Università degli Studi di Palermo

AREA QUALITÀ, PROGRAMMAZIONE E SUPPORTO STRATEGICO
SETTORE STRATEGIA PER LA RICERCA
U. O. DOTTORATI

Dottorato in Scienze Molecolari e Biomolecolari.
Dipartimento di Scienze e Tecnologie Biologiche, Chimiche e Farmaceutiche (STEBICEF).
Settore Scientifico Disciplinare CHIM/08.

Correlation between cell line chemosensitivity and protein expression pattern as new approach for the design of targeted anticancer small molecules

IL DOTTORE
GABRIELE LA MONICA

IL COORDINATORE
PROF.SSA GIOVANNA PITARRESI

IL TUTOR
PROF. ANTONINO LAURIA

CO TUTOR
PROF.SSA ANNAMARIA MARTORANA

CICLO XXXV.
ANNO CONSEGUIMENTO TITOLO 2022.

Table of Contents

Abstract

***Background and rationale**:* Over the past few decades, several databases with a significant amount of biological data related to cancer cells and anticancer agents (e.g.: National Cancer Institute database, NCI; Cancer Cell Line Encyclopedia, CCLE; Genomic and Drug Sensitivity in Cancer portal, GDSC) have been developed. The huge amount of heterogeneous biological data extractable from these databanks (among all, drug response and protein expression) provides a real foundation for predictive cancer chemogenomics, which aims to investigate the relationships between genomic traits and the response of cancer cells to drug treatment with the aim to identify novel therapeutic molecules and targets. In very recent times many computational and statistical approaches have been proposed to integrate and correlate these heterogeneous biological data sequences (protein expression – drug response), with the aim to assign the putative mechanism of action of anticancer small molecules with unknown biological target/s. The main limitation of all these computational methods is the need for experimental drug response data (after screening data). From this point of view, the possibility to predict *in silico* the antiproliferative activity of new/untested small molecules against specific cell lines, could enable correlations to be found between the predicted drug response and protein expression of the desired target from the very earliest stages of research. Such an innovative approach could allow to select the compounds with molecular mechanisms that are more likely to be connected with the target of interest preliminary to the *in vitro* assays, which would be a critical aid in the design of new targeted anticancer agents.

***Results**:* In the present study, we aimed to develop a new innovative computational protocol based on the correlation of drug activity and protein expression data to support the discovery of new targeted anticancer agents. Compared with the approaches reported in the literature, the main novelty of the proposed protocol was represented by the use of predicted antiproliferative activity data, instead of experimental ones.

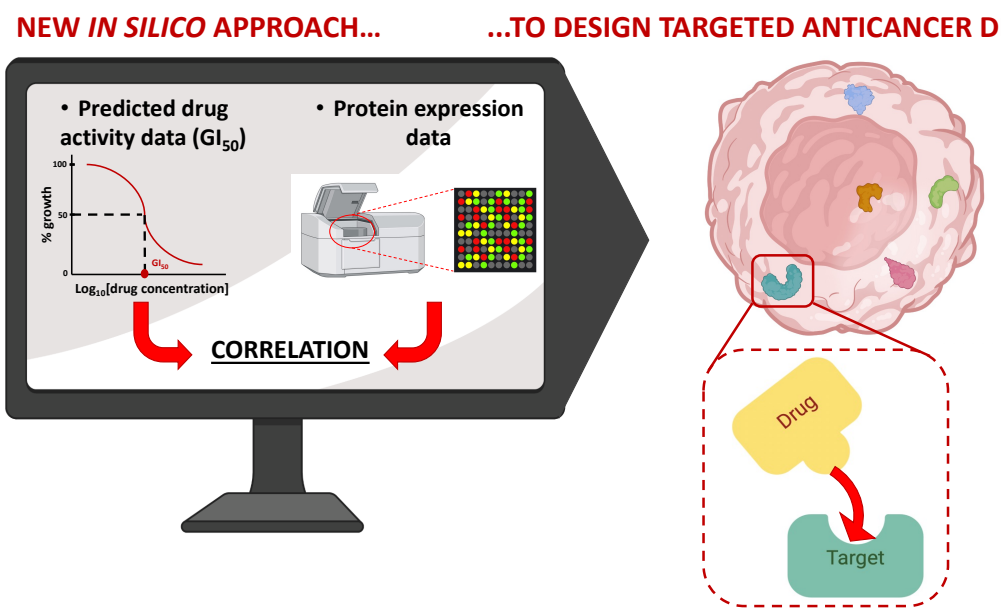
To this aim, in the first phase of the research the new *in silico* Antiproliferative Activity Predictor (AAP) tool able to predict the anticancer activity (expressed as GI_{50}) of new/untested small molecules against the NCI-60 panel was developed. The ligand-based

tool, which took the advantages of the consolidated expertise of the research group in the manipulation of molecular descriptors, was adequately validated and the reliability of the prediction was further confirmed by the analysis of an in-house database and subsequent evaluation of a set of molecules selected by the NCI for the one-dose/five-doses antiproliferative assays.

In the second part of the study, a new computational method to correlate drug activity data and protein expression pattern data was proposed and evaluated by analyzing several case studies of targeted drugs tested by NCI, confirming the reliability of the proposed method for the biological data analysis.

In the last part of the project the proposed correlation approach was applied to design new small molecules as selective inhibitors of Cdc25 phosphatase, a well-known protein involved in carcinogenic processes. By means of this innovative approach, integrated with other classical ligand/structures-based techniques, it was possible to screen a large database of molecular structures, and to select the ones with optimal relationship with the focused target. *In vitro* antiproliferative and enzymatic inhibition assays of the selected compounds led to the identification of new structurally heterogeneous inhibitors of Cdc25 proteins and confirmed the results of the *in silico* analysis.

***Conclusions**:* Collectively, the obtained results showed that the correlation between protein expression pattern and chemosensitivity is an innovative, alternative, and effective method to identify new modulators for the selected targets. In contrast to traditional *in silico* methods, the proposed protocol allows for the selection of molecular structures with heterogeneous scaffolds, which are not strictly related to the binding sites and with chemical-physical features that may be more suitable for all the pathways involved in the overall mechanism. The biological assays further corroborate the robustness and the reliability of this new approach and encourage its application in the anticancer targeted drug discovery field.



Graphical abstract.

1. Introduction

1.1. Biological data correlation as innovative approach for the design of targeted anticancer small molecules

The positive impact of targeted therapy and precision oncology in clinical treatments is driving medicinal chemistry research toward the identification of new and/or repurposed molecules with optimal ligand–target relationships. In this context, both the identification of target protein/s and the investigation of the potential mechanism of action of new small molecules exerting antiproliferative activity represent a great challenge. The unceasing development of databases storing a considerable amount of biological data related to cancer cells and anticancer drugs (gene expression, protein activity, mutation status, drug sensitivity data, etc.) undoubtedly contributes to this goal and represents an invaluable source of information for the entire scientific community [1-3].

Among the existing databases, the National Cancer Institute (NCI) database is probably the best known. Indeed, since 1990, NCI has been developing and improving within the Developmental Therapeutics Program (DTP) an experimental project named 'NCI-60 Tumor Cell Lines Screen'. Through this, the anticancer efficacy of both synthetic and natural molecules from laboratories around the world can be tested monthly under standardized conditions against approximately sixty cell lines, derived from nine different tumor types (leukemia, melanoma, lung, colon, brain, ovarian, breast, prostate, and kidney cancers). To date, biological data related to the anticancer activity of more than 250 000 compounds, expressed as GI_{50}, TGI, and LC_{50}, have been provided to the scientific community [4]. In addition, as a part of the so-called 'Molecular Characterization Program', NCI determines the gene and protein expression profile of thousands of molecular targets across NCI-60 [5].

Other important databases include Cancer Cell Line Encyclopedia (CCLE) and Genomic and Drug Sensitivity in Cancer portal (GDSC), which store information related to both the target profiling of >1000 cell lines (more than those of NCI) and drug activity data of a few compounds (24 and ~150 for CCLE and GDSC, respectively) [6,7].

The large amount of data extractable from these databases provides a real basis for predictive cancer chemogenomics, which studies the relationships between genomic traits

and the response of cancer cells to drug treatment with the aim to identify novel therapeutic molecules and targets. In other words, the antiproliferative effect of small molecules against cancer cell lines could be interpreted as a signature of the drugs at both the transcriptional and molecular level. Consequently, finding a correlation between the two types of data (e.g: drug activity and expression of the target protein) could provide new insight into the mechanism of action and putative biological target of several compounds, thus offering a valuable aid in anticancer drug design [8].

In this context, the most recent generation of high-performing statistical, computational, and machine-learning techniques play a crucial role, allowing to process a great amount of biological information quickly and to elucidate, with remarkable reliability, the correlation between the biological features of a specific cancer and the observed sensitivity to a given pharmacological agent. To highlight the great opportunity given by these approaches in the design of new selective anticancer agents, the most important literature applications deserve to be described in detail.

1.2. *In silico* approaches based on the correlation between cell line chemosensitivity and gene/protein expression pattern

Protocols able to integrate and correlate rapidly and efficiently the high number of biological data, collected across multiple screenings, could enable the extrapolation of information and development of hypotheses for the possible mechanism of action/ biological target of a given small molecule (**Figure 1**) [3,9-11].

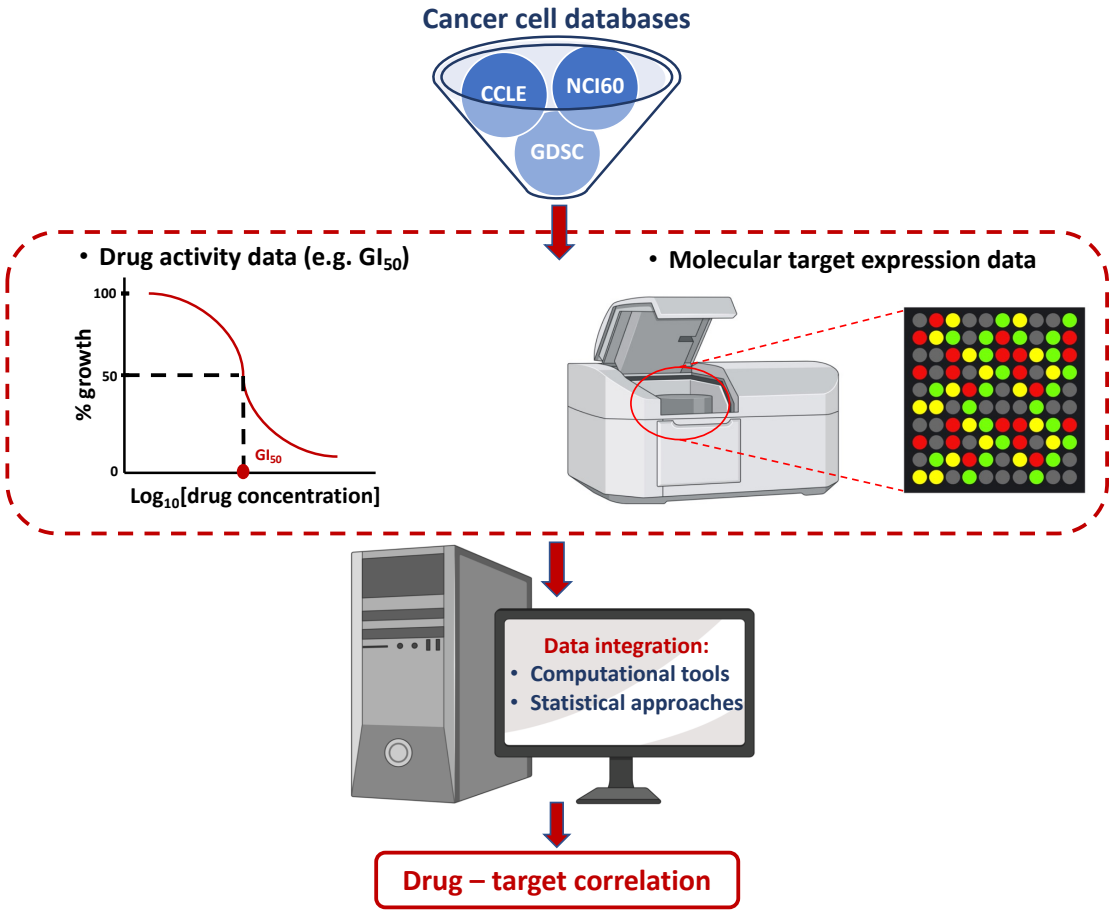


Figure 1. *Importance of the available computational techniques for the integration of the biological data (drug activity and protein expression) stored on the online databases.*

One of the best known tools developed by the NCI to correlate the biological data stored in its database is COMPARE [12]. The development of this algorithm has enhanced the value of the anticancer screening program as a discovery tool. Using COMPARE, a probe or 'seed' compound, identified by its NCI accession number, can be compared with all the compounds of the database. All these molecules are then ranked according to the similarity of their biological activity, expressed as the cell sensitivity of each of the 60 cell lines in the screening protocol, in terms of the biological response, toward the same 60 cell lines of the seed compound.

Starting from the hypothesis that compounds with similar biological activity patterns share a similar mechanism of action, the pairwise correlations provided by COMPARE suggest that the highest compounds in this ranking act through a mechanism similar to that of the seed compound. By contrast, because the proteome profile is different in different

cell lines, the COMPARE algorithm could also give an estimation of the affinity of a given compound toward a specific molecular target by correlating the drug sensitivity of a set of cell lines with their proteome profile. The requirement and also the limit of COMPARE is the need for anticancer screening results.

Two other well-known projects aimed at integrating biological data of cancer cell lines are CellMiner and CellMiner Companion (RCellMiner). These free-access web-based suites, representing the evolution of COMPARE, enable comparative analyses (pattern comparison and cross-correlation tools) exploiting the resources of the NCI database (genomic, molecular, phenotypic, and pharmacological data). These tools are based on the calculation of z-scores for the biological data of interest and on the ability to perform integration and find correlations (calculation of correlation coefficients) between drug activity (GI_{50} against NCI-60) and protein expression levels, transcription levels of miRNAs, and DNA copy or gene mutation data related to NCI-60 cancer cell lines. These analyses enable the extrapolation of information about possible mechanisms of action or resistance to chemicals of interest [13-17]. The evolution of these suites is the CellMiner Cross-Database (CellMiner-CDB), which, recently updated to version 1.2, is the product of four different cancers cell line pharmacogenomic databases (NCI-60, CCLE, GDSC, and CTRP). The integration and leverage of these databanks by means of statistical methods reinforces the reliability and reproducibility of the performed integrative analysis, making the CellMiner-CDB a unique resource [18,19]. The flexibility of this tool allows correlation analysis for specific cancer cell lines, such as the specific tool focused on small cell lung cancer cell lines (SCLC-CellMiner) developed by Tlemsani et al [20].

Recently, Krushkal et al., considering the importance of natural sources for new possible lead compounds, correlated the GI_{50} values and protein expression data stored in NCI and CellMiner to elucidate the potential mechanism of action of numerous natural products (from plants, marine invertebrates, and microbes) [21].

The lack of recurring updates of the CellMiner database induced Arroyo and co-workers to download all the available data from NCI-60 and undertake large-scale screening to find a connection between standardized cancer cell line sensitivity to antiproliferative agents

and corresponding standardized protein expression. The results obtained showed that the correlation between the antiproliferative activity of a given drug and the expression of a given protein could be exploited to: (i) confirm the biological target of known drugs; (ii) associate a mechanism of action with a new small molecule; (iii) repurpose an already developed drug against another target with which it is 'correlated'; (iv) explain the pleiotropic effects of drugs (multi/off target view); and (v) find targets that can be influenced by a given drug. Interestingly, all the data were uploaded to an open-access database called GEDA (Gene Expression and Drug Activity), which allowed the correlations to be analyzed graphically by means of bar plots representing the level of gene expression per drug [22].

Two other databases developed to perform integrative and correlation analysis are PharmacoDb and GDA (Genomic and Drugs Integrated Analysis). The former includes data from seven different databases and is coupled with a bioinformatic tool called PharmacoGx, which can be used to perform correlation analysis [23,24]. In the same way, GDA, comprising data from >50 000 compounds and >70 cancer cell lines, was developed to provide the scientific community with an easy-to-use web-based tool for integrative genomic determinant–drug activity data integration [25].

Rees et al. developed a computational tool based on statistical analysis that correlates the biological data of basal gene expression with experimental chemosensitivity, expressed as the area under the curve (AUC) of concentration–response graphs, of a group of known drugs. They found that matching the two biological data types could aid in identifying the biological target/s of new molecules with unknown mechanism of action and the additional targets responsible for determining the antiproliferative effect of known targeted drugs, as well as clarifying the resistance mechanism (high correlation with expression of proteins involved in metabolic inactivation or in the import/export of chemicals into/from the cell). The advantage of this method is the free availability of data with a detailed explanation of the analytical tool made available online by the Cancer Therapeutics Response Portal [26-28]. This protocol led to the identification of PDE3A as a potential target for the compound DNMDP, demonstrating that predictive chemogenomics based on the correlation between

protein expression and chemosensitivity could represent an important resource to gain insight into small-molecule targets [29]. Recently, taking advantage of this method, the putative molecular targets/mechanism of action of several non-oncology drugs (e.g., disulfiram) endowed with anticancer properties have been proposed [30].

Gonçalves and co-workers recently used an indirect approach to identify the biological target(s) and mechanism of action of antiproliferative compounds. They proposed a method comprising the correlation between CRISPR-based gene knockout (loss of-function screen) of numerous target proteins with drug sensitivity across >400 cell lines. They correlated/integrated the effect of expression arrest of a target with exposition to a drug of a given cell line under the hypothesis that a correlation in cell viability between a genic knockout and the antiproliferative effect of a drug would support the hypothesis that the considered target could be linked with the mechanism of action of the small molecule [31,32].

In **Table 1** an overview of the most representative examples of correlation tools developed to date is reported, highlighting which online databases are considered and which types of correlation analysis can be performed by each one.

Table 1. *Overview of the most representative examples of correlation tools developed to date.*

Tool/Computational Protocol	Databases considered	Type of correlation analysis performed	Ref.
COMPARE	NCI-60	Compound vS compound, compound vS target	[12]
RCellMiner	NCI-60, CCLE, GDSC, CTRP	Drug activity vS protein expression levels/transcriptions levels of miRNAs/DNA copy/gene mutations	[14-16]
GEDA	NCI-60	Drug activity vS protein expression	[22]
GDA	CCLE, NCI60	Drug activity (50 000 cpds) vS cell lines data (> 70)	[25]
PharmacoDB/ PharmacoGx	CCLE, GDSC, gCSI, CTRP, OHSU, FIMM, UHNBreast	Drug activity data vS Target data across seven different databases	[24]

The overview of the most interesting protocols reported in the literature showed how the integration of drug activity and molecular target data (protein/gene expression), which

are currently stored in several databases, represents a useful approach for the identification of both biological target/s and putative action mechanisms involved in the antiproliferative effect of new small molecules designed as targeted anticancer agents.

However, the main limitation of all these computational methods is the need for experimental drug response data (post-screen application). In this light, a turning point could be the ability to predict, using computational tools, the antiproliferative activity of newly designed compounds. Such an approach, if built and validated appropriately, could enable new correlations to be found between the predicted drug response and protein expression of the desired target from the very earliest stages of research, providing a valuable help in anticancer targeted drug discovery field.

2. Aim of the work

The correlation of biological data, such as drug activity – protein expression, by means of computational tools could represent an attractive frontier in the anticancer targeted drug discovery field. In this context, the development of tools capable to predict/estimate reliably and rapidly the anticancer activity of new compounds could represent a turning point, enabling new correlations to be found between the predicted drug response and protein expression of the desired target from the very earliest stages of research. Such an innovative approach could allow to select the compounds with molecular mechanisms that are more likely to be connected with the target of interest preliminary to the *in vitro* assays.

In the light of these considerations, this PhD study aims to propose an innovative and reliable *in silico* protocol capable to design new selective targeted anticancer agents by exploiting the approach based on the correlation drug activity - protein expression.

The objectives of the research can be summarized as follows:

1. Development and validation of a computational tool for predicting the anticancer activity of new/untested chemical structures against the NCI-60 panel, one of the most characterized sets of cancer cell lines. This preliminary step of the project served as foundation for the next steps;
2. Development of a new *in silico* method able to correlate different types of biological data, such as drug activity and protein expression, and to extrapolate connections between drugs and target/s;
3. Application of the proposed method to analyze several case studies of known targeted anticancer agents, by exploiting the biological data stored in the NCI database. This step served as validation for the correlation method;
4. Application of the developed correlation method to design new targeted anticancer drugs and consequent biological assays to evaluate the effectiveness of the proposed *in silico* predictions.

3. Results and Discussion

3.1. Antiproliferative Activity Predictor: a new open-access ligand-based tool implemented in DRUDIT^{online} web-platfom for drug response prediction against NCI-60 panel

3.1.1. Computational protocols to predict cell line chemosensitivity: state of art

To apply the proposed innovative idea of correlation drug activity – protein expression for the design of new targeted anticancer compounds, a computational protocol able to predict the antiproliferative activity of untested compounds is required.

In this context, in the last decades several examples of such protocols and machine learning approaches have been proposed, as recently well reviewed by Firoozbakht et al [33].

For instance, Lind et al. developed a regression random forest model by the integration of mutational status data of 145 selected oncogenes from the Genomics of Drug Sensitivity in Cancer (GDSC) and more than 1200 molecular descriptors to reliably calculate the GI_{50} values of selected compounds against cancer cells [34]. Similarly, Zhang et al. drew up a dual-layer integrated cell-drug network capable of predicting drug response using gene expression profiles of numerous cancer cell types available in the Cancer Cell Line Encyclopedia (CCLE) and the Comprehensive Genomic Profiling (CGP), and providing chemical proprieties of drugs captured by molecular descriptors [35]. In addition, numerous web-based tools have been described such as: the PaccMann web-service, which can estimate the chemosensitivity of cancer cell line by integrating both cancer cell and chemical structures features [36]; CDRscan [37], a deep learning model that assesses anticancer drug responsiveness based on large-scale drug screening assay data. It employs a two-step convolution architecture, where the genomic mutational fingerprints of cell lines and the molecular fingerprints of drugs are processed individually, then merged by 'virtual docking'. Analysis of the extrapolation capability revealed a high accuracy ($R^2 > 0.84$; AUROC > 0.98). The tool was applied to a large set of approved drugs, allowing the identification of 14 oncology and 23 non-oncology drugs; $DeepIC_{50}$ [38], a 1-dimensional convolution neural

network model, designed to predict drug responsiveness classes, based on a large set of features. As a result, it showed better cell viability IC_{50} prediction accuracy in pan-cancer cell lines over two independent cancer cell line datasets; and, more recently, pdCSM-cancer, which uses graph-based signature representation to estimate antiproliferative activity against multiple cancer cell lines [39].

However, despite these great efforts, a protocol capable of reliably, accurately, and rapidly screening and predicting the anticancer activity of huge structure databases against large panels of cell lines, such those of the NCI, has not been developed yet to date.

In the light of these considerations, in the first part of this PhD thesis, based on the available antiproliferative data collected by NCI and the previous knowledge of the research group on molecular descriptor manipulation [40-42], the new *in silico* ligand-based Antiproliferative Activity Predictor (AAP) tool, is proposed. AAP can predict the GI_{50} values (half-maximal cell growth inhibitory concertation) of input structures for the entire NCI-60 panel.

In the next sections, the rationale, the computational details, the validation, and the application of the AAP are described and discussed in detail.

3.1.2. Database selection

The first step in establishing the AAP protocol was to select a reliable set of chemical structures, with known anticancer activity (mainly estimated as GI_{50} values), to be used as foundation for the models. For this purpose, the NCI-60 database proved to be one the most suitable [43].

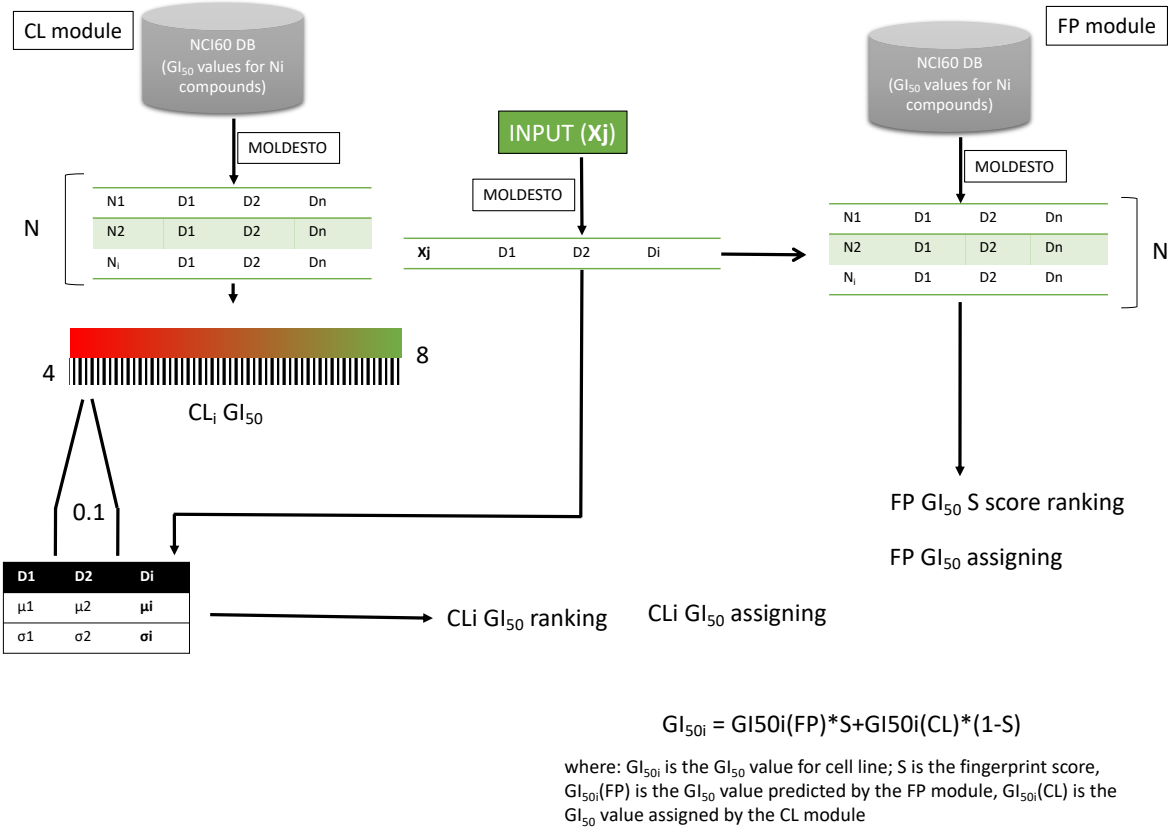
In particular, from the thousands of structures tested by the NCI, those that were tested in a 5-dose assay, and for which the experimental GI_{50} values were therefore available, were selected. According to the publication date (see Materials and Methods section, paragraph 4.1.1), the structure database was divided in two groups [44]:

- the compounds tested and published up to September 2014 (hereinafter referred to as NCI2014DB) were selected as training set and used to build the model;
- the others, tested and published between September 2014 and June 2016 (referred to as NCI2016DB) were used as a test set to validate the tool.

The next two sections describe the creation of the protocol and the subsequent validation.

3.1.3. Fingerprint (FP) and Cell Line (CL) modules: the two pillars of the AAP tool

The AAP protocol is able to assign -logGI_{50} values (indicated as GI_{50} in the following sections) to the input structures against the sixty cell lines of the NCI-60 panel (L_n). The assignment of each predicted GI_{50} value results from the weighed contribution of two "modules", called Finger-Print (FP) and Cell-Lines (CL), which work synergistically together through a series of well-considered steps, as shown in **Figure 2**.



$$GI_{50i} = GI50i(FP)*S+GI50i(CL)*(1-S)$$

where: GI_{50i} is the GI_{50} value for cell line; S is the fingerprint score, GI_{50i}(FP) is the GI_{50} value predicted by the FP module, GI_{50i}(CL) is the GI_{50} value assigned by the CL module

Figure 2. *Flow-chart of the Antiproliferative Activity Predictor (AAP) protocol: the chemical structures of the compounds belonging to the NCI2014DB (training set, N) are submitted to the molecular descriptor calculation performed by MOLDESTO software to build the two modules FP and CL (FingerPrint and CellLine, respectively). For each input structure submitted to the tool, each module separately assigns a GI_{50} value, GI_{50}(FP) and GI_{50}(CL), according to the workflow depicted in the scheme. These two parameters are weighed by means of the formula showed in the lower part of the scheme to assign the predicted GI_{50} value against each of the sixty NCI cell lines (i).*

Preliminarily, a set of molecular descriptors (D_1, D_2, D_3, D_n) was calculated for the training set (34k structures, N, from the NCI2014DB). The calculation of molecular

descriptors was performed by the *in-house* MOLDESTO software which calculates more than 1000 1D, 2D, and 3D molecular descriptors (D_i) for the input structures (Supplementary Material S1; see Materials and Methods section, paragraph 4.1.1). Then, the above-described modules (FP and CL) were built.

FP module relies on the molecular descriptors calculated for the training set structures and on the associated GI_{50} experimental values. The idea behind the FP module is that similar structures might have similar biological activities (FingerPrint matching). In details, the FP module matches the molecular descriptors of the input structure X_j with those of the structures, one by one, of the training set and assigns the S score, in the range 0-1. This parameter measures the degree of similarity between the input structure and each structure included in the training set. By ranking according to S score, the protocol assigns to the input structure the experimental GI_{50} values of the best scored structure of the training set (GI_{50}(FP)). If experimental GI_{50} data are missing, no GI_{50} value is assigned to the input structure for that specific cell line.

The **CL module**, on the other hand, is based on the cell lines. For each of the sixty NCI cell lines, 42 templates (CL_i) were built: 40 of which for the GI_{50} values in the range 4-8 (0.1 unit each) and 2 for GI_{50} values <4 and >8, respectively.

With this aim, the structures of the training set were assigned to each template according to their experimental GI_{50} values. Then, a couple of values, mean (μ_i) and the standard deviation (σ_i), was computed for all molecular descriptors, considering the structures belonging to each template.

Thus, the molecular descriptor values of the input structure are matched with the 42 templates for the sixty cell lines: to the input structure is assigned the GI_{50} value (GI_{50}CL) of the corresponding template with the higher score for each cell line (see **Figure 3** for further details about the CL scoring).

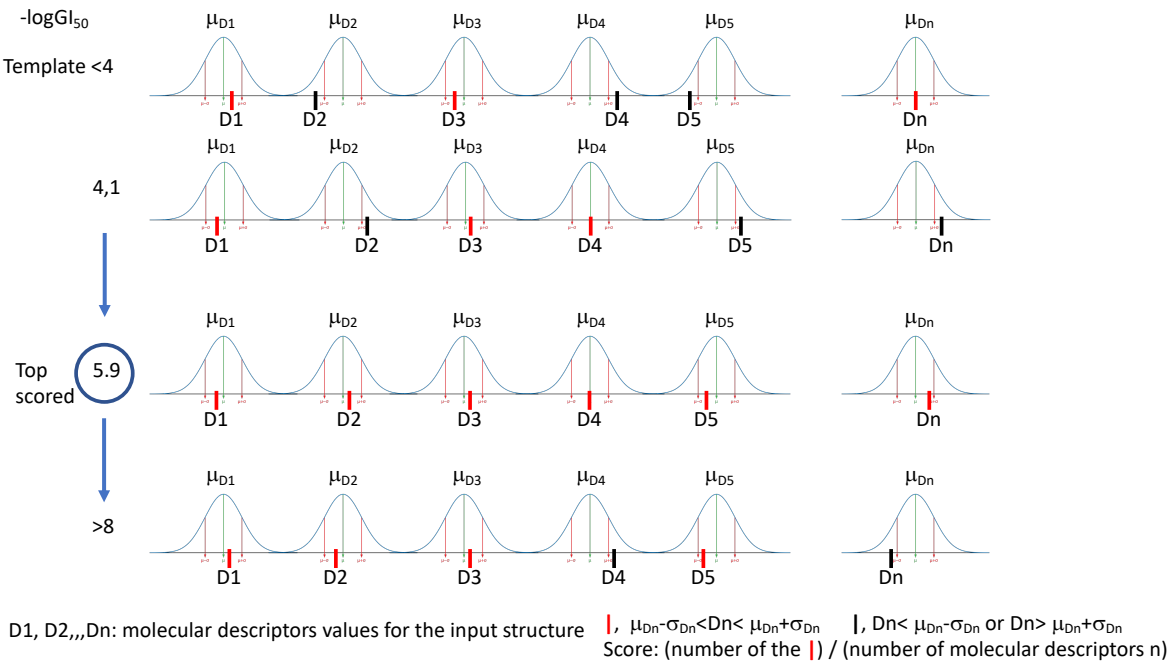


***Figure 3**. Graphical representation of GI_{50} prediction by the CL module for a cell line.*

Thus, once an input structure X_j is uploaded into DRUDIT tools interface, MOLDESTO optimizes the geometry *in vacuo*, and calculates the molecular descriptors described above for the training set. Then, the molecular descriptors values are submitted to FP and CL modules.

The output data from these modules are weighted as shown below:

$$GI_{50i} = GI_{50i}(FP)*S+GI_{50i}(CL)*(1-S)$$

where: GI_{50i} is the GI_{50} value for that cell line; S is the fingerprint score, $GI_{50i}(FP)$ is the GI_{50} value predicted by the FP module, $GI_{50i}(CL)$ is the GI_{50} value assigned by the CL module.

From this formula, if the structural similarity between the input structure and the best scored structure of the training set is high (S score close to 1), it is assumed that the biological activity of the input structure is very similar to that of the compound of the training set (similar structure could correspond to a similar biological activity). When S=1, the input structure is included in the training set, thus the predicted GI_{50} correspond to the experimental values. Instead, if S is not close to 1, $GI_{50}(CL)$ contributes more to the overall

result, according to the S value. In case the GI_{50}(FP) is not available (unavailable data from NCI screening), the GI_{50} will correspond to the GI_{50}(CL).

3.1.4. Validation of the AAP tool

The predictive ability of AAP was validated by internal and external datasets. Both required a set of NCI-60 compounds tested at five-doses. The first set was generated by randomly selecting compounds with available antiproliferative activity data belonging to the NCI2014DB database. These molecules were used to validate the CL module, the predictive part of the protocol.

The second set of molecules was assembled from the compounds added to the NCI database after 2014 (NCI2016DB, test set) and therefore not included as training set in the CL and FP modules; this cluster of compounds was used to validate the Antiproliferative Activity Predictor tool (GI_{50}).

Internal validation: 5 ‰ of the training database structures (193 molecules), randomly selected from the NCI2014DB (Supplementary Material S2), were used to validate the CL module by matching the calculated GI_{50}(CL) values with the experimental GI_{50} data. Because these structures were used to generate the AAP protocol, their experimental GI_{50} values were indicated by the FP protocol, except for those that were not available (the experimental GI_{50} are listed in Supplementary Material S3; an empty box in the matrix indicates unavailable experimental data). The 193 structures were clustered in three groups, according to GI_{50} values: the most active compounds (more than 40/60 GI_{50} values equal to 8); the structures with GI_{50} values in the range of 4-8; and the cluster of low active/inactive compounds, with GI_{50} values close to 4.

Therefore, GI_{50}(CL)s were first calculated for the selected structures by setting the DRUDIT parameters (see Materials and Methods section for the meaning of the DRUDIT parameters). This step has two aims: it allows to verify the predictive capability of the CL module and, more importantly, to optimize the DRUDIT parameters for the best prediction of antiproliferative activity for new compounds. Thus, runs 1-18 were performed by modulating the values of the parameters N, Z, and G (see Materials and Methods, paragraph

4.1.1 for further details), as reported in **Table 2**. The 18 outputs for CL module are reported in Supplementary Material S4.

Table 2. *Overview of DRUDIT runs for the parameters tuning, absolute deviation from true values (|DTV(GI_{50})|) is reported, run code (1-18) is given in parentheses.*

Z	N	G		
		a	b	c
50	240	1.22 (1)	1.23 (2)	1.23 (3)
	500	1.22 (7)	1.30 (8)	1.31 (9)
	760	1.32 (13)	1.44 (14)	1.42 (15)
100	240	1.31 (4)	1.64(5)	1.72 (6)
	500	1.23 (10)	1.51 (11)	1.53 (12)
	760	1.28 (16)	1.42 (17)	1.44 (18)

The eighteen matrices from CL were matched with the experimental GI_{50} to obtain 18 new matrices in which the |DTV (GI_{50})|, i.e., the absolute deviation of the calculated GI_{50}(CL) from the experimental GI_{50} value, was reported for each structure. Furthermore, the average |DTV(GI_{50})| for all runs was examined for each entry.

From the analysis of these data, it appears that the protocol allows to identify the potentially inactive or on-average active structures (below <4 or in the range 4-7) with a remarkable degree of reliability, while it is less effective for structures with high activity (in the range 7-8), but with acceptable errors.

Moreover, it is demonstrated that the quality of the prediction is closely related to the amount of available biological data used to build the model: since the number of high active compounds (7<GI_{50}<8) is very low with respect to inactive or medium active compounds, the prediction is negatively affected (higher error).

Moreover, the average |DTV(GI_{50})| for each cancer cell line was calculated for every run. We decided to exclude the M19-MEL melanoma cancer cell line from the analysis due to lack of sufficient biological data.

The matrix values provided by each run were further elaborated to calculate the overall average |DTV(GI_{50})| for each cancer cell line. SK-MEL-28, of the melanoma panel, yielded the best predictions (overall average |DTV(GI_{50})| of 1.14).

To select the optimized DRUDIT parameters, the overall average |DTV(GI_{50})| was calculated for each GI_{50}(CL) matrix (**Table 2**). The parameters of run 1 (N=240, Z=50, G=a) were identified as the best, with an overall average error of 1.22 (**Table 2**). In this run, the renal cancer panel was the best predicted, with an average error of 1.18. The full results are reported in Supplementary Material S5.

External test validation of Antiproliferative Activity Predictor: To validate the Antiproliferative Activity Predictor tool, a set of 99 molecules was collected from NCI2016DB (see Materials and Methods section, paragraph 4.1.1, for database selection and Supplementary Material S6). Their known GI_{50} values were compared with the predicted ones, calculated using the optimized DRUDIT parameters (run 1, see internal validation). The output matrix showed an interesting *scenario* with a total average |DTV(GI_{50})| of 0.87 and excellent predictions for structures with low activity (GI_{50}>4 for at least 40 cell lines). On the other hand, significant errors were observed for highly active structures (GI_{50} upper or close to 8). This evidence confirmed the capability of the protocol to better predict GI_{50} values for low activity molecules. The average errors for each cancer cell line were also calculated and showed excellent prediction for the breast cancer panel (average error of 0.77 for the panel) and especially against MDA-MB-231-ATCC (average error for the cancer line of 0.64).

Analyzing the |DTV(GI_{50})| for all selected structures, considered by ranges, it was found that the protocol was able to assign the correct value returning ÷1 for 65% of the data (**Figure 4**).

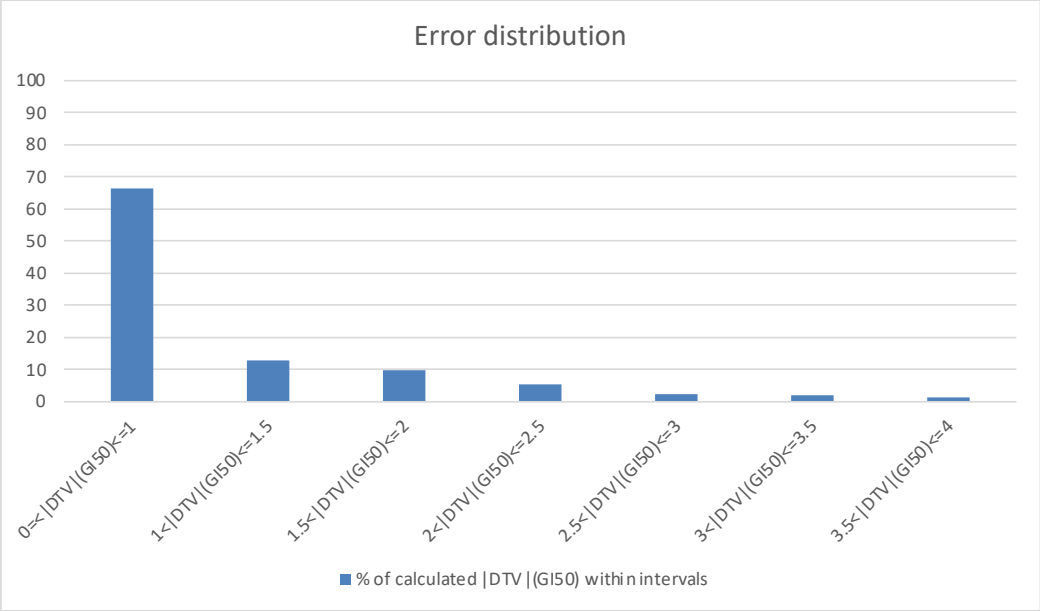


Figure 4. *Error distribution for the external validation of the AAP protocol.*

The matrix comparing experimental GI_{50} with calculated ones for all structures is reported in Supplementary Material S7, in addition to the full analysis data.

To demonstrate the relevant contribution of the CL module in the prediction, we also compared experimental GI_{50} values with calculated GI_{50}(FP) ones. The total average $|DTV(GI_{50})|$ of 0.95, which is higher than the one obtained by combining the predicted GI_{50} values of both modules, shows that the CL module improves accuracy and leads to better predictions (Supplementary Material S8).

3.1.5. Parameters optimization for cell line/subpanels activity prediction

Tuning of DRUDIT parameters (N, Z, and G) could also allow optimization of the prediction for a particular cell line or sub-panel (all the following results are shown in Supplementary Material S9).

With this aim, the analysis was addressed to determinate the best combination of parameters (1-18) for the nine NCI subpanels and for each cancer cell line as described above.

Then, the Average $|DTV(GI_{50})|$ obtained for the cancer cell lines in all runs were analyzed (Supplementary Material S5).

Regarding the optimization of parameters for each cancer cell line, SK-MEL-28 (melanoma panel) provided the best prediction, with an average error of 0.97 for parameter combinations 2 and 3. Full results are shown in **Table 3**.

***Table 3**. Tuning of DRUDIT parameters tuning for each cancer cell line.*

PANELS	CELL LINES	RUN	AVERAGE |DTV(GI_{50})|
Breast Cancer	BT-549	4	1.35
	HS-578T	3	1.30
	MCF7	1 or 7	1.30
	MDA-MB-231-	7	1.22
	T-47D	7	1.16
CNS Cancer	SF-268	7	1.17
	SF-295	1	1.25
	SF-539	4	1.18
	SNB-19	7	1.15
	SNB-75	1	1.16
	U251	1	1.24
Colon Cancer	COLO-205	10	1.13
	HCC-2998	1	1.09
	HCT-116	2 or 7	1.13
	HCT-15	2	1.21
	HT29	1	1.14
	KM12	1	1.19
	SW-620	2	1.14
Leukemia	CCRF-CEM	7	1.13
	HL-60TB	7	1.22
	K-562	2	1.27
	MOLT-4	3	1.12
	RPMI-8226	10	1.12
	SR	2	1.28
Melanoma	LOX-IMVI	3	1.16
	M14	1 or 3	1.20
	MALME-3M	10	1.19
	MDA-MB-435	3	1.22
	SK-MEL-2	3	1.03

	SK-MEL-28	2 or 3	0.97
	SK-MEL-5	2	1.26
	UACC-257	2	1.07
	UACC-62	10	1.31
Non-Small Lung Cancer	A549-ATCC	3	1.18
	EKVX	1	1.02
	HOP-62	1 or 8	1.19
	HOP-92	10	1.21
	NCI-H226	7	1.07
	NCI-H23	4	1.16
	NCI-H322M	1	1.10
	NCI-H460	2	1.26
	NCI-H522	7	1.09
Ovarian Cancer	IGROV1	1 or 3	1.25
	NCI-ADR-RES	4	1.31
	OVCAR-3	1 or 4	1.22
	OVCAR-4	7	1.00
	OVCAR-5	16	1.02
	OVCAR-8	1	1.14
	SK-OV-3	7	1.18
Prostate Cancer	DU-145	10	1.19
	PC-3	2	1.19
Renal Cancer	786-0	10	1.16
	A498	1	1.21
	ACHN	7	1.19
	CAKI-1	1	1.11
	RXF-393	10	1.12
	SN12C	1 or 10	1.16
	TK-10	10	0.99
	UO-31	2	1.16

To identify the best parameter combination for each panel, the averages $|DTV(GI_{50})|$ of cell lines obtained in each run were grouped by the panel and then the average $|DTV(GI_{50})|$ for the entire panel was calculated for runs 1-18. The results as best combination of parameters for each panel are given in **Table 4**.

Table 4. *Best parameter combinations for each of the NCI panel.*

PANELS	RUN	AVERAGE |DTV(GI50)|
Breast Cancer	1 or 3	1.37
CNS Cancer	1	1.23
Colon Cancer	1	1.19
Leukemia	2	1.23
Melanoma	3	1.18
Non-Small Cell Lung Cancer	2 or 7	1.20
Ovarian Cancer	1	1.21
Prostate Cancer	1	1.23
Renal Cancer	10	1.15

Renal cancer panel resulted the best predicted of all, with an average error of 1.15 by using parameter combination 10.

3.1.6. Application of the AAP tool for the virtual screening of an *in-house* structure database

Considering the high predictive power of the proposed protocol, it seemed interesting to exploit it for the analysis of an *in-house* small molecules database to select compounds to be submitted to the NCI-DTP screening program. This permitted to further confirm the reliability of the tool in the prediction of the anticancer activity of new small molecules.

In detail, this part of the study was carried out in collaboration with the organic chemistry group supervised by Prof. Antonio Palumbo Piccionello (University of Palermo), which has been working for many years on the synthesis of new biologically active curcumin-like compounds [45-47], and has provided an interesting database of structures that could be screened by the AAP tool.

The evaluation of curcumin-like compounds as antiproliferative agents against the NCI-60 panel appeared remarkably interesting. Indeed, curcumin has also been intensively studied in recent decades for its significant anticancer activity against various malignant cell types [48], and many curcumin analogues (e.g. GO-Y030, CDF, UBS109, EF24, EF31, **Figure 5**)

have already been studied as antiproliferative agents, showing even higher activity and improved drug-like properties compared to curcumin [49-51].

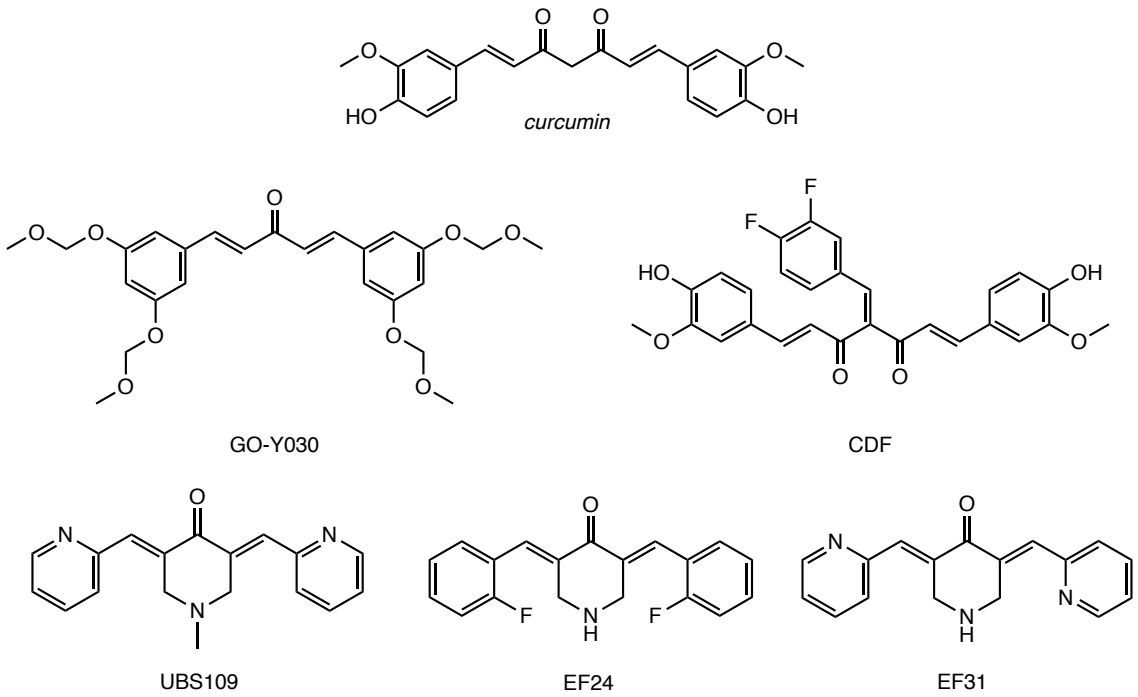
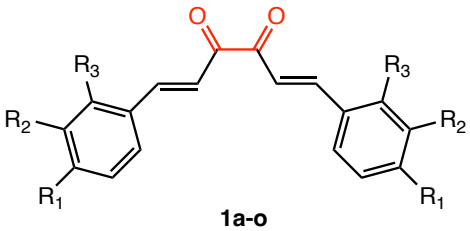


Figure 5. *Curcumin and curcumin-like biologically active compounds.*

In details, the selected *in-house* database included three different subclasses of curcumin-like derivatives, as reported in **Figure 6**: 1,2-diones (**1a-o**); 1,2,4-oxadiazoles (**2a-h**); 1,3,4-oxadiazoles (**3a-h**). They were developed by replacing the symmetrical β-diketone core of curcumin, which is responsible for unfavorable physicochemical properties and a weak pharmacokinetic profile [52-57], with stacked moieties, such as heterocycles or α-diketones; these replacements have been shown to improve stability, solubility, oral absorption, and bioavailability [58]. The three classes of structures are variously substituted, as depicted in **Figure 6**.

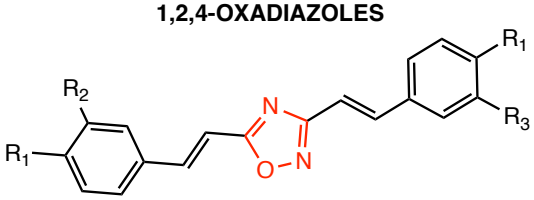
DIKETONES



1a-o

cpd	R_1	R_2	R_3
a	OCH_3	OCH_3	H
b	CH_3	H	H
c	OCH_2Ph	OCH_3	H
d	OCH_3	H	H
e	H	H	H
f	F	H	H
g	$N(CH_3)_2$	H	H
h	Cl	H	H
i	F	F	H
j	H	Cl	H
k	H	H	Cl
l	H	OCH_3	H
m	OCH_2OCH_3	OCH_3	H
n	OCH_2O	OCH_2O	H
o	COOH	H	H

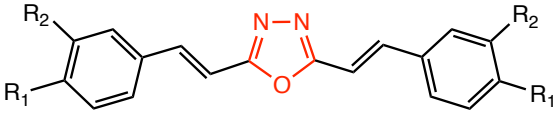
1,2,4-OXADIAZOLES



2a-h

cpd	R_1	R_2	R_3
a	OCH_3	OCH_3	OCH_3
b	OCH_3	H	H
c	H	H	H
d	OCH_3	Cl	OCH_3
e	$N(CH_3)_2$	H	H
f	Cl	H	H
g	F	OCH_3	OCH_3
h	F	H	H

1,3,4-OXADIAZOLES



3a-h

cpd	R_1	R_2
a	F	F
b	Cl	H
c	OCH_2O	OCH_2O
d	$N(CH_3)_2$	H
e	OCH_3	OCH_3
f	OCH_3	H
g	OCH_2OCH_3	OCH_3
h	F	OCH_3

Figure 6. *In house structures database of curcumin-like compounds studied with AAP tool.*

After selecting the molecule database, the next step was to tune the parameters of the AAP protocol. As mentioned earlier, the GI_{50} calculation can be targeted to a specific cell line or class of compounds by optimizing the parameters N, G, and Z. In this light, curcumin, tested by NCI (NSC code 32982) and included in the training set (NCI2014DB), was selected as reference compound to determinate the best combination of parameters for the CL module. The tuning was performed with the parameters in the ranges $250 < N < 800$ and $50 < Z < 100$ and considering a, b, or c for the G function. Eighteen runs were started, following the procedure described above for the internal validation (the combinations are reported

in **Table 2**). Consequently, the total absolute deviations from the experimental GI_{50} values were calculated for each run and the set N= 760, Z= 50, and G= c was identified (total $|DTV(GI_{50})|$= 0.44) and applied to the AAP tool for the selected database (see Supplementary Material S10).

