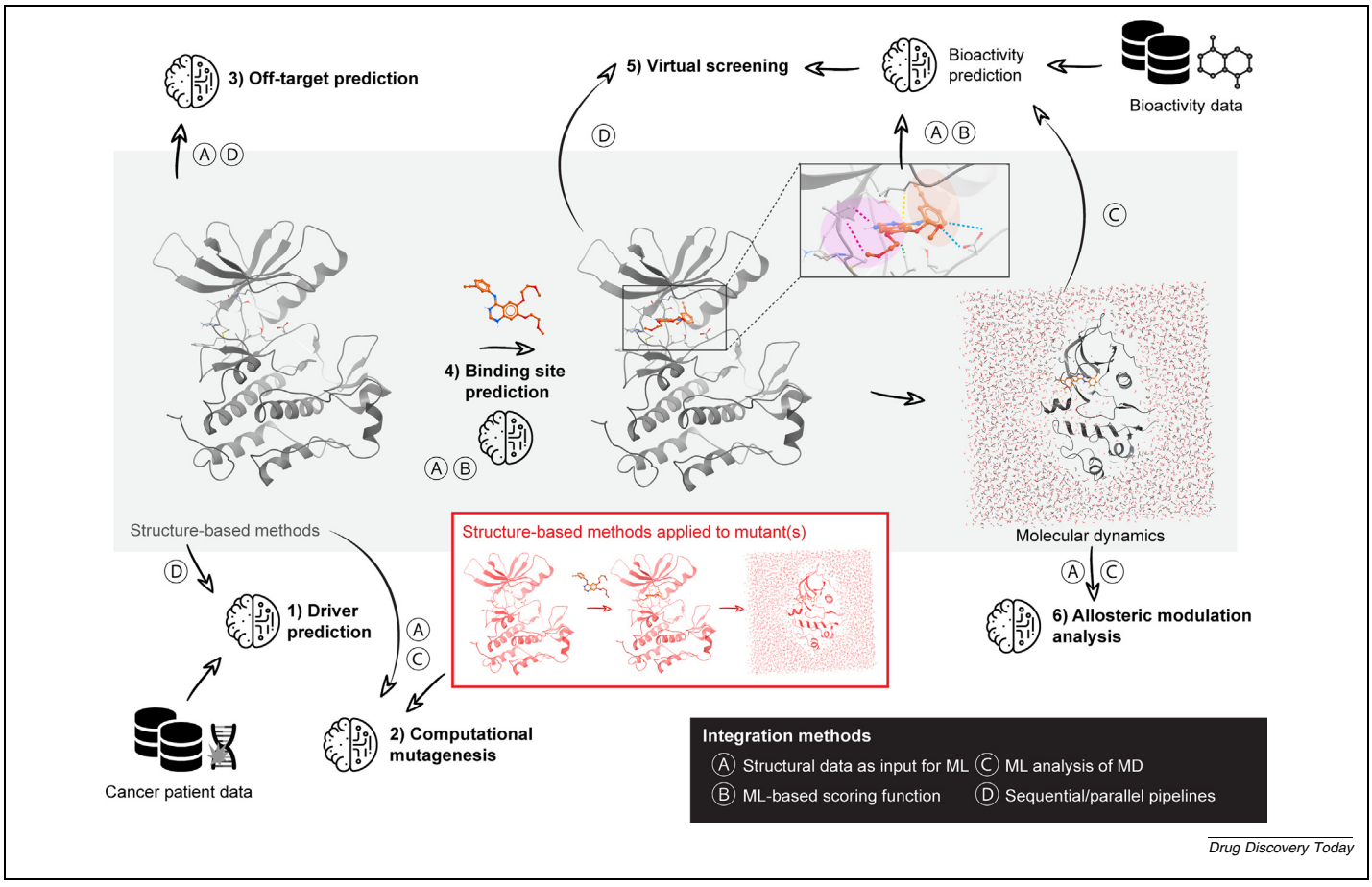


FIGURE 1

Use case scenarios for integrated structure-based (SB) and machine learning (ML) methods in oncological drug discovery and the integration methods employed. We address six use case scenarios: 1) driver prediction, 2) computational mutagenesis, 3) (off)-target prediction, 4) binding site prediction, 5) virtual screening, and 6) allosteric modulation analysis. Integration approaches that achieve full integration include those where: (A) structural data derived from SB methods is used as input for ML models, with emphasis on the predicted output; (B) docking poses are analyzed with ML-based scoring functions; and (C) output trajectories from molecular dynamics (MD) simulations are analyzed with ML. It is still more common, however, to combine SB and ML methods without full integration, with the implementation done in a sequential or parallel way (D), in which ML acts as a pre-filter for the SB phase, or vice versa.

mutant of interest for which there are enough data. In order to increase the impact of methods developed for members of families that have highly conserved binding pockets and activation mechanisms, such as kinases (Babbitt *et al.*,[18] Aldeghi *et al.*,[19] Patil *et al.*[20]) or G protein-coupled receptors, the training sets could be enriched with data from other members of the family. Therefore, computational mutagenesis efforts in general could benefit from more extensive experimentally validated mutagenesis datasets. These should be deposited in publicly available databases, following FAIR principles to favor the creation of relevant training and validation datasets.

