Bioorg. Med. Chem. Lett. 48 (2021) 128240

Contents lists available at ScienceDirect

Bioorganic & Medicinal Chemistry Letters

journal homepage: www.elsevier.com/locate/bmcl

Synthesis, docking, machine learning and antiproliferative activity of the 6-ferrocene/heterocycle-2-aminopyrimidine and 5-ferrocene-1H-Pyrazole derivatives obtained by microwave-assisted Atwal reaction as potential anticancer agents

Eclair Venturini Filho [a], Jorge W.S. Pina [a], Mariana K. Antoniazi [a], Laiza B. Loureiro [a,f], Marcos A. Ribeiro [a], Carlos B. Pinheiro [b], Celina J. Guimarães [c,d], Fátima C.E. de Oliveira [c], Claudia Pessoa [c], Alex G. Taranto [e], Sandro J. Greco [a,*]

[a] Chemistry Department, Federal University of Espírito Santo, Vitória, Espírito Santo CEP.:29075-910, Brazil
[b] Physical Department, Minas Gerais Federal University, Av. Antônio Carlos 6627, Pampulha, Belo Horizonte, Minas Gerais CEP.: 30161-970 Brazil
[c] Department of Physiology and Pharmacology, Faculty of Medicine, Federal University of Ceará., Fortaleza, Ceará CEP 60430-275, Brazil
[d] Pharmacy Sector, Foundation of Oncology Control of the State of Amazonas, Manaus, Amazonas, CEP 69040-010, Brazil
[e] Laboratory of Drug Design and Bioinformatics, Federal University of São João del-Rei, São João del-Rei, Minas Gerais CEP: 36307-352, Brazil

ARTICLE INFO

Keywords:
Pyrimidine
Ferrocene
Atwal reaction
Antiproliferative activity
Computational studies

ABSTRACT

A simple and fast methodology under microwave irradiation for the synthesis of 2-aminopyrimidine and pyrazole derivatives using Atwal reaction is reported. After the optimization of the reaction conditions, eight 2-aminolpyrimidines containing ferrocene and heterocycles and three ferrocene pyrazoles were synthesized from the respective chalcones in good yields. Eight compounds had their structure determined by X-ray diffraction. The molecular hybrid **6a-h** and **9a-c** were tested on four cancer cell lines - HCT116, PC3, HL60 and SNB19 - where four pyrimidine **6a**, **6f-h** and one pyrazole **9c** derivatives show promising antiproliferative activity. In addition, docking simulation and machine learning methods were carried out to explain the biological activity achieved by the synthetized compounds.

Cancer has become a threat to human beings as millions of people are suffering by this disease in all world. Although efforts have been made in reducing cancer incidence, the numbers of cancer patient continue to increase, probably due to the population aging and greater exposure to carcinogenic agents. Therefore, development of effective and safe anticancer agents with less toxicity and high potency remains critically important.

In this context, 2-aminopyrimidine and its derivatives have a special place in cancer chemotherapy for presenting anticancer activity against different types of cancer and for acting through various mechanisms of action. As examples, we can highlight the recent works of Ding and coworkers that designed and synthesized a series of 2-aminopyrimidine derivatives including the compound **2** as highly selective FGFR4 inhibitors in breast cancer cells (Fig. 1).[1] In another example Alam research group synthetized 4-(3-(aryl)-1-phenyl-1H-pyrazol-4-yl)-6-(pyridine-3-yl)pyrimidin-2-amine **3** and evaluated *in vitro* cytotoxicity against a panel of human cancer cell lines namely, HeLa (human cervix), NCI-H460 (human lung), PC-3 (human prostate), and NIH-3T3 (mouse embryo fibroblasts) cell lines.[2] Molecule **4** was prepared in the work of Yoo and coworkers and presented antiproliferative activities against HeLa by inhibitory activities in AXL kinase.[3] Several other examples can be easily found in the literature.[4-9]

The use of ferrocene in medicinal applications is an active research area and several reports have shown that some ferrocene derivatives are highly active *in vitro* and *in vivo*, against several diseases including cancer.[10-13] The anticancer potential of ferrocene derivatives was first reported in the 1970s, by Brynes and co-workers, when they described the antitumor activity of ferrocene compounds bearing amine or amide groups against lymphocytic leukemia P-388.[14] Since then efforts have been made to synthesize ferrocene derivatives with anticancer activity,

* Corresponding author.
E-mail address: sandro.greco@ufes.br (S.J. Greco).
[f] Currently at the State University of Campinas.