The output of the AAP tool with the calculated GI_{50} values of the 39 in-house compounds is provided in Supplementary Material S11. The AAP GI_{50} values of the curcumin-like analogues were compared with the experimental ones determined by the NCI for the curcumin lead compound (Supplementary Material S11). The analysis of the average GI_{50}s of each compound for the full panel highlighted several curcumin-like molecules with a predicted antiproliferative activity better than that of the reference curcumin (average GI_{50} value of 5.17), such as: **1a** (5.47), **1e** (5.65), **1m** (5.82) for the diones class; **2a** (5.49), **2d** (5.57), and **2g** (5.44) for the 1,2,4-oxadiazole class; **3a** (5.72), **3e** (5.49), **3g** (6.27), and **3h** (5.47) for the 1,3,4-oxadiazole class. To test the consistency of the AAP protocol, compounds classified as both active and inactive (with a GI_{50} mean of less than 4.5) were proposed to the NCI for the *in vitro* evaluation of the antiproliferative activity against the NCI-60 human tumor cell lines, assuming that a reliable protocol must be able to identify both active and inactive compounds.

3.1.7. Biological assays: NCI-60 Human Tumor Cell Lines Screen for selected compounds

All the analyzed curcumin-like compounds were submitted to NCI cell line-based *in vitro* screening for anticancer drugs. As described in the Materials and Methods section (paragraph 4.2.1), the NCI applies specific criteria for compound selection. In the case of analogues, the selected compounds are those that are the most representative of the series and have significant structural novelty, compared to the NCI collection.

From the three series of curcumin-like derivatives, five molecules were selected by NCI for the *in vitro* biological screening: **1a** (NSC785541), **1b** (NSC785539), and **1c** (NSC785540) for the dione series; **2a** (NSC785542) for the 1,2,4-oxadiazole series; and **3e** (NSC785543) for the 1,3,4-oxadiazole series (**Figure 7**).

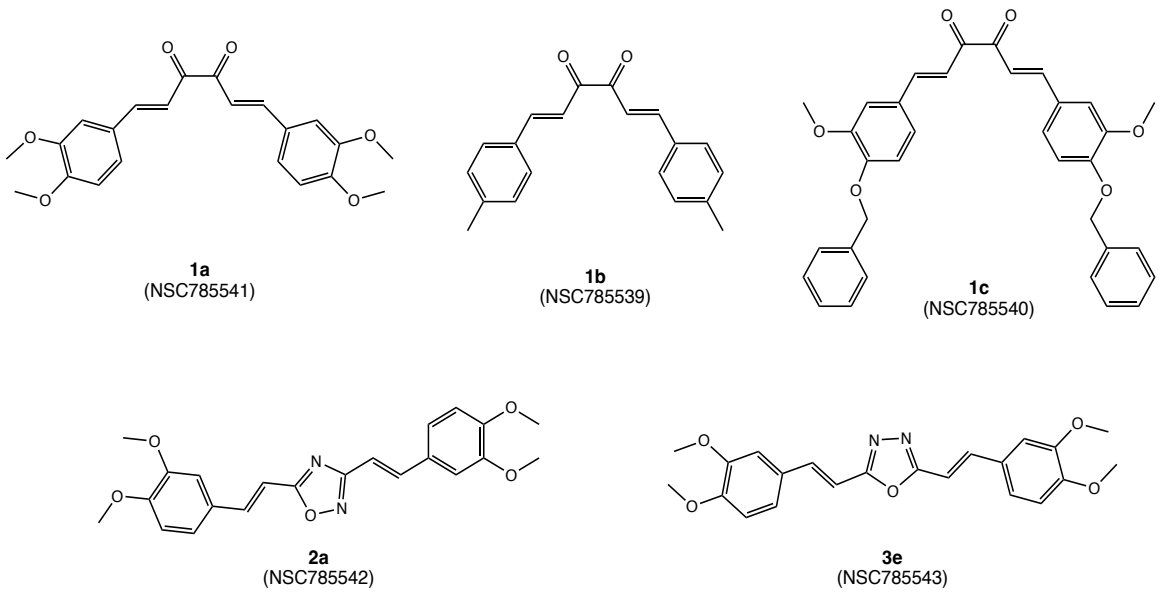


Figure 7. *Chemical structures of the five curcumin-like compounds selected by NCI for the one-dose antiproliferative assay.*

One-dose antiproliferative assay

The NCI screening protocol consists of a preliminary one-dose assay (concentration at 10 µM) against the full NCI-60 panel. Compounds that meet the NCI selection criteria and have a significant growth inhibitory effect on a minimum number of cell lines proceed to the 5-dose screening (experimental details are described in Materials and Methods section, paragraph 4.2.2). Results are expressed as percent of growth (G%) of the treated cells when compared to the untreated control cells. This parameter accurately expresses the anticancer potential of the drug. At G%>100, the compound has no effect on cancer cell proliferation (inactive). In the range, the compound inhibits cell proliferation by a percentage, which is expressed as 100-G%. When G% is <0, the compound is cytotoxic and lethal to the cancer cells. To graphically represent the most sensitive panels/cell lines, a mean growth percentage is also provided.

The mean G% for each of the nine subgroups is given in **Table 5** for the five compounds; the full results and the mean graphs at one dose screening are reported in Supplementary Material S12.

Table 5. *G% values determined for the five selected compounds against the NCI-60 panel at one dose assay.*

PANEL[1]	1a	1b	1c	2a	3e
Leukemia	14.84	78.71	96.37	77.47	18.53
Non-Small Cell Lung Cancer	61.79	98.02	95.47	84.39	29.86
Colon Cancer	-16.06	80.65	95.84	86.52	21.07
CNS Cancer	32.98	98.55	101.40	98.19	18.66
Melanoma	24.26	97.49	100.59	96.74	22.40
Ovarian Cancer	46.48	103.35	101.45	95.29	32.68
Renal Cancer	21.07	100.02	100.31	94.36	34.17
Prostate Cancer	33.06	102.04	101.62	88.89	40.34
Breast Cancer	18.21	86.77	99.62	88.93	24.71
Overall average	26.29	93.96	99.19	90.09	26.93

[1] For each compound the average G% values against the nine subpanels are reported.

Consistent with GI_{50} values predicted by AAP for these compounds, biological data confirm the dimethoxy-dione **1a** (NSC785541) and the dimethoxy-1,3,4-oxadiazole **3e** (NSC785543) as the most active curcumin-like derivatives. They showed a remarkable overall average G% (26.29 and 26.93 respectively), and an average inhibition of cell growth of about 75% compared to the full NCI-60.

In particular, compound **1a** demonstrates high lethality against colon cancer panel (average G% of -16.06), with the highest cytotoxic effect on cell lines HCT-116 (G%= -74.85) and HT29 (G%= -35.20), and against leukemia panel, with an average G% of 14.84 and almost complete blockade of cell growth (G% ~ 0) of cell lines K562, RPMI-8226, and SR (see Supplementary Material S12). Notable data are obtained for the LOX IMVI cell line (melanoma panel), with a G% of -66.41, and against the RXF 393 (renal cancer panel), with a G% of -63.27.

Similarly, dimethoxy-1,3,4-oxadiazole **3e** shows high antiproliferative effect against colon cancer panel, but with lower toxicity; it exhibits a G% close to 0, against the most sensitive cell lines, HCT116 and HT-29, implying an arrest of cell growth with low/no lethality. Furthermore, the leukemia, melanoma, and CNS cancer panels also show interesting sensitivity to the tested compound. Therefore, according to the selection criteria of the DTP NCI protocol, these two molecules progressed to the full five-dose assay.

The other three compounds **1b**, **1c** and **2a**, tested in the NCI one-dose protocol, generally exhibited less inhibitory activity, with an average G% next to 100.

These results confirm the AAP *in silico* data, according to the compounds **1b** and **1c** were predicted to be almost inactive, with a mean GI_{50} of 4.38 and 4.54 (low millimolar range), respectively.

On the other hand, when it is analyzed the antiproliferative effect on specific human cancer cell lines, dione **1b** reduces the growth of RPMI-8226 and MCF-7 cell lines by 55% and 76%, respectively (G% = 45.15 and 23.84), and induces a remarkable death in HCT-116 colon cancer cell line (G%= -36.84); **1c**, instead, exhibits a G% of 69.01 against HCT-116 colon cancer cells.

The 1,2,4-oxadiazole **2a**, which was predicted to be more active than curcumin, showed no significant anticancer activity against the NCI-60 database, except for the colon cancer HT29 cell line (G% of 56.32). This is an unexpected result considering that the isomer **3e**, was selected for the 5-dose screening. Probably, it could be hypothesized that the switch from the 1,3,4-oxadizole to the 1,2,4-oxadiazole core affect notably the capability of the compound to interfere with biological target/s. Thus, the two compounds **1a** and **3e** were selected for the five-dose screening to measure the GI_{50}s, which permitted us to further validate our tool.

*Five-dose antiproliferative assay for the most active derivatives **1a** and **3e***

The two selected compounds **1a** and **3e** were tested with the five-dose assay by measuring the percentage of cell growth at five different concentrations (from 10^{-8} to 10^{-4} M), as described in detail in the Materials and Methods section paragraph 4.2.3. For each

selected compound, NCI provided the measured GI_{50}, TGI, and LC_{50} values against the NCI-60 cell lines, with the corresponding mean graphs (see Supplementary Material S13).

In the first part of this section, attention is focused exclusively on the GI_{50} values, as these data allowed us to further assess the predictive ability of the proposed AAP protocol.

The comparison of the average predicted GI_{50} with the experimental values obtained by NCI for the two compounds confirmed that the protocol was able to predict with high accuracy the range of activity of both compounds against the full NCI-60 (for **1a**, the average predicted GI_{50} was 5.39, whereas the average experimental GI_{50} was 5.49; for **3e**, the average predicted GI_{50} was 5.41, whereas the average experimental GI_{50} was 5.28).

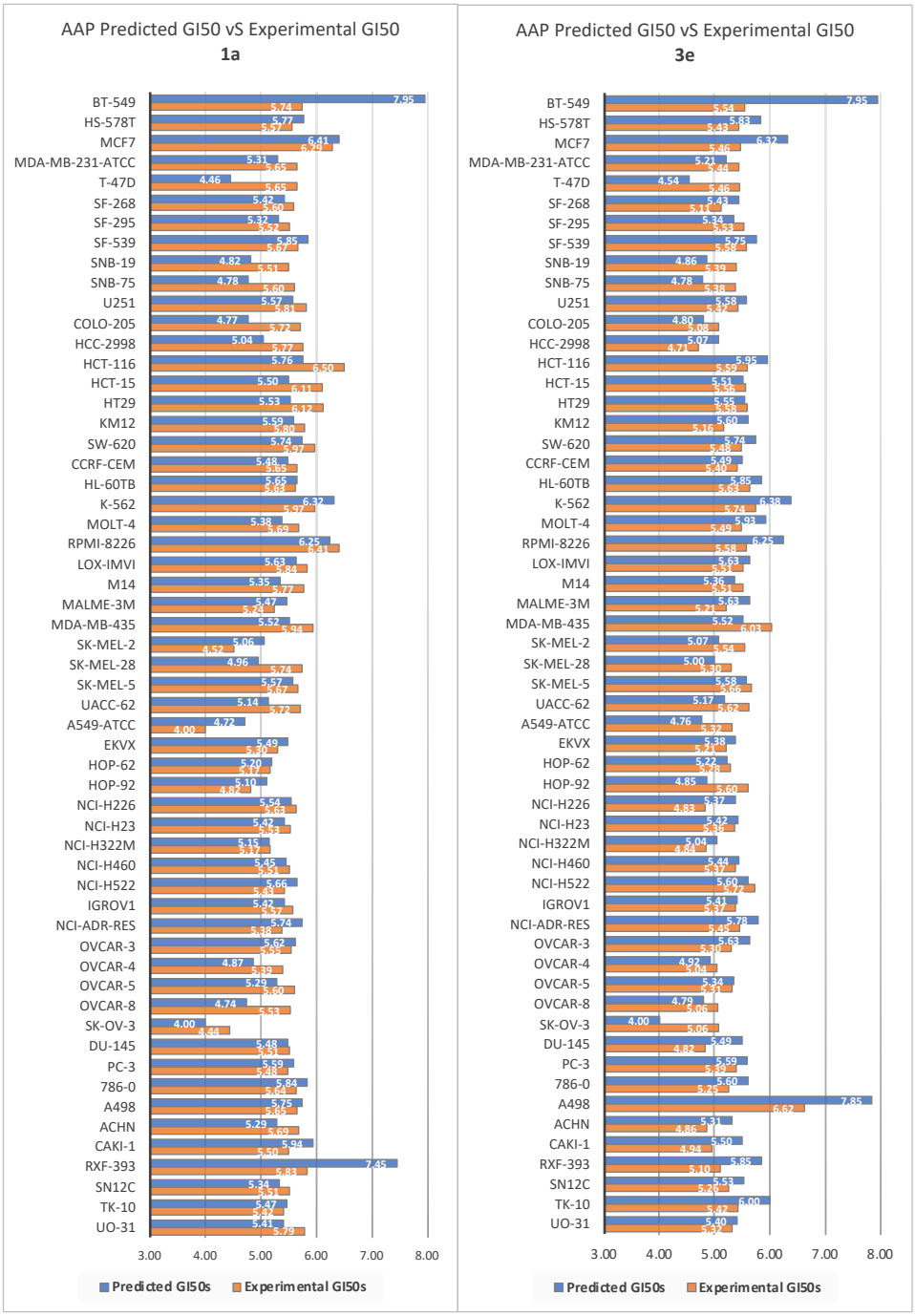
To further analyze the performance of the AAP protocol, the predicted GI_{50} were matched to the experimental ones; moreover, the $|DTV(GI_{50})|$ was computed for the two tested compounds. This allowed to calculate the average absolute error for the compounds, which was of 0.39 and 0.40 for **1a** and **3e**, respectively (see Supplementary Material S14), meaning capability to assign the GI_{50} value with an error of less than one order of magnitude.

In general, considering the mean $|DTV(GI_{50})|$ for each panel, AAP returned very low errors in activity prediction against specific panels, as, for example, the prostate cancer panel for **1a** (average $|DTV(GI_{50})|$ for the panel of 0.07) and colon cancer panel for **3e** (average $|DTV(GI_{50})|$ for the panel of 0.25). Furthermore, a detailed analysis of specific cell lines revealed that the protocol was able to predict GI_{50} against some specific cell lines with quite good precision: for **1a**, the $|DTV(GI_{50})|$ against HL-60TB and NCI-H322M was only 0.02; for **3e**, a $|DTV(GI_{50})|$ of only 0.03 was computed for HT-29 and OVCAR-5.

On the other hand, BT-549 (breast cancer panel) gave the worst predictions for both compounds, with a $|DTV(GI_{50})|$ of 2.21 and 2.41; this evidence is consistent with the results presented in the previous section (tool validation), where this cell line showed the highest error. As previously demonstrated, high prediction error can be attributed to the lack of biological data for selected cell lines. This is confirmed by looking at the $GI_{50}(FP)$ assigned to both compounds: the structure selected as the best score in the FP module was not tested against BT-549, thus the final GI_{50} relied solely on the CL protocol, rather than

combining outputs from both modules, which may have drastically affected the quality of the prediction.

In **Figure 8a** the two bar graphs depicting the comparisons between predicted GI_{50} *versus* experimental GI_{50} are reported to graphically appreciate the excellent performance of our protocol; in **Figure 8b**, instead, the error mean graphs are reported to highlight the cell lines for which the highest/lowest errors were recorded.



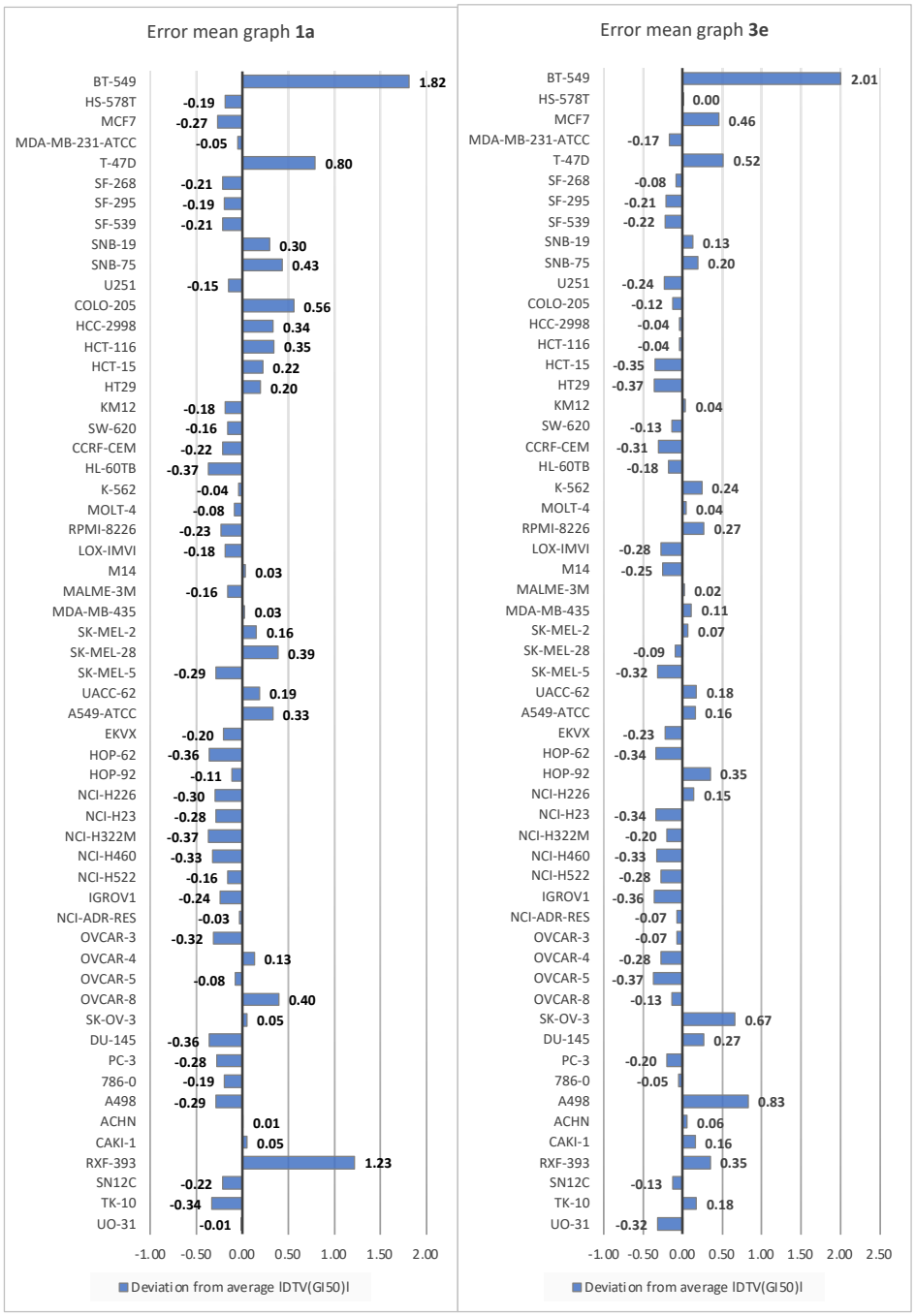


Figure 8. *(**a**) Comparison between AAP predicted GI_{50} values and the corresponding experimental GI_{50} values measured by NCI for the two selected compounds **1a** and **3e** (inside each bar, the corresponding GI_{50} value is indicated); (**b**) Error mean graphs for the two selected compounds **1a** and **3e**.*

To further evaluate the prediction capability of AAP, the pdCSM-cancer tool was selected as a comparative approach [39]. Thus, structures **1a** and **3a** were submitted to pdCSM-cancer and compared with the results obtained by AAP. The obtained data are reported in **Figure 9**. It is noteworthy that the AAP tool showed a $|DTV(GI_{50})|$ of 0.39 and

0.41 for the compounds **1a** and **3a** , while the pdCSM-cancer tool showed 0.60 and 0.63 value, respectively. The cell line that showed a low grade of reliability is BT-549, not only in these cases in the study but also in the external test screening. The change of the dataset could solve the problem.

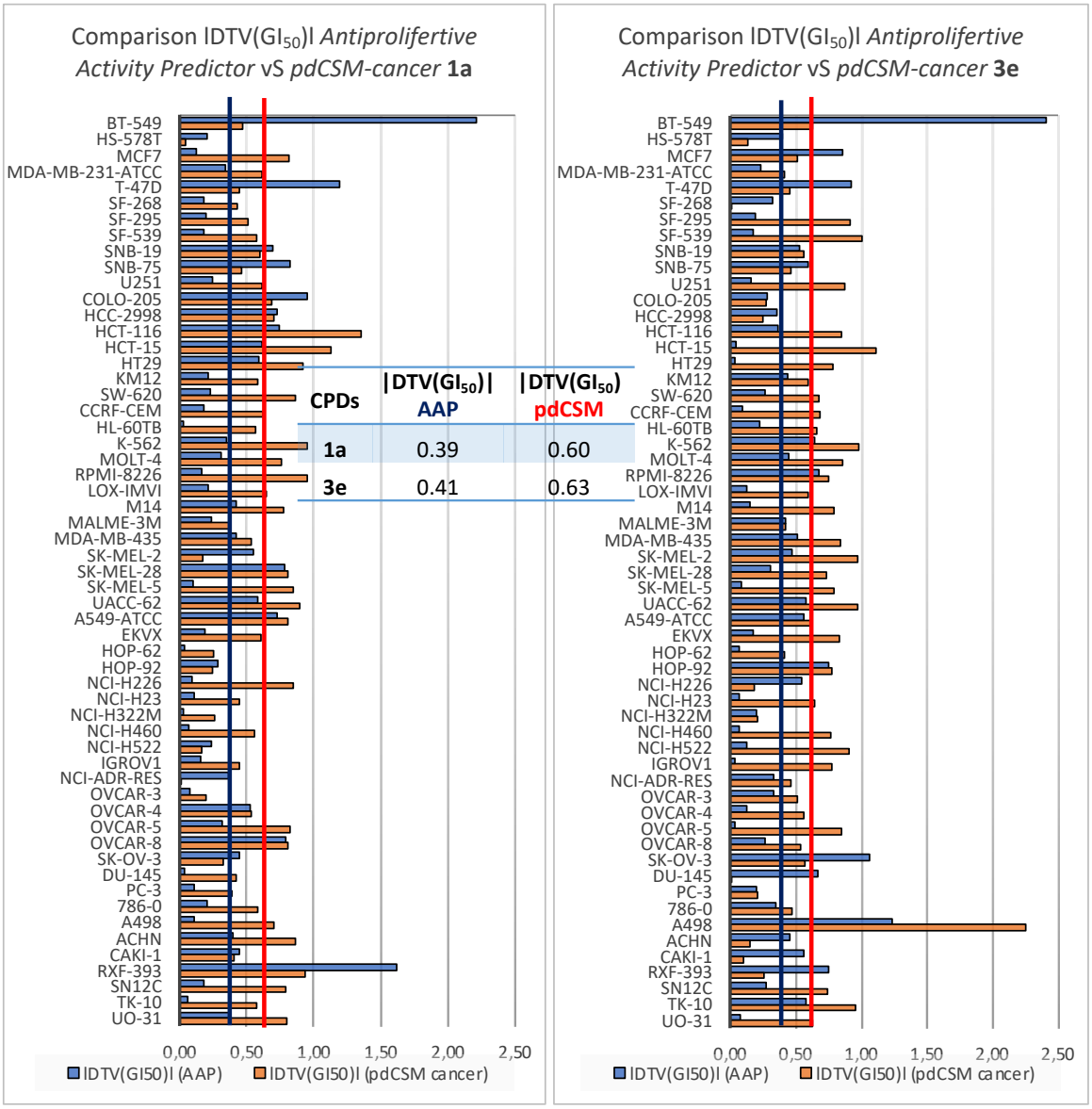


Figure 9. *Performance of AAP versus pdCSM-cancer tools: colored blue and red vertical lines indicate the mean |DTV(GI_{50})| for AAP and pdCSM tools, respectively.*

It is noteworthy the antiproliferative effects of two selected curcumin-like compounds in comparison with curcumin; the protocol, indeed, predicted a higher activity with respect to the parent compound. To this aim, the full NCI output data, which, apart from GI_{50}s, included also TGI and LC_{50} measurements, are reported in **Table 6** (the mean graphs and

the full NCI schedules are reported in Supplementary Material S13). Furthermore, for comparison, the five-dose results of curcumin are also reported.

Firstly, according to the GI_{50} values (the most diagnostic parameter used to compare antiproliferative potential), it emerged that the average GI_{50} is higher for both the tested compounds, in the high micromolar range, compared to curcumin (5.59 for **1a**, 5.37 for **3e** vs 5.16 for curcumin), confirming the *in silico* predictions.

With respect to the tumor subpanels, the most active compound, the dione **1a**, proves to be particularly effective against leukemia, colon, and breast cancer. In fact, the calculated average GI_{50} values for these subpanels (5.87, 6.00, and 5.79, respectively) were always much higher than the overall average GI_{50} (5.59). In details, among these sub-panels, several cell lines showed remarkable sensitivity to the compound, with excellent GI_{50} values in the low micromolar range: RPMI-8226 (6.41), HCT-116 (6.5, the most sensitive cell line), HCT-15 (6.11), HT-29 (6.12) and MCF-7 (6.29). Also, at TGI level it resulted the most active (overall average TGI 4.81), confirming its selectivity against the aforementioned sub-panels and cell lines, especially against colon cancer (average TGI of 5.48). This trend was also confirmed at the LC_{50} level, with high cytotoxic effect against colon cancer cell lines (average GI_{50} of 4.91 for colon cancer). Interestingly, it must be underlined the very low toxicity against RPMI-8226 ($LC_{50} < 4$), which emerged as one of the most sensitive (GI_{50} in the sub-micromolar range) demonstrating the high potency and low toxicity of the compound against this cell line.

The curcumin-like **3e**, although it was less active than the previous one, proved to be more effective than curcumin. Remarkable results were obtained against leukemia sub-panel, with a panel average of 5.57, much higher than the average value for the full NCI-60 (5.37). Moreover, the oxadiazole derivative exhibited high potency against two cell lines, MDA-MB-435 (melanoma) and A498 (renal cancer), with excellent GI_{50} values in the low micromolar range (6.03 and 6.61, respectively). Regarding TGI level, it resulted slightly less potent than curcumin, but the average LC_{50} of 4.01, also against the most susceptible cells, suggests high potency with low cytotoxicity, even at high concentrations.

Table 6. *DTP NCI five-dose screening for compounds* ***1a*** *and* ***3e****.*

PANEL	CELL LINE	1a			3e			Curcumin		
		GI_{50}	TGI	LC_{50}	GI_{50}	TGI	LC_{50}	GI_{50}	TGI	LC_{50}
Leukemia	CCRF-CEM	5.65	4.7	4	5.4	4	4	5.52	4.81	4
	HL-60(TB)	5.63	4.77	4	5.63	5.07	4	5.14	4.60	4.04
	K-562	5.97	4	4	5.74	4	4	5.51	4.26	4
	MOLT-4	5.69	4.72	4	5.49	4	4	5.33	4.75	4.12
	RPMI-8226	6.41	5.63	4	5.58	4	4	5.68	5.20	4
	Panel average	**5.87**	**4.76**	**4**	**5.57**	**4.21**	**4**	**5.43**	**4.72**	**4.03**
Non-Small Cell Lung Cancer	A549/ATCC	4	4	4	5.32	4	4	4.89	4.50	4.11
	EKVX	5.3	4	4	5.21	4	4	4.82	4.45	4.10
	HOP-62	5.17	4	4	5.28	4	4	5.44	4.72	4.24
	HOP-92	4.82	4.08	4	5.6	4.63	4	NT	NT	NT
	NCI-H226	5.63	NT[1]	4	4.83	4	4	4.73	4.27	4
	NCI-H23	5.52	4	4	5.36	4	4	5.25	4.50	4
	NCI-H322M	5.17	4	4	4.84	4	4	4.78	4.49	4.21
	NCI-H460	5.51	4.95	4	5.37	4	4	5.09	4.64	4.22
	NCI-H522	5.43	4.72	4	5.72	5.17	4.02	5.27	4.78	4.07
	Panel average	**5.17**	**4.22**	**4.00**	**5.28**	**4.20**	**4**	**5.03**	**4.54**	**4.12**
Colon Cancer	COLO-205	5.72	5.31	4.4	5.08	4.01	4	4.87	4.54	4.21
	HCC-2998	5.77	5.49	5.22	4.71	4	4	5.52	5.09	4.53
	HCT-116	6.5	5.88	5.39	5.59	4.75	4	5.53	5.03	4.28

	HCT-15	6.11	5.21	4.25	5.56	4	4	5.39	4.73	4.14
	HT-29	6.12	5.58	5.09	5.58	4.97	4	5.29	4.49	4
	KM12	5.8	5.43	5.07	5.16	4	4	5.27	4.71	4.19
	SW-620	5.97	5.48	4.97	5.48	4	4	5.38	4.67	4.07
	Panel average	**6.00**	**5.48**	**4.91**	**5.31**	**4.25**	**4**	**5.32**	**4.75**	**4.20**
CNS Cancer	SF-268	5.6	5.07	4	5.11	4	4	5.15	4.44	4
	SF-295	5.52	4.51	4	5.53	4.75	4	5.10	4.68	4.32
	SF-539	5.67	5.29	4.22	5.57	5.03	4.09	5.55	5.05	4.48
	SNB-19	5.51	4.63	4	5.39	4.04	4	5.05	4.61	4.20
	SNB-75	5.6	4.47	4	5.38	4.34	4	5.17	4.74	4.35
	U251	5.81	5.44	5.07	5.42	4.73	4	5.33	4.78	4.31
	Panel average	**5.62**	**4.90**	**4.22**	**5.40**	**4.48**	**4.02**	**5.22**	**4.72**	**4.28**
Melanoma	LOX IMVI	5.84	5.49	5.15	5.51	4.54	4	5.57	5.07	4
	MALME-3M	5.24	4.1	4	5.21	4	4	4.85	4.56	4.27
	M14	5.77	5.36	4.24	5.51	4.58	4	5.42	4.80	4.35
	MDA-MB-435	5.94	5.53	5.11	6.03	5.41	4.19	5.53	4.92	4.40
	SK-MEL-2	4.51	4	4	5.54	4.73	4	4.78	4.39	4.06
	SK-MEL-28	5.74	5.4	NT	5.3	4	4	5.35	4.80	4.30
	SK-MEL-5	5.67	5.21	4	5.66	4.99	4	5.06	4.65	4.28
	UACC-257	5.61	5.13	4	4.97	4	4	4.94	4.62	4.31
	UACC-62	5.72	5.31	4.52	5.62	5	4	5.19	4.69	4.26
	Panel average	**5.56**	**5.06**	**4.38**	**5.48**	**4.58**	**4.02**	**5.19**	**4.72**	**4.25**

Ovarian Cancer	IGROV-1	5.57	NT	4	5.37	4	4	5.10	4.57	4.09
	OVCAR-3	5.55	5	4	5.3	4.09	4	5.18	4.61	4.17
	OVCAR-4	5.39	4	4	5.04	4	4	5.03	4.44	4
	OVCAR-5	5.6	5.09	4	5.31	4.27	4	4.78	4.45	4.12
	OVCAR-8	5.53	4	4	5.06	4	4	5.13	4.55	4.08
	NCI/ADR-RES	5.38	4	4	5.45	4	4	5.14	4.12	4
	SK-OV-3	4.44	4	4	5.06	4.03	4	5.05	4.68	4.33
	Panel average	**5.35**	**4.35**	**4.00**	**5.23**	**4.06**	**4**	**5.06**	**4.49**	**4.11**
Renal Cancer	786-0	5.64	5.1	4	5.25	4	4	5.48	4.97	4.42
	A-498	5.65	4.7	4	6.62	5.04	4	4.80	4.48	4.16
	ACHN	5.69	5.27	4	4.86	4	4	4.91	4.54	4.17
	CAKI-1	5.5	4	4	4.94	4	4	4.92	4.60	4.30
	RXF-393	5.83	5.52	5.2	5.1	4.11	4	5.52	4.95	4.27
	SN12C	5.51	4.75	4	5.26	4	4	5.08	4.60	4.20
	TK-10	5.42	4.6	4	5.42	4.56	4	4.85	4.51	4.18
	UO-31	5.79	5.47	5.14	5.32	4	4	4.95	4.61	4.27
	Panel average	**5.63**	**4.93**	**4.29**	**5.35**	**4.21**	**4**	**5.06**	**4.66**	**4.25**
Prostate Cancer	PC-3	5.48	4	4	5.39	4	4	5.06	4.59	4.15
	DU-145	5.51	4.85	4	4.82	4	4	4.81	4.53	4.25
	Panel average	**5.50**	**4.43**	**4.00**	**5.11**	**4.00**	**4**	**4.93**	**4.56**	**4.20**
Breast Cancer	MCF7	6.29	5.02	4	5.46	4	4	5.48	4.46	4
	MDA-MB-231/ATCC	5.65	5.22	4	5.44	4.36	4	4.75	4.25	4

HS 578T	5.57	4	4	5.43	4.32	4	4.96	4.23	4
BT-549	5.74	5.35	4.66	5.54	4.45	4	5.30	4.86	4.37
T-47D	5.65	4	4	5.46	4.09	4	5.08	4.33	4
MDA-MB-468	5.81	5.43	4	5.54	4.56	4	NT	NT	NT
Panel average	**5.79**	**4.84**	**4.11**	**5.48**	**4.30**	**4**	**5.11**	**4.43**	**4.07**
Overall average	5.59	4.81	4.24	5.37	4.28	4.01	5.16	4.63	4.17
Range	4-6,5	4-5.9	4-5.4	4.7-6.6	4-5.4	4-4.2	4.7-5.9	4-5.20	4-4.53

[1] NT= not tested against the cell line.

3.1.8. The AAP protocol: a tool implemented in the open-acces DRUDITonline web-platform

The results presented in the previous paragraphs proved the extremely reliability of the developed AAP protocol. In order to provide free access to the scientific community, it has been included as a tool in DRUDIT^{online}, a web-service (accessible at https://www.drudit.com, **Figure 10a**) created by the research group to support the research in medicinal chemistry field, which includes the already developed Biotarget finder tool [40-42].

In **Figure 10b** the "tools" page of the website is shown, and panel corresponding to the AAP is highlighted. The easy-to-use interface permit to upload or draw the chemical structures directly on the platform and to select, by an ON/OFF button, the required tools (Biotarget finder or AAP, or both). By a dedicated window, it is possible to vary and select the desired combination of parameters (N, Z and G) to direct the calculation accordingly. Once the calculations have been completed, the output matrices with the results (e.g.: predicted GI_{50} if one considers the AAP tool) can be downloaded from the "Results" page as a .csv file.

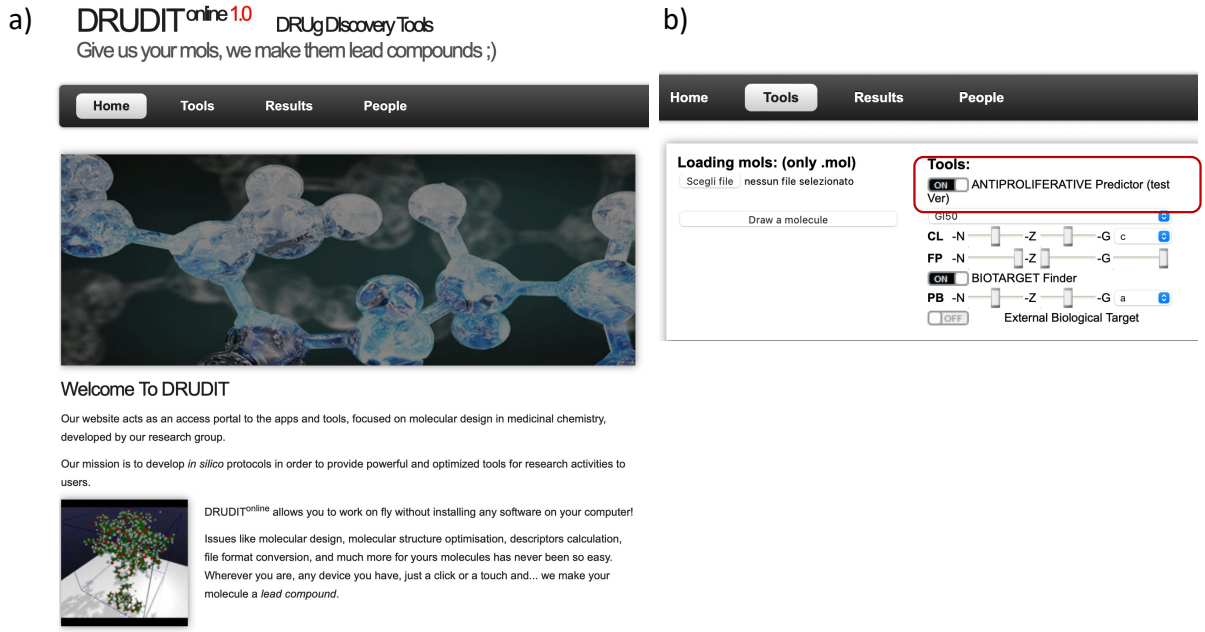


Figure 10. *(**a**) Homepage of the open-access web-service DRUDIT^{online}; (**b**) "Tools" page in DRUDIT^{online} web-platform.*

3.2. A new *in silico* method for the correlation between cell line chemosensitivity and gene/protein expression patterns: application to known targeted anticancer agents

After the development of the AAP tool to predict the antiproliferative activity of new compounds, the second phase of this PhD project aimed to develop a new and quick in-house computational method to correlate drug activity and protein expression pattern. Before its application to the design of new anticancer agents, to demonstrate that such a correlation analysis could be applied to the biological data related to the NCI-60 panel, the proposed method needed to be validated. In this light, a correlation analysis between the antiproliferative activity of several anticancer compounds tested by NCI and the expression of the corresponding target in the NCI-60 panel was performed.

The proposed method and the case studies are described in detail in the next paragraph.

3.2.1 Method to correlate drug response and protein expression across NCI-60 panel: mathematical rationale and case studies

Given the impossibility of comparing the NCI antiproliferative and protein/gene expression data, because of the various experimental conditions applied, a protocol that allowed their comparison and consequent evaluation through normalization of all available data was established. Briefly, for each target, the experiments of protein/gene expression were downloaded from NCI database, normalized, and compared with the normalized GI_{50} values of the corresponding structure tested by NCI using the *in silico* procedure explained in **Figure 11**.

The proposed normalization process consists of several steps: for each sequence of biological data, GI_{50}s (G) and Protein/gene expressions (EP), the mean value μ is estimated and the deviation from this is calculated for all cell lines χ. The values are then normalized (NGI_{50} and NEP). To match the sequences of the normalized data, the signs of the values are considered. For each cell line, the score $\delta_i = 1$ is assigned when the signs are concord,

otherwise $\delta_i = 0$. The sum of these values divided by the number of cell lines gives the final score.

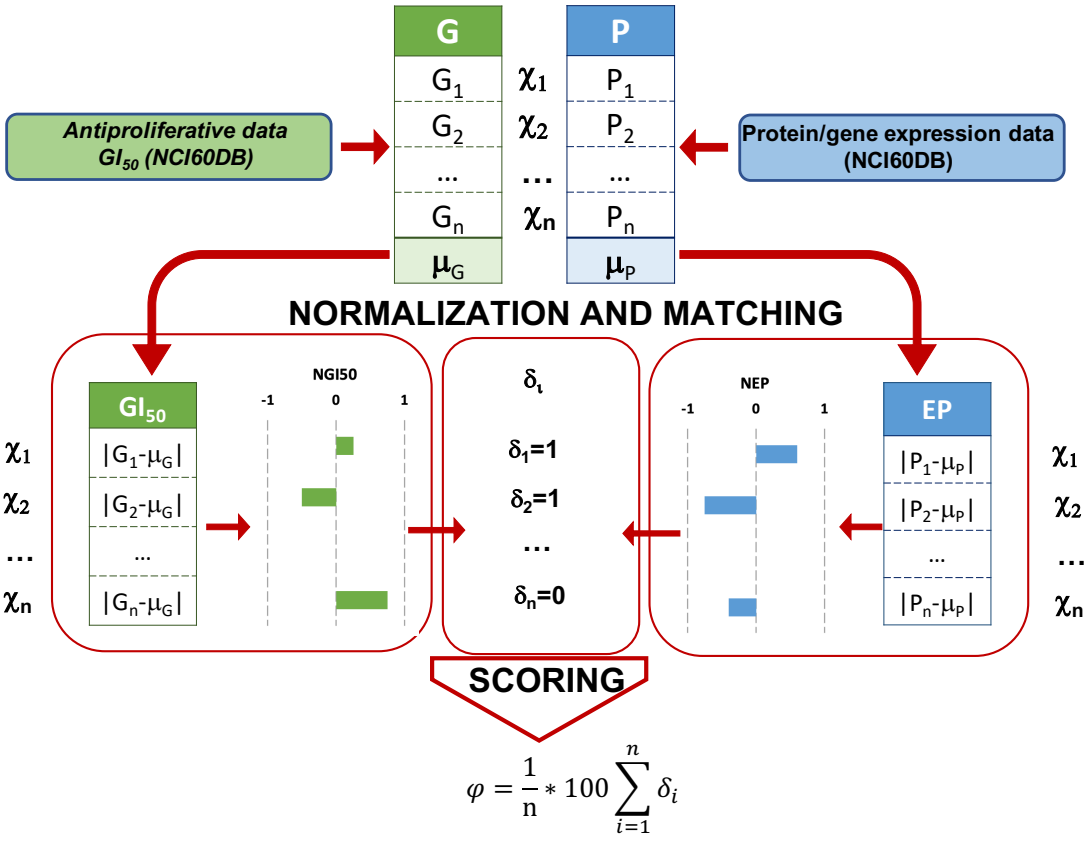


Figure 11. *Workflow for the assessment of the correlation between GI_{50} and protein expression.*

The output data were the percentage of matchings recovered for each structure in relation to the expression of the corresponding target/s in cancer cells.

Several cases studies of known targeted drugs tested by NCI were analyzed to confirm the capability of the proposed mathematical approach to match these different types of biological data sequences. The chemical structures of drugs analyzed in the case studies are reported in **Figure 12**, whereas the overall results of the performed correlation analysis are included in Supplementary Material S15.

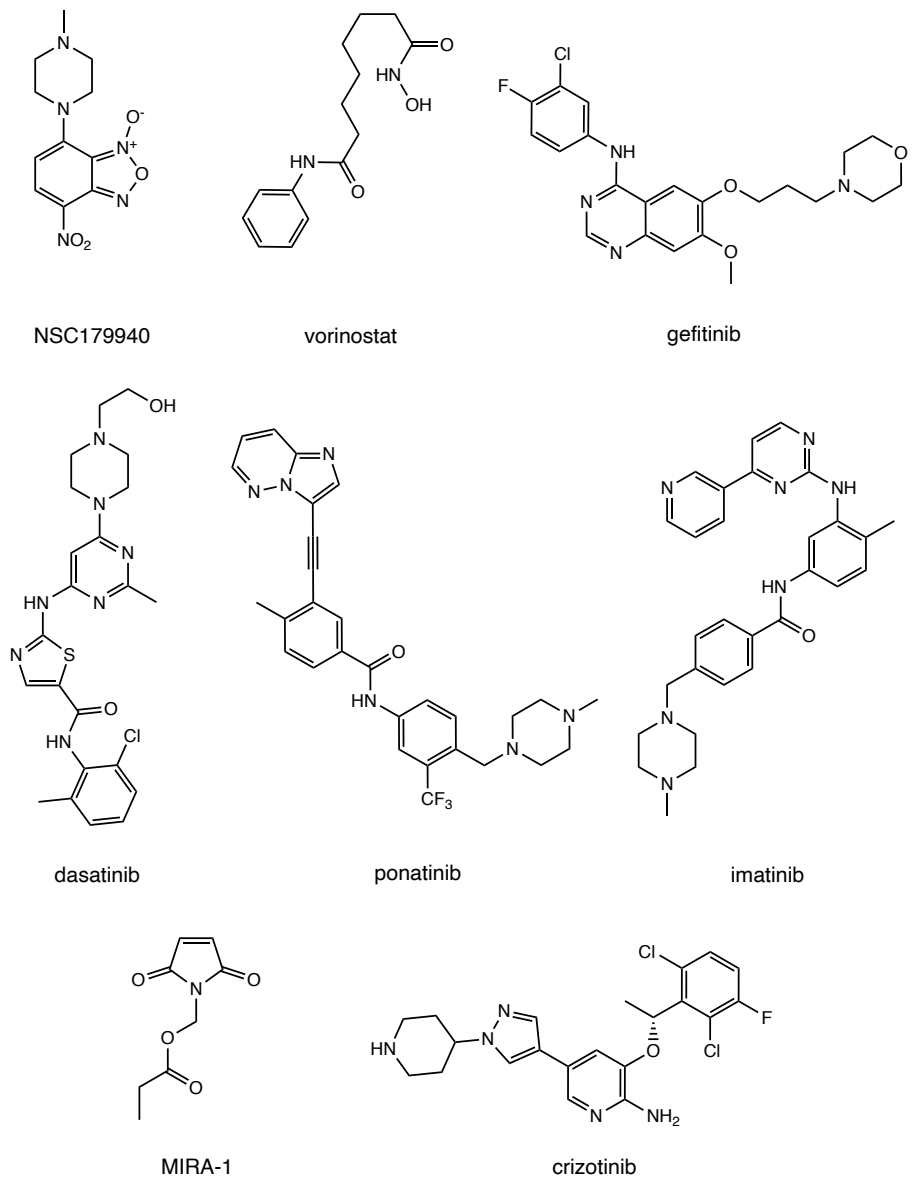


Figure 12. *Examples of targeted drugs the antiproliferative activity of which (expressed as GI_{50}) against National Cancer Institute NCI-60 is highly correlated with the expression of the target protein within the same cells.*

The first case study is focused on NSC179940, a benzofuroxan derivative that selectively modulates the MDM4 protein [59], a p53-interacting protein that is frequently upregulated in cancer cells, to prevent p53 tumor-suppressor function [60].

The high expression of MDM4 is crucial to MCF7 cell growth. The MDM4 gene silencing decreases cell proliferation [61,62]. By using the above-mentioned protocol, the percentage of matching score was 72% between the NCI-60 GI_{50} values and the corresponding MDM4 expression (according to the model, 42 out of 58 GI_{50} values matched the MDM4 expression

values in the corresponding cancer cell lines after administration of NSC179940). **Figure 13a** shows the normalized mean graph describing this correlation. Given their specific proteomic profile, the MCF7, CCRF-CEM, and HCT-116 cell lines are representative of a high correlation between antiproliferative activity and MDM4 expression.

Histone acetyl transferase (HAT) and histone deacetylase (HDAC) both modulate gene expression by controlling the pattern of histone acetylation. Several biological studies have shown that the balance between HAT and HDAC is altered in different human cancers [63]. In contrast, it has also been shown that inhibition of HDAC causes growth arrest, differentiation, and apoptosis of tumor cells, rendering HDACs promising targets for cancer therapy [64].

Vorinostat (NSC701852), a US Food and Drug Administration (FDA)-approved HDAC4 inhibitor [65], showed a significant percentage of matching between HDAC4 expression and GI_{50} values (72%). In detail, optimal data were obtained for the CCRF-CEM and MOLT-4 cell lines, belonging to the leukemia NCI subpanel (see the normalized mean graph in **Figure 13b**). The *in silico* results are consistent with the experimental evidence reported in the literature. In particular, it has been demonstrated that, in both cell lines, increased HDAC4 expression is crucial for uncontrolled cell proliferation, while small interfering (si)RNA-mediated knocking down of HDAC4 expression or HDAC inhibition leads to growth inhibition and apoptosis [66,67], with the intensification of cell sensitivity to cytotoxic drugs [68].

