



Outubro de 2022

Bruna Filipa Ferreira da Silva

***In silico* and *in vitro* assessment of
novel purine derivatives as anticancer
agents**

Universidade do Minho
Escola de Ciências





Universidade do Minho
Escola de Ciências

Bruna Filipa Ferreira da Silva

***In silico* and *in vitro* assessment of
novel purine derivatives as
anticancer agents**

Dissertação de Mestrado
Mestrado em Bioquímica Aplicada

Trabalho realizado sob a orientação de
Doutora Alice Dias
Doutor Manuel Bañobre-Lopéz

Outubro 2022

DIREITOS DE AUTOR E CONDIÇÕES DE UTILIZAÇÃO DO TRABALHO POR TERCEIROS

Este é um trabalho académico que pode ser utilizado por terceiros desde que respeitadas as regras e boas práticas internacionalmente aceites, no que concerne aos direitos de autor e direitos conexos.

Assim, o presente trabalho pode ser utilizado nos termos previstos na licença abaixo indicada.

Caso o utilizador necessite de permissão para poder fazer um uso do trabalho em condições não previstas no licenciamento indicado, deverá contactar o autor, através do RepositóriUM da Universidade do Minho.



Atribuição-NãoComercial-SemDerivações
CC BY-NC-ND

<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Só é permitido que outros façam download dos trabalhos e os compartilhem desde que sejam atribuídos aos autores os devidos créditos, mas sem que possam alterá-los de nenhuma forma ou utilizá-los para fins comerciais.

Acknowledgments

Words cannot express my gratitude to my supervisors, Doctors Alice Dias and Manuel Bañobre-López, for their valuable patience and feedback. I am grateful to have had the opportunity to work with you on this project, an experience I will never forget and will carry proudly throughout my professional career.

I would also like to thank my colleagues from the Nanomedicine and Nanosafety research groups at INL, the Biodiversity team, and the Center of Biological Engineering at the University of Minho, who have also helped me throughout this journey and provided me with the tools I needed for this work.

Lastly, I would be remiss in not mentioning my family, especially my parents. Their belief in me and unconditional love and support have kept my spirits and motivation high during this process. I would also like to thank my boyfriend - my partner in life and lab -, and my close friends for all the good memories and emotional support. You truly made my academic path unforgettable.

My fascination for the biological processes that compose life as we know it began in my childhood years. More recently, and although the true complexity and concept of life remain unknown, it has become my dream to apply this knowledge to help and improve other people's lives. To all of you, thank you for being a part of my dream.

“Nothing truly valuable arises from ambition or from a mere sense of duty; it stems rather from love and devotion towards men and towards objective things.”

– Albert Einstein, *the Human Side* (1979)

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the University of Minho.

Avaliação *in silico* e *in vitro* de novos derivados de purina como agentes anticancerígenos

As purinas desempenham um papel importante na maioria dos processos biológicos. Por esta razão, pesquisas sobre o potencial dos derivados de purina para serem aplicados como agentes terapêuticos levaram ao desenvolvimento de vários fármacos. Assim, a síntese de novos análogos de purina é uma abordagem valiosa para examinar o efeito dos substituintes na atividade biológica.

O cancro é uma das principais causas de morte e uma importante barreira para o aumento da esperança de vida em todos os países do mundo. A acumulação de adenosina extracelular e de outros análogos de purina tem sido reportada em contexto tumoral e tem-se mostrado crucial para o seu desenvolvimento. Além disso, outras purinas biológicas como o ATP, também desempenham um papel mediador nas vias de sinalização intracelular através da fosforilação de uma variedade de substratos que medeiam a divisão celular. Posto isto, novas famílias de análogos de purina foram desenhadas e sintetizadas para aplicações no tratamento do cancro. O mecanismo de ação das purinas já conhecidas como anticancerígenos foi inicialmente relacionado com a sua integração no DNA mas, presentemente, estudos mostram que estas moléculas poderão na verdade comportar-se como agentes polifarmacológicos (agentes que atuam em múltiplos alvos). Embora a polifarmacologia costumasse ter uma conotação negativa, já que ter efeito sobre mais do que um alvo possa levar a efeitos colaterais inesperados, agora a química medicinal tende a desenhar agentes menos seletivos para a terapia. Atualmente, os cientistas têm motivos para acreditar que a combinação da terapia polifarmacológica com sistemas de *drug delivery* possa ser a resposta para o tratamento de doenças complexas como o cancro.

Recentemente, o rápido crescimento de recursos de dados tornou-se um fator determinante para capturar aspetos da eficácia de fármacos anticancerígenos. Neste trabalho, ferramentas computacionais como inteligência artificial e simulações moleculares foram usadas para avaliar a atividade e o mecanismo de ação de novos derivados de purina. O efeito das moléculas na viabilidade celular também foi avaliado em duas linhas de células cancerígenas e uma saudável. Foram também realizados testes enzimáticos e a capacidade de encapsular as moléculas em nanopartículas lipídicas sólidas.

Em termos gerais, os resultados mostraram que estas moléculas recém-desenhadas têm potencial para serem aplicadas não apenas como agentes anticancerígenos, mas também como agentes terapêuticos para outras doenças complexas. Dado o seu perfil polifarmacológico, mais estudos devem ser feitos para avaliar plenamente o potencial oculto destes compostos como agentes terapêuticos.

Palavras chave: cancro, design de fármacos auxiliado computacionalmente, derivados de purina, polifarmacologia

***In silico* and *in vitro* assessment of novel purine derivatives as anticancer agents**

Purines play an important role in most biological processes as components of ATP, cAMP, etc. For this reason, research on purine derivative's potential to be applied as therapeutic agents resulted on the continuous development of multiple drugs still used today. Thus, modification of substituents on the purine ring is a valuable approach to examine the corresponding effect on biological activity.

Cancer is a leading cause of death and an important barrier to increasing life expectancy in every country in the world. Accumulation of extracellular adenosine and other compounds containing purines have been reported in the tumor microenvironment and it has been proven to be crucial to its development. Additionally, these types of compounds also mediate intracellular signaling pathways through phosphorylation of a range of downstream substrates that mediate cell division. Knowing this, novel families of purine analogues have been designed and synthesized for applications in cancer treatment. The mechanism of action of known purines as anticancer agents was first thought to be related to their integration in DNA, but today, studies have shown that these molecules could actually behave as polypharmacological agents. Although polypharmacology has been linked to a negative connotation, as having an effect on more than one target could lead to unexpected side effects, now medicinal chemistry tends to design less selective agents for therapy. At present, scientists have a reason to believe that combining polypharmacological therapy with drug delivery systems might be the answer to treat complex diseases such as cancer.

Recently, the rapid growth of data resources resulted from the application of omics technologies, which have become a very determinant factor to capture different aspects of anticancer drug efficacy. In the present work, computational tools such as artificial intelligence and molecular simulations have been used to assess the activity and mechanism of action of newly designed purine derivatives. The molecules' effect on cell viability was also assessed in two different cancer cell lines and a healthy cell line. In addition to the viability assays, enzymatic tests were performed and ability to encapsulate the molecules in solid lipid nanoparticles was also assessed.

Overall, results have shown that these newly designed molecules have the potential to be applied not only as anticancer agents, but also as therapeutic agents for other complex diseases. Given their polypharmacological profile, further studies should be done to fully assess their lurking potential as therapeutic agents.

Keywords: cancer, computer-aided drug design, polypharmacology, purine derivatives

Contents

Abbreviation list.....	viii
1 Introduction	10
1.1 Cancer: state of the art and new findings.....	11
1.2 Purine derivatives: applications and possible targets according to literature	13
1.2.1 Extracellular signaling pathways as targets	14
1.2.2 Intracellular signaling pathways.....	16
1.2.3 Using purine derivatives to target metabolic pathways and DNA synthesis.....	20
1.3 Computer-aided drug design: using artificial intelligence techniques for activity and multi-target prediction	21
1.4 Purine derivatives as polypharmacological agents: multitarget drug discovery for the future	24
1.5 Drug formulations: solid lipid nanoparticles (SLNs) as drug delivery systems	27
1.6 Previous work	29
2 Work relevance and objectives.....	31
3 Materials and methods.....	34
3.1 Prediction of biological activity spectrum of purine derivatives	35
3.2 Protein-ligand interactions	39
3.3 Test compounds and reagents	40
3.4 Cell lines and culture conditions	40
3.5 2D cell plating.....	41
3.6 Cell viability.....	42
3.7 Cell lysis, protein extraction and quantification	43
3.8 Measurement of ATPase activity (inorganic phosphate (P _i) quantification)	44
3.8 Drug encapsulation using SLNs.....	45
3.9 Statistical analysis.....	46
4 Results and discussion	48
4.1 Prediction of activity and targets	49
4.3 2D <i>in vitro</i> assessments	54
4.4 Insertion in drug delivery systems.....	57
4.5 P _i quantification.....	58
5 Concluding remarks and future perspectives.....	59
6 References.....	65
8 Appendix.....	78

Abbreviation list

AC adenylyl cyclase	HSP90 heat shock protein 90
ADME absorption, distribution, metabolism and excretion	IC₅₀ half maximal inhibitory concentration
ADP adenosine diphosphate	IMP inositol monophosphate
ADSL adenylosuccinate lyase	JNK1 c-Jun N-terminal kinase
AKT protein kinase B	MAPK mitogen-activated protein kinase
ALP alkaline phosphatase	MD molecular docking
AMP adenosine monophosphate	MDM2 murine double minute 2
ATIC 5-aminoimidazole-4-carboxamide ribonucleotide transformylase; inosine monophosphate cyclohydrolase	MEK mitogen-activated extracellular signal-regulated kinase
ATP adenosine triphosphate	MNA multilevel neighbors of atoms
AURK aurora kinase	MR molecular refractivity
CADD computer-aided drug design	NAD nicotinamide adenine dinucleotide
cAMP cyclic adenosine monophosphate	NM II non-muscle myosin II
CDK cyclin-dependent kinase	PAICS phosphoribosylaminoimidazole carboxylase; phosphoribosylaminoimidazolesuccinocarboxamide synthetase
cGMP cyclic guanosine monophosphate	PAINS pan-assay interference compounds
DLS diffusion light scattering	PAP prostatic acid peptidase
DMSO dimethyl sulfoxide	PAPS 3'-Phosphoadenosine-5'-phosphosulfate
DNA deoxyribonucleic acid	PBS phosphate buffer solution
ECM extracellular matrix	PDB protein data bank
EGFR endothelial growth factor receptor	PDGFR platelet-derived growth factor receptor
EPR enhanced permeability retention	PEG polyethylene glycol
FAD flavin adenine nucleotide	PFAS/FGAS phosphoribosylformylglycinamide synthetase
FBS fetal bovine serum	PI3K phosphoinositide 3-kinase
FDA U.S. Food and Drug Administration	PPAT phosphoribosyl pyrophosphate amidotransferase
GART glycinamide ribonucleotide synthetase; glycinamide ribonucleotide transformylase; aminoimidazole ribonucleotide synthetase	PRPP phosphoribosyl pyrophosphate
GFP green fluorescent protein	QSAR quantitative structure-activity relationship
GTP guanosine triphosphate	

R&D research and development

RT room temperature

SAR structure-activity relationship

SLN solid lipid nanoparticle

TME tumor microenvironment

TPSA topological polar surface area

VEGFR vascular endothelial growth factor
receptor

1 Introduction

1.1 Cancer: state of the art and new findings

The first description of cancer appears in the Edwin Smith Papyrus and dates back to 3000 BC. In this copy of a part of an ancient Egyptian textbook, it says that “There is no treatment” for the disease¹. The origin of the word “cancer” is credited to Greek physician Hippocrates (460-370 BC), considered to be the “Father of Medicine”. He used the terms *carcinus* (tumor), *carcinoma* (malignant tumor), and *cancer* (nonhealing malignant ulcer) to describe the abnormal growths that resembled a crab, leading to the terms². About 140 years ago, German microscopist Johannes Mueller showed that cancers were made of cells, a discovery which began the search for changes that would help pinpoint the differences between healthy and cancerous tissues³. Today, we know that cancer is a genetic disorder that results from genetic or epigenetic alterations in the somatic cells⁴. This dysregulated balance of cell proliferation and death can lead to the development of a population of cells that can evade tissues and metastasize to other sites. If left untreated, it will potentially lead to significant morbidity and ultimately death of the patient⁵ (**Figure 1**). In the intervening period, a huge amount of information has been acquired about cancer cells, their mechanisms of survival and development. In the past decades in particular, rapid technological progress has allowed scientists to dissect the cancer genome, transcriptome, proteome and interactome, and a variety of new techniques and models to assess anticancer activity have been developed. Now that almost anything seems technically possible to observe and perform, the key issue for the twenty-first century scientists is to identify the right questions to ask⁶.

Cancer is considered to be a leading cause of death and an important barrier to increasing life expectancy in every country in the world. According to estimates from GLOBOCAN, 19.3 million new cancer cases and almost 10.0 million cancer deaths occurred in 2020⁷. Female breast cancer has surpassed lung cancer as the most commonly diagnosed cancer, with an estimated 2.3 million new cases (11.7%), followed by lung (11.4%), colorectal (10.0%), prostate (7.3%), and stomach (5.6%) cancers. Lung cancer remained the leading cause of cancer death, with an estimated 1.8 million deaths (18%), followed by colorectal (9.4%), liver (8.3%), stomach (7.7%), and female breast (6.9%) cancers⁷. It is important to note that these estimates do not reflect the impact of severe acute respiratory syndrome coronavirus 2⁸, as they are based on extrapolations of cancer data collected in earlier years before the pandemic. Although the full extent of the impact of the COVID-19 pandemic in different world regions is currently unknown, delays in diagnosis and treatment associated with the concerns of individuals, health system closures, including suspension of screening programs, and reduced availability of and access to care are

expected to cause a decline in cancer incidence followed by increases in advanced-stage diagnoses and cancer mortality⁹⁻¹¹.

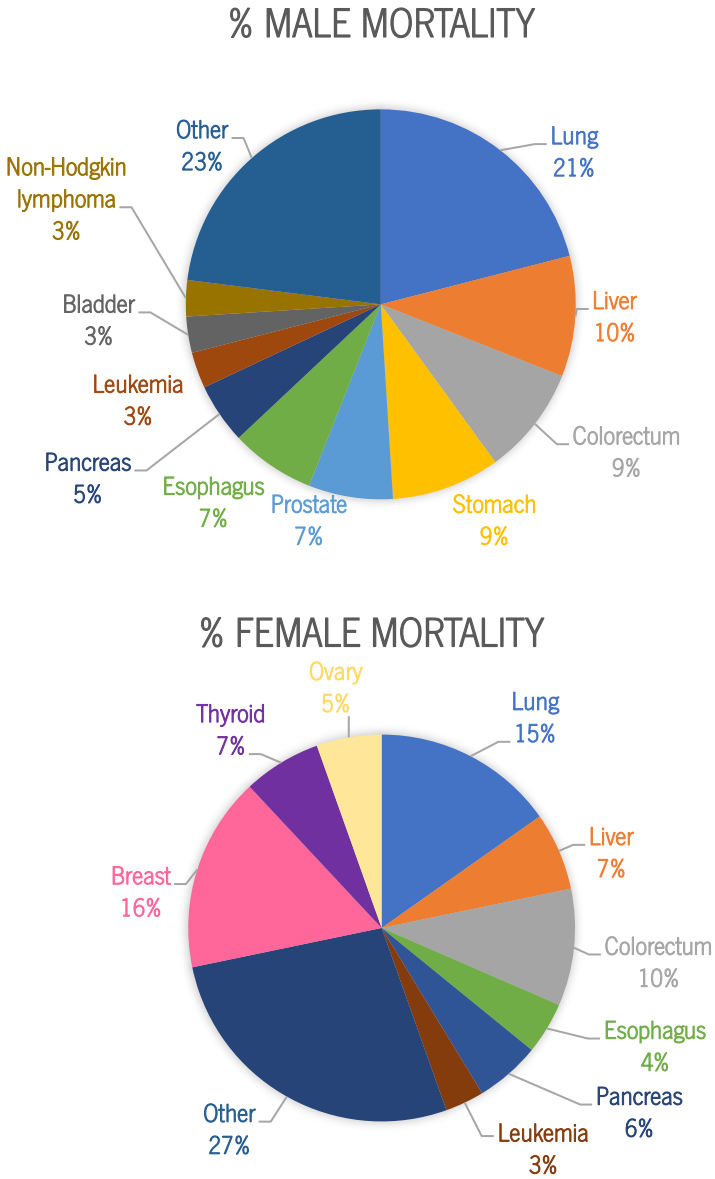


Figure 1: Distribution of cancer deaths for the top 10 most common cancers in 2020 for men and women, according to GLOBOCAN.

Standard treatments for cancer management involve procedures such as surgery, radiotherapy and chemotherapy. One of the main reasons why cancer is so hard to treat is the limitations associated with the use of conventional chemotherapeutic agents. For instance, many anticancer drugs provoke debilitating temporary or permanent side effects for the patient¹² and lowering the dose of these drugs to minimize these side effects may significantly reduce their effectiveness¹³. In addition, many cancer cells usually acquire chemoresistance (lack of response to drug-induced growth inhibition), which constitutes

a persistent problem during treatment¹². Even with the significant progress made in immunotherapy in recent years, only a minority of cancer patients benefit from it and many of them relapse during therapy¹⁴. This, along with the unavailability of effective tools for early cancer diagnosis⁴, makes the cure for cancer the Holy Grail of biomedicine.

Recently, the rapid growth of data resources resulted from the application of omics technologies has become a very determinant factor to capture different aspects of anticancer drug efficacy. Computational methods are essential for the analysis of these big data resources to predict drug activity and to generate clinically useful results¹⁵. Significant advances in machine learning and molecular modeling techniques have afforded a variety of tools to effectively identify drug targets and build structure–activity relationship models, to map key interactions between ligands and their binding sites, and to predict absorption/distribution/metabolism/excretion (ADME) properties of candidate compounds leading to the creation of a drug candidate. The properties of novel therapeutic agents can be investigated and ultimately optimized via expert science behind medicinal chemistry and methods of computer-aided drug design (CADD)¹⁶.

1.2 Purine derivatives: applications and possible targets according to literature

Purines play an important role in most biological processes as components of ATP, GTP, cAMP, cGMP, NAD, FAD, PAPS, etc^{17,18}. Thus, modification of substituents on the purine ring is a valuable approach to examine the corresponding effect on biological activity. The chemistry of purines allows for the manipulation of substituents in a variety of ways and, as a consequence, extensive and diverse libraries of purine derivatives have been prepared bearing different types and combinations of functionality¹⁹. Recent studies have shown advances in exploration of purine-based compounds as biological tools and their potential as therapeutic agents to treat an impressively wide range of diseases such as cancer, rheumatoid arthritis, asthma, diabetes, Parkinson's disease and depression¹⁹ (**Figure 2**).

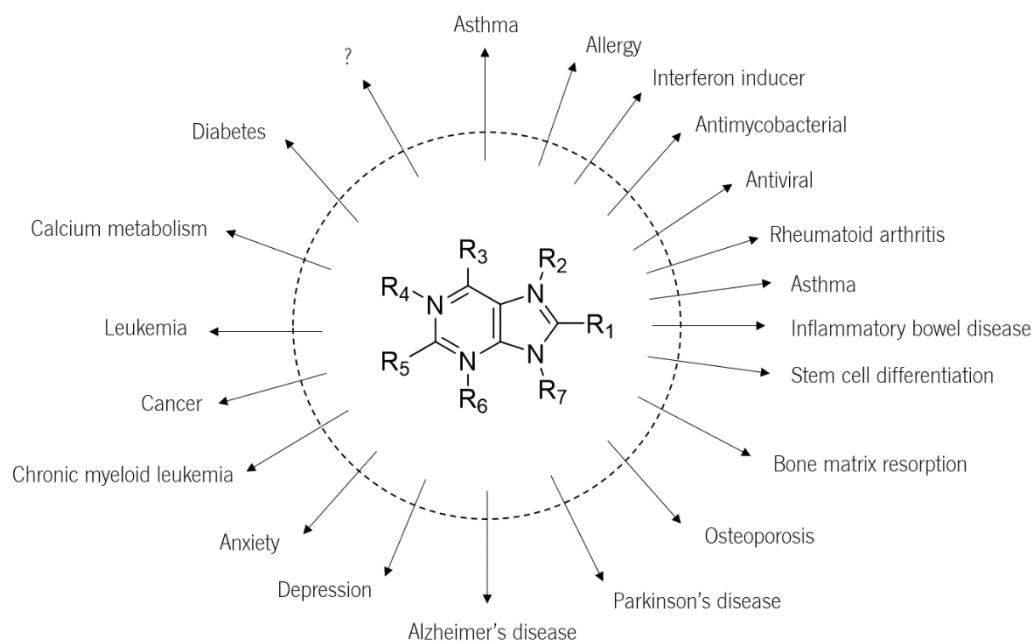


Figure 2: General structure of purine derivatives and potential applications. Combinations of various substituents (1, 2, 3, 4...) on the different positions (R1–R7) of the purine ring lead interactions with different targets¹⁹.

1.2.1 Extracellular signaling pathways as targets

Accumulation of extracellular ATP often reflects metabolic changes under conditions of cellular stress. The enzymes CD39 and CD73 catalyze hydrolysis of extracellular ATP to adenosine. CD39 is a two-transmembrane domain-embedded ecto-enzyme that hydrolyzes ATP into AMP in a two-step process, whereas CD73 is a GPI-anchored ectoenzyme highly selective for AMP that hydrolyzes it into adenosine²⁰. Tissue hypoxia, which is elevated in tumor microenvironments, is an important driver of extracellular adenosine accumulation as it promotes transcription regulation of CD73²¹. Elevated levels of adenosine generally reflect high levels of CD73 expression on cancer cells, fibroblasts and/or immune cells. CD73 constitutes a promising therapeutic target to re-establish anti-tumor immunity dampened by adenosine receptors²². Although the CD39/CD73 axis is one of the most highly studied pathways and is thought to account for the bulk of adenosine production, alternative pathways are also present. CD38 and CD203a are able to sequentially convert NAD⁺ into AMP, which can again be converted into adenosine via CD73. Additionally, adenosine can also be generated through two additional pathways: alkaline phosphatase (ALP), which can directly convert AMP, ATP, or ADP into adenosine; and prostatic acid peptidase (PAP), which converts AMP into adenosine²³ (**Figure 4**).

The multifaceted nature of adenosinergic signaling provides multiple potential targets that have been shown to alleviate the immunosuppressive TME in a variety of preclinical models²⁴. Extracellular adenosine exerts its biological activity through four G-protein-coupled receptors: A1, A2a, A2b and A3²⁵, resulting in different effects on the intracellular signaling pathways (**Figure 4**).

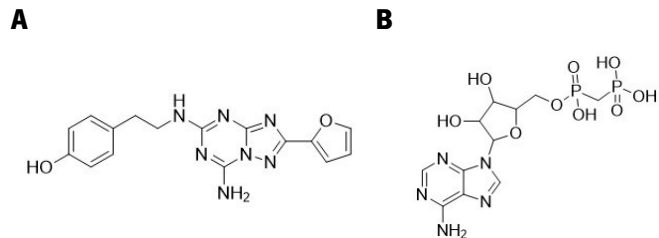


Figure 3: (A) A2a inhibitor **ZM241385**²⁶ and (B) CD73 inhibitor **adenosine 5'-(α,β -methylene)diphosphate**²⁷

In addition to adenosine, other signaling molecules such as growth factors are also present extracellularly. These growth factors activate receptor tyrosine kinases (RTKs), leading to the regulation of cell division and proliferation^{28,29} (**Figure 4**).

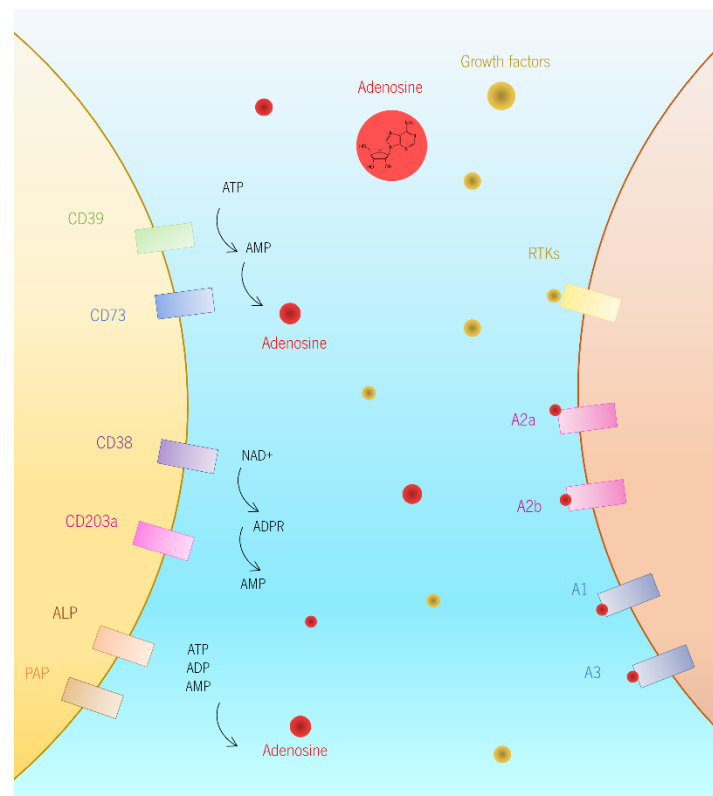


Figure 4: Extracellular signaling between cells by adenosine and growth factors. Extracellular adenosine can be formed directly due to the action of CD39, CD73, PLP and PAP via ATP dephosphorylation. This adenosine serves as a ligand and interact with adenosine receptors (A2a, A2b, A1 and A3). Additionally, growth factors liberated by neighboring cells serve as ligands to RTKs, leading to cell proliferation and division.