https://doi.org/10.1016/j.bmcl.2021.128240
Received 3 February 2021; Received in revised form 2 June 2021; Accepted 28 June 2021
Available online 2 July 2021
0960-894X/

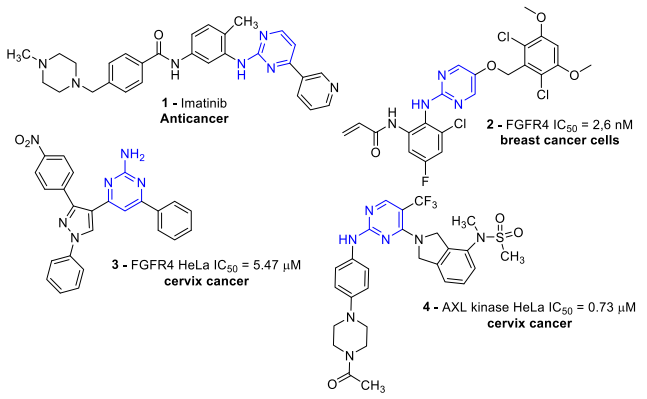


Fig. 1. Selected examples of 2-aminopyrimidines with anticancer activity.

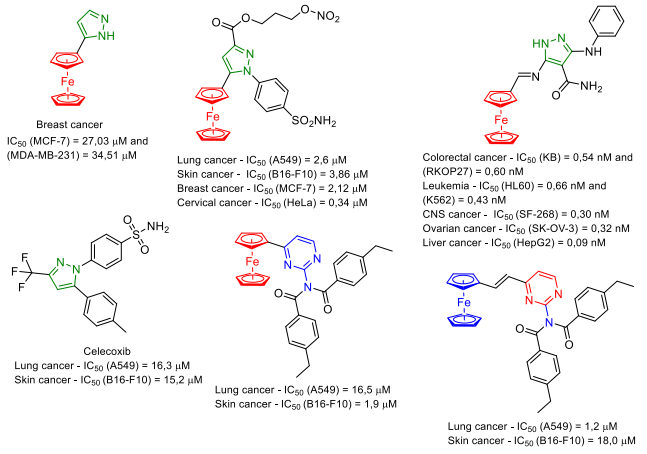
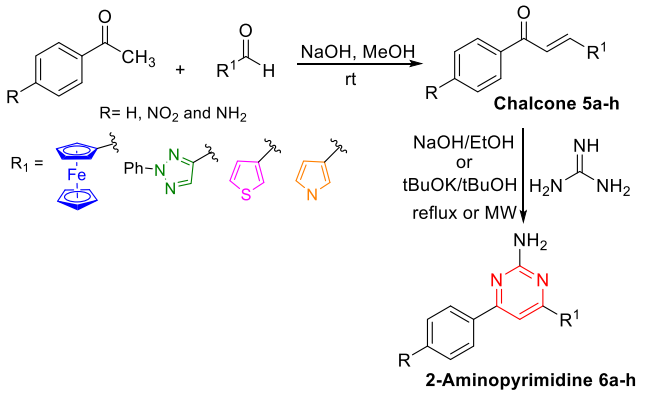


Fig. 2. Molecular Hybrids pyrimidine and pyrazole/ferrocene with anticancer activity.

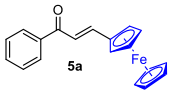
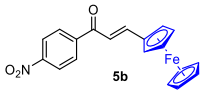
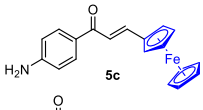
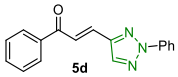
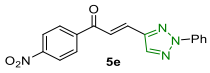
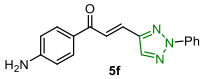
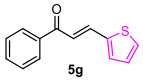
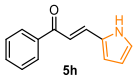


Scheme 1. Synthesis of 2-aminopyrimidines **6a-h**.

and among the great structural diversity, molecular hybrids containing the pyrimidine and pyrazole group can be highlighted. Some examples are shown in Fig. 2.[15–18]

In this work we have explored the traditional and microwave assisted methodology of the Atwal reaction to obtain 6-ferrocene/heterocycle-2-aminopyrimidine and 5-ferrocene-1H-Pyrazole derivatives. Our strategy was based on an easy and rapid two-step synthesis pathway shown in Scheme 1. We started with the preparation of ferrocene and heterocycle chalcones **5a-h** using the classical Claisen-Schmidt condensation reaction.[19] Then, 2-aminopyrimidines **6a-h** were prepared via Atwal reaction by of the respective chalcones with the guanidine chloride in basic media (Scheme 1).[20]

Table 1
Synthesis of the precursor chalcones.