(Off)-target prediction

Defining the (off)-target space of drugs in development is important for obtaining a selectivity profile, but also for the rational design of polypharmacological candidates that have a multi-target profile. Moreover, re-analyzing the target space of approved drugs is key to obtaining a better understanding of their mode of action, or to starting re-purposing efforts. These endpoints are highly relevant in oncological drug discovery, where off-target effects are often responsible for grave adverse reactions. Integrated ML-SB methods have proven useful in predicting (off)-target spaces.

Efforts to predict the target space usually start from known information, such as ligand–protein[21,22] or protein–protein interactions.[23] Pande *et al.*[21] set up an SB-ML integrated pipeline to identify the most likely target of the natural compound resveratrol, for which the mode of action is still unknown. This study was possible because of the (recent) resolution of nine proteins in complex with resveratrol. A set of 40 anti-breast cancer resveratrol derivatives from the literature were used for docking, and a 3D quantitative activity–structure relationship (QSAR) CoMFA/CoMSIA PLS model was created for target-derived results from docking studies. Based on the performance of the models, MDM2 and QR2 were suggested as potential targets for resveratrol derivatives.

Computational methods can also be used to rationally propose polypharmacological approaches for the discovery of novel drugs[23] or the repurposing of existing drugs.[22] The implementation by Lim *et al.*[22] used the original crystal structure of an

approved drug as the template to search the genome for genes encoding proteins with an appropriate ligand-binding space. Subsequently, docking was performed and the results used, together with bioactivity data, as input for an ML algorithm that predicted genome-wide ligand–protein interactions in a fully integrated fashion. RIOK1, among other kinases, was predicted to be an off-target of PDE3 inhibitors such as levosimendan, and consequently, this kinase was proposed as a target for drug repurposing for anti-cancer therapies. Conversely, Zhi *et al.*[23] used a sequential SB-ML pipeline to identify novel targets related to dihydroorotate dehydrogenase (DHODH) and to screen drug candidates for their potential to inhibit multiple targets in small cell lung cancer. First, protein–protein interaction information was leveraged for network pharmacology analysis. This allowed the identification of related proteins that could also be affected by drugs that inhibit DHODH, such as uridine monophosphate synthase (UMPS), which like DHODH is involved in pyrimidine biosynthesis. Docking data for both DHODH and UMPS identified eight potential multi-target compounds, which were prioritized on the basis of their predicted binding affinity towards DHODH using three multi-GNN (Graph Neural Network) regression models. The top three candidates were subjected to MD validation, which confirmed that they show stable interactions with both targets.

Integrated approaches for the prediction of (off)-targets can have a direct impact in lead prioritization in oncological drug discovery. The application of the methodologies, however, mostly depends on the available data. Approaches such as those of Pande *et al.*[21] and Lim *et al.*[22] are relevant when true binding modes have been identified. In the case of Zhi *et al.*,[23] both rich interactome databases and bioactivity data for the identified targets of interest were needed.

Prediction of binding site

Once the relevant targets have been defined, the binding sites need to be characterized for drug discovery purposes. In oncological drug discovery, this task can be made more complicated by mutated binding sites or transformed protein–protein interactions. An extensive array of tools is available for the prediction of small molecule binding sites, as recently reviewed by Krivák and Hoksza.[24] In their independent benchmark, they showed that some methods in which SB and ML techniques were integrated performed as well or better than SB-exclusive methods. Nevertheless, they urged caution over the prediction of complex features from ML analysis of structural data when using relatively small training datasets. Of particular interest in anti-cancer drug development is the discovery of allosteric binding sites that can be targeted selectively in cancer cells, thereby reducing off-site adverse effects triggered by events at orthosteric binding sites. Most of the binding site prediction methods summarized by Krivák and Hoksza[24] can be used to predict allosteric binding sites, but these sites share a number of differential characteristics that have triggered the development of tools that specifically predict allosteric binding sites.[25] Some of these methods are built on top of general binding site predictors with, for example, an added layer of ML classification.[26] The application of these methods and the analysis of the effects caused by allosteric modulators are discussed in more detail below under the heading 'Allosteric modulation analysis'.

While the information and software needed for binding site prediction is extensively available for small molecules, the prediction of binding regions in protein–protein binding modelling is still challenging.[27] Protein–protein interactions are crucial in certain aspects of cancer pathogenicity,[8] and integrated SB-ML approaches have proven beneficial in this area.[28,29] Kawaguchi *et al.*[28] used a Bayesian active learning-based protein–protein docking approach to predict the conformation of the dimerization interface of CD44 and the residues involved. Similarly, the approach developed by Taherzadeh *et al.*[29] uses ML to predict protein–peptide binding residues from protein sequence and structural-data-derived features. The predicted residues from the RF classifier are used as input for a density-based clustering algorithm that defines the binding region on the protein surface. Taherzadeh *et al.*[29] showed that the performance of this approach is better than that of non-ML methods applied to the same dataset. In general, however, the exploratory nature of the applications in this use case scenario makes it challenging to assess the performance of the methods reviewed. To counterbalance this problem and to reduce the effect of false positives, one option would be to use a consensus approach in which several tools are employed and sites that are predicted by more than one of them are investigated further.

The feasibility of the approaches reviewed here largely depends on the availability of structural data. The use of homology models can be useful here, with some authors showing that their integrated ML-SB methods perform equally well in experimental structures as in homology models.[29,30] Moreover, the recent release of AlphaFold[31] to predict protein structures with high accuracy opens doors for the implementation of many of these methods on a genome-wide scale. The distribution of AlphaFold as open source code has facilitated the development of related tools that will improve the biological relevance of this tool. An example is AlphaFill,[32] a tool that enriches AlphaFold models with ligands and co-factors. Of very high relevance in oncological drug discovery, these tools could enable the prediction of binding sites in mutants that have not being experimentally determined.