1.2.2 Intracellular signaling pathways

Proliferation, differentiation, and survival are all tightly regulated by the RAS/PI3K/AKT and RAS/PKA/ERK/MAPK pathways. These pathways are induced as a result of ligand-activated RTKs, which are a family of cell surface receptors that serve as receptors for external signaling molecules such as growth factors, hormones, cytokines, neurotrophic factors, and others²⁸. The RTK family comprises various classes of receptors, such as endothelial growth factor receptor (EGFR), vascular endothelial growth factor receptors (VEGFR), platelet-derived growth factor receptors (PDGFR), among others²⁹. These pathways consist of an intracellular signal transduction mechanism that regulate metabolism, proliferation, cell survival, growth and angiogenesis in response to extracellular signals. This is mediated through serine and/or threonine phosphorylation of a range of downstream substrates³⁰ (**Figure 8**). As single agents, most inhibitors are cytostatic rather than cytotoxic to cancer cells³¹. However, using strategies for multi-targeted kinase inhibition may be more effective than inhibiting either target alone³². Preclinical experiments support this idea, and drugs such as **PP121** (**Figure 5F**) that target multiple steps in this pathway have been designed³³. When adenylyl cyclase (AC) activity is induced by RTK or AR signaling, the enzyme transforms ATP into cAMP, and allows for the transfer of a phosphate group to next substrate, resulting in a cascade signaling effect. One of these substrates, EPAC, plays a crucial role in cell division. Recently, studies have found that EPAC-specific inhibitor treatment or silencing the EPAC gene expression rendered cells resistant to viral infection³⁴.

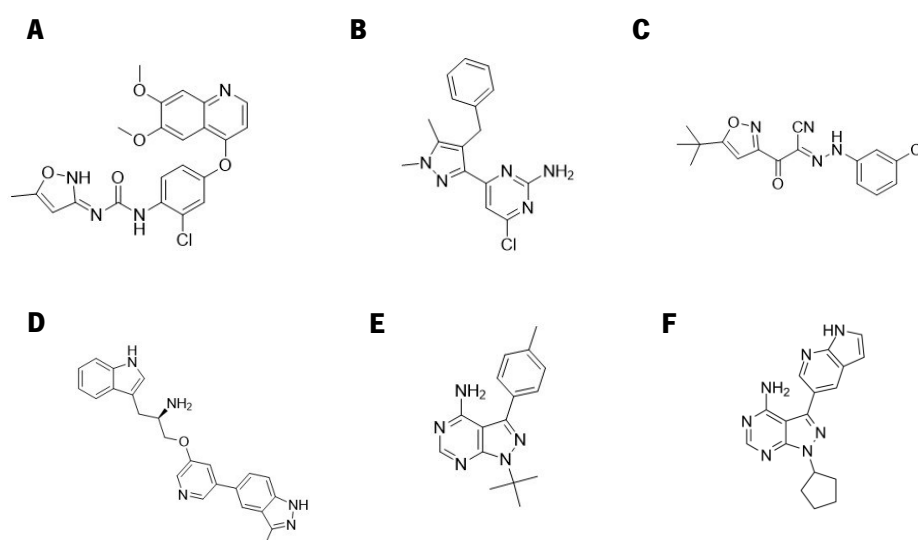


Figure 5: (A) VEGFR inhibitor **tivozanib**³⁵, (B) AC inhibitor **TDI10229**³⁶, (C) EPAC inhibitor **ESI-09**^{37,38}, (D) PKA inhibitor **(2S)-1-(1H-indol-3-yl)-3-[5-(3-methyl-1H-indazol-5-yl)pyridin-3-yl]oxy-propan-2-amine**³⁹, (E) HCK protein tyrosin kinase **1-ter-butyl-3-p-tolyl-1H-pyrazolo[3,4-d]pyrimidin-4-ylamine**⁴⁰ and (F) **PP121**⁴¹.

The p53 proteins (originally thought to be, and often spoken of as, a single protein) are crucial in vertebrates, where they prevent cancer formation⁴¹. As such, p53 is classified as a tumor suppressor gene and has been described as "the guardian of the genome"⁴²⁻⁴⁶. AKT mediates control of P53 levels through enhancing MDM2 (murine double minute 2)-mediated targeting of p53 for degradation⁴⁷. Multiple phosphorylation events trigger the tetrameric transcription factor p53's activation in response to cellular stress. This activation causes overexpression or repression of genes involved in cell-cycle arrest, DNA repair, apoptosis, and senescence depending on the type of stress⁽⁴⁸⁾. P53 induces the expression of p21, an inhibitor of the cyclin dependent kinases (CDKs)2, 3, 4 and 6 (**Figure 8**) (49).

CDKs are the catalytic subunits of a family of mammalian heterodimeric serine/threonine kinases that are implicated in the control of cell-cycle progression, transcription and neuronal function^{48,49}. The dysregulation of cell cycle control in cancer has been studied for many years and, consequently, the central role played by CDKs has led them to be investigated as drug targets^{50,51}. Many malignancies overexpress CDKs, ultimately leading to the development of CDK inhibitors.

The 2,6,9-trisubstituted purines, such as **roscovitine (Figure 6A)**, were among the first low molecular weight inhibitors of CDKs. However, its short half-life and rapid metabolism to inactive derivatives, its rather weak potency on CDKs and, consequently, the large quantities required to treat patients, constitutes a limiting factor in its clinical use. Therefore, analogues of roscovitine are quite desirable. **CR8 (Figure 6B)**, an optimized compound, induces apoptotic cell death with about 50-fold enhanced potency compared to **roscovitine**^{52,53}. **Olomoucine (Figure 6C)** has only weak cytotoxic activity in comparison to **olomoucine II (Figure 6D)**, which is more potent *in vitro* against tumor cells⁵⁴. In previous studies, structure-based design was used to optimize the ATP-competitive inhibition of CDK1 and CDK2 by O6-cyclohexylmethylguanines. The resulting optimized compound, **NU6102 (Figure 6E)**, demonstrated inhibition of cell growth and target protein phosphorylation, consistent with CDK1 and CDK2 inhibition³⁹. This compound is selective and one of the most active CDK2 inhibitors described so far.

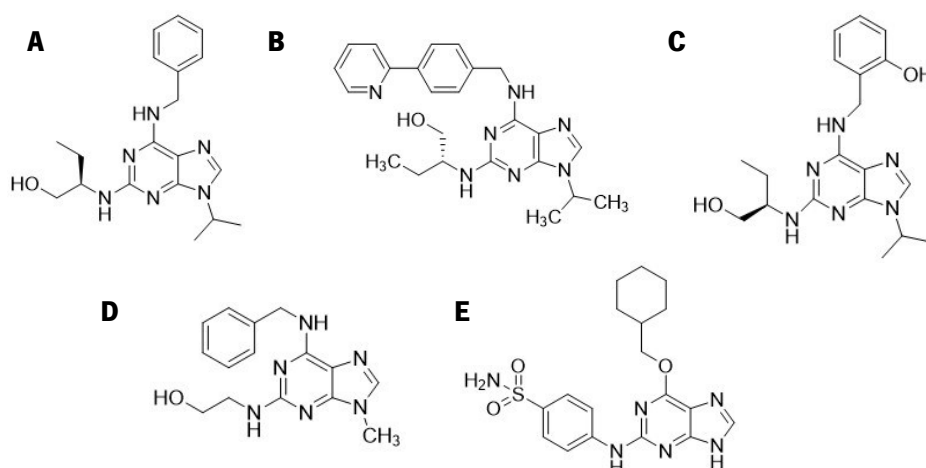


Figure 6: Examples of purine derivatives as CDK inhibitors. (A) **roscovitine**, (B) **CR8**, (C) **olomoucine**, (D) **olomoucine II** and (E) **NU6102**.

Serine/threonine kinase mammalian Aurores (Aurora-A, -B, and -C) are the most prevalent mitotic progression regulators and are often abundantly expressed in human cancers. Recent evidence suggests that the Aurora A and B (the two major types of Aurora kinase), tend to play a key role in the regulation of chromatid divergence^{55,56} (**Figure 8**). More specifically, it allows the splitting cell to provide its daughter cell with its genetic material^{55,57-59}. Inadequacies genetic instability, strongly linked to tumorigenesis can occur in this segregation⁶⁰. A temporary spindle checkpoint-dependent mitotic arrest results from abnormalities in mitotic spindle assembly caused by Aurora-A inhibition. As a result, Aurora-A inhibited cells exit from mitosis and undergo apoptosis either by inducing a G1 arrest followed by apoptosis or by a p53-independent mechanism. Contrarily, blocking Aurora-B disrupts the normal alignment of the chromosomes during mitosis and overcomes the mitotic spindle checkpoint, leading to polyploidy, the failure of cytokinesis, and endoreduplication, which is followed by cell death after more than 48 hours^{61,62}.

Thus, the development of Aurora oncogenic inhibitors may boost cancer patients' clinical outcomes. In previous studies^{63,64}, it has been shown that purine derivatives could potentially be used as Aurora kinase (AURK) inhibitors (**Figure 7**).

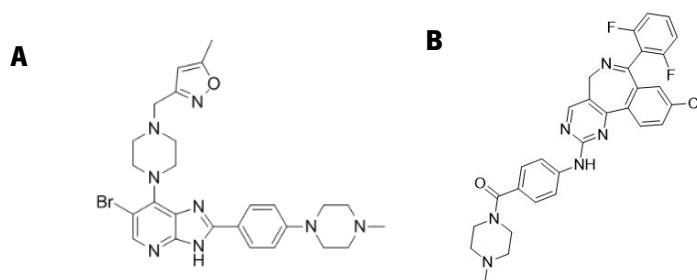


Figure 7: Aurora kinase inhibitors (A) **CCT137690**^{62,64} and (B) **9-chloro-7-(2,6-difluorophenyl)-N-4-[(4-methylpiperazin-1-yl)carbonyl]phenyl]-5H-pyrimido[5,4-d][2]benzazepin-2-amine**⁶⁵.

Although targeted therapies often elicit profound initial patient responses, these effects are transient due to residual disease leading to acquired resistance. How tumors transition between drug responsiveness, tolerance and resistance, remains unclear. In EGFR-mutant lung adenocarcinoma cells, residual disease and acquired resistance in response to EGFR inhibitors requires AURKA activity. Aurora kinase inhibitors suppress this adaptive survival program, increasing the magnitude and duration of EGFR inhibitor response in preclinical models⁶⁶.

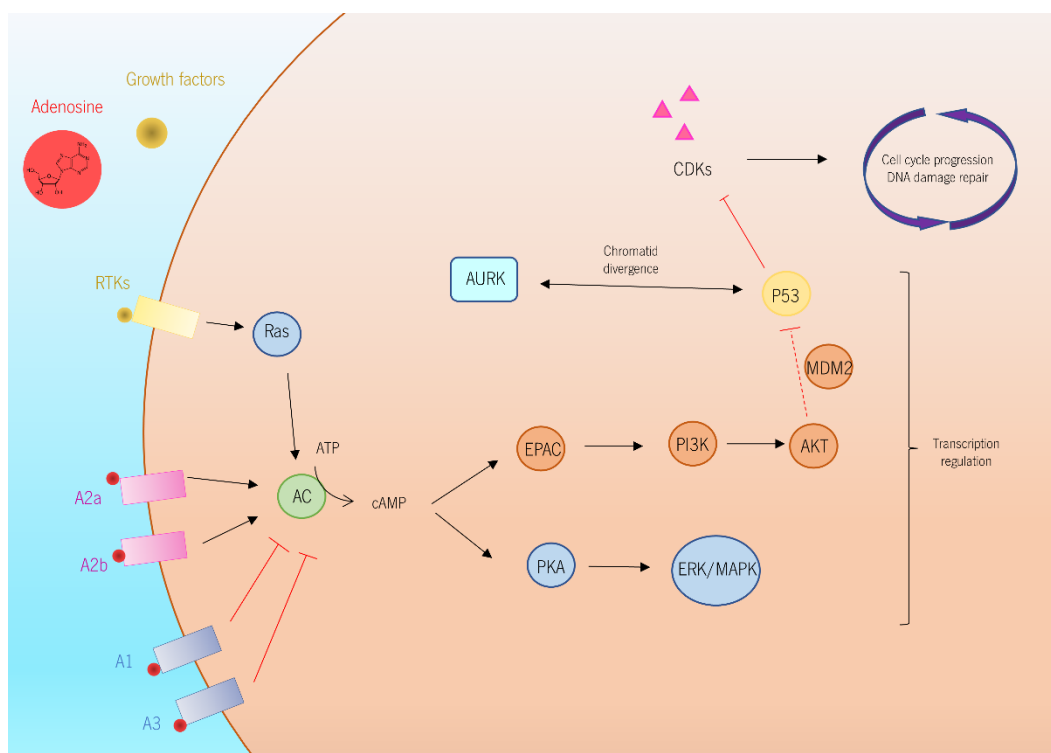


Figure 8: As a result of activation through A2a and A2b receptors, adenylate cyclase leads to dephosphorylation and conversion of ATP to cAMP. Then, this second messenger engages pathways that control transcription, which are essential for cell survival. Ras can be activated by growth factor signaling via RTKs to further increase adenyl cyclase activity, despite the mechanisms underlying the RTK-mediated enhancement of AC function remaining unclear⁽⁸⁶⁾. These intracellular signaling pathways can influence how p53 functions, a focal protein that inhibits CDK activity. In reaction to both internal and external influences, CDKs are responsible for regulating and ensuring proper cell cycle progression.

1.2.3 Using purine derivatives to target metabolic pathways and DNA synthesis

Reprogrammed metabolic pathways are essential for cancer cell survival and growth. These reprogrammed activities allow tumor cells to take up nutrients and use them to produce ATP, generate biosynthetic precursors and macromolecules, and tolerate stresses associated with malignancy (for example, redox stress and hypoxia)^{67,68}.

In cancer cells, high concentrations of purine metabolites have been indicated, and this discovery led to the development of the earliest antitumor drugs – purine antimetabolites⁶⁹. These molecules are chemical analogues that share similar structures to the metabolites within the purine metabolism. This allows them to compete with purine nucleotides to be incorporated into DNA during cell cycle, inhibiting rapid division and proliferation^{70,71}. Currently, many antimetabolites have been approved for cancer treatment, such as **6-thioguanine**^{72,73} (**Figure 9A**) and **6-mercaptopurine**⁷³⁻⁷⁵ (**Figure 9B**).



Figure 9: Purine derivatives as antimetabolites: (A) **6-thioguanine** and (B) **6-mercaptopurine**.

Under cellular conditions of high purine demand, the *de novo* purine biosynthetic enzymes cluster near mitochondria and microtubules to form dynamic multienzyme complexes called purinosomes⁷⁶ (**Figure 10**). These purinosomes have been identified within purine metabolism, and since they are closely related to the cell cycle^{77,78}, these results provide a novel therapeutic strategy for cancers by targeting purinosome assembly. In numerous cancers, specific pathway enzymes are usually upregulated, reflecting its importance in tumorigenesis^{79,80}. Mutations of these enzymes may affect purinosome formation, which is likely to mediate cell cycle and enhance sensitivity to cancer chemotherapeutics⁸⁰. In addition to mutations, purinosome disassembly can also be induced by inhibition of microtubule polymerization and purine supplementation⁸¹.

Studies have shown that microtubule depolymerization induced by nocodazole resulted in the loss of purinosome-mitochondria colocalization, suggesting that the association of purinosomes with

mitochondria is facilitated by microtubule-directed transport, and thereby supporting the notion of an interdependency between these two components in order to maximize purine production⁸². The use of HSP90 inhibitors to treat cancer can also be advantageous, since their therapeutic effect may result from the combined effects on multiple oncogenic proteins that are responsible for tumor progression, including the ones that form the purinosome⁸³.

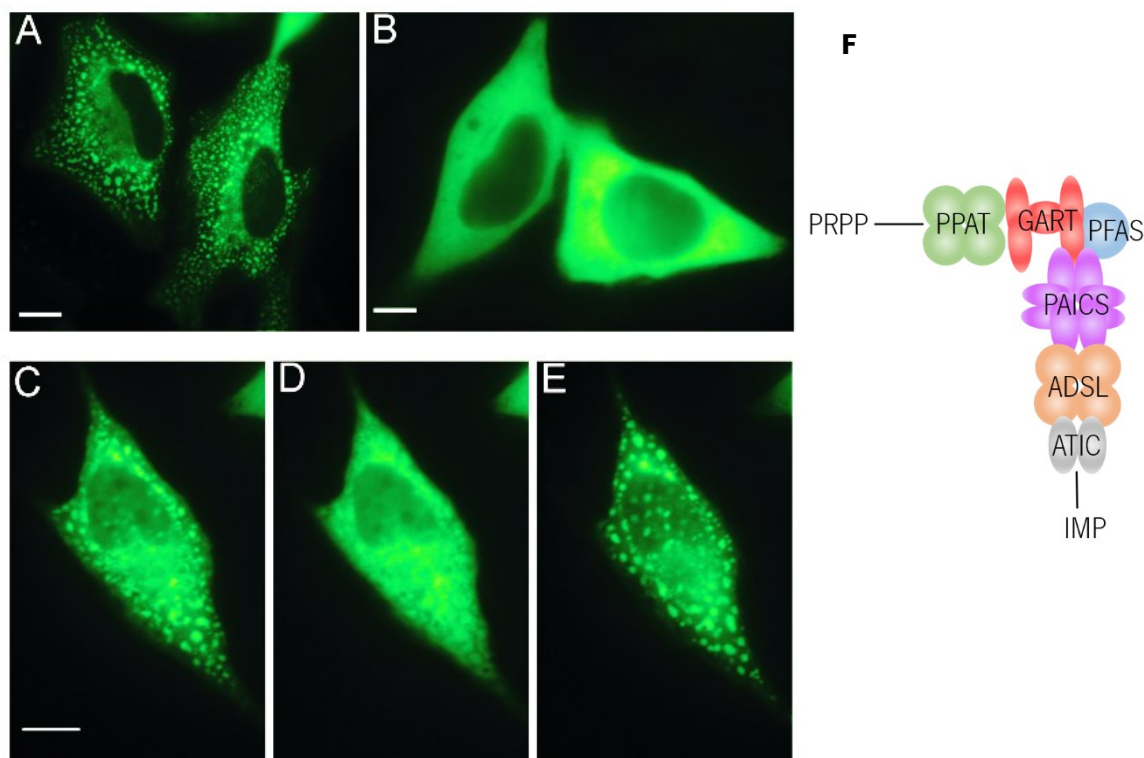


Figure 10: Images of the purinosome and purinosome multienzyme complex. The pathway enzyme FGAMS (also known as PFAS) was fused with green fluorescent protein (GFP). (A) Purinosomes formed in HeLa cells transiently transfected with FGAMS-GFP in purine-depleted medium. (B) Diffuse fluorescence signal of FGAMS-GFP in HeLa cells in purine-rich medium. Reversible formation of clusters by FGAMS-GFP is shown in (C)–(E). (C) Purinosome formed when HeLa cells are cultured in purine-depleted medium. (D) Purinosomes disperse within 2 hours upon incubation with purine-rich medium. (E) Purinosomes reformed after returning to purine-depleted medium for 1 hour (Scale bar, 10 μ m)⁸⁴. (F) Six enzymes are involved in *de novo* purine synthesis and the purinosome complex: PPAT, the trifunctional enzyme GART, PFAS, the bifunctional enzyme PAICS, ADSL, and the bifunctional enzyme ATIC⁸⁴.

1.3 Computer-aided drug design: using artificial intelligence techniques for activity and multi-target prediction

The development of a chemical entity and its testing, evaluation, and authorization to become a marketed drug is a laborious and expensive process that is prone to failure⁸⁵. Indeed, it is estimated that just 5 in 5000 drug candidates make it through preclinical testing to human testing and just one of those tested in humans reaches the market⁸⁶. The discovery of novel chemical entities with the desired biological activity is crucial to keep the discovery pipeline going⁸⁷. Thus, the design of novel molecular structures for

synthesis and *in vitro* testing is vital for the development of novel therapeutics for future patients. CADD methods have become a powerful tool in the process of drug discovery and development⁸⁸. These methods include structure-based design such as molecular docking and dynamics, and ligand-based design such as quantitative structure–activity relationships (QSAR) and pharmacophore modeling. Molecular modeling and cheminformatics have made notable contributions to drug discovery⁸⁹. Several important drugs have been developed with computational methods, including imatinib (a kinase inhibitor used to treat certain types of cancer)^{90,91}.

Cheminformatics is the application of computational methods to chemical problems, with particular emphasis on the manipulation of structural information. Cheminformatics is a relatively new field of information technology that focuses on the collection, storage, analysis, and manipulation of chemical data⁹². The chemical data of interest typically includes information on small molecule formulas, structures, properties, and biological activities, which can then be used in drug design or other applications.

The origins of the biological activity of interest's mechanism of action have been uncovered in large part thanks to QSAR. This field determines the crucial chemical characteristics that lead to good or bad biological activity by using historical data on biological activity to predict the activity of novel substances⁹³⁻⁹⁵. The concept of structure-activity cliffs demonstrated that even a minor change in the chemical structure such as addition or deletion of functional groups or the stereoisomeric placement of those groups may be a deciding factor whether the compound can or cannot bind to a specific target, which can give rise to significant alteration in activity^{96,97}. The SAR can be quantified by using a machine learning approach. Machine learning is an evolving branch of algorithms that are designed to emulate human intelligence by learning from the surrounding environment. These techniques have been applied in diverse fields ranging from pattern recognition, engineering, finance, entertainment, and computational biology including biomedical and medical applications⁹⁸. In drug design, the algorithm can learn from a training dataset comprising previously studied molecules, their structure, properties and biological activity. After the learning stage, the algorithm is then tested using a different set of molecules and its predictions/labels are evaluated and validated. Then, the final algorithm can be used to predict labels of new cases based on what it has previously “learned” (**Figure 11**).

This approach can be used not only to predict the biological activity of compounds that have already been studied or approved to pave the way for drug repurposing, but also to give leads on the targets of newly designed molecules. It's important to note that, since the algorithm makes predictions based on previously known information, when it is introduced to a new compound that is of contrastive structure and properties included in the training dataset, the predictions made might not be the most confident. This means that, if a molecule is too “novel”, the algorithm might not be so sure of the molecule's activity or target. In addition, most open-source algorithms for drug activity prediction are only based on 2D structure and therefore, the compounds' 3D structure does not weigh on the algorithms' decision-making. For these reasons, the results obtained by using machine learning tools designed for understanding chemical-biological interactions should always be complemented with molecular docking studies and experimental results.

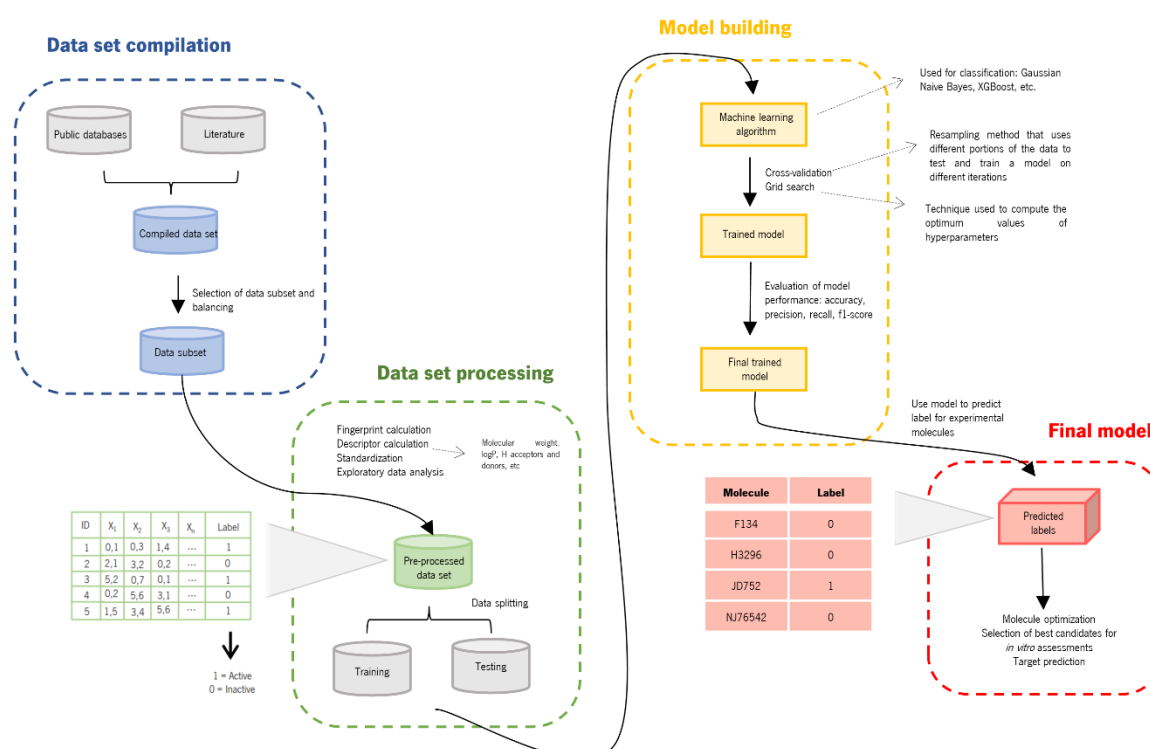


Figure 11: Using machine learning approaches to predict compound activity. First, data about active and inactive compounds for various targets are extracted from public databases (like ChEMBL, PubChem, etc) and literature. This data is then balanced, processed, and divided into a training dataset and a testing dataset. The first one will be used to “train” the algorithm and the second one to see if the algorithm is making accurate predictions based on what it “learned” from its training. The final model can then be used to predict the activity or targets for new compounds, paving the way for structure and experimental optimization.

Molecular docking is a key tool in computer-aided drug design, which consists of a computer simulation procedure to predict the lowest energy conformation of a ligand in a receptor- ligand complex. The goal of ligand-protein docking is to predict the predominant binding mode(s) of a ligand with a protein

of known three-dimensional structure^{99,100}. Additionally, these tools can predict the final binding energy and inhibition constants of a molecular conformation, which is modeled in terms of dispersion and repulsion, hydrogen bond, desolvation, electrostatic, torsion free energy, final total internal energy and unbound systems' energy. Therefore, detailed understanding of the principles that contribute to the predicted binding free energy provides information about the nature of ligand-protein interactions¹⁰¹.

1.4 Purine derivatives as polypharmacological agents: multitarget drug discovery for the future

Multitargeting compounds comprising activity on more than a single biological target have gained remarkable relevance in drug discovery owing to the complexity of multifactorial diseases such as cancer, inflammation, or the metabolic syndrome. The concept of polypharmacology is related to the interaction of drug molecules with multiple targets, which may interfere with a single or multiple disease pathways¹⁰². Polypharmacological drug profiles can produce additive or synergistic effects while reducing side effects and significantly contribute to the high therapeutic success of indispensable drugs. While their identification has long been the result of serendipity, medicinal chemistry now tends to design polypharmacology. Studies have concluded that despite some characteristic challenges remain unresolved, designing polypharmacology and multitargeting compounds holds enormous potential to secure future therapeutic innovation¹⁰³.