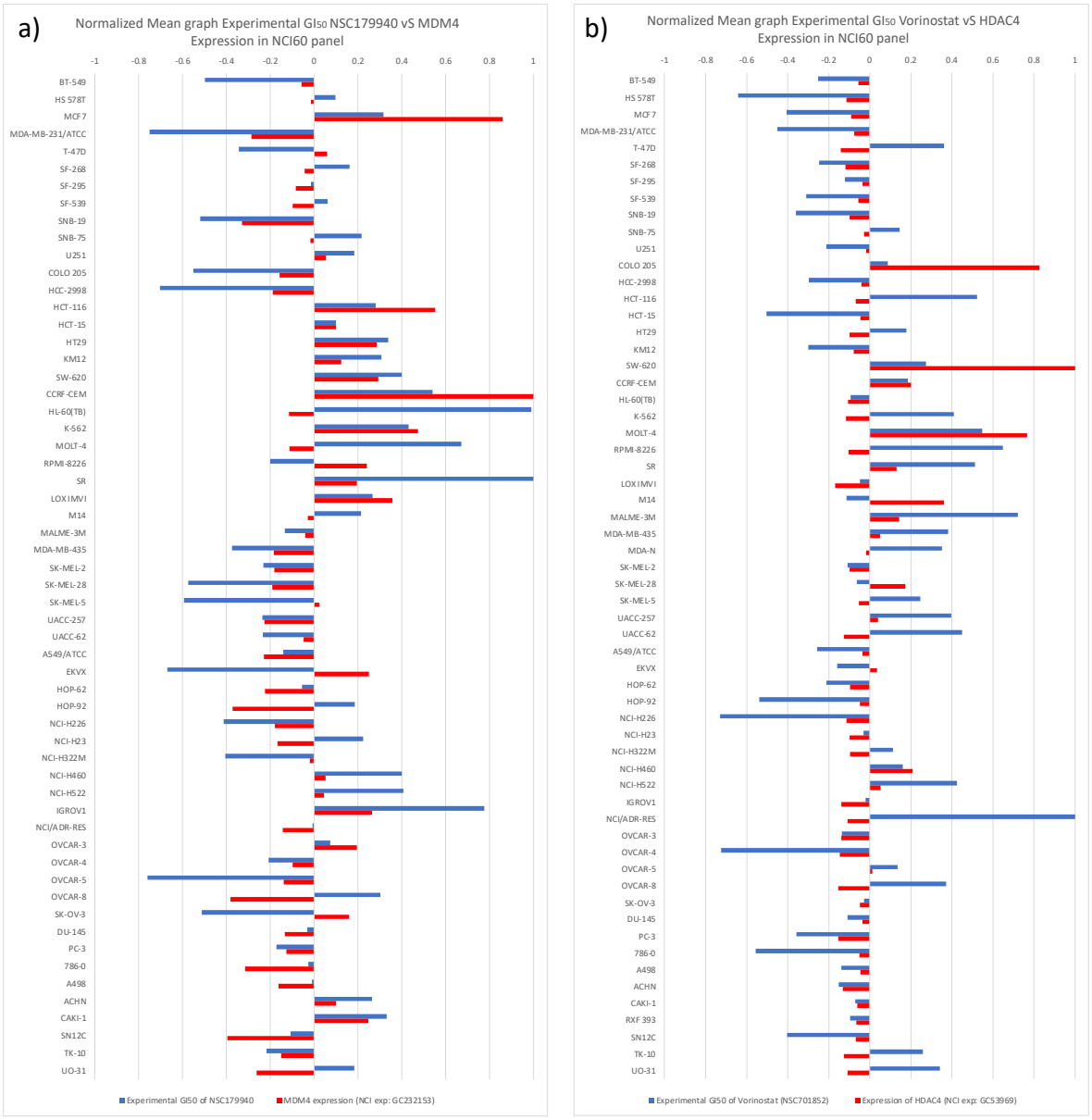


Figure 13. *(**a**) Mean graph representing the trend of normalized experimental GI_{50} of the MDM4 inhibitor NSC179940, available in the National Cancer Institute (NCI) database, compared with MDM4 expression in NCI-60. The high correlation between the two sets of biological data is evident if the MCF7, CCRF-CEM, and HCT-116 cell lines are considered, in which MDM4 has a pivotal role in cancer cell proliferation. (**b**) Mean graph representing the trend of normalized experimental GI_{50} of the HDAC4 inhibitor vorinostat, available in the National Cancer Institute (NCI) database, compared with HDAC4 expression in NCI-60. The high correlation between the two sets of biological data is evident if the CCRF-CEM and MOLT-4 cell lines are considered, in which HDAC4 has a pivotal role in cancer cell proliferation.*

Tyrosine kinase receptors are key proteins in signal transduction pathways, controlling cell proliferation, differentiation, metabolism, and apoptosis. Chronic activation of tyrosine kinase receptors is observed in most human cancers. Their inhibition severely impairs the proliferation and survival of tumor cells [69]. Among the tyrosine kinase receptors,

epidermal growth factor receptor (EGFR) is one of the most important targets in anticancer therapy. The roles of the EGFR signaling pathway in various cancers have been investigated since the 1980s and multiple inhibitors have been developed, many of which are in clinical use. Gefitinib (NSC759856) was the first EGFR inhibitor approved for clinical use [70]. As shown in the mean graph of the correlation of GI_{50}/EGFR expression after treatment with gefitinib (see **Supplementary Figure 1**), most of the GI_{50} values matched the related normalized values of EGFR expression in the NCI-60 panel (percentage matching = 73%). High correspondence was obtained against the A498 and ACHN renal cancer cell lines, in accordance with the crucial role of EGFR in the survival of both cell lines [71].

Dasatinib (NSC759877) is a multityrosine kinase inhibitor, active on both the tyrosine kinase nonreceptors Bcr-Abl, c-Src, and the tyrosine kinase receptors c-KIT and PDGFR-b. It was approved by the FDA in 2006 for the treatment of adults with chronic myeloid leukemia (CML) and Philadelphia chromosome-positive acute lymphoblastic leukemia (ALL) [72]. A good correlation has been obtained with Abl and c-Src expression (matching of 70% and 71%, respectively, the normalized mean graph is shown in the **Supplementary Figure 2**). The notable antiproliferative effect of dasatinib against K-562, a leukemia cell line positive for the Philadelphia chromosome [73,74], results from the key role of Abl kinase in cancer progression: numerous studies demonstrated that inhibition of Abl expression determined cell growth inhibition and apoptosis [75,76]. By contrast, the predicted antiproliferative activity of dasatinib against 786-0 and A498 renal cancer cell lines is correlated with the inhibition of c-Src transcription and the related mRNA translation by miRNA, which suppresses cell growth [77].

For several other compounds, a significant correlation has been obtained for particular cell lines, especially those in which the target protein has a pivotal role. Two examples are ponatinib (NSC758487) and imatinib (NSC743414), two Bcr-Abl inhibitors approved by the FDA for the therapy of CML and ALL [78]. A high correlation between GI_{50} and Abl expression was observed especially for the K-562 cell line (see **Supplementary Figure 3**, **Supplementary Figure 4**). Other well-correlated results have been obtained for MIRA-1 (NSC19630), a mutant p53-reactivating small molecule [79], especially for the CCRF-CEM cell lines, in which

p53 is mutated (NCI experiment for p53 mutation status DNA: MT171, MT2924) (**Supplementary Figure 5**) [80,81].

Similar results were observed for crizotinib, a small-molecule ALK inhibitor approved by the FDA [82]. The strong correlation between experimental GI_{50} and ALK expression in the SR leukemic cell line is supported by evidence of the involvement of ALK in SR cancer cell survival and growth (**Supplementary Figure 6**, GC28955) [83].

3.3. *In silico* identification of small molecules as new Cdc25 inhibitors through the correlation between chemosensitivity and protein expression pattern

The final part of this study aimed to apply the proposed new approach based on the correlation between chemosensitivity and protein expression data to screen a database of small molecules and identify new anticancer agents with optimal and selective inhibitory activity against a target involved in carcinogenic processes. In this light, the integration of a tool able to predict in advance the anticancer activity of new compounds, the AAP, gave us the opportunity to select the structures whose anticancer activity was optimally correlated with the expression of the desired target already in the first phases of drug design, before the *in vitro* antiproliferative assays. To reinforce the reliability and the robustness of the proposed *in silico* protocol, the new developed tools were integrated with other well-established ligand-based (as the Biotarget finder tool, developed by the research group [42]) and structure-based techniques (such as Induced Fit Docking).

The first application of this innovative approach herein reported was focused on the design new potential inhibitors of Cdc25, a phosphatase involved in the regulation of cell cycle, thus in cell growth, as described in detail in the following paragraph. The full computational protocol will be described, highlighting the advantages of such an approach over other conventional virtual screenings.

Finally, the biological results of the best molecules identified *in silico* are reported (in collaboration with the biochemistry research group supervised by Prof. Carla Gentile – University of Palermo), permitting to verify the reliability of the full protocol herein proposed.

3.3.1. Cdc25, a brief overview of its role in cancer

The cell division cycle 25 (Cdc25) protein family, firstly identified in yeasts as a mitotic inducer [84], comprises three members (Cdc25A, B, and C) of dual-specificity protein phosphatases [85]. In mammalian cells, these proteins are involved in the activation of cyclin-dependent kinases 1 and 2 (Cdk1-2), through the dephosphorylation of specific

threonine and tyrosine residues located in the ATP-binding loop (Tyr^{15} and Thr^{14}). All three forms cooperate to regulate cell division in human cells [86]. Particularly, Cdc25A controls both early and late cell-cycle transitions (G1-S, S, and G2-M) by acting on CyclinE–Cdk2, CyclinA–Cdk2 and CyclinB–Cdk1, meanwhile Cdc25B and Cdc25C promote mitosis activating CyclinB–Cdk1 and CyclinA-Cdk2 (**Figure 14**) [87-89]. Following DNA damage and environmental stresses, Cdc25 proteins are phosphorylated and inactivated by Checkpoint 1–2 (Chk1-2) and MAPKAPK-2 kinases, leading to their export outside the nucleus, and consequent cell-cycle arrest [90,91].

Considering their role in the activation of cyclin-Cdk complexes and, therefore, in cell-cycle progression, Cdc25s have become interesting targets in the search for anticancer drugs. In particular, due to their role in the activation of cyclin-Cdk complexes downstream the G1 phase and involving cyclin A, cyclin M, and cyclin B, interference with Cdc25s activity could effectively block tumor cell proliferation. Indeed, although there are growth factor-dependent and nutrient-dependent checkpoints in the G1 phase, the vast majority of mutations that contribute to cell transformation involve genes regulating G1-progression. Those mutations make tumor cells autonomous from growth factor stimulation but, through the dysregulation of the cellular metabolism, also able to override nutritional sensing [92,93]. Consequently, while in the absence of growth factor instructions and nutrients cells commonly arrest at the G1 phase and undergo apoptosis, tumor cells are able to overcome all G1-checkpoints and progress along the G1 phase. In particular, in cancer cells, two signaling pathways activate G1 progression. The first involves *ras* mutations that, in a growth factor independent way, activate MAP kinase pathways increasing cyclin D expression and G1 complexes (cyclin D-Cdk4 and cyclin D-Cdk6) activation. The second is the mTOR pathway, which is very sensitive to the presence of the energy and nutrients required for activation of cyclin E complexes [94].

The possibility of intervention downstream of G1 checkpoints, through the inhibition of cyclin E, A, and B partners, strengthens Cdc25 inhibition as an anticancer strategy. From a structural point of view, human Cdc25A, B, and C include 524, 560, and 473 amino acids, respectively [95-97]. All three proteins comprise two main regions: the N-terminal region,

which is extremely variable and acts as a regulatory domain (as the site of phosphorylation and ubiquitination, or the sequencing of nuclear localization and exportation), and the C-terminal region, which is extremely homologous and contains the catalytic site [98,99]. The catalytic domain includes the HCX_5R motif, characteristic of tyrosine phosphatase and composed of a highly conserved histidine; a catalytic cysteine (namely Cys^{430}, Cys^{473}, and Cys^{377} in Cdc25A, Cdc25B, and Cdc25C, respectively); five residues (X_5), whose amide groups form hydrogen bonds with phosphate residues; and a conserved arginine, required for binding to a phosphorylated amino acid of the substrate [96,97,100-102].

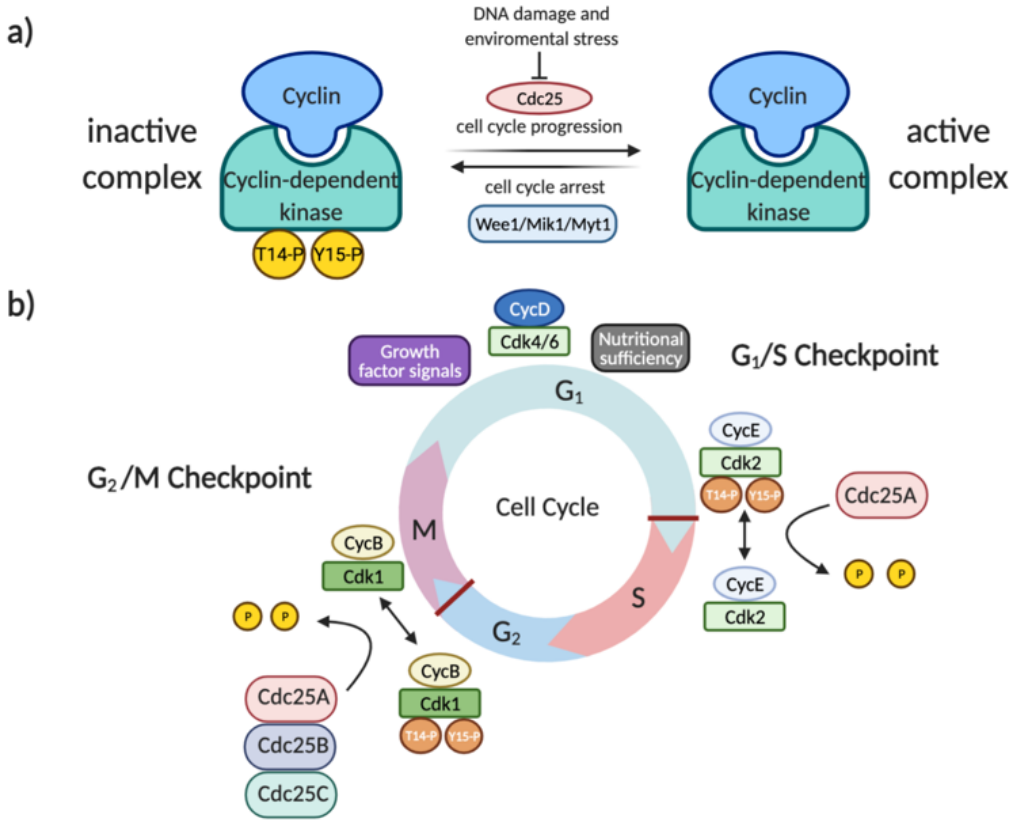


Figure 14. *(**a**) Inactivation of cyclin-dependent kinases (Cdks) by the Wee1/Mik1/Myt1 protein kinase family through the phosphorylation of T14 and Y15. (**b**) The promotion of the entrance of cell division cycle 25 A (Cdc25A) in the S-phase cell cycle through the activation of the Cdk2/CycE complex (on the right); the promotion of mitosis by Cdc25A-B-C through the activation of the Cdk1/CycB complex (on the left).*

The analysis of the crystal structures of the catalytic domains of Cdc25A and Cdc25B (**Figure 15**, panel (**a**) and (**b**); Protein Data Bank (PDB) id: 1C25 and 1QB0, respectively) shows that the active sites appear flat and shallow, in contrast to other phosphatases [103,104].

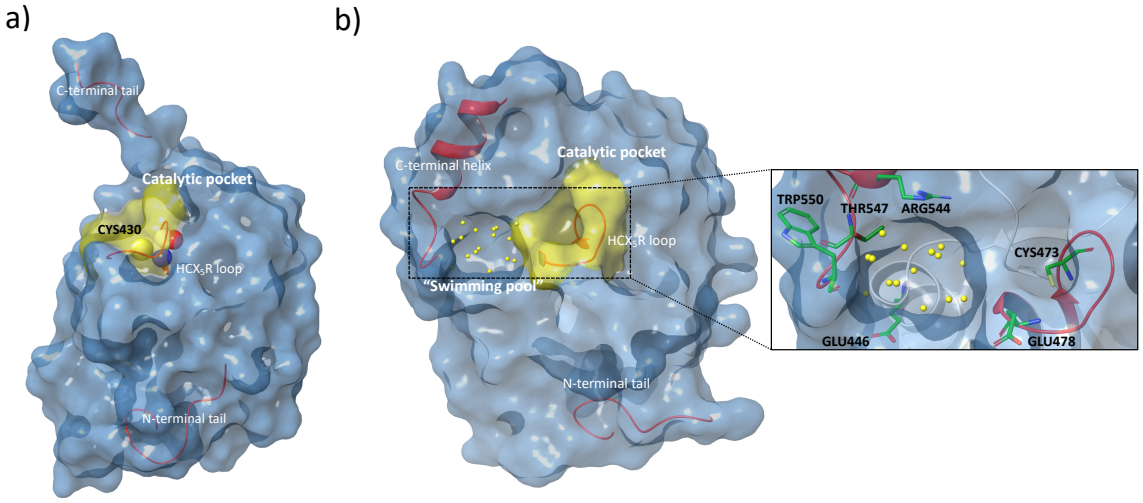


Figure 15. *(**a**) Surface view of the Cdc25A crystal structure (PDB id: 1C25) [103] with the catalytic Cys^{430} in the HCX_5R loop (CPK representation) and the C/N terminal tails highlighted. (**b**) Surface view of the Cdc25B crystal structure (PDB id: 1QB0) [104] with the catalytic HCX_5R loop, the water molecules of the "swimming pool" region (yellow dots), the C-terminal helix, and the N-terminal tail highlighted. On the right, special focus is given to several of the most important residues (thick tube representation) within the catalytic pocket and the "swimming pool" involved in the catalytic process and the interactions with ligands.*

However, a well-ordered C-terminal helix adjacent to the catalytic pocket in the structure of Cdc25B contributes to the formation of the so-called "swimming pool", an extended and deep protein-sequence occupied by a significant amount of water molecules. This region contains several key residues that, in collaboration with those present in the catalytic domain, participate in the stabilization of protein-inhibitor complexes (**Figure 15**, panel (b)) [98,105].

As reported by Lavecchia et al. [106,107], several molecules have been developed as selective inhibitors of Cdc25s. The most studied classes are quinonoids, phosphate surrogates, and electrophilic entities [108]. In particular, NSC663284 and BN82685 (**Figure 16**), belonging to the quinonoid class, showed a remarkable Cdc25 inhibition activity, with IC_{50} values in the nanomolar range [109,110]. For many years, NSC663284 (**Figure 16**) has been used as a lead compound for the design of new Cdc25 inhibitors, and its mechanism of action has been extensively investigated [111,112]. In 2017, Ge et al. identified by *in silico* analysis the "swimming pool" region as the potential binding site of NSC663284 in the Cdc25B phosphatase [113].

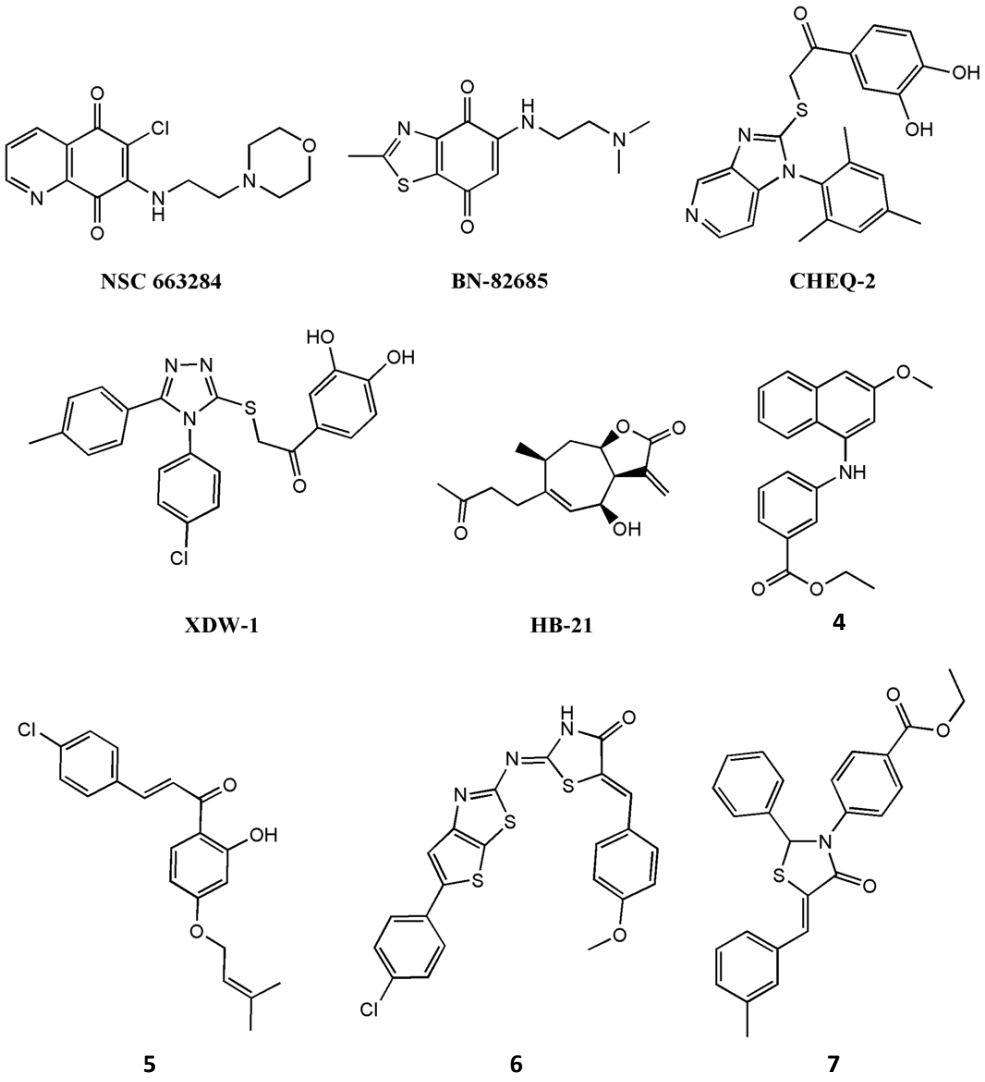


Figure 16. *The chemical structure of some well-known Cdc25 inhibitors.*

Moreover, Tao et al., in analyzing Cdc25s catalytic domains and pharmacophoric moieties [114], reviewed more suitable molecules (imidazopyridine CHEQ-2 [115], 1,2,4-triazole XDW-1 [116], sesquiterpene HB-21 [117], naphthyl-phenylamine (**Figure 16**, molecule **4**) [118,119], chalcone (**Figure 16**, molecule **5**) [120], 1,3-thiazolidin-4-ones (**Figure 16**, molecules **6** and **7**) [121,122]), with interesting inhibitory activity on Cdc25 enzymes.

Due to the important role of Cdc25 as a checkpoint component of the cell cycle, the deregulation of its proteins at transcriptional, translational, and post translational levels can cooperate with oncogenic transformation and the progression of disease [108], especially in breast, ovarian, colorectal, esophageal, gastric, lung head, and neck cancer cells [123-130]. Furthermore, the overexpression of Cdc25A and Cdc25B is frequently linked with poor

clinical prognosis [98,131]. For these reasons, Cdc25s can be considered attractive targets from the development of specific inhibitors for targeted therapeutic treatment [108,132].

3.3.2. *In silico* protocol applied to screen the small molecules database

The *in silico* mixed ligand-structure based protocol herein applied to design new selective Cdc25 inhibitors involved several sequential steps, as shown in the workflow reported in **Figure 17**:

- Ligand-based studies:
    - First screening of a large database of small molecules against the Cdc25 template implemented in the *Biotarget finder*, a tool of the DRUDIT^{online} platform that can estimate the biological affinity for input structures against biological target/s;
    - Application of the newly developed AAP tool to predict the antiproliferative activity (GI_{50}) of the best inhibitors;
    - Application of the proposed approach based on the correlation between predicted drug activity and expression of target protein (Cdc25) across the NCI-60 panel.
- Structure-based studies:
    - Induced-Fit Docking (IFD): the best ranked compounds selected through the ligand-based phase were then further examined by means of IFD simulations, to verify their capability to fit well into the binding pocket of the target of interest.

The most interesting compounds were then submitted for *in vitro* assays.

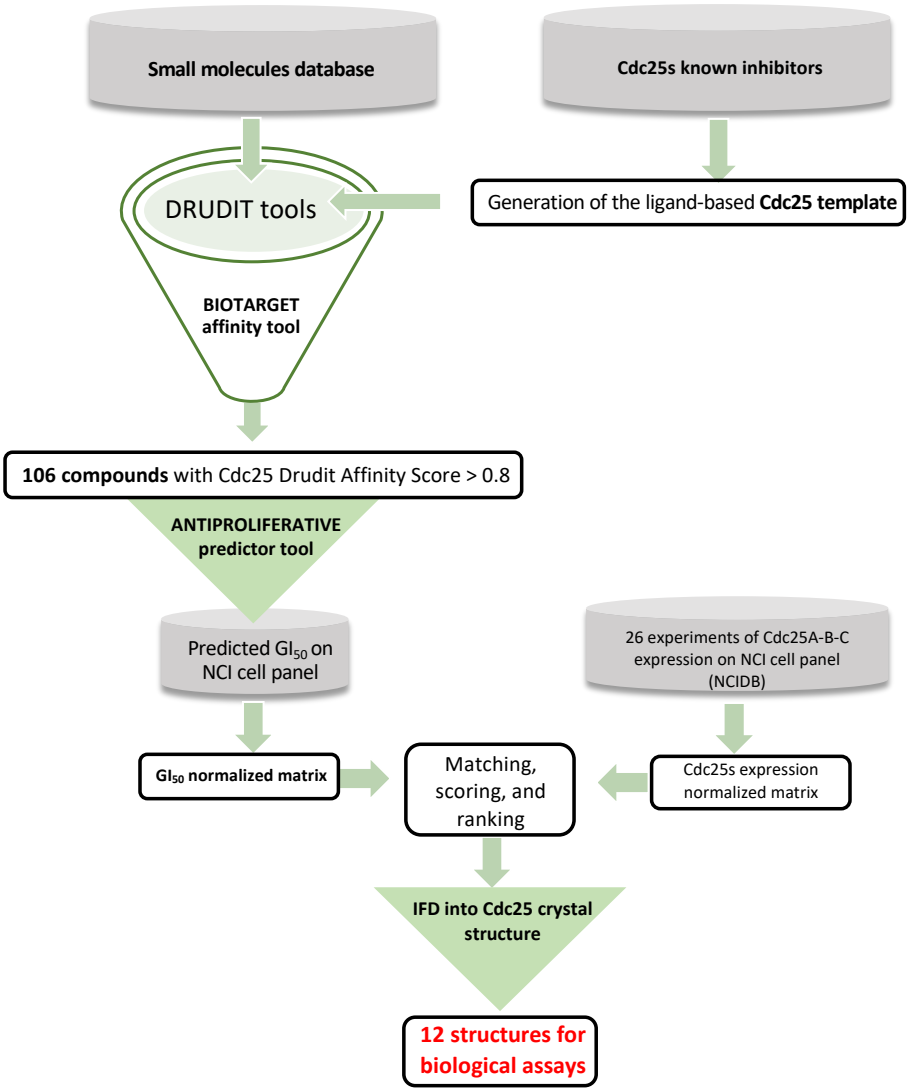


Figure 17. *The in silico protocol aimed at identifying new Cdc25 inhibitors from a database of purchasable compounds (Sigma-Aldrich, St. Louis, MO, USA).*

3.3.2.1 Ligand-based studies

In the first step of the *in silico* protocol (workflow in **Figure 17**) the Biotarget finder tool was used [42]. As anticipated, this ligand-based tool, already developed by the research group, is capable to predict the biological affinity of input structure against a given target, which is virtualized by the molecular descriptors of the corresponding inhibitors [42].

Thus, the Cdc25 template was built and integrated as an external biological target in the DRUDIT^{online} platform, following the procedure reported in the literature [42]. In details, a set of small molecules containing 117 various Cdc25s inhibitors with IC_{50} values lower than

10 µM was collected from BindingDB [133] (Supplementary Material S16). The set of 117 structures was processed with MOLDESTO (the molecular descriptors tool, implemented in DRUDIT). The output matrix (structures versus molecular descriptors) was converted into a sequence (the template) of a pair of values for each molecular descriptor (Di) (**Figure 18**): mean (µ) and standard deviation (σ).

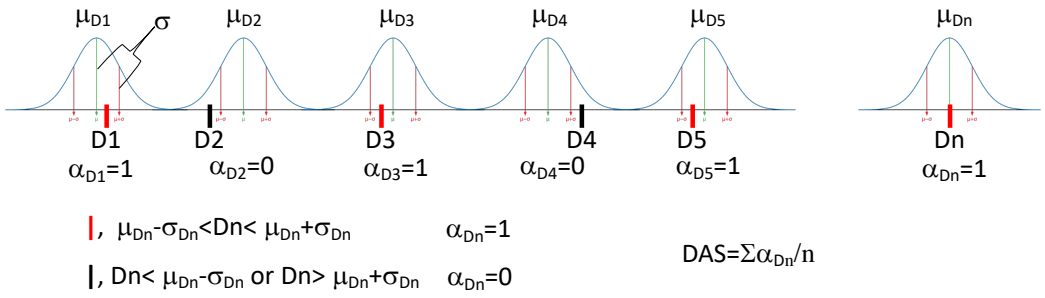


Figure 18. *DAS (DRUDIT affinity score) calculation: D1, D2, …, Dn: molecular descriptor values for the input structure; n: number of molecular descriptors.*

Then, a database of purchasable compounds including more than 10.000 small molecules (Sigma-Aldrich catalogue, St. Louis, MO, USA) was selected and submitted to the Biotarget Finder tool in DRUDIT in order to rank each structure according to their DRUDIT Affinity Score (DAS) against the Cdc25 template, as reported in **Figure 18**. The protocol assigns the α_n binary score to each molecular descriptor (D_i). This value is 1 when the molecular descriptor (D_i) is in the range $\mu(D_i) \pm \sigma(D_i)$, and 0 otherwise (**Figure 18**). The DAS score is assigned as $\Sigma\alpha_n/n$, thus it is in the range 0 ÷ 1 (low ÷ high affinity).

Supplementary Material S17 reports the full BIOTARGET matrix produced by DRUDIT and the ranking of the input structures against the Cdc25 template. The application of the 0.8 cutoffs to DAS reduced the database to 106 molecules (Supplementary Material S18).

In order to select the best structures for biological assay, further analysis was conducted by the application of the *in silico* approach based on the correlation drug activity–protein expression, proposed in this PhD work, (see workflow in **Figure 17**). For this purpose, a correlation method similar to the one used previously (see **Figure 11**, case studies) was established (**Figure 19**).

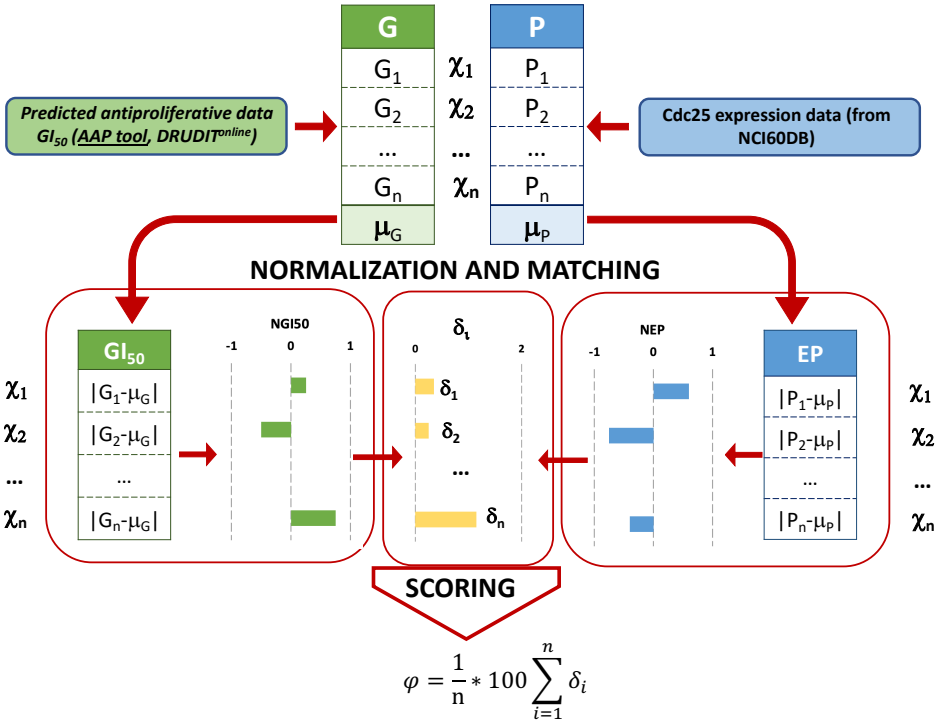


Figure 19. *Workflow for the assessment of the correlation between the antiproliferative activity values (expressed as GI_{50}s) and expression patterns (Eps) of Cdc25 on NCI-60 cancer cell lines χ_i.*

As shown, the NCI database was used as a source of Cdc25 expression patterns against the sixty NCI cell lines (hereafter, we denote the i-nth cell line by χ_i) [5]. Therefore, a set of protein expression patterns of Cdc25s against χ_i was downloaded from the NCI database. Then, 26 experiments reporting the measures of the expression pattern (EP) for each Cdc25 form against χ_i (10 for Cdc25A, 6 for Cdc25B, and 10 for Cdc25C, Supplementary Material S19) were selected, obtaining the molecular target expression pattern values (P_i) and their mean value μ_P. The deviation of each P_i from the μ_P normalized against the highest absolute value was computed to obtain the NEP_i for each of the Cdc25 forms.

Because the experimental GI_{50} values for the selected structures were not measured, we used the Antiproliferative Activity Predictor tool in DRUDIT, to predict the GI_{50} values against χ_i for the input structures (Supplementary Material S20). Each of the 106 selected input structures was processed as follows. The mean value μ_G of the predicted GI_{50} values G_i against χ_i was computed. Then, the deviation of each G_i from μ_G was calculated and normalized against the highest absolute value in order to obtain NGI_{50i} (Supplementary Material S20).

Then, a scoring method was established: the differences, $\delta_i = |NEP_i - NGI_{50i}|$, and the fitting score, $\Phi = \Sigma\delta_i$, were computed for each of the Cdc25 A, B, and C form structures. Finally, each structure was ranked based on the mean of the three ϕ values (Φ). The highest scoring structures were those that reported lower values of Φ, indicating the best correlation between protein expression pattern and sensitivity.

Among the 106 structures, the first 24 were selected for structure-based study in the next *in silico* step (**Figure 20** , Supplementary Material S21).

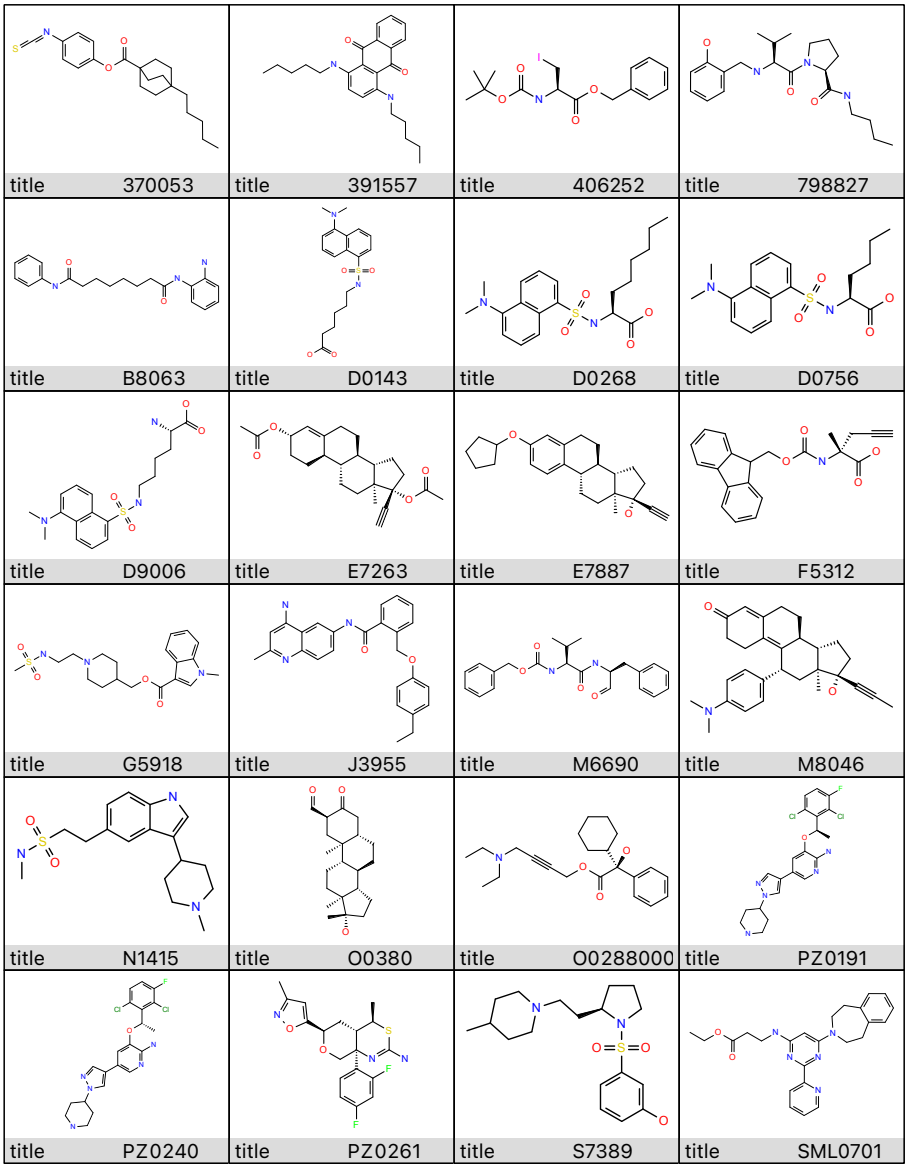


Figure 20. *Representation of the chemical structure of the 24 small molecules selected as potential Cdc25s modulators and identified by the correlation between protein expression pattern (EPs) and antiproliferative activity (GI_{50}s) data.*

3.3.2.2 Structure-Based Studies

The top 24 ranked ligands were further analyzed through structure-based studies in order to select the compounds that best fitted to the binding site of Cdc25s. A range of possibilities for the binding site and binding mode of various structures into Cdc25s binding sites were reported in the literature, identifying two potential binding regions for the target inhibition: the "swimming pool" pocket and the substrate catalytic site (**Figure 15**) [105,134,135]. Furthermore, the suggested binding modes are various, and they are obtained by the use of different molecular docking programs. For these reasons, and in order to cover all the possibilities, the docking grid was extended to both the catalytic and the "swimming pool" zones. Then, the 24 top-scoring molecules (**Figure 20**) were submitted to induced-fit study into the binding site of Cdc25B (see Materials and Methods section, paragraph 4.1.2), whose crystal structure is available at the PDB as 1QB0 [104]. The induced-fit docking (IFD) results (**Table 7**) allowed for the selection of the top 50% of scored molecules for their investigation, as Cdc25s modulators, in wet screenings (bold in **Table 7**).

Table 7. *Ligand and structure-based output results. DAS: DRUDIT affinity score; IFD: induced-fit docking.*

Cpd #	DAS	Docking Score	Prime Score	IFD Score
370053	0.824	−6.067	−5898.431	−396.631
391557	0.832	−7.523	−5938.661	−398.557
406252	0.842	−5.433	−5950.001	−398.655
798827	0.81	−10.426	−5945.922	−403.413
B8063	0.824	−8.394	−5988.759	−401.905
D0143	0.8	−6.39	−5915.872	−398.467
D0268	0.822	−6.625	−5846.066	−398.491
D0756	0.802	−5.706	−5925.995	−398.421
D9006	0.806	−6.233	−5900.211	−398.686
E7263	0.838	−7.806	−5939.597	−398.736
E7887	0.82	−8.818	−5930.375	−399.411
F5312	0.854	−7.861	−5973.455	−401.703
G5918	0.814	−7.472	−5939.175	−401.572
J3955	0.814	−7.846	−5957.553	−400.302

Cpd #	DAS	Docking Score	Prime Score	IFD Score
M6690	0.836	−8.443	−5995.928	−404.249
M8046	0.810	−6.245	−5931.352	−396.262
N1415	0.822	−4.851	−5934.807	−398.316
O0288000	0.826	−10.547	−5942.99	−402.148
O0380	0.800	−7.945	−5958.136	−398.68
PZ0191	0.824	−9.439	−5997.385	−405.087
PZ0240	0.808	−9.103	−5983.756	−404.548
PZ0261	0.820	−5.406	−5967.376	−397.571
S7389	0.892	−7.65	−5875.452	−398.318
SML0701	0.834	−7.211	−6128.861	−407.205

The top 50% of the scored molecules selected for the *in vitro* screenings are shown in bold.

Further analyses were performed on the selected hits considering a series of well-consolidated parameters for the search of bioactive compounds, such as PAINS filters [136]; Lipinski's rule [137]; Veber rules [138]; and Egan rules [139]. Thus, the 12 previously selected molecules were submitted to SwissADME web-tools (**http://www.swissadme.ch**) [140]. The results reported in **Table 8** show that, generally, the selected compounds meet expectations in terms of bioactivity. In particular, seven of the twelve structures have no violations, and only the Cpd 798827 presented two rule violations (PAINs and Veber).

Table 8. *Drug-likeness parameters calculated for the selected compounds.*

Cpd *	P	L	V	E	Cpd *	P	L	V	E
M6690	0	0	1	0	PZ0240	0	0	0	0
E7887	0	1	0	0	PZ0191	0	0	0	0
E7263	0	1	0	0	J3955	0	0	0	0
F5312	0	0	0	0	798827	1	0	1	0
O0288000	0	0	0	0	SML0701	0	0	0	0
B8063	0	0	1	0	G5918	0	0	0	0

* **P**: PAINS #alert; **L**: Lipinski #violations; **V**: Veber #violations; **E**: Egan #violations.

Finally, the molecular descriptors matrix of the selected compounds was merged with that of the known Cdc25 inhibitors used to build the template, and multivariate analysis

was performed. From this analysis emerged the method behind which the protocol was able to select molecules that are structurally different from those used in the building of the template. In fact, by applying principal component analysis (PCA) to the matrix (Supplementary Material S22), the visual inspection on the PCA 2D representation (**Figure 21**) allows for the identification of a central region where the selected structures are clustered, while the template structures are spread in the left or right area, in approximatively two clusters.

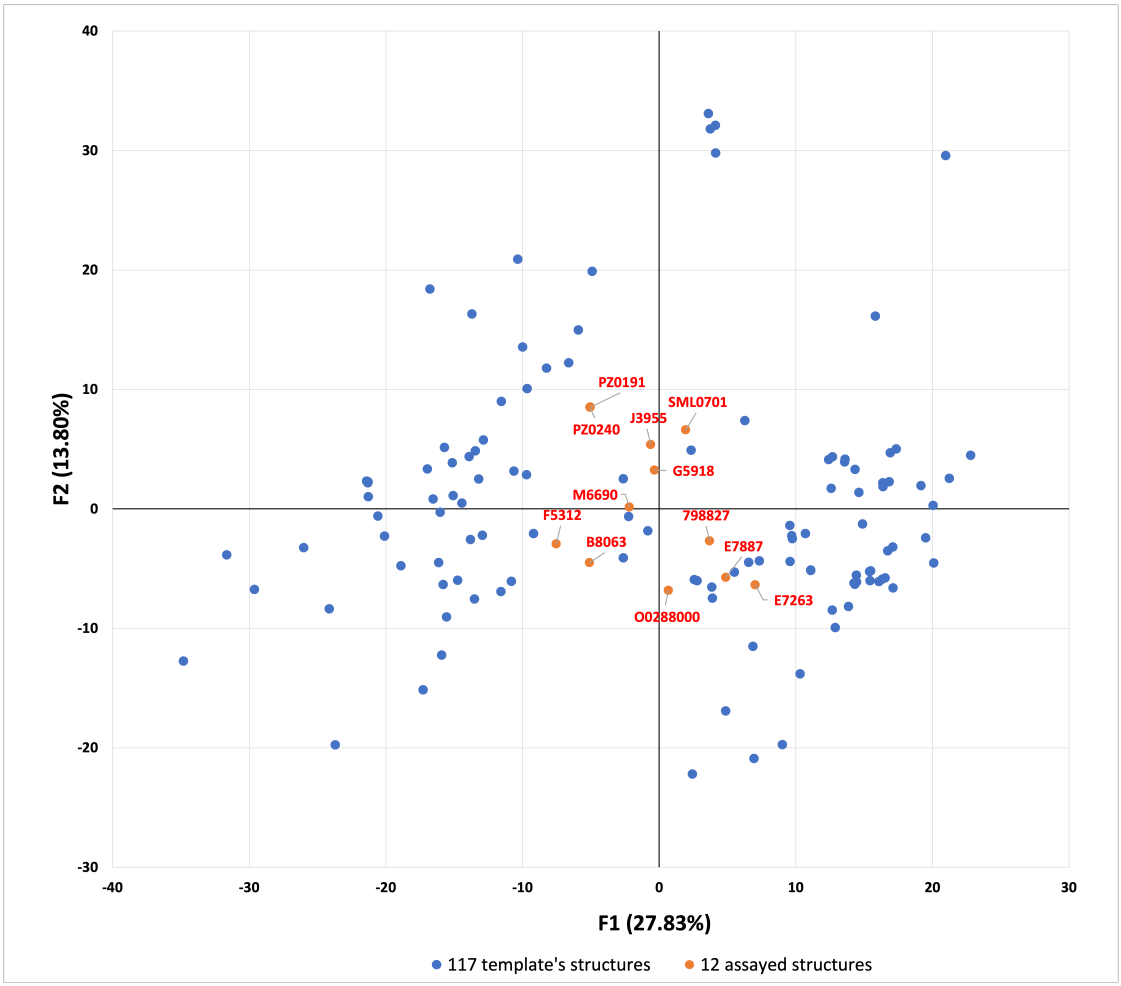


Figure 21. *Principal component analysis (F1 versus F2) applied to the molecular descriptor matrix of the selected compounds merged with the known Cdc25 inhibitors.*

These results suggested that the ligand-based step of the protocol was able to select molecules endowing various scaffolds, in a different manner from that of the classical ligand-based methods.

3.3.3. Cdc25s Phosphatase Inhibitory Activity

The selected compounds, whose structures and purity consistency were checked by HMRS analysis (Supplementary Material S23), were tested in a dose-response assay on Cdc25 phosphatases. Eight of the twelve tested compounds (798827, B8063, E7887, F5312, J3955, M6690, O0288000, and PZ0240) inhibited *in vitro* the recombinant human Cdc25s phosphatases in a concentration-dependent manner. Concerning Cdc25A, the recorded IC_{50} values ranged from 1.12 ± 0.09 to 19.88 ± 2.07 µM (**Table 9** and Supplementary Material S24). A similar pattern of inhibition was observed on Cdc25B and Cdc25C, even though a lower inhibition activity was displayed. The determined IC_{50} values for the most active compounds (E7887 and J3955) were comparable to the IC_{50} of the menadione, a Cdc25 quinonoid inhibitor used as a positive control in this study. On the contrary, the compounds E7263, G5918, PZ0191, and SML0701 did not reduce the activity of the phosphatase in the tested concentration range (0.25–25 µM).

Table 9. *IC_{50} values of the selected compounds for the inhibition of Cdc25 A, B, and C phosphatases.*

Cpd	Cdc25A (µM)	Cdc25B (µM)	Cdc25C (µM)
798827	12.02 ± 1.03	16.73 ± 1.71	14.43 ± 1.33
B8063	15.23 ± 1.37	18.03 ± 1.11	17.73 ± 1.36
E7263	>25	>25	>25
E7887	7.41 ± 0.79	8.91 ± 1.02	8.12 ± 0.87
F5312	17.12 ± 1.91	19.11 ± 1.88	19.34 ± 1.65
G5918	>25	>25	>25
J3955	1.12 ± 0.09	2.19 ± 0.07	2.22 ± 0.07
M6690	17.16 ± 1.14	19.83 ± 1.41	17.79 ± 1.83
O0288000	12.03 ± 1.17	14.14 ± 1.51	12.93 ± 1.56
PZ0191	>25	>25	>25
PZ0240	19.88 ± 2.07	22.13 ± 2.17	20.37 ± 2.34
SML0701	>25	>25	>25
menadione	4.48 ± 0.17	5.97 ± 0.75	4.49 ± 0.27

All values are the mean ± S.D. of three independent determinations.

To the best of our knowledge, this is the first time that the Cdc25 inhibition activity of the tested compounds has been assayed. On the other hand, among the tested compounds, phosphatase inhibition activity is documented only for the clinically approved anticancer agent PZ0240 ((S)-crizotinib), which, in addition to its effect as a kinase inhibitor, is a stereospecific inhibitor of 2-hydroxy-dATP diphosphatase1 or MutT homolog 1 (MHT1) phosphatase [141], an enzyme required for the survival of cancer cells and involved in DNA repair processes to maintain genome stability under oxidative stress [142,143].

3.3.4. Antiproliferative Screening

To assess Cdc25 inhibitors in cells, the antiproliferative activity of the most promising compounds (798827, B8063, E7887, F5312, J3955, M6690, O0288000, and PZ0240) was evaluated on the HepG2 tumor cell line for 48 h via MTT-based cell viability assay. Three of the tested compounds (E7887, J3955, and PZ0240) showed concentration-dependent antiproliferative activity with GI_{50} in the low micromolar range (**Table 10**). On the contrary, the other molecules showed low or no activity in the tested concentration range. Moreover, among the inactive molecules, only for B8063 (BML-210), an inhibitor of histone deacetylase, the antiproliferative activity was previously reported on leukemic and cervical cancer cells [144,145].

Table 10. *The antiproliferative activity of the selected compounds at 48 h against HepG2 cell lines expressed as GI_{50} values (GI_{50} ± SE (μM)).*

Cpd	GI_{50} (μM)
798827	>25
B8063	23.03 ± 2.13
E7887	13.03 ± 0.85
F5312	>25
J3955	1.50 ± 0.37
M6690	20.01 ± 1.87
O0288000	>25
PZ0240	7.35 ± 0.77

No data on the antiproliferative activity of E7887 (quinestrol), a synthetic estrogen used to treat postmenopausal syndrome and as a contraceptive component [146], had been reported in the literature prior to this study. Instead, the antiproliferative activity of J3955, a high-affinity and selective opioid receptor-like 1 (ORL1) receptor antagonist, was recently explored and its effects on osteosarcoma and hepatocarcinoma cells were demonstrated [147,148].

However, while we recorded a GI_{50} value of 1.50 ± 0.37 µM on HepG2 cells, the data available in the literature showed antiproliferative activity at much higher concentrations. In particular, Zhao B. and Hu T. found antiproliferative effects on HepG2 cells at concentrations above 20 µM [148]. The reason for this discrepancy is probably due to the dissimilar cancer cell lines used for the analysis, with genetic and mutational profiles that are not completely superimposable.

In their studies, Zhao B. and Hu T. employed a type of HepG2 cells characterized by deep invasive properties. In our assays, we used poorly invasive cells with phenotypic requirements generally described in the literature for the HepG2 line. The low invasiveness of hepatocarcinoma cells is justified by a high expression of epithelial cadherins and by a lack of expression of mesenchymal cadherins [149].

The biological results suggested an interesting correlation between the antiproliferative effect on HepG2 cell lines and the inhibition properties of Cdc25 enzymes. In particular, compounds E7887 and J3955, characterized by the highest Cdc25 modulation activities (Table 8) showed notable cellular growth inhibition with GI_{50} values of 13.03 ± 0.85 and 1.50 ± 0.37 µM, respectively (**Table 10**).

The antiproliferative data, observed for the PZ0240 compound (GI_{50} values 7.35 ± 0.77 µM), are in line with its anticancer therapeutic use as tyrosine kinase and an MTH1 inhibitor. The observed Cdc25s inhibition activities do not exclude multi-target effects.

3.3.5. Cell Cycle Distribution and Phosphorylation of Cdk1

Based on the results described above, among all the tested molecules compound J3955 was identified as the one with the best inhibition effect on the Cdc25 phosphatase and with

the highest antiproliferative activity. With the aim to further elucidate the mechanism of action of J3955 as a Cdc25 inhibitor, we performed flow cytometric analysis on HepG2 cells.

Since the Cdc25s enzymes control the cell cycle through the dephosphorylation of their natural substrate Cdks, the inhibition of Cdc25s results in the hyperphosphorylation of Cdks with consequent cell-cycle arrest. Therefore, the impact of cell exposure to J3955 on cell-cycle progress and the phosphorylation state of Cdks was investigated.

The flow cytometric analysis, for cell-cycle perturbation experiments, was executed in order to detect the shifts in cell-cycle distribution before a significant number of cells underwent apoptosis. The working concentrations of the compound J3955 were fixed at 1× and 2× of its GI_{50} value used in the cell proliferation assay at 48 h.