Compound	Chalcone	Time (h)	Solvent	Yield (%)
5a		24	MeOH	97
5b		12	MeOH	95
5c		120	EtOH	89
5d		21	MeOH	97
5e		3	MeOH	54
5f		27	MeOH	45
5g		3	MeOH	90
5h		2	MeOH	75

In the first step, a Claisen-Schmidt condensation reaction was used to prepare chalcones following the standard methodology described in the literature.[19,21] This condensation procedure is attractive since it predominantly generates the (*E*)-isomer.[22] The eight chalcones obtained, the reaction times and the yield results for each one of them are shown in Table 1.

For all chalcones, except **5c,** an enolizable ketone (acetophenone or *p*-nitro or *p*-amino acetophenone) was reacted in the presence of aqueous sodium hydroxide in a methanol solution of the respective non-enolizable aldehyde (ferrocene carbaldehyde or 2-phenyl-2H-1,2,3-trazol-4-carbaldehyde or 1H-pyrrole-2-carbaldehyde or thiophene-2-carbaldehyde). The aldehyde 2-phenyl-2*H*-1,2,3-triazole was prepared in three steps from *D*-(+)-glucose, as previously reported.[23]

Using the Claisen-Schmidt condensation with NaOH as the base and methanol as the solvent, the chalcone **5c** was obtained with yield around 50% and a long reaction time (120 h) was required to consume the reagents. In other to improve both reaction time and yield in **5c** production, alternative methodologies were explored: reactions with different bases (EtONa, tBuONa, DIPEA and DBU), base excess, reflux heating and different solvents (EtOH, ^{i}PrOH, CH_2Cl_2 and THF) were tested. The best result obtained in terms of yield was using EtOH, two equivalents of NaOH and room temperature (89% yield).

Comparing the reaction times for obtaining the chalcones **5a-c** and **5d-f**, the effect of the substituent at the *para* position of the aromatic ring is clearly perceived, since the chalcones **5b** and **5e** containing the nitro group, strongly electrons withdrawer, increases the speed of the reaction. This can be explained for the greater enolate stability.[24]

Through reaction times of chalcones **5a**, **5d**, **5g** and **5h**, we can also note the influence of the aldehyde reactivity. The experimental results show that the pyrrole aldehyde presents the most electrophilic carbonyl group, which is consistent with the electronic effects characteristic of these heterocycles.[25]

Structures of compounds **5a-h** are consistent with IR, ^{1}H and ^{13}C NMR spectra. All compounds showed bands near 1630 cm^{-1} in the IR spectra for the C=O stretch of chalcones. In contrast with the literature,[22] all compounds were obtained, after purification, as single

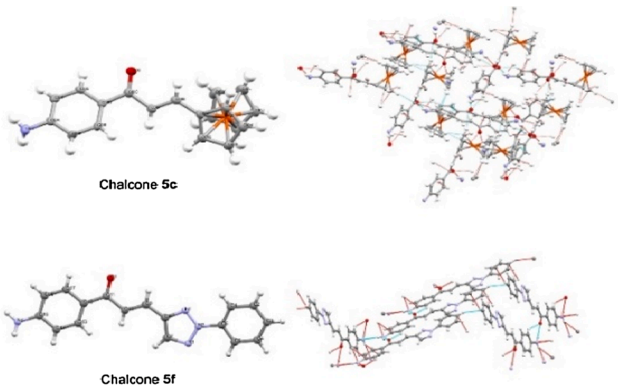
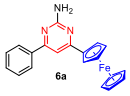
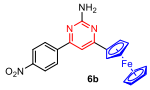
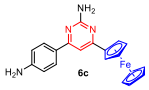
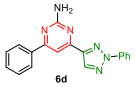
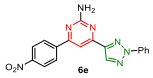
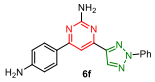
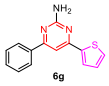
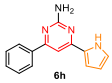