As purines are involved in most essential biological processes, it is important to note that novel purine derivatives might be able to interact with different targets. Different investigations have firmly established that many active compounds interact with multiple targets and that in most cases, targets of promiscuous compounds are related to each other (compound promiscuity across different target families is rare)¹⁰⁴. Multi-target activities of drugs provide the basis for polypharmacological effects, which are frequently responsible for therapeutic efficacy, with protein kinase inhibitors used in oncology being a paradigmatic example³². In this case, **PP121 (Figure 12A)** was shown to target both tyrosine kinases and PI3K family members. **Reversine (Figure 12B)**, in addition to inhibiting MEK, also has an inhibitory effect on receptors from different families, such as JNK1 (a protein kinase involved in various processes such as cell proliferation, differentiation, migration, transformation and programmed cell death)¹⁰⁵, NM II (non-muscle myosin II, an actin-binding protein that controls cell protrusion, adhesion and polarity)¹⁰⁶ and AUKRB¹⁰⁷.

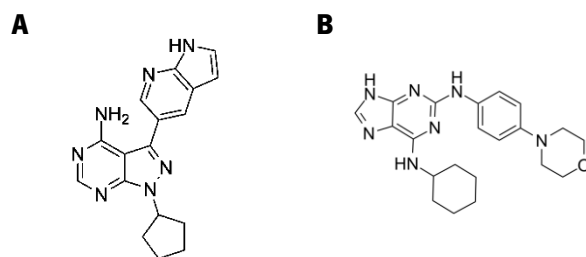


Figure 12: Purine derivatives as polypharmacological agents: (A) **PP121** and (B) **reversine**.

Molecules with polypharmacological effects were considered for many years to be a "double-edged sword", since unintended polypharmacology can lead to adverse and unexpected effects as the compounds progress in clinical trials. However, improved understanding of signaling pathways and other biological mechanisms that are crucial for cellular survival are leading scientists to believe that multitarget compound development is the key for future drug discovery. For example, the first purine derivatives to be developed as kinase inhibitors were initially thought to elicit their therapeutic effect by inhibiting a particular kinase, but nowadays, scientists have found that these compounds actually owe their success to a more complex pattern of activities.

The molecular and genetic complexity of advanced-stage diseases such as cancer suggests that targeting a single oncogenic pathway may not be sufficient to achieve durable remissions in patients¹⁰⁸. Accordingly, novel drug discovery and development strategies are focusing on targeting multiple signaling pathways, either with drug combinations or through the design and development of a single compound able to target multiple oncoproteins¹⁰⁹. Since purine derivatives comprehend a wide range of modifications and targets, it is necessary to characterize the different signaling pathways that are hyperactive in the TME, in order to design and optimize new compounds for a specific target or a multi-target strategy, according to its SAR.

The general structure of purines has a wide range of possible modifications and their derivatives are involved in most biological mechanisms essential for the survival of cells and tissues. For this reason, and in combination with new and effective biological assessment techniques, the study of molecules derived from purines has an enormous impact in drug design and treatment not only of cancer, but also of other complex pathologies, paving the way for drug repurposing. Scientists now have the reason to believe that important therapeutic applications of newly designed and optimized purine derivatives are on the edge of being discovered.

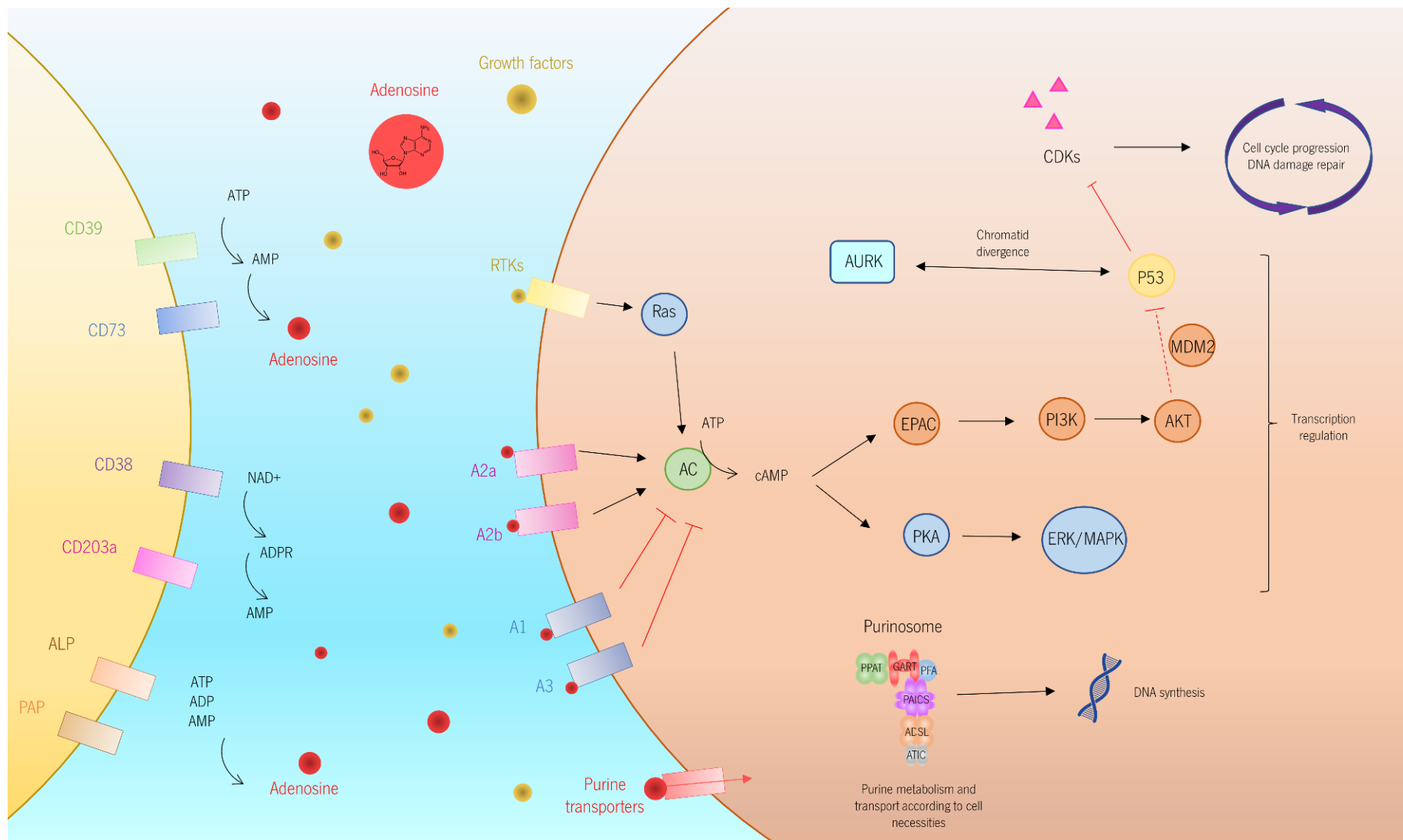


Figure 13: Possible target pathways for anticancer purine derivatives, according to literature. Purine derivatives may interfere with these pathways by, for instance, acting as adenosine or ATP agonists or as inhibitors of different receptors, according to the literature.

1.5 Drug formulations: solid lipid nanoparticles (SLNs) as drug delivery systems

Over the past three decades, significant advances have been made in drug delivery technology. Drug delivery technologies represent a vast, vital area of R&D of pharmaceuticals and the demand for innovative drug delivery systems continues to grow¹¹⁰. A drug delivery system is defined as a formulation or a device that allows the introduction of a therapeutic compound in the cell and improves its effectiveness and safety by controlling the rate, time and place of release¹¹¹. Depending on the characteristics of the active compounds to be administered, it may be possible to insert them in drug delivery systems such as encapsulation. After verifying whether the molecules can be encapsulated, the release and internalization of the therapeutic molecules in the cell can be evaluated and the cytotoxic effects of the free vs encapsulated drug can be compared. In this way, it is possible to improve the therapeutic efficacy of the molecules, not by optimizing the structure of the compounds themselves, but by improving their internalization and cell localization system.

Solid lipid nanoparticles (SLNs) are colloidal nanoparticles composed of a lipid matrix (solid at both room and body temperatures¹¹²), and surfactants used as stabilizing and solvating agents¹¹³. Different lipid and surfactant compositions can have an impact on the formulations' size, polydispersity, surface charge, stability, and drug release profile¹¹⁴. Frequently, waxes, fatty alcohols, and fatty acids like mono-, di-, and triglycerides are utilized to prepare SLNs¹¹⁵. The small size of the formulations (ranging from 10 to 1000 nm) and the high drug encapsulation efficiency are the key advantages of SLNs. Additionally, by enhancing their bioavailability and regulating the drug release rate¹¹⁶ (avoiding the "burst effect"¹¹⁷, which is a major drawback of the drug delivery systems since they could expose the patient to a drug overdose¹¹⁸), these formulations can enhance the therapeutic efficacy of hydrophobic agents. SLNs can also accumulate in the tumor regions through active delivery mechanisms: the surface of the SLNs is functionalized with ligands that can specifically identify overexpressed receptors on the surface of cancer cells and, ultimately, be translocated into the cells. (**Figure 14**). Passive mechanisms of targeting include different charges or PEGylation^{119,120}.

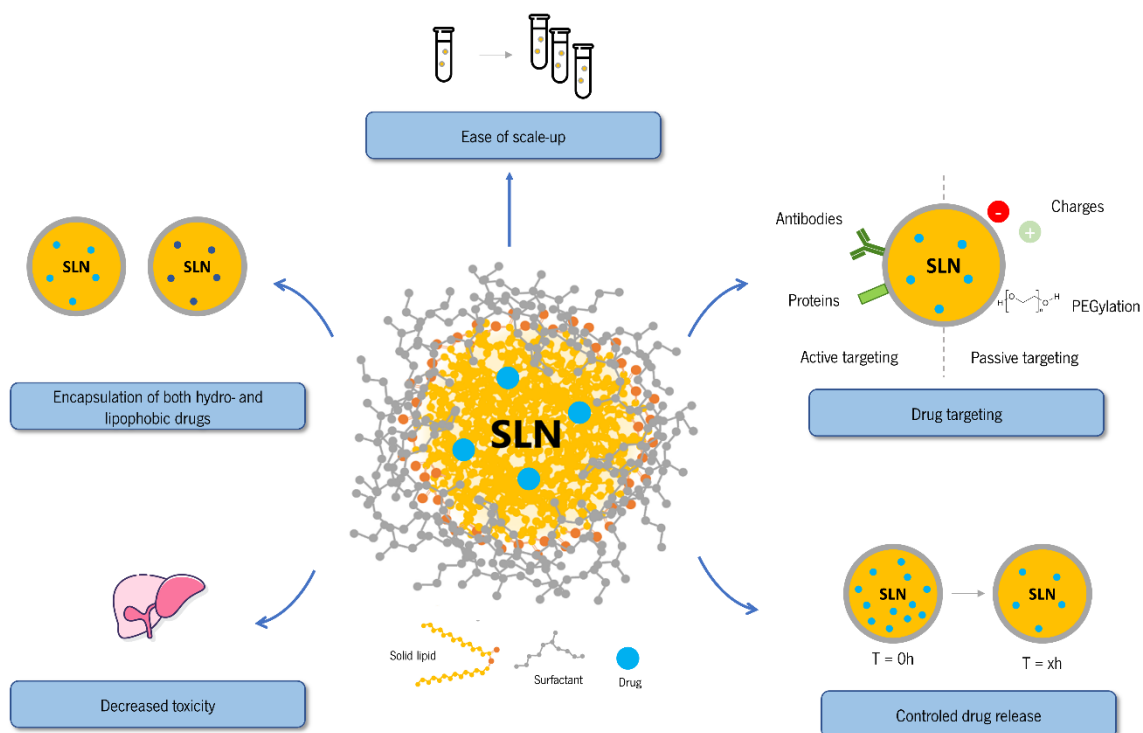


Figure 14: Main applications and advantages of SLNs (Adapted from¹²¹).

When the SLNs deliver the medicine specifically to its site of action, the therapeutic impact of the encapsulated drug is expected to be more effective. However, in addition to the physicochemical qualities of the nanoparticles, the effective accumulation of nanoparticles in solid tumors also depends on the characteristics of the tumors' microenvironment. Since tumors have increased vascular permeability and poor lymphatic drainage, it has been argued that the enhanced permeability and retention (EPR) effect might theoretically lead to the passive accumulation of nanoparticles, liposomes, or other carriers and macromolecules in tumors¹²². The EPR effect is a term used to describe a universal pathophysiological phenomenon and mechanism in which macromolecular compounds (above 40 kDa) can gradually accumulate in the tumor vascularized area, allowing for the targeted delivery and retention of anticancer agents into solid tumor tissue¹²³. This high vascularization of the tumor results from the own tumor's development requirements, which demands a high and constant supply of nutrients and oxygen to be able to sustain its uncontrolled proliferation. The process of inducing new blood vessels is known as angiogenesis, and it is carried out by malignant cells secreting proteins and growth factors¹²⁴. The rapid formation of new capillaries combined with the absence of the basal membrane, which serves as the vasculature's support tissue, can result in the formation of aberrant vascular architecture with

endothelium gaps that range in size from 200 nm to 2 μm ¹²⁵. Due to their characteristically small sizes compared to pore size, circulating nanoparticles can now easily enter the tumor location through the gaps seen in the surrounding blood vessels^{119,122,125-127}. Due to a lack of the lymphatic system, an elevated retention can be seen in conjugation with an enhanced permeability. This is because the increased hydrodynamic size of the nanoparticles prevents them from returning to the nearby capillaries, which ultimately lengthens their retention time in the tumor^{119,120,128} (**Figure 15**).

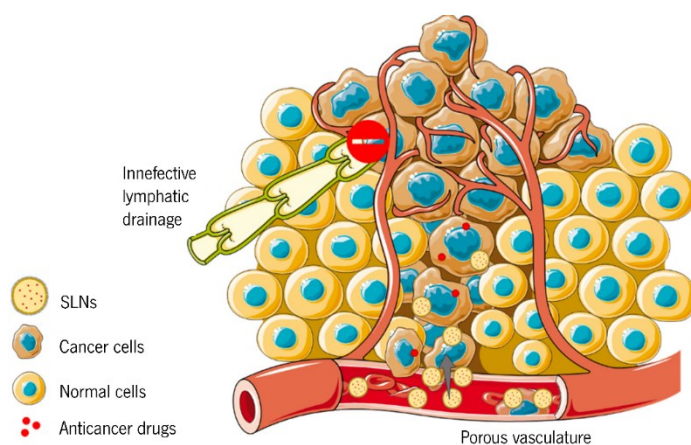


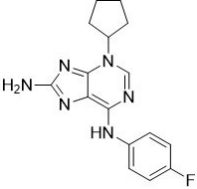
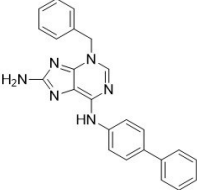
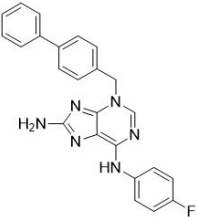
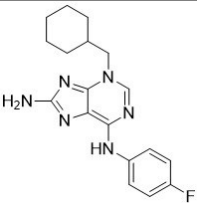
Figure 15: The EPR effect and the absorption of nanoparticles across cancerous tissues are shown schematically. Due to the leaky vasculature in the tumor location and the malfunctioning lymphatic system, the EPR effect encourages an enhanced concentration of nanoparticles in cancer cells in contrast to normal cells (Adapted from ^[31]).

1.6 Previous work

A novel class of purine derivatives has been prepared as new anticancer agents at the University of Minho Chemistry Centre. The synthesis of 6,8-diaminopurine class compounds has been described in previous studies¹²⁹. Additionally, the compounds were screened against HCT116 and p53-wt cell lines previously¹³⁰. In the present work, the top 4 compounds with the best IC_{50} values determined previously in HCT116 cell lines were chosen for further studies (**Table 1**). Additionally, the compounds' molecular properties were calculated using SwissADME¹³¹ and all molecules passed Lipinski, Ghose, Egan, Veber and Muegge drug-likeness filters¹³²⁻¹³⁶ with 0 violations.

Based on these results, these compounds are expected to be promising candidates for drug development as anticancer agents. However, there is still a need to screen and evaluate these molecules on different human tumor cell lines and healthy cell lines.

Table 2: Purine derivatives used in biological studies and their molecular properties.

Compound #	Structure	LogP	MW (g/mol)	MR	Heavy atoms	Rotatable bonds	Fraction Csp3	Num. H acceptors	Num. H donors	TPSA (Å ²)
1		2.61	312.15	87.61	23	3	0.31	4	2	81.65
2		3.82	392.46	120.46	30	5	0.04	3	2	81.65
3		4.13	410.45	120.41	27	5	0.04	4	2	81.65
5		2.92	326.37	92.41	24	3	0.35	4	2	81.65

2 Work relevance and objectives

A study conducted by Hofmarcher et al.¹³⁷ presents evidence on the economic burden of cancer in Europe, considering health expenditure on cancer care, cancer drugs expenditure (not including funding for cancer research), informal costs of time forgone by relatives and friends to provide unpaid care, and costs from productivity loss from premature mortality and morbidity. Altogether, cancer caused an estimated cost of 199 billion € to society in Europe in 2018. According to the Special Committee on Beating Cancer of the European Parliament's background note on cancer research of 2020¹³⁸, the rising number of cancer cases combined with the predicted increase in advanced-stage diagnoses and cancer mortality caused by the pandemic⁹⁻¹¹, makes basic cancer research crucial for the continuous advancement in cancer prevention, diagnosis, treatment and follow-up care.

As previously mentioned, research on purine derivative's potential as therapeutic agents since the 1950's resulted on the continuous development of multiple drugs still used today, not only to treat cancer, but also other diseases. The ability of these compounds to interact with different targets, resulting in polypharmacological effects, makes them strong candidates to treat not only complex diseases such as cancer, but to also be used in drug-repurposing. This means that, even if a specific class of purine derivatives has no positive anticancer effect, its study is still relevant as the compounds may be useful in other applications.

However, it is known that drug development is expensive and time-consuming. In a recent study, which included new therapeutic drugs and biologic agents approved by the U.S. Food and Drug Administration (FDA) between 2009 and 2018, estimated that the mean cost of developing a new drug ranges from \$314 million to \$2.8 billion¹³⁹. Additionally, according to the FDA drug approval process, the full research, development and approval process can take an average of 12 years¹⁴⁰. To overcome this problem, it is crucial to use and develop new artificial intelligence (AI) and other informatic techniques. Machine learning is an application or subset of AI that allows machines to learn from data, and its growing use in patient diagnosis and drug development, optimization and discovery is changing the future of medical sciences¹⁴¹. When it comes to drug development, these techniques can be used to filter the best candidates for *in vitro* and *in vivo* assessments by predicting their activity and mechanism of action, to select the best models for clinical research based on their insertion in the interactome, to optimize the therapeutic agents to be studied, and to facilitate data management and interpretation, which may lead to a tremendous decrease in drug development duration and cost.

In addition to informatic tool application, the use of models that better mimic the diseases' physiology (in this case, the tumor microenvironment) is a very important aspect that is crucial to obtain

realistic results. In combination with newly developed drug delivery techniques, it allows for a more precise and effective treatment, reducing the side effects that are commonly described in cancer treatment and other diseases. According to results of some studies, the addition of computer-aided design techniques, including docking studies, cheminformatics and bioinformatics, to the R&D could lead to a reduction in the cost of new drug development by up to 50%¹⁴².

Taking all of the previous aspects into consideration, the present work aims to:

- Use machine learning and other informatic tools to predict the activity and targets of the novel purine derivatives and plan future options for drug repurposing;
- Assess their effect on healthy and cancer cell lines using 2D *in vitro* models;
- Evaluate the ability to encapsulate the compounds into drug delivery systems.

3 Materials and methods

3.1 Prediction of biological activity spectrum of purine derivatives

The biological activity spectrum of the selected compounds was predicted using the PASS (Prediction of Activity Spectra for Substances) online server¹⁴³. PASS is a software product used as a tool to determine the biological potential of a given organic drug-like molecule, prior to their biological testing. Prediction of biological activity is based on the analysis of SAR of more than 250000 biologically active substances (including drugs, drug candidates, lead and toxic compounds) compiled in a single training dataset. The approach used in PASS is based on the assumption that $\text{Activity} = f(\text{Structure})$. Thus, by "comparing" the structure of a new compound with structures of well-known biologically active substance, it is possible to estimate if a new compound may have a particular biological effect. It uses 2D molecular fragments known as multilevel neighbors of atoms (MNA) descriptors and attributes to each compound tested a value of probability to be active (Pa) and probability to be inactive (Pi). According to its developers, the PASS tool may be useful in:

- Drug repurposing: revealing new effects and mechanisms of action for the old substances in corporate and private data bases.
- Providing the basis for selection of the most prospective compounds for high throughput screening from the set of available samples.
- Determining the assays that are more relevant for a particular compound.

Average accuracy of prediction is estimated using a leave-one-out cross validation procedure for the whole PASS training set, which was determined to be about 95%. Since the PASS tool is open access and has been used in many studies for several years, there are many publications where PASS predictions were confirmed by biological assessments. However, as previously mentioned, PASS cannot predict the activity spectrum for essentially new compound if all its descriptors are new and so they don't occur in the training set, which means that, if a molecule is too "novel", it might not be so sure if it's active. Additionally, it makes predictions based only on 2D representation of molecules.

Novel compounds belonging to five different families (**Figure 16**) were submitted to the PASS online server in SMILES format and their biological activity and targets were predicted.

Additional methods and details are available at the PASS online website <http://www.way2drug.com/PASSOnline/index.php>.

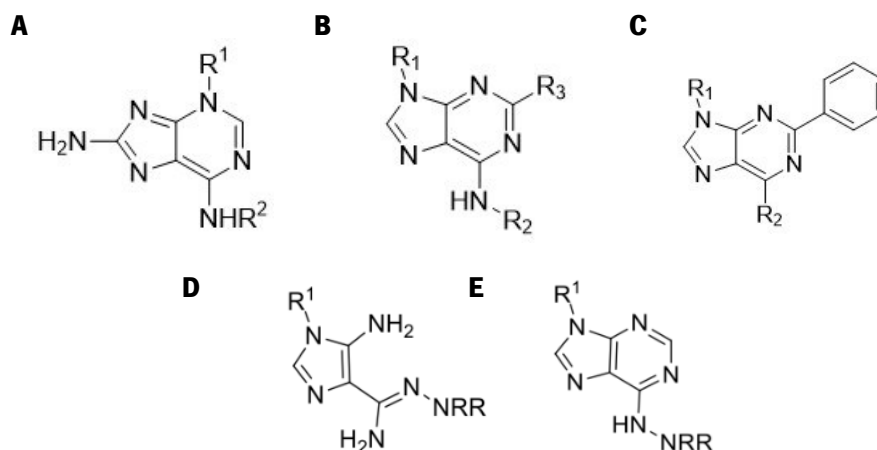


Figure 16: General structure of purine derivatives of different families: (A) Family 1 (6,8-diaminopurines), (B) Family 2 (6,9-substituted adenine), (C) Family 3 (6,9-disubstituted 2-aryl purines), (D) Family B6 (amidrazones), and (E) Family B7 (6-hydrazinopurines).

Similarly to the PASS tool, a similar machine learning technique was applied. In this case, the datasets used comprised of:

- Anticancer compounds from 3 different databases (PubChem, ChEMBL and ChemSpider). Due to the absence of inactive molecules described in literature and databases, a series of decoys were generated using DUD-E¹⁴⁴ platform.
- Experimental purine derivatives according to previous studies, where molecules with $IC_{50} > 30$ μ M were considered inactive
- Mixture of both previously mentioned
- RTK, AURK, CDK, AR and RAS inhibitors from PubChem and decoys generated with DUD-E for every target¹⁴⁴.

Decoys consist of molecules that are designed based on active ones but with certain modifications that would assumedly make them inactive. The previously mentioned information was compiled using the SMILES notation for every molecule in the dataset, each one labeled as “active” or “inactive”. Using RDkit, the next step involved obtaining the molecular fingerprints, transform them into an array and concatenate them into the dataset. Molecular fingerprints are a way to represent molecules as mathematical objects. By doing this, it is possible to perform statistical analysis and/or machine learning techniques on the set of molecules to gain new insights that we could not gain as humans. The fingerprints used included:

- Morgan fingerprints¹⁴⁵

- MACCs keys¹⁴⁶
- Atom pair¹⁴⁷

Next, some of the molecules' descriptors relevant for drug design¹³²⁻¹³⁶ and have an influence on the SAR were calculated:

- LogP
- TPSA
- MW
- Number of hydrogen acceptors and hydrogen donors
- Number of rotatable bonds
- Number of heteroatoms
- Number of valence electrons

The entire information of the dataset was then standardized and balanced using a synthetic minority oversampling technique (SMOTE), if applicable. In this technique, a random example from the minority class (in this case, active molecules) is first chosen. Then k of the nearest neighbors for that example are found. k-nearest neighbors is a classifier, which uses proximity to make classifications or predictions about the grouping of an individual data point, working off the assumption that similar points can be found near one another. In SMOTE, a randomly selected neighbor is chosen and a synthetic example is created at a randomly selected point between the two examples in feature space¹⁴⁸. With the dataset now complete, the information was split into different portions for training and testing and various algorithms were applied:

- XGBoost¹⁴⁹
- Linear regression¹⁵⁰
- Decision trees¹⁵¹
- Random forest¹⁵²
- Naïve Bayes¹⁵³

Then, the confusion matrix was obtained (**Table 2**) and a k-fold cross validation was applied. The cross-validation uses a limited sample to estimate how the model is expected to perform in general when used to make predictions on data not used during the training of the model. Generally, it results in a less biased or less optimistic estimate¹⁵⁴.

Table 2: Confusion matrix.