Virtual screening

The most common scenario in computational drug discovery is virtual screening (VS). Like computational mutagenesis, VS can be seen as a tool for prioritizing compounds for experimental analysis. VS has been extensively explored using SB and ML methods independently, but their combination—either in a fully integrated or in a sequential way—allows for the use of as much available data as possible and, expectedly, provides more accurate results. This use case scenario is certainly not unique to oncological drug discovery, but the advances made in computational drug discovery in this area could very well power successful anti-cancer drug discovery stories.

A classic way to integrate SB and ML learning methods in VS is the use of ML-based scoring functions in docking.[33–38] These can be directly integrated in the docking software or, more commonly, used *a posteriori* for re-scoring. Moreover, ML scoring functions are often target-specific,[33–35] but not necessarily so.[38] One of the simplest approaches is to include docking scores as features for an ML classifier.[33] The slightly more complex approach developed by Yang *et al.*[34] begins with a similarity-based docking method, which was able to reduce the challenges presented by the large conformational space of Cathepsin S inhibitors. Subsequently, a fragmentation method was applied to the predicted poses.[34] In addition, Berishvili *et al.* demonstrated the added value of including not only docking-derived features for the ML scoring function,[35] but also MD-derived features.[36] In retrospect, however, they showed that ML-based target-specific scoring functions were not accurate in identifying active tankyrase compounds. More complex methods, such as FEP, were needed in order to properly correlate the predicted binding affinity to the pIC_{50} values determined experimentally. As for other ML applications, the development of accurate ML scoring functions is highly dependent on the quality of the datasets available for training and validation. Adeshina *et al.*[38] focused on the development of a high-quality dataset (D-COID, publicly available) for training ML re-scoring functions. Importantly, they included challenging decoy complexes from the DUD-E dataset and tried to keep the dataset balanced. They also refrained from using docked poses in the training set.

Similar approaches might not necessarily be coined ML scoring functions, even though they also use ligand–protein interaction data as input for ML models.[39,40] Kalaki and Asadollahi-Baboli[39] used an approach in which docking was performed as a first step to discern relevant interactions and to derive ML descriptors. Using a slightly different approach, Li *et al.*[40] constructed a pharmacological space that accounted for ligand, protein, and ligand–protein interaction descriptors. The last were generated by combining the average fingerprint per protein data from known binders.

In general, however, the most typical approach in VS is still the use of SB and ML methods in a sequential or parallel way.[41–50] These methods often include the development of a ligand-based QSAR classification[41–47] or regression[48,49] model from experimental bioactivity data in order to prioritize compounds from a chemical database on the basis of their predicted binding affinity. The wide array of models and databases reviewed here is collated in Table 1. Subsequently, the selected hits are filtered on the basis of different criteria, which depend on the scope of the project (such as reverse pharmacophore mapping,[43] or ΔG calculation with MM-GBSA[44]). Finally an SB method such as docking[41,42,44–46,49,50] and/or MD[41–43,46,48–51] is deployed to rationalize the results of the ML model and to propose compounds for *in vitro* validation. Sometimes, the SB phase is a filter on its own, with a docking-based VS,[41,46] and occasionally it is used before the ML phase.[44,49] Moreover, the ML model is not always built to predict only binding affinity, but sometimes also anti-cancer activity[50] or mode of action.[45] When focused on multiple on- and off-targets, sequential pipelines can also be used to prioritize polypharmacological compounds, as was the case in the work on kinase inhibitors by Burggraaff *et al.*[47] Even though VS strategies are more common in the screening of small molecules, there are also examples from peptide VS campaigns, such as that of Junaid *et al.*[51].

One of the main limitations of VS approaches lies in the definition of relevant training and validation sets for ML. Even though databases such as ChEMBL and PubChem contain a very large amount of bioactivity data, target-specific applications usually still end up with datasets that are too small to support generalizable predictions. This is an even more relevant bottleneck when considering cancer-related mutants, for which VS campaigns would be extremely beneficial in the prioritization of personalized medicine drugs. Moreover, target-specific applications present an important challenge in avoiding learnt biases and overfitting.[52] The inclusion of decoys in the datasets (such as data from the DUD-E dataset) is a good way to balance the presence of active and inactive compounds.[53] Consequently, although the D-COID dataset[38] is a good starting point for the development of re-scoring functions, it might require experimental expansion via collaborative work to improve its suitability for target-specific applications.

Allosteric modulation analysis

So far, we have mostly referred to orthosteric ligand binding (i.e. binding at the site where the endogenous ligand or substrate binds) when describing ligand binding. However, allosteric modulation has been described as a powerful tool to increase the selectivity of targeted compounds and to overcome drug-resistant mutations, and it is therefore worth exploring in cancer research. Indeed, unraveling the mechanisms that underlie allosteric effects can be a key step in proposing new therapeutic routes. Moreover, allosteric binding sites and modulators have been shown to exhibit characteristics that are different to those of their orthosteric counterparts.[54] This calls for the development of allosteric-specific tools for most of the use case scenarios described in the sections above.