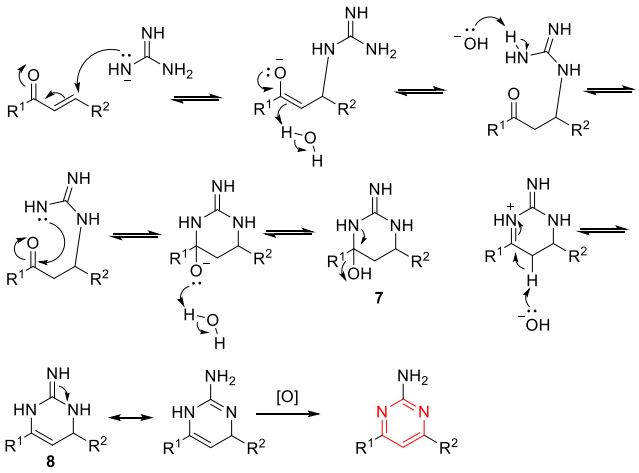
Fig. 3. Molecular representation and atomic labeling scheme of chalcones **5c** and **5f** obtained from single crystal X-ray diffraction technique.

Table 2
Conventional and microwave heating Atwal reaction.

		Conventional Heating		Microwave Heating	
Compound	Pyrimidine	Time	Yield (%)	Time	Yield (%)
6a		15 h 9 h	57[a] 25[b]	1 h 20 min	45[c] 62[d]
6b		4 h 3 h	63[a] 10[b]	10 min 15 min	15[c] 71[d]
6c		35 h 12 h	23[a] 54[b]	2 h 40 min	10[c] 75[d]
6d		6 h 3 h	55[a] 10[b]	2 h 3 h	78[c] 42[d]
6e		1 h 13 h	82[a] 15[b]	30 min 35 min	57[c] 22[d]
6f		3 h 2 h	23 [a] 34[b]	1 h 50 min	52[c] 27[d]
6g		5 h 1 h	35[a] 20[b]	20 min 45 min	58[c] 25[d]
6h		4 h 1 h	10[a] 50[b]	15 min 15 min	42[c] 80[d]

diastereomers. To triazole chalcones **5d-f** one characteristic singlet was observed for hydrogen in the 1,2,3-triazole ring, around 8.10 ppm. The *E*-geometry of the double bonds in **5a-h** were assigned with basis on their ^{1}H NMR spectra. For example, compound **5a** showed a signal for its vinyl hydrogen at 7.10 ppm with a coupling constant of 12.0 Hz, which agrees with the geometry *trans*. Furthermore, crystallization of **5c** and **5f** with hot methanol provided a single crystal for X-ray diffraction study revealed the (*E*) geometry for these chalcones, which in solid state adopts a s-*trans* conformation (Fig. 3).

Initially, the synthesized chalcones were used to obtain the respective 2-aminopyrimidines **6a-h** by Atwal reaction under traditional heating conditions (Table 2). In this procedure, the cyclocondensation occurs usually using guanidine chloride, NaOH as the base and ethanol as solvent. Low yields were observed for all compounds, except for **6e**, which was obtained in 82% yield. An explanation for this is, probably, due to the formation of by-products as described by Chebanov and co-workers.[26]



Scheme 2. Mechanistic proposal for Atwal reaction.

Considering the low yields obtained, different reaction conditions were tested varying the solvent (H_2O, MeOH, EtOH, tBuOH, THF, DMF and DMSO) and the bases (NaOH, EtOK, tBuOK, DBU, Et_3N, DIPEA and Py). For synthesis of **6a**, **6b**, **6d**, **6e** and **6g** the best combination was NaOH/EtOH under reflux whereas for **6c**, **6f** and **6h** better yields were obtained using tBuOK/tBuOH. The detailed experimental procedures for these reactions are described in the supplementary material.

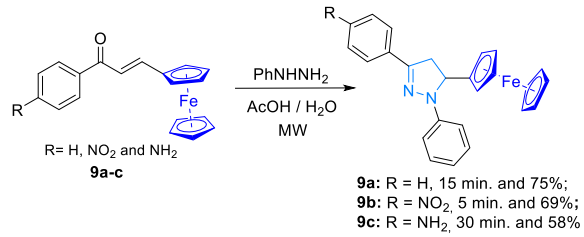
All Atwal reactions described previously using NaOH/EtOH and tBuOK/tBuOH as base/solvent combinations were performed under microwave irradiation and the best results are shown in the Table 2. These findings showed that under microwave assisted conditions the reaction times decreased considerably, and similar or higher yields were obtained when compared with the traditional heating method. This can be explained by the first stage of the cyclocondensation reaction involving an 1,4-addition between guanidine and chalcone. The formed adduct is thermodynamically favored at higher temperatures,[27] and the transfer of energy to the reaction system in heating by microwave increases the efficiency of the process.