		Predicted class	
		Inactive	Active
Actual class	Inactive	True inactive	False active
	Active	False inactive	True active

Additionally, a grid-search was performed in order to find the best model and parameters. Grid-search is used to find the optimal hyperparameters of a model, resulting in the most effective predictions¹⁵⁵. The trained models using different combinations of parameters were evaluated on their scores of precision, accuracy, recall and f1 score for each class (active – label 1 - and inactive – label 0):

- Precision (class) = True (class) / (True (class) + False (class))
 - This formula as follows: within everything that has been predicted as a positive, precision counts the percentage that is correct, and therefore measures the quality of a positive prediction made by the model.
- Accuracy = (True (Class 0) + True (Class 1)) / Total Sample Size
 - Measures the number of predictions that are correct as a percentage of the total number of predictions that are made. As an example, if 90% of predictions are correct, the accuracy is simply 90%. However, accuracy is not a good metric to use when there is class imbalance. For example, if the training dataset has 90 inactive and 10 active molecules, the model could be predicting that every molecule tested is inactive and therefore, would have 90% accuracy. In this case, accuracy is not a good performance metric, since it is high although the model is incorrectly predicting all the active molecules.
- Recall (class) = True (class) / (True (class) + False (class))
 - When the recall is high, it means that the model accurately classifies all positive samples as positive. As a result, the model may be relied on to identify positive samples. For example, it is OK to classify a non-cancerous tumor as malignant, however, a cancerous growth should not be termed non-cancerous. In this case, when making output-sensitive predictions, models must have a high recall.
- F-1 scores = 2 * (Precision * Recall) / (Precision + Recall)

- Especially valuable when working on classification models in which your data set is imbalanced, as it combines precision and recall into a single metric.

Different combinations of the previously mentioned steps were applied in order to obtain the best performing model. The final models were saved into a file (“pickling” in Python) to later be used for predicting the activity of newly developed purine derivatives.

3.2 Protein-ligand interactions

Two of the most popular protein-ligand docking programs are AutoDock and Vina. Because these tools are free and distributed under an open-source license, they are a popular choice for many users in academia. In general, previous results demonstrated that the overall performance of Vina and AutoDock in distinguishing between actives and decoys was comparable. The findings, however, differed greatly depending on the sort of target. AutoDock performed better at distinguishing ligands and decoys in more hydrophobic, poorly polar, and poorly charged pockets, whereas Vina performed better in polar and charged binding pockets. The tendency for the type of ligand was the same for both Vina and AutoDock¹⁵⁶. For this reason, the protein-ligand simulations were performed by using both softwares.

The design of the purine derivatives and known ligands was performed by using PyMOL Version 2.5.2¹⁵⁷. Structure was checked and optimized using Avogadro Version 1.2.0^{158,159}. The crystalline structures of the relevant targets and most known inhibitors were extracted from Protein Data Bank (PDB)^{160,161} and docking simulations were carried out using Autodock Vina Version 1.1.2 and AutoDock 4 Version 1.5.6¹⁶² by following the recommended procedure in the user guide¹⁶³ using AutoDock Tools Version 1.5.6¹⁶² as an auxiliary tool (**Figure 17**). The grid box was set at the active site and the Lamarckian genetic algorithm was used to perform the simulation. The obtained conformations and their respective interactions with the surrounding aminoacids were observed ranked by energy and the final images were obtained by submitting the results into the Proteins Plus server^{164,165} and Chimera Version 1.16¹⁶⁶.

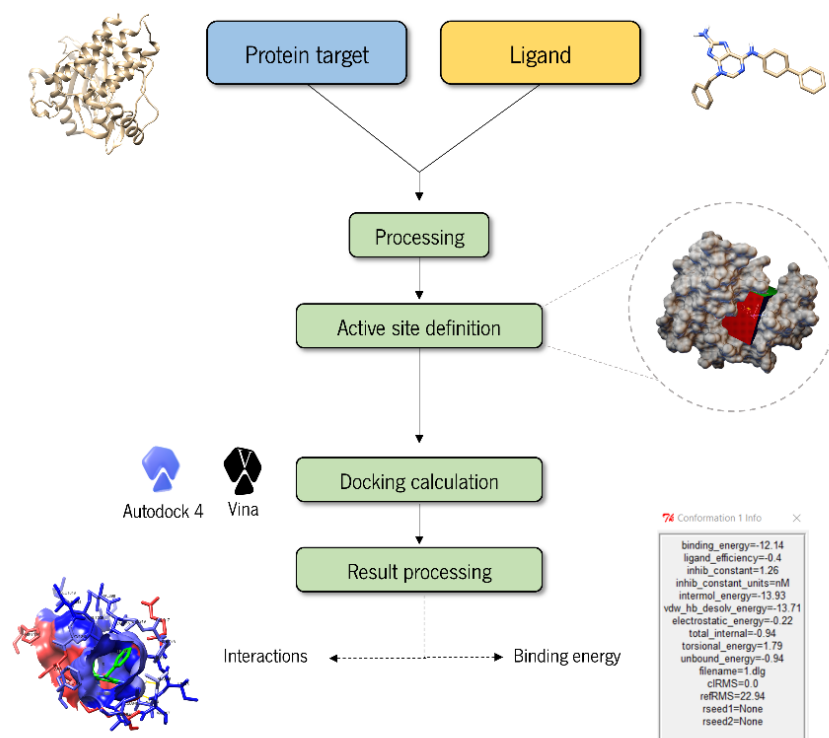


Figure 17: Overview of the docking simulation procedure. Initially, the crystallographic structure of the protein is extracted from PDB, along with its inhibitor, if present. If not present, the ligand is designed and optimized using other software. After obtaining their structure, both components are processed (water molecules and other undesirable components are removed, charges are attributed, etc.) and checked for errors. Then, the active site where the simulation will take place is defined and the calculations begin. As a result, it is possible to study the protein-ligand interactions and their affinity. Illustrative images extracted from Autodock Tools and Chimera.

3.3 Test compounds and reagents

All the compounds used in the present work were designed by random screening starting from a core of purine ring. All the compounds used in the present work were synthesized in the Chemistry Department of the University of Minho by Nádia Senhorães, Soraia Fernandes and Bruna Leite, under the supervision of Dr. Alice Dias.

Stock solutions of each compound with the concentration of 10 mM were made by dissolving the compounds into dimethyl sulfoxide (DMSO) and stored at 4°C.

Reagents, suppliers, equipment and respective models are listed in supplementary tables 6 and 7 in Appendix.

3.4 Cell lines and culture conditions

The compounds' toxicity was assessed in the following cell lines:

- HCT-15: cells isolated from the large intestine of a male Dukes C colorectal cancer patient (ATCC CCL-225) <https://www.atcc.org/products/ccl-225>;
- Hep G2: cells with epithelial-like morphology isolated from a hepatocellular carcinoma of a 15-year-old, white, male liver cancer patient (ATCC HB-8065) <https://www.atcc.org/products/hb-8065>;
- THLE-3: cell lines were derived from primary normal liver cells by infection with SV40 large T antigen (ATCC CRL-11233) <https://www.atcc.org/products/crl-11233>.

All cell lines were carefully unfrozen and maintained at 37°C in a humidified 5% CO₂ atmosphere. HCT-15 cells were maintained in DMEM 1X base culture medium supplemented with 10% (v/v) fetal bovine serum (FBS) and 1% (v/v) penicillin/streptomycin solution; Hep G2 cells were maintained in RPMI base culture with the same supplementation; THLE-3 cells were maintained in BEGM medium, also with the same supplementation.

Cancer cells were maintained and grown in T75 cm² flasks and passaged 3 times a week, while normal THLE-3 healthy cells were maintained and grown in T25 cm² flasks previously coated with 1.5 mL of a mixture of 0.01 mg/ml fibronectin, 0.03 mg/ml bovine collagen type I and 0.01 mg/ml bovine serum albumin dissolved in BEBM medium. The coated flasks were incubated overnight and the remaining mixture was suctioned off before use. The coating solution was stored at 4°C in cold room for up to 3 months.

All solutions were heated up to 37°C and mixed under sterile conditions in laminar flow chambers before use. Handling of the cells was also performed in a sterile environment using laminar flow chambers. All the cell lines were purchased on ATCC and the reagents, suppliers, equipment used and respective models are listed in supplementary tables 1 and 2 in Appendix.

3.5 2D cell plating

At 80% confluence, the culture medium was carefully removed from the culture flasks and the cells were washed with 3 mL of phosphate buffer solution (PBS, enough to cover the flask) and the solution was immediately discarded. 2.5 mL of trypsin 1% were added ensuring that all the cells were covered with the solution and then incubated for 5 minutes in the previously mentioned conditions. After incubation, 7.5 mL of previously warmed culture medium was added in order to inactivate the trypsin. A

pipette was used to pass the solution and mix it, ensuring that most cells were in solution instead of adherent to the flask. The solution was then centrifuged at 1200 rpm for 5 minutes. The supernatant was discarded and the cells were resuspended in 3 mL of fresh medium. A portion of this cell suspension was used to continue with the cell culture according to the manufacturer's recommendations. 20 μ L of the remaining cell suspension were taken and put on the haematocytometer to count the cells using an inverted optical microscope.

All the solutions mentioned were previously warmed at 37°C before use. Reagents, suppliers, equipment used and respective models are listed in supplementary tables 6 and 7 in Appendix.

3.6 Cell viability

Resazurin has been broadly used as indicator of cell viability in several types of assays and it was the selection of choice for this work. Resazurin is a redox indicator that is cell permeable and can be used to monitor viable cell number. When cells are metabolically active, mitochondrial enzymes are responsible for the transference of electrons, reducing resazurin (blue and not fluorescent) to resorufin (pink and fluorescent). The quantity of resorufin produced is proportional to the number of viable cells and can be quantified using a microplate fluorometer^{167,168}.

HCT-15 cells were seeded at 6.5×10^4 cells/mL and Hep G2 cells were seeded at 5×10^4 cells/mL in a 96-well plate and incubated for 24h in the previously mentioned conditions. The medium was then removed and new one was added together with the compounds to be tested into each well, with concentrations ranging from 10 μ M to 100 μ M. Controls with 10% DMSO, 1% DMSO, and media only were also added. For THLE-3 cells, the wells were previously coated as stated in section 3.4 before seeding at 5×10^4 cells/mL. In this cell line, the treatment was added directly to the well after 24h incubation, without removing the media. The plates were then incubated for 24h and 48h. After each time point, 6.97 % (v/v) of resazurine at 10X concentration was added to each well and incubated again for 3.5 hours. Fluorescence intensity was measured with 535 nm excitation and 595 nm emission using a plate reader (**Figure 18**).

Reagents, suppliers, equipment and respective models are listed in supplementary tables 6 and 7 in Appendix.

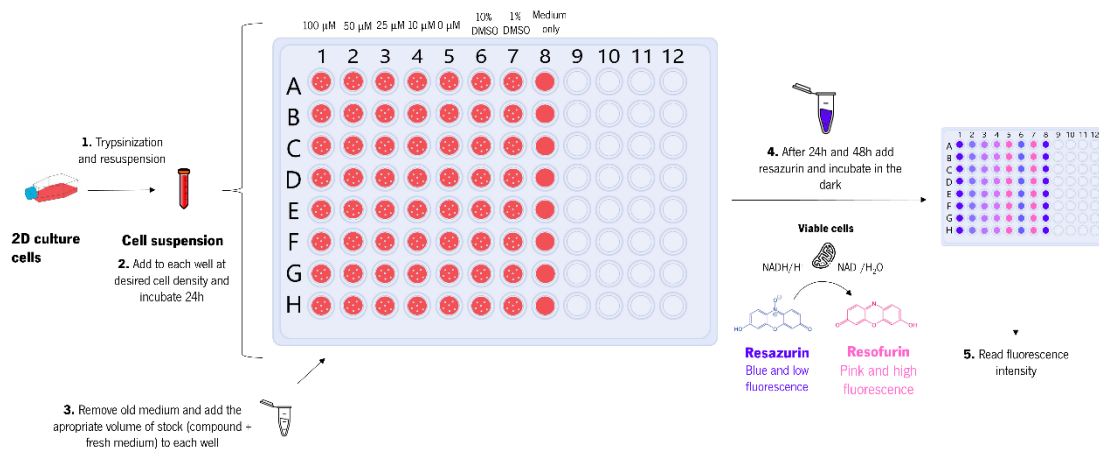


Figure 18: 2D cell plating and resazurin viability assay procedure.

3.7 Cell lysis, protein extraction and quantification

The plated Hep G2 cells were washed with pre-heated PBS and trypsin was added, according to section 3.5. After incubation, medium was added, and the cells were collected. The cell suspension was then centrifuged at 1200 rpm at 4°C for 5 minutes to pellet the cells. The supernatant was removed, and the pellet was resuspended in 1 mL of ice-cold PBS with 1X protease inhibitor cocktail. This suspension was then frozen at -80 °C for 5 minutes and thawed at room temperature (RT) for 3 cycles. The suspension was then centrifuged at 10000 g at 4 °C for 15 minutes and the supernatant (whole cell protein extract) was collected and kept in ice for the steps ahead (**Figure 19**).

The protein was quantified using the Bradford method. The whole protein sample was diluted in PBS buffer 1:5 and 5 μL of the dilution were added to a microplate. Different concentrations of bovine serum albumin (BSA) were used to establish a standard curve. 200 μL of Bradford reagent¹⁶⁹ were added to each well and the absorbance was read at 595 nm (**Figure 19**).

Reagents, suppliers, equipment and respective models are listed in supplementary tables 6 and 7 in Appendix.

3.8 Measurement of ATPase activity (inorganic phosphate (P_i) quantification)

50 µg of protein was mixed with Y-ATP solution composed of 300 µL of 3 mM ATP, 0.02% Triton X-100, 50 mM KCl, 1 mM sodium molybdate, 6 mM MgSO₄ in 30 mM Tris pH 8, with 100 µM of test compounds. Sample without compound and a sample with 1% DMSO were used as controls. After a 30-minute incubation at 37°C with slow agitation (80 rpm), color stop solution was added in a 1:1 proportion. This solution is meant to stop the enzymatic reaction and acts as a coloring agent, turning the samples with more P_i into a bright yellow color. The color stop reagent was prepared previously by mixing 25 mL ammonium molybdate solution [10 g ammonium molybdate, 90 mL: distilled water, 1 mL NH₃ (25 %) adjusted to 100 ml] with 25 mL ammonium vanadate solution (0.235 g ammonium vanadate, 40 ml distilled water at 60 °C, slowly adding 2 mL nitric acid before adjustment to 100 mL with distilled water) and then slowly adding 16.5 mL nitric acid (65%). Then, the samples were centrifuged at 2400 g for 5 minutes at RT. 200 µL of supernatant was collected and transferred to a 96-well plate, in order to measure the absorbance at 415 nm using a blank control performed without protein and NaH₂PO₄ as standard to establish a calibration curve (**Figure 19**)^{170,171}.

Reagents, suppliers, equipment and respective models are listed in supplementary tables 6 and 7 in Appendix.

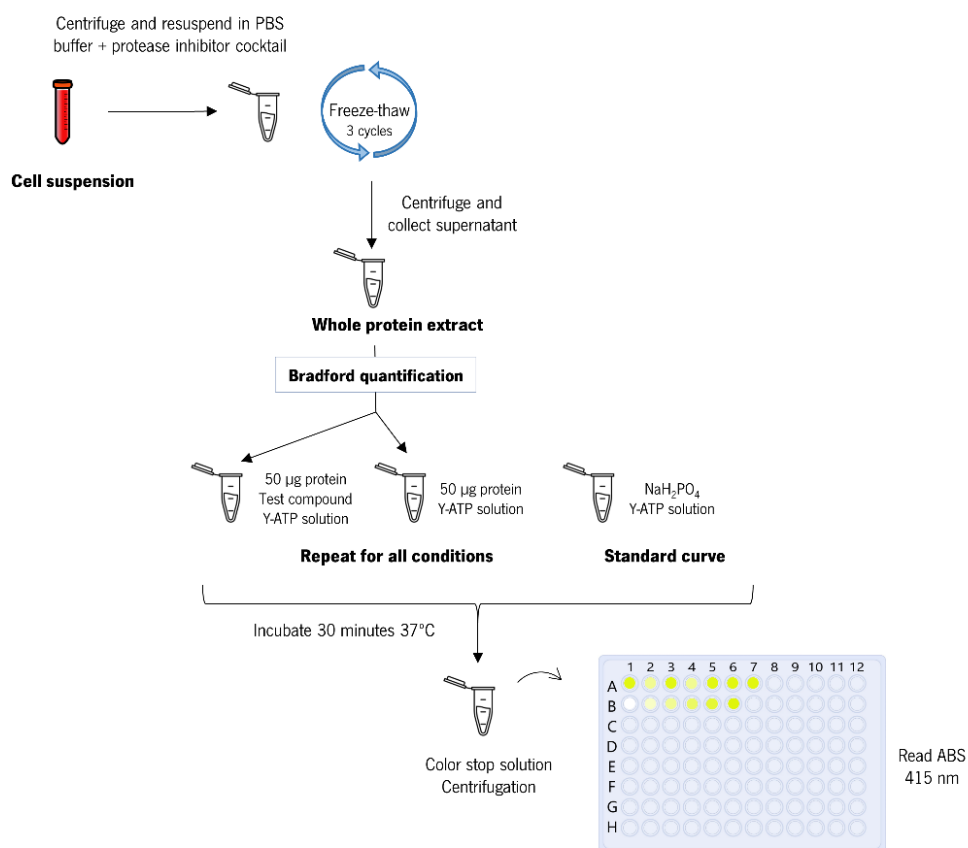


Figure 19: Pi quantification assay procedure.

3.8 Drug encapsulation using SLNs

100 mg of carnauba wax were in the water bath at around 90°C. After melting, 10 mg of compound were added to the carnauba wax and mixed with a hot spatula. Tween-80 solution at 25 mg/mL was preheated at the same temperature in the water bath. 0,5 mL of this solution and 1,75 mL MiliQ water (also pre-heated) were added to the wax/drug mixture. The sample was then sonicated for 6 on/off cycles of 20 seconds with 25% amplitude. The samples were left to cool down on running water before collecting the supernatant by centrifugation. After a 10-minute centrifugation at 3000 rpm, the supernatant was collected and kept in the fridge. The final formulation was then analyzed using dynamic light scattering (DLS) by diluting the samples 1:100 and using a 6 mm carbon electrode cell (particle refractive index = 1.460-0.000i; refractive index of the dispersion medium = 1.333; duration = 90 seconds; number of repeated measurements = 3).

DLS is an optical technique used for analyzing dynamic properties and size distribution of a broad variety of physical, chemical and biological materials suspended in a liquid, which could be colloidal particles, bubbles, droplets or macromolecules. DLS measures the speed of the particles in suspension undergoing Brownian motion to obtain a translational diffusion coefficient. Brownian motion is the random movement of particles due to their collision with solvent molecules that surround them. The particles' size is dependent on diffusion speed: small particles diffuse faster, while larger particles diffuse slower. The instrumental setup involves the illumination of the sample by a laser beam. The particles scatter the light and the interference of scattered waves generate scattered light intensity signals. Since the suspended particles undergo Brownian motion, it results in fluctuations of the scattered light intensity. As a result, plots are obtained showing intensity changes with time (slower fluctuations for larger particles and rapid for smaller). The systems' correlator takes rapid snapshots of scattered light and allows diffusion coefficient calculation, which can later be used to calculate the particles' hydrodynamic diameter according to the Stokes-Einstein equation¹⁷².

In a colloidal system, dispersed particles have two layers of oppositely charged ions on the surface, called the stern and double layers. The zeta potential is defined as the voltage at the edge of the double layer. If two particles have high enough zeta potentials of the same sign, they will not agglomerate because of those like charges repelling each other. One important use of the zeta potential is to predict the long-term stability of particles: if zeta is less than -60 mV or above +60 mV, the particles have excellent stability. Conversely, zeta potentials between -10 and +10 mV are likely to agglomerate¹⁷³ (PEG ligands can protect them sterically, avoiding agglomeration¹⁷⁴).

Reagents, suppliers, equipment and respective models are listed in supplementary tables 6 and 7 in Appendix.

3.9 Statistical analysis

For the viability assay, the mean of the fluorescence intensity values for the cell control (0 μ M) and medium only control were calculated using Excel. Then, for each of the replicates of each condition, the percentage of viability was calculated using the following formula:

$\% \text{ viability} = 100 * \frac{\text{Condition-Medium}}{\text{Cell control-medium}}$, where "Condition" is equal to the fluorescence intensity of every replicate for every condition, "Medium" is equal to the mean of the fluorescence intensity values

for the wells containing medium only, and “Cell control” is equal to the mean of the fluorescence intensity values for the wells containing no testing compound. After normalization, the concentrations were converted into log (M) and a non-linear fit curve was obtained for calculation of IC₅₀.

For the ATPase activity assay, the absorbance value for each condition without compound was subtracted to the sample in the same condition with testing compound, to remove the absorbance caused by spontaneous hydrolysis of ATP to ADP + P_i. This way, only the P_i formation caused by the presence of the testing compound is considered. The results for each condition were converted to % and compared to the control samples (corresponding to 100% activity).

For all tests, a minimum of 3 independent replicates with 3 technical replicates were performed. The values obtained were then inserted into GraphPad Prism 9 Version 9.3.0. This software was used to normalize the data and perform an ordinary ANOVA and checking for significance and to identify outliers for every condition of every test.

4 Results and discussion

4.1 Prediction of activity and targets

The complete list of predicted activities for each family is listed on supplementary tables 1 to 5 in appendix. Overall, according to the PASS online server (considering $P_a < 0,4$ and $P_a > P_i$), the types of targets and biological activity seem to depend on the purine family. This means that, when the position and type of substituent groups change, so does the activity. In general, results show that:

- Families 1 (6,8-diaminopurines) and 2 (6,9-substituted adenine) were the ones with more predicted targets related to cancer and other inflammatory diseases, and also antimicrobial activity.
- All the compounds tested belonging to family 1 were predicted to have both antineoplastic activity (non-Hodgkin's lymphoma) and to be CDK9/cyclin T1 inhibitors.
- Family 2 was the only family to hit on adenosine 2 receptors.
- Family B7 (6-hydrazinopurines) were the one with more predicted activity related to neurological or heart diseases by targeting adenosine 1 receptors.

The attempt to build a more accurate machine learning model optimized for purine derivatives described in section 3.1 was most successful when using a more restrictive dataset comprised of known inhibitors and generated decoys from literature, namely RTK, AURK, CDK, AR and RAS. In this attempt, the dataset included 48 active inhibitors and 300 decoys (approximately 50 for each target), labeled as active (corresponding to label 1) or inactive (label 0). Then, every structure was converted into a Morgan molecular fingerprint with radius 2. The previously mentioned descriptors were calculated and concatenated into the fingerprints. The dataset was then standardized and split into training and tested datasets in a 80%/20% proportion. Given the unbalanced data (more inactive samples than active samples), SMOTE was applied. This technique significantly improved the model's final recall metric. The resampled training dataset was then used to train the XGBoost algorithm. Finally, a k-fold cross validation and grid search were performed. The metrics for this procedure are represented on **tables 3 and 4**.

Tables 3 and 4: Confusion matrix (left) and respective metrics (right) obtained.

		Predicted	
		0	1
Actual	0	59	0
	1	3	7

Metric \ Label	Accuracy	Precision	Recall	F1 score
0 (inactive)	0,96	0,95	1	0,98
1 (active)		1	0,70	0,82

According to the metrics, it is possible to infer that the final model has a high accuracy, as 96% of the data is correctly labeled. Additionally, each type of prediction is very precise, reflecting the quality of the model's predictions. However, the percentage of recall for the inactive class is higher than the active class (100 vs 70%). This means that the model is more likely to correctly classify inactive molecules than active ones. As we can see, the model incorrectly classified 3 active molecules as inactive, leading to a lower recall value.

Taken these metrics into account, the model was saved using pickle and then used to predict the activity of the novel purine derivatives. The dataset containing all the purine derivatives and the respective SMILES notation was treated the same way as the train/test dataset and their predictions were obtained. Out of a total of 67 novel purine derivatives tested, 64 of them were predicted as active. None of the 3 molecules predicted as inactive were part of the ones used in biological assays.

It is important to note that, given the nature of the dataset used to train the model, it only allows for prediction of activity. Therefore, it does not allow us to take conclusions about the specific targets it might be interacting with. Since the data was composed of information about active and inactive molecules against the five proteins (RTK, AURK, CDK, AR and RAS), it means that the molecules predicted as active, are likely to be active in these proteins overall.

The attempt to build a machine learning model using other datasets was not successful. The various combinations of different methods used (such as molecular fingerprints, algorithms, databases, etc) did not lead to a final model that was capable of predicting already existing anticancer agents' or novel agents' activity precisely. This could be the result of the lack of inactive compounds described in literature, as without that information, the model is not able to establish a correlation between the molecules' characteristics and its activity. Therefore, it was unable to determine a QSAR value. The use of the generated decoys is not the most reliable method as they are generated automatically and are not experimentally validated as inactive. This means that, when using decoys, we cannot know for certain if they are good examples of inactive molecules. Another reason that could be leading to the model's

inefficiency could be the dataset itself: using a wide range of targets and their respective inhibitors as examples to train the model could be an obstacle to establish a correlation between molecular descriptors and activity. For example, if a cancer protein requires a drug with big bulky groups in a certain position to be inhibited, but another protein target does not, the model cannot use this criterion to determine if a compound can be labeled as an anticancer agent. If this happens when considering other molecular descriptors, it will lead to an inaccurate model. This is especially relevant in the case of developing a model whose goal is to predict the activity of a molecule in the treatment of complex diseases, such as cancer. Since there are so many diverse targets to be considered, it is particularly difficult to build a “one size fits all” algorithm. Additionally, it is important to consider that if the novel molecules whose label is being predicted are very contrasting in terms of structure and properties when compared to other molecules described in databases, any model will most likely perceive them with a lower probability of being active. In these cases, in particular, the use of machine learning methods might not be the best approach to determine a novel molecular entity’s activity. However, by restricting the training dataset to be composed of targets already known to interact with purine derivatives, it was possible to predict the novel compounds’ activity effectively.

Although the PASSonline predictions did match the probable targets found in literature for these new molecules, the method used is not described clearly. For this reason, although public access tools like PASS could in fact give an indication of a target pathway, it should not be considered as an accurate or definite result. These predictions should always be compared to previous literature search, and the targets’ interactions with the molecules being studied should be assessed through molecular simulations in order to determine which molecules should be used in further biological assays.

As aforesaid, compounds from family 1 (**Figure 19A**) has been previously tested on HCT 116 cell line, and the target predictions for this family seem to be mainly related to the same signaling pathway. This “homogenous” prediction makes the compounds belonging to this family the best candidates for docking studies.