The work by Lu *et al.*[25] provides a very complete review of the currently available SB methods for allosteric modulator discovery. Some of these methods integrate SB and ML techniques for allosteric binding site prediction,[26] allosteric interaction scoring,[55] and allosteric effect analysis of mutations.[56] Lu *et al.*[25] demonstrate the applicability of these tools in oncological drug discovery by describing the prioritization of allosteric activators and inhibitors for anti-cancer (potential) targets SIRT6 and STAT3, respectively. In both cases, allosteric binding pockets were predicted and used for VS of commercial libraries. The compounds that were identified by these computational efforts were confirmed by either experimental assays or crystallographic studies. Of direct application in oncological drug discovery is AlloDriver,[56] a driver prediction tool that maps mutations from clinical cancer samples to their 3D structures, labels them as orthosteric or (potentially) allosteric, and classifies targets as driver or passenger using a combination of RF and multi-layer neural networks. Although it is periodically updated, this tool relies

on the availability of annotated allosteric sites (and driver mutations), which is a common bottleneck in ML-based allostery prediction methods.

Specific to allosteric modulation analyses is the exploration of the allosteric pathways that drive the observed effects. Given the complex conformational landscape of proteins that is often responsible for allosteric pathways, these aspects are often better explored in a dynamic setting.[25,57] Hence, the efforts reviewed below use ML techniques to analyze MD trajectories and to find patterns that help to explain the observed effects.[58–60] For example, the work of Uyar *et al.*[58] was able to identify differential dynamic patterns in apo and allosteric inhibitor-bound neurolysin structures, as well as the key residues involved. Moreover, the analysis of MD trajectories with Markov state models using time-structure based independent component analysis (tICA) allowed Chen *et al.*[59] and Hu *et al.*[60] to identify conformational microstates. These microstates were then related to mutation-driven allosteric effects on the catalytic activity of SEDT8, and to energetic differences in Na^+ translocation and metastable states in active and inactive MOR, respectively. These effects were further validated experimentally.

Even though the concept of allostery has been known for 50 years, it has recently gained more attention in drug discovery thanks to an exponential increase in the number of known allosteric modulators over the past two decades.[25] Of the 19 allosteric modulators currently approved by the FDA, three are indicated as anti-cancer drugs.[61] The use of computational tools, and more specifically ML-based methods, still suffers from the lack of experimentally determined allosteric interactions and mechanisms. In the near future, we expect this area of research, in combination with experimental validation, to play a more important role in oncological drug discovery because promises to bring more selective anticancer drugs to the market.

Conclusions

Integrated ML-SB methods are useful for investigating different aspects in oncological drug discovery. These computational drug discovery methods apply to a variety of use case scenarios, which can be cancer-specific or more general but with potential application in oncological research. There is no rule of thumb for selecting an approach because the most appropriate methods depend largely on the scope of the study. However, some ML-SB integration methods are primarily leveraged in specific use case scenarios, such as ML-based scoring functions in VS or the use of ML to analyze MD simulations in allosteric modulation analyses. VS is still the most commonly used strategy, but integrated methods are also gaining relevance in fields such as driver prediction and computational mutagenesis, where the use of structural data has proven to be a significant complement to omics data. Despite their broad domain of applicability, the approaches reviewed here still present certain limitations that are worth discussing. Data availability and computational requirements are common bottlenecks that need to be assessed on a project-specific basis.

Box 2. Open questions on present and future directions

The articles reviewed here exemplify the added value of integrated AI-SB methods in oncological drug discovery. However, some questions that are worth exploring in the future arise from their interpretation:

- Structural data availability is a common bottleneck. How beneficial is its inclusion in pan-target analyses when it results in a reduced target space? Will approaches such as AlphaFold31 be able to solve this issue?
- At present, the analysis of trajectories from MD with ML is rather restricted to cases with small datasets (i.e. allosteric modulation analyses). However, we expect that with increasing amounts of data and computing power, this approach will become more relevant in big-scale virtual screening.
- Is it pertinent to continue expanding the research into integrated approaches without conducting exhaustive benchmarking against classical individual methods?
- Are there enough resources devoted to enlarging and standardizing publicly available datasets for computational oncological drug discovery? Will these expand into aspects that are often neglected, such as the tumor microenvironment?
- We hypothesize the rise of allosteric modulation analyses to bring more selective drugs to the market. Will we also see a boom in publicly available allosteric structural and experimental data for ML applications?
- Is the potential added value of more complex approaches worth the probable resulting increase in computing power/time and data storage needs? Will this aspect limit the use of deep learning approaches in the near future?

A common drawback in computational drug discovery is the lack of experimental validation. We strongly advise an increase in collaborative work leading to both validated tools and larger datasets for ML training.

Moreover, it has been shown that less expensive approaches sometimes outperform more complex ones in the same tasks. Future research will probably extend towards the use of more complex algorithms that are currently underrepresented, such as deep neural networks (DNNs), in order to capture all of the relevant information from structural data. Finally, a common drawback in computational drug discovery, which can be observed in the articles reviewed here, is the lack of experimental validation. These aspects trigger some open questions on the use of integrated computational methods in oncological drug research, which we address in Box 2. Nevertheless, the approaches presented here provide a good way to prioritize targets and small molecules in the field, and their combination with experimental validation is likely to be a key factor in bringing drugs for personalized oncological therapies to the market more rapidly. During the revision of our manuscript, a proposal for a further concep-

tual extension of the hallmarks of cancer was published.[62] This exemplifies the fast pace at which oncological research advances and the need to revisit constantly the biological relevance of the methods applied in oncological drug discovery.