The spectral data for the known pyrimidines **6a, 6b, 6g and 6h** shown in the supplementary information agree with those previously reported.[28-30] The novel 2-aminopyrimidine (**6c**, **6d**, **6e** and **6f**) were characterized by IR, NMR spectroscopy and mass spectrometry. IR spectroscopic data showed the formation of new bands around 3500–3200 cm^{-1} typical NH_2 group, besides the absence of the characteristic stretching band of the carbonyl group. According to the ^{1}H NMR, the formation of pyrimidine ring was verified through the signals referring to the amino group and the singlet attributed to the methinic hydrogen present in the pyrimidine ring around 6.20 and 7.20 ppm, respectively.

Furthermore, according to the ^{13}C NMR spectra the signal at 100.6 ppm is attributed to the methinic carbon of the pyrimidine, where the couplings that corroborate with this characterization can be seen in the HSQC and HMBC contour maps. In case of **6c** the characteristic signs of the ferrocene group were observed in ^{1}H NMR data at 4.03 (5H), 4.41 (2H) and 5.01 ppm (2H), in addition to the chemical shifts in ^{13}C NMR at 68.1, 69.9, 70.4 and 82.6 ppm referring to the unsubstituted cyclopentadienyl ring of ferrocene. The ^{1}H NMR of triazole pyrimidines **6d-f** showed an important singlet around 8.50 ppm referring to triazole hydrogen, this being the most deshielding hydrogen in the molecule. The ^{13}C NMR signals present in the triazole ring are assigned around 135.0 and 148.0 ppm.

The electrospray ionization mass spectrometry (ESI-MS) analysis of novel 2-aminopyrimidine **6c**, **6d**, **6e** and **6f** showed a molecular ion peak in 371.09523 *m/z*, 315.13524 *m/z*, 437.19340 *m/z* and 330.14617 *m/z*, respectively, corroborating the spectroscopic results.

The mechanism proposed is initiated by the conjugated addition of



Scheme 3. Ferrocene pyrazole synthesis.