4.2 Protein-compound interactions

The novel purine derivatives’ interactions with some of their predicted targets have been studied using Autodock 4 and Autodock Vina and the results obtained were compared with those for known inhibitors using the same protocol. The molecules used for these studies were:

- Cyclin-dependent kinase 2 (CDK2) and inhibitor **roscovitine (Figure 6A)** (PDB entry 2a4l)
- Protein tyrosine kinase HCK and inhibitor **1-ter-butyl-3-*p*-tolyl-1*H*-pyrazolo[3,4-*d*]pyrimidin-4-ylamine (Figure 5E)** (PDB entry 1qcf)
- Aurora kinase A (AURKA) and inhibitor **9-chloro-7-(2,6-difluorophenyl)-*N*{4-[(4-methylpiperazin-1-yl)carbonyl]phenyl}-5*H*-pyrimido[5,4-*d*][2]benzazepin-2-amine (Figure 7B)** (PDB entry 3h10)
- PKA and inhibitor **(2*S*)-1-(1*H*-indol-3-yl)-3-[5-(3-methyl-1*H*-indazol-5-yl)pyridin-3-yl]oxy-propan-2-amine (Figure 5D)** (PDB entry 2jds)
- AC and inhibitor **TDI10229³⁶ (Figure 5B)** (PDB entry 7ovd)
- EPAC and inhibitor **ESI-09³⁷ (Figure 5C)** (PDB entry 3cf6)
- VEGFR2 and inhibitor **tivozanib (Figure 5A)** (PDB entry 4ase)
- Adenosine receptor A2a and inhibitor **ZM241385 (Figure 2A)** (PDB entry 2ydv)

It is important to note that a more negative value of binding energy corresponds to a more favorable ligand-protein interaction. The smaller the inhibition constant (K_i), the greater the binding affinity and the smaller amount of ligand needed in order to inhibit the activity of that enzyme. Knowing this, the results of the docking simulations indicate that these compounds might be interacting strongly with the targets, validating the predicted results using the PASS server and therefore confirming that compounds belonging to this family are good candidates for further biological assays (**Tables 5 and 6**).

Table 5: Autodock Vina results.

Ligand	Lowest binding energy (Kcal/mol)				
	Known inhibitor	1	2	3	5
Receptor					
PKA	-10.0	-8.8	-10.3	-10.6	-9.1
AC	-10.6	-9.5	-10.1	-9.9	-9.3
EPAC	-8.3	-9.2	-7.0	-8.8	-9.0
CDK2	-7.5	-8.6	-10.0	-9.8	-9.1
AURKA	-10.0	-7.9	-9.0	-8.7	-8.4
HCK	-9.0	-8.9	-10.7	-9.1	-9.4
VEGFR	-10.3	-8.1	-9.1	-10.6	-7.6
A2a	-9.3	-8.5	-10.8	-11.2	-9.2

Table 6: Autodock 4 results.

		Lowest binding energy (Kcal/mol)				
Ligand \ Receptor	Known inhibitor	1	2	3	5	
PKA	-11.07	-7.6	-10.14	-10.09	-8.59	
AC	-11.34	-9.83	-10.82	-11.16	-10.07	
EPAC	-9.3	-7.97	-10.07	-7.82	-8.37	
CDK2	-7.87	-7.43	-8.62	-8.65	-7.88	
AURKA	-10.26	-7.25	-8.62	-7.92	-7.3	
HCK	-7.89	-8.18	-9.62	-9.22	-8.77	
VEGFR	-11.55	-9.08	-12.14	-10.62	-9.31	
A2a	-9.3	-8.2	-10.05	-9.59	-8.68	

Overall, most compounds tested have a similar binding energy compared to known inhibitors. In some cases, the binding energy is even more negative than the one predicted for the known inhibitor, which means they could have a greater inhibitory effect. Notably, the best performing molecule was compound 2, which complexed with VEGFR is able to form 3 hydrogen bonds in the binding pocket with the residues of valine, alanine and glutamine and interact hydrophobically with 11 other residues. These interactions combined, result in a noteworthy binding energy of -12.14 Kcal/mol (**Figure 20**).

Given these results, we can see that compounds 1 and 5 seem to present similar binding modes in each protein target (**Supplementary figure 1**). Generally, compound 5 presents a more negative binding energy, possibly due to the presence of the bulkier cycloalkyl chain in the N3 position of the purine ring, revealing the determining effects of Van der Waals interactions between ligand and protein. Compounds 2 and 3 seem to bind to the targets in dissimilar ways due to the different position of the biphenyl substituent, suggesting that its presence is important to the interactions in the binding site.

It is also possible to infer that the compounds have the potential to be exerting their effect by inhibiting kinase activity, specially VEGFR, which seems to be the targets with the best binding affinities towards the compounds in general. Kinases are enzymes that catalyze the transfer of phosphate groups from high energy molecules (such as ATP) to a substrate¹⁷⁵. If kinases are inhibited, a decrease in ATP dephosphorylation should be observed. Further enzymatic studies should be conducted to confirm the interaction between the purine derivatives studied and their possible targets and their effect of ATP dephosphorylation and consequent inorganic phosphate formation.

Additional figures displaying predicted interactions of ligands and proteins are shown in Appendix.

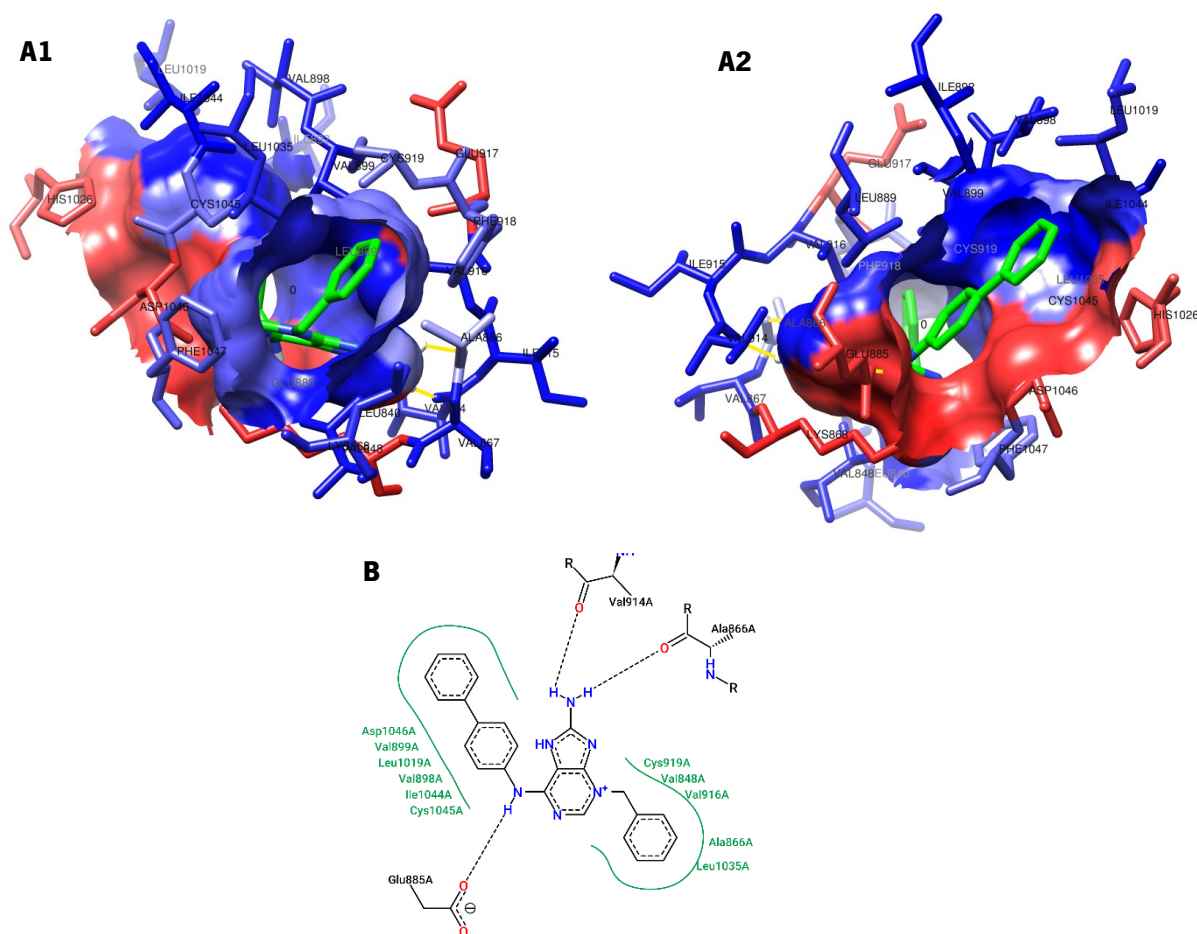


Figure 20: (A1 and A2) Less energetic docking pose predicted of the VEGFR–compound 2 complex in the active binding site. The surface of VEGFR is displayed in different tones from red to blue depending on residue lipophilicity (more lipophilic residues are shown in blue and less lipophilic residues are shown in red), whereas compound 3 is displayed as in green. Hydrogen bonds are displayed as yellow lines. (B) Predicted binding mode and interactions between compound 3 and the surrounding residues in the active cleft. Figures A1 and A2 were obtained using Chimera Version 1.16 (139) and figure B was obtained through the Protein-Plus Server(137,138).

4.3 2D *in vitro* assessments

Results have shown that the purine derivatives tested have a greater effect in Hep G2 hepatocarcinoma cell line than in HCT-15 colorectal cancer cell line after a 24- and 48-hour incubation period. In the hepatocarcinoma cell line in particular, it is possible to observe a reduction of the percentage of viable cells compared to the control without treatment as the compound concentration increases and establish a dose-response curve (**Figure 21C** and **21D**). Additionally, relatively to this cell line, it was also possible to determine IC₅₀ values (**Table 7**).

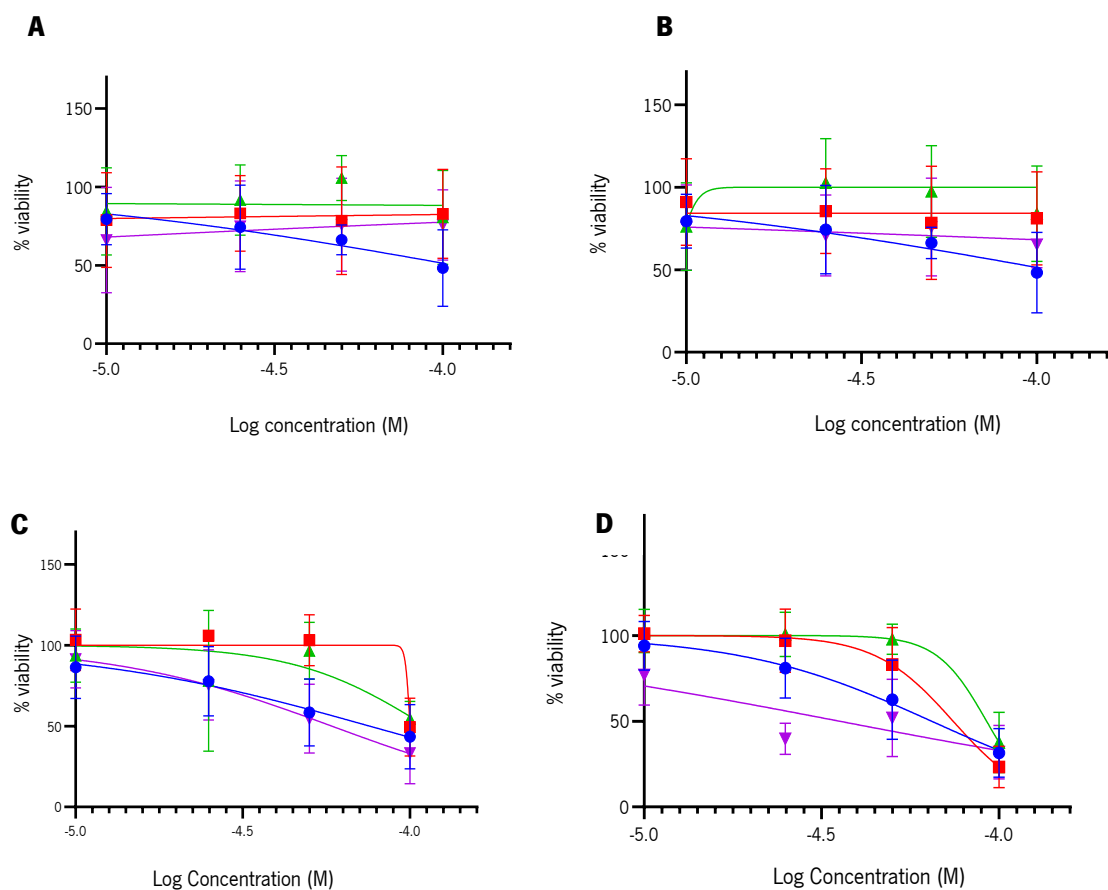


Figure 21: Effect on cell viability in HCT-15 and Hep G2 cell lines after 24h and 48h of incubation with purine derivatives: (A) HCT-15 24h incubation, (B) HCT-15 48h incubation, (C) Hep G2 24h incubation, and (D) Hep G2 48h incubation with compounds 1 (blue circle), 2 (red square), 3 (green upwards triangle) and 5 (purple downwards triangle).

Table 7: IC₅₀ for each compound tested in different cell lines (*R²<0,4).

Compound	IC ₅₀ (μM)			
	1	2	3	5
HCT-15 24h	*			
HCT-15 48h	*			
Hep G2 24h	75.63	99.96	*	58.30
Hep G2 48h	64.52	74.19	92.19	35.52

In both cell lines, it is possible to observe changes in the cells' normal morphology: naturally, HCT-15 cells appear to have a flat shape, while Hep G2 usually seem slightly more elongated. However, when exposed to both the compounds and 10% DMSO controls, the cells of both hepatocarcinoma and

colorectal cancer cell lines shrink and become more spherical in shape, suggesting cell death by apoptosis¹⁷⁶ (**Figure 22**).

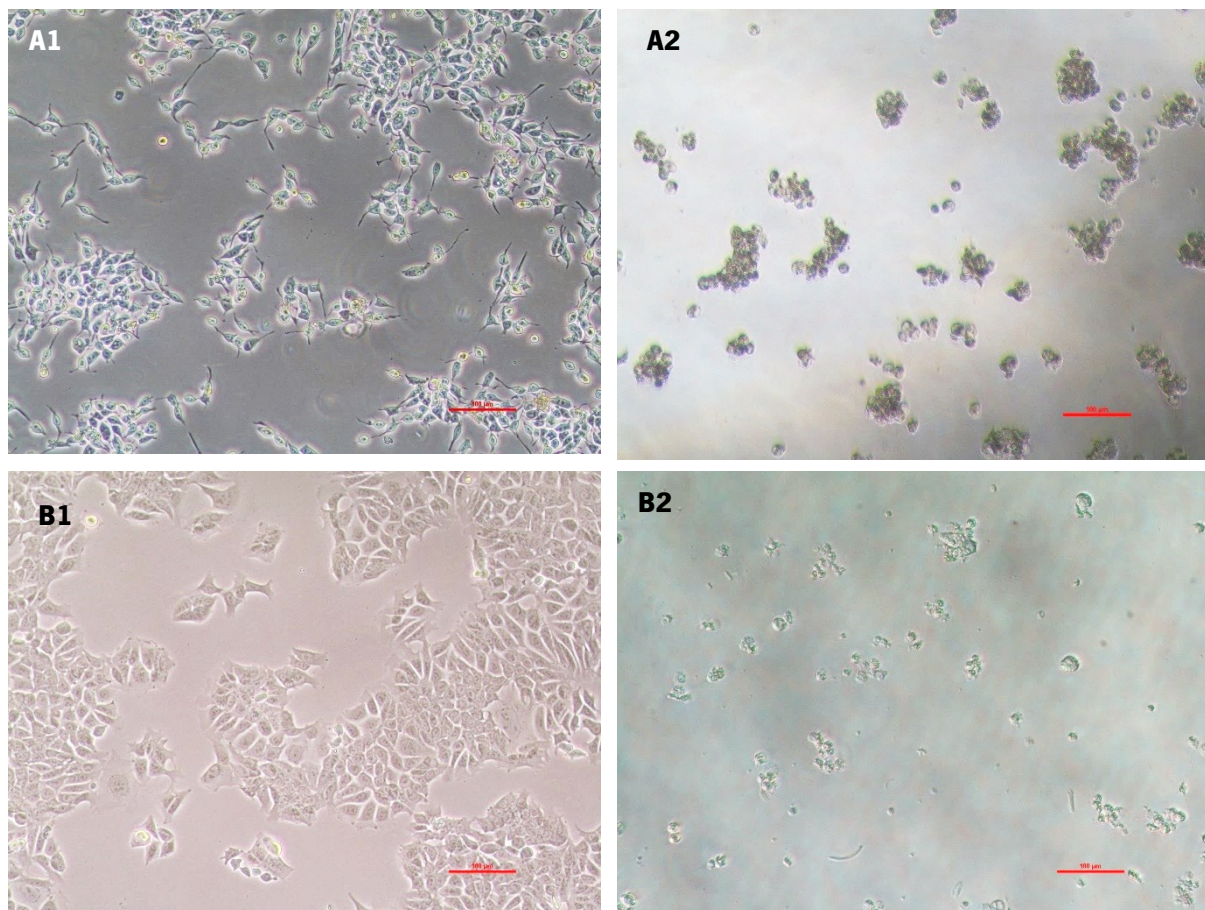


Figure 22: Changes in Hep G2 (A) and HCT-15 (B) cell morphology before (1) and after treatment (2) with 100 μM of compound 1 after 48h (B). In both cell lines, the cells become spherical in shape, suggesting apoptosis. The same tendency is observed for all compounds tested. Scale bars = 100 μm .

Although the compounds studied have shown an IC_{50} value between 30 and 100 μM , which profiles them as moderately active¹⁷⁷, it is important to note that using 2D cell models to assess the effect of newly developed drugs might not represent their full potential. By using a 3D approach using cell spheroids and organoids, it is possible to better mimic the tumor's microenvironment and cell conditions. In these models, it is possible to observe gradients of O_2 , CO_2 , nutrients, growth factors, cytokines, pH and waste products, which could affect the drug's performance. In moderately soluble drugs, like these novel molecules, having an increased pH could increase their solubility and consequent availability for cells, resulting in a better IC_{50} value. Further, as we can see by comparing the results in both cell lines, the type of cell line used has a great influence on the compound's IC_{50} . This means that it is of extreme

importance to keep testing the molecules' effect on different cell lines to fully take conclusions on their application as anticancer drugs, as a molecule could be effective on only certain types of cancers.

Additionally, by using drug delivery systems such as SLNs, it could be possible to improve the drugs efficiency. For this reason, the synthesis of deliverers should be performed, and the effect of the free drug should be compared with the effect of the encapsulated drug in different cancer models.

In addition to these studies, it is important to verify the occurrence of apoptosis and assess the phase of the cell cycle stop induced by these novel compounds. This can be achieved with flow cytometry and multimodal holographic microscopy¹⁷⁶ and can give further information on the action mechanism of this family of purine derivatives.

4.4 Insertion in drug delivery systems

Carnauba wax is produced in Brazil from leaves of the palm tree *Copernica cerifera* and it is considered as generally regarded as safe (GRAS) substance. In its composition there a predominance of various types of esters, free carboxylic acids, and fatty alcohols, which are mostly inert and stable components¹⁷⁸. Due to its physico-chemical characteristics, this wax is commonly used in food, usually in coating formulations for fruits and vegetables¹⁷⁹.

During SLN synthesis, it was observed that the Carnauba wax became solid after the compound addition, forming lumps of compound/wax aggregations upon mixing. The addition of hot Tween solution and water did not change the wax's state, and after sonication and cooling, it was still possible to observe solid wax lumps and flakes. The solidification of the wax could be occurring as a result of decrease in temperature. On a different note, it is possible that the nanoparticles are not observed in the DLS, or that the reaction did not take place as expected. This can be related to the establishment of strong intermolecular interactions between the wax and the compound, or it could mean that the compound reacts with some of the components of the wax, in particular the carboxylic acids.

Consequently, a different lipid core composition should be tried instead of carnauba wax. New forms of encapsulation might be also tested to perform encapsulation of these molecules.

4.5 P_i quantification

Results regarding the quantification of P_i in a protein sample after incubation with novel purine derivatives show that these molecules do not have a significant effect on P_i depletion or increase compared to controls (**Figure 23**).

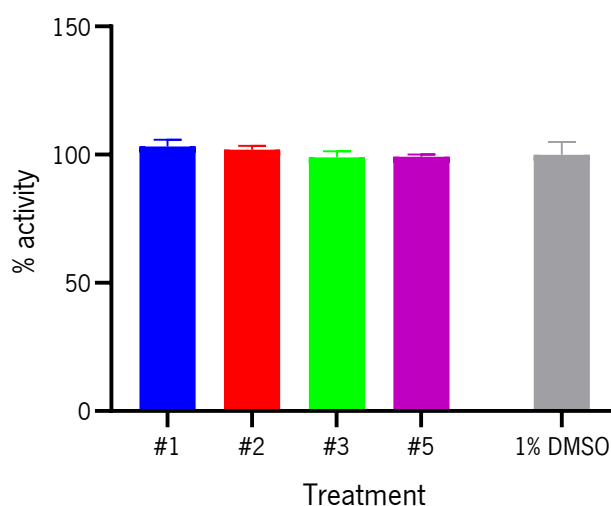


Figure 23: Percentage of ATPase activity in Hep G2 whole protein extracts after treatment with 100 μ M of test compounds and 1% DMSO relative to control, where the control with no treatment corresponds to 100% of activity.



These results lead to the conclusion that the compounds do not significantly inhibit enzymes that produce P_i, like, for example, proton pumps. However, it is important to note that in a whole cell extract, phosphatases are also present. Contrarily to kinases, phosphatases are enzymes that catalyze the hydrolysis of phospho-ester bonds in organic P-containing substrates releasing P_i¹⁸⁰. In that case, the hypothesis that the compounds tested could actually be inhibiting these phosphatases should not be ruled out. Additionally, it should be pointed out that the action of the predicted targets for the novel compounds, namely protein kinases, do not result in P_i formation but rather in its transfer to the next substrate on the signaling cascade. Therefore, it is only possible to conclude that the presence of the novel compounds at the concentration mentioned does not result in inhibition of P_i formation within the cell's proteome.

In addition to these studies, other enzymatic assays should be performed to assess the true effectiveness of the compounds on a specific protein. Further assessment on quantification of specific proteins and the confirmation of their activity within the protein extract should also be performed.

5 Concluding remarks and future perspectives

New anticancer drug approval rates are 5%¹⁸¹⁻¹⁸³, despite significant investments in cancer research, drug discovery and development¹⁸⁴. Most of the published data regarding known cell-based processes is derived from experiments performed in two-dimensional (2D) conditions where cells are grown as one single layer on a rigid surface. These conventional cell monolayer cultures, grown under simplified and unrealistic conditions, do not fully reflect the essential physiology of real tissues: they modify the tissue-specific architecture (forced polarity, flattened cell shape), mechanical/biochemical signals, and subsequent cell-to-cell communication¹⁸⁵. Despite these drawbacks, 2D cultures remain very attractive for laboratory purposes because of their simplicity and low cost¹⁸⁶. One strategy to improve the success rate of new anticancer drugs transitioning into the clinic would be to more closely align the cellular models used in the early lead discovery with pre-clinical animal models and patient tumors. For solid tumors, the development and implementation of three-dimensional (3D) *in vitro* tumor models that more accurately replicate human solid tumor architecture and biology could be the answer (**Table 8**). Recent advances in tissue engineering and regenerative medicine have provided new techniques for 3D spheroid generation and a variety of *in vitro* 3D cancer models are being explored for cancer drug discovery¹⁸⁴.

Table 8: Comparison between a monolayer (2D) tumor cell culture and a 3D tumor spheroid and main features¹⁸⁷.

Model	2D 	3D 
Cell interactions	Not able to reproduce cell-cell and cell-ECM interactions	Cell-cell and cell-ECM interactions preserved
Tumor architecture	Not able to replicate a 3D mass	Reproduction of a real tumor structure
Cell morphology and polarity	Changes cell morphology and polarity	Cell morphology and polarity is preserved
Molecular mechanisms	High differences in gene expression compared to <i>in vivo</i> tumors	Gene expression profile is similar to <i>in vivo</i> tumors
Gradients	Unlimited access to O ₂ , nutrients and metabolites	Gradients of O ₂ , CO ₂ , nutrients, growth factors, cytokines, pH and waste products
Biological zones	Unable to form proliferative, quiescent and necrotic zones.	Able to form proliferative, quiescent and necrotic zones.
Cost and complexity	Cheaper and less complex to perform and analyze	More expensive, complex analysis
Cell lines	Suitable for most cell lines	Most suitable for solid tumor cell lines

Tumor cells cultured in 3D microenvironments are exposed to a variety of biological stimuli that affect their responses and behaviors. The cell-cell interactions, cell-ECM interactions and local gradients of nutrients, growth factors, secreted factors and oxygen regulate cell function¹⁸⁸⁻¹⁹². This microenvironment alters numerous cellular and functional activities including morphology, signal transduction, histone acetylation, gene expression, protein expression, drug metabolism, differential zones of proliferation, viability, hypoxia, pH, differentiation, migration, and drug sensitivity^{188-190,192-200}. Additionally, cells in 3D cultures are exposed to different adhesive, topographical and mechanical forces than cells growing in 2D^{188,189,192,195,197,198}, and the cell-cell and cell-ECM interactions of cells in multi-layer tumor spheroids constitute a permeability barrier through which therapeutic agents must penetrate^{188-190,192-200} (**Figure 24**).

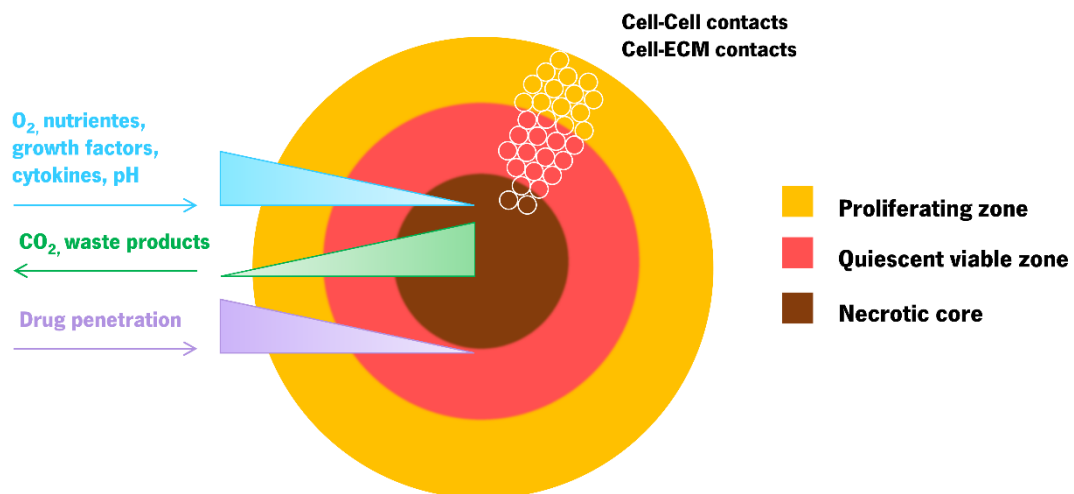


Figure 24: 3D tumor model microenvironment¹⁸⁴.