Acknowledgements

The authors thank Anastassis Perrakis for critically reading the manuscript. MGG, APAJ, and LHH thank the Oncode Institute for funding and support.

References

1. X. Yang, Y. Wang, R. Byrne, G. Schneider, S. Yang, Concepts of artificial intelligence for computer-assisted drug discovery, Chem Rev 119 (2019) 10520–10594, https://doi.org/10.1021/acs.chemrev.8b00728.
2. F. Azuaje, Artificial intelligence for precision oncology: beyond patient stratification, NPJ Precis Oncol 3 (2019) 6, https://doi.org/10.1038/s41698-019-0078-1.
3. M. Duran-Frigola, R. Mosca, P. Aloy, Structural systems pharmacology: the role of 3D structures in next-generation drug development, Chem Biol 20 (2013) 674–684, https://doi.org/10.1016/j.chembiol.2013.03.004.
4. H. Li, K.H. Sze, G. Lu, P.J. Ballester, Machine-learning scoring functions for structure-based drug lead optimization, Wiley Interdiscip Rev Comput Mol Sci 10 (2020), https://doi.org/10.1002/wcms.1465 e1465.
5. M. Batool, B. Ahmad, S. Choi, A structure-based drug discovery paradigm, Int J Mol Sci 20 (2019) 2783, https://doi.org/10.3390/ijms20112783.
6. D. Sydow, L. Burggraaff, A. Szengel, et al., Advances and challenges in computational target prediction, J Chem Inf Model 59 (2019) 1728–1742, https://doi.org/10.1021/acs.jcim.8b00832.
7. D. Hanahan, R.A. Weinberg, The hallmarks of cancer, Cell 100 (2000) 57–70, https://doi.org/10.1016/S0092-8674(00)81683-9.
8. D. Hanahan, R.A. Weinberg, Hallmarks of cancer: the next generation, Cell 144 (2011) 646–674, https://doi.org/10.1016/j.cell.2011.02.013.
9. P. Krzyszczyk, A. Acevedo, E.J. Davidoff, L.M. Timmins, I. Marrero-Berrios, M. Misaal Patel, et al., The growing role of precision and personalized medicine for cancer treatment, Technology 6 (2018) 79–100, https://doi.org/10.1142/S2339547818300020.
10. F. Wu, Y. Zhou, L. Li, X. Shen, G. Chen, X. Wang, et al., Computational approaches in preclinical studies on drug discovery and development, Front Chem 8 (2020) 726, https://doi.org/10.3389/fchem.2020.00726.
11. Bailey MH, Tokheim C, Porta-Pardo E, Sengupta S, Bertrand D, Weerasinghe A, *et al*. Comprehensive characterization of cancer driver genes and mutations. *Cell* 2018;173:371–85.e18. doi:10.1016/j.cell.2018.02.060.
12. Knijnenburg TA, Wang L, Zimmermann MT, Nyasha Chambwe 1, Galen F Gao 4, Andrew D Cherniack AD, *et al*. Genomic and molecular landscape of DNA damage repair deficiency across The Cancer Genome Atlas. *Cell Rep* 2018;23:239–54.e6. doi:10.1016/j.celrep.2018.03.076.
13. J. Liñares-Blanco, C.R. Munteanu, A. Pazos, C. Fernandez-Lozano, Molecular docking and machine learning analysis of Abemaciclib in colon cancer, BMC Mol Cell Biol 21 (2020) 52, https://doi.org/10.1186/s12860-020-00295-w.
14. M.A. Jensen, V. Ferretti, R.L. Grossman, L.M. Staudt, The NCI Genomic Data Commons as an engine for precision medicine, Blood 130 (2017) 453–459, https://doi.org/10.1182/blood-2017-03-735654.
15. M. Masso, A. Bansal, A. Bansal, A. Henderson, Structure-based functional analysis of BRCA1 RING domain variants: concordance of computational mutagenesis, experimental assay, and clinical data, Biophys Chem 266 (2020), https://doi.org/10.1016/j.bpc.2020.106442 106442.
16. A.P. Pandurangan, T.L. Blundell, Prediction of impacts of mutations on protein structure and interactions: SDM, a statistical approach, and mCSM, using machine learning, Protein Sci 29 (2020) 247–257, https://doi.org/10.1002/pro.3774.
17. K.N. Chitrala, M. Nagarkatti, P. Nagarkatti, S. Yeguvapalli, Analysis of the TP53 deleterious single nucleotide polymorphisms impact on estrogen receptor alpha-p53 interaction: a machine learning approach, Int J Mol Sci 20 (2019) 2962, https://doi.org/10.3390/ijms20122962.
18. G.A. Babbitt, M.L. Lynch, M. McCoy, E.P. Fokoue, A.O. Hudson, Function and evolution of B-Raf loop dynamics relevant to cancer recurrence under drug inhibition, J Biomol Struct Dyn 40 (2022) 468–483, https://doi.org/10.1080/07391102.2020.1815578.
19. M. Aldeghi, V. Gapsys, B.L. De Groot, Predicting kinase inhibitor resistance: physics-based and data-driven approaches, ACS Cent Sci 5 (2019) 1468–1474, https://doi.org/10.1021/acscentsci.9b00590.
20. K. Patil, E.J. Jordan, J.H. Park, K. Suresh, C.M. Smith, A.A. Lemmon, et al., Computational studies of anaplastic lymphoma kinase mutations reveal common mechanisms of oncogenic activation e2019132118, Proc Natl Acad Sci USA 118 (2021), https://doi.org/10.1073/pnas.2019132118.