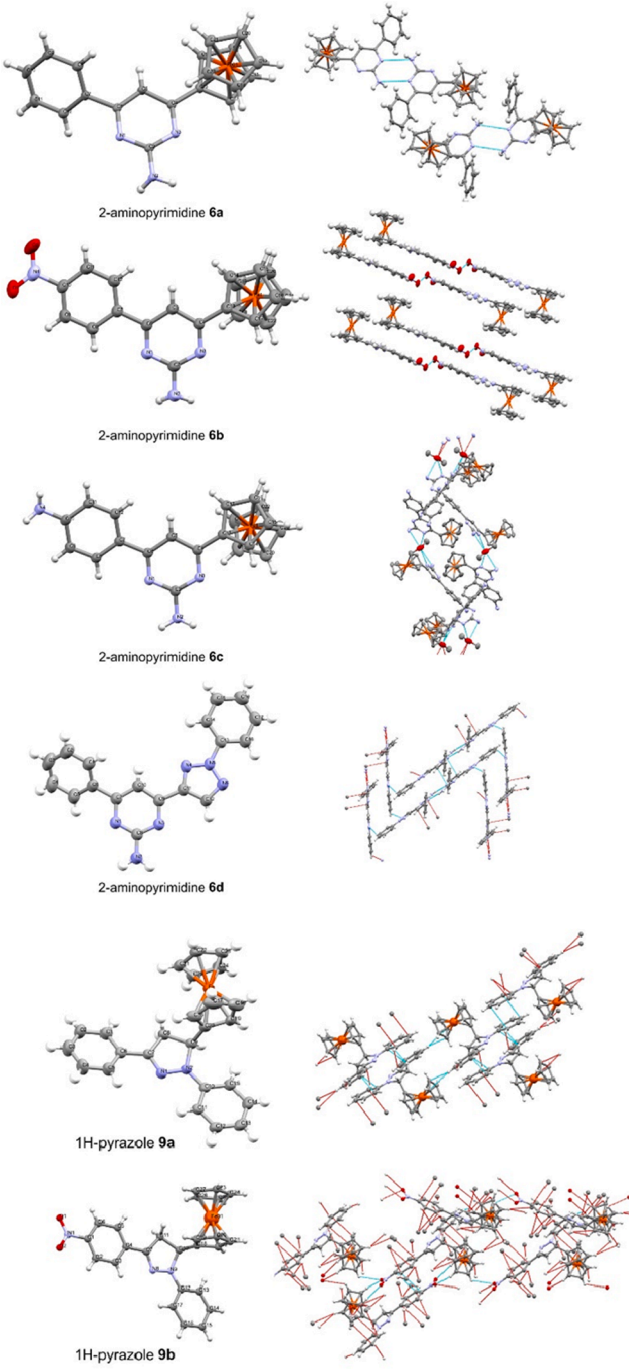


Fig. 4. Molecular representation and atomic labeling scheme of pyrimidine hybrids **6a**, **6b**, **6c** and **6d** obtained from single crystal X-ray diffraction technique. Atomic displacement parameters at 70%. Hydrogen atoms labels were omitted for sake of clarity.

Table 3
Anti-proliferative activities of selected compounds against HCT116, PC3, HL60 and SNB19 cell lines.

Compound	Cell lines - IC_{50} (μM)			
	HCT116	PC3	HL60	SNB19
6a	102.10	100.50	**3.98**	>200
6b	>200	>200	>200	>200
6c	115.00	96.40	75.79	>200
6d	120.20	142.30	66.47	>200
6e	132.80	>200	> 200	132.10
6f	95.27	58.23	68.19	106.50
6g	56.99	**33.56**	70.26	85.11
6h	55.47	81.55	29.33	173.50
9a	>200	101.30	>200	>200
9b	> 200	>200	>200	>200
9c	**3.12**	124.40	6.81	**60.44**

guanidine to chalcone, as suggested by Atwal (Scheme 2).[20,26,31] Subsequently, 1,2-addition occurs between other guanidine nitrogen to the chalcone carbonyl, forming the hydroxylated cyclic intermediate **7**. This hydroxyl is eliminated via anchimeric assistance and E1cb mechanism to former dihydropyrimidinone **8**. This compound is found in imine/enamine tautomeric equilibrium, that after is easily oxidized in solution, leading to the formation of the corresponding amine pyrimidine.

To attest the Atwal methodology under microwave irradiation and increase its scope, 5-ferrocene-1H-pyrazole hybrids were synthesized using phenyl hydrazine instead of guanidine in acid media (Scheme 3).

The spectral data for the known ferrocene 1H-pyrazole **9a** agree with those previously reported.[32] The novel hybrid **9b** and **9c** were characterized by IR, NMR spectroscopy and mass spectrometry.

IR spectroscopic data showed the formation of new bands around 1590 and 1490 cm^{-1} typical ν(C═N and C═C) pyrazole ring. Besides, compounds **9b** and **9c** shows characteristic bands around 1330 cm^{-1} attributed to ν (NO_2) and bands around 3500–3200 cm^{-1} typical NH_2 group, respectively. As well the absence of the stretching band of the carbonyl group. According to the ^{1}H NMR, **9b** and **9c** had similar behavior for 1H-pyrazole ring signals. Both spectral shows evidence for ring formation through the two signals around 3.60–3.90 ppm referring to the methylene hydrogens also one signal around 5.00–5.20 ppm attributed for methinic hydrogen, all signals with double doublet coupling.

Furthermore, according to the ^{13}C NMR spectra all ferrocene-1H-pyrazole hybrids presents characteristic signals of the ferrocene group. The ^{13}C NMR signals present in the pyrazoline ring are assigned 40.97, 59.73 and 143.73 ppm for **9b** and 42.08, 58.49 and 145.71 ppm for **9c**. The electrospray ionization mass spectrometry (ESI-MS) analysis of novel ferrocene-pyrazole **9b** and **9c** showed a molecular ion peak in 450.09154 *m/z* and 422.13103 *m/z*, respectively.

The geometrical properties of some of the compounds obtained could be investigated by single crystal X-ray diffraction techniques. The crystal properties, the data collection, and the structure refinement parameters as well as the geometrical parameters are summarized in the supplementary information. The molecular structures of chalcones were previously shown in Fig. 2 and structures of the **6a**, **6b, 6c, 6d**, **9a** and **9b** are shown in Fig. 4.

X-ray structure determination of compounds **6a**, **6b, 6d, 9a** and **9b** have been obtained from ethanol and **5c, 5f** and **6c** from methanol under slow evaporation. Chalcone **5c** and pyrimidine **6a** and **6c** crystallized in orthorhombic system with $P2_12_12_1$, *Pbca* and *Pbcn* space groups, respectively, and compound **6b** crystallized in a triclinic *P*1 *space group*. Monoclinic crystal system was obtained to **5f**, **6d**, **9a** and **9b** in space group $P2_1$, $P2_1/n$, $P2_1/n$ and $P2_1/c$, respectively.