As previously mentioned, purine derivatives play a role in most biological processes in normal conditions. This means that they don't only have a potential to act as anticancer drugs, where the metabolism and proliferation are overactive, but also in the context of other complex diseases such as Parkinson's, Alzheimer's and inflammatory disease. For example, adenosine receptors, which might be one of the targets of action of novel purine derivatives according to results, modulate neuronal and synaptic function in a range of ways that may make them relevant to the occurrence, development and treatment of brain ischemic damage and degenerative disorders²⁰¹. In cardiac tissue, adenosine acts as an autacoid that plays a critical role in regulating cardiac function, including heart rate, contractility, and coronary flow²⁰². Adenosine's combined effects on neuronal viability and inflammatory processes have also prompted researchers to think about how they might contribute to the development of diseases like Lesch-Nyhan syndrome, Creutzfeldt-Jakob disease, Huntington's disease, multiple sclerosis, and brain

damage brought on by stroke. Adenosine receptors may have pathological significance, and serious efforts are being made to develop ligands that would target adenosine receptors as therapeutic agents to treat some of these conditions²⁰¹.

. The increasing popularity of gene-modified models with targeted deletion or overexpression of a single protein subtype has helped to elucidate the roles of each receptor subtype²⁰². Knowing this, and despite the only moderate effect demonstrated by the cell viability assays, it is of extreme importance to test these novel compounds not only in the context of other types of cancer, but also in neuronal and healthy tissues, micro-organisms, and viruses. By using gene-editing tools, it is possible to better understand the roles of each target and the effect on their inhibition. Further, using natural and genetically modified proteins, it is possible to better understand the effect of a compound on a specific enzyme.

Over the past three decades, significant advances have been made in drug delivery technology. Drug delivery technologies represent a vast, vital area of research and development in pharmaceuticals and the demand for innovative drug delivery systems continues to grow¹¹⁰. A drug delivery system is defined as a formulation or a device that allows the introduction of a therapeutic compound in the cell and improves its effectiveness and safety by controlling the rate, time and place of release¹¹¹. Depending on the characteristics of the active compounds to be administered, it may be possible to incorporate them in drug delivery systems of different nature (lipidic, polymeric, etc). After verifying whether the molecules can be encapsulated, the release and internalization of the therapeutic molecules in the cell can be evaluated and the cytotoxic effects of the free vs encapsulated drug can be compared. In this way, it is possible to improve the therapeutic efficacy of the molecules, not by optimizing the structure of the compounds themselves, but by improving their internalization and cell localization system. Further studies are needed in order to better understand the physicochemical properties of the compounds and to design a more appropriate technology and carrier that allows its incorporation into a delivery system. Even if the compound is hydrophobic, we hypothesize that some chemical interaction with some of the components of the carnauba wax could be hampering the formation of the SLNs. Other wax compositions or carriers of different nature (lipid or polymeric) could be explored.

In order to fully complete the protein-ligand interactions, simulations using molecular dynamics should also be performed. In this case, simulations are performed using even bigger systems with more realistic boundary conditions and better sampling due to longer sampling times. Recently, realistic simulations of systems as complex as transmembrane channels have become possible to perform²⁰³. Based on a general model of the physics driving interatomic interactions, molecular dynamics (MD)

simulations predict how each atom in a protein or other molecular system would move over time²⁰⁴. These simulations may capture a wide range of critical biomolecular processes, such as conformational change, ligand binding, and protein folding, and disclose the positions of all atoms with femtosecond temporal resolution. Importantly, such simulations may also anticipate how biomolecules will respond to perturbations like mutation, phosphorylation, protonation, or the addition or removal of a ligand at the atomic level²⁰⁵.

Finally, when it comes to using machine learning methods to assess the activity of novel or already known compounds for drug repurposing, there is a wide range of techniques and algorithms that could be implemented in order to develop a more accurate model. In the future, we aim to build a tool that is able to process much more information about anticancer agents in general (not just purine derivatives) and to correlate that information with the compounds' activity. Recently, due to the COVID-19 pandemic, researchers have been able to demonstrate that the integration of extensive interactions, deep neural networks, and multiple evidence can facilitate the rapid identification of candidate drugs for COVID-19 treatment²⁰⁶. In a different study, it was demonstrated that deep learning methods and the results are useful for the study of signatures and markers of drug response²⁰⁷. The main goal of using machine learning in the context of drug design is to build better, more effective, open-access tools and platforms that can be used for multiple purposes, ultimately leading to the implementations of novel therapeutic agents.

To summarize, and taking into account the polypharmacological profile of purine derivatives, it is of extreme relevance to continue with studies regarding these novel molecules, as we have demonstrated that they have the potential to be used not only as therapeutic agents in the context of cancer, but also as antivirals (as we've seen that, for example, EPAC inhibition induces cell resistance to viral infections³⁴), antimicrobial agents²⁰⁸, among others. Having a dual antimicrobial and anticancer effect is of special importance, as evidence shows that intratumoral bacteria are characteristic for each tumor type, resulting in a "microbial signature" for each cancer type. A dysbiotic microbiota predisposes the body to develop cancer by inducing genetic instability, initiating DNA damage and proliferation of the damaged progeny, eliciting favorable immune response, metabolic dysregulation and altered response to therapy. Designing a single agent that is able to act in both aspects, and on top of that using drug delivery systems, is one of the main goals for future assessments.

In the present work, we were able to develop an appropriate working methodology by coming into contact with several emerging technologies of interest in the field of drug development in order to confirm the potential of novel purine derivatives as therapeutic agents.

6 References

1. Breasted, J. H. & University of Chicago. Oriental Institute. The Edwin Smith surgical papyrus, published in facsimile and hieroglyphic transliteration with translation and commentary in two volumes. (1930).
2. Hippocrates & Adams, F. *The genuine works of Hippocrates*. (William Wood and Company, 1886).
3. Haggard, H. W. & Smith, G. M. Johannes Müller and the Modern Conception of Cancer. *Yale J. Biol. Med.* **10**, 419.b1-436 (1938).
4. Chakraborty, S. & Rahman, T. The difficulties in cancer treatment. *Ecancermedicalscience* **6**, ed16 (2012).
5. Ruddon, R. *Cancer Biology*. (Oxford University Press, 2007).
6. Knowles, M. A. & Selby, P. *Introduction to the cellular and molecular biology of cancer*. (Oxford University Press, 2005).
7. Sung, H. *et al.* Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA. Cancer J. Clin.* **0**, 1–41 (2021).
8. Valencia, D. N. Brief Review on COVID-19: The 2020 Pandemic Caused by SARS-CoV-2. *Cureus* **12**, (2020).
9. Corley, D. A. *et al.* Cancer Screening during COVID-19: A Perspective from NCI's PROSPR consortium. *Gastroenterology* **160**, 999–1002 (2021).
10. Sharpless, N. E. COVID-19 and cancer. *Science* **368**, 1290 (2020).
11. Kutikov, A. *et al.* A War on Two Fronts: Cancer Care in the Time of COVID-19. *Ann. Intern. Med.* **172**, 756–758 (2020).
12. Luqmani, Y. A. Mechanisms of Drug Resistance in Cancer Chemotherapy. *Med. Princ. Pract.* **14**, 35–48 (2005).
13. Widakowich, C., de Castro, G., de Azambuja, E., Dinh, P. & Awada, A. Review: side effects of approved molecular targeted therapies in solid cancers. *Oncologist* **12**, 1443–1455 (2007).
14. Sharma, P., Hu-Lieskovan, S., Wargo, J. A. & Ribas, A. Primary, Adaptive, and Acquired Resistance to Cancer Immunotherapy. *Cell* **168**, 707–723 (2017).
15. Jiang, P., Sellers, W. R. & Liu, X. S. Big data approaches for modeling response and resistance to cancer drugs. *Annu. Rev. Biomed. Data Sci.* **1**, 1–27 (2018).
16. Ban, F. *et al.* Best Practices of Computer-Aided Drug Discovery: Lessons Learned from the Development of a Preclinical Candidate for Prostate Cancer with a New Mechanism of Action. *J. Chem. Inf. Model.* **57**, 1018–1028 (2017).
17. Chapman, T. Drug discovery: the leading edge. *Nature* **430**, 109–115 (2004).
18. Graves, P. R. & Haystead, T. A. J. Molecular biologist's guide to proteomics. *Microbiol. Mol. Biol. Rev.* **66**, 39–63 (2002).
19. Legraverend, M. & Grierson, D. S. The purines: Potent and versatile small molecule inhibitors

- and modulators of key biological targets. **14**, 3987–4006 (2006).
20. Allard, B., Turcotte, M. & Stagg, J. CD73-generated adenosine: Orchestrating the tumor-stroma interplay to promote cancer growth. *J. Biomed. Biotechnol.* **2012**, (2012).
 21. Sitkovsky, M. V. T regulatory cells: hypoxia-adenosinergic suppression and re-direction of the immune response. *Trends Immunol.* **30**, 102–108 (2009).
 22. Allard, D., Allard, B., Gaudreau, P. O., Chrobak, P. & Stagg, J. CD73-adenosine: A next-generation target in immuno-oncology. *Immunotherapy* vol. 8 145–163 at <https://doi.org/10.2217/imt.15.106> (2016).
 23. Zimmermann, H. Prostatic acid phosphatase, a neglected ectonucleotidase. *Purinergic Signal.* **5**, 273–275 (2009).
 24. Thompson, E. A. & Powell, J. D. Inhibition of the Adenosine Pathway to Potentiate Cancer Immunotherapy: Potential for Combinatorial Approaches. *Annu. Rev. Med.* **72**, 331–348 (2021).
 25. Fredholm, B. B., IJzerman, A. P., Jacobson, K. A., Linden, J. & Müller, C. E. International union of basic and clinical pharmacology. LXXXI. Nomenclature and classification of adenosine receptors - An update. *Pharmacol. Rev.* **63**, 1–34 (2011).
 26. Doré, A. S. *et al.* Structure of the Adenosine A2A Receptor in Complex with ZM241385 and the Xanthines XAC and Caffeine. *Structure* **19**, 1283–1293 (2011).
 27. Caiazzo, E., Morello, S., Carnuccio, R., Ialenti, A. & Cicala, C. The Ecto-5'-Nucleotidase/CD73 Inhibitor, α,β -Methylene Adenosine 5'-Diphosphate, Exacerbates Carrageenan-Induced Pleurisy in Rat. *Front. Pharmacol.* **0**, 775 (2019).
 28. Regad, T. Targeting RTK Signaling Pathways in Cancer. *Cancers (Basel)*. **7**, 1758–1784 (2015).
 29. Ségaliny, A. I., Tellez-Gabriel, M., Heymann, M. F. & Heymann, D. Receptor tyrosine kinases: Characterisation, mechanism of action and therapeutic interests for bone cancers. *J. Bone Oncol.* **4**, 1–12 (2015).
 30. Peltier, J., O'Neill, A. & Schaffer, D. V. PI3K/Akt and CREB regulate adult neural hippocampal progenitor proliferation and differentiation. *Dev. Neurobiol.* **67**, 1348–1361 (2007).
 31. Fruman, D. A. & Rommel, C. PI3K and cancer: Lessons, challenges and opportunities. *Nat. Rev. Drug Discov.* **13**, 140–156 (2014).
 32. Knight, Z. A., Lin, H. & Shokat, K. M. Targeting the cancer kinome through polypharmacology. *Nat. Rev. Cancer* **10**, 130–137 (2010).
 33. Apsel, B. *et al.* Targeted polypharmacology: Discovery of dual inhibitors of tyrosine and phosphoinositide kinases. *Nat. Chem. Biol.* **4**, 691–699 (2008).
 34. Tao, X. *et al.* Blocking of Exchange Proteins Directly Activated by cAMP Leads to Reduced Replication of Middle East Respiratory Syndrome Coronavirus. *J. Virol.* **88**, 3902–3910 (2014).
 35. McTigue, M. *et al.* Molecular conformations, interactions, and properties associated with drug efficiency and clinical performance among VEGFR TK inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 18281–18289 (2012).

36. Fushimi, M. *et al.* Discovery of TDI-10229: A Potent and Orally Bioavailable Inhibitor of Soluble Adenylyl Cyclase (sAC, ADCY10). *ACS Med. Chem. Lett.* **12**, 1283–1287 (2021).
37. Wang, X., Luo, C., Cheng, X. & Lu, M. Lithium and an EPAC-specific inhibitor ESI-09 synergistically suppress pancreatic cancer cell proliferation and survival. *Acta Biochim. Biophys. Sin. (Shanghai)*. **49**, 573–580 (2017).
38. Zhu, Y. *et al.* Biochemical and Pharmacological Characterizations of ESI-09 Based EPAC Inhibitors: Defining the ESI-09 “Therapeutic Window”. *Sci. Reports 2015 51* **5**, 1–8 (2015).
39. Davies, T. G. *et al.* Structure-based design of a potent purine-based cyclin-dependent kinase inhibitor. *Nat. Struct. Biol.* **9**, 745–749 (2002).
40. Schindler, T. *et al.* Crystal structure of Hck in complex with a Src family-selective tyrosine kinase inhibitor. *Mol. Cell* **3**, 639–648 (1999).
41. Surget, S., Khoury, M. P. & Bourdon, J. C. Uncovering the role of p53 splice variants in human malignancy: a clinical perspective. *Onco. Targets. Ther.* **7**, 57 (2014).
42. Matlashewski, G. *et al.* Isolation and characterization of a human p53 cDNA clone: expression of the human p53 gene. *EMBO J.* **3**, 3257–3262 (1984).
43. Isobe, M., Emanuel, B. S., Givol, D., Oren, M. & Croce, C. M. Localization of gene for human p53 tumour antigen to band 17p13. *Nat. 1986 3206057* **320**, 84–85 (1986).
44. Kern, S. E. *et al.* Identification of p53 as a Sequence-Specific DNA-Binding Protein. *Science (80-)*. **252**, 1708–1711 (1991).
45. McBride, O. W., Merry, D. & Givol, D. The gene for human p53 cellular tumor antigen is located on chromosome 17 short arm (17p13). *Proc. Natl. Acad. Sci.* **83**, 130–134 (1986).
46. Bourdon, J. C. *et al.* p53 isoforms can regulate p53 transcriptional activity. *Genes Dev.* **19**, 2122–2137 (2005).
47. Abraham, A. G. & O’Neill, E. PI3K/Akt-mediated regulation of p53 in cancer. *Biochem. Soc. Trans.* **42**, 798–803 (2014).
48. Malumbres, M. & Barbacid, M. Mammalian cyclin-dependent kinases. *Trends Biochem. Sci.* **30**, 630–641 (2005).
49. Hochegger, H., Takeda, S. & Hunt, T. Cyclin-dependent kinases and cell-cycle transitions: does one fit all? *Nat. Rev. Mol. Cell Biol.* *2008 911* **9**, 910–916 (2008).
50. Malumbres, M. Cyclin-dependent kinases. *Genome Biol.* **15**, 122 (2014).
51. Heptinstall, A. B., Adiyasa, I., Cano, C. & Hardcastle, I. R. Recent advances in CDK inhibitors for cancer therapy. *Future Med. Chem.* **10**, 1369–1388 (2018).
52. Bettayeb, K. *et al.* CR8, a potent and selective, roscovitine-derived inhibitor of cyclin-dependent kinases. *Oncogene* **27**, 5797–5807 (2008).
53. Bukanov, N. O. *et al.* CDK inhibitors R-roscovitine and S-CR8 effectively block renal and hepatic cystogenesis in an orthologous model of ADPKD. *Cell Cycle* **11**, 4040–4046 (2012).
54. Kryštof, V., Lenobel, R., Havlíček, L., Kuzma, M. & Strnad, M. Synthesis and biological activity of Olomoucine II. *Bioorganic Med. Chem. Lett.* **12**, 3283–3286 (2002).

55. Andrews, P. D. Aurora kinases: Shining lights on the therapeutic horizon? *Oncogene* **24**, 5005–5015 (2005).
56. Jackson, J. R., Patrick, D. R., Dar, M. M. & Huang, P. S. Targeted anti-mitotic therapies: Can we improve on tubulin agents? *Nat. Rev. Cancer* **7**, 107–117 (2007).
57. Meraldi, P., Honda, R. & Nigg, E. A. Aurora kinases link chromosome segregation and cell division to cancer susceptibility. *Curr. Opin. Genet. Dev.* **14**, 29–36 (2004).
58. Marumoto, T., Zhang, D. & Saya, H. Aurora-A - A guardian of poles. *Nat. Rev. Cancer* **5**, 42–50 (2005).
59. Giet, R., Petretti, C. & Prigent, C. Aurora kinases, aneuploidy and cancer, a coincidence or a real link? *Trends Cell Biol.* **15**, 241–250 (2005).
60. Bolanos-Garcia, V. M. Aurora kinases. *Int. J. Biochem. Cell Biol.* **37**, 1572–1577 (2005).
61. Kaestner, P., Stolz, A. & Bastians, H. Determinants for the efficiency of anticancer drugs targeting either Aurora-A or Aurora-B kinases in human colon carcinoma cells. *Mol. Cancer Ther.* **8**, 2046–2056 (2009).
62. Bavetsias, V. & Linardopoulos, S. Aurora kinase inhibitors: Current status and outlook. *Front. Oncol.* **5**, 278 (2015).
63. Khalifa, M. E. Design, synthesis and molecular docking study of new purine derivatives as Aurora kinase inhibitors. *J. Mol. Struct.* **1229**, 129843 (2021).
64. Bavetsias, V. *et al.* Optimization of Imidazo[4,5-b]pyridine-Based Kinase Inhibitors: Identification of a Dual FLT3/Aurora Kinase Inhibitor as an Orally Bioavailable Preclinical Development Candidate for the Treatment of Acute Myeloid Leukemia kinase (K_d = 6.2 nM), and FLT3 m. **55**, 8721–8734 (2012).
65. Aliagas-Martin, I. *et al.* A class of 2,4-bisanilinopyrimidine Aurora A inhibitors with unusually high selectivity against Aurora B. *J. Med. Chem.* **52**, 3300–3307 (2009).
66. Shah, K. N. *et al.* Aurora kinase A drives the evolution of resistance to third-generation EGFR inhibitors in lung cancer. *Nat. Med.* **25**, 111–118 (2019).
67. De Berardinis, R. J. & Chandel, N. S. Fundamentals of cancer metabolism. *Sci. Adv.* **2**, (2016).
68. Hanahan, D. & Weinberg, R. A. Hallmarks of Cancer: The Next Generation. *Cell* **144**, 646–674 (2011).
69. Yin, J. *et al.* Potential Mechanisms Connecting Purine Metabolism and Cancer Therapy. *Front. Immunol.* **9**, 1697 (2018).
70. Peters, G. J. Novel Developments in the Use of Antimetabolites. *Nucleosides, Nucleotides and Nucleic Acids* **33**, 358–374 (2014).
71. Tiwari, M. Antimetabolites: Established cancer therapy. *J. Cancer Res. Ther.* (2012).
72. Brox, L. W., Birkett, L. & Belch, A. Clinical pharmacology of oral thioguanine in acute myelogenous leukemia. *Cancer Chemother. Pharmacol.* **6**, 35–38 (1981).
73. Karran, P. & Attard, N. Thiopurines in current medical practice: Molecular mechanisms and contributions to therapy-related cancer. *Nat. Rev. Cancer* **8**, 24–36 (2008).

74. Bennett, L. L. *et al.* Activity and mechanism of action of 6-methylthiopurine ribonucleoside in cancer cells resistant to 6-mercaptopurine. *Nature* **205**, 1276–1279 (1965).
75. Huang, Y. *et al.* Voltammetric studies of the interaction of 6-mercaptopurine with cucurbit[7]uril and DNA. *J. Incl. Phenom. Macrocycl. Chem.* **69**, 131–137 (2011).
76. Pedley, A. M. & Benkovic, S. J. A New View into the Regulation of Purine Metabolism: The Purinosome. *Trends Biochem. Sci.* **42**, 141–154 (2017).
77. Chan, C. Y. *et al.* Purinosome formation as a function of the cell cycle. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 1368–1373 (2015).
78. An, S., Kumar, R., Sheets, E. D. & Benkovic, S. J. Reversible compartmentalization of de novo purine biosynthetic complexes in living cells. *Science* **320**, 103–106 (2008).
79. Goswami, M. T. *et al.* Role and regulation of coordinately expressed de novo purine biosynthetic enzymes PPAT and PAICS in lung cancer. *Oncotarget* **6**, 23445–23461 (2015).
80. Baresova, V. *et al.* Mutations of ATIC and ADSL affect purinosome assembly in cultured skin fibroblasts from patients with AICA-ribosiduria and ADSL deficiency. *Hum. Mol. Genet.* **21**, 1534–1543 (2012).
81. French, J. B., Fang, Y. & Benkovic, S. J. The purinosome, a multi-protein complex involved in the de novo biosynthesis of purines in humans. *Chem. Commun.* **49**, 4444–4452 (2013).
82. Microtubule-directed transport of purine metabolons drives their cytosolic transit to mitochondria *CELL BIOLOGY.* **18**, 13009–13014 (2018).
83. French, J. B. *et al.* Hsp70/Hsp90 chaperone machinery is involved in the assembly of the purinosome. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 2528–2533 (2013).
84. Baresova, V. *et al.* Study of purinosome assembly in cell-based model systems with de novo purine synthesis and salvage pathway deficiencies. *PLoS One* **13**, e0201432 (2018).
85. Kola, I. & Landis, J. Can the pharmaceutical industry reduce attrition rates? *Nat. Rev. Drug Discov.* **3**, 711–715 (2004).
86. Drug development: the journey of a medicine from lab to shelf - The Pharmaceutical Journal. <https://pharmaceutical-journal.com/article/feature/drug-development-the-journey-of-a-medicine-from-lab-to-shelf>.
87. Fischer, T., Gazzola, S. & Riedl, R. Approaching Target Selectivity by De Novo Drug Design. *Expert Opin. Drug Discov.* **14**, 791–803 (2019).
88. Mouchlis, V. D., Melagraki, G., Zacharia, L. C. & Afantitis, A. Computer-Aided Drug Design of β -Secretase, γ -Secretase and Anti-Tau Inhibitors for the Discovery of Novel Alzheimer's Therapeutics. *Int. J. Mol. Sci.* **21**, (2020).
89. Martinez-Mayorga, K., Madariaga-Mazon, A., Medina-Franco, J. L. & Maggiora, G. The impact of chemoinformatics on drug discovery in the pharmaceutical industry. <https://doi.org/10.1080/17460441.2020.1696307> **15**, 293–306 (2020).
90. Talele, T., Khedkar, S. & Rigby, A. Successful Applications of Computer Aided Drug Discovery: Moving Drugs from Concept to the Clinic. *Curr. Top. Med. Chem.* **10**, 127–141 (2010).

91. Leelananda, S. P. & Lindert, S. Computational methods in drug discovery. *Beilstein J. Org. Chem.* **12**, 2694–2718 (2016).
92. Prakash, N. & Gareja, D. A. Cheminformatics. *J. Proteomics Bioinforma.* **3**, 249–252 (2010).
93. Nantasenamat, C. Best Practices for Constructing Reproducible QSAR Models. *Methods Pharmacol. Toxicol.* 55–75 (2020).
94. Nantasenamat, C., Isarankura-Na-Ayudhya, C. & Prachayasittikul, V. Advances in computational methods to predict the biological activity of compounds. *Expert Opin. Drug Discov.* **5**, 633–654 (2010).
95. Nantasenamat, C., Isarankura-Na-Ayudhya, C., Naenna, T. & Prachayasittikul, V. A practical overview of quantitative structure-activity relationship. *EXCLI J.* **8**, 74–88 (2009).
96. Guha, R. & Van Drie, J. H. Structure - Activity landscape index: Identifying and quantifying activity cliffs. *J. Chem. Inf. Model.* **48**, 646–658 (2008).
97. Sisay, M. T., Peltason, L. & Bajorath, J. Structural Interpretation of Activity Cliffs Revealed by Systematic Analysis of Structure–Activity Relationships in Analog Series. *J. Chem. Inf. Model.* **49**, 2179–2189 (2009).
98. El Naqa, I. & Murphy, M. J. What Is Machine Learning? *Mach. Learn. Radiat. Oncol.* 3–11 (2015).
99. Morris, G. M. & Lim-Wilby, M. Molecular Docking. *Methods Mol. Biol.* **443**, 365–382 (2008).
100. Dias, R. & de Azevedo Jr., W. Molecular Docking Algorithms. *Curr. Drug Targets* **9**, 1040–1047 (2008).
101. Agarwal, S. & Mehrotra, R. An overview of Molecular Docking. *JSM Chem* **4**, 1024 (2016).
102. Reddy, A. S. & Zhang, S. Polypharmacology: drug discovery for the future. *Expert Rev. Clin. Pharmacol.* **6**, 41–47 (2013).
103. Proschak, E., Stark, H. & Merk, D. Polypharmacology by Design: A Medicinal Chemist's Perspective on Multitargeting Compounds. *J. Med. Chem.* **62**, 420–444 (2019).
104. Hu, Y. & Bajorath, J. Compound promiscuity: What can we learn from current data? *Drug Discov. Today* **18**, 644–650 (2013).
105. Jemaà, M. *et al.* Reversine inhibits Colon Carcinoma Cell Migration by Targeting JNK1. *Sci. Rep.* **8**, 11821 (2018).
106. Vicente-Manzanares, M., Ma, X., Adelstein, R. S. & Horwitz, A. R. Non-muscle myosin II takes centre stage in cell adhesion and migration. *Nat. Rev. Mol. Cell Biol.* **10**, 778–790 (2009).
107. D'Alise, A. M. *et al.* Reversine, a novel Aurora kinases inhibitor, inhibits colony formation of human acute myeloid leukemia cells. *Mol. Cancer Ther.* **7**, 1140–1149 (2008).
108. Altieri, D. C. Mitochondrial Hsp90 chaperones as novel molecular targets in prostate cancer. *Futur. Oncol.* **6**, 487–489 (2010).
109. Anighoro, A., Bajorath, J. & Rastelli, G. Polypharmacology: Challenges and Opportunities in Drug Discovery. *J. Med. Chem.* **57**, 7874–7887 (2014).