21. A. Pande, M. Manchanda, H.R. Bhat, P.S. Bairy, N. Kumar, P. Gahtori, Molecular insights into a mechanism of resveratrol action using hybrid computational docking/CoMFA and machine learning approach, J Biomol Struct Dyn (2021) 1–15, https://doi.org/10.1080/07391102.2021.1910572.
22. H. Lim, D. He, Y. Qiu, P. Krawczuk, X. Sun, L. Xie, Rational discovery of dual-indication multi-target pde/kinase inhibitor for precision anti-cancer therapy using structural systems pharmacology, PLoS Comput Biol 15 (2019) e1006619, https://doi.org/10.1371/journal.pcbi.1006619.
23. H.Y. Zhi, L. Zhao, C.C. Lee, C.Y.C. Chen, A novel graph neural network methodology to investigate dihydroorotate dehydrogenase inhibitors in small cell lung cancer, Biomolecules 11 (2021) 477, https://doi.org/10.3390/biom11030477.
24. R. Krivák, D. Hoksza, P2Rank: machine learning based tool for rapid and accurate prediction of ligand binding sites from protein structure, J Cheminform 10 (2018) 39, https://doi.org/10.1186/s13321-018-0285-8.
25. S. Lu, X. He, D. Ni, J. Zhang, Allosteric modulator discovery: from serendipity to structure-based design, J Med Chem 62 (2019) 6405–6421, https://doi.org/10.1021/acs.jmedchem.8b01749.
26. W. Huang, S. Lu, Z. Huang, X. Liu, L. Mou, Y. Luo, et al., Allosite: a method for predicting allosteric sites, Bioinformatics 29 (2013) 2357–2359, https://doi.org/10.1093/bioinformatics/btt399.
27. I.A. Vakser, Challenges in protein docking, Curr Opin Struct Biol 64 (2020) 160–165, https://doi.org/10.1016/j.sbi.2020.07.001.
28. M. Kawaguchi, N. Dashzeveg, Y. Cao, Y. Jia, X. Liu, Y. Shen, H. Liu, Extracellular Domains I and II of cell-surface glycoprotein CD44 mediate its trans-homophilic dimerization and tumor cluster aggregation, J Biol Chem 295 (2020) 2640–2649, https://doi.org/10.1074/jbc.RA119.010252.
29. G. Taherzadeh, Y. Zhou, A.W.C. Liew, Y. Yang, Structure-based prediction of protein-peptide binding regions using Random Forest, Bioinformatics 34 (2018) 477–484, https://doi.org/10.1093/bioinformatics/btx614.
30. L. Li, M. Khanna, I. Jo, F. Wang, N.M. Ashpole, A. Hudmon, S.O. Meroueh, Target-specific support vector machine scoring in structure-based virtual screening: computational validation, in vitro testing in kinases, and effects on lung cancer cell proliferation, J Chem Inf Model 51 (2011) 755–759, https://doi.org/10.1021/ci100490w.
31. J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, et al., Highly accurate protein structure prediction with AlphaFold, Nature 596 (2021) 583–589, https://doi.org/10.1038/s41586-021-03819-2.
32. Hekkelman ML, Vries I de, Joosten RP, Perrakis A. AlphaFill: enriching the AlphaFold models with ligands and co-factors. *bioRxiv* 2021:2021.11.26.470110. https://doi.org/10.1101/2021.11.26.470110.
33. J. Che, R. Feng, J. Gao, H. Yu, Q. Weng, Q. He, et al., Evaluation of artificial intelligence in participating structure-based virtual screening for identifying novel Interleukin-1 receptor associated kinase-1 inhibitors, Front Oncol 10 (2020) 1769, https://doi.org/10.3389/fonc.2020.01769.
34. Y. Yang, J. Lu, C. Yang, Y. Zhang, Exploring fragment-based target-specific ranking protocol with machine learning on cathepsin S, J Comput Aided Mol Des 33 (2019) 1095–1105, https://doi.org/10.1007/s10822-019-00247-3.
35. V.P. Berishvili, A.E. Voronkov, E.V. Radchenko, V.A. Palyulin, Machine learning classification models to improve the docking-based screening: a case of PI3K-tankyrase inhibitors, Mol Inform 37 (2018) e1800030, https://doi.org/10.1002/minf.201800030.
36. V.P. Berishvili, V.O. Perkin, A.E. Voronkov, E.V. Radchenko, R. Syed, C. Venkata Ramana Reddy, et al., Time-domain analysis of molecular dynamics trajectories using deep neural networks: application to activity ranking of tankyrase inhibitors, J Chem Inf Model 59 (2019) 3519–3532, https://doi.org/10.1021/acs.jcim.9b00135.
37. V.P. Berishvili, A.N. Kuimo, A.E. Voronkov, E.V. Radchenko, P. Kumar, Y.E. Choonara, et al., Discovery of novel tankyrase inhibitors through molecular docking-based virtual screening and molecular dynamics simulation studies, Molecules 25 (2020) 3171, https://doi.org/10.3390/molecules25143171.
38. Y. Adeshina, E. Deeds, J. Karanicolas, Machine learning classification can reduce false positives in structure-based virtual screening, Proc Natl Acad Sci USA 117 (2020) 18477–18488, https://doi.org/10.1073/pnas.2000585117.