The histograms in **Figure 22** represent the percentage of cells in respective cell-cycle phases (G1, S, and G2/M), along with the percentage of cells in the sub-G1 (dead cells) obtained by flow cytometry after either a 12 h (**Figure 22**, panel A) or a 24 h (**Figure 22**, panel B) treatment. In the absence of J3955, HepG2 cells showed a normal diploid distribution with fast proliferation characteristics, with S + G2/M phase cells accounting for about 45% of the total cells. A 12 h treatment with J3955 arrested the cell cycle at the G2/M phase in a dose-dependent manner (**Figure 22**A). An increase in G2/M phase cells from 24% to 27% and from 24% to 36% ($p < 0.0001$) was observed as a result of cell exposure to J3955 at 1 × GI_{50} (1.5 µM) and at 2 × GI_{50} (3.0 µM), respectively. The cell accumulation in the G2/M cell-cycle phase was coupled to a decrease in the G0/G1 phase cells rather than a decrease in the S phase.

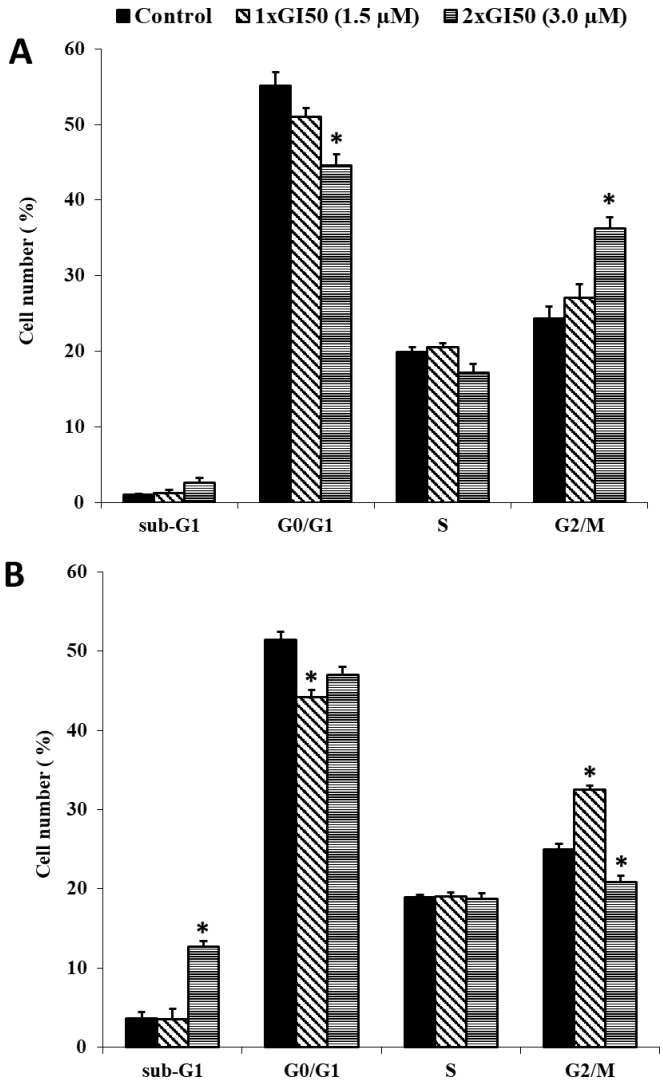


Figure 22. *The effects of J3955 at 2× and 1× its GI_{50} value on the cell-cycle distribution of HepG2 cells following 12 h (**a**) and 24 h (**b**) treatments. Results are expressed as the mean of two independent experiments, performed in duplicate. Statistical analyses were performed using the Student's t-test to determine the differences between the datasets. * Denotes significant differences ($p < 0.0001$) from untreated cells (control).*

A similar trend was observed after 24 h treatments with J3955 at 1 × GI_{50} (1.5 µM): G2/M phase cells increase from 25% to 32% ($p < 0.0001$) (**Figure 22**B). However, when the cells were exposed to J3955 at 2 × GI_{50} for 24 h, a new sub-G0/G1 population appeared, indicative of apoptotic cells, with a parallel decrease of the population in the G2/M phase (**Figure 22**B).

Dephosphorylated Cdc25s are activated by both Cdk1 and Cdk2. Then, Cdc25s activate both the G1/S transition and S-phase Cdk-cyclin complexes (Cdk2-cyclinE and Cdk2/cyclinA), but also the Cdk1/cyclin B complex involved in the G2/M transition. Consequently, although

Cdc25 inhibition induces cell-cycle arrest, the stage of cell-cycle block by Cdc25 inhibitors cannot be predicted and literature data show that it is cell line-dependent.

In the same cell line, it is observed that several molecules displaying Cdc25 inhibition effects trigger different cell-cycle arrests. For example, Kabakci et al., studying the inhibition effect of various naphthoquinone compounds on Cdc25s, observed different cell-cycle arrests in HeLa cells (G1/S or G2/M arrest) [150] and an imidazopyridine Cdc25 inhibitor triggered S-phase arrest in MCF-7, HepG2, and HT-29 cell lines [115].

On the other hand, it has been observed that two structurally unrelated Cdc25 inhibitors arrested melanoma cell lines in the G2/M cell-cycle phase and activated an apoptotic program [118,119,151]. In addition, in previously published works some terpenoid compounds were tested on A375.S2 human melanoma cell lines. The authors described a different stopping phase of the cell cycle via Cdc25, strictly dependent on the effect of the tested triterpenoid compound [152,153].

Endogenous Cdc25s control cell cycle progression through dephosphorylation via the activation of their natural substrate Cdks. Thus, to directly assess Cdk1 activity in HepG2 cells we used antibodies recognizing phosphorylated Thr14 or Tyr15, two amino acid residues selectively dephosphorylated by Cdc25 in the Cdk1 catalytic domain [150].

In order to investigate the involvement of the inhibition of Cdc25s in the antiproliferative action of J3955, the phosphorylation status of Cdk1 was analyzed after cell exposure to J3955.

Western blot analysis of the lysate from HepG2 cells treated with J3955 at 0.75 µM, 1.5 µM, and 3 µM for 6 h showed a significant dose-dependent accumulation of the phosphorylated form of Cdk1, when compared to control cell lysates (**Figure 23**). Considering that an increase in the phosphorylated protein fraction results in a decrease in the non-phosphorylated and active protein fraction, with the amount of total protein being almost unchanged along the different treatments, our results indicate that J3955 may impair Cdk1 activity in exposed cells and suggests its specific influence in molecular mechanisms involving Cdc25 proteins.

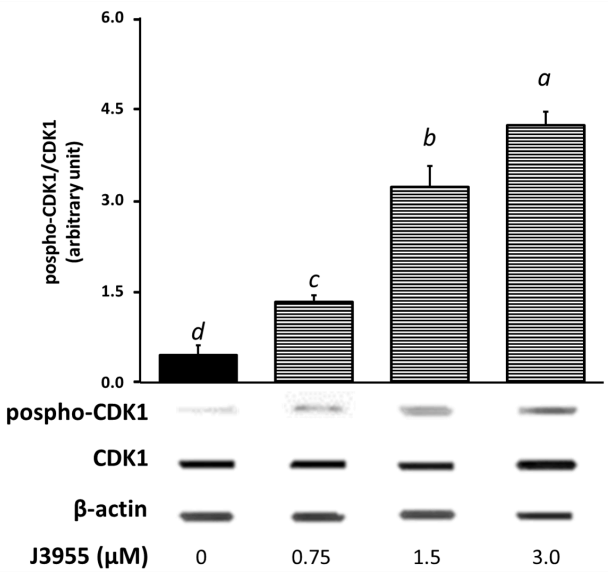


Figure 23. *Effects of a 6 h treatment with J3955 (0.75, 1.5, and 3 μM) on Cdk1 phosphorylation in HepG2 cells. After the treatment, cells were collected, and the proteins were isolated for Western blot analysis as described in ''Materials and methods''. The panel shows a representative Western blot and densitometric analysis. The values represent the ratio between phospho-Cdk1 and total Cdk1, both previously normalized for the corresponding β-actin. Values are expressed as the mean ± S.D. of three separate experiments with similar results. Different lowercase letters on the top of each histogram indicate statistical ($p < 0.05$) differences among the tested samples, as measured by one-way ANOVA followed by the Tuckey test. The letter "a" marks the highest value. Bars not sharing the same letter were significantly different.*

3.3.6. Binding poses of the most interesting Cdc25 inhibitors identified in the study

To gain further insight into the structure-based results, the binding modes of the most interesting compounds were analyzed. The best example, observed by the IFD study (J3955 **Figure 24a**,**b**), shows several interactions at the junction between the catalytic pocket and the "swimming pool". In particular, the flat aromatic quinoline scaffold of J3955 is accommodated in the surface of the shallow active site, whereas the flexible and hydrophobic phenylethyl moiety is able to penetrate deeply into the adjacent "pool", where it encounters several apolar residues (Met^{483}, Leu^{443}, Pro^{444}, Leu^{445}, Cys^{426}, and Tyr^{428}). The amidic portion forms several H-bonds with Glu^{478} (in the catalytic HCX_5R motif) and Arg^{544} (C-terminus, in the "swimming pool") side chains.

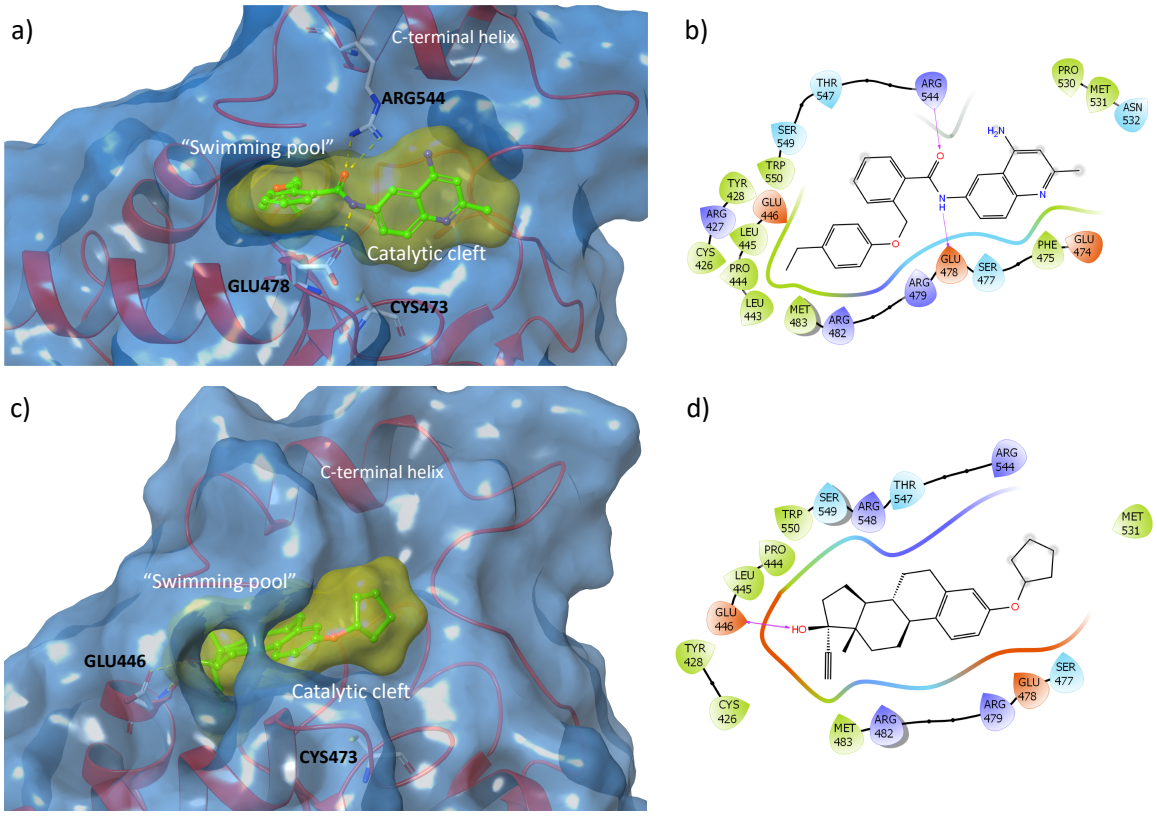


Figure 24. *Predicted binding modes of J3955 (**a**) and of E7887 (**c**) into the "swimming pool"-catalytic cleft of Cdc25B (PDB id: 1QB0); ligand–protein interaction diagram for J3955–Cdc25B (**b**) and E7887–Cdc25B (**d**) complexes, with the hydrogen bond shown as violet arrows.*

Similarly, but with completely different structural features, compound E7887 also could inhibit Cdc25 by binding both to the catalytic cleft (partially occupied by the cyclopentyl ring) and to the "swimming pool", in which the hydrophobic steroid scaffold is inserted (**Figure 24c**, **d**). Inside this cavity, 17β-OH stabilized the complex by forming H-bonds with the side COOH and the backbone NH of Glu[446].

The lack of reactive groups could permit a reversible binding, avoiding the well-established toxicity of the quinonoid agents and electrophilic entities.

4. Materials and Methods

4.1 *In silico* studies

4.1.1. Ligand-based

- *Hardware:* The DRUDIT WEB service runs on 4 servers that are automatically selected according to the number of jobs and online availability. Each server can support up to 10 simultaneous jobs, while the exceeding jobs are placed in a queue.
- *Software:* DRUDIT consists of several software modules implemented in C and JAVA and running on MacOS Mojave.
- *Database selection and data sets building.* The NCI-60 database, containing both antiproliferative and chemical data of thousands of compounds, was selected as a reliable source to construct the AAP protocol. In details, since the presented tool is based on molecular descriptors, the 2D chemical structures of the NCI tested compounds (.mol files, available only till the June 2016 release) and the corresponding growth inhibition data were retrieved from the NCI website (284.176 chemical structures) [44,154]. Among these thousands of compounds, only those tested in the five-dose assay, thus with available GI_{50} data, were selected to build and validate the model. In particular, the structures were split in two sets: a training set, containing more than 34k compounds released till 2014 (NCI2014DB), was used to construct the protocol; a test set, containing about one hundred compounds that were first released in 2016 (NCI2016DB), was used to validate the AAP tool. The protein expression data used in the correlation analysis were downloaded from the NCI-60 database - Molecular Characterization Program [155]. The Binding database (Bdb) [133] storing Ki, Kd, IC_{50}, and EC_{50} values of known inhibitors of well-defined protein targets [156,157], was used as reliable source to download the 2D structures of the known Cdc25 inhibitors reported in literature. The selected 117 Cdc25 inhibitors with IC_{50} values lower than 10 μM were used to build the corresponding template in the Biotarget finder tool implemented in $DRUDIT^{online}$, following the procedure reported in literature [42]. The database of purchasable molecules submitted to *in silico* screening against Cdc25 contains 10.715 structures, was retrieved from the Sigma-Aldrich repository.

- *MOLDESTO: a new software for molecular descriptors calculation.* MOLDESTO (MOLecular DEScriptor Tool), as we described in a previous work [42], is a software tool implemented in DRUDIT that represents the evolution of the expertise of the research group in the calculation/ manipulation of molecular descriptors [158]. It is currently able to calculate more than 1000 1D, 2D, and 3D molecular descriptors for each input structure (the full list of molecular descriptors calculated by MOLDESTO is reported in Supplementary Material S1). The input structures can be drawn directly in the web interface or uploaded as commonly used molecule file formats as external files (e.g. SMILES, SDF, Inchi, Mdl, Mol2). The software is provided with a caching system to boost the calculation speed of previously submitted structures.
- *DRUDIT settings for Antiproliferative Activity Predictor (AAP) tool.* The AAP tool comprises the fingerprint (FP) and cell lines (CL) modules, which cooperate simultaneously to assign the predict GI_{50} values to an input structure. In each module, the performed calculation is dynamic; indeed, it could be modulated by appropriately varying the values of the available parameters (three for each module, see below).

  FP module: The FP module parameters are: choice of biological activity as GI_{50}, TGI, LC_{50}, or G% (in this work only the first choice was considered); N (-b) the best number of dynamically selected molecular descriptors; Z (-m) number of descriptors for which |v m|/m < <value> applies. v: descriptor value, m: target mean; G (-c) max number of zero percentage values per descriptor.

  CL module: The DRUDIT parameters for the CL module are: choice of biological activity as GI_{50}, TGI, LC_{50}, or G% (in this work only the first parameter is considered); N (-b) the best number of dynamically selected molecular descriptors; Z (-m) the max number of zeros percentage per descriptor; G (-f) the Gaussian smoothing function to be used (a, b, or c mode).

4.1.2. Induced Fit Docking on Cdc25 X-ray structure

The IFD (induced-fit docking) was applied by means of the Schrödinger software suite [159-162] by using the settings from previous works [163,164]. Cdc25B atomic coordinates were downloaded from the Protein Data Bank (PDB id, 1QB0) and refined by the Protein

Preparation Wizard module to apply default parameters [165]. The IFD score, which accounts for both the protein-ligand interaction energy and the total energy of the system, is calculated weighing 95% of Glide Gscore and 5% of Prime Energy. It is used to rank the IFD poses considering that the more negative the IFD score, the more favorable the binding [166-170].

4.2. NCI-60 antiproliferative screenings

4.2.1 Compounds selection guidelines

The compounds to be screened are selected according to precise and rigorous guidelines; in general, the submission is encouraged for: molecules which bring some novelties (novel heterocyclic ring systems, privileged scaffolds) to the NCI collection; compounds emerged from computer aided drug design. In addition, in the case of series of analogues, it is preferred to select only the one which is expected to provide the greatest information. On the other hand the submission of compounds with the following features is discouraged: excessive flexibility; presence of drug-unlike function groups (nitro, nitroso, diazo, imine…); presence of chemical portions which could affect the reliability of the assays (PAINS) [171].

4.2.2. One dose assay

All compounds submitted to NCI are firstly assayed against the NCI-60 DB in a one dose screen (concentration of 10^{-5} M); this kind of assay aims to determine the G% (growth inhibition percent) of the compounds against the considered cells. The results are plotted in a one-dose graph showing the G% of the single compound against the 60 cell lines. This first assay is considered passed only for the most promising compounds (satisfaction of predetermined threshold criteria); in this case the compound pass to the 5 doses screen (for further experimental details about the standardized assay procedures see ref. [172,173]).

4.2.3. Five-doses assay

The most active compounds are submitted to multiple dose screen, using 5 different concentrations (ranging from 10^{-8} to 10^{-4} M). The dose response curves obtained from this assay permit to extrapolate the GI_{50} (the molar concentration of the compound that inhibits 50% of cell growth), TGI (the molar concentration of the compound leading to total

inhibition of cell growth) and LC_{50} (the molar concentration of the compound that induces 50% cell death) values of the selected compounds against each cancer cell lines. For each of the mentioned parameters a Mean Graph MIDpoint (MG_MID) is calculated, providing an average activity parameter over all the cell lines (for further experimental details about the standardized assay procedures see ref.[172,173]).

4.3. Biological procedures

4.3.1. Cdc25s phosphatase inhibitory activity assay

The inhibitory activity of the selected compounds for Cdc25s was assessed using the CycLex protein phosphatase Cdc25A, -B, and -C fluorometric assay Kit (CycLex, Cat. No. CY-1355) in accordance with the manufacturer's protocol. The assay is based on the competition of the test compound for O-methyl-fluorescein phosphate (OMFP), an exclusive fluorescence Cdc25 substrate.

An assay mixture containing OMFP was freshly prepared following the kit instructions. Test compounds were previously dissolved in DMSO to obtain stock solutions at 20 mM and kept at −20 °C. Working solutions of each compound were freshly prepared in the assay buffer. In each well, 40 µL of assay mixture was mixed with 5 µL of the test compound. The reaction was initiated by adding 5 µL (0.1 µg/µL) of the purified recombinant Cdc25 (Cdc25A, -B, and -C) proteins and mixing thoroughly. The plate was incubated at room temperature for 15 min. Then, 25 µL of stop solution was added. Phosphatase activity was measured in a 96-well microtiter plate using a Cdc25s substrate. Fluorescence intensity (FI) was measured using a GloMax®-Multi Microplate Reader equipped with a GloMax®-Multi Fluorescence Module (Promega Corporation, Madison, WI, USA) with an excitation wavelength of 485 nm and an emission wavelength of 580 nm. The background was defined as the FI generated from the wells that did not contain Cdc25s but were incubated with the assay mixture. The percentage enzyme activity of the test sample with respect to the control (OMFP wells) was calculated using the following equation:

$$\% \text{ FI} = (\text{FI test sample}/\text{FI control}) \times 100 \qquad (1)$$

IC_{50} was defined as the concentration of the compound at which there was 50% FI of the OMFP wells.

4.3.2. Cell Culture

The cancer cell line HepG2 (hepatocarcinoma cells) was obtained from the American Type Culture Collection (ATCC) (Rockville, MD, USA). The cells were cultured in RPMI supplemented with 5% (*v*/*v*) FBS, 2 mM L-glutamine, 50 IU/mL penicillin, and 50 µg/mL streptomycin, and maintained in a humidified atmosphere with 5% CO_2 at 37 °C. Cells were routinely cultured in 75 cm^2 culture flasks and were trypsinized using trypsin-EDTA. Exponentially growing cells were used for experiments.

4.3.3. Anticancer evaluation assay

The selected derivatives were submitted to the MTT assay to assess the growth inhibition activity against HepG2 cells. The MTT assay is a measurement of cell metabolic activity, quite effective in estimating cell proliferation, that is based on the protocol first described by Mosmann [174]. The assay was performed as previously described [175]. Briefly, the cells were seeded into a series of standard 96-well plates in 100 µL of complete culture medium at 1.5×10^4cells/cm^2. Cells were incubated for 24 h under 5% CO_2 at 37 °C and the medium was then replaced with 100 µL of fresh medium supplemented by 5% (*v*/*v*) FBS containing the treatments. The stock solutions (20 mM) were prepared by dissolving selected compounds in DMSO. Working solutions were freshly prepared on the day of testing by diluting the stock solutions in the complete culture medium. For the experiment, we used a concentration range from 20 to 0.02 µM. Twenty-four hours after seeding, aliquots of 100 µL of different solutions at the appropriate concentrations were added to the appropriate wells and the cells were incubated for 48 h without the renewal of the medium. In each experiment, the DMSO concentration never exceeded 0.25% and a culture medium with 0.25% DMSO was used as control. After the incubation time, cells were washed and 50 µL FBS-free medium containing 0.5 mg/mL of MTT was added. The medium was discarded after a 3 h incubation at 37 °C and formazan blue formed in the cells was dissolved in DMSO. The absorbance (OD, optical density) at 570 nm of the MTT-formazan was measured in a

microplate reader. As the absorbance is directly proportional to the number of living, metabolically active cells, the percentage of growth (PG) with respect to the untreated cell control for each drug concentration was calculated according to one of the following two equations:

$$\text{if } (ODtest - ODtzero) \geq 0, \text{ then } PG = 100 \times (ODtest - ODtzero)/(ODctr - ODtzero) \quad (2)$$

$$\text{if } (ODtest - ODtzero) < 0, \text{ then } PG = 100 \times (ODtest - ODtzero)/ODtzero \quad (3)$$

where: ODtzero is the average of the optical density measurements before the exposure of cells to the test compound; ODtest is the average of the optical density measurements after the desired period of time; and ODctr is the average of the optical density measurements after the desired period of time with no exposure of the cells to the test compound. The concentration necessary for 50% of growth inhibition (GI_{50}) for each derivative was calculated from concentration−response curve using linear regression analysis, by fitting the test concentrations that give PG values above and below the reference value (50%). Each result was the mean value of three separate experiments performed in quadruplicate. Finally, in order to exclude potential cytotoxic effects at the concentration range used for our experiments, the Trypan blue exclusion method was employed.

4.3.4. Cell-cycle analysis

DNA staining with propidium iodide (PI) and flow cytometry analysis were applied as previously described with the aim to evaluate the effects of the selected derivatives on cell-cycle progression [176]. Briefly, HepG2 cells were seeded on 12-well plates at a density of 2.0×10^4 cells/cm^2, and treated 24 h after seeding without or with the indicated concentrations of the test compound for 12 or 24 h. Following the treatments, cells were collected, washed in PBS, and stained with staining solution (20 µg/mL propidium iodide, 200 µg/mL RNAse A, and 0.1% Triton X-100 in PBS) for 30 min at 37 °C. The DNA contents of more than 10.000 cells were subjected to fluorescence-activated cell sorting (FACS) analysis (Coulter Epics XLTM, Beckman, Brea, CA, USA), and the percentage of cells belonging to the different compartments of the cell cycle was determined. All experiments were performed in duplicate and reproduced at least two times.

4.3.5. Western Blotting

The phosphorylation status of the Cdk1 was analyzed by Western blotting, as previously reported [177]. Briefly, HepG2 cells were treated with J3955 (1.5, 3, and 6 μM) for 6 h and after treatment cells were rinsed twice with ice-cold PBS and harvested by scraping in ice-cold hypotonic lysis buffer (10 mM Hepes, 1.5 mM MgCl2, 10 mM KCl, 0.5 mM phenylmethylsulfonyl fluoride (PMSF), 1.5 lg/mL soybean trypsin inhibitor, 7 lg/mL pepstatin A, 5 lg/mL leupeptin, 0.1 mM benzamidine, and 0.5 mM dithioerythritol (DTT)) and incubated for 15 min on ice. The lysates were centrifuged at 13,000× *g* for 5 min, and supernatants were immediately portioned and stored at −80 °C. The protein concentration was determined using the Bradford protein assay reagent (Bio-Rad, Hercules, CA, USA). Aliquots of cell extracts containing 5–15 μg protein were separated on 8–12% sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis and transferred to a nitrocellulose membrane. Colored molecular weight standards (Amersham) were run simultaneously. The immunoblot was incubated overnight at 4 °C with blocking solution (5% skim milk), followed by incubation with either an anti-Cdk1 monoclonal antibody (Invitrogen, Carlsbad, CA, USA, Cat: 33–1800), anti-phospho-Cdk1 (Thr14, Tyr15) polyclonal antibody (Invitrogen, Cat: 710840), or anti β-actin monoclonal antibody (Invitrogene, Cat: MA1-744) as control, for 1 h at room temperature. Blots were washed two times with Tween 20/Tris-buffered saline (TTBS) and incubated with a 1:1000 dilution of horseradish peroxidase (HRP)-conjugated polyclonal goat anti-mouse IgG antibody (Dako, Glostrup, Denmark), or with a 1:2000 dilution of horseradish peroxidase (HRP)-conjugated polyclonal goat anti-rabbit IgG antibody (Dako) for 1 h at room temperature.

4.3.6. Statistical Analysis

All data are expressed as mean ± S.D. Three independent observations were made for each experiment. Statistical difference was calculated using an unpaired Student's *t*-test. Tukey was used to examine the difference between group means.

5. Conclusions and Future Perspectives

In the last decades, a lot of computational and statistical approaches have been proposed with the aim of integrating and correlating the huge amount of experimental biological data (e.g.: antiproliferative activity and protein expression) stored in the most famous databases focused on cancer research. From this point of view, these types of correlations could allow to gain insight into the mechanism of action and to identify the biological target of anticancer compounds.

The main limitation of all these computational methods is the need for experimental drug response data (post-screen application). The possibility to predict *in silico* the antiproliferative activity of new/untested small molecules against specific cell lines, could enable correlations to be found between the predicted drug response and protein expression of the desired target from the very earliest stages of research. Such an innovative approach could allow to select the compounds with molecular mechanisms that are more likely to be connected with the target of interest preliminary to the *in vitro* assays, providing a crucial help in the design of new targeted anticancer agents.

Based on these premises, this PhD Thesis aimed to develop and propose a new *in silico* protocol based on the correlation between chemosensitivity and protein expression pattern to design new targeted anticancer drugs. To this end, the project comprised different phases:

1. The first step of the study consisted of the development of the Antiproliferative Activity Predictor (AAP), a reliable computational tool able to predict the anticancer activity of new compounds against the NCI-60 panel. This ligand-based protocol, validated by both an internal and external set of structures, has proven to be high reliable and robust. The encouraging results were further supported by the examination of an in-house database and subsequent evaluation of a set of molecules selected by the NCI for the one-dose/five-dose antiproliferative assays (paragraph 3.1).
2. After the development of the AAP tool to predict the antiproliferative activity of new compounds, the second phase of this PhD project aimed to develop a new

computational method to correlate drug activity and protein expression pattern. Before its application to the design of new anticancer agents, the proposed method was validated by the analysis of some case studies to demonstrate that such a correlation analysis could be applied to the biological data related to the full NCI-60 panel (paragraph 3.2).

3. The final part of the study aimed at applying the proposed approach to design new targeted anticancer drug. Hence, the attention was focused on Cdc25 enzymes, crucial targets to halt tumor proliferation. The new mixed ligand structure-based approach permitted, from the evaluation of a database containing more than 10.000 small molecules, the identification of 12 compounds as potential inhibitors of Cdc25s. The biological screenings of the selected structures consolidated and confirmed the *in silico* results. In particular, the enzymatic inhibition assays showed interesting Cdc25s IC_{50} values for most of the tested molecules. Among them, J3955, the most active inhibitor of Cdc25s, exhibited antiproliferative activity against HepG2 cells, with GI_{50} values in the low micromolar range. The flow cytometric analysis, for cell-cycle perturbation experiments, highlighted, after treatment with J3955, cell-cycle arrest and the accumulation of the phosphorylated form of Cdk1 (paragraph 3.3).

Collectively, our results demonstrated that the correlation between protein expression pattern and chemosensitivity revealed an innovative, alternative, and robust method in the identification of new modulators for the selected targets. Differently from the traditional *in silico* methods, the proposed protocol allows for the selection of molecular structures with heterogeneous scaffolds, which are not strictly related to the binding sites and with chemical-physical features that can be more suitable for all the pathways involved in the overall mechanism. The biological results further corroborate the robustness and the reliability of this new approach and encourage its application in the anticancer targeted drug discovery field.

6. Appendix: Visiting research period at the University of Oxford, Department of Pharmacology (March 2022 - August 2022), Supervisor: Dr. Thomas Lanyon-Hogg

6.1. Design, synthesis, and biological evaluation of new heterocyclic analogues of the bacterial DNA repair inhibitor IMP-1700: Summary of the section

In the last section of this PhD thesis, a separate chapter is dedicated to present the results of the visiting research period carried out at the Department of Pharmacology of University of Oxford (March-August 2022), under the supervision of Dr. Thomas Lanyon-Hogg. The research project was focused on the design, synthesis, and biological evaluation of new heterocyclic small molecules as inhibitors of bacterial DNA strand break repair pathways, a key element involved in the development of antimicrobial resistance (AMR).

IMP-1700 (**Figure 25**) is one of the most potent bacterial DNA repair inhibitors developed to date (single digit nanomolar potency) [178,179]. Despite the encouraging biological activities, its fluoroquinolone core, subject to an FDA warning for "black-box toxicity", makes the progression to the subsequent phase of drug development unlikely. From a medicinal chemistry point of view, few SAR studies have been conducted on the IMP-1700 structure, especially on the central quinolone scaffold. Thus, a central question to be answered is: can we substitute the 4-quinolone core in IMP-1700 structure with other heterocycles, without affecting its biological activity?

Trying to elucidate, in part, this aspect, the main aim of this study was the design, synthetize and the evaluate *in vitro* some non-fluoroquinolone heterocyclic analogues of IMP-1700 (**Figure 25**).

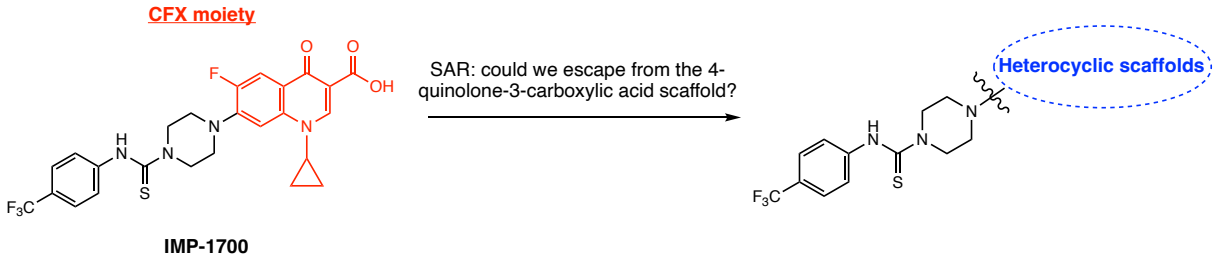


Figure 25. *Overview of the project: design, synthesis, and biological evaluation of new heterocyclic analogues of the bacterial DNA repair inhibitor IMP-1700.*

6.2 Introduction

6.2.1 The emergence of antimicrobial resistance (AMR) and the decline of antibiotics: where are we going?

The discovery of the first antibiotics during the last century have drastically transformed the modern medicine; indeed, these drugs represented a turning point not only for an effective treatment of infectious diseases (in the pre-antibiotic era more than half of deaths were attributed to infections), but also to make many modern medical therapies and procedures (such as surgery, cancer treatment, organ transplant) possible and safe [180-182].

However, despite the initial success (the so called "Golden Age" of antibiotics, ended in the early '70), their overuse/misuse in human health, but also in agriculture and livestock, evolved a rapid rise of antimicrobial resistance (AMR), resulting in a fast erosion of the antibiotic arsenal over time. As evidenced in **Figure 26**, resistance to almost all the approved antibiotics emerged immediately after their approval for clinical use [180,182-184].

In addition, some bacteria, such as *Enterococcus faecium, Staphylococcus aureus, Klebsiella pneumoniae, Acinetobacter baumannii, Pseudomonas aeruginosa,* and *Enterobacter* (ESKAPE pathogens), have developed resistance to almost all the clinically antibiotic classes, making these infections difficult to treat. Thus, the feared scenario of the "post-antibiotic era" is starting to become a real possibility in the near future [185].

This dramatic scenario is reflected in the most recent statistics: it has been estimated that more than 700.000 deaths occur annually as a consequence of multidrug resistant bacteria, a figure that could increase to 10 million a year by 2050 [182,186].

Furthermore, additional factors combine to make AMR even more challenging and threatening:

- Insufficient development of new antibacterial drugs in the last three decades (antibiotic discovery gap): as shown in **Figure 26**, only few new antibiotic drugs have been approved in the last twenty years and, in addition, all of them belong to already known class of therapeutics, such as tetracyclines, aminoglycosides and

fluoroquinolones (after the "Golden Age", only two new classes of antibiotics have been discovered) [187-191];

- withdraw of already approved antibiotics from the market [192];
- rapid drop of interest and investments by pharmaceutical companies in the antibacterial discovery field (limited financial market as new antibiotics kept as agent of last resort for most resistant infections) [181,192].

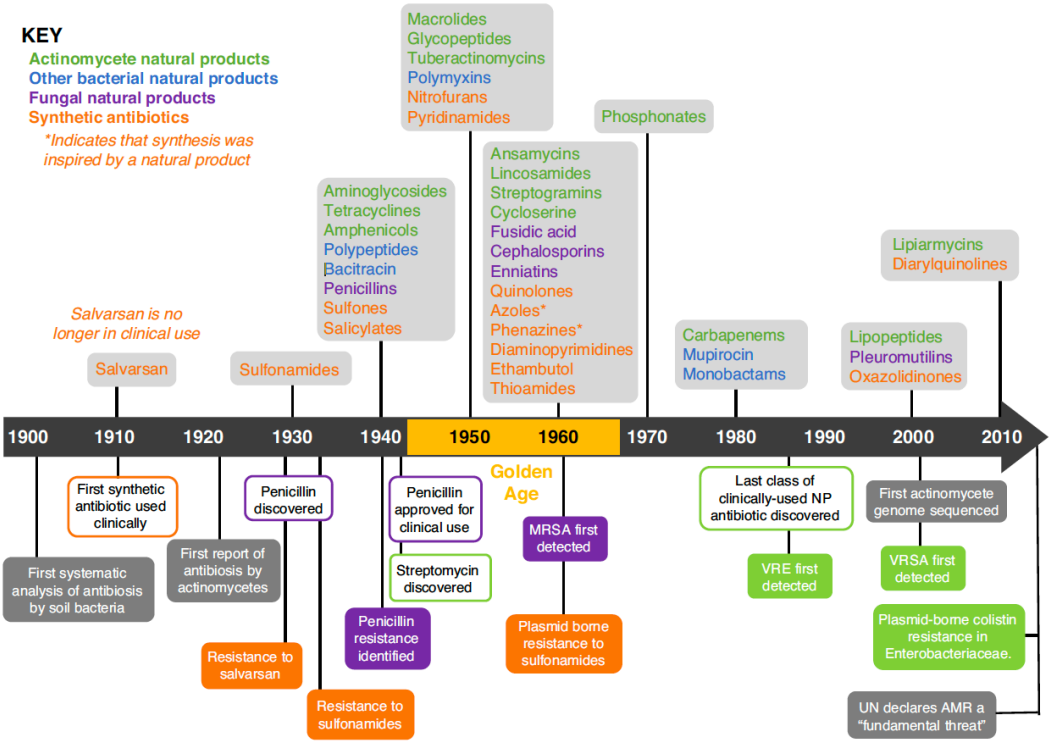


Figure 26. *Timeline of antibiotic discovery, with a special focus on the development of resistance [180].*

In this light, AMR represents one of the most serious challenges threatening global health during this century; hence, the discovery of drugs capable to treat resistant bacteria and maintain their effectiveness over time represents an urgent priority for the whole scientific community [185,193,194].

Many approaches have been undertaken during the last decades to tackle the AMR. One of the most interesting strategies could be to improve/prolong the efficacy of existing antimicrobial drugs by directly countering bacterial mechanisms of drug resistance [182]. Indeed, molecules able to block the bacterial pathways responsible of resistance could have both advantages of increasing the lifetime of the existing therapies (adjuvants) and,

meanwhile, decreasing the capability of pathogens to develop resistance and survive [178,195].

In the light of this, understanding the main mechanisms by which bacteria develop resistance to antibiotics is critical to try to find an effective solution to this crisis. Among all, the bacterial DNA repair and SOS pathways are probably the most characterized elements involved in these processes.

6.2.2. The inhibition of bacterial DNA repair mechanisms and SOS response pathway: possibility to re-sensitize multidrug resistant bacteria to existing antibiotic

Considering the crucial role of the bacterial DNA repair pathways in the rise of AMR, in this section the detailed description of the main components of this network is needed.

During the bacterial life cycle, DNA double strand breaks (DSBs) can occur; these could be due to ionizing radiation, metabolic perturbations, mechanical stress during DNA replication, oxidative bursts generated by the host immune response during infections (production of ROS), antibiotics targeting DNA structure (e.g: quinolones). A single DSB could be lethal for the bacterial cell if left unrepaired before cell division; however, the procaryotes, like eukaryotes, have developed complex DNA repair and recombination mechanisms to protect themselves [196]. Among the most famous components involved in these processes, we can find the RecBCD (in gram-negative bacteria) and the equivalent AddAB (for gram-positive bacteria), two multi-protein complexes with helicase-nuclease activity, which are able to find and repair the DSBs by the mechanism shown in **Figure 27**.

In detail, the DSB (damaged sites of bacterial DNA) is loaded with RecBCD complex that starts to unwind the double strand in an ATP- dependent manner. Once RecBCD reaches a so-called ‘crossover hot-spot instigator (Chi)’ sequence, RecB subunit cuts the strand with this sequence, continuing to unwind it. At this point RecA, a monomeric strand-exchange recombinase, is loaded onto the newly generated 3’-ended strand and starts the homologous recombination process by invading an intact homologous DNA strand, determining the DSBs repair. However, RecA activity is not limited just to DSBs repair; indeed, RecA, once complexed to a DNA strand, is able to recognize and bind the repressor proteins LexA, causing LexA autoproteolysis. This process determines the activation of the

SOS-responsive genes (more than 50 elements), with consequent expression of low fidelity DNA-polymerase, error prone DNA replication, increased rate of mutagenesis, resistance and tolerance to stresses, persistence, virulence, host immune response adaption, biofilm formation, antibiotic evasion [178,196-200].

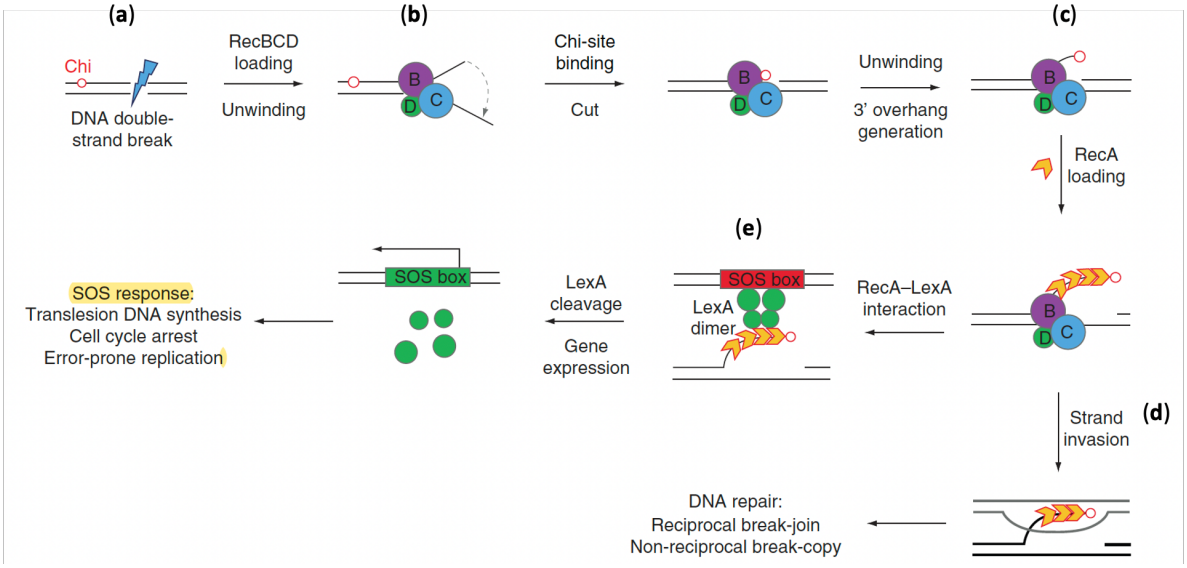


Figure 27. *Mechanism of activation of the bacterial DNA repair mechanisms and of the SOS response pathway: (**a**) external stresses (antibiotics, ionizing radiation, oxidative burst generated by host immune response, etc) cause DNA damage; (**b**) RecBCD, a multiprotein complex with a double helicase-nuclease activity, recognizes and binds the DSB site, starting to process the double strand; (**c-d**) after the recognition of the Chi recombination hotspot, RecA is recruited on the 3' overhang generated, starting the homologous recombination process; (**e**) RecA also binds to the repressor LexA complexes, causing LexA autocleavage and consequent activation of SOS response genes [178].*

Many studies proved that some of the most used antibiotics, such as trimethoprim (DHFR inhibitor), β-lactams (cell wall biosynthesis inhibitors), aminoglycosides (protein synthesis inhibitors) and fluoroquinolones (DNA gyrase inhibitors) are potent inducers of SOS response, which, in turn, has a direct role in the development of acquired resistance to these drugs [198,201,202]. From the other point of view, the inactivation of the DSB repair mechanisms and of the SOS pathway proved to enhance/restore the sensibility to the common antibiotic treatments, even in the most feared "superbugs" (e.g.: *Escherichia coli, Pseudomonas aeruginosa, Staphylococcus aureus, and Mycobacterium tuberculosis*) [203-210].

In the light of these considerations, the bacterial DSB repair mechanisms and the SOS pathway, due to their key role in the acquisition of resistance, represent therefore an

appealing area of antibiotic discovery [208,211,212]. Indeed, a small molecule able to halt these processes could have the advantages both of control the spread of resistance and to prolong the shelf life of the current available antibiotics, with a decisive impact even on antibiotic therapy *in vivo*.

Among all the possible targets within the SOS pathway, the following discussion will be focused especially on the AddAB/RecBCD complex inhibitors. Indeed, to this class belong some of the most potent bacterial DNA repair inhibitors currently discovered.

6.2.3. IMP-1700: from the potent inhibition of the bacterial DNA repair pathways and to the issue of the quinolone core

Few examples of inhibitors of AddAB/RecBCD complexes have been reported in literature in the last decades, and ML328 (CID1517823) is one of the first members of this class (**Figure 28**) [213]. Identified through a cellular DNA-repair high throughput screen, this pipemidic acid derivative exhibited interesting inhibitory activity in biochemical assays but limited cellular potency. From a structural point of view, the features that may be related to biological activity are: a central heterocyclic aromatic scaffold, as the pipemidic one, which is able to intercalate in the DNA double strand; a cyclic amine (e.g. piperazine); an arylthiourea moiety substituted at 3-position with a CF_3 group [213].

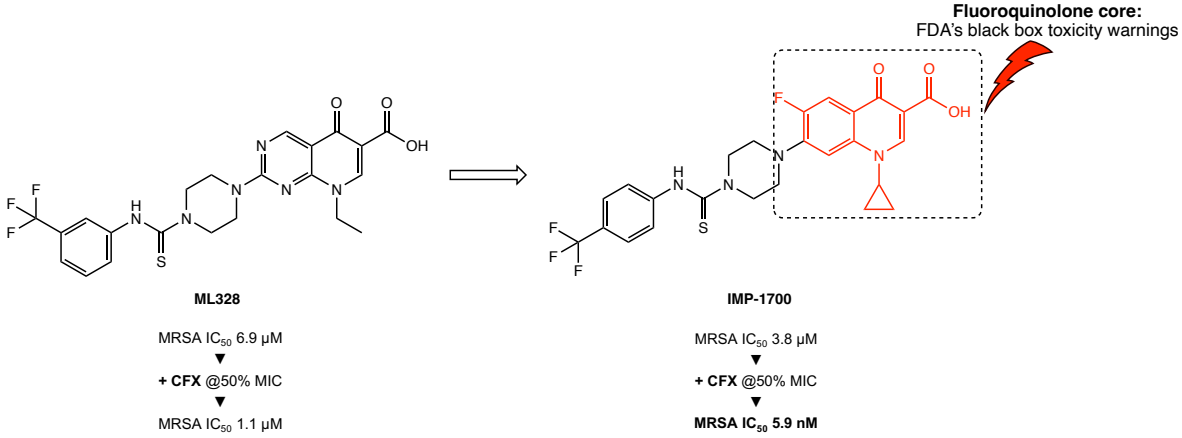


Figure 28. *Chemical structure of the pipemidic derivative ML238 and the corresponding optimized CFX derivative IMP-1700 [178,179,213].*

Starting from these assumptions, Lim et al. performed a SAR study to optimize both the AddAB/RecBCD inhibition and the antibacterial activities of the lead compound ML328 [179].

In detail, hypothesizing that an improved intercalation in the bacterial dsDNA could led to an improved inhibition of the AddAB/RecBCD complex, they substituted the pipemidic core with a ciprofloxacin (CFX) core (scaffold-hop), where the C6-fluorine and N1-cyclopropyl substituents modulate electron density to favor a better insertion in the DNA double strand. The obtained series of CFX derivatives was further optimized by guiding the substitution of the aryl moiety with the Topliss decision tree. Among all, IMP-1700, with a 4-CF_3 electron withdrawing group on the aryl thiourea moiety (**Figure 28**), resulted the most promising compound, as emerged from the biological assays.

In detail, from the MIC assay performed against the CFX-resistant MRSA strain USA300 JE2, IMP-1700 exhibited a ~650 increase in potency in the presence of CFX (EC_{50}=5.9 ± 0.6 nM) compared to its use as single agent (EC_{50}=3.8 ± 0.5 μM). This represents ~160-fold increase in potency compared to the lead compound ML328 (EC_{50}=1.1 ± 0.1 μM) for potentiation against JE2 by CFX. In addition, it resulted that IMP-1700 quantitatively synergized with CFX indicating a mechanism of action that amplifies the effect of each compound. To support the hypothesis that this synergism was due to interference on the DNA repair mechanisms, a SOS response assay and a RecBCD pulldown on lysate were conducted: IMP-1700 was capable to inhibit the SOS response in JE2 bacterial cells treated with CFX and to pull down recombinant AddAB from lysates [179]

Thus, even if the exact molecular target/ mechanism of action is still unclear, the acquired biological data strongly support a mechanism of inhibition of the AddAB/RecBCD complex, thus the SOS response pathway [179].

Despite the outstanding potency in biological assays, the fluoroquinolone nature of IMP-1700 (CFX core) is not ideal for a lead compound. Firstly, even if no appreciable inhibition of DNA gyrase and topoisomerase IV (fluoroquinolone targets) was observed, a sort of unwanted cipro-like activity (DNA damages, activation of SOS response), potentially due to hydrolysis or metabolism of the compound to CFX, was evident in the assays [179]. In addition, the fluoroquinolone family of antibiotics is associated with several serious adverse effects, which are typical of the class (so called “class effects”; e.g.: tendinitis and tendon rupture; headache; dizziness; liver failure; QT interval prolongation). In order to limit the

use of these drugs, the FDA added a black-box toxicity warning in their packages, to inform the patients about the risks associated to their use [214,215]. As a result, new therapeutic drugs are less likely to be approved if they contain a quinolone core.

In the light of these considerations, IMP-1700, because of its fluoroquinolone structure, is not an ideal drug candidate for further development.

6.3. Aim of the project

IMP-1700, as introduced previously, is one of the most potent inhibitors of bacterial DNA repair pathways published to date. Despite this, the FDA black box warnings (unwanted side effects) and potential off-target interactions (e.g: DNA gyrase inhibition) ascribed to its central CFX scaffold (4-oxo-quinoline-3-carboxylic acid), makes the progression to subsequent phases of drug development unlikely.

From a medicinal chemistry point of view, SAR studies on the central quinolone core have not been conducted [179], thus the critical question whether this part of the molecule is required/essential for the biological activity remains unanswered. The possibility to obtain new, non-CFX derivatives which maintain the potent SOS inhibitory activity, without the off-target inhibition of DNA gyrase, could represent an important starting point for the design of a new class of more selective compounds.

In the light of the above consideration, the main aim of this study was to conduct a SAR analysis on the bicyclic CFX scaffold of IMP-1700. In particular, the study consisted of the design, synthesis, and biological evaluation of new benzo-fused six membered heterocyclic analogues of the lead compound (**Figure 29**), in order to analyze whether the changes made on the central core have effects on the biological activity. On the other hand, the piperazine-thiourea-(4-trifluoromethyl) phenyl moiety at the C7 position of the central core, which appeared a key pharmacophoric portion from previous SAR studies [179], was maintained.

In details, the cores selected for this SAR study were:

- 4-quinazolinone core, reported in literature as a privileged scaffold in antibacterial drug design [216], and in some compounds active as antibiotic potentiator with SOS inhibitory activity [217]. More interestingly, one quinazoline analogue of IMP-1700 have been tested already by the research group, showing encouraging SOS inhibitory activity and lack of the off-target effects of the CFX core.
- 2,4-(1*H*,3*H*)-quinazolinedione core, which constitutes the central core of many compounds endowed with antibacterial activity, even against multidrug resistant bacteria (e.g. MRSA) [218,219].
- 4-quinolone core, to analyze how the lack of -COOH, affect the activity.