110. Ranade, V. & Cannon, J. *Drug Delivery Systems*. (CRC Press, 2011).
111. Gupta, R. & Rai, B. Computer-Aided Design of Nanoparticles for Transdermal Drug Delivery. in *Methods in Molecular Biology* vol. 2059 225–237 (Humana Press, 2020).
112. Kumar, M., Kakkar, V., Mishra, A. K., Chuttani, K. & Kaur, I. P. Intranasal delivery of streptomycin sulfate (STRS) loaded solid lipid nanoparticles to brain and blood. *Int. J. Pharm.* **461**, 223–233 (2014).
113. Wissing, S. A., Kayser, O. & Müller, R. H. Solid lipid nanoparticles for parenteral drug delivery. *Adv. Drug Deliv. Rev.* **56**, 1257–1272 (2004).
114. Gao, P. *et al.* Designing multifunctional cancer-targeted nanosystem for magnetic resonance molecular imaging-guided theranostics of lung cancer. *Drug Deliv.* **25**, 1811 (2018).
115. Katouzian, I., Faridi Esfanjani, A., Jafari, S. M. & Akhavan, S. Formulation and application of a new generation of lipid nano-carriers for the food bioactive ingredients. *Trends Food Sci. Technol.* **68**, 14–25 (2017).
116. Din, F. U. *et al.* Effective use of nanocarriers as drug delivery systems for the treatment of selected tumors. *Int. J. Nanomedicine* **12**, 7291–7309 (2017).
117. Tapeinos, C., Battaglini, M. & Ciofani, G. Advances in the design of solid lipid nanoparticles and nanostructured lipid carriers for targeting brain diseases. *J. Control. Release* **264**, 306–332 (2017).
118. Brazel, C. S. & Huang, X. The Cost of Optimal Drug Delivery: Reducing and Preventing the Burst Effect in Matrix Systems. in *American Chemical Society* 267–282 (2004).
119. Bayón-Cordero, L., Alkorta, I. & Arana, L. Application of Solid Lipid Nanoparticles to Improve the Efficiency of Anticancer Drugs. *Nanomater. (Basel, Switzerland)* **9**, (2019).
120. Bertrand, N., Wu, J., Xu, X., Kamaly, N. & Farokhzad, O. C. Cancer Nanotechnology: The impact of passive and active targeting in the era of modern cancer biology. *Adv. Drug Deliv. Rev.* **66**, 2 (2014).
121. Cerqueira, M., Belmonte-Reche, E., Gallo, J., Baltazar, F. & Bañobre-López, M. Advanced (Magnetic) Theranostic Nanostructures Lab. *Pharmaceutics* **14**, (2022).
122. Stylianopoulos, T. EPR-effect: utilizing size-dependent nanoparticle delivery to solid tumors. *Ther. Deliv.* **4**, 421–423 (2013).
123. Wu, J. The Enhanced Permeability and Retention (EPR) Effect: The Significance of the Concept and Methods to Enhance Its Application. *J. Pers. Med.* **11**, (2021).
124. Fouad, Y. A. & Aanei, C. Revisiting the hallmarks of cancer. *Am. J. Cancer Res.* **7**, 1016 (2017).
125. Greish, K. Enhanced permeability and retention (EPR) effect for anticancer nanomedicine drug targeting. *Methods Mol. Biol.* **624**, 25–37 (2010).
126. Sun, T. *et al.* Engineered nanoparticles for drug delivery in cancer therapy. *Angew. Chem. Int. Ed. Engl.* **53**, 12320–12364 (2014).
127. Golombek, S. K. *et al.* Tumor targeting via EPR: Strategies to enhance patient responses. *Adv. Drug Deliv. Rev.* **130**, 17–38 (2018).

128. Natfji, A. A., Ravishankar, D., Osborn, H. M. I. & Greco, F. Parameters Affecting the Enhanced Permeability and Retention Effect: The Need for Patient Selection. *J. Pharm. Sci.* **106**, 3179–3187 (2017).
129. Senhorães, N. & Dias, A. Síntese de derivados de adenina e guanina. (PhD in Sciences, Specialization in Chemistry, University of Minho, 2015).
130. Sousa, A., Pereira-Wilson, C. & Dias, A. Evaluation of the anticancer potential of novel purine derivatives in human colon cancer. (MSc in Applied Biochemistry, University of Minho, 2014).
131. Daina, A., Michielin, O. & Zoete, V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. Reports 2017 7* **7**, 1–13 (2017).
132. Lipinski, C. A. Lead- and drug-like compounds: the rule-of-five revolution. *Drug Discov. Today Technol.* **1**, 337–341 (2004).
133. Ghose, A. K., Viswanadhan, V. N. & Wendoloski, J. J. A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. A qualitative and quantitative characterization of known drug databases. *J. Comb. Chem.* **1**, 55–68 (1999).
134. Veber, D. F. *et al.* Molecular properties that influence the oral bioavailability of drug candidates. *J. Med. Chem.* **45**, 2615–2623 (2002).
135. Egan, W. J., Merz, K. M. & Baldwin, J. J. Prediction of drug absorption using multivariate statistics. *J. Med. Chem.* **43**, 3867–3877 (2000).
136. Muegge, I., Heald, S. L. & Brittelli, D. Simple selection criteria for drug-like chemical matter. *J. Med. Chem.* **44**, 1841–1846 (2001).
137. Hofmarcher, T., Lindgren, P., Wilking, N. & Jönsson, B. The cost of cancer in Europe 2018. *Eur. J. Cancer* **129**, 41–49 (2020).
138. European Parliament. *Background note on cancer research.* (2020).
139. Wouters, O. J., McKee, M. & Luyten, J. Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009-2018. *JAMA* **323**, 844–853 (2020).
140. Van Norman, G. A. Drugs, Devices, and the FDA: Part 1: An Overview of Approval Processes for Drugs. *JACC Basic to Transl. Sci.* **1**, 170–179 (2016).
141. Basu, K., Sinha, R., Ong, A. & Basu, T. Artificial Intelligence: How is It Changing Medical Sciences and Its Future? *Indian J. Dermatol.* **65**, 365 (2020).
142. Veljkovic, V., Veljkovic, N., Este, J., Huther, A. & Dietrich, U. Application Of The EIIP/ISM Bioinformatics Concept in Development of New Drugs. *Curr. Med. Chem.* **14**, 441–453 (2007).
143. Filimonov, D. A. *et al.* Prediction of the biological activity spectra of organic compounds using the pass online web resource. *Chem. Heterocycl. Compd.* **50**, 444–457 (2014).
144. Mysinger, M. M., Carchia, M., Irwin, J. J. & Shoichet, B. K. Directory of useful decoys, enhanced (DUD-E): Better ligands and decoys for better benchmarking. *J. Med. Chem.* **55**, 6582–6594 (2012).
145. Rogers, D. & Hahn, M. Extended-connectivity fingerprints. *J. Chem. Inf. Model.* **50**, 742–754

- (2010).
146. Durant, J. L., Leland, B. A., Henry, D. R. & Nourse, J. G. Reoptimization of MDL keys for use in drug discovery. *J. Chem. Inf. Comput. Sci.* **42**, 1273–1280 (2002).
 147. Awale, M. & Reymond, J. L. Atom pair 2D-fingerprints perceive 3D-molecular shape and pharmacophores for very fast virtual screening of ZINC and GDB-17. *J. Chem. Inf. Model.* **54**, 1892–1907 (2014).
 148. Chawla, N. V., Bowyer, K. W., Hall, L. O. & Kegelmeyer, W. P. SMOTE: Synthetic Minority Over-sampling Technique. *J. Artif. Intell. Res.* **16**, 321–357 (2011).
 149. Chen, T. & Guestrin, C. XGBoost: A Scalable Tree Boosting System. in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2016).
 150. Su, X., Yan, X. & Tsai, C. L. Linear regression. *Wiley Interdiscip. Rev. Comput. Stat.* **4**, 275–294 (2012).
 151. Kingsford, C. & Salzberg, S. L. What are decision trees? *Nat. Biotechnol.* *2008 269* **26**, 1011–1013 (2008).
 152. Cutler, A., Cutler, D. R. & Stevens, J. R. Random Forests. *Ensemble Mach. Learn.* 157–175 (2012).
 153. Ontivero-Ortega, M., Lage-Castellanos, A., Valente, G., Goebel, R. & Valdes-Sosa, M. Fast Gaussian Naïve Bayes for searchlight classification analysis. *Neuroimage* **163**, 471–479 (2017).
 154. Oneto, L. The ‘K’ in K-fold Cross Validation. in *Proceedings of the 20th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning* (eds. G. Balint et al.) 441–446 (BELGIO, 2012).
 155. Liashchynskiy, P. & Liashchynskiy, P. Grid Search, Random Search, Genetic Algorithm: A Big Comparison for NAS. (2019).
 156. Vieira, T. F. & Sousa, S. F. Comparing AutoDock and Vina in Ligand/Decoy Discrimination for Virtual Screening. *Appl. Sci.* (2019).
 157. Schrödinger, L. *The {PyMOL} Molecular Graphics System, Version ~2.* (2015).
 158. Avogadro - Free cross-platform molecular editor - Avogadro. <https://avogadro.cc/>.
 159. Hanwell, M. D. *et al.* Avogadro: An advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminform.* **4**, 1–17 (2012).
 160. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
 161. RCSB PDB: Homepage. <https://www.rcsb.org/>.
 162. Morris, G. M. *et al.* Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function. *J. Comput. Chem.* **19**, 1639–1662 (1998).
 163. Morris, G. M. *et al.* User Guide AutoDock Version 4.2 Updated for version 4.2.6 Automated Docking of Flexible Ligands to Flexible Receptors. at <http://autodock.scripps.edu/> (1991).
 164. Schöning-Stierand, K. *et al.* ProteinsPlus: interactive analysis of protein–ligand binding

- interfaces. *Nucleic Acids Res.* **48**, W48 (2020).
165. Zentrum für Bioinformatik: Universität Hamburg - Proteins Plus Server. <https://proteins.plus/>.
 166. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
 167. Borra, R. C., Lotufo, M. A., Gaglioti, S. M., Barros, F. de M. & Andrade, P. M. A simple method to measure cell viability in proliferation and cytotoxicity assays. *Braz. Oral Res.* **23**, 255–262 (2009).
 168. Riss, T. L. *et al.* Cell Viability Assays. *Assay Guid. Man.* (2016).
 169. Kruger, N. J. The Bradford Method For Protein Quantitation. 17–24 (2009).
 170. Dan, S. K. & Ray, A. K. Characterization and identification of phytase-producing bacteria isolated from the gastrointestinal tract of four freshwater teleosts. *Ann. Microbiol.* **64**, 297–306 (2014).
 171. Djafarzadeh, S., Vuda, M., Jeger, V., Takala, J. & Jakob, S. M. The Effects of Fentanyl on Hepatic Mitochondrial Function. *Anesth. Analg.* **123**, 311–325 (2016).
 172. Hassan, P. A., Rana, S. & Verma, G. Making sense of Brownian motion: Colloid characterization by dynamic light scattering. *Langmuir* **31**, 3–12 (2015).
 173. Clogston, J. D. & Patri, A. K. Zeta potential measurement. *Methods Mol. Biol.* **697**, 63–70 (2011).
 174. Syue, M. R., Wei, F. J., Chou, C. S. & Fu, C. M. Magnetic and electrical properties of Mn-Zn ferrites synthesized by combustion method without subsequent heat treatments. *J. Appl. Phys.* **109**, (2011).
 175. Fabbro, D., Cowan-Jacob, S. W. & Moebitz, H. Ten things you should know about protein kinases: IUPHAR Review 14. *Br. J. Pharmacol.* **172**, 2675–2700 (2015).
 176. Balvan, J. *et al.* Multimodal holographic microscopy: Distinction between apoptosis and oncosis. *PLoS One* **10**, (2015).
 177. Indrayanto, G., Putra, G. S. & Suhud, F. Validation of in-vitro bioassay methods: Application in herbal drug research. *Profiles Drug Subst. Excipients Relat. Methodol.* **46**, 273–307 (2021).
 178. Vandenburg, L. E. & Wilder, E. A. The structural constituents of carnauba wax. *J. Am. Oil Chem. Soc.* 1970 4712 **47**, 514–518 (1970).
 179. Weller, C. L., Gennadios, A. & Saraiva, R. A. Edible bilayer films from zein and grain sorghum wax or carnauba wax. *LWT* **31**, 279–285 (1998).
 180. Liberti, S. *et al.* HuPho: the human phosphatase portal. *FEBS J.* **280**, 379–387 (2013).
 181. Ocana, A., Pandiella, A., Siu, L. L. & Tannock, I. F. Preclinical development of molecular-targeted agents for cancer. *Nat. Rev. Clin. Oncol.* **8**, 200–209 (2010).
 182. Hait, W. N. Anticancer drug development: the grand challenges. *Nat. Rev. Drug Discov.* **9**, 253–254 (2010).
 183. Hutchinson, L. & Kirk, R. High drug attrition rates—where are we going wrong? *Nat. Rev. Clin. Oncol.* 2011 84 **8**, 189–190 (2011).

184. Sant, S. & Johnston, P. A. The production of 3D tumor spheroids for cancer drug discovery. *Drug Discov. Today Technol.* **23**, 27–36 (2017).
185. Yamada, K. M. & Cukierman, E. Modeling Tissue Morphogenesis and Cancer in 3D. *Cell* **130**, 601–610 (2007).
186. Hoarau-Véchet, J., Rafii, A., Touboul, C. & Pasquier, J. Halfway between 2D and animal models: Are 3D cultures the ideal tool to study cancer-microenvironment interactions? *Int. J. Mol. Sci.* **19**, (2018).
187. Pires Celeiro, S., Baltazar, F. & Viana-Pereira, M. 3D culture systems as models for solid tumors and cancer metabolism. in *Biomaterials for 3D Tumor Modeling* 131–155 (2020).
188. Baker, B. M. & Chen, C. S. Deconstructing the third dimension: how 3D culture microenvironments alter cellular cues. *J. Cell Sci.* **125**, 3015–3024 (2012).
189. Friedrich, J., Seidel, C., Ebner, R. & Kunz-Schughart, L. A. Spheroid-based drug screen: considerations and practical approach. *Nat. Protoc.* **4**, 309–324 (2009).
190. Lovitt, C. J., Shelper, T. B. & Avery, V. M. Advanced Cell Culture Techniques for Cancer Drug Discovery. *Biology (Basel)*. **3**, 345 (2014).
191. Lovitt, C. J., Shelper, T. B. & Avery, V. M. Miniaturized three-dimensional cancer model for drug evaluation. *Assay Drug Dev. Technol.* **11**, 435–448 (2013).
192. Wang, C. *et al.* Three-dimensional in vitro cancer models: a short review. *Biofabrication* **6**, (2014).
193. Wenzel, C. *et al.* 3D high-content screening for the identification of compounds that target cells in dormant tumor spheroid regions. *Exp. Cell Res.* **323**, 131–143 (2014).
194. Yip, D. & Cho, C. H. A multicellular 3D heterospheroid model of liver tumor and stromal cells in collagen gel for anti-cancer drug testing. *Biochem. Biophys. Res. Commun.* **433**, 327–332 (2013).
195. Ekert, J. E. *et al.* Three-Dimensional Lung Tumor Microenvironment Modulates Therapeutic Compound Responsiveness In Vitro – Implication for Drug Development. *PLoS One* **9**, e92248 (2014).
196. Fischbach, C. *et al.* Engineering tumors with 3D scaffolds. *Nat. Methods* **4**, 855–860 (2007).
197. Härmä, V. *et al.* A Comprehensive Panel of Three-Dimensional Models for Studies of Prostate Cancer Growth, Invasion and Drug Responses. *PLoS One* **5**, e10431 (2010).
198. Hongisto, V. *et al.* High-throughput 3D screening reveals differences in drug sensitivities between culture models of JIMT1 breast cancer cells. *PLoS One* **8**, (2013).
199. Horman, S. R., To, J. & Orth, A. P. An HTS-compatible 3D colony formation assay to identify tumor-specific chemotherapeutics. *J. Biomol. Screen.* **18**, 1298–1308 (2013).
200. Shin, C. S., Kwak, B., Han, B. & Park, K. Development of an in vitro 3D tumor model to study therapeutic efficiency of an anticancer drug. *Mol. Pharm.* **10**, 2167–2175 (2013).
201. Stone, T. W., Ceruti, S. & Abbracchio, M. P. Adenosine receptors and neurological disease: neuroprotection and neurodegeneration. *Handb. Exp. Pharmacol.* **193**, 535–587 (2009).

202. Mustafa, S. J., Morrison, R. R., Teng, B. & Pelleg, A. Adenosine Receptors and the Heart: Role in Regulation of Coronary Blood Flow and Cardiac Electrophysiology. *Handb. Exp. Pharmacol.* **193**, 161 (2009).
203. Hansson, T., Oostenbrink, C. & Van Gunsteren, W. F. Molecular dynamics simulations. *Curr. Opin. Struct. Biol.* **12**, 190–196 (2002).
204. Karplus, M. & McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* **2002 99** **9**, 646–652 (2002).
205. Hollingsworth, S. A. & Dror, R. O. Molecular Dynamics Simulation for All. *Neuron* **99**, 1129–1143 (2018).
206. Hsieh, K. *et al.* Drug repurposing for COVID-19 using graph neural network and harmonizing multiple evidence. *Sci. Reports 2021 111* **11**, 1–13 (2021).
207. Jia, P. *et al.* Deep generative neural network for accurate drug response imputation. *Nat. Commun. 2021 121* **12**, 1–16 (2021).
208. Tunçbilek, M., Ateş-Alagöz, Z., Altanlar, N., Karayel, A. & Özbey, S. Synthesis and antimicrobial evaluation of some new substituted purine derivatives. *Bioorg. Med. Chem.* **17**, 1693–1700 (2009).

8 Appendix

Supplementary table 1: Predicted activity for a total of 13 compounds belonging to family 1.

Predicted activity	Number of hit compounds	Frequency (%)
Abl kinase inhibitor	3	23
Adenosine receptor antagonist	3	23
Adenylate cyclase V inhibitor	4	31
ADP-thymidine kinase inhibitor	8	62
Aldehyde oxidase substrate	6	46
Alopecia treatment	5	36
Antischemic, cerebral	1	8
Antineoplastic	5	36
Antineoplastic (multiple myeloma)	5	36
Antineoplastic (non-Hodgkin's lymphoma)	13	100
Antineoplastic (solid tumors)	3	23
Antiprotozoal (Trypanosoma)	11	86
Antipsoriatic	1	8
Antiviral (Adenovirus)	1	8
Antiviral (Hepatitis B)	8	62
Antiviral (Herpes)	7	54
Antiviral (Picornavirus)	7	54
Antiviral (Poxvirus)	8	62
Apyrase inhibitor	3	23
Benzoate-CoA ligase inhibitor	1	8
CDK9/cyclin T1 inhibitor	13	100
Cyclic AMP agonist	8	62
Cyclin-dependent kinase 1 inhibitor	1	8
Cyclin-dependent kinase 2 inhibitor	1	8
Cyclin-dependent kinase 5 inhibitor	1	8
Cyclin-dependent kinase inhibitor	1	8
DNA directed DNA polymerase inhibitor	1	8
DNA synthesis inhibitor	3	23
DNA-3-methyladenine glycosylase I inhibitor	5	36
Epidermal growth factor receptor kinase inhibitor	1	8
Erythropoiesis stimulant	2	15
Fibroblast growth factor agonist	2	15
Glutamate-5-semialdehyde dehydrogenase inhibitor	2	15
Glycerol-ether monooxygenase inhibitor	1	8
Immunomodulator	3	23
Immunostimulant	1	8
Immunosuppressant	10	77
Intermittent claudication treatment	3	23
Janus tyrosine kinase 2 inhibitor	6	46
Kidney function stimulant	5	36
Lck kinase inhibitor	2	12
Macrophage stimulant	1	8
Mannotetraose 2-alpha-N-acetylglucosaminyltransferase inhibitor	2	15
Na ⁺ -transporting two-sector ATPase inhibitor 3	3	23
Nootropic	1	8
p21-activated kinase 4 inhibitor	1	8
p21-activated kinase inhibitor	2	15
Polarisation stimulant	4	31
Prenyl-diphosphatase inhibitor	1	8
Protein kinase inhibitor	8	62
Protein-tyrosine kinase (PTK, not ETK, WZC) inhibitor	3	23
Pterin deaminase inhibitor	1	8
Serum-glucocorticoid regulated kinase 1 inhibitor	8	62
Signal transduction pathways inhibitor	8	62
TP53 expression enhancer	12	92
Transcription factor NF kappa A inhibitor	5	38
Tyrosine kinase inhibitor	8	62
Vasodilator, peripheral	1	8

Supplementary table 2: Predicted activity for a total of 15 compounds belonging to family 2.

Predicted activity	Number of hit compounds	Frequency (%)
1-Acylglycerol-3-phosphate O-acyltransferase inhibitor	3	20
2-Dehydropantoate 2-reductase inhibitor	7	47
3'-Demethylstaurosporine O-methyltransferase inhibitor	7	47
3'-Nucleotidase inhibitor	1	7
4-Coumarate-CoA ligase inhibitor	1	7
4-Hydroxyphenylacetate 3-monooxygenase inhibitor	6	40
5 Hydroxytryptamine release inhibitor	2	13
Abl kinase inhibitor	7	47
Acrocyllindropepsin inhibitor	1	7
Adenine deaminase inhibitor	1	7
Adenosine receptor antagonist	2	13
Adenosine regulator	13	87
Adenylate cyclase V inhibitor	4	27
Adenylyl-sulfate reductase inhibitor	7	47
ADP-thymidine kinase inhibitor	8	53
Alcohol dehydrogenase (acceptor) inhibitor	6	40
Aldehyde oxidase inhibitor	1	7
Aldehyde oxidase substrate	1	7
Alpha-1,6-mannosyl-glycoprotein 4-beta-N-acetylglucosaminyltransferase inhibitor	2	13
Alpha-Methylacyl-CoA racemase inhibitor	1	7
Angiogenesis inhibitor	3	20
Antiepileptic	5	33
Antineoplastic	10	67
Antineoplastic (multiple myeloma)	5	33
Antineoplastic (non-Hodgkin's lymphoma)	12	80
Antineoplastic (solid tumors)	9	60
Antineurotic	3	20
Antiprotozoal (Trypanosoma)	12	80
Antiviral (Adenovirus)	1	7
Antiviral (Hepatitis B)	1	7
Antiviral (Herpes)	1	7
Antiviral (Picornavirus)	4	27
Antiviral (Poxvirus)	4	27
APOA1 expression enhancer	2	13
Apyrase inhibitor	9	60
Aryl sulfotransferase inhibitor	1	7
Aryl-acylamidase inhibitor	1	7
Aspergillus nuclease S1 inhibitor	3	20
Aspulinone dimethylallyltransferase inhibitor	3	20
ATP phosphoribosyltransferase inhibitor	4	27
Benzoate-CoA ligase inhibitor	4	27
Biotinidase inhibitor	6	40
Cardioprotectant	3	20
CDK9/cyclin T1 inhibitor	8	53
CDP-glycerol glycerophosphotransferase inhibitor	1	7
Channel-conductance-controlling ATPase inhibitor	7	47
Chaperonin ATPase inhibitor	5	33
Chloride peroxidase inhibitor	4	27
Chymosin inhibitor	1	7
Cyclic AMP agonist	9	60
Cyclic AMP modulator	1	7
Cyclic AMP phosphodiesterase inhibitor	9	60
Cyclin-dependent kinase 1 inhibitor	1	7
Cyclin-dependent kinase inhibitor	1	7
Cyclohexanone monooxygenase inhibitor	6	40

CYP2A1 substrate	1	7
CYP2C19 inhibitor	2	13
CYP2D16 substrate	1	7
CYP2H substrate	1	7
Cytochrome-b5 reductase inhibitor	2	13
Diamine N-acetyltransferase inhibitor	1	13
DNA-(apurinic or apyrimidinic site) lyase inhibitor	6	40
DNA-3-methyladenine glycosylase I inhibitor	5	33
Dynein ATPase inhibitor	2	13
Ecdysone 20-monooxygenase inhibitor	1	7
Endothelial growth factor antagonist	8	53
Epidermal growth factor receptor kinase inhibitor	7	47
ErbB-1 antagonist	7	47
Fibroblast growth factor agonist	1	7
Focal adhesion kinase 2 inhibitor	1	7
Formaldehyde dehydrogenase (glutathione) inhibitor	1	7
Formate-dihydrofolate ligase inhibitor	7	47
General pump inhibitor	1	7
Gluconate 2-dehydrogenase (acceptor) inhibitor	6	40
Glucose oxidase inhibitor	8	53
Glutamate-5-semialdehyde dehydrogenase inhibitor	9	60
Glutamate-tRNA ligase inhibitor	1	7
Glutaminy-peptide cyclotransferase inhibitor	1	7
Glutathione peroxidase inhibitor	1	7
Glycine-tRNA ligase inhibitor	1	7
Growth factor agonist	7	47
Growth factor antagonist	8	53
Guanidinoacetate kinase inhibitor	1	7
Guanidinoacetate N-methyltransferase inhibitor	3	20
H ⁺ -transporting two-sector ATPase inhibitor	9	60
HIF1A expression inhibitor	4	27
Histidine N-acetyltransferase inhibitor	5	33
Imidazoline I1 receptor agonist	4	27
Immunosuppressant	8	53
Inosine nucleosidase inhibitor	4	27
Isofenicillin-N epimerase inhibitor	2	13
Janus tyrosine kinase 2 inhibitor	9	60
Janus tyrosine kinase inhibitor	5	33
Kidney function stimulant	7	47
Kinase inhibitor	2	13
Lck kinase inhibitor	9	60
Leukopoiesis stimulant	4	27
L-threonine 3-dehydrogenase inhibitor	1	7
Lysase inhibitor	9	60
Lysine 2,3-aminomutase inhibitor	7	47
Macrophage stimulant	1	7
Malate dehydrogenase (acceptor) inhibitor	8	53
Malate oxidase inhibitor	8	53
Methylenetetrahydrofolate dehydrogenase (NADP ⁺) inhibitor	7	47
Mitochondrial processing peptidase inhibitor	1	7
Myosin ATPase inhibitor	1	7
N6-methyl-lysine oxidase inhibitor	1	7
Na ⁺ -transporting two-sector ATPase inhibitor	10	66
N-acetylneuraminate 4-O-acetyltransferase inhibitor	1	7
N-acetylneuraminate 7-O(or 9-O)-acetyltransferase inhibitor	7	47
NAD(P) ⁺ -arginine ADP-ribosyltransferase inhibitor	7	47
NADPH peroxidase inhibitor	6	40
NADPH-ferrihemoprotein reductase inhibitor	2	13
N-hydroxyarylamine O-acetyltransferase inhibitor	3	20
Nicotinic alpha2beta2 receptor antagonist	1	7
Nicotinic alpha6beta3beta4alpha5 receptor antagonist	1	7

Nitrate reductase inhibitor	1	7
Nitrite reductase [NAD(P)H] inhibitor	3	20
Nucleoside oxidase (H2O2-forming) inhibitor	8	53
Nucleoside-diphosphatase inhibitor	2	13
Nucleotide metabolism regulator	8	53
Octopamine antagonist	1	7
Pancreatic ribonuclease inhibitor	1	7
Para amino benzoic acid antagonist	2	13
P-benzoquinone reductase (NADPH) inhibitor	1	7
Phosphodiesterase II inhibitor	1	7
Phospholipid-translocating ATPase inhibitor	4	27
Phthalate 4,5-dioxygenase inhibitor	4	27
Preneoplastic conditions treatment	3	20
Prenyl-diphosphatase inhibitor	2	13
Protein kinase inhibitor	11	73
Protein-tyrosine kinase (PTK, not ETK, WZC) inhibitor	7	47
Protein-tyrosine kinase Lyn inhibitor	2	13
Proto-oncogene tyrosine-protein kinase Fgr inhibitor	11	73
Pyroglutamyl-peptidase II inhibitor	1	7
RELA expression inhibitor	1	7
Respiratory distress syndrome treatment	2	13
Retinoic acid metabolism inhibitor	4	27
Rubredoxin-NAD+ reductase inhibitor	4	27
Saccharopepsin inhibitor	1	7
Serum-glucocorticoid regulated kinase 1 inhibitor	6	40
Signal transduction pathways inhibitor	13	87
Sphinganine kinase inhibitor	7	47
Src kinase inhibitor	6	40
Succinate-semialdehyde dehydrogenase [NAD(P)+] inhibitor	2	13
Sulfate adenylyltransferase (ADP) inhibitor	1	7
Sulfite dehydrogenase inhibitor	5	33
Sulfur dioxygenase inhibitor	1	7
Thiamine-triphosphatase inhibitor	2	13
Thiopurine S-methyltransferase inhibitor	1	7
Thromboxane B2 antagonist	2	13
TP53 expression enhancer	13	87
Trans-acenaphthene-1,2-diol dehydrogenase inhibitor	10	67
Transcription factor NF kappa A inhibitor	1	7
Transplant rejection treatment	6	40
tRNA adenylyltransferase inhibitor	1	7
Tyrosine kinase inhibitor	13	87
Vanilloid 1 agonist	1	7
Vascular endothelial growth factor 2 antagonist	2	13
Vascular endothelial growth factor antagonist	2	13
Vasodilator	1	7
Vasodilator, coronary	6	40
Vasodilator, peripheral	4	27
Venom exonuclease inhibitor	4	27
Wound healing agent	1	7
X-methyl-His dipeptidase inhibitor	2	13

Supplementary table 3: Predicted activity for a total of 5 compounds belonging to family 3.