Further crystallographic details for the structure reported in this paper may be obtained from the Cambridge Crystallographic Data Center, on quoting the depository numbers CCDC- 2041688, 2042671, 1874652, 1874676, 1874661, 2042672, 2042673 and 2042416 for the

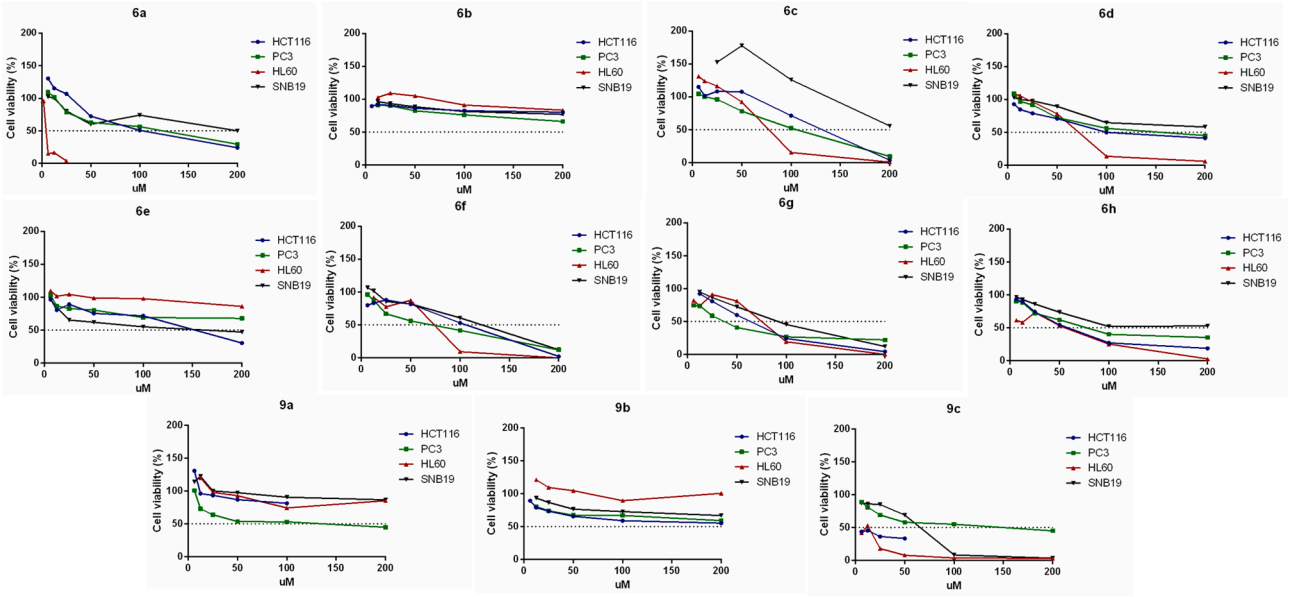
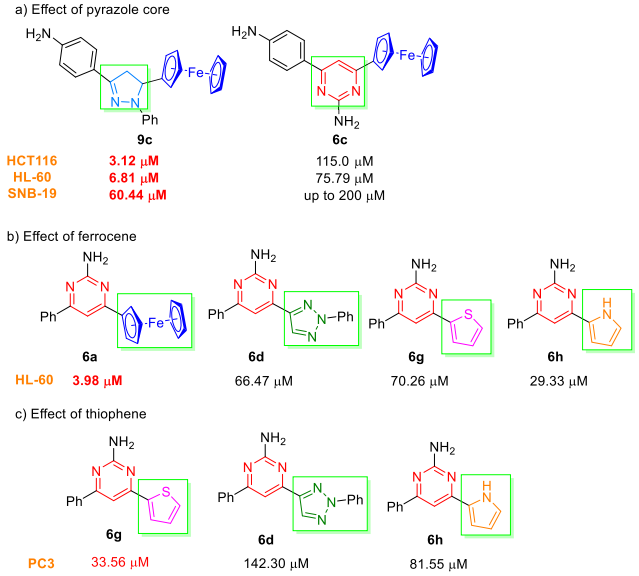


Fig. 5. Inhibitory concentration means (IC50) of the synthesized compound **6a-6h** and **9a-9c** against HCT116, PC3, HL60 and SNB19 tumoral cell lines.