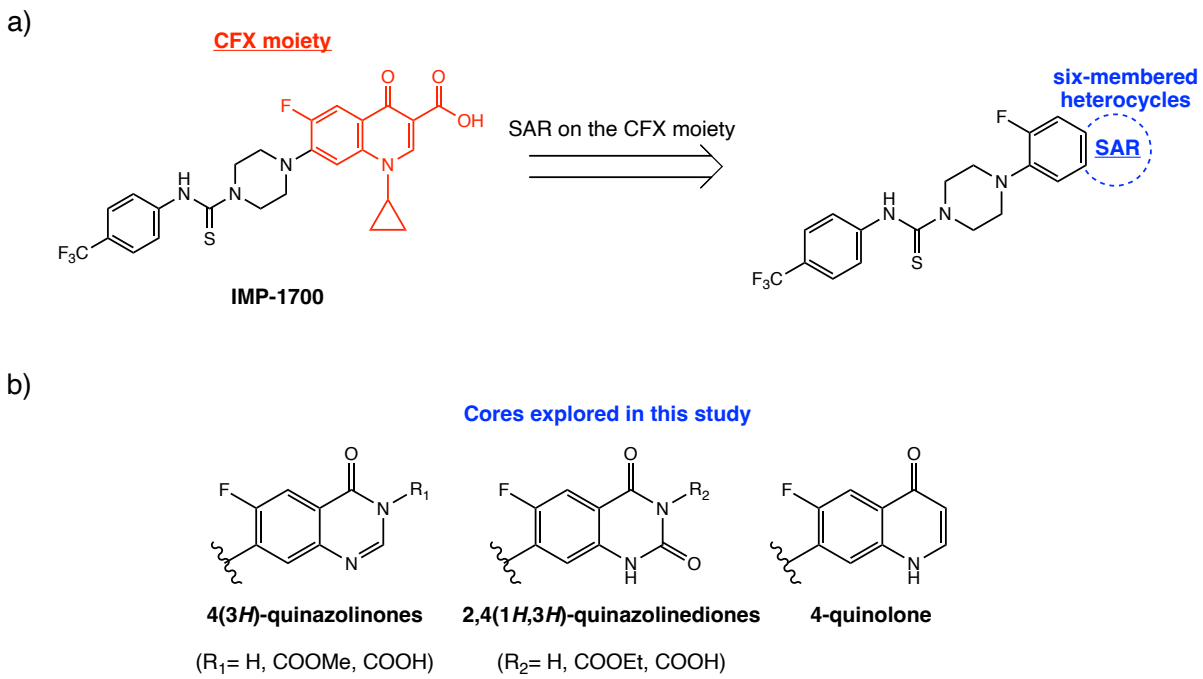
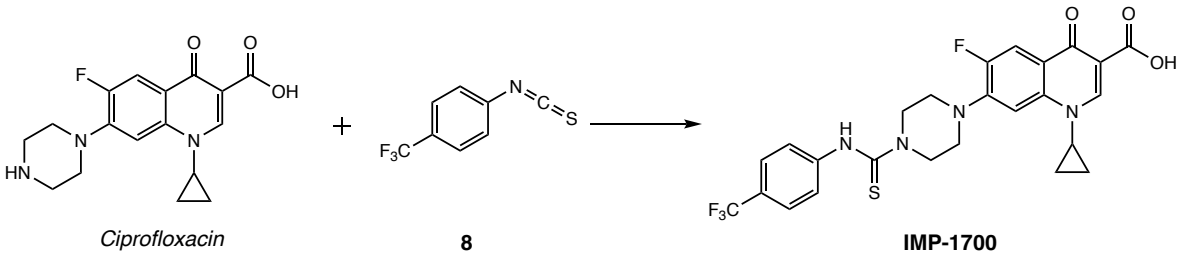


Figure 29. *(**a**) Aim of the project: is it possible to rationally substitute the CFX moiety with other cores without affecting the activity? General scheme of the SAR analysis conducted on the CFX core of IMP-1700; (**b**) General structures of the three benzo-fused six membered heterocyclic systems considered in this study: 4(3H)-quinazolinones; 2,4(1H,3H)-quinazolinediones; 4-quinolone.*

6.4. Results and discussion

6.4.1. Chemistry

As starting point, the synthesis of the lead compound IMP-1700 was considered (**Scheme 1**): this consisted of a one-step coupling reaction between ciprofloxacin (CFX) and 4-(trifluoromethyl) phenyl isothiocyanate **8**, using sodium hydrogen carbonate as base and dry DMF as solvent [179]. From a mechanistic point of view, the nucleophilic attack of the free secondary amine of the piperazine ring of the electrophilic isothiocyanate carbon leads to the formation of the required thiourea moiety.



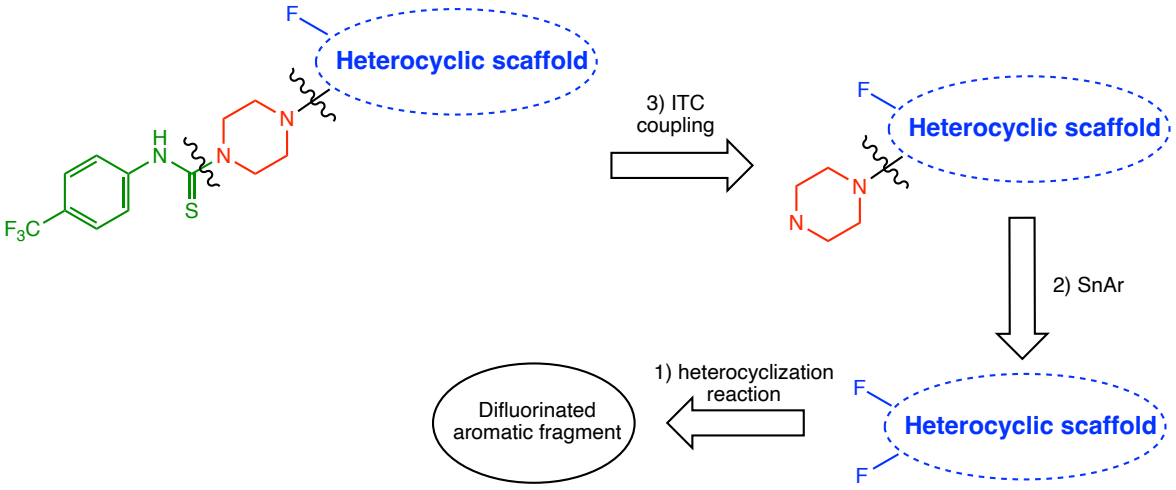
Scheme 1. *One-step coupling reaction used for the synthesis of the lead compound IMP-1700; Reagents and conditions: (i) 4-(trifluoromethyl) phenyl isothiocyanate **8**, $NaHCO_3$, DMF_{dry}, RT, 24h [179].*

Taking this into account, the first step consisted of planning the most convenient synthetic route to have access to the selected heterocycles analogues of IMP-1700.

In **Scheme 2**, the general retrosynthetic approach proposed for the synthesis of the required compounds is shown. In detail, it comprised three different steps:

1. an heterocyclization reaction to afford the required difluorinated scaffolds;
2. a selective nucleophilic aromatic substitution reaction (SnAr) at the C7 position of the heterocyclic core to introduce the piperazine fragment;
3. a coupling reaction between the 7-piperazynil derivative and the 4-(trifluoromethyl) phenyl isothiocyanate **8** to introduce the key thiourea moiety, as for the synthesis of IMP-1700 (see **Scheme 1**).

According to the explained retrosynthetic scheme, the detailed discussion of the synthetic pathways was divided in three different sections (6.4.1.1-6.4.1.3), each one describing one of the three steps. All the experimental details, including procedures and characterization of the synthetized products, are reported in the Materials and Methods section (paragraph 6.6)



Scheme 2. *Proposed retrosynthetic approach to afford the required heterocyclic analogues of IMP-1700.*

6.4.1.1. Synthesis of the 6-7-difluorinated heterocyclic scaffolds

In this section the heterocyclization reactions performed to obtain the required difluorinated cores (4-quinazolinones, 2,4(1*H*,3*H*)-quinazolinediones and 4-quinolone, depicted in **Figure 30**) are described in detail.

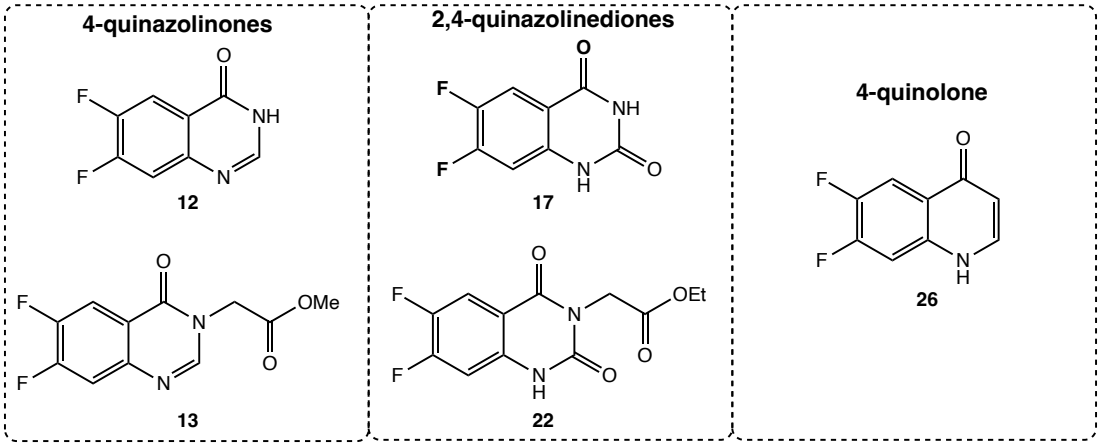


Figure 30. *Overview of the heterocyclic cores described in this work.*

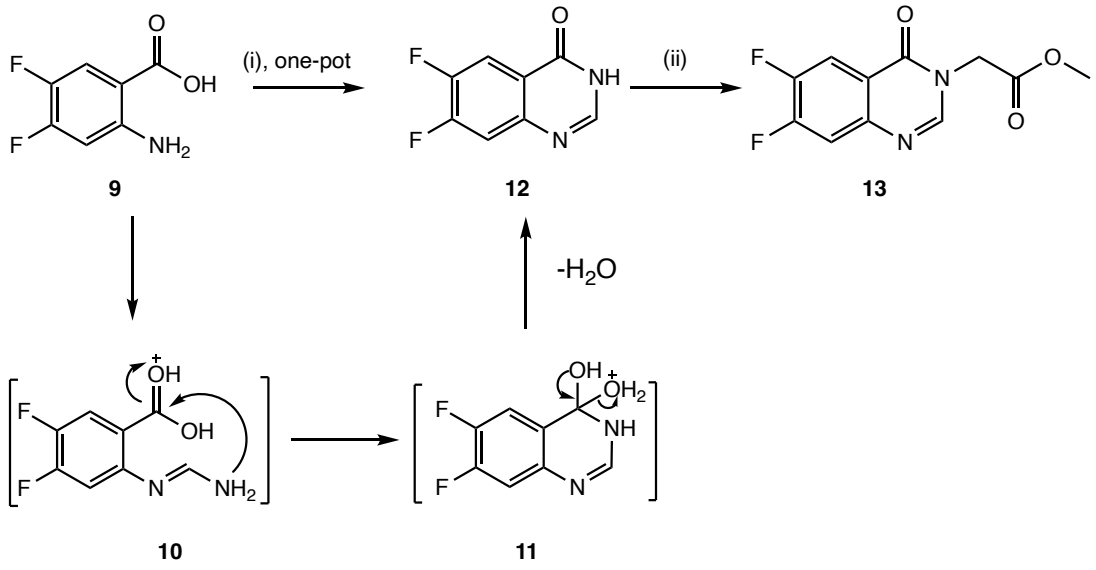
The most common synthetic approach for the synthesis of both 4-quinazolinones and of 2,4-quinazolinediones comprises a heterocyclization of a 2-anthranilic acid derivative with an appropriate reagent, such as an amide or a cyanate/isocyanate, respectively [220,221].

In the **Scheme 3**, the synthetic route applied for the synthesis of the two 4-quinazolinone cores **12** and **13** is reported. The 6,7-difluoroquinazolin-4(1H)-one intermediate **12**, was obtained with quantitative yields through a one-pot heterocyclization

reaction between the commercially available 2-amino-4,5-difluorobenzoic acid (**9**, anthranilic acid) and formamide [222]. In detail, the procedure involved the fusion of **9** and formamide at 125°C (solvent free), according to the Niementowski procedure, with an organic acid as catalyst (acetic acid) [220]. In the lower part of **Scheme 3**, the detailed mechanism of cyclization is reported: the NH_2 of the anthranilic acid attacks the carbonyl of formamide, forming the o-amidobenzamide intermediate **10**; this, under acid catalysis, readily undergoes a cyclization-dehydration process (**11**) to form the required quinazolinone **12** [220].

With the aim of introducing the carboxylate at 3-position of the ring, as in the lead compound IMP-1700, the N3 of the quinazolinone intermediate **12** was subjected to alkylation with methyl-2-bromoacetate, in the presence of K_2CO_3 as base and DMF_{dry} as solvent, to afford methyl 2-(6,7-difluoro-4-oxoquinazolin-3(4H)-yl)acetate **13** as pure off-white crystals in quantitative yields (adapted from literature procedure [223]).

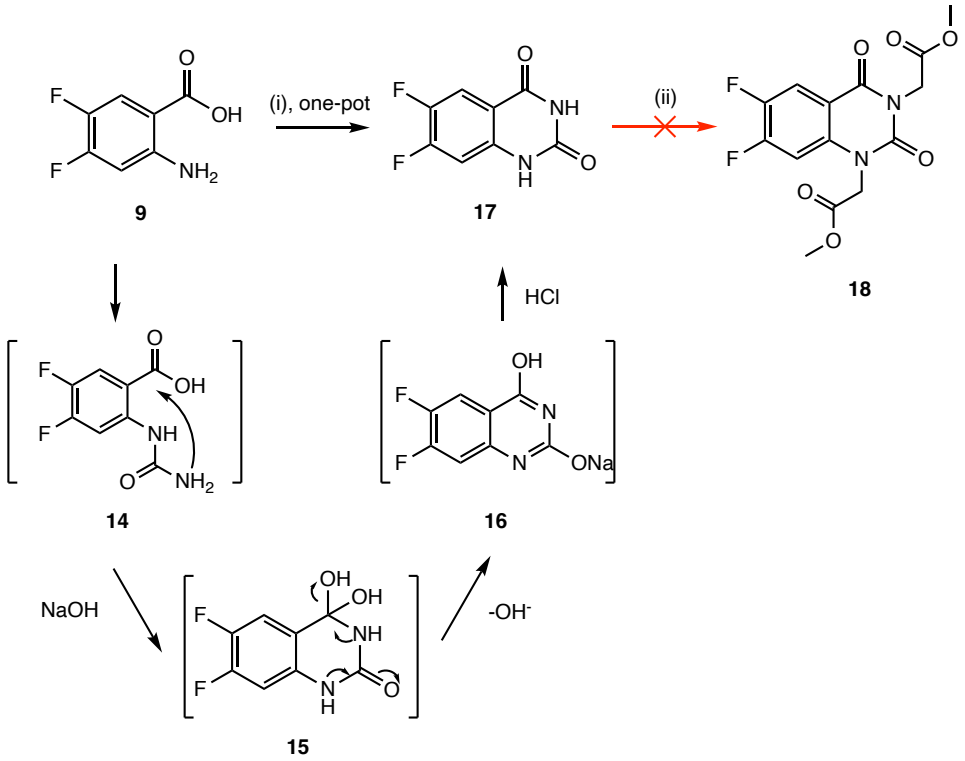
By observing the structure of **13** in comparison to the lead compound, a methylene bridge (CH_2) between the COOH and the heterocyclic moiety was added. Several attempts are reported in literature to directly bind a carboxy group to the N3 of quinazolines, but, due to the instability of the nitrogen-to-carboxyl bond, they were unsuccessful [224].



Scheme 3. *Synthesis of the 4-quinazolinone scaffolds **12**, **13**, with the detailed mechanism of cyclization. Reagents and conditions: (i) formamide, CH_3COOH, 125 °C, 24 h; (ii) methyl 2-bromoacetate, K_2CO_3, DMF_{dry}, 45-50 °C, 2 h.*

Regarding to the synthesis of the 2,4(1*H*,3*H*)-quinazolinedione core, the same starting material **9** was treated with potassium cyanate in water under acid catalysis, following a procedure reported in literature, to afford the intermediate **17** with good yields (**Scheme 4**) [225]. As shown in the lower part of **Scheme 4**, the mechanism of this one-pot cyclization reaction comprised three steps: nucleophilic attack of the NH_2 on the electrophilic carbon of KOCN to form the urea intermediate **14**; cyclization under basic conditions (NaOH) to obtain derivative **16** as sodium salt (enol form); treatment with HCl to give the required diketo product **17**.

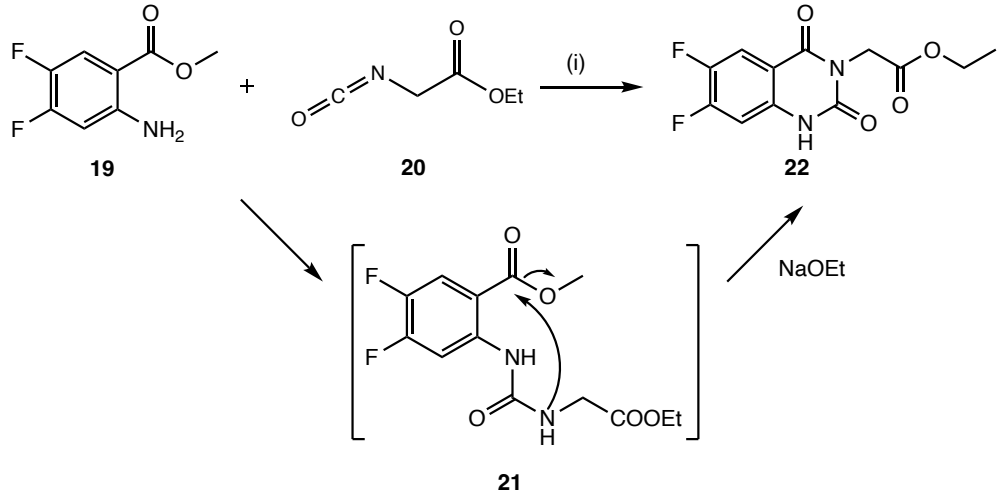
With the aim of introducing the N3 carboxymethyl portion, as seen for the synthesis of compound **13**, the selective alkylation of the N3 with methyl 2-bromoacetate was attempted. As the N3 between the two electron-withdrawing keto groups, was the most acidic, it was most prone to activation by the potassium carbonate. Unfortunately, even with a stochiometric amount of alkylating agent, both the nitrogen atoms reacted, leading to a mixture of the starting material and of the dialkylated product **18** (experimental data not shown).



Scheme 4. *Synthesis of the 2,4-quinazolinedione core **17**. Reagents and conditions: (i) KOCN, CH_3COOH, water, RT, 24 h, then NaOH and HCl; (ii) methyl 2-bromoacetate, K_2CO_3, DMF_{dry}, 45-50 °C, 2 h.*

Thus, to overcome this issue, the alternative synthetic strategy reported in **Scheme 5** was performed. In detail, methyl 2-amino-4,5-difluorobenzoate (**19**) and ethyl isocyanatoacetate (**20**), both commercially available, were reacted in pyridine to afford the key intermediate ethyl 2-(6,7-difluoro-2,4-dioxo-1,4-dihydroquinazolin-3(2*H*)-yl)acetate **22** with excellent yields (91%). The use of the methyl ester **19** instead of the carboxylic acid was justified by its higher reactivity and its higher solubility in the organic solvent used. A similar literature procedure was applied for the synthesis of several 2,4-dioxothieno[2,3-*d*]pyrimidine derivatives [226].

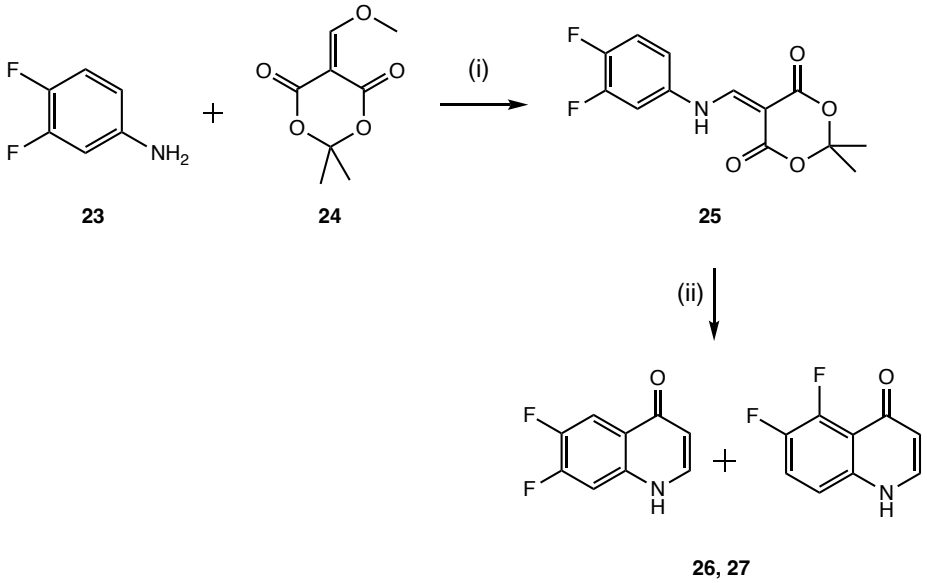
The cyclization mechanism proposed is very similar to the one reported for the synthesis of **17**: initial formation of the urea adduct **21**, with consequent ring closure under basic conditions. In this case, instead of a hydroxide base, the alkoxide NaOEt was used: this permitted to afford the cyclized product without the undesired hydrolysis of the ethyl ester.



Scheme 5. *Synthesis of the ethyl 2-(6,7-difluoro-2,4-dioxo-1,4-dihydroquinazolin-3(2H)-yl)acetate* ***22****. Reagents and conditions: (i) ethyl isocyanatoacetate, pyridine, 45 °C, 2h; then add NaOEt 21% in EtOH; then neutralize with HCl 2N.*

The two ester derivatives **13** and **22**, designed to mimic the 3-COOH group of the central core of lead compound, were subjected to hydrolysis to obtain the corresponding carboxylic acids. Nevertheless, the presence of a free COOH impeded the following SnAr (formation of a salt with the piperazine); thus, it appeared more convenient to hydrolyze the ester functions after the SnAr reaction (see next section).

In the **Scheme 6** the synthetic route to obtain the required decarboxylated 4-quinolone core is shown. In details, the first reaction consisted of the condensation between the 3,4-difluoroaniline **23** and 5-methoxymethylene-2,2-dimethyl-[1,3]dioxa,n-4,6-dione **24** in 2-propanol to obtain the intermediate **25** as a pale-yellow solid [227]. The subsequent thermic intramolecular cyclization was performed at 220°C (a eutectic mixture of diphenyl ether – biphenyl was used as heating media) and led to formation of the two difluoro regioisomers **26** (6,7-difluoroquinol-4-one, 80%), **27** (5,6-difluoroquinol-4-one, 20%). Having the same polarity (almost identical R_f in TLC), they were impossible to separate by flash column chromatography. Hypothesizing a different reactivity towards the nucleophilic attack of piperazine, the mixture was used in the following step without further purification (see next section).



Scheme 6. *Synthesis of the decarboxylated quinolone cores **26**, **27**. Reagents and conditions: (i) 2-propanol, 70 °C, 1 h; (ii) diphenyl ether – biphenyl (eutectic ratio), 220 °C, 2 h.*

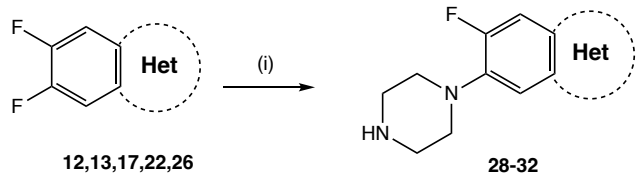
6.4.1.2. Synthesis of the piperazinyl derivatives 28-35

By following the retrosynthetic **Scheme 2**, the obtained difluoro heterocycles **12**, **13**, **17**, **22**, **26/27** were then subjected to the nucleophilic aromatic substitution reaction (SnAr) to afford selectively the 7-piperazinyl derivatives **28-32** (**Scheme 7a**).

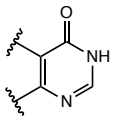
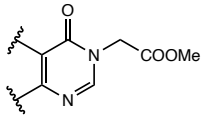
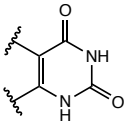
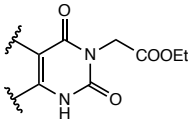
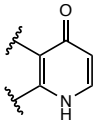
In addition, in order to obtain another compound for biological assays, also the piperazinyl derivative of the starting material **19** was prepared (compound **33**, **Scheme 7b**).

A large excess of piperazine (3-4 eq.) and a moderate heating (80°C, b.p. acetonitrile) were necessary to drive the reaction to completion. All the compounds were purified by silica gel flash column chromatography with a mixture DCM-MeOH (5-20%) + 1% triethylamine.

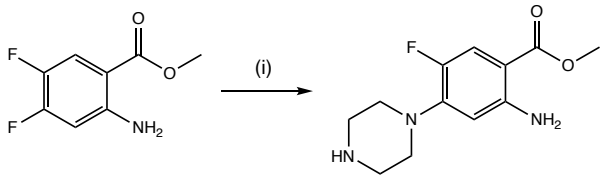
a)



12,13,17,22,26 28-32

Compound	12, 28	13, 29	17, 30	22, 31	26, 32
Het					

b)



19 33

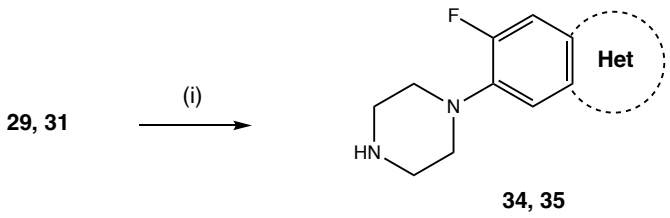
Scheme 7. *(**a**) Synthesis of the piperazinyl derivatives **28-32**; (**b**) Synthesis of the piperazinyl derivatives **33**. Reagents and conditions: (i) piperazine, MeCN, 80°C, overnight.*

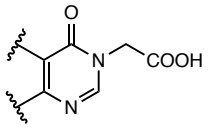
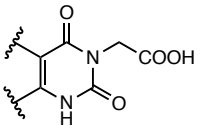
From a mechanistic point of view, the selective displacement of the 7-*F* over the 6-*F* was explained by the electronic effects of the neighboring functional groups: the presence of an electron withdrawing carbonyl group in para to the 7-*F* was responsible of an increased electrophilicity of the C-7, thus increasing reactivity towards the SnAr. This evidence was further confirmed by both ^{1}HNMR (H-5, more deshielded compared to H-8, has a stronger coupling with fluorine) and ^{13}CNMR (long range coupling between the C-4 quaternary carbonyl group and the C-6 fluorine, $^{4}J_{C\text{-}F}$ = 2.5 Hz) and previous works, too [228,229].

This higher reactivity of the C-7 position compared to the C-5/6 allowed also explanation of the different reactivity of the two regioisomers **26**, **27** towards the nucleophilic attack of piperazine: **27**, with a 5,6-difluoro substitution (m/o position to the

carbonyl group), was completely unreactive towards the SnAr, and was easily separated from the required regioisomer piperazinyl derivative **32** by column chromatography.

Derivatives **29** and **31** were subjected to hydrolyzed successfully affording the compounds **34**, **35** (**Scheme 8**). Despite the very mild conditions (use of LiOH, weaker then NaOH; room temperature), the reaction proceeded quickly to completion (30 min). The hydrolyzed compounds, given their zwitterionic nature, were completely water soluble and were precipitated as salts by carefully adjusting the pH to 7 with HCl. The collected white solids were directly used in the next step without purification.

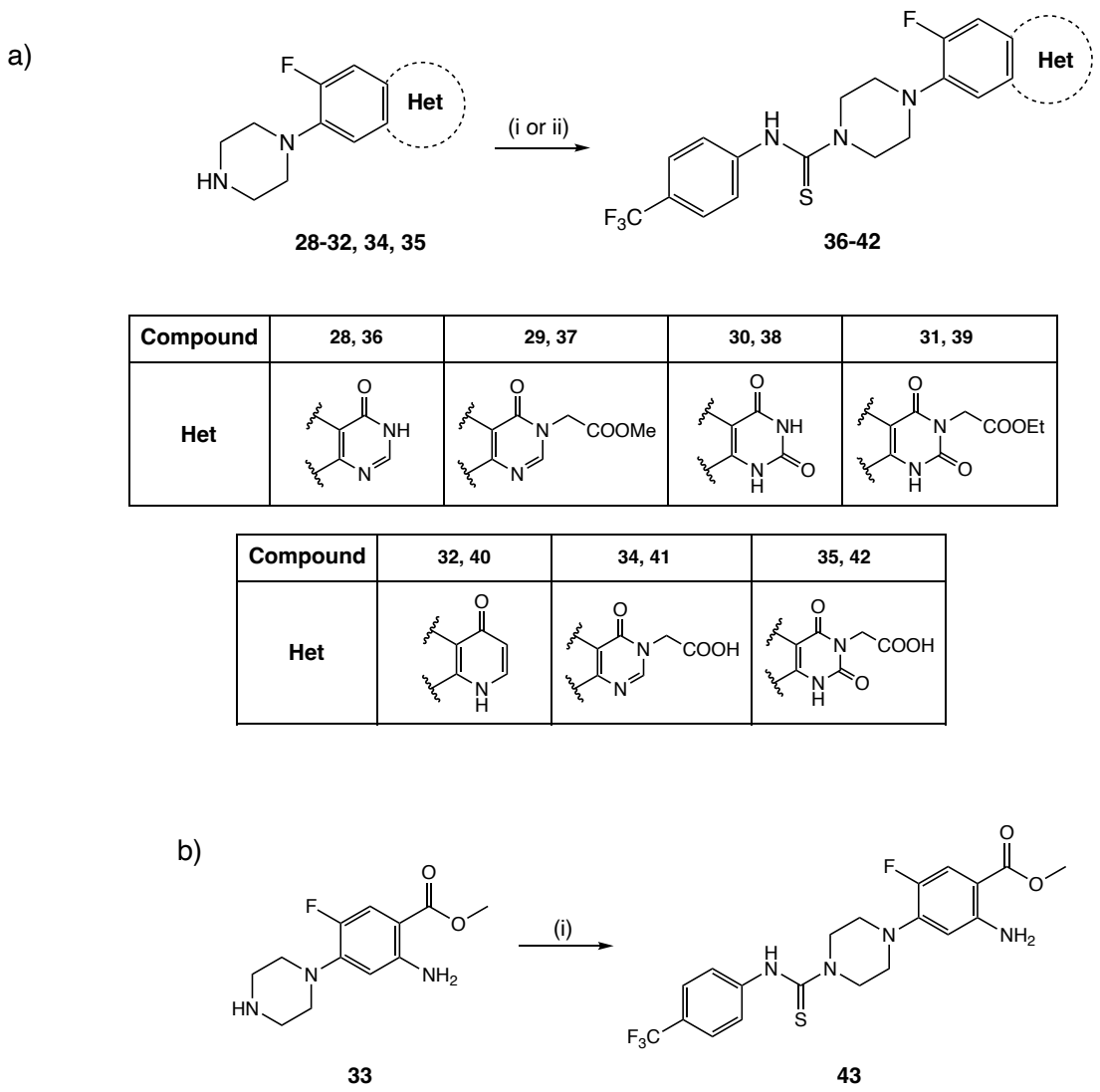


Compound	34	35
Het		

Scheme 8. *Synthesis of the carboxyl acid derivatives* ***34, 35****. Reagents and conditions: (i) LiOH, water, RT, 30 min.*

6.4.1.3. Synthesis of piperazine-thioureatrifluoromethylphenyl compounds 36-43, analogues of IMP-1700

Analogously to the synthesis of IMP-1700, the last synthetic step was formation of the piperazine-thiourea compounds **36-43** *via* a coupling reaction between the piperazinyl derivatives **28-35** and the 4-(trifluoromethyl)phenyl isothiocyanate **8** (**Scheme 9a,b**), under strictly dry (anhydrous DMF was used as solvent) and inert (argon) conditions.



Scheme 9. *(**a**) Synthesis of the piperazine-thioureatrifluoromethylphenyl derivatives **36-42**; (**b**) Synthesis of the piperazine-thioureatrifluoromethylphenyl derivatives **43**. Reagents and conditions: (i) for compounds **36-40**, **42**, **43**, 4-(trifluoromethyl) phenyl isothiocyanate, DMF_{dry}, RT, overnight; (ii) for compound **41**, 4-(trifluoromethyl) phenyl isothiocyanate, base (Cs_2CO_3 or DIPEA), DMF_{dry}, RT, overnight.*

In the case of derivative **41**, the addition of a base (Cs_2CO_3 or DIPEA) was necessary to activate the amino group (likely present as an HCl salt) of the piperazine ring. Unfortunately, the instability of the 4-(trifluoromethyl)phenyl isothiocyanate under basic conditions led to the formation of multiple side products, which remained in mixture with the required compound, even after two flash column chromatography purifications. For this reason, it was not possible to test **41** in the biological assays.

On the other hand, compounds **36-40** and **42,43** were shown to have purity >95% by HPLC-UV and were therefore progressed to biological testing.

6.5. Conclusions

AMR represents one of the most serious challenges threatening global health during this century. Several approaches have been undertaken to tackle it and, among all, one of the most interesting strategies could be to improve/prolong the efficacy of existing antimicrobial drugs by directly countering bacterial mechanisms of drug resistance. Inhibition of the bacterial DNA repair mechanisms is one potential method to achieve this goal.

IMP-1700 is one of the most potent inhibitors of bacterial DNA repair mechanism developed to date. However, its CFX core, responsible for unwanted cipro-like activity (DNA damages, activation of SOS response) and subject to an FDA warning for "black-box toxicity", is not ideal for a lead compound and makes the progression to the subsequent phase of drug development unlikely.

The main aim of this project was to conduct a SAR analysis on the bicyclic CFX scaffold of IMP-1700, to understand whether this part of the molecule is required to maintain the biological activity; on the other hand, no changes were made on the piperazine-thiourea-(4-trifluoromethyl) phenyl moiety. In particular, in this study new benzo-fused six membered heterocyclic analogues (4-quinazolinones; 2,4-quinazolinediones; 4-quinolone) of the lead compound (compounds **36-43**) have been designed and successfully synthetized.

Seven of them (compounds **36-40** and **42,43**) were isolated with appropriate purity and progressed to biological testing, which are still in progress. The results of the MRSA assay (CFX potentiation), as well as the SOS response and RecBCD inhibition assays will show whether the changes made on the IMP-1700 structure allows to maintain the potent SOS inhibitory activity, without the unwanted side effects. This could represent an important starting point for the design of a new class of more selective compounds.

6.6. Materials and Methods

6.6.1. Chemistry

Unless otherwise indicated, all reagents and solvents were purchased from commercial sources (Merck, Alfa Aesar, Thermo Fisher and Fluorochem) and used without further purification. Water was purified by a Suez® Select Fusion system. Organic layers were dried over $MgSO_4$. *In vacuo* refers to the use of a rotatory evaporator attached to either a rotary vane or diaphragm pump. Thin layer chromatography (TLC) was performed on Merck aluminium plates coated with 60 F254 silica and were visualized using UV light (254 or 364 nm) or by ninhydrin. Flash column chromatography was performed on a Teledyne ISCO Combiflash® Rf+ using Redisep® standard prepacked cartridge system.

^1H NMR, ^{13}C NMR spectra were recorded on Bruker Advance spectrometers (Bruker AVIIIHD 400 nanobay), at 400 and 101 MHz respectively, in the specified deuterated solvent. Chemical shifts (δ) are reported in parts per million (ppm) and referred as the internal standard to the residual solvent peaks; the coupling constants J are expressed in Hz and were measured through MestReNova software. Due to the poor solubility, the carbon assignments for the SnAr derivatives were not possible. Compound peaks overlapped with water peaks were assigned using COSY NMR experiments. The NMR of the most representative compounds are reported in the **Supplementary Figures S7-19**. The following abbreviations are used: br s = broad signal, s = singlet, d = doublet, dd = double doublet, t = triplet, q = quartet, m = multiplet, RT = room temperature.

High resolution mass spectra (HRMS) were recorded on a Waters BioAccord system, from solutions of MeOH, water or mixture. The purity of the tested compounds was determined to be >95% by analytical reverse phase HPLC performed with a dual-pump LC-20AD system equipped with an Atlantis dC18 Column (100Å, 5 µm, 4.6 mm X 150 mm) and eluted with a gradient 20-100% MeCN in water (15 min run), using Shimidzu UV/VIS detector SPD-20A (selected wavelengths: 254nm, 280 nm).

Starting materials 2-amino-4,5-difluobenzoic acid (**9**), methyl 2-amino-4,5-difluorobenzoate (**19**) and 3,4-difluoaniline (**23**) were commercially available. The intermediated **12**, **17**, **26**/**27** were prepared following previously reported procedures, as

indicated in the corresponding sections; the spectra data determined were consistent with the ones reported in literature.

6.6.1.1. Synthesis of the difluorinated heterocyclic scaffolds

6,7-difluoroquinazolin-4(1H)-one (**12**)

2-Amino-4,5-difluorobenzoic acid **9** (1.0 g, 5.8 mmol), glacial acetic acid (2.5 mL), and formamide (9 mL) were mixed at 125 °C for 24h. The mixture was cooled to RT and ice water (15 mL) was added. The precipitate was filtered, collected, and dried under vacuum to obtain **12** as a white solid (950 mg, 90%). ^{1}H NMR (400 MHz, DMSO-d_6) δ: 12.48 (br s, 1H, NH), 8.14 (s, 1H, H_2), 8.02 (dd, J = 10.5, 8.8 Hz, 1H, H_5 or H_8), 7.74 (dd, J = 11.4, 7.3 Hz, 1H, H_5 or H_8) [222].

Methyl 2-(6,7-difluoro-4-oxoquinazolin-3(4H)-yl)acetate (**13**)

To a suspension of 6,7-difluoroquinazolin-4(3H)-one **12** (1 mmol) in anhydrous DMF (10 mL), K_2CO_3 (1.5 mmol) and methyl 2-bromoacetate (1.1 mmol) were added. The mixture was heated at 50°C for 2 h, then allowed to cool to RT. $LiCl_{(aq)}$ 5% (30 mL) was added, and the product extracted with EtOAc (3x25mL), dried, and the solvent removed *in vacuo* to give **13** as a light brown crystal. Yield: 95%. ^{1}H NMR (400 MHz, DMSO-d_6) δ: 8.43 (s, 1H, H_2), 8.08 (dd, J = 10.4, 8.6 Hz, 1H, H_5 or H_8), 7.83 (dd, J = 11.3, 7.3 Hz, 1H, H_5 or H_8), 4.86 (s, 2H, CH_2), 3.72 (s, 3H, -OCH_3). ^{13}C NMR (101 MHz, DMSO-d_6) δ: 168.67, 159.37 (d, J_{C-F} = 2.8 Hz), 154.44 (dd, J_{C-F} = 254.7, 14.6 Hz), 149.39 (dd, J_{C-F} = 249.1, 13.9 Hz), 149.30 (d, J_{C-F} = 2.2 Hz), 146.48 (dd, J_{C-F} = 11.4, 2.3 Hz), 119.03 (dd, J_{C-F} = 6.8, 1.8 Hz), 116.01 (d, J_{C-F} = 17.8 Hz), 114.18 (dd, J_{C-F} = 19.0, 2.1 Hz), 52.96, 47.74. HRMS (ESI+) m/z calculated for $C_{11}H_9F_2N_2O_3$ 255.0576, found 255.0574.

6,7-difluoroquinazoline-2,4(1H,3H)-dione (**17**)

KOCN (5.22 mmol) in water (5 mL) was added to a solution of 2-amino-4,5-difluorobenzoic acid **9** (4.04 mmol) in water/AcOH (18 mL/0.25 mL). The mixture was stirred overnight at RT. NaOH (0.092 mol) was added slowly in portions first forming a clear purple solution, and then precipitation. After stirring for 10 mins the precipitate was filtered and resuspended in H_2O, acidified to pH 4 with 4N HCl, and stirred for another minute. The precipitate was

filtered, washed with cold water, and dried to give **17** as a white powder (Yield: 61 %). ^{1}H NMR (400 MHz, DMSO-d_6) δ 11.48 (br s, 1H, NH), 11.27 (br s, 1H, NH), 7.83 (td, *J* = 9.6, 2.3 Hz, 1H, H_8), 7.15 – 7.05 (m, 1H, H_5) [225].

Ethyl 2-(6,7-difluoro-2,4-dioxo-1,4-dihydroquinazolin-3(2H)-yl)acetate (**22**)

To a suspension of methyl-2-amino-4,5-difluorobenzoate **19** (4.27 mmol) in pyridine (5 mL), was added ethyl isocyanatoacetate **20** (0.74 mL, 1.55 mmol) dropwise. The reaction mixture was stirred at 50°C for 5 h then allowed to cool to RT. The solvent was removed *in vacuo*, with the residue resuspended in EtOH and NaOEt 21% w/w in EtOH (3.2 mL, 2 eq. of sodium ethoxide) added. After stirring for 1h at room temperature, the mixture was slowly neutralized with HCl 2N in an ice bath. The volatiles were removed *in vacuo* and the resulting solid was collected by filtration, washed thoroughly with water and EtOH and dried under vacuum to obtain ethyl (**22**) as a fluffy orange solid. Yield: 91%. ^{1}H NMR (400 MHz, DMSO-d_6) δ: 11.76 (br s, 1H, NH), 7.92 (dd, *J* = 10.2, 8.4 Hz, 1H, H_8), 7.17 (dd, *J* = 10.9, 6.6 Hz, 1H, H_5), 4.63 (s, 2H, CH_2), 4.15 (q, *J* = 7.1 Hz, 2H, -O$\underline{CH_2}$$CH_3$), 1.20 (t, *J* = 7.1 Hz, 3H, -$OCH_2\underline{CH_3}$). ^{13}C NMR (101 MHz, DMSO-d_6) δ: 168.28, 160.71 (d, J_{C-F} = 2.4 Hz), 154.50 (dd, J_{C-F} = 254.9, 14.6 Hz), 149.97, 146.21 (dd, J_{C-F} = 243.7, 13.9 Hz), 137.53 (d, J_{C-F} = 11.2 Hz), 116.16 (d, J_{C-F} = 19.2 Hz), 110.48 (d, J_{C-F} = 4.2 Hz), 104.81 (d, J_{C-F} = 21.9 Hz), 61.63, 41.93, 14.46. HRMS (ESI+) calculated for $C_{12}H_{10}F_2N_2O_4$ $[M+H]^+$ 285.0681, found 285.0677.

6,7-difluoroquinolin-4(1H)-one (**26**)*/5,6-difluoroquinolin-4(1H)-one* (**27**)

3,4-Difluoroaniline (**23**, 1.55 mmol) and 5-methoxymethylene-2,2-dimethyl-[1,3]-dioxa,n-4,6-dione (**24**, 1.565 mmol) were dissolved in 2-propanol (5 ml), and the solution was stirred at 70°C for 30 min. The reaction mixture was filtered and was washed with methanol and ether, and dried to obtain a yellow solid. This was combined with biphenyl (5.8 g) and diphenyl ether (20 ml) and the suspension was stirred at 220 °C for 1h. The reaction mixture was filtered and washed with $CHCl_3$, with the resulting residue containing a mixture of the two regioisomers **26** and **27** and used in the next reaction without purification [227].

6.6.1.2. Synthesis of the 7-piperazinyl derivatives

Genaral procedure for Piperazine SnAr:

Piperazine (4 eq.) was added to a solution of difluoro heteroyclic compound (1 eq.) in anhydrous MeCN. The reaction was stirred at 80 °C overnight. The solvent was removed *in vacuo* and the residue was purified via flash silica column chromatography DCM-MeOH (5-25%) with 1% Et_3N.

6-fluoro-7-(piperazin-1-yl)quinazolin-4(1H)-one (**28**)

Prepared following the general procedure from the difluoro heterocyclic intermediate **12**. Yield: 89%. ^{1}H NMR (400 MHz, DMSO-d_6) δ: 8.02 (s, 1H, H_2), 7.67 (d, *J* = 13.2 Hz, 1H, H_5), 7.10 (d, *J* = 8.1 Hz, 1H, H_8), 3.11 (m, 4H, 2xCH_2, $H_{2'}$, $H_{-6'}$), 2.88 (s, 4H, 2xCH_2, $H_{3'}$, $H_{5'}$), NHs missing. HRMS (ESI+) m/z calculated for $C_{12}H_{14}FN_4O$ 249.1146 $[M+H]^+$, found 249.1146.

Methyl 2-(6-fluoro-4-oxo-7-(piperazin-1-yl)quinazolin-3(4H)-yl)acetate (**29**)

Prepared following the general procedure from the difluoro heterocyclic intermediate **13**. Yield 75%. ^{1}H NMR (400 MHz, DMSO-d_6) δ: 8.32 (s, 1H, H_2), 7.73 (d, *J* = 13.1 Hz, 1H, H_5), 7.20 (d, *J* = 8.0 Hz, 1H, H_8), 4.82 (s, 2H, CH_2), 3.71 (s, 3H, -OCH_3), 3.33 – 3.26 (m, 4H, 2xCH_2, $H_{2'}$, $H_{-6'}$), 3.11 – 3.04 (m, 4H, 2xCH_2, $H_{3'}$, $H_{5'}$), NH missing. HRMS (ESI+) m/z calculated for $C_{15}H_{18}FN_4O_3$ $[M+H]^+$ 321.1357, found 321.1359.

6-fluoro-7-(piperazin-1-yl)quinazoline-2,4(1H,3H)-dione (**30**)

Prepared following the general procedure from the difluoro heterocyclic intermediate **17**. Yield 60%. ^{1}H NMR (400 MHz, DMSO-d_6) δ: 7.48 (d, *J* = 13.0 Hz, 1H, H_5), 6.68 (d, *J* = 7.3 Hz, 1H, H_8), 3.19 – 3.12 (m, 4H, 2xCH_2, $H_{2'}$, $H_{-6'}$), 3.02 – 2.95 (m, 4H, 2xCH_2, $H_{3'}$, $H_{5'}$), NHs missing. HRMS (ESI+) m/z calculated for $C_{12}H_{13}FN_4O_2$ $[M+H]^+$ 264.1022, found 264.1021.

Ethyl 2-(6-fluoro-2,4-dioxo-7-(piperazin-1-yl)-1,4-dihydroquinazolin-3(2H)-yl)acetate (**31**)

Prepared following the general procedure from the difluoro heterocyclic intermediate **22**. Yield: 92%. ^{1}H NMR (400 MHz, DMSO-d_6) δ: 7.54 (d, *J* = 13.0 Hz, 1H, H_5), 6.71 (d, *J* = 7.3 Hz, 1H, H_8), 4.61 (s, 2H, CH_2), 4.13 (q, *J* = 7.1 Hz, 2H, -O$\underline{CH_2}CH_3$), 3.23 – 3.17 (m, 4H, 2xCH_2, $H_{2'}$,

$H_{-6'}$), 3.04 – 2.97 (m, 4H, 2xCH_2, $H_{3'}$, $H_{5'}$), 1.20 (t, *J* = 7.1 Hz, 3H, -$OCH_2\underline{CH_3}$), NHs missing. HRMS (ESI+) m/z calculated for $C_{16}H_{19}FN_4O_4$ $[M+H]^+$ 351.1463, found 351.1478.

6-fluoro-7-(piperazin-1-yl)quinolin-4(1H)-one (**32**)

Prepared following the general procedure from the two difluoro heterocyclic intermediate **26**/**27**. Yield: 68%. ^{1}H NMR (400 MHz, DMSO-d_6) δ: 7.83 (d, J = 7.2 Hz, 1H, H_1), 7.62 (d, J = 13.7, 1H, H_5), 6.99 (t, J = 9.1 Hz, 1H, H_8), 5.93 (d, J = 7.2 Hz, 1H, H_3), , 3.03 (m 4H, 2xCH_2, $H_{2'}$, $H_{-6'}$), 2.87 (m, 4H, 2xCH_2, $H_{3'}$, $H_{5'}$), NHs missing.

Methyl 2-amino-5-fluoro-4-(piperazin-1-yl)benzoate (**33**)

Prepared following the general procedure from the difluoro starting material **19**. Yield: 74%. ^{1}H NMR (400 MHz, DMSO-d_6) δ: 7.28 (d, J = 14.7 Hz, 1H, H_6), 6.49 (br s, 2H, NH_2), 6.28 (d, J = 8.0 Hz, 1H, H_3), 3.73 (s, 3H, -OCH_3), 3.01 – 2.94 (m, 4H, 2xCH_2, $H_{2'}$, $H_{-6'}$), 2.85 – 2.79 (m, 4H, 2xCH_2, $H_{3'}$, $H_{5'}$).

6.6.1.3. General procedure for the ester hydrolysis (synthesis of compounds 34, 35)

To a suspension of the appropriate ester (**29** or **31**, 1 eq.) in water was added LiOH (2.5 eq.) and the mixture was stirred at RT. After 30 min, the mixture was slowly neutralized to pH 6-7 with 2N $HCl_{(aq)}$ and filtered under vacuum. The obtained white precipitate was dried under vacuum and used for the next step without purification.

6.6.1.4. Synthesis of the piperazine-thioureatrifluoromethylphenyl derivatives 36-43 (Isothiocyanate coupling reaction)

- **General Procedure A**

To suspension of piperazine derivative (1 eq.) in anhydrous DMF was added 4-(trifluoromethyl)phenyl isothiocyanate **8** (1.1 eq.) and the resulting mixture was allowed to stir at RT overnight under argon. The solvent was removed *in vacuo* and the residue was washed with LiCl (aq) 5% w/v and then with Et_2O. The crude residue was then purified by flash silica column chromatography.

- **General Procedure B**

To suspension of piperazine derivative (1 eq.) in anhydrous DMF was added 4-(trifluoromethyl) phenyl isothiocyanate **8** (1.1 eq.), and base (1 eq). The resulting mixture was allowed to stir at RT overnight under argon. The solvent was removed *in vacuo* and the residue was washed with $NH_4Cl_{(aq)}$, $LiCl_{(aq)}$ 5% w/v, and then with Et_2O. The crude residue was then purified by flash silica column chromatography.

- **General Procedure C**

To suspension of piperazine derivative (1 eq.) in anhydrous DMF was added 4-(trifluoromethyl) phenyl isothiocyanate **8** (1.1 eq.). The resulting mixture was allowed to stir at RT overnight under argon. The solvent was removed *in vacuo* and the residue was washed with Et_2O, $LiCl_{(aq)}$ 5% w/v, and then 1N HCl. There were no further purification steps.