Predicted activity	Number of hit compounds	Frequency (%)
(R)-6-hydroxynicotine oxidase inhibitor	1	20
[acyl-carrier-protein] S-acetyltransferase inhibitor	1	20
1-Acylglycerol-3-phosphate O-acyltransferase inhibitor	1	20
27-Hydroxycholesterol 7alpha-monoxygenase inhibitor	2	40
2-Dehydropantoate 2-reductase inhibitor	2	40
3'-Demethylstaurosporine O-methyltransferase inhibitor	1	20

4-Hydroxyphenylacetate 3-monooxygenase inhibitor	2	40
5 Hydroxytryptamine release inhibitor	2	40
5-O-(4-coumaroyl)-D-quininate 3'-monooxygenase inhibitor	2	40
Acute neurologic disorders treatment	1	20
Adenosine receptor antagonist	2	40
Adenosine regulator	3	60
Adenylyl-sulfate reductase inhibitor	2	40
ADP-thymidine kinase inhibitor	2	40
Alcohol dehydrogenase (acceptor) inhibitor	1	20
Alkane 1-monooxygenase inhibitor	1	20
Alopecia treatment	1	20
Anticonvulsant	2	40
Antieczematic	2	40
Antiepileptic	2	40
Antiischemic, cerebral	2	40
Antineoplastic (non-Hodgkin's lymphoma)	3	60
Antiprotozoal (Trypanosoma)	1	20
Antiviral (Herpes)	1	20
Antiviral (Picornavirus)	2	40
Antiviral (Poxvirus)	3	60
Apyrase inhibitor	2	40
Aspergillus nuclease S1 inhibitor	1	20
Aspulinone dimethylallyltransferase inhibitor	2	40
ATP phosphoribosyltransferase inhibitor	1	20
Benzoate-CoA ligase inhibitor	1	20
Biotinidase inhibitor	1	20
CDK9/cyclin T1 inhibitor	1	20
Channel-conductance-controlling ATPase inhibitor	1	20
Chaperonin ATPase inhibitor	1	20
Creatininase inhibitor	1	20
Cyclic AMP agonist	3	60
Cyclic AMP modulator	1	20
Cyclic AMP phosphodiesterase inhibitor	4	80
Cyclohexanone monooxygenase inhibitor	1	20
CYP2A1 substrate	1	20
Cytochrome-b5 reductase inhibitor	1	20
Diamine N-acetyltransferase inhibitor	1	20
DNA-(apurinic or apyrimidinic site) lyase inhibitor	2	40
DNA-3-methyladenine glycosylase I inhibitor	2	40
Dynein ATPase inhibitor	1	20
Ecdysone 20-monooxygenase inhibitor	2	40
Erythropoiesis stimulant	1	20
Fibroblast growth factor agonist	1	20
FMO1 substrate	1	20
Formate-dihydrofolate ligase inhibitor	1	20
GABA A receptor agonist	1	20
GABA receptor agonist	1	20
Glucose oxidase inhibitor	2	40
Glutamate-5-semialdehyde dehydrogenase inhibitor	2	40
Glutamyl endopeptidase II inhibitor	2	40
Guanidinoacetate N-methyltransferase inhibitor	1	20
H ⁺ -transporting two-sector ATPase inhibitor	2	40
HIF1A expression inhibitor	1	20
Histamine release inhibitor	3	60
Histidine N-acetyltransferase inhibitor	2	40
Immunomodulator	2	40
Immunosuppressant	2	40
Insulin promoter	1	20
Insulysin inhibitor	1	20
Interferon alpha agonist	1	20
Isopenicillin-N epimerase inhibitor	2	40

Kidney function stimulant	4	80
Leukopoiesis stimulant	3	60
L-threonine 3-dehydrogenase inhibitor	1	20
Lysase inhibitor	1	20
Lysine 2,3-aminomutase inhibitor	3	60
Malate dehydrogenase (acceptor) inhibitor	3	60
Malate oxidase inhibitor	2	40
Mannotetraose 2-alpha-N-acetylglucosaminyltransferase inhibitor	2	40
Mediator release inhibitor	1	20
Membrane permeability inhibitor	1	20
Methylenetetrahydrofolate reductase (NADPH) inhibitor	3	60
Mitochondrial processing peptidase inhibitor	2	40
Myosin ATPase inhibitor	1	20
Na+-transporting two-sector ATPase inhibitor	2	40
N-acetylneuraminate 7-O(or 9-O)-acetyltransferase inhibitor	4	80
NAD(P)+-arginine ADP-ribosyltransferase inhibitor	2	40
NADPH peroxidase inhibitor	3	60
N-hydroxyarylamine O-acetyltransferase inhibitor	1	20
Nicotinic alpha6beta3beta4alpha5 receptor antagonist	1	20
Nitrate reductase (cytochrome) inhibitor	2	40
Nitrate reductase inhibitor	1	20
Nootropic	1	20
Nucleoside oxidase (H2O2-forming) inhibitor	2	40
Nucleoside-diphosphatase inhibitor	2	40
Nucleotide metabolism regulator	5	100
Octopamine antagonist	1	20
Omptin inhibitor	1	20
P-benzoquinone reductase (NADPH) inhibitor	1	20
Phospholipid-translocating ATPase inhibitor	2	40
Phthalate 4,5-dioxygenase inhibitor	3	60
Prenyl-diphosphatase inhibitor	1	20
Respiratory analeptic	1	20
Respiratory distress syndrome treatment	1	20
Rubredoxin-NAD+ reductase inhibitor	1	20
Sphinganine kinase inhibitor	2	40
Sulfate adenyltransferase (ADP) inhibitor	1	20
Sulfite dehydrogenase inhibitor	2	40
Tetrahydroxynaphthalene reductase inhibitor	1	20
Thiamine-triphosphatase inhibitor	1	20
Thromboxane B2 antagonist	1	20
TP53 expression enhancer	1	20
Trans-acenaphthene-1,2-diol dehydrogenase inhibitor	2	40
Trimethylamine-oxide aldolase inhibitor	1	20
tRNA adenyltransferase inhibitor	1	20
Undecaprenyl-phosphate mannosyltransferase inhibitor	2	40
Vasodilator	1	20
Vasodilator, coronary	1	20
Vasodilator, peripheral	2	40
Venom exonuclease inhibitor	1	20
Venombin AB inhibitor	1	20
X-methyl-His dipeptidase inhibitor	1	20

Supplementary table 4: Predicted activity for a total of 8 compounds belonging to family B6.

Predicted activity	Number of hit compounds	Frequency (%)
5 Hydroxytryptamine uptake stimulant	2	25
Adenosine regulator	5	63
ADP-thymidine kinase inhibitor	1	13
Albendazole monooxygenase inhibitor	1	13
Antianginal	6	75

Anticonvulsant	1	13
Antieczematic	1	13
Antiischemic, cerebral	4	50
Antineoplastic	2	25
Antineoplastic (multiple myeloma)	2	25
Antineurotic	6	75
Antinociceptive	6	75
Antiprotozoal	2	25
Antiprotozoal (Trypanosoma)	1	13
Antiviral (Adenovirus)	3	38
Antiviral (Picornavirus)	3	38
Antiviral (Poxvirus)	3	38
Autoimmune disorders treatment	3	38
Biliary tract disorders treatment	3	38
CDP-glycerol glycerophosphotransferase inhibitor	5	63
Chloride peroxidase inhibitor	1	13
CYP2C19 inducer	8	100
CYP2C8 inducer	1	13
Dementia treatment	2	25
DNA-3-methyladenine glycosylase I inhibitor	3	38
GABA receptor agonist	1	13
Gluconate 2-dehydrogenase (acceptor) inhibitor	2	25
Glutamate-5-semialdehyde dehydrogenase inhibitor	1	13
Heat shock protein 27 antagonist	3	38
Kidney function stimulant	3	38
Leukopoiesis inhibitor	4	50
Lysine 2,3-aminomutase inhibitor	1	13
Mannotetraose 2-alpha-N-acetylglucosaminyltransferase inhibitor	5	63
Na+-transporting two-sector ATPase inhibitor	3	38
NADPH peroxidase inhibitor	2	25
Nicotinic alpha6beta3beta4alpha5 receptor antagonist	3	38
Nucleotide metabolism regulator	2	25
Octopamine antagonist	2	25
p38 MAP kinase inhibitor	1	13
Phobic disorders treatment	4	50
Phospholipid-translocating ATPase inhibitor	4	50
Phthalate 4,5-dioxygenase inhibitor	1	13
Platelet adhesion inhibitor	3	38
Polarisation stimulant	1	13
Rheumatoid arthritis treatment	2	25
Taurine-2-oxoglutarate transaminase inhibitor	2	25
Venombin AB inhibitor	1	13

Supplementary table 5: Predicted activity for a total of 7 compounds belonging to family B7.

Predicted activity	Number of hit compounds	Frequency (%)
2-Dehydropantoate 2-reductase inhibitor	2	29
3'-Demethylstaurosporine O-methyltransferase inhibitor	2	29
5 Hydroxytryptamine uptake stimulant	1	14
Acute neurologic disorders treatment	5	71
Adenosine A1 receptor agonist	7	100
Adenosine regulator	7	100
Adenylyl-sulfate reductase inhibitor	2	29
ADP-thymidine kinase inhibitor	6	86
Alcohol dehydrogenase (acceptor) inhibitor	1	14
Antianginal	7	100
Anticonvulsant	7	100
Antiepileptic	6	86
Antiischemic	7	100
Antiischemic, cerebral	7	100

Antineoplastic	7	100
Antineoplastic (lymphocytic leukemia)	2	29
Antiviral (Picornavirus)	3	43
Antiviral (Poxvirus)	4	57
Anxiolytic	7	100
Apyrase inhibitor	4	57
CDK9/cyclin T1 inhibitor	1	14
Channel-conductance-controlling ATPase inhibitor	5	71
Chloride peroxidase inhibitor	1	14
Cyclic AMP agonist	1	14
Cyclic AMP phosphodiesterase inhibitor	4	57
Cyclohexanone monooxygenase inhibitor	1	14
CYP2C19 inducer	2	29
DNA-(apurinic or apyrimidinic site) lyase inhibitor	1	14
DNA-3-methyladenine glycosylase I inhibitor	3	43
Glucose oxidase inhibitor	6	86
Glutamate-5-semialdehyde dehydrogenase inhibitor	7	100
H ⁺ -transporting two-sector ATPase inhibitor	7	100
Histidine N-acetyltransferase inhibitor	2	29
HMGCS2 expression enhancer	4	57
Inosine nucleosidase inhibitor	2	29
Isopenicillin-N epimerase inhibitor	1	14
Kidney function stimulant	1	14
Leukopoiesis stimulant	1	14
Lysase inhibitor	1	14
Lysine 2,3-aminomutase inhibitor	6	86
Macrophage stimulant	1	14
Malate dehydrogenase (acceptor) inhibitor	1	14
Malate oxidase inhibitor	3	43
Na ⁺ -transporting two-sector ATPase inhibitor	7	100
N-acetylneuraminate 7-O(or 9-O)-acetyltransferase inhibitor	1	14
NAD(P) ⁺ -arginine ADP-ribosyltransferase inhibitor	3	43
Nucleoside oxidase (H ₂ O ₂ -forming) inhibitor	6	86
Nucleotide metabolism regulator	6	86
Octopamine antagonist	2	29
Protein kinase inhibitor	6	86
Sphinganine kinase inhibitor	2	29
Stroke treatment	7	100
Trans-acenaphthene-1,2-diol dehydrogenase inhibitor	6	86
Tyrosine kinase inhibitor	7	100
Vasodilator, peripheral	1	14
X-methyl-His dipeptidase inhibitor	2	29

Supplementary table 6: List of reagents used and suppliers.

Reagent	Supplier
Ammonium molybdate	Merk
Ammonium vanadate	Merk
ATP	Sigma
BEGM medium	Lonza
Bovine collagen serum	Sigma
Bovine serum albumin	Sigma
Bradford reagent	PanReac AppliChem ITW Reagents
Carnauba wax	Sigma
DMEM 1X medium	Gibco
DMSO	Labkem
FBS	Sigma
Fibronectin	Sigma
KCl	Panreac
MgSO ₄	Panreac
NaH ₂ PO ₄	Merk
NH ₃	Sigma
Nitric acid	Sigma
PBS	Sigma
Penicillin/streptomycin	Gibco
Protease inhibitor cocktail	Sigma
Resazurine	Sigma
RPMI medium	Gibco
Sodium molybdate	Sigma
Tris	Carlo Erba
Triton X-100	Sigma
Trypsin	Merk
Tween-80	Sigma

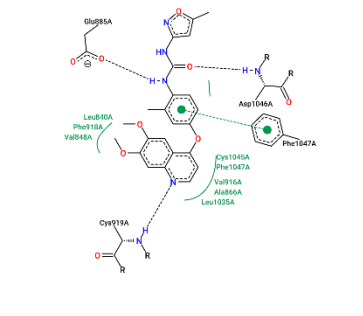
Supplementary table 7: List of equipment and model used.

Equipment	Brand and model
Centrifuge	Eppendorf 5810R Eppendorf MiniSpin
DLS	Horiba Scientific nanopartica nano particle analyzer SZ-100
Freezer -80°C	Thermoscientific TXS Series
Incubator	Binder
Inverted microscopes	Nikon Eclipse TS2 Nikon Eclipse TS100
Microscope camera	Nikon Digital Sight DS-Fi1 Nikon Digital Sight DS-U3
Plate readers	Synergy H1 Hybrid Reader, Biotek Thermoscientific Multiscan Sky
Refrigerator 4°C	Samsung
Sonicator	Branson Digital Sonifier
Vertical flow chambers	Telstar Bio II Advance
Water baths	VWR VWB 12

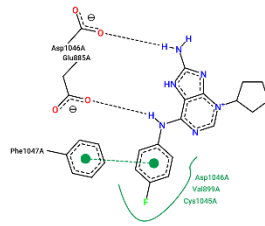
Supplementary figure 1: Predicted ligand-receptor interactions by Autodock 4. Images obtained with ProteinPlus Server.

VEGFR

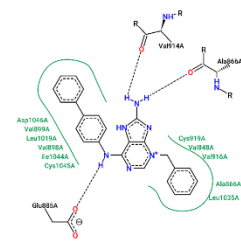
Inhibitor



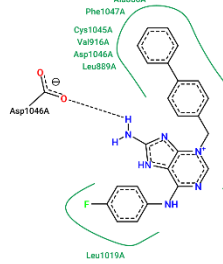
#1



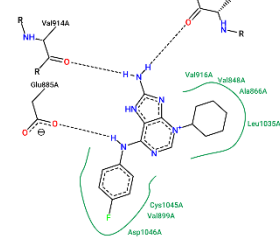
#2



#3

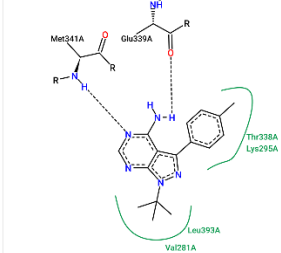


#5

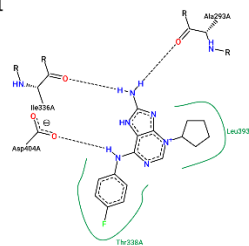


RTK

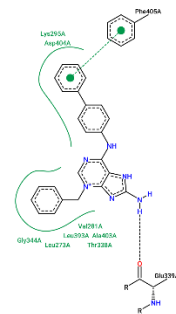
Inhibitor



#1

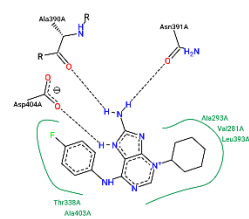


#2



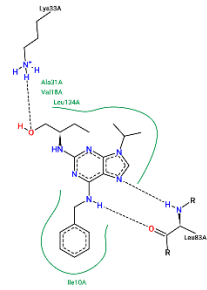
#3

#5

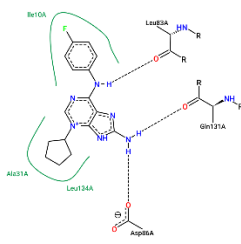


CDK2

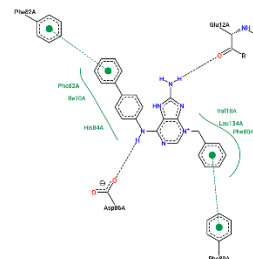
Inhibitor



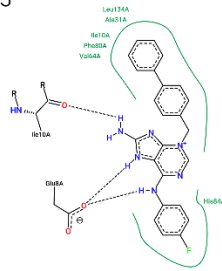
#1



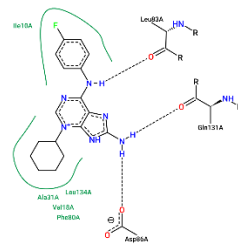
#2



#3

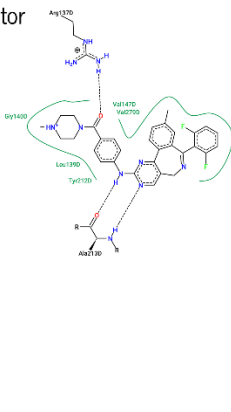


#5

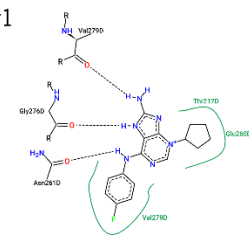


AURKA

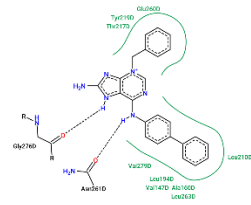
Inhibitor



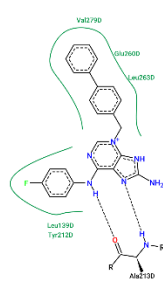
#1



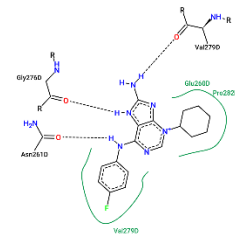
#2



#3

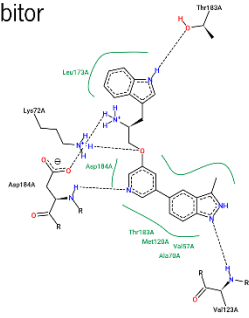


#5

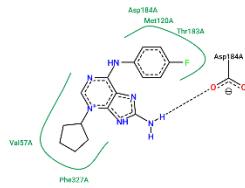


PKA

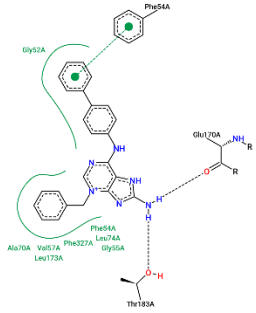
Inhibitor



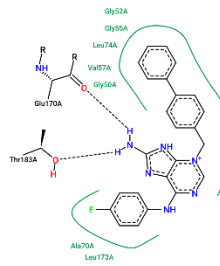
#1



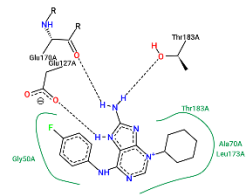
#2



#3

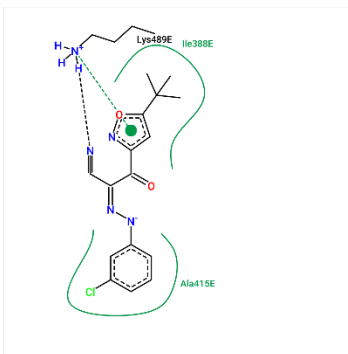


#5

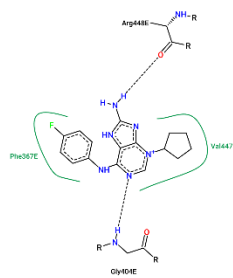


EPAC

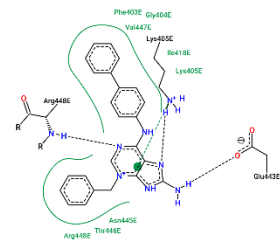
Inhibitor



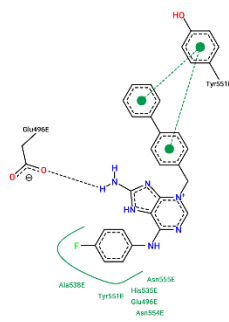
#1



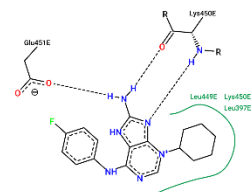
#2



#3

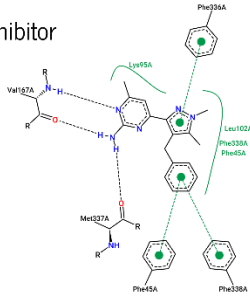


#5

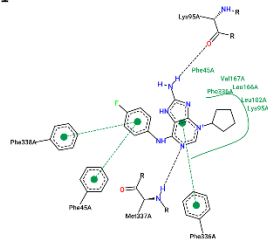


AC

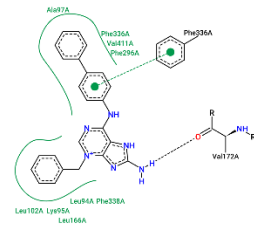
Inhibitor



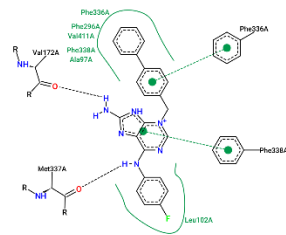
#1



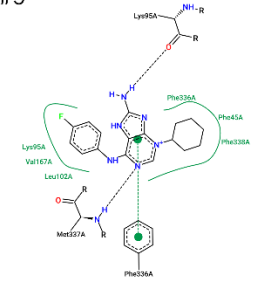
#2



#3

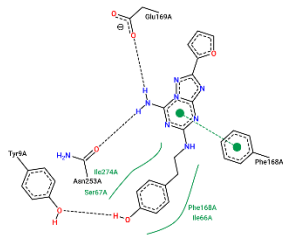


#5

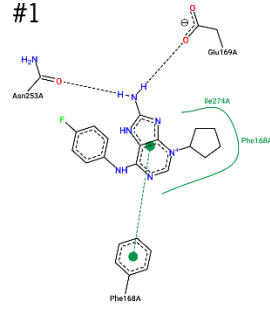


A2a

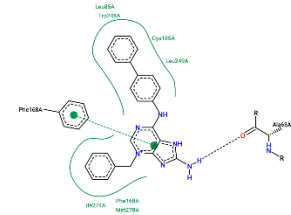
Antagonist



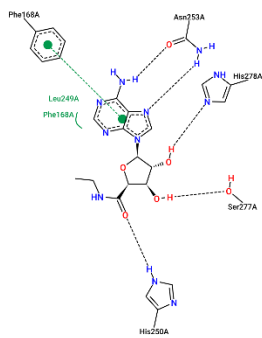
#1



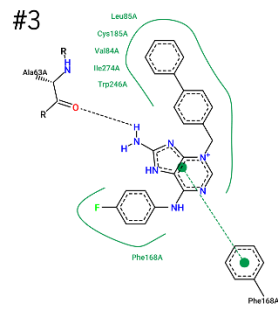
#2



Agonist



#3



#5