39. Z. Kalaki, M. Asadollahi-Baboli, Molecular docking-based classification and systematic QSAR analysis of indoles as Pim kinase inhibitors, SAR QSAR Environ Res 31 (2020) 399–419, https://doi.org/10.1080/1062936X.2020.1751277.
40. L. Li, C.C. Koh, D. Reker, J.B. Brown, H. Wang, N.K. Lee, et al., Predicting protein-ligand interactions based on bow-pharmacological space and Bayesian additive regression trees, Sci Rep 9 (2019) 7703, https://doi.org/10.1038/s41598-019-43125-6.
41. Raju B, Verma H, Narendra G, Sapra B, Silakari O. Multiple machine learning, molecular docking, and ADMET screening approach for identification of selective inhibitors of CYP1B1. *J Biomol Struct Dyn* 2021;Mar 26:1–16. doi:10.1080/07391102.2021.1905552.
42. H. Chen, Z. Chen, Z. Zhang, Y. Li, S. Zhang, F. Jiang, et al., Discovery of new LXRβ agonists as glioblastoma inhibitors, Eur J Med Chem 194 (2020), https://doi.org/10.1016/j.ejmech.2020.112240 112240.
43. A.K. Halder, M.N.D.S. Cordeiro, Akt inhibitors: the road ahead to computational modeling-guided discovery, Int J Mol Sci 22 (2021) 3944, https://doi.org/10.3390/ijms22083944.
44. E.R. Azhagiya Singam, P. Tachachartvanich, D. Fourches, A. Soshilov, J.C.Y. Hsieh, M.A. La Merrill, et al., Structure-based virtual screening of perfluoroalkyl and polyfluoroalkyl substances (PFASs) as endocrine disruptors of androgen receptor activity using molecular docking and machine learning, Environ Res 190 (2020), https://doi.org/10.1016/j.envres.2020.109920 109920.
45. O. Kadioglu, T. Efferth, A machine learning-based prediction platform for P-glycoprotein modulators and its validation by molecular docking, Cells 8 (2019) 1286, https://doi.org/10.3390/cells8101286.
46. Q. Guo, H. Zhang, Y. Deng, S. Zhai, Z. Jiang, D. Zhu, L. Wang, Ligand- and structural-based discovery of potential small molecules that target the colchicine site of tubulin for cancer treatment, Eur J Med Chem 196 (2020), https://doi.org/10.1016/j.ejmech.2020.112328 112328.
47. L. Burggraaff, E.B. Lenselink, W. Jespers, J. van Engelen, B.J. Bongers, M.G. González, et al., Successive statistical and structure-based modeling to identify chemically novel kinase inhibitors, J Chem Inf Model 60 (2020) 4283–4295, https://doi.org/10.1021/acs.jcim.9b01204.
48. J.Q. Chen, H.Y. Chen, W.J. Dai, Q.J. Lv, C.Y.C. Chen, Artificial intelligence approach to find lead compounds for treating tumors, J Phys Chem Lett 10 (2019) 4382–4400, https://doi.org/10.1021/acs.jpclett.9b01426.
49. X. Chen, H.Y. Chen, Z.D. Chen, J.N. Gong, C.Y.C. Chen, A novel artificial intelligence protocol for finding potential inhibitors of acute myeloid leukemia, J Mater Chem B 8 (2020) 2063–2081, https://doi.org/10.1039/d0tb00061b.
50. Q. Guo, Y. Luo, S. Zhai, Z. Jiang, C. Zhao, J. Xu, L. Wang, Discovery, biological evaluation, structure-activity relationships and mechanism of action of pyrazolo [3,4-b] pyridin-6-one derivatives as a new class of anticancer agents, Org Biomol Chem 17 (2019) 6201–6214, https://doi.org/10.1039/c9ob00616h.
51. M. Junaid, M. Shah, A. Khan, C. Li, M.T. Khan, A.C. Kaushik, et al., Structural-dynamic insights into the *H. pylori* cytotoxin-associated gene A (CagA) and its abrogation to interact with the tumor suppressor protein ASPP2 using decoy peptides, J Biomol Struct Dyn 37 (2019) 4035–4050, https://doi.org/10.1080/07391102.2018.1537895.
52. J. Sieg, F. Flachsenberg, M. Rarey, In need of bias control: evaluating chemical data for machine learning in structure-based virtual screening, J Chem Inf Model 59 (2019) 947–961, https://doi.org/10.1021/acs.jcim.8b00712.
53. B.K. Allen, S. Mehta, S.W.J. Ember, E. Schonbrunn, N. Ayad, S.C. Schürer, Large-scale computational screening identifies first in class multitarget inhibitor of EGFR kinase and BRD4, Sci Rep 5 (2015) 16924, https://doi.org/10.1038/srep16924.
54. G.J.P. van Westen, A. Gaulton, J.P. Overington, Chemical, target, and bioactive properties of allosteric modulation, PLoS Comput Biol 10 (2014), https://doi.org/10.1371/journal.pcbi.1003559 e1003559.
55. S. Li, Q. Shen, M. Su, X. Liu, S. Lu, Z. Chen, et al., Alloscore: a method for predicting allosteric ligand-protein interactions, Bioinformatics 32 (2016) 1574–1576, https://doi.org/10.1093/bioinformatics/btw036.