4-(6-fluoro-4-oxo-1,4-dihydroquinazolin-7-yl)-N-(4-(trifluoromethyl)phenyl)piperazine-1-carbothioamide (**36**)

Prepared following General Procedure A from intermediate **28**. Purification via flash silica column chromatography DCM-MeOH (2-15%). Yield 83%. ^{1}H NMR (400 MHz, DMSO-d_6) δ: 12.19 (br s, 1H, NH), 9.79 (s, 1H, H_{SCN-H}), 8.04 (s, 1H, H_2), 7.72 (d, J = 12.9 Hz, 1H, H_5), 7.65 (d, J = 8.4 Hz, 2H, $H_{2''}$, $H_{6''}$), 7.59 (d, J = 8.4 Hz, 2H, $H_{3''}$, $H_{5''}$), 7.17 (d, J = 8.0 Hz, 1H, H_8), 4.12 (t, J = 5.0 Hz, 4H, 2xCH_2, $H_{3'}$, $H_{5'}$), 3.35 (t, J = 4.9 Hz, 4H, 2xCH_2, $H_{2'}$, $H_{6'}$). ^{13}C NMR (101 MHz, DMSO-d_6) δ: 181.79, 160.20 (d, $^4J_{C-F}$ = 2.9 Hz), 153.57 (d, $^1J_{C-F}$ = 247.2 Hz), 147.35, 145.72, 145.52 (d, $^2J_{C-F}$ = 9.76 Hz), 145.37 (d, $^4J_{C-F}$ = 1.33 Hz), 125.55 (q, $^3J_{C-F}$ = 3.84 Hz), 124.93 (q, $^1J_{C-F}$ = 271.21 Hz), 124.68, 124.17 (q, $^2J_{C-F}$ = 32.13 Hz), 116.49 (d, $^3J_{C-F}$ = 8.2 Hz), 115.99 (d, $^3J_{C-F}$ = 3.1 Hz), 111.78 (d, $^2J_{C-F}$ = 23.0 Hz), 49.49 (d, $^4J_{C-F}$ = 4.3 Hz), 48.38. HRMS (ESI+) m/z calculated for $C_{20}H_{17}F_4N_5OS$ $[M+H]^+$ 452.1163, found 452.1161.

Methyl 2-(6-fluoro-4-oxo-7-(4-((4-(trifluoromethyl)phenyl)carbamothioyl)piperazin-1-yl)quinazolin-3(4H)-yl)acetate (**37**)

Prepared following General Procedure A from intermediate **29**. Purification via flash silica column chromatography DCM-MeOH (2-10%). Yield: 15%. ^{1}H NMR (400 MHz, DMSO-d_6) δ: 9.73 (s, 1H, H_{SCN-H}), 8.32 (s, 1H, H_2), 7.75 (d, J = 12.9 Hz, 1H, H_5), 7.66 (d, J = 8.4 Hz, 2H, $H_{3''}$, $H_{5''}$), 7.58 (d, J = 8.4 Hz, 2H, $H_{2''}$, $H_{6''}$), 7.21 (d, J = 7.9 Hz, 1H, H_8), 4.82 (s, 2H, CH_2), 4.13 (t, J

= 4.9 Hz, 4H, 2xCH_2, $H_{3'}$, $H_{5'}$), 3.71 (s, 3H, -OCH_3), 3.40 (m, 4H, 2xCH_2, $H_{2'}$, $H_{6'}$). ^{13}C NMR (101 MHz, DMSO-d_6) δ: 181.79, 168.94, 159.56 (d, $^4J_{C-F}$ = 2.5 Hz), 153.76 (d, $^1J_{C-F}$ = 248.0 Hz), 148.28, 146.64, 145.79 (d, $^2J_{C-F}$ = 9.9 Hz), 145.30, 125.63 (q, $^3J_{C-F}$ = 3.76 Hz), 124.96 (q, $^1J_{C-F}$ = 271.4 Hz), 124.70, 124.26 (q, $^2J_{C-F}$ = 31.9 Hz), 115.94 (d, $^3J_{C-F}$ = 3.2 Hz), 114.95 (d, $^3J_{C-F}$ = 8.5 Hz), 112.03 (d, $^2J_{C-F}$ = 23.0 Hz), 52.89, 49.36 (d, $^4J_{C-F}$ = 4.27 Hz), 48.29, 47.50. HRMS (ESI+) m/z calculated for $C_{23}H_{21}F_4N_5O_3S$ $[M+H]^+$ 524.1374, found 524.1373.

4-(6-fluoro-2,4-dioxo-1,2,3,4-tetrahydroquinazolin-7-yl)-N-(4-(trifluoromethyl)phenyl) piperazine-1-carbothioamide (**38**)

Prepared following General Procedure A from intermediate **30**. Purification via flash silica column chromatography DCM-MeOH (10-25%). Yield: 65%. ^{1}H NMR (400 MHz, DMSO-d_6) δ: 11.18 (br s, 1H, NH), 11.02 (br s, 1H, NH), 9.71 (s, 1H, H_{SCN-H}), 7.66 (d, *J* = 8.6 Hz, 2H, $H_{2''}$, $H_{6''}$), 7.58 (d, *J* = 8.5 Hz, 2H, $H_{3''}$, $H_{5''}$), 7.51 (d, *J* = 12.9 Hz, 1H, H_5), 6.67 (d, *J* = 7.3 Hz, 1H, H_8), 4.10 (t, *J* = 4.8 Hz, 4H, 2xCH_2, $H_{3'}$, $H_{5'}$), 3.31 (t, *J* = 5.2 Hz, 4H, 2xCH_2, $H_{2'}$, $H_{6'}$). ^{13}C NMR (101 MHz, DMSO-d_6) δ: 181.79, 162.36 (d, $^4J_{C-F}$ = 2.5 Hz), 150.86, 150.38 (d, $^1J_{C-F}$ = 241.13 Hz), 145.65 (d, $^2J_{C-F}$ = 9.82 Hz), 145.27 (d, $^4J_{C-F}$ = 1.18 Hz), 139.05, 125.62 (q, $^3J_{C-F}$ = 4.06 Hz), 124.91 (q, $^1J_{C-F}$ = 271.82 Hz), 124.67, 124.41 (q, $^2J_{C-F}$ = 32.07 Hz), 113.24 (d, $^2J_{C-F}$ = 23.5 Hz), 107.31 (d, $^3J_{C-F}$ = 7.5 Hz), 104.16 (d, $^3J_{C-F}$ = 2.07 Hz), 49.16 (d, $^4J_{C-F}$ = 4.3 Hz), 48.26. HRMS (ESI+) m/z calculated for $C_{20}H_{17}F_4N_5O_2S$ $[M+H]^+$ 468.1112, found 468.1119.

Ethyl 2-(6-fluoro-2,4-dioxo-7-(4-((4-(trifluoromethyl)phenyl)carbamothioyl)piperazin-1-yl)-1,4-dihydroquinazolin-3(2H)-yl)acetate (**39**)

Prepared following General Procedure A from intermediate **31**. Purification via flash silica column chromatography DCM-MeOH (2-15%). Yield: 55%. ^{1}H NMR (400 MHz, DMSO-d_6) δ: 11.52 (br s, 1H, NH), 9.72 (s, 1H, H_{SCN-H}), 7.66 (d, *J* = 8.3 Hz, 2H, $H_{2''}$, $H_{6''}$), 7.62 – 7.53 (m, 3H, H_5, $H_{3''}$, $H_{5''}$), 6.69 (d, *J* = 7.2 Hz, 1H, H_8), 4.62 (s, 2H, CH_2), 4.13 (m, 6H, 3xCH_2, $H_{2'}$, $H_{6'}$, -O$\underline{CH_2}$$CH_3$), 3.34 (m, 4H, 2x$CH_2$, $H_{3'}$, $H_{5'}$), 1.20 (t, *J* = 7.1 Hz, 3H, -$OCH_2\underline{CH_3}$). ^{13}C NMR (101 MHz, DMSO-d_6) δ: 181.80, 168.53, 161.01 (d, $^4J_{C-F}$ = 2.4 Hz), 150.52 (d, $^1J_{C-F}$ = 242.1 Hz), 150.32, 146.04 (d, $^2J_{C-F}$ = 9.9 Hz), 145.28, 137.69, 125.62 (q, $^3J_{C-F}$ = 3.8 Hz), 124.91 (q, $^1J_{C-F}$ = 271.3 Hz), 124.66, 124.27 (q, $^2J_{C-F}$ = 31.7 Hz), 113.69 (d, $^2J_{C-F}$ = 23.5 Hz), 106.09 (d, $^3J_{C-F}$ = 7.9 Hz),

103.96, 61.51, 49.07 (d, $^{4}J_{C-F}$ = 4.32 Hz), 48.22, 41.72, 14.48. HRMS (ESI+) m/z calculated for $C_{24}H_{23}F_4N_5O_4S$ $[M+H]^+$ 554.1480, found 554.1476.

4-(6-fluoro-4-oxo-1,4-dihydroquinolin-7-yl)-N-(4-(trifluoromethyl)phenyl)piperazine-1-carbothioamide (**40**)

Prepared following General Procedure A from intermediate **32**. Purification via flash silica column chromatography DCM-MeOH (5-20%). Yield: 75%. ^{1}H NMR (400 MHz, DMSO-d_6) δ: 12.15 (br s, 1H, NH), 9.77 (s, 1H, H_{SCN-H}), 7.98 (d, J = 7.3 Hz, 1H, H_2), 7.72 (d, J = 13.4 Hz, 1H, H_5), 7.67 (d, J = 8.5 Hz, 2H, $H_{3''}$, $H_{5''}$), 7.59 (d, J = 8.4 Hz, 2H, $H_{2''}$, $H_{6''}$), 7.10 (d, J = 7.5 Hz, 1H, H_8), 6.13 (d, J = 7.3 Hz, 1H, H_3), 4.14 (t, J = 5.1 Hz, 4H, 2xCH_2, $H_{2'}$, $H_{6'}$), 3.31 (d, J = 5.0 Hz, 4H, 2xCH_2, $H_{3'}$, $H_{5'}$). ^{13}C NMR (101 MHz, DMSO-d_6) δ: 181.84, 175.49 (d, $^{4}J_{C-F}$ = 2.15 Hz), 152.51 (d, $^{1}J_{C-F}$ = 245.2 Hz), 145.32, 143.96 (d, $^{2}J_{C-F}$ = 11.3 Hz), 140.03, 138.25, 125.61 (q, $^{3}J_{C-F}$ = 3.60 Hz), 124.92 (q, $^{1}J_{C-F}$ = 271.37 Hz), 124.67, 124.23 (q, $^{2}J_{C-F}$ = 31.81 Hz), 120.28 (d, $^{3}J_{C-F}$ = 6.5 Hz), 110.32 (d, $^{2}J_{C-F}$ = 21.8 Hz), 107.87, 107.04 (d, $^{3}J_{C-F}$ = 2.3 Hz), 49.64 (d, $^{4}J_{C-F}$ = 3.75 Hz), 48.40. HRMS (ESI+) m/z calculated for $C_{21}H_{18}F_4N_4OS$ $[M+H]^+$ 451.1210, found 451.1210.

2-(6-fluoro-4-oxo-7-(4-((4-(trifluoromethyl)phenyl)carbamothioyl)piperazin-1-yl)quinazolin-3(4H)-yl)acetic acid (**41**)

Prepared following General Procedure B from intermediate **34**. Purification via flash silica column chromatography DCM-MeOH (5-20%) + 1% AcOH. Yield: 10%, impure. ^{1}H NMR (400 MHz, DMSO-d_6) δ: 9.77 (s, 1H, H_{SCN-H}), 8.33 (s, 1H, H_2), 7.76 (d, J = 13.0 Hz, 1H, H_5), 7.66 (d, J = 8.5 Hz, 2H, $H_{2''}$, $H_{6''}$), 7.60 (d, *J* = 8.5 Hz, 2H, $H_{3''}$, $H_{5''}$), 7.21 (d, J = 8.0 Hz, 1H, H_8), 4.72 (s, 2H, CH_2), 4.13 (t, J = 5.0 Hz, 4H, 2xCH_2, $H_{3'}$, $H_{5'}$), 3.39 (t, J = 5 Hz, 4H, 2xCH_2, $H_{2'}$, $H_{6'}$), OH missing. HRMS (ESI-) m/z calculated for $C_{22}H_{19}F_4N_5O_3S$ $[M-H]^-$ 508.1072, found 508.1076.

2-(6-fluoro-2,4-dioxo-7-(4-((4-(trifluoromethyl)phenyl)carbamothioyl)piperazin-1-yl)-1,4-dihydroquinazolin-3(2H)-yl)acetic acid (**42**)

Prepared following General Procedure C from intermediate **35**. Yield: 82 %. ^{1}H NMR (400 MHz, DMSO-d_6) δ: 12.89 (br s, 1H, -COOH), 11.50 (s, 1H, NH), 9.78 (s, 1H, H_{SCN-H}), 7.66 (d, J = 8.6 Hz, 2H, $H_{2''}$, $H_{6''}$), 7.62 – 7.54 (m, 3H, H_5, $H_{3''}$, $H_{5''}$), 6.72 (d, J = 7.3 Hz, 1H, H_8), 4.53 (s, 2H, CH_2), 4.12 (t, *J* = 4.8 Hz, 4H, 2xCH_2, $H_{3'}$, $H_{5'}$), 3.34 (m, 4H, 2xCH_2, $H_{2'}$, $H_{6'}$). ^{13}C NMR (101

MHz, DMSO-d_6) δ: 181.80, 169.87, 161.07 (d, $^4J_{C-F}$ = 2.0 Hz), 150.50 (d, $^1J_{C-F}$ = 242.0 Hz), 150.39, 145.90 (d, $^2J_{C-F}$ = 9.8 Hz), 145.35, 137.71, 125.56 (q, $^3J_{C-F}$ = 3.9 Hz), 124.92 (q, $^1J_{C-F}$ = 271.5 Hz), 124.68, 124.20 (q, $^2J_{C-F}$ = 31.9 Hz), 113.65 (d, $^2J_{C-F}$ = 23.5 Hz), 106.32 (d, $^3J_{C-F}$ = 8.2 Hz), 104.00, 49.13 (d, $^4J_{C-F}$ = 4.4 Hz), 48.26, 41.79. HRMS (ESI-) m/z calculated for $C_{22}H_{19}F_4N_5O_4S$ $[M-H]^-$ 524.1022, found 524.1021.

Methyl 2-amino-5-fluoro-4-(4-((4-(trifluoromethyl)phenyl)carbamothioyl)piperazin-1-yl)benzoate (**43**)

Prepared following General Procedure A from intermediate **33**. Purification via flash column chromatography DCM-MeOH (2-10%). Yield: 60%. ^{1}H NMR (400 MHz, DMSO-d_6) δ: 9.69 (s, 1H, H_{SCN-H}), 7.66 (d, J = 8.6 Hz, 2H, $H_{2''}$, $H_{6''}$), 7.58 (d, J = 8.2 Hz, 2H, $H_{3''}$, $H_{5''}$), 7.34 (d, J = 14.5 Hz, 1H, H_5), 6.53 (br s, 2H, NH_2), 6.33 (d, J = 7.9 Hz, 1H, H_8), 4.07 (m, 4H, 2xCH_2, $H_{2'}$, $H_{6'}$), 3.75 (s, 3H, -OCH_3), 3.24 – 3.17 (m, 4H, 2xCH_2, $H_{3'}$, $H_{5'}$). HRMS (ESI+) m/z calculated for $C_{20}H_{20}F_4N_4O_2S$ $[M+H]^+$ 457.1316, found 457.1323.

References

1. Schenone, M.; Dančík, V.; Wagner, B.K.; Clemons, P.A. Target identification and mechanism of action in chemical biology and drug discovery. *Nat Chem Biol* **2013**, *9*, 232-240, doi:10.1038/nchembio.1199.
2. Roszik, J.; Subbiah, V. Mining Public Databases for Precision Oncology. *Trends Cancer* **2018**, *4*, 463-465, doi:10.1016/j.trecan.2018.04.008.
3. Li, K.; Du, Y.; Li, L.; Wei, D.Q. Bioinformatics Approaches for Anti-cancer Drug Discovery. *Curr Drug Targets* **2020**, *21*, 3-17, doi:10.2174/1389450120666190923162203.
4. Shoemaker, R.H. The NCI60 human tumour cell line anticancer drug screen. In *Nat Rev Cancer*; England, 2006; Volume 6, pp. 813-823.
5. https://dtp.cancer.gov/databases_tools/molecular_target/default.htm. Available online: (accessed on November 10, 2022
6. Barretina, J.; Caponigro, G.; Stransky, N.; Venkatesan, K.; Margolin, A.A.; Kim, S.; Wilson, C.J.; Lehár, 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, doi:10.1038/nature11003.
7. 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*, D955-961, doi:10.1093/nar/gks1111.
8. Bredel, M.; Jacoby, E. Chemogenomics: an emerging strategy for rapid target and drug discovery. *Nat Rev Genet* **2004**, *5*, 262-275, doi:10.1038/nrg1317.
9. Caroli, J.; Dori, M.; Bicciato, S. Computational Methods for the Integrative Analysis of Genomics and Pharmacological Data. *Front Oncol* **2020**, *10*, 185, doi:10.3389/fonc.2020.00185.
10. Reinhold, W.C.; Varma, S.; Rajapakse, V.N.; Luna, A.; Sousa, F.G.; Kohn, K.W.; Pommier, Y.G. Using drug response data to identify molecular effectors, and molecular "omic" data to identify candidate drugs in cancer. *Hum Genet* **2015**, *134*, 3-11, doi:10.1007/s00439-014-1482-9.
11. Belizário, J.E.; Sangiuliano, B.A.; Perez-Sosa, M.; Neyra, J.M.; Moreira, D.F. Using Pharmacogenomic Databases for Discovering Patient-Target Genes and Small Molecule Candidates to Cancer Therapy. *Front Pharmacol* **2016**, *7*, 312, doi:10.3389/fphar.2016.00312.
12. Zaharevitz, D.W.; Holbeck, S.L.; Bowerman, C.; Svetlik, P.A. COMPARE: a web accessible tool for investigating mechanisms of cell growth inhibition. *J Mol Graph Model* **2002**, *20*, 297-303, doi:10.1016/s1093-3263(01)00126-7.
13. http://discover.nci.nih.gov/cellminer/. Available online: (accessed on November 10, 2022
14. Reinhold, W.C.; Sunshine, M.; Liu, H.; Varma, S.; Kohn, K.W.; Morris, J.; Doroshow, J.; Pommier, Y. CellMiner: a web-based suite of genomic and pharmacologic tools to explore transcript and drug patterns in the NCI-60 cell line set. *Cancer Res* **2012**, *72*, 3499-3511, doi:10.1158/0008-5472.CAN-12-1370.

15. Luna, A.; Rajapakse, V.N.; Sousa, F.G.; Gao, J.; Schultz, N.; Varma, S.; Reinhold, W.; Sander, C.; Pommier, Y. rcellminer: exploring molecular profiles and drug response of the NCI-60 cell lines in R. *Bioinformatics* **2016**, *32*, 1272-1274, doi:10.1093/bioinformatics/btv701.
16. Wang, S.; Gribskov, M.; Hazbun, T.R.; Pascuzzi, P.E. CellMiner Companion: an interactive web application to explore CellMiner NCI-60 data. *Bioinformatics* **2016**, *32*, 2399-2401, doi:10.1093/bioinformatics/btw162.
17. Reinhold, W.C.; Sunshine, M.; Varma, S.; Doroshow, J.H.; Pommier, Y. Using CellMiner 1.6 for Systems Pharmacology and Genomic Analysis of the NCI-60. *Clin Cancer Res* **2015**, *21*, 3841-3852, doi:10.1158/1078-0432.CCR-15-0335.
18. Rajapakse, V.N.; Luna, A.; Yamade, M.; Loman, L.; Varma, S.; Sunshine, M.; Iorio, F.; Sousa, F.G.; Elloumi, F.; Aladjem, M.I.; et al. CellMinerCDB for Integrative Cross-Database Genomics and Pharmacogenomics Analyses of Cancer Cell Lines. *iScience* **2018**, *10*, 247-264, doi:10.1016/j.isci.2018.11.029.
19. Luna, A.; Elloumi, F.; Varma, S.; Wang, Y.; Rajapakse, V.N.; Aladjem, M.I.; Robert, J.; Sander, C.; Pommier, Y.; Reinhold, W.C. CellMiner Cross-Database (CellMinerCDB) version 1.2: Exploration of patient-derived cancer cell line pharmacogenomics. *Nucleic Acids Res* **2021**, *49*, D1083-D1093, doi:10.1093/nar/gkaa968.
20. Tlemsani, C.; Pongor, L.; Elloumi, F.; Girard, L.; Huffman, K.E.; Roper, N.; Varma, S.; Luna, A.; Rajapakse, V.N.; Sebastian, R.; et al. SCLC-CellMiner: A Resource for Small Cell Lung Cancer Cell Line Genomics and Pharmacology Based on Genomic Signatures. *Cell Rep* **2020**, *33*, 108296, doi:10.1016/j.celrep.2020.108296.
21. Krushkal, J.; Negi, S.; Yee, L.M.; Evans, J.R.; Grkovic, T.; Palmisano, A.; Fang, J.; Sankaran, H.; McShane, L.M.; Zhao, Y.; et al. Molecular genomic features associated with in vitro response of the NCI-60 cancer cell line panel to natural products. *Mol Oncol* **2021**, *15*, 381-406, doi:10.1002/1878-0261.12849.
22. Arroyo, M.M.; Berral-González, A.; Bueno-Fortes, S.; Alonso-López, D.; Rivas, J.L. Mining Drug-Target Associations in Cancer: Analysis of Gene Expression and Drug Activity Correlations. *Biomolecules* **2020**, *10*, doi:10.3390/biom10050667.
23. Smirnov, P.; Safikhani, Z.; El-Hachem, N.; Wang, D.; She, A.; Olsen, C.; Freeman, M.; Selby, H.; Gendoo, D.M.; Grossmann, P.; et al. PharmacoGx: an R package for analysis of large pharmacogenomic datasets. *Bioinformatics* **2016**, *32*, 1244-1246, doi:10.1093/bioinformatics/btv723.
24. Smirnov, P.; Kofia, V.; Maru, A.; Freeman, M.; Ho, C.; El-Hachem, N.; Adam, G.A.; Ba-Alawi, W.; Safikhani, Z.; Haibe-Kains, B. PharmacoDB: an integrative database for mining in vitro anticancer drug screening studies. *Nucleic Acids Res* **2018**, *46*, D994-D1002, doi:10.1093/nar/gkx911.
25. Caroli, J.; Sorrentino, G.; Forcato, M.; Del Sal, G.; Bicciato, S. GDA, a web-based tool for Genomics and Drugs integrated analysis. *Nucleic Acids Res* **2018**, *46*, W148-W156, doi:10.1093/nar/gky434.
26. Rees, M.G.; Seashore-Ludlow, B.; Cheah, J.H.; Adams, D.J.; Price, E.V.; Gill, S.; Javaid, S.; Coletti, M.E.; Jones, V.L.; Bodycombe, N.E.; et al. Correlating chemical sensitivity and basal gene expression reveals mechanism of action. *Nat Chem Biol* **2016**, *12*, 109-116, doi:10.1038/nchembio.1986.

27. 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, doi:10.1158/2159-8290.CD-15-0235.
28. Rees, M.G.; Seashore-Ludlow, B.; Clemons, P.A. Computational Analyses Connect Small-Molecule Sensitivity to Cellular Features Using Large Panels of Cancer Cell Lines. *Methods Mol Biol* **2019**, *1888*, 233-254, doi:10.1007/978-1-4939-8891-4_14.
29. de Waal, L.; Lewis, T.A.; Rees, M.G.; Tsherniak, A.; Wu, X.; Choi, P.S.; Gechijian, L.; Hartigan, C.; Faloon, P.W.; Hickey, M.J.; et al. Identification of cancer-cytotoxic modulators of PDE3A by predictive chemogenomics. *Nat Chem Biol* **2016**, *12*, 102-108, doi:10.1038/nchembio.1984.
30. Corsello, S.M.; Nagari, R.T.; Spangler, R.D.; Rossen, J.; Kocak, M.; Bryan, J.G.; Humeidi, R.; Peck, D.; Wu, X.; Tang, A.A.; et al. Discovering the anti-cancer potential of non-oncology drugs by systematic viability profiling. *Nat Cancer* **2020**, *1*, 235-248, doi:10.1038/s43018-019-0018-6.
31. Gonçalves, E.; Segura-Cabrera, A.; Pacini, C.; Picco, G.; Behan, F.M.; Jaaks, P.; Coker, E.A.; van der Meer, D.; Barthorpe, A.; Lightfoot, H.; et al. Drug mechanism-of-action discovery through the integration of pharmacological and CRISPR screens. *Mol Syst Biol* **2020**, *16*, e9405, doi:10.15252/msb.20199405.
32. Vazquez, F.; Boehm, J.S. The Cancer Dependency Map enables drug mechanism-of-action investigations. *Mol Syst Biol* **2020**, *16*, e9757, doi:10.15252/msb.20209757.
33. Firoozbakht, F.; Yousefi, B.; Schwikowski, B. An overview of machine learning methods for monotherapy drug response prediction. *Brief Bioinform* **2021**, doi:10.1093/bib/bbab408.
34. Lind, A.P.; Anderson, P.C. Predicting drug activity against cancer cells by random forest models based on minimal genomic information and chemical properties. *PLoS One* **2019**, *14*, e0219774, doi:10.1371/journal.pone.0219774.
35. Zhang, N.; Wang, H.; Fang, Y.; Wang, J.; Zheng, X.; Liu, X.S. Predicting Anticancer Drug Responses Using a Dual-Layer Integrated Cell Line-Drug Network Model. *PLoS Comput Biol* **2015**, *11*, e1004498, doi:10.1371/journal.pcbi.1004498.
36. Cadow, J.; Born, J.; Manica, M.; Oskooei, A.; Rodríguez Martínez, M. PaccMann: a web service for interpretable anticancer compound sensitivity prediction. *Nucleic Acids Res* **2020**, *48*, W502-W508, doi:10.1093/nar/gkaa327.
37. Chang, Y.; Park, H.; Yang, H.J.; Lee, S.; Lee, K.Y.; Kim, T.S.; Jung, J.; Shin, J.M. Cancer Drug Response Profile scan (CDRscan): A Deep Learning Model That Predicts Drug Effectiveness from Cancer Genomic Signature. *Sci Rep* **2018**, *8*, 8857, doi:10.1038/s41598-018-27214-6.
38. Joo, M.; Park, A.; Kim, K.; Son, W.J.; Lee, H.S.; Lim, G.; Lee, J.; Lee, D.H.; An, J.; Kim, J.H.; et al. A Deep Learning Model for Cell Growth Inhibition IC50 Prediction and Its Application for Gastric Cancer Patients. *Int J Mol Sci* **2019**, *20*, doi:10.3390/ijms20246276.
39. Al-Jarf, R.; de Sá, A.G.C.; Pires, D.E.V.; Ascher, D.B. pdCSM-cancer: Using Graph-Based Signatures to Identify Small Molecules with Anticancer Properties. *J Chem Inf Model* **2021**, *61*, 3314-3322, doi:10.1021/acs.jcim.1c00168.

40. Lauria, A.; Tutone, M.; Almerico, A.M. Virtual lock-and-key approach: the in silico revival of Fischer model by means of molecular descriptors. *Eur J Med Chem* **2011**, *46*, 4274-4280, doi:10.1016/j.ejmech.2011.06.033.
41. Lauria, A.; Tutone, M.; Barone, G.; Almerico, A.M. Multivariate analysis in the identification of biological targets for designed molecular structures: the BIOTA protocol. *Eur J Med Chem* **2014**, *75*, 106-110, doi:10.1016/j.ejmech.2014.01.025.
42. Lauria, A.; Mannino, S.; Gentile, C.; Mannino, G.; Martorana, A.; Peri, D. DRUDIT: web-based DRUgs DIscovery Tools to design small molecules as modulators of biological targets. *Bioinformatics* **2020**, *36*, 1562-1569, doi:10.1093/bioinformatics/btz783.
43. https://dtp.cancer.gov/discovery_development/nci-60/. Available online: (accessed on November 10, 2022
44. NCI DTP Chemical Data. Available online: https://wiki.nci.nih.gov/display/NCIDTPdata/Chemical+Data (accessed on November 10, 2022
45. Kayed, R.; Lo Cascio, F.; Piccionello Palumbo, A.; Pace, A. Novel small molecules that bind and/or modulate different forms of tau oligomers, 23 April 2020, WO2020/219714 A1.
46. Battisti, A.; Palumbo Piccionello, A.; Sgarbossa, A.; Vilasi, S.; Ricci, C.; Ghetti, F.; Spinozzi, F.; Marino Gammazza, A.; Giacalone, V.; Martorana, A.; et al. Curcumin-like compounds designed to modify amyloid beta peptide aggregation patterns. *RSC Advances* **2017**, *7*, 31714-31724, doi:10.1039/c7ra05300b.
47. Lo Cascio, F.; Puangmalai, N.; Ellsworth, A.; Bucchieri, F.; Pace, A.; Palumbo Piccionello, A.; Kayed, R. Toxic Tau Oligomers Modulated by Novel Curcumin Derivatives. *Sci Rep* **2019**, *9*, 19011, doi:10.1038/s41598-019-55419-w.
48. Sharma, R.A.; Gescher, A.J.; Steward, W.P. Curcumin: the story so far. *Eur J Cancer* **2005**, *41*, 1955-1968, doi:10.1016/j.ejca.2005.05.009.
49. Adeluola, A.; Zulfiker, A.H.M.; Brazeau, D.; Amin, A.R.M.R. Perspectives for synthetic curcumins in chemoprevention and treatment of cancer: An update with promising analogues. *Eur J Pharmacol* **2021**, *906*, 174266, doi:10.1016/j.ejphar.2021.174266.
50. Ahsan, M.J.; Choudhary, K.; Jadav, S.S.; Yasmin, S.; Ansari, M.Y.; Sreenivasulu, R. Synthesis, antiproliferative activity, and molecular docking studies of curcumin analogues bearing pyrazole ring. *Medicinal Chemistry Research* **2015**, *24*, 4166-4180, doi:10.1007/s00044-015-1457-y.
51. Ahsan, M.J. Evaluation of Anticancer Activity of Curcumin Analogues Bearing a Heterocyclic Nucleus. *Asian Pac J Cancer Prev* **2016**, *17*, 1739-1744, doi:10.7314/apjcp.2016.17.4.1739.
52. Anand, P.; Kunnumakkara, A.B.; Newman, R.A.; Aggarwal, B.B. Bioavailability of curcumin: problems and promises. *Mol Pharm* **2007**, *4*, 807-818, doi:10.1021/mp700113r.
53. Sabet, S.; Rashidinejad, A.; Melton, L.D.; McGillivray, D.J. Recent advances to improve curcumin oral bioavailability. *Trends in Food Science and Technology* **2021**, *110*, 253-266, doi:10.1016/j.tifs.2021.02.006.

54. Sanidad, K.Z.; Sukamtoh, E.; Xiao, H.; McClements, D.J.; Zhang, G. Curcumin: Recent Advances in the Development of Strategies to Improve Oral Bioavailability. *Annu Rev Food Sci Technol* **2019**, *10*, 597-617, doi:10.1146/annurev-food-032818-121738.
55. Kotha, R.R.; Luthria, D.L. Curcumin: Biological, Pharmaceutical, Nutraceutical, and Analytical Aspects. *Molecules* **2019**, *24*, doi:10.3390/molecules24162930.
56. Nelson, K.M.; Dahlin, J.L.; Bisson, J.; Graham, J.; Pauli, G.F.; Walters, M.A. The Essential Medicinal Chemistry of Curcumin. *J Med Chem* **2017**, *60*, 1620-1637, doi:10.1021/acs.jmedchem.6b00975.
57. Olotu, F.; Agoni, C.; Soremekun, O.; Soliman, M.E.S. An Update on the Pharmacological Usage of Curcumin: Has it Failed in the Drug Discovery Pipeline? *Cell Biochem Biophys* **2020**, *78*, 267-289, doi:10.1007/s12013-020-00922-5.
58. Rodrigues, F.C.; Kumar, N.A.; Thakur, G. The potency of heterocyclic curcumin analogues: An evidence-based review. *Pharmacol Res* **2021**, *166*, 105489, doi:10.1016/j.phrs.2021.105489.
59. Wang, H.; Ma, X.; Ren, S.; Buolamwini, J.K.; Yan, C. A small-molecule inhibitor of MDMX activates p53 and induces apoptosis. *Mol Cancer Ther* **2011**, *10*, 69-79, doi:10.1158/1535-7163.MCT-10-0581.
60. Wade, M.; Li, Y.C.; Wahl, G.M. MDM2, MDMX and p53 in oncogenesis and cancer therapy. *Nat Rev Cancer* **2013**, *13*, 83-96, doi:10.1038/nrc3430.
61. Lam, S.; Lodder, K.; Teunisse, A.F.; Rabelink, M.J.; Schutte, M.; Jochemsen, A.G. Role of Mdm4 in drug sensitivity of breast cancer cells. *Oncogene* **2010**, *29*, 2415-2426, doi:10.1038/onc.2009.522.
62. Danovi, D.; Meulmeester, E.; Pasini, D.; Migliorini, D.; Capra, M.; Frenk, R.; de Graaf, P.; Francoz, S.; Gasparini, P.; Gobbi, A.; et al. Amplification of Mdmx (or Mdm4) directly contributes to tumor formation by inhibiting p53 tumor suppressor activity. *Mol Cell Biol* **2004**, *24*, 5835-5843, doi:10.1128/mcb.24.13.5835-5843.2004.
63. Lakshmaiah, K.C.; Jacob, L.A.; Aparna, S.; Lokanatha, D.; Saldanha, S.C. Epigenetic therapy of cancer with histone deacetylase inhibitors. *J Cancer Res Ther* **2014**, *10*, 469-478, doi:10.4103/0973-1482.137937.
64. Ficner, R. Novel structural insights into class I and II histone deacetylases. *Curr Top Med Chem* **2009**, *9*, 235-240, doi:10.2174/156802609788085304.
65. Siegel, D.; Hussein, M.; Belani, C.; Robert, F.; Galanis, E.; Richon, V.M.; Garcia-Vargas, J.; Sanz-Rodriguez, C.; Rizvi, S. Vorinostat in solid and hematologic malignancies. *J Hematol Oncol* **2009**, *2*, 31, doi:10.1186/1756-8722-2-31.
66. Sanchez-Gonzalez, B.; Yang, H.; Bueso-Ramos, C.; Hoshino, K.; Quintas-Cardama, A.; Richon, V.M.; Garcia-Manero, G. Antileukemia activity of the combination of an anthracycline with a histone deacetylase inhibitor. *Blood* **2006**, *108*, 1174-1182, doi:10.1182/blood-2005-09-008086.
67. Jasek, E.; Gajda, M.; Lis, G.J.; Jasińska, M.; Litwin, J.A. Combinatorial effects of PARP inhibitor PJ34 and histone deacetylase inhibitor vorinostat on leukemia cell lines. *Anticancer Res* **2014**, *34*, 1849-1856.
68. Gruhn, B.; Naumann, T.; Gruner, D.; Walther, M.; Wittig, S.; Becker, S.; Beck, J.F.; Sonnemann, J. The expression of histone deacetylase 4 is associated with

prednisone poor-response in childhood acute lymphoblastic leukemia. *Leuk Res* **2013**, *37*, 1200-1207, doi:10.1016/j.leukres.2013.07.016.
69. Gentile, C.; Martorana, A.; Lauria, A.; Bonsignore, R. Kinase Inhibitors in Multitargeted Cancer Therapy. *Curr Med Chem* **2017**, *24*, 1671-1686, doi:10.2174/0929867324666170112112734.
70. Herbst, R.S.; Fukuoka, M.; Baselga, J. Gefitinib--a novel targeted approach to treating cancer. *Nat Rev Cancer* **2004**, *4*, 956-965, doi:10.1038/nrc1506.
71. Ciardiello, F.; Caputo, R.; Bianco, R.; Damiano, V.; Pomatico, G.; Pepe, S.; Bianco, A.R.; Agrawal, S.; Mendelsohn, J.; Tortora, G. Cooperative inhibition of renal cancer growth by anti-epidermal growth factor receptor antibody and protein kinase A antisense oligonucleotide. *J Natl Cancer Inst* **1998**, *90*, 1087-1094, doi:10.1093/jnci/90.14.1087.
72. Kantarjian, H.; Jabbour, E.; Grimley, J.; Kirkpatrick, P. Dasatinib. *Nat Rev Drug Discov* **2006**, *5*, 717-718, doi:10.1038/nrd2135.
73. Lozzio, C.B.; Lozzio, B.B. Human chronic myelogenous leukemia cell-line with positive Philadelphia chromosome. *Blood* **1975**, *45*, 321-334.
74. Klein, E.; Ben-Bassat, H.; Neumann, H.; Ralph, P.; Zeuthen, J.; Polliack, A.; Vánky, F. Properties of the K562 cell line, derived from a patient with chronic myeloid leukemia. *Int J Cancer* **1976**, *18*, 421-431, doi:10.1002/ijc.2910180405.
75. Honma, Y.; Okabe-Kado, J.; Kasukabe, T.; Hozumi, M.; Umezawa, K. Inhibition of abl oncogene tyrosine kinase induces erythroid differentiation of human myelogenous leukemia K562 cells. *Jpn J Cancer Res* **1990**, *81*, 1132-1136, doi:10.1111/j.1349-7006.1990.tb02524.x.
76. McGahon, A.J.; Brown, D.G.; Martin, S.J.; Amarante-Mendes, G.P.; Cotter, T.G.; Cohen, G.M.; Green, D.R. Downregulation of Bcr-Abl in K562 cells restores susceptibility to apoptosis: characterization of the apoptotic death. *Cell Death Differ* **1997**, *4*, 95-104, doi:10.1038/sj.cdd.4400213.
77. Majid, S.; Saini, S.; Dar, A.A.; Hirata, H.; Shahryari, V.; Tanaka, Y.; Yamamura, S.; Ueno, K.; Zaman, M.S.; Singh, K.; et al. MicroRNA-205 inhibits Src-mediated oncogenic pathways in renal cancer. *Cancer Res* **2011**, *71*, 2611-2621, doi:10.1158/0008-5472.CAN-10-3666.
78. Rossari, F.; Minutolo, F.; Orciuolo, E. Past, present, and future of Bcr-Abl inhibitors: from chemical development to clinical efficacy. *J Hematol Oncol* **2018**, *11*, 84, doi:10.1186/s13045-018-0624-2.
79. Bou-Hanna, C.; Jarry, A.; Lode, L.; Schmitz, I.; Schulze-Osthoff, K.; Kury, S.; Bezieau, S.; Mosnier, J.F.; Laboisse, C.L. Acute cytotoxicity of MIRA-1/NSC19630, a mutant p53-reactivating small molecule, against human normal and cancer cells via a caspase-9-dependent apoptosis. *Cancer Lett* **2015**, *359*, 211-217, doi:10.1016/j.canlet.2015.01.014.
80. Geley, S.; Hartmann, B.L.; Hattmannstorfer, R.; Löffler, M.; Ausserlechner, M.J.; Bernhard, D.; Sgonc, R.; Strasser-Wozak, E.M.; Ebner, M.; Auer, B.; et al. p53-induced apoptosis in the human T-ALL cell line CCRF-CEM. *Oncogene* **1997**, *15*, 2429-2437, doi:10.1038/sj.onc.1201399.

81. Massumi, M.; Ziaee, A.A.; Sarbolouki, M.N. Apoptosis induction in human lymphoma and leukemia cell lines by transfection via dendrosomes carrying wild-type p53 cDNA. *Biotechnol Lett* **2006**, *28*, 61-66, doi:10.1007/s10529-005-4689-y.
82. Shaw, A.T.; Yasothan, U.; Kirkpatrick, P. Crizotinib. *Nat Rev Drug Discov* **2011**, *10*, 897-898, doi:10.1038/nrd3600.
83. Wan, W.; Albom, M.S.; Lu, L.; Quail, M.R.; Becknell, N.C.; Weinberg, L.R.; Reddy, D.R.; Holskin, B.P.; Angeles, T.S.; Underiner, T.L.; et al. Anaplastic lymphoma kinase activity is essential for the proliferation and survival of anaplastic large-cell lymphoma cells. *Blood* **2006**, *107*, 1617-1623, doi:10.1182/blood-2005-08-3254.
84. Russell, P.; Nurse, P. cdc25+ functions as an inducer in the mitotic control of fission yeast. *Cell* **1986**, *45*, 145-153, doi:10.1016/0092-8674(86)90546-5.
85. Eckstein, J.W. Cdc25 as a potential target of anticancer agents. *Invest New Drugs* **2000**, *18*, 149-156, doi:10.1023/a:1006377913494.
86. Lindqvist, A.; Källström, H.; Lundgren, A.; Barsoum, E.; Rosenthal, C.K. Cdc25B cooperates with Cdc25A to induce mitosis but has a unique role in activating cyclin B1-Cdk1 at the centrosome. *J Cell Biol* **2005**, *171*, 35-45, doi:10.1083/jcb.200503066.
87. Karlsson-Rosenthal, C.; Millar, J.B. Cdc25: mechanisms of checkpoint inhibition and recovery. *Trends Cell Biol* **2006**, *16*, 285-292, doi:10.1016/j.tcb.2006.04.002.
88. Donzelli, M.; Draetta, G.F. Regulating mammalian checkpoints through Cdc25 inactivation. *EMBO Rep* **2003**, *4*, 671-677, doi:10.1038/sj.embor.embor887.
89. Lammer, C.; Wagerer, S.; Saffrich, R.; Mertens, D.; Ansorge, W.; Hoffmann, I. The cdc25B phosphatase is essential for the G2/M phase transition in human cells. *J Cell Sci* **1998**, *111 (Pt 16)*, 2445-2453.
90. Lopez-Girona, A.; Furnari, B.; Mondesert, O.; Russell, P. Nuclear localization of Cdc25 is regulated by DNA damage and a 14-3-3 protein. *Nature* **1999**, *397*, 172-175, doi:10.1038/16488.
91. Pines, J. Four-dimensional control of the cell cycle. *Nat Cell Biol* **1999**, *1*, E73-79, doi:10.1038/11041.
92. Foster, D.A.; Yellen, P.; Xu, L.; Saqcena, M. Regulation of G1 Cell Cycle Progression: Distinguishing the Restriction Point from a Nutrient-Sensing Cell Growth Checkpoint(s). *Genes Cancer* **2010**, *1*, 1124-1131, doi:10.1177/1947601910392989.
93. Mukhopadhyay, S.; Saqcena, M.; Foster, D.A. Synthetic lethality in KRas-driven cancer cells created by glutamine deprivation. *Oncoscience* **2015**, *2*, 807-808, doi:10.18632/oncoscience.253.
94. Saqcena, M.; Menon, D.; Patel, D.; Mukhopadhyay, S.; Chow, V.; Foster, D.A. Amino acids and mTOR mediate distinct metabolic checkpoints in mammalian G1 cell cycle. *PLoS One* **2013**, *8*, e74157, doi:10.1371/journal.pone.0074157.
95. Sur, S.; Agrawal, D.K. Phosphatases and kinases regulating CDC25 activity in the cell cycle: clinical implications of CDC25 overexpression and potential treatment strategies. *Mol Cell Biochem* **2016**, *416*, 33-46, doi:10.1007/s11010-016-2693-2.
96. Brenner, A.K.; Reikvam, H.; Lavecchia, A.; Bruserud, Ø. Therapeutic targeting the cell division cycle 25 (CDC25) phosphatases in human acute myeloid leukemia--the possibility to target several kinases through inhibition of the various CDC25 isoforms. *Molecules* **2014**, *19*, 18414-18447, doi:10.3390/molecules191118414.

97. Liu, K.; Zheng, M.; Lu, R.; Du, J.; Zhao, Q.; Li, Z.; Li, Y.; Zhang, S. The role of CDC25C in cell cycle regulation and clinical cancer therapy: a systematic review. *Cancer Cell Int* **2020**, *20*, 213, doi:10.1186/s12935-020-01304-w.
98. Rudolph, J. Cdc25 phosphatases: structure, specificity, and mechanism. *Biochemistry* **2007**, *46*, 3595-3604, doi:10.1021/bi700026j.
99. Källström, H.; Lindqvist, A.; Pospisil, V.; Lundgren, A.; Rosenthal, C.K. Cdc25A localisation and shuttling: characterisation of sequences mediating nuclear export and import. *Exp Cell Res* **2005**, *303*, 89-100, doi:10.1016/j.yexcr.2004.09.012.
100. Keyse, S.M.; Ginsburg, M. Amino acid sequence similarity between CL100, a dual-specificity MAP kinase phosphatase and cdc25. *Trends Biochem Sci* **1993**, *18*, 377-378, doi:10.1016/0968-0004(93)90092-2.
101. Aressy, B.; Ducommun, B. Cell cycle control by the CDC25 phosphatases. *Anticancer Agents Med Chem* **2008**, *8*, 818-824, doi:10.2174/187152008786847756.
102. Savitsky, P.A.; Finkel, T. Redox regulation of Cdc25C. *J Biol Chem* **2002**, *277*, 20535-20540, doi:10.1074/jbc.M201589200.
103. Fauman, E.B.; Cogswell, J.P.; Lovejoy, B.; Rocque, W.J.; Holmes, W.; Montana, V.G.; Piwnica-Worms, H.; Rink, M.J.; Saper, M.A. Crystal structure of the catalytic domain of the human cell cycle control phosphatase, Cdc25A. *Cell* **1998**, *93*, 617-625, doi:10.1016/s0092-8674(00)81190-3.
104. Reynolds, R.A.; Yem, A.W.; Wolfe, C.L.; Deibel, M.R.; Chidester, C.G.; Watenpaugh, K.D. Crystal structure of the catalytic subunit of Cdc25B required for G2/M phase transition of the cell cycle. *J Mol Biol* **1999**, *293*, 559-568, doi:10.1006/jmbi.1999.3168.
105. Rudolph, J. Targeting the neighbor's pool. *Mol Pharmacol* **2004**, *66*, 780-782, doi:10.1124/mol.104.004788.
106. Lavecchia, A.; Coluccia, A.; Di Giovanni, C.; Novellino, E. Cdc25B phosphatase inhibitors in cancer therapy: latest developments, trends and medicinal chemistry perspective. *Anticancer Agents Med Chem* **2008**, *8*, 843-856, doi:10.2174/187152008786847783.
107. Lavecchia, A.; Di Giovanni, C.; Novellino, E. CDC25 phosphatase inhibitors: an update. *Mini Rev Med Chem* **2012**, *12*, 62-73, doi:10.2174/138955712798868940.
108. Boutros, R.; Lobjois, V.; Ducommun, B. CDC25 phosphatases in cancer cells: key players? Good targets? *Nat Rev Cancer* **2007**, *7*, 495-507, doi:10.1038/nrc2169.
109. Lazo, J.S.; Aslan, D.C.; Southwick, E.C.; Cooley, K.A.; Ducruet, A.P.; Joo, B.; Vogt, A.; Wipf, P. Discovery and biological evaluation of a new family of potent inhibitors of the dual specificity protein phosphatase Cdc25. *J Med Chem* **2001**, *44*, 4042-4049, doi:10.1021/jm0102046.
110. Brezak, M.C.; Quaranta, M.; Contour-Galcera, M.O.; Lavergne, O.; Mondesert, O.; Auvray, P.; Kasprzyk, P.G.; Prevost, G.P.; Ducommun, B. Inhibition of human tumor cell growth in vivo by an orally bioavailable inhibitor of CDC25 phosphatases. *Mol Cancer Ther* **2005**, *4*, 1378-1387, doi:10.1158/1535-7163.MCT-05-0168.
111. Cossy, J.; Belotti, D.; Brisson, M.; Skoko, J.J.; Wipf, P.; Lazo, J.S. Biological evaluation of newly synthesized quinoline-5,8-quinones as Cdc25B inhibitors. *Bioorg Med Chem* **2006**, *14*, 6283-6287, doi:10.1016/j.bmc.2006.05.053.

112. Lavecchia, A.; Di Giovanni, C.; Pesapane, A.; Montuori, N.; Ragno, P.; Martucci, N.M.; Masullo, M.; De Vendittis, E.; Novellino, E. Discovery of new inhibitors of Cdc25B dual specificity phosphatases by structure-based virtual screening. *J Med Chem* **2012**, *55*, 4142-4158, doi:10.1021/jm201624h.
113. Ge, Y.; van der Kamp, M.; Malaisree, M.; Liu, D.; Liu, Y.; Mulholland, A.J. Identification of the quinolinedione inhibitor binding site in Cdc25 phosphatase B through docking and molecular dynamics simulations. *J Comput Aided Mol Des* **2017**, *31*, 995-1007, doi:10.1007/s10822-017-0073-y.
114. Tao, Y.; Hao, X.; Ding, X.; Cherukupalli, S.; Song, Y.; Liu, X.; Zhan, P. Medicinal chemistry insights into novel CDC25 inhibitors. *Eur J Med Chem* **2020**, *201*, 112374, doi:10.1016/j.ejmech.2020.112374.
115. Song, Y.; Lin, X.; Kang, D.; Li, X.; Zhan, P.; Liu, X.; Zhang, Q. Discovery and characterization of novel imidazopyridine derivative CHEQ-2 as a potent CDC25 inhibitor and promising anticancer drug candidate. *Eur J Med Chem* **2014**, *82*, 293-307, doi:10.1016/j.ejmech.2014.05.063.
116. Park, H.; Bahn, Y.J.; Jung, S.K.; Jeong, D.G.; Lee, S.H.; Seo, I.; Yoon, T.S.; Kim, S.J.; Ryu, S.E. Discovery of novel Cdc25 phosphatase inhibitors with micromolar activity based on the structure-based virtual screening. *J Med Chem* **2008**, *51*, 5533-5541, doi:10.1021/jm701157g.
117. Zhang, S.; Jia, Q.; Gao, Q.; Fan, X.; Weng, Y.; Su, Z. Dual-Specificity Phosphatase CDC25B Was Inhibited by Natural Product HB-21 Through Covalently Binding to the Active Site. *Front Chem* **2018**, *6*, 531, doi:10.3389/fchem.2018.00531.
118. Cerchia, C.; Nasso, R.; Mori, M.; Villa, S.; Gelain, A.; Capasso, A.; Aliotta, F.; Simonetti, M.; Rullo, R.; Masullo, M.; et al. Discovery of Novel Naphthylphenylketone and Naphthylphenylamine Derivatives as Cell Division Cycle 25B (CDC25B) Phosphatase Inhibitors: Design, Synthesis, Inhibition Mechanism, and in Vitro Efficacy against Melanoma Cell Lines. *J Med Chem* **2019**, *62*, 7089-7110, doi:10.1021/acs.jmedchem.9b00632.
119. Aliotta, F.; Nasso, R.; Rullo, R.; Arcucci, A.; Avagliano, A.; Simonetti, M.; Sanità, G.; Masullo, M.; Lavecchia, A.; Ruocco, M.R.; et al. Inhibition mechanism of naphthylphenylamine derivatives acting on the CDC25B dual phosphatase and analysis of the molecular processes involved in the high cytotoxicity exerted by one selected derivative in melanoma cells. *J Enzyme Inhib Med Chem* **2020**, *35*, 1866-1878, doi:10.1080/14756366.2020.1819257.
120. Zhang, J.; Ji, F.J.; Gu, Y.; Zhang, X.Y.; Qiao, S.X. Chalcones derivatives as potent Cell division cycle 25B phosphatase inhibitors. *Pharmacol Rep* **2014**, *66*, 515-519, doi:10.1016/j.pharep.2013.08.016.
121. Huber-Villaume, S.; Revelant, G.; Sibille, E.; Philippot, S.; Morabito, A.; Dunand, S.; Chaimbault, P.; Bagrel, D.; Kirsch, G.; Hesse, S.; et al. 2-(Thienothiazolylimino)-1,3-thiazolidin-4-ones inhibit cell division cycle 25 A phosphatase. *Bioorg Med Chem* **2016**, *24*, 2920-2928, doi:10.1016/j.bmc.2016.04.063.
122. Li, H.L.; Ma, Y.; Li, Y.; Chen, X.B.; Dong, W.L.; Wang, R.L. The design of novel inhibitors for treating cancer by targeting CDC25B through disruption of CDC25B-CDK2/Cyclin

A interaction using computational approaches. *Oncotarget* **2017**, *8*, 33225-33240, doi:10.18632/oncotarget.16600.
123. Kristjánsdóttir, K.; Rudolph, J. Cdc25 phosphatases and cancer. *Chem Biol* **2004**, *11*, 1043-1051, doi:10.1016/j.chembiol.2004.07.007.
124. Galaktionov, K.; Lee, A.K.; Eckstein, J.; Draetta, G.; Meckler, J.; Loda, M.; Beach, D. CDC25 phosphatases as potential human oncogenes. *Science* **1995**, *269*, 1575-1577, doi:10.1126/science.7667636.
125. Cangi, M.G.; Cukor, B.; Soung, P.; Signoretti, S.; Moreira, G.; Ranashinge, M.; Cady, B.; Pagano, M.; Loda, M. Role of the Cdc25A phosphatase in human breast cancer. *J Clin Invest* **2000**, *106*, 753-761, doi:10.1172/JCI9174.
126. Nishioka, K.; Doki, Y.; Shiozaki, H.; Yamamoto, H.; Tamura, S.; Yasuda, T.; Fujiwara, Y.; Yano, M.; Miyata, H.; Kishi, K.; et al. Clinical significance of CDC25A and CDC25B expression in squamous cell carcinomas of the oesophagus. *Br J Cancer* **2001**, *85*, 412-421, doi:10.1054/bjoc.2001.1934.
127. Takemasa, I.; Yamamoto, H.; Sekimoto, M.; Ohue, M.; Noura, S.; Miyake, Y.; Matsumoto, T.; Aihara, T.; Tomita, N.; Tamaki, Y.; et al. Overexpression of CDC25B phosphatase as a novel marker of poor prognosis of human colorectal carcinoma. *Cancer Res* **2000**, *60*, 3043-3050.
128. Broggini, M.; Buraggi, G.; Brenna, A.; Riva, L.; Codegoni, A.M.; Torri, V.; Lissoni, A.A.; Mangioni, C.; D'Incalci, M. Cell cycle-related phosphatases CDC25A and B expression correlates with survival in ovarian cancer patients. *Anticancer Research* **2000**, *20*, 4835-4840.
129. Kudo, Y.; Yasui, W.; Ue, T.; Yamamoto, S.; Yokozaki, H.; Nikai, H.; Tahara, E. Overexpression of cyclin-dependent kinase-activating CDC25B phosphatase in human gastric carcinomas. *Japanese Journal of Cancer Research* **1997**, *88*, 947-952, doi:10.1111/j.1349-7006.1997.tb00313.x.
130. Gasparotto, D.; Maestro, R.; Piccinin, S.; Vukosavljevic, T.; Barzan, L.; Sulfaro, S.; Boiocchi, M. Overexpression of CDC25A and CDC25B in head and neck cancers. *Cancer Research* **1997**, *57*, 2366-2368.
131. Zhang, Z.; Zhang, G.; Kong, C. High expression of Cdc25B and low expression of 14-3-3σ is associated with the development and poor prognosis in urothelial carcinoma of bladder. *Tumour Biol* **2014**, *35*, 2503-2512, doi:10.1007/s13277-013-1331-9.
132. Boutros, R.; Dozier, C.; Ducommun, B. The when and wheres of CDC25 phosphatases. *Curr Opin Cell Biol* **2006**, *18*, 185-191, doi:10.1016/j.ceb.2006.02.003.
133. Liu, T.; Lin, Y.; Wen, X.; Jorissen, R.N.; Gilson, M.K. BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities. *Nucleic Acids Res* **2007**, *35*, D198-201, doi:10.1093/nar/gkl999.
134. Lavecchia, A.; Cosconati, S.; Limongelli, V.; Novellino, E. Modeling of Cdc25B dual specifity protein phosphatase inhibitors: docking of ligands and enzymatic inhibition mechanism. *ChemMedChem* **2006**, *1*, 540-550, doi:10.1002/cmdc.200500092.
135. Lund, G.; Dudkin, S.; Borkin, D.; Ni, W.; Grembecka, J.; Cierpicki, T. Inhibition of CDC25B phosphatase through disruption of protein-protein interaction. *ACS Chem Biol* **2015**, *10*, 390-394, doi:10.1021/cb500883h.

136. Baell, J.B.; Holloway, G.A. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J Med Chem* **2010**, *53*, 2719-2740, doi:10.1021/jm901137j.
137. Lipinski, C.A.; Lombardo, F.; Dominy, B.W.; Feeney, P.J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* **2001**, *46*, 3-26, doi:10.1016/s0169-409x(00)00129-0.
138. Veber, D.F.; Johnson, S.R.; Cheng, H.Y.; Smith, B.R.; Ward, K.W.; Kopple, K.D. Molecular properties that influence the oral bioavailability of drug candidates. *J Med Chem* **2002**, *45*, 2615-2623, doi:10.1021/jm020017n.
139. Egan, W.J.; Merz, K.M.; Baldwin, J.J. Prediction of drug absorption using multivariate statistics. *J Med Chem* **2000**, *43*, 3867-3877, doi:10.1021/jm000292e.
140. Daina, A.; Michielin, O.; Zoete, V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci Rep* **2017**, *7*, 42717, doi:10.1038/srep42717.
141. Huber, K.V.; Salah, E.; Radic, B.; Gridling, M.; Elkins, J.M.; Stukalov, A.; Jemth, A.S.; Göktürk, C.; Sanjiv, K.; Strömberg, K.; et al. Stereospecific targeting of MTH1 by (S)-crizotinib as an anticancer strategy. *Nature* **2014**, *508*, 222-227, doi:10.1038/nature13194.
142. Hashiguchi, K.; Hayashi, M.; Sekiguchi, M.; Umezu, K. The roles of human MTH1, MTH2 and MTH3 proteins in maintaining genome stability under oxidative stress. *Mutat Res* **2018**, *808*, 10-19, doi:10.1016/j.mrfmmm.2018.01.002.
143. McPherson, L.A.; Troccoli, C.I.; Ji, D.; Bowles, A.E.; Gardiner, M.L.; Mohsen, M.G.; Nagathihalli, N.S.; Nguyen, D.M.; Robbins, D.J.; Merchant, N.B.; et al. Increased MTH1-specific 8-oxodGTPase activity is a hallmark of cancer in colon, lung and pancreatic tissue. *DNA Repair (Amst)* **2019**, *83*, 102644, doi:10.1016/j.dnarep.2019.102644.
144. Savickiene, J.; Borutinskaite, V.V.; Treigyte, G.; Magnusson, K.E.; Navakauskiene, R. The novel histone deacetylase inhibitor BML-210 exerts growth inhibitory, proapoptotic and differentiation stimulating effects on the human leukemia cell lines. *Eur J Pharmacol* **2006**, *549*, 9-18, doi:10.1016/j.ejphar.2006.08.010.
145. Borutinskaite, V.V.; Magnusson, K.E.; Navakauskiene, R. Histone deacetylase inhibitor BML-210 induces growth inhibition and apoptosis and regulates HDAC and DAPC complex expression levels in cervical cancer cells. *Mol Biol Rep* **2012**, *39*, 10179-10186, doi:10.1007/s11033-012-1892-5.
146. Jaschevatzky, O.E.; Anderman, S.; Shalit, A.; Kampf, D.; Grünstein, S. The treatment of postmenopausal syndrome by monthly oral doses of quinestrol. *Acta Obstet Gynecol Scand* **1979**, *58*, 175-178, doi:10.3109/00016347909154577.
147. Zheng, C.J.; Yang, L.L.; Liu, J.; Zhong, L. JTC-801 exerts anti-proliferative effects in human osteosarcoma cells by inducing apoptosis. *J Recept Signal Transduct Res* **2018**, *38*, 133-140, doi:10.1080/10799893.2018.1436561.
148. Zhao, B.; Hu, T. JTC-801 inhibits the proliferation and metastasis of the Hep G2 hepatoblastoma cell line by regulating the phosphatidylinositol 3-kinase/protein

kinase B signalling pathway. *Oncol Lett* **2019**, *17*, 1939-1945, doi:10.3892/ol.2018.9780.
149. Riou, P.; Saffroy, R.; Chenailler, C.; Franc, B.; Gentile, C.; Rubinstein, E.; Resink, T.; Debuire, B.; Piatier-Tonneau, D.; Lemoine, A. Expression of T-cadherin in tumor cells influences invasive potential of human hepatocellular carcinoma. *FASEB J* **2006**, *20*, 2291-2301, doi:10.1096/fj.06-6085com.
150. Kabakci, Z.; Käppeli, S.; Cantù, C.; Jensen, L.D.; König, C.; Toggweiler, J.; Gentili, C.; Ribaudo, G.; Zagotto, G.; Basler, K.; et al. Pharmacophore-guided discovery of CDC25 inhibitors causing cell cycle arrest and tumor regression. *Sci Rep* **2019**, *9*, 1335, doi:10.1038/s41598-019-38579-7.
151. Capasso, A.; Cerchia, C.; Di Giovanni, C.; Granato, G.; Albano, F.; Romano, S.; De Vendittis, E.; Ruocco, M.R.; Lavecchia, A. Ligand-based chemoinformatic discovery of a novel small molecule inhibitor targeting CDC25 dual specificity phosphatases and displaying in vitro efficacy against melanoma cells. *Oncotarget* **2015**, *6*, 40202-40222, doi:10.18632/oncotarget.5473.
152. Hsiao, Y.P.; Tsai, C.H.; Wu, P.P.; Hsu, S.C.; Liu, H.C.; Huang, Y.P.; Yang, J.H.; Chung, J.G. Cantharidin induces G2/M phase arrest by inhibition of Cdc25c and Cyclin A and triggers apoptosis through reactive oxygen species and the mitochondria-dependent pathways of A375.S2 human melanoma cells. *Int J Oncol* **2014**, *45*, 2393-2402, doi:10.3892/ijo.2014.2689.
153. Hung, F.M.; Chen, Y.L.; Huang, A.C.; Hsiao, Y.P.; Yang, J.S.; Chung, M.T.; Chueh, F.S.; Lu, H.F.; Chung, J.G. Triptolide induces S phase arrest via the inhibition of cyclin E and CDC25A and triggers apoptosis via caspase- and mitochondrial-dependent signaling pathways in A375.S2 human melanoma cells. *Oncol Rep* **2013**, *29*, 1053-1060, doi:10.3892/or.2013.2230.
154. NCI60 Growth Inhibition Data - Download NCI Cell Line Data. Available online: https://wiki.nci.nih.gov/display/NCIDTPdata/NCI-60+Growth+Inhibition+Data (accessed on November 10, 2022
155. https://dtp.cancer.gov/discovery_development/nci-60/characterization.htm. Available online: (accessed on November 10, 2022
156. Gaulton, A.; Bellis, L.J.; Bento, A.P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B.; et al. ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res* **2012**, *40*, D1100-1107, doi:10.1093/nar/gkr777.
157. Wang, Y.; Xiao, J.; Suzek, T.O.; Zhang, J.; Wang, J.; Bryant, S.H. PubChem: a public information system for analyzing bioactivities of small molecules. *Nucleic Acids Res* **2009**, *37*, W623-633, doi:10.1093/nar/gkp456.
158. Lauria, A.; Patella, C.; Abbate, I.; Martorana, A.; Almerico, A.M. Lead optimization through VLAK protocol: new annelated pyrrolo-pyrimidine derivatives as antitumor agents. *Eur J Med Chem* **2012**, *55*, 375-383, doi:10.1016/j.ejmech.2012.07.046.
159. Sherman, W.; Day, T.; Jacobson, M.P.; Friesner, R.A.; Farid, R. Novel procedure for modeling ligand/receptor induced fit effects. *J Med Chem* **2006**, *49*, 534-553, doi:10.1021/jm050540c.

160. Sherman, W.; Beard, H.S.; Farid, R. Use of an induced fit receptor structure in virtual screening. *Chem Biol Drug Des* **2006**, *67*, 83-84, doi:10.1111/j.1747-0285.2005.00327.x.
161. *Maestro, version 9.3, Schrödinger, LLC, New York, NY, 2012.*
162. Zhong, H.; Tran, L.M.; Stang, J.L. Induced-fit docking studies of the active and inactive states of protein tyrosine kinases. *J Mol Graph Model* **2009**, *28*, 336-346, doi:10.1016/j.jmgm.2009.08.012.
163. Martorana, A.; Lauria, A. Design of antitumor drugs targeting c-kit receptor by a new mixed ligand-structure based method. *J Mol Graph Model* **2020**, *100*, 107666, doi:10.1016/j.jmgm.2020.107666.
164. Martorana, A.; Gentile, C.; Lauria, A. In Silico Insights into the SARS CoV-2 Main Protease Suggest NADH Endogenous Defences in the Control of the Pandemic Coronavirus Infection. *Viruses* **2020**, *12*, doi:10.3390/v12080805.
165. Wang, H.; Aslanian, R.; Madison, V.S. Induced-fit docking of mometasone furoate and further evidence for glucocorticoid receptor 17alpha pocket flexibility. *J Mol Graph Model* **2008**, *27*, 512-521, doi:10.1016/j.jmgm.2008.09.002.
166. Friesner, R.A.; Banks, J.L.; Murphy, R.B.; Halgren, T.A.; Klicic, J.J.; Mainz, D.T.; Repasky, M.P.; Knoll, E.H.; Shelley, M.; Perry, J.K.; et al. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem* **2004**, *47*, 1739-1749, doi:10.1021/jm0306430.
167. Friesner, R.A.; Murphy, R.B.; Repasky, M.P.; Frye, L.L.; Greenwood, J.R.; Halgren, T.A.; Sanschagrin, P.C.; Mainz, D.T. Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J Med Chem* **2006**, *49*, 6177-6196, doi:10.1021/jm051256o.
168. Halgren, T.A.; Murphy, R.B.; Friesner, R.A.; Beard, H.S.; Frye, L.L.; Pollard, W.T.; Banks, J.L. Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J Med Chem* **2004**, *47*, 1750-1759, doi:10.1021/jm030644s.
169. Jacobson, M.P.; Friesner, R.A.; Xiang, Z.; Honig, B. On the role of the crystal environment in determining protein side-chain conformations. *J Mol Biol* **2002**, *320*, 597-608, doi:10.1016/s0022-2836(02)00470-9.
170. Jacobson, M.P.; Pincus, D.L.; Rapp, C.S.; Day, T.J.; Honig, B.; Shaw, D.E.; Friesner, R.A. A hierarchical approach to all-atom protein loop prediction. *Proteins* **2004**, *55*, 351-367, doi:10.1002/prot.10613.
171. Compound Submission for NCI60 Testing - Selection Guidelines. Available online: https://dtp.cancer.gov/organization/dscb/compoundSubmission/structureSelection.htm (accessed on November 10, 2022
172. NCI60 Screening Methodology - NCI60 Cell One/Five Doses Screen. Available online: https://dtp.cancer.gov/discovery_development/nci-60/methodology.htm (accessed on November 10, 2022
173. NCI-60 Human Cancer Cell Line Screen - Standard Operating Procedures for Sample Preparation for NCI60 Screen. Available online: https://dtp.cancer.gov/discovery_development/nci-60/handling.htm (accessed on November 10, 2022

174. Mosmann, T. Rapid colorimetric assay for cellular growth and survival: Application to proliferation and cytotoxicity assays. *Journal of Immunological Methods* **1983**, *65*, 55-63, doi:10.1016/0022-1759(83)90303-4.
175. Lauria, A.; Alfio, A.; Bonsignore, R.; Gentile, C.; Martorana, A.; Gennaro, G.; Barone, G.; Terenzi, A.; Almerico, A.M. New benzothieno[3,2-d]-1,2,3-triazines with antiproliferative activity: synthesis, spectroscopic studies, and biological activity. *Bioorg Med Chem Lett* **2014**, *24*, 3291-3297, doi:10.1016/j.bmcl.2014.06.007.
176. Martorana, A.; Gentile, C.; Perricone, U.; Piccionello, A.P.; Bartolotta, R.; Terenzi, A.; Pace, A.; Mingoia, F.; Almerico, A.M.; Lauria, A. Synthesis, antiproliferative activity, and in silico insights of new 3-benzoylamino-benzo[b]thiophene derivatives. *Eur J Med Chem* **2015**, *90*, 537-546, doi:10.1016/j.ejmech.2014.12.002.
177. Lauria, A.; Abbate, I.; Gentile, C.; Angileri, F.; Martorana, A.; Almerico, A.M. Synthesis and biological activities of a new class of heat shock protein 90 inhibitors, designed by energy-based pharmacophore virtual screening. *J Med Chem* **2013**, *56*, 3424-3428, doi:10.1021/jm4002023.
178. Lanyon-Hogg, T. Targeting the bacterial SOS response for new antimicrobial agents: drug targets, molecular mechanisms and inhibitors. *Future Med Chem* **2021**, *13*, 143-155, doi:10.4155/fmc-2020-0310.
179. Lim, C.S.Q.; Ha, K.P.; Clarke, R.S.; Gavin, L.A.; Cook, D.T.; Hutton, J.A.; Sutherell, C.L.; Edwards, A.M.; Evans, L.E.; Tate, E.W.; et al. Identification of a potent small-molecule inhibitor of bacterial DNA repair that potentiates quinolone antibiotic activity in methicillin-resistant Staphylococcus aureus. *Bioorg Med Chem* **2019**, *27*, 114962, doi:10.1016/j.bmc.2019.06.025.
180. Hutchings, M.I.; Truman, A.W.; Wilkinson, B. Antibiotics: past, present and future. *Curr Opin Microbiol* **2019**, *51*, 72-80, doi:10.1016/j.mib.2019.10.008.
181. Cook, M.A.; Wright, G.D. The past, present, and future of antibiotics. *Sci Transl Med* **2022**, *14*, eabo7793, doi:10.1126/scitranslmed.abo7793.
182. Ali, J.; Rafiq, Q.A.; Ratcliffe, E. Antimicrobial resistance mechanisms and potential synthetic treatments. *Future Sci OA* **2018**, *4*, FSO290, doi:10.4155/fsoa-2017-0109.
183. Schrader, S.M.; Vaubourgeix, J.; Nathan, C. Biology of antimicrobial resistance and approaches to combat it. *Sci Transl Med* **2020**, *12*, doi:10.1126/scitranslmed.aaz6992.
184. Hwang, A.Y.; Gums, J.G. The emergence and evolution of antimicrobial resistance: Impact on a global scale. *Bioorg Med Chem* **2016**, *24*, 6440-6445, doi:10.1016/j.bmc.2016.04.027.
185. Church, N.A.; McKillip, J.L. Antibiotic resistance crisis: challenges and imperatives. *Biologia* **2021**, *76*, 1535-1550, doi:10.1007/s11756-021-00697-x.
186. Matos de Opitz, C.L.; Sass, P. Tackling antimicrobial resistance by exploring new mechanisms of antibiotic action. *Future Microbiol* **2020**, *15*, 703-708, doi:10.2217/fmb-2020-0048.
187. Brown, D.G.; Wobst, H.J. A Decade of FDA-Approved Drugs (2010-2019): Trends and Future Directions. *J Med Chem* **2021**, *64*, 2312-2338, doi:10.1021/acs.jmedchem.0c01516.

188. Mullard, A. 2020 FDA drug approvals. *Nat Rev Drug Discov* **2021**, *20*, 85-90, doi:10.1038/d41573-021-00002-0.
189. Mullard, A. 2021 FDA approvals. *Nat Rev Drug Discov* **2022**, *21*, 83-88, doi:10.1038/d41573-022-00001-9.
190. Annunziato, G. Strategies to Overcome Antimicrobial Resistance (AMR) Making Use of Non-Essential Target Inhibitors: A Review. *Int J Mol Sci* **2019**, *20*, doi:10.3390/ijms20235844.
191. Thakare, R.; Kesharwani, P.; Dasgupta, A.; Srinivas, N.; Chopra, S. Chapter 1 - Antibiotics: past, present, and future. In *Drug Discovery Targeting Drug-Resistant Bacteria*; Academic Press: 2020; pp. 1-8.
192. Kinch, M.S.; Patridge, E.; Plummer, M.; Hoyer, D. An analysis of FDA-approved drugs for infectious disease: antibacterial agents. *Drug Discov Today* **2014**, *19*, 1283-1287, doi:10.1016/j.drudis.2014.07.005.
193. Fisher, J.F.; Mobashery, S. Endless Resistance. Endless Antibiotics? *Medchemcomm* **2016**, *7*, 37-49, doi:10.1039/C5MD00394F.
194. Roope, L.S.J.; Smith, R.D.; Pouwels, K.B.; Buchanan, J.; Abel, L.; Eibich, P.; Butler, C.C.; Tan, P.S.; Walker, A.S.; Robotham, J.V.; et al. The challenge of antimicrobial resistance: What economics can contribute. *Science* **2019**, *364*, doi:10.1126/science.aau4679.
195. Melander, R.J.; Melander, C. The Challenge of Overcoming Antibiotic Resistance: An Adjuvant Approach? *ACS Infect Dis* **2017**, *3*, 559-563, doi:10.1021/acsinfecdis.7b00071.
196. Wigley, D.B. Bacterial DNA repair: recent insights into the mechanism of RecBCD, AddAB and AdnAB. *Nat Rev Microbiol* **2013**, *11*, 9-13, doi:10.1038/nrmicro2917.
197. Karabulut, A.C.; Cirz, R.T.; Taylor, A.F.; Smith, G.R. Small-molecule sensitization of RecBCD helicase-nuclease to a Chi hotspot-activated state. *Nucleic Acids Res* **2020**, *48*, 7973-7980, doi:10.1093/nar/gkaa534.
198. Memar, M.Y.; Yekani, M.; Celenza, G.; Poortahmasebi, V.; Naghili, B.; Bellio, P.; Baghi, H.B. The central role of the SOS DNA repair system in antibiotics resistance: A new target for a new infectious treatment strategy. *Life Sci* **2020**, *262*, 118562, doi:10.1016/j.lfs.2020.118562.
199. Maslowska, K.H.; Makiela-Dzbenska, K.; Fijalkowska, I.J. The SOS system: A complex and tightly regulated response to DNA damage. *Environ Mol Mutagen* **2019**, *60*, 368-384, doi:10.1002/em.22267.
200. Qin, T.T.; Kang, H.Q.; Ma, P.; Li, P.P.; Huang, L.Y.; Gu, B. SOS response and its regulation on the fluoroquinolone resistance. *Ann Transl Med* **2015**, *3*, 358, doi:10.3978/j.issn.2305-5839.2015.12.09.
201. Pourahmad Jaktaji, R.; Pasand, S. Overexpression of SOS genes in ciprofloxacin resistant Escherichia coli mutants. *Gene* **2016**, *576*, 115-118, doi:10.1016/j.gene.2015.09.069.
202. Aldred, K.J.; Kerns, R.J.; Osheroff, N. Mechanism of quinolone action and resistance. *Biochemistry* **2014**, *53*, 1565-1574, doi:10.1021/bi5000564.
203. Amundsen, S.K.; Fero, J.; Salama, N.R.; Smith, G.R. Dual nuclease and helicase activities of Helicobacter pylori AddAB are required for DNA repair, recombination,

and mouse infectivity. *J Biol Chem* **2009**, *284*, 16759-16766, doi:10.1074/jbc.M109.005587.
204. Cirz, R.T.; Chin, J.K.; Andes, D.R.; de Crécy-Lagard, V.; Craig, W.A.; Romesberg, F.E. Inhibition of mutation and combating the evolution of antibiotic resistance. *PLoS Biol* **2005**, *3*, e176, doi:10.1371/journal.pbio.0030176.
205. Diaz-Diaz, S.; Recacha, E.; Machuca, J.; García-Duque, A.; Docobo-Pérez, F.; Blázquez, J.; Pascual, A.; Rodríguez-Martínez, J.M. Synergistic Quinolone Sensitization by Targeting the *recA* SOS Response Gene and Oxidative Stress. *Antimicrob Agents Chemother* **2021**, *65*, doi:10.1128/AAC.02004-20.
206. Alam, M.K.; Alhhazmi, A.; DeCoteau, J.F.; Luo, Y.; Geyer, C.R. RecA Inhibitors Potentiate Antibiotic Activity and Block Evolution of Antibiotic Resistance. *Cell Chem Biol* **2016**, *23*, 381-391, doi:10.1016/j.chembiol.2016.02.010.
207. Recacha, E.; Machuca, J.; Díaz-Díaz, S.; García-Duque, A.; Ramos-Guelfo, M.; Docobo-Pérez, F.; Blázquez, J.; Pascual, A.; Rodríguez-Martínez, J.M. Suppression of the SOS response modifies spatiotemporal evolution, post-antibiotic effect, bacterial fitness and biofilm formation in quinolone-resistant Escherichia coli. *J Antimicrob Chemother* **2019**, *74*, 66-73, doi:10.1093/jac/dky407.
208. Podlesek, Z.; Žgur Bertok, D. The DNA Damage Inducible SOS Response Is a Key Player in the Generation of Bacterial Persister Cells and Population Wide Tolerance. *Front Microbiol* **2020**, *11*, 1785, doi:10.3389/fmicb.2020.01785.
209. Recacha, E.; Machuca, J.; Díaz de Alba, P.; Ramos-Güelfo, M.; Docobo-Pérez, F.; Rodriguez-Beltrán, J.; Blázquez, J.; Pascual, A.; Rodríguez-Martínez, J.M. Quinolone Resistance Reversion by Targeting the SOS Response. *mBio* **2017**, *8*, doi:10.1128/mBio.00971-17.
210. Mo, C.Y.; Culyba, M.J.; Selwood, T.; Kubiak, J.M.; Hostetler, Z.M.; Jurewicz, A.J.; Keller, P.M.; Pope, A.J.; Quinn, A.; Schneck, J.; et al. Inhibitors of LexA Autoproteolysis and the Bacterial SOS Response Discovered by an Academic-Industry Partnership. *ACS Infect Dis* **2018**, *4*, 349-359, doi:10.1021/acsinfecdis.7b00122.
211. Yakimov, A.; Bakhlanova, I.; Baitin, D. Targeting evolution of antibiotic resistance by SOS response inhibition. *Comput Struct Biotechnol J* **2021**, *19*, 777-783, doi:10.1016/j.csbj.2021.01.003.
212. Culyba, M.J.; Mo, C.Y.; Kohli, R.M. Targets for Combating the Evolution of Acquired Antibiotic Resistance. *Biochemistry* **2015**, *54*, 3573-3582, doi:10.1021/acs.biochem.5b00109.
213. Amundsen, S.K.; Spicer, T.; Karabulut, A.C.; Londoño, L.M.; Eberhart, C.; Fernandez Vega, V.; Bannister, T.D.; Hodder, P.; Smith, G.R. Small-molecule inhibitors of bacterial AddAB and RecBCD helicase-nuclease DNA repair enzymes. *ACS Chem Biol* **2012**, *7*, 879-891, doi:10.1021/cb300018x.
214. Liu, H.H. Safety profile of the fluoroquinolones: focus on levofloxacin. *Drug Saf* **2010**, *33*, 353-369, doi:10.2165/11536360-000000000-00000.
215. Tanne, J.H. FDA adds "black box" warning label to fluoroquinolone antibiotics. *BMJ* **2008**, *337*, a816, doi:10.1136/bmj.a816.

216. Gatadi, S.; Lakshmi, T.V.; Nanduri, S. 4(3H)-Quinazolinone derivatives: Promising antibacterial drug leads. *Eur J Med Chem* **2019**, *170*, 157-172, doi:10.1016/j.ejmech.2019.03.018.
217. Jones, M.L.; Lilly, J.C.; Ankala, S.V.; Singleton, S.F. Preparation of quinazolinones, indazoles and other heterocyclic compounds as SOS inhibitors and antibiotic potentiators. WO2016094730 A1 2016-06-16.
218. Ellsworth, E.L.; Showalter, H.D.H.; Powell, S.A.; Sanchez, J.P.; Kerschen, J.A.; Stier, M.A.; Tran, T.P. Preparation of quinazolinediones as antibacterial agents for quinolone-resistant bacteria. WO2002102793 A2 2002-12-27.
219. Bird, P.; Ellsworth, E.L.; Nguyen, D.Q.; Sanchez, J.P.; Showalter, H.D.H.; Singh, R.; Stier, M.A.; Tran, T.P.; Watson, B.M.; Yip, J. Preparation of 3-aminoquinazoline-2,4-dione antibacterial agents. WO2001053273 A1 2001-07-26.
220. He, L.; Li, H.; Chen, J.; Wu, X.F. Recent advances in 4(3H)-quinazolinone syntheses. *RSC Advances* **2014**, *4*, 12065-12077, doi:10.1039/c4ra00351a.
221. Gheidari, D.; Mehrdad, M.; Maleki, S. Recent advances in synthesis of quinazoline-2,4(1H,3H)-diones: Versatile building blocks in N-heterocyclic compounds. *Applied Organometallic Chemistry* **2022**, *36*, doi:10.1002/aoc.6631.
222. Zhang, Y.; Hou, Q.; Li, X.; Zhu, J.; Wang, W.; Li, B.; Zhao, L.; Xia, H. Enrichment of novel quinazoline derivatives with high antitumor activity in mitochondria tracked by its self-fluorescence. *Eur J Med Chem* **2019**, *178*, 417-432, doi:10.1016/j.ejmech.2019.06.015.
223. Schenkel, L.B.; Olivieri, P.R.; Boezio, A.A.; Deak, H.L.; Emkey, R.; Graceffa, R.F.; Gunaydin, H.; Guzman-Perez, A.; Lee, J.H.; Teffera, Y.; et al. Optimization of a Novel Quinazolinone-Based Series of Transient Receptor Potential A1 (TRPA1) Antagonists Demonstrating Potent in Vivo Activity. *J Med Chem* **2016**, *59*, 2794-2809, doi:10.1021/acs.jmedchem.6b00039.
224. Heppell, J.T.; Al-Rawi, J.M.A. Synthesis, antibacterial, and DNA-PK evaluation of some novel 6-fluoro-7-(cyclic amino)-2-(thioxo or oxo)-3-substituted quinazolin-4-ones as structural analogues of quinolone and quinazolin-2,4-dione antibiotics. *Medicinal Chemistry Research* **2015**, *24*, 2756-2769, doi:10.1007/s00044-015-1336-6.
225. Sekiguchi, Y.; Kanuma, K.; Omodera, K.; Busujima, T.; Tran, T.-A.; Han, S.; Casper, M.; Kramer, B.A. Preparation of novel quinazolines as MCH receptor antagonists. WO2004087680 A1 2004-10-14.
226. Furuya, S.; Choh, N.; Ohtaki, T.; Watanabe, T. Thienopyrimidine derivatives, their production, and use as endothelin antagonists. EP640606 A1 1995-03-01.
227. Shimizu, K.; Shimizu, T.; Kawakami, K.; Nakoji, M.; Sakai, T. Preparation of pyridines and related compounds as TGF-β inhibitors. EP1724268A1 2006-11-22.
228. Tobe, M.; Isobe, Y.; Tomizawa, H.; Nagasaki, T.; Obara, F.; Hayashi, H. Structure-activity relationships of 6-fluoroquinazolines: Dual-Acting compounds with inhibitory activities toward both TNF-α production and T cell proliferation. *Bioorganic and Medicinal Chemistry* **2002**, *11*, 609-616, doi:10.1016/S0968-0896(02)00338-3.

229. Shan, C.; Yan, J.W.; Wang, Y.Q.; Che, T.; Huang, Z.L.; Chen, A.C.; Yao, P.F.; Tan, J.H.; Li, D.; Ou, T.M.; et al. Design, Synthesis, and Evaluation of Isaindigotone Derivatives To Downregulate c-myc Transcription via Disrupting the Interaction of NM23-H2 with G-Quadruplex. *J Med Chem* **2017**, *60*, 1292-1308, doi:10.1021/acs.jmedchem.6b01218.

List of Abbreviations

AAP *– Antiproliferative Activity Predictor tool*

ABL *– Abelson kinase*

ALL *– Acute Lymphoblastic Leukemia*

AMR *– Antimicrobial Resistance*

AUC *– Area Under the Curve*

BCR *- Breakpoint Cluster Region protein*

CCLE *– Cancer Cell Line Encyclopedia*

Cdc25 *- Cell division cycle 25 phosphatase*

Cdk *– Cyclin dependent kinase*

CFX *- Ciprofloxacin*

CGP *– Comprehensive Genomic Profiling*

CML *- Chronic Myeloid Leukemia*

CL *– Cell Line module (DRUDIT)*

CNS *– Central Nervous System*

CRISPR *- Clustered Regularly Interspaced Short Palindromic Repeats*

CTRP *– Cancer Therapeutics Response Portal*

DAS *– Drudit Affinity Score*

DCM *- Dichloromethane*

DIPEA *– N,N-Diisopropylethylamine*

DMF *- Dimethylformamide*

DMSO *- Dimethylsulfoxide*

DNMDP *- 6-(4-(diethylamino)-3-nitrophenyl)-5-methyl-4,5-dihydropyridazin-3(2H)-one*

DRUDIT^{online} *- DRUg DIscovery Tools online platform*

DSB *– DNA double strand break*

DTP *– Developmental Therapeutic Program (NCI)*

|DTV (GI_{50})| *- Absolute deviation of the predicted GI_{50} from the experimental GI_{50}*

EGFR *– Epidermal Growth Factor Receptor*

EP *– Protein expression pattern*

FDA *– Food and Drug Administration*

FP *– Finger Print module (DRUDIT)*

GDA *- Genomic and Drugs Integrated Analysis database*

GDSC *– Genomic and Drug Sensitivity in Cancer portal*

GEDA *- Gene Expression and Drug Activity*

G% *- Percentage of cancer cell growth inhibition*

GI_{50} *– Growth inhibition 50%, Half-maximal cell growth inhibitory concertation*

GI_{50}(CL) *- Growth inhibition 50% predicted by the CL module (AAP tool).*

GI_{50}(FP) *- Growth inhibition 50% predicted by the FP module (AAP tool).*

HAT *– Histone acetyl transferase*

HDAC *– Histone Deacetylase*

HPLC *– High-Performance Liquid Chromatography*

IC_{50} *– Half-maximal inhibitory concentration*

IFD *– Induced-Fit Docking simulation*

LC_{50} *– lethal concentration 50*

MeOH *- Methanol*

MG_MID *– Mean Graph MIDpoint*

MOLDESTO *– MOLecular DEScriptor TOols*

MRSA *– Methycillin resistant Staphilococcus Aureus*

NCI *- National Cancer Institute*

NCI2014DB *– Structure database released by NCI till 2014, training set*

NCI2016DB *– Structure database released by NCI after 2014, test set*

NEP *– Normalized Protein Expression pattern data*

NGI_{50} *– Normalized GI_{50} data*

OD *– Optical Density*

ORL1 *- Opioid receptor-like 1 receptor*

PAINS *- Pan-assay interference compounds*

PCA *– Principal Component Analysis*

PDB *– Protein Data Bank*

PDE3A *– Phosphodiesterase 3A*

Rf *– Retention factor (chromatography)*

ROS *– Reactive oxygen species*

RT *– Room temperature*

SAR *– Structure-activity relationship*

SnAr *– Nucleophilic Aromatic Substitution*

TGI *– Concentration for total inhibition of cancer cell proliferation*

TLC *- Thin Layer Chromatography*

UV/vis *– Ultraviolet/visible light*

Supplementary Materials:

The following are available online at:

https://drive.google.com/drive/folders/14k6HZl4CqB0FH7lhGjCissiZ1W1f0Z7v?usp=sharing

S1: list of molecular descriptors implemented in MOLDESTO (.pdf).

S2: list of the selected structures from NCI2014DB for internal validation and relative .mol files (.zip).

S3: experimental GI50s of the NCI2014DB structures selected for internal test (.xlsx).

S4: 18 output matrices of GI50(CL) calculation for internal validation (.xlsx).

S5: analysis of $|DTV(GI_{50})|$ for the 18 matrices (.xlsx).

S6: list of the selected structures from NCI2016DB for external validation and relative .mol files (.zip).

S7: $|DTV(GI_{50})|$ obtained for the external validation (.xlsx).

S8: $|DTV(GI_{50})|$ obtained for the FP external validation (.xlsx).

S9: Parameter optimization for cell line-specific activity prediction (.xlsx).

S10: APP parameter optimization for curcumin (.xlsx).

S11: AAP output matrix for the in-house database of curcumin-like compounds (.xlsx).

S12: One dose response overview and mean graphs for the 5 selected compounds (.zip).

S13: Five-dose response data for **1a** and **3e** (.zip).

S14: Comparison GI_{50} experimental- predicted for **1a** and **3e** (.xlsx).

S15: Case studies from NCI database for the correlation drug activity - protein expression (.xlsx);

S16: 117 Cdc25 inhibitors selected to build the template (.pdf).

S17: Cdc25 DAS calculated for purchasable structures database (.xlsx).

S18: 106 selected structures with Cdc25 DAS above 0.8 (.pdf).

S19: Cdc25s normalized NCI expression (.xlsx).

S20: GI_{50} predicted values for the 106 selected structures (.xlsx).

S21: Matching of Cdc25s EP-GI_{50} for the 106 selected structures (.xlsx).

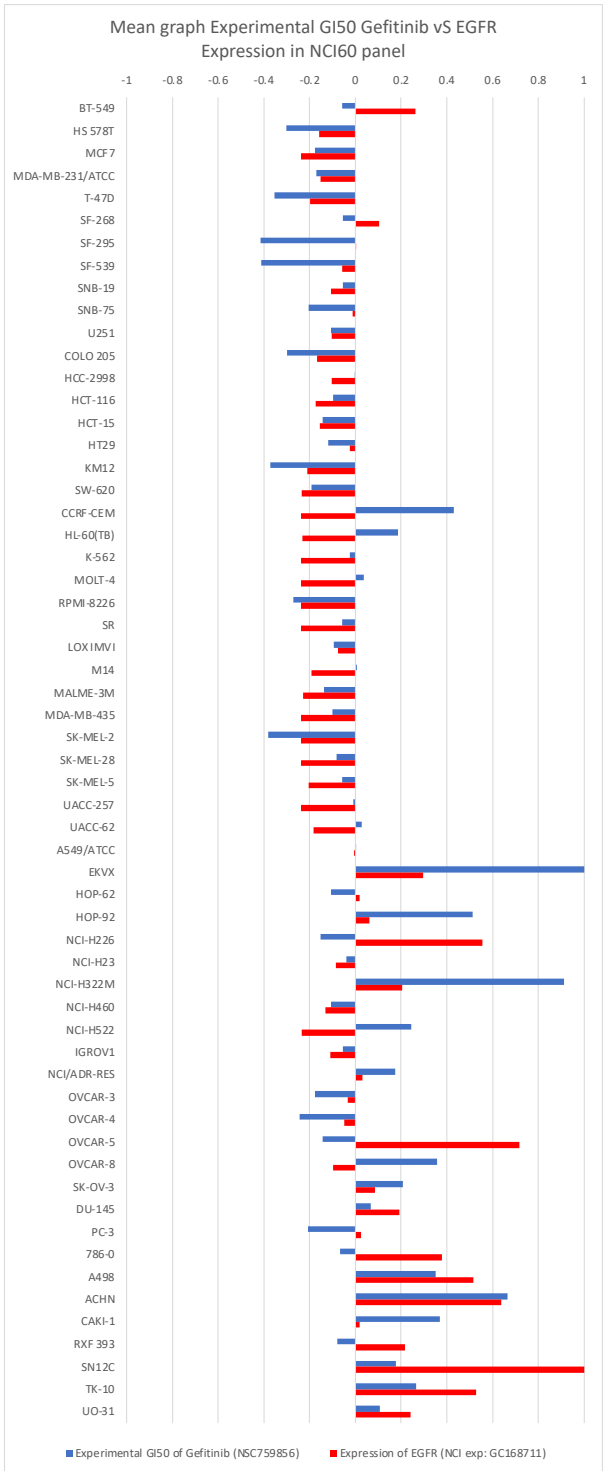
S22: Molecular descriptors matrix of the template and selected structures and PCA results (.xlsx).

S23: HRMS data for the *in vitro* tested compounds (.pdf).

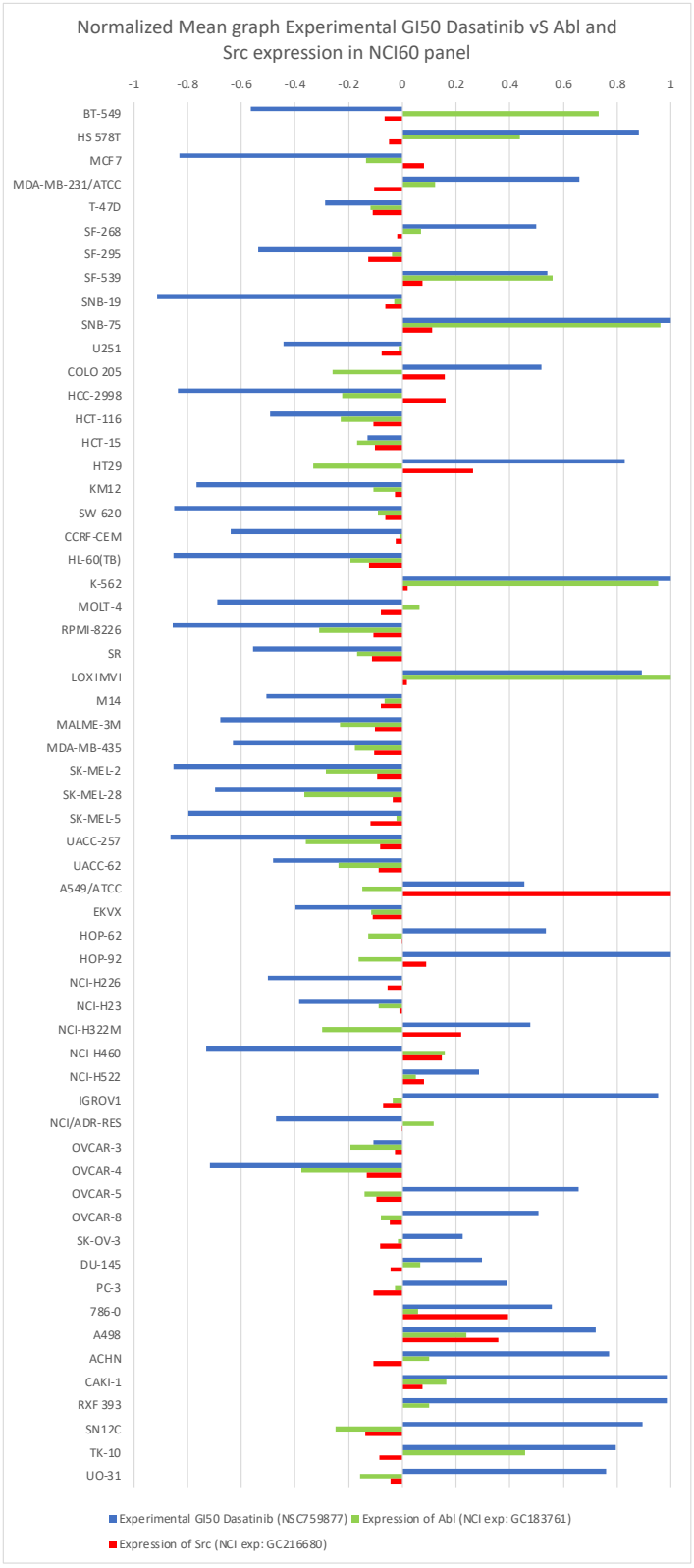
S24: Dose-response curves for the data presented in Table 8 (.pdf).