56. K. Song, Q. Li, W. Gao, S. Lu, Q. Shen, X. Liu, et al., AlloDriver: a method for the identification and analysis of cancer driver targets, Nucleic Acids Res 47 (2019) W315–W321, https://doi.org/10.1093/nar/gkz350.
57. R. Nussinov, C.J. Tsai, H. Jang, Dynamic protein allosteric regulation and disease, Adv Exp Med Biol 1163 (2019) 25–43, https://doi.org/10.1007/978-981-13-8719-7_2.
58. A. Uyar, V.T. Karamyan, A. Dickson, Long-range changes in neurolysin dynamics upon inhibitor binding, J Chem Theory Comput 14 (2018) 444–452, https://doi.org/10.1021/acs.jctc.7b00944.
59. S. Chen, R.P. Wiewiora, F. Meng, N. Babault, A. Ma, W. Yu, et al., The dynamic conformational landscape of the protein methyltransferase setd8, Elife 8 (2019) e45403, https://doi.org/10.7554/eLife.45403.
60. X. Hu, Y. Wang, A. Hunkele, D. Provasi, G.W. Pasternak, M. Filizola, Kinetic and thermodynamic insights into sodium ion translocation through the μ-opioid receptor from molecular dynamics and machine learning analysis, PLoS Comput Biol 15 (2019) e1006689, https://doi.org/10.1371/journal.pcbi.1006689.
61. O.S. Amamuddy, W. Veldman, C. Manyumwa, A. Khairallah, S. Agajanian, O. Oluyemi, et al., Integrated computational approaches and tools for allosteric drug discovery, Int J Mol Sci 21 (2020) 847, https://doi.org/10.3390/ijms21030847.
62. D. Hanahan, Hallmarks of cancer: new dimensions, Cancer Discov 12 (2022) 31–46, https://doi.org/10.1158/2159-8290.CD-21-1059.
63. F. Weinberg, N. Reischmann, L. Fauth, S. Taromi, J. Mastroianni, M. Köhler, et al., The atypical kinase RIOK1 promotes tumor growth and invasive behavior, EBioMedicine 20 (2017) 79–97, https://doi.org/10.1016/j.ebiom.2017.04.015.
64. Z.N. Wee, S.M.J.M. Yatim, V.K. Kohlbauer, M. Feng, J.Y. Goh, Y. Bao, et al., IRAK1 is a therapeutic target that drives breast cancer metastasis and resistance to paclitaxel, Nat Commun 6 (2015) 8746, https://doi.org/10.1038/ncomms9746.
65. N. Fuchs, M. Meta, D. Schuppan, L. Nuhn, T. Schirmeister, Novel opportunities for cathepsin S inhibitors in cancer immunotherapy by nanocarrier-mediated delivery, Cells 9 (2020) 2021, https://doi.org/10.3390/cells9092021.
66. H.J. Xi, R.P. Wu, J.J. Liu, L.J. Zhang, Z.S. Li, Role of acetylcholinesterase in lung cancer, Thorac Cancer 6 (2015) 390–398, https://doi.org/10.1111/1759-7714.12249.
67. N.A. Keane, M. Reidy, A. Natoni, M.S. Raab, M. O'Dwyer, Targeting the Pim kinases in multiple myeloma, Blood Cancer J 5 (2015) e325, https://doi.org/10.1038/bcj.2015.46.
68. J. Zhou, W.R. Chen, L.C. Yang, J. Wang, J. Sun, W. Zhang, et al., KIF11 functions as an oncogene and is associated with poor outcomes from breast cancer, Cancer Res Treat 51 (2019) 1207–1221, https://doi.org/10.4143/crt.2018.460.
69. A.A. Shafi, A.E. Yen, N.L. Weigel, Androgen receptors in hormone-dependent and castration-resistant prostate cancer, Pharmacol Ther 140 (2013) 223–238, https://doi.org/10.1016/j.pharmthera.2013.07.003.
70. V. Dolhyi, D. Avierin, M. Hojouj, I. Bondarenko, Tubulin role in cancer development and treatment, Asploro J Biomed Clin Case Rep 2 (2019) 15–22, https://doi.org/10.36502/2019/asjbccr.6154.
71. M.D. Castellone, R.M. Melillo, RET-mediated modulation of tumor microenvironment and immune response in multiple endocrine neoplasia type 2 (MEN2), Endocr Relat Cancer 25 (2018) T105–T119, https://doi.org/10.1530/ERC-17-0303.
72. Y. Kudo, S. Iizuka, M. Yoshida, T. Tsunematsu, T. Kondo, A. Subarnbhesaj, et al., Matrix metalloproteinase-13 (MMP-13) directly and indirectly promotes tumor angiogenesis, J Biol Chem 287 (2012) 38716–38728, https://doi.org/10.1074/jbc.M112.373159.
73. V.T. Karamyan, The role of peptidase neurolysin in neuroprotection and neural repair after stroke, Neural Regen Res 16 (2021) 21–25, https://doi.org/10.4103/1673-5374.284904.
74. D.T. Chen, J.H. Pan, Y.H. Chen, W. Xing, Y. Yan, Y.F. Yuan, W.A. Zeng, The mu-opioid receptor is a molecular marker for poor prognosis in hepatocellular carcinoma and represents a potential therapeutic target, Br J Anaesth 122 (2019) e157–e167, https://doi.org/10.1016/j.bja.2018.09.030.