Supplementary figures:

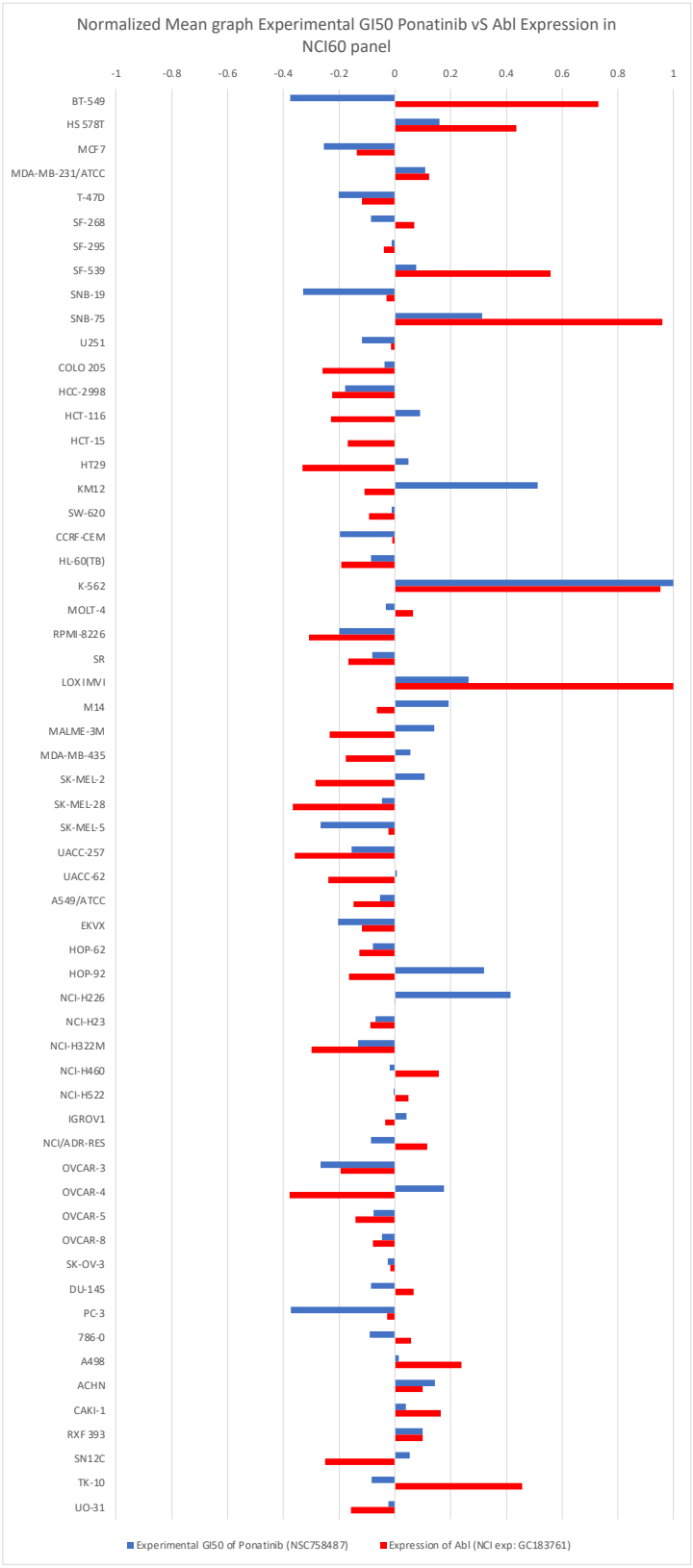
Supplementary Figure 1. *Normalized Mean graph Experimental GI_{50} Gefitinib vS EGFR Expression in NCI-60 panel.*



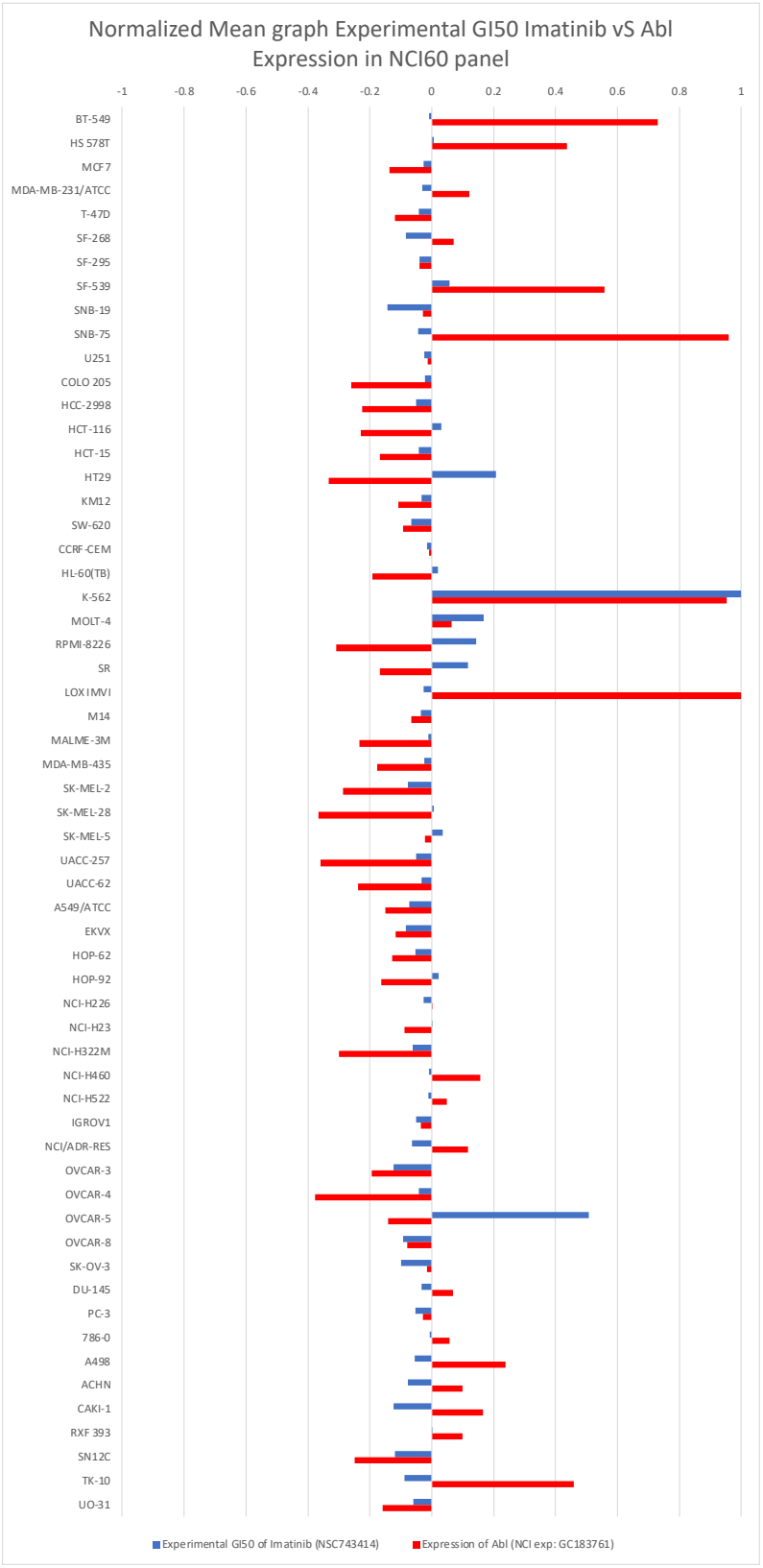
Supplementary Figure 2. *Normalized Mean graph Experimental GI_{50} Dasatinib vS Abl and Src expression in NCI-60 panel.*



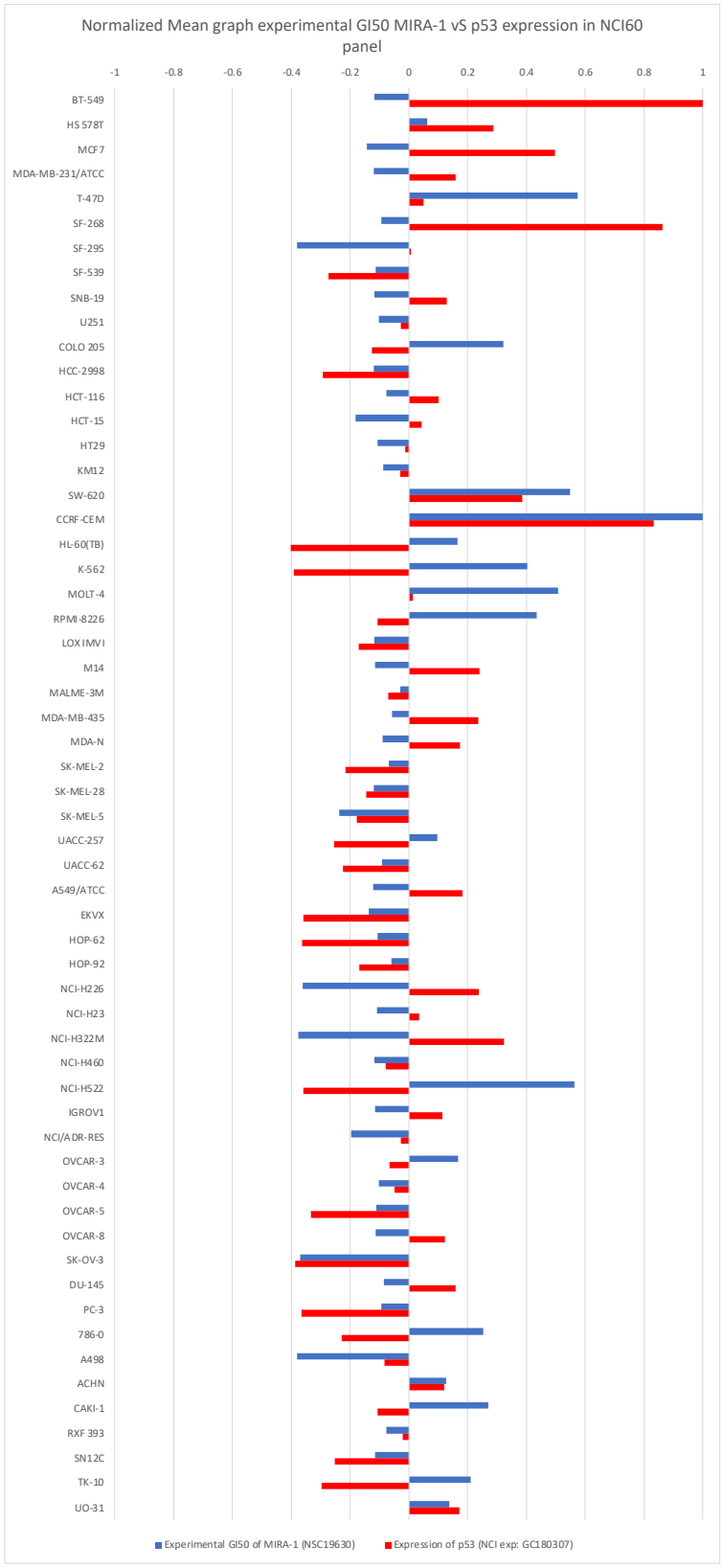
Supplementary Figure 3. *Normalized Mean graph Experimental GI_{50} Ponatinib vS Abl Expression in NCI-60 panel.*



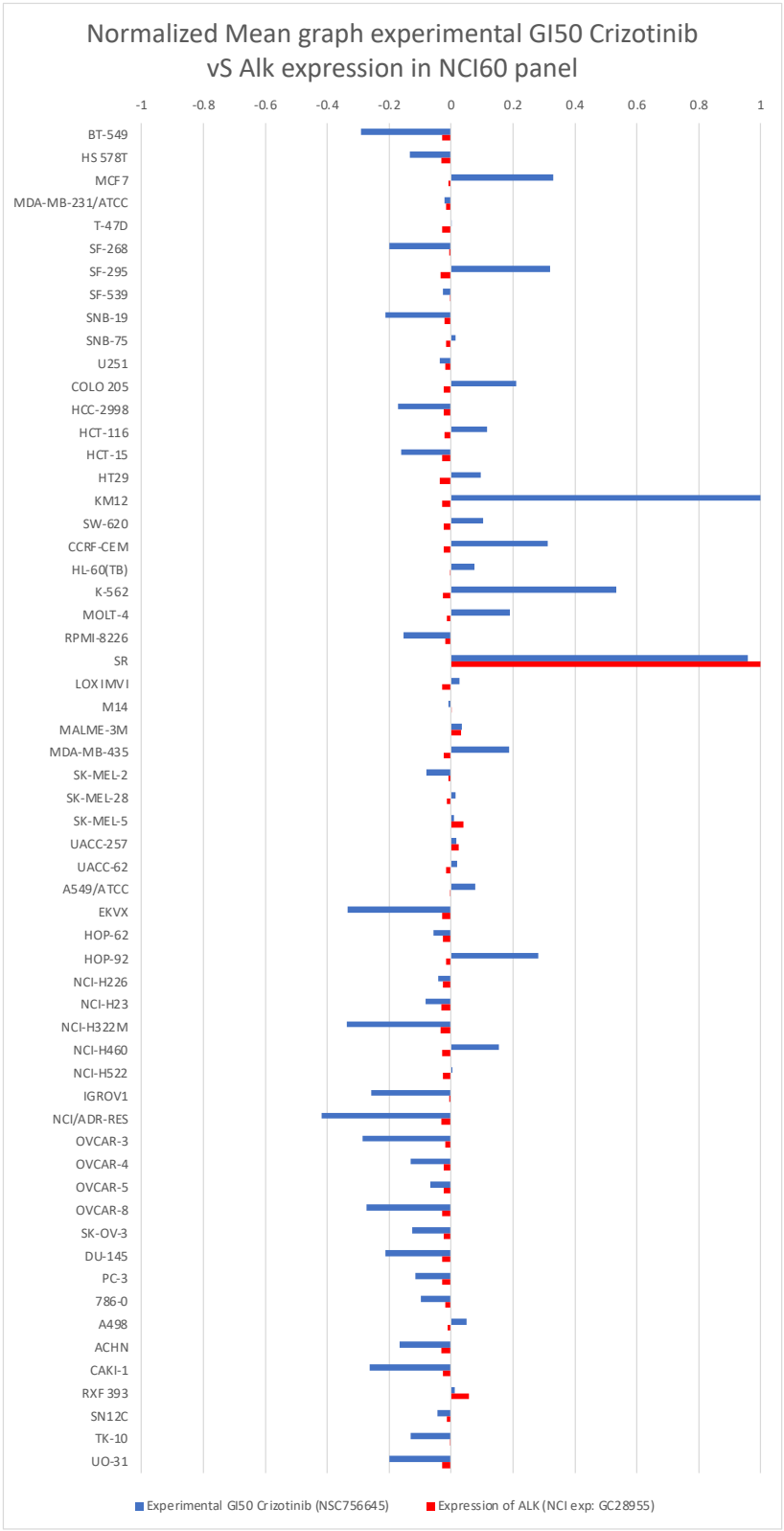
Supplementary Figure 4. *Normalized Mean graph Experimental GI_{50} Imatinib vS Abl Expression in NCI-60 panel.*



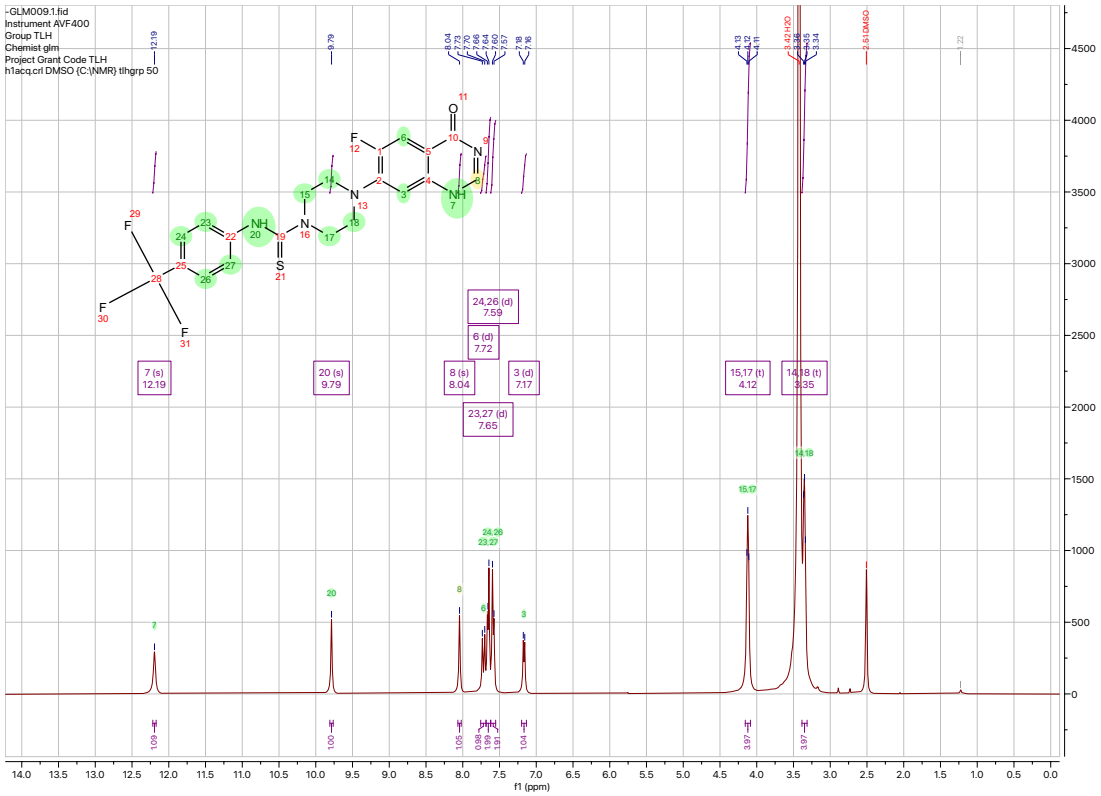
Supplementary Figure 5. *Normalized Mean graph experimental GI_{50} MIRA-1 vS p53 expression in NCI-60 panel.*



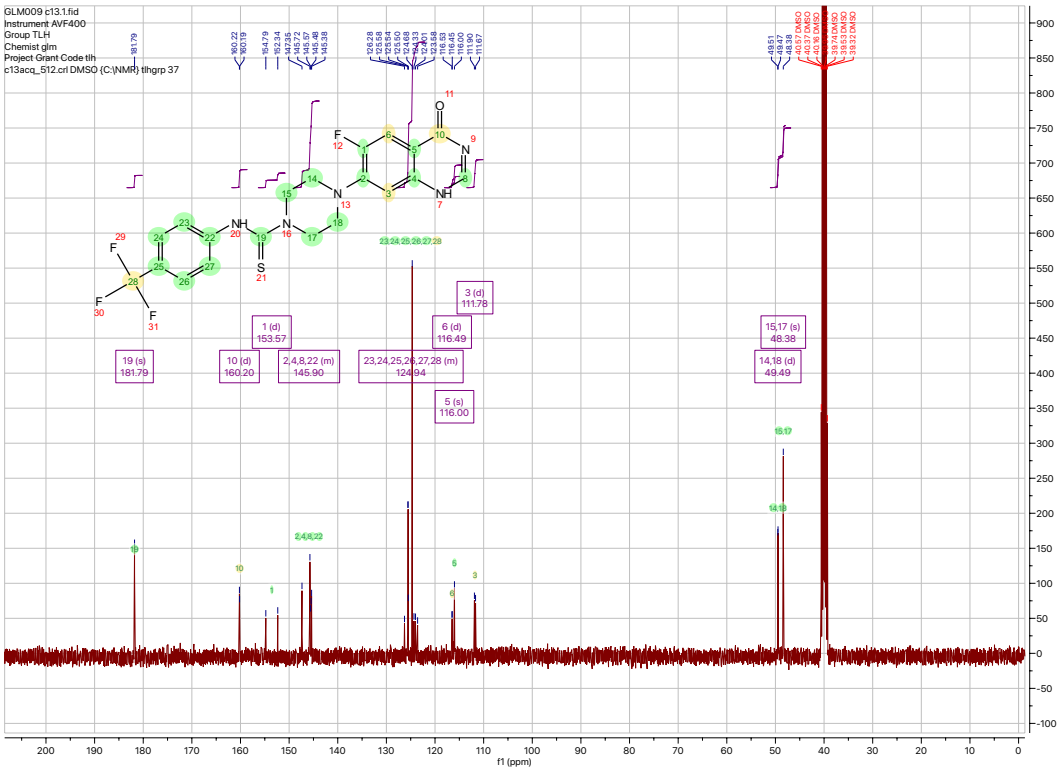
Supplementary Figure 6. *Normalized Mean graph experimental GI_{50} Crizotinib vS Alk expression in NCI-60 panel.*



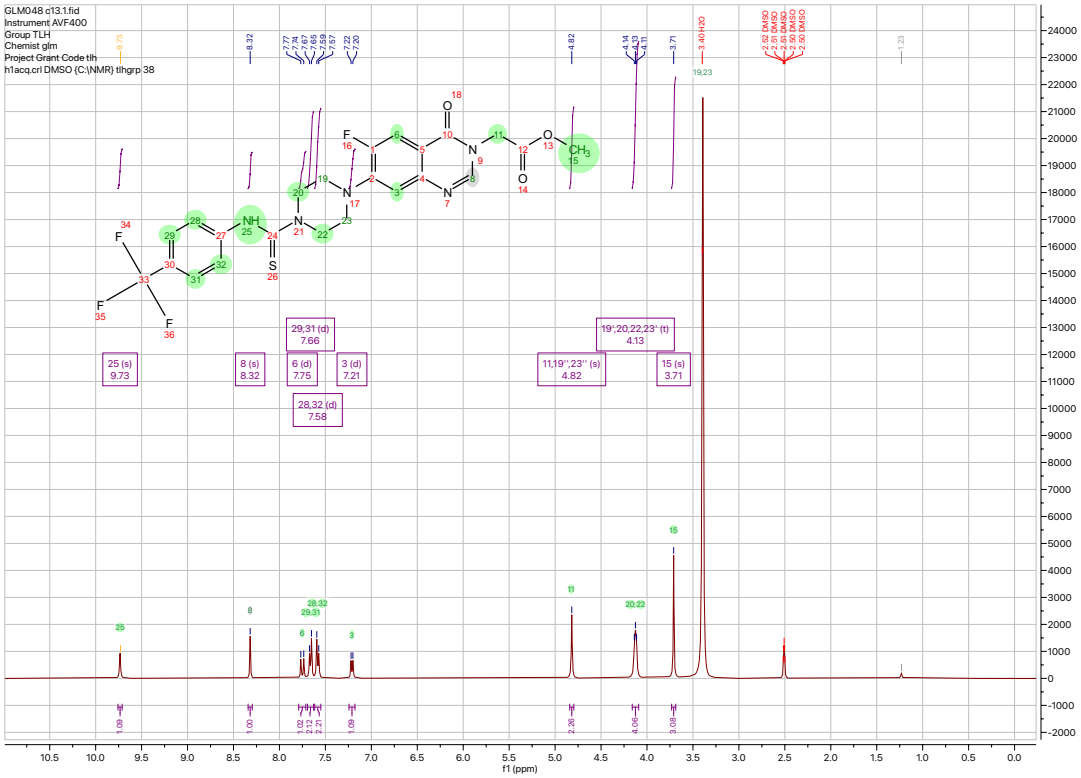
Supplementary Figure 7. *^{1}H NMR of compound **36** (DMSO-d_6).*



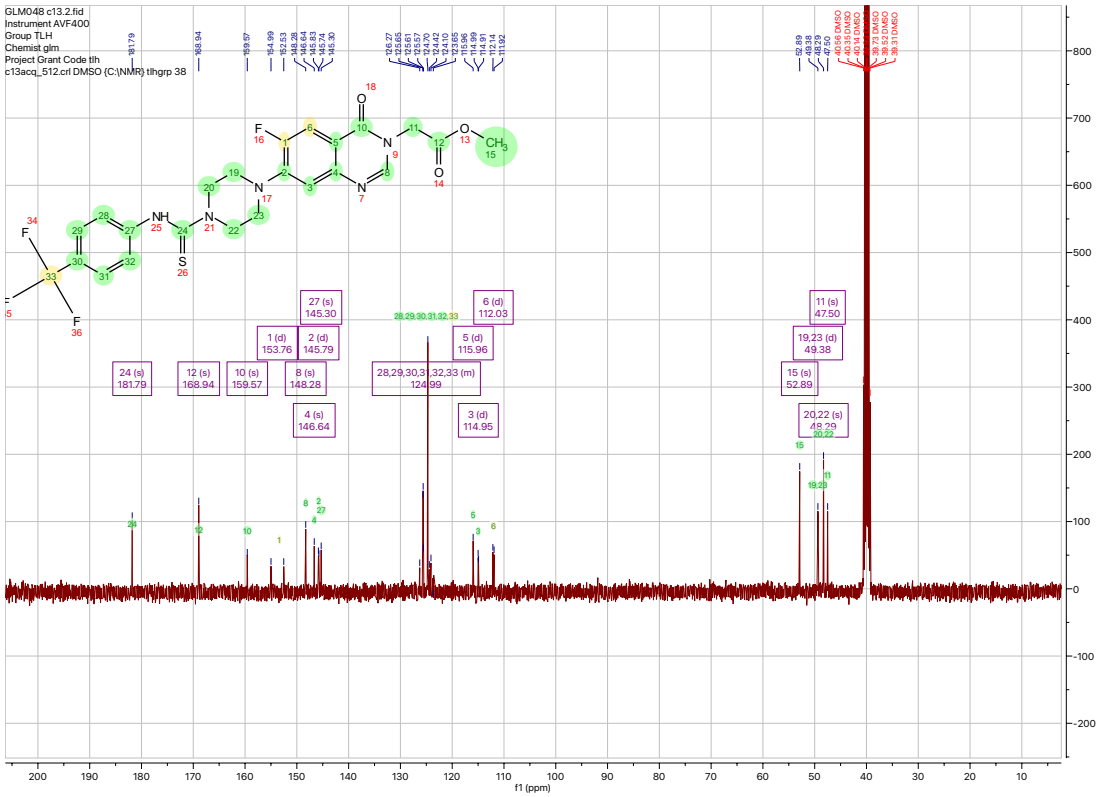
Supplementary Figure 8. *^{13}C NMR of compound **36** (DMSO-d_6).*



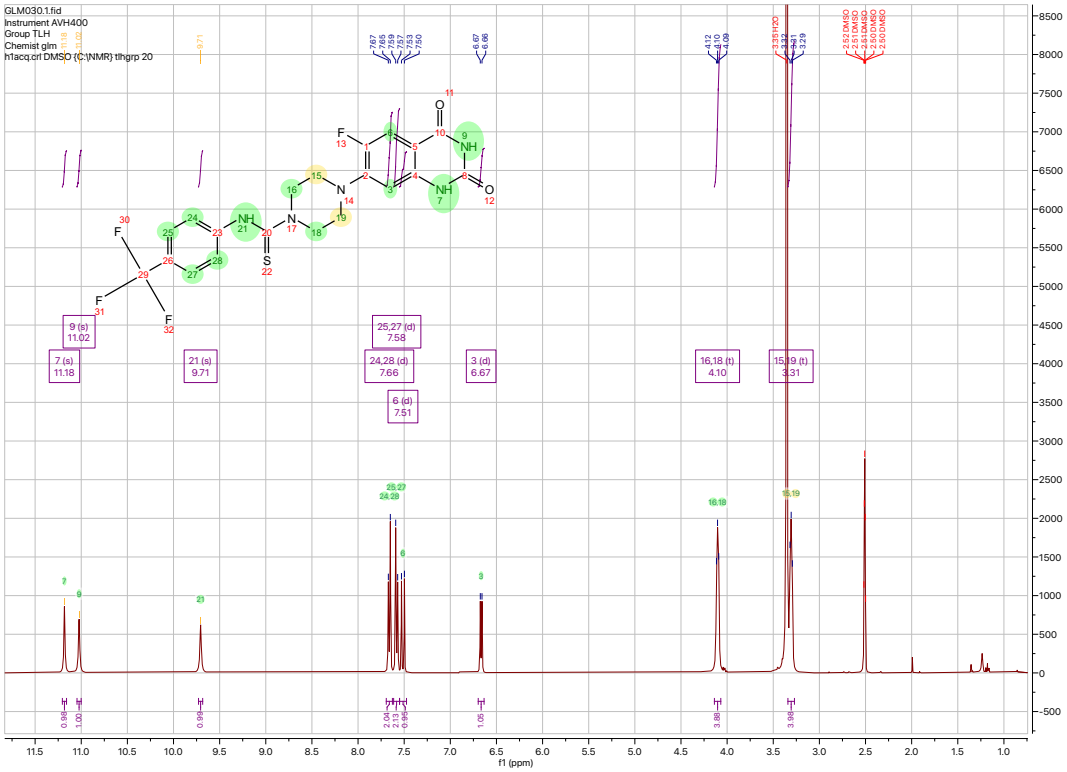
Supplementary Figure 9. *^{1}H NMR of compound **37** (DMSO-d_6).*



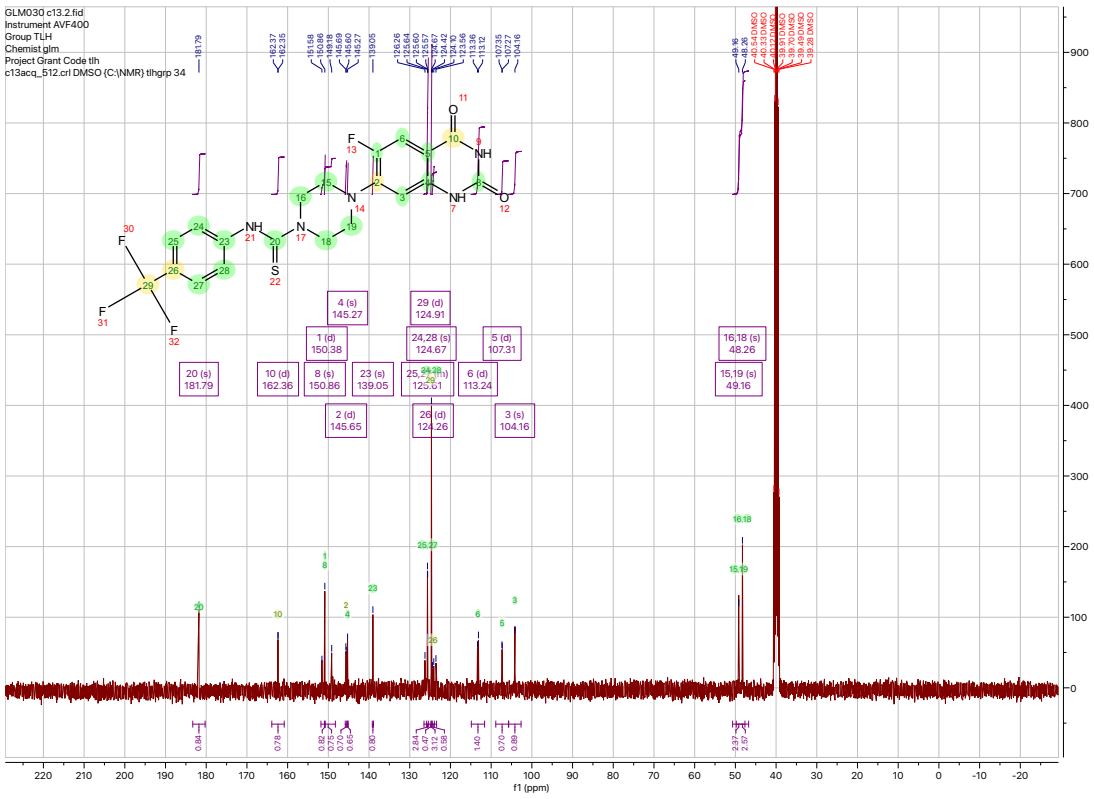
Supplementary Figure 10. *^{13}C NMR of compound **37** (DMSO-d_6).*



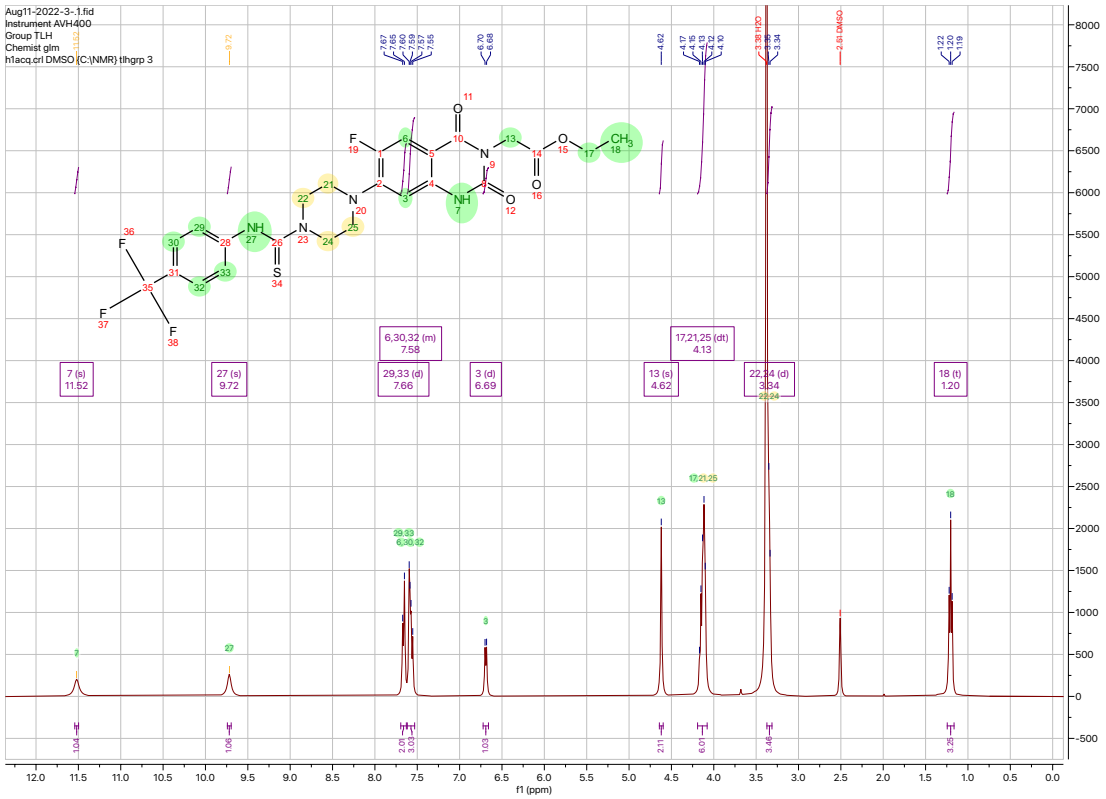
Supplementary Figure 11. *^{1}H NMR of compound **38** (DMSO-d_6).*



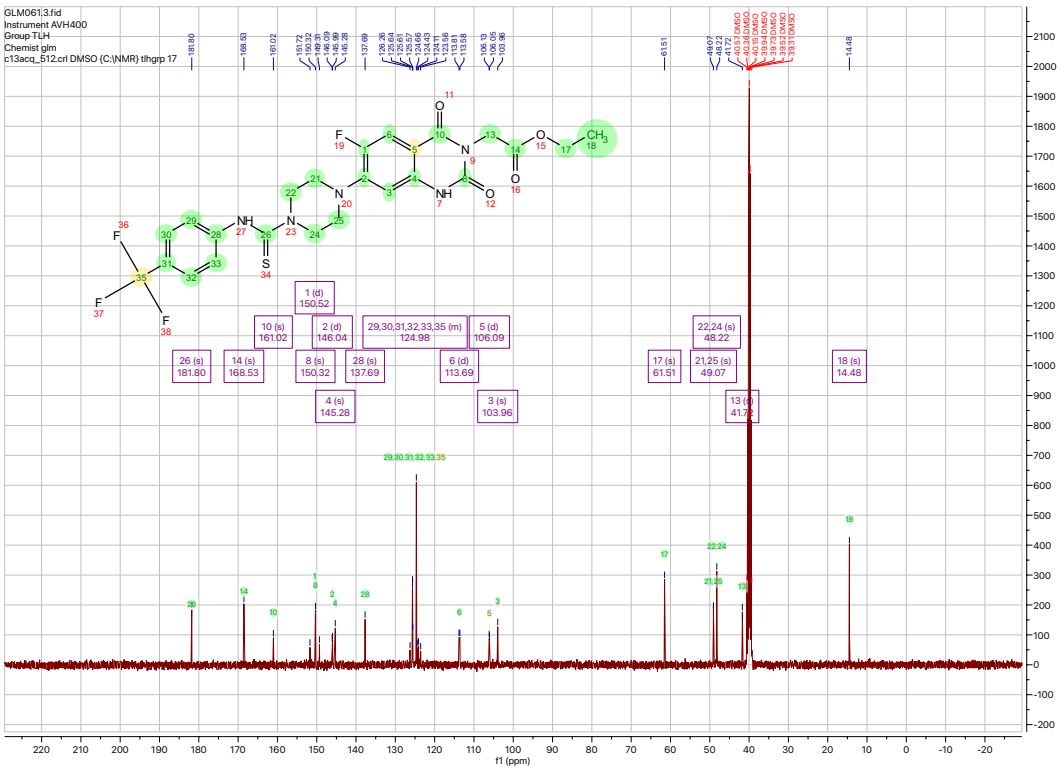
Supplementary Figure 12. *^{13}C NMR of compound **38** (DMSO-d_6).*



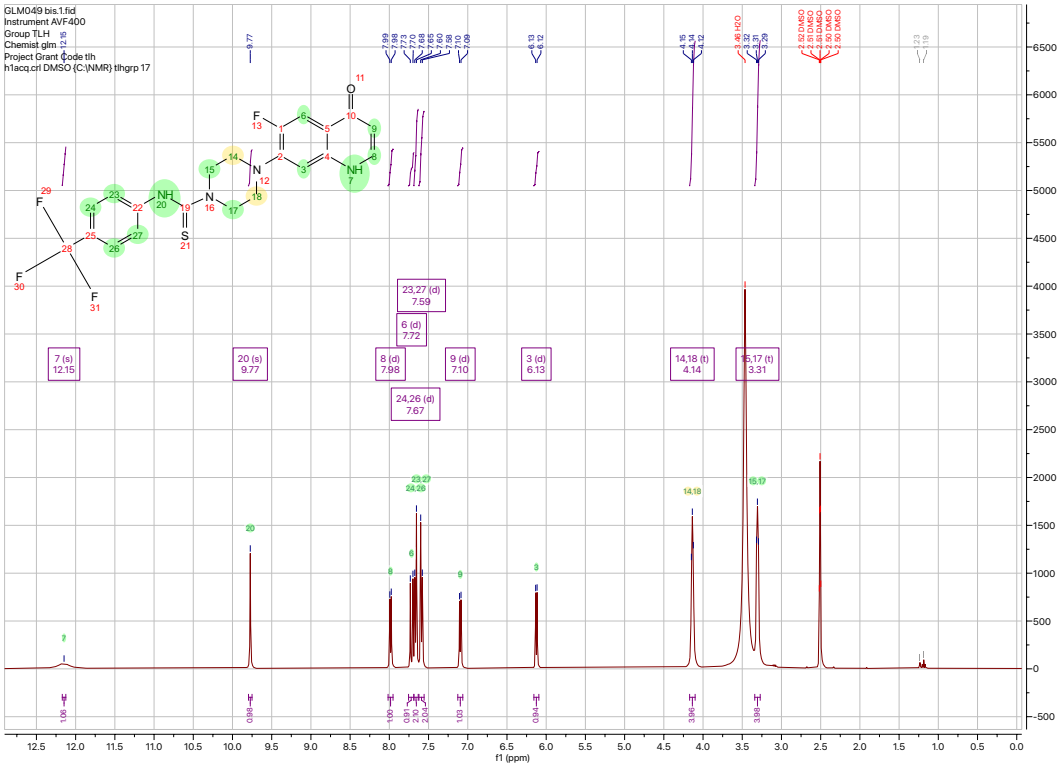
Supplementary Figure 13. *^{1}H NMR of compound **39** (DMSO-d_6).*



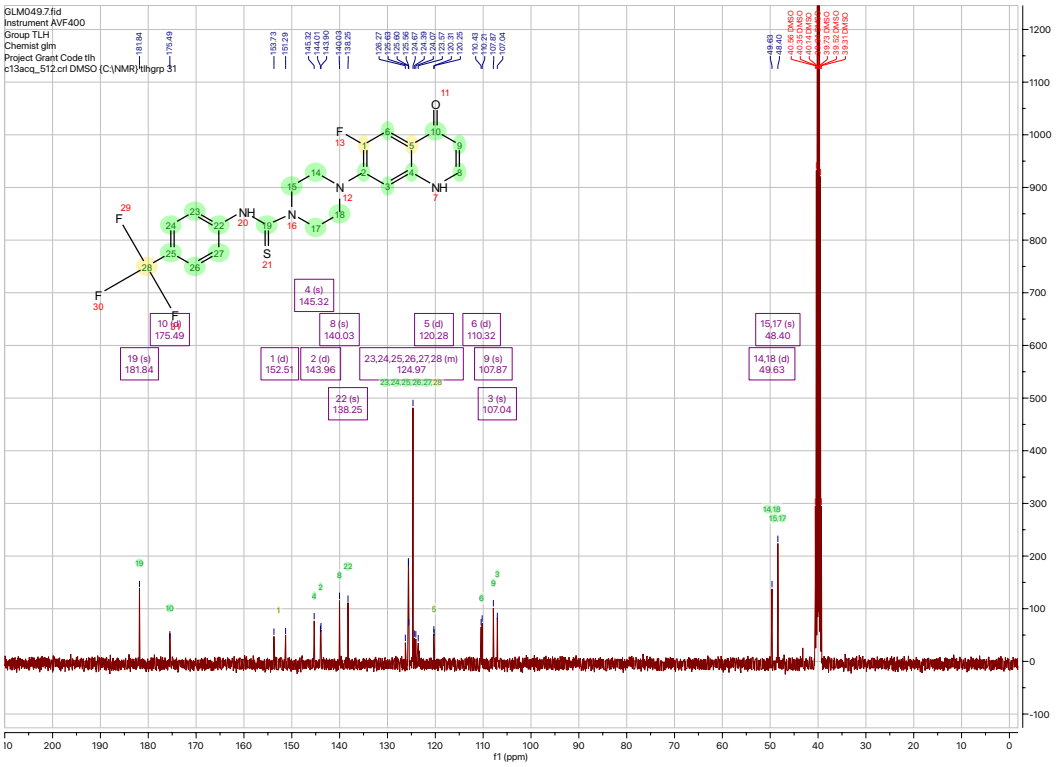
Supplementary Figure 14. *^{13}C NMR of compound **39** (DMSO-d_6).*



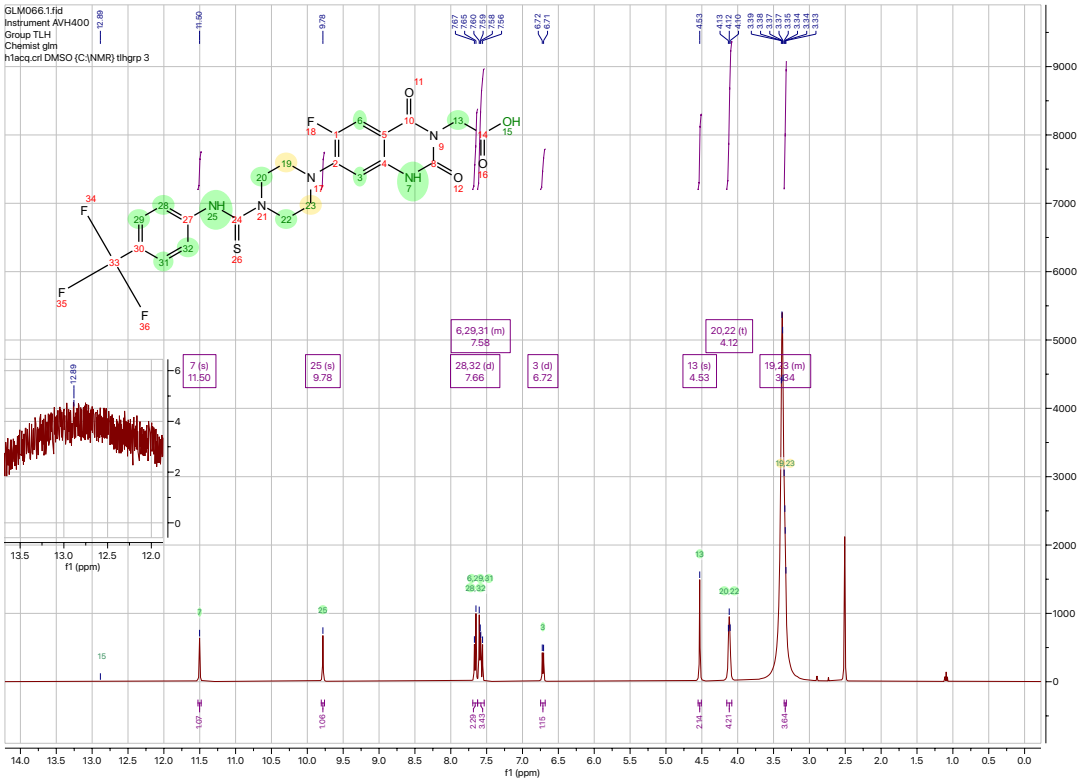
Supplementary Figure 15. *^{1}H NMR of compound **40** (DMSO-d_6).*



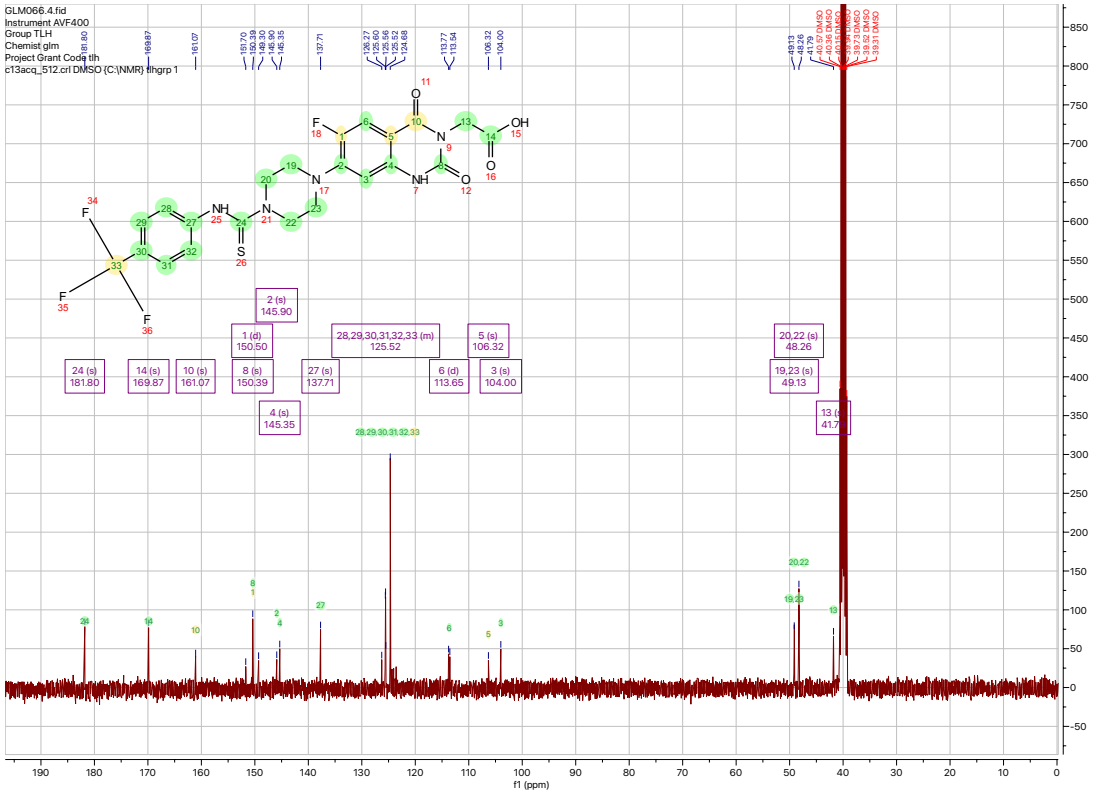
Supplementary Figure 16. *^{13}C NMR of compound **40** (DMSO-d6).*



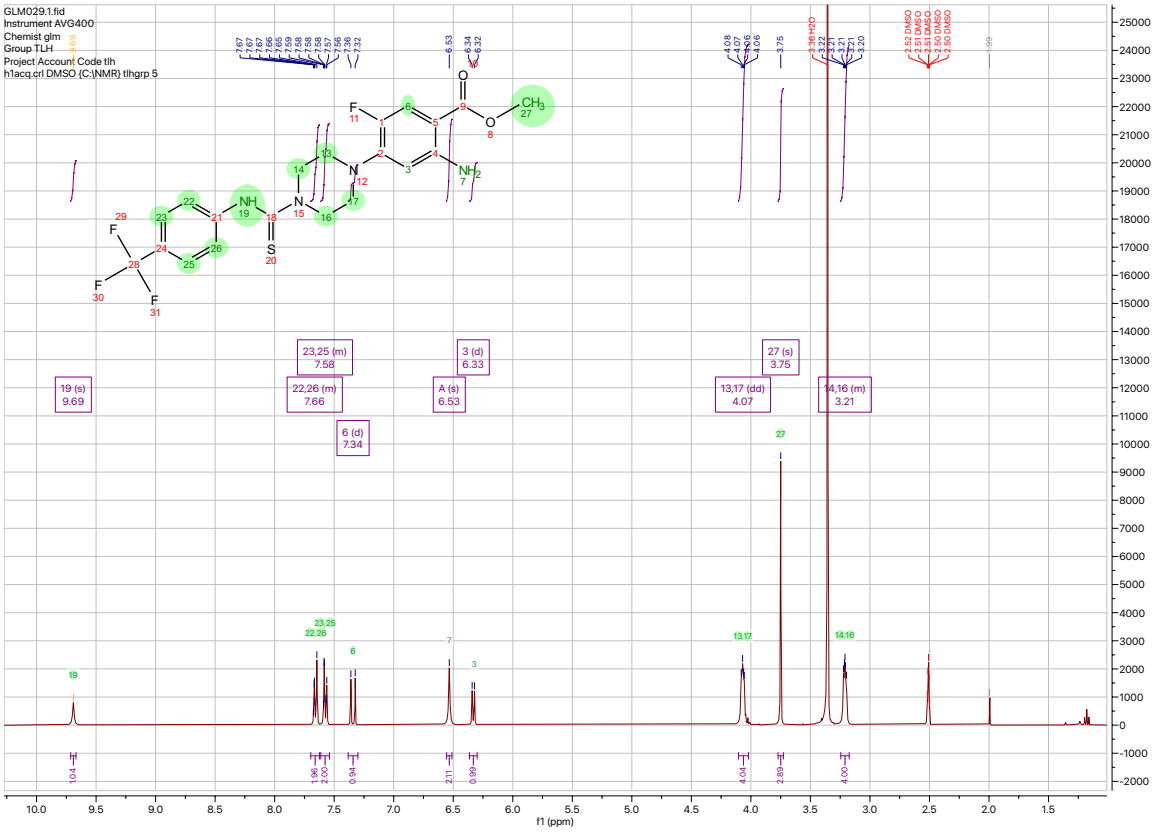
Supplementary Figure 17. *^{1}H NMR of compound **42** (DMSO-d_6).*



Supplementary Figure 18. *^{13}C NMR of compound **42** (DMSO-d_6).*



Supplementary Figure 19. *^{1}H NMR of compound* ***43*** *(DMSO-d_6)*



List of Figures

List of Tables

List of Schemes

List of Supplementary Figures

Acknowledgements

I would like to express my sincere gratitude firstly Prof. Antonino Lauria and Prof. Annamaria Martorana for their supervision, teaching, and mentoring and to my colleague Alessia Bono for her kind and constant support; Prof. Antonio Palumbo Piccionello for providing the in house structure database for the AAP validation and the NCI screenings; Prof. Carla Gentile for *in vitro* antiproliferative and enzymatic inhibition assays (Cdc25); Dr. Daniele Peri and Salvatore Mannino for software development. Thanks are due to the NCI, Bethesda, MD (USA) for performing the antitumor testing of the curcumin-like compounds.

I would like particularly to thank Dr. Thomas Lanyon-Hogg for his infinite hospitality in University of Oxford, his constant support, kindness, and niceness and for being always prompt and available to provide me suggestions. Thanks to Jacob, for his constant support in lab and for running the biological assays for my compounds. Furthermore, a special thanks to all the other members of the "TLH group", Tom H., Vanessa, Om, Vasiliki, Keira, Adam T., for making me feel home in Oxford.

Part of the text and figures were reproduced or adapted from published manuscripts (Lauria, A. et al. Identification of biological targets through the correlation between cell line chemosensitivity and protein expression pattern. *Drug Discovery Today* **2021**, *26*, 2431-2438, doi:10.1016/j.drudis.2021.05.013 for the introduction and the Results, section 3.2; Lauria, A. et al. In Silico Identification of Small Molecules as New Cdc25 Inhibitors through the Correlation between Chemosensitivity and Protein Expression Pattern. *Int. J. Mol. Sci.* **2021**, *22*, doi:10.3390/ijms22073714 for the Results, section 3.3).

This work was in part supported by the PJ_RIC_FFABR_2017_005832 Grant—University of Palermo and by the European Union 2014–2020 PON Ricerca e Innovazione grant from the Italian Ministry of Education, University and Research, entitled "PROGEMA—Processi Green per l'Estrazione di Principi Attivi e la Depurazione di Matrici di Scarto e Non" (ARS01_00432).

Biography and CV of Gabriele La Monica

October 2019: Graduated in Chemistry and Pharmaceutical Technology with full marks and honors at the University of Palermo with a thesis entitled "Studi di modellistica molecolare e sintesi di inibitori selettivi di ZAP-70 nell'approccio terapeutico target mirato" (supervisors Proff. A. Lauria and A. Martorana).

November 2019: He started the Ph.D. degree in Molecular and Biomolecular Sciences (University of Palermo), under the supervision of Prof. A. Lauria and A. Martorana, focusing his research on the design and synthesis of new heterocyclic small molecules endowed with biological activities.

March-August 2022: Visiting student at the Department of Pharmacology, University of Oxford under the guidance of Dr. Thomas Lanyon-Hogg.

List of peer-reviewed publications

- Martorana, A.; La Monica, G.; Lauria, A. Quinoline-Based Molecules Targeting c-Met, EGF, and VEGF Receptors and the Proteins Involved in Related Carcinogenic Pathways. *Molecules* **2020**, *25*, doi:10.3390/molecules25184279.

- Lauria, A.; La Monica, G.; Terenzi, A.; Mannino, G.; Bonsignore, R.; Bono, A.; Almerico, A.M.; Barone, G.; Gentile, C.; Martorana, A. Antiproliferative Properties and G-Quadruplex-Binding of Symmetrical Naphtho[1,2-b:8,7-b']dithiophene Derivatives. *Molecules* **2021**, *26*, doi:10.3390/molecules26144309.

- Lauria, A.; La Monica, G.; Gentile, C.; Mannino, G.; Martorana, A.; Peri, D. Identification of biological targets through the correlation between cell line chemosensitivity and protein expression pattern. *Drug Discovery Today* **2021**, *26*, 2431-2438, doi:10.1016/j.drudis.2021.05.013.*

- Lauria, A.; Martorana, A.; La Monica, G.; Mannino, S.; Mannino, G.; Peri, D.; Gentile, C. In Silico Identification of Small Molecules as New Cdc25 Inhibitors through the Correlation between Chemosensitivity and Protein Expression Pattern. *Int J Mol Sci* **2021**, *22*, doi:10.3390/ijms22073714.*****

- Lauria, A.; La Monica, G.; Bono, A.; Martorana, A. Quinoline anticancer agents active on DNA and DNA-interacting proteins: From classical to emerging therapeutic targets. *Eur J Med Chem* **2021**, *220*, 113555, doi:10.1016/j.ejmech.2021.113555.

- La Monica, G.; Lauria, A.; Bono, A.; Martorana, A. Off-target-based design of selective hiv-1 protease inhibitors. *International Journal of Molecular Sciences* **2021**, *22*, doi:10.3390/ijms22116070.

- Mingoia, F.; Panzeca, G.; Vitale, M.C.; La Monica, G.; Bono, A.; Lauria, A.; Martorana, A. One pot-like regiospecific access to 1-aryl-1*H*-pyrazol-3(2*H*)-one derivatives and evaluation of the anticancer activity. *Arkivoc* **2022**, *2022*, 191-203, doi:10.24820/ark.5550190.p011.739.

- La Monica, G.; Bono, A.; Lauria, A.; Martorana, A. Targeting SARS-CoV-2 Main Protease for Treatment of COVID-19: Covalent Inhibitors Structure-Activity Relationship Insights and Evolution Perspectives. *J Med Chem* **2022**, *65*(19), 12500-12534, doi:10.1021/acs.jmedchem.2c01005.

***** *Scientific production arisen from this thesis*

List of conference proceedings

- La Monica G., Martorana A., Lauria A., STUDI DI MODELLISTICA MOLECOLARE E SINTESI DI INBITORI SELETTIVI DI ZAP-70 NELL'APPROCCIO TERAPEUTICO TARGET MIRATO; Società Chimica Italiana, Workshop della Sezione Sicilia 2020: Poster Communication.

- La Monica G., CORRELATION BETWEEN CHEMOSENSITIVITY AND PROTEIN EXPRESSION TO THE IDENTIFICATION OF NEW CDC25 INHIBITORS; Proceedings of PhD student poster session, pp.132-133; European School of Medicinal Chemistry (ESMEC), (40th Advanced Course of Medicinal Chemistry and 'E. Duranti' National Seminar for PhD Students)- Poster communication;

- Bono A., La Monica G., Martorana A., Lauria A., VALUTAZIONE IN SILICO E SINTESI DI DERIVATI TIENO[3,2-c]CHINOLINICI: POTENZIALI ANTITUMORALI RET-TARGETED NEL CARCINOMA MIDOLLARE TIROIDEO; Società Chimica Italiana, Workshop della Sezione Sicilia 2021 - Catania: Poster Communication (P10);

- Mingoia F., Vitale M.C., La Monica G., Bono A., Lauria A., Martorana A., New Pyrazolo[1,2-a]benzo[1,2,4]triazine-3-one derivatives as potential SARS-CoV-2 protease modulators; XXVII CONGRESSO NAZIONALE DELLA SOCIETÀ CHIMICA ITALIANA FAR PO011.