Scheme 4. Comparative summary of the biological activity of molecular hybrids: a) compares and proves that 1H-pyrazole is more active than the pyrimidine core containing the same substituents in the HCT116, HL60 and SNB19 cell lines; b) shows that among the four aminopyrimidines containing the phenyl group on carbon C4 and different pharmacophore groups on carbon C6 (ferrocene, triazole, thiophene and pyrrole), the one with ferrocene is the most active against HL60; c) shows that among three aminopyrimidines containing the phenyl group on carbon C4 and different heterocycles (triazole, thiophene and pyrrole) on carbon C6, the one with thiophene is the most active against PC3 human cancer line.

compounds **5c**, **5f**, **6a-d** and **9a-b**, respectively.

The newly synthesized compounds **6a-h** and **9a-9c** were screened for their *in vitro* antiproliferative activities against four types of human cancer cell lines including HCT116 (Human Colon Carcinoma), PC3 (Prostate Carcinoma), HL60 (Promyelocytic leukemia) and SNB19 (Astrocytoma) by MTT based assay. The results are represented as the half maximal inhibitory concentration (IC_{50}-μM) values in Table 3.

Graph representation of inhibitory concentration mean (IC_{50}) of the compound **6a-h** and **9a-c** against HCT116 (human colorectal carcinoma), PC3 (prostate adenocarcinoma), HL60 (acute promyelocytic leukemia) and SNB19 (astrocytoma) tumoral cell lines are shown in the Fig. 5.

Molecular hybrids **9c**, **6g** and **6h**, showed good antiproliferative activities with IC_{50} of 3.12, 55.47 and 56.99 μM, respectively, for HCT116. For the PC3 cancer cell line, **6f**, **6g** and **6h**, exhibited good antiproliferative activities with the IC_{50} values of 33.56, 58.23 and 81.55 μM, respectively. The compounds **6a**, **6f** and **6h**, inhibited Promyelocytic leukemia cell lines with IC_{50} of 3.98; 6.81 and 29.33 μM and lastly, for SNB19 cancer cell line, compounds **9c** with an IC_{50} value of 60.44 μM, **6g** 85.11 μM and **6f** 106.5 μM, presented the best results.

The screening data indicated that compound **9c** with pyrazole core and substituted with ferrocene and NH_2Ph was more effective that the similar compound **6c** with pyrimidine core in the HCT116, HL60 and SNB19 human cancer cell line (Scheme 4a). In addition, as previously highlighted, **9c** exhibited the best antiproliferative activities to HCT116 and SNB19 cell lines when compared to all compounds synthesized. Furthermore, this hybrid showed predominant cytotoxic potential (3.12 to 124.4 μM) in all tested cell lines. The other pyrazole derivatives **9a** and **9b** substituted with Ph and NO_2Ph, respectively presented low inhibition activity.

The promising anticancer activity of the hybrids containing pyrimidine and ferrocene **6a-c** was confirmed only against the HL60 cancer cells, with **6a** being the most powerful – IC_{50} 3,98 μM – among all compounds. The Hybrid **6a** when compared to structurally similar compounds – **6d**-triazole, **6g**-thiophene, **6h**-pyrrole – the synergetic effect of ferrocene is evident (Scheme 4b). The other pyrimidine/ferrocene hybrids **6b** and **6c** substituted on carbon C4 by p-NO_2Ph and p-NH_2Ph showed low and moderate antiproliferative activities, respectively. The pyrimidine derivatives with phenyl triazole on carbon C6 **6d-f** showed only moderate anticancer activity and the most significant IC_{50} values were determined for the **6f** with NH_2Ph substituent in the prostate carcinoma (58,23 μM). The hybrids **6g** and **6h** with thiophene and pyrrole, respectively, were effective in practically all tested cell lines, highlighting the compound **6g** in the colon carcinoma, prostate carcinoma, and astrocytoma cancers and hybrid **6h** in the colon carcinoma, prostate carcinoma, and Promyelocytic leukemia. It should be noted that **6g** presented the best results, this is, the low IC_{50} values (33,56 μM) to prostate carcinoma – PC3 (Scheme 4c).

For computational studies, initially, all 2-aminopyrimidines **6a-h** and pyrazole **9a-c** structures were generated using the academic version

Table 4
Grid box parameters for all molecular targets. All values are in Å.

	Receptors			
	1 M17	4HLW	5LGE	5QGF
Center X	22.108	16.542	−14.976	43.588
Center Y	0.234	17.872	−83.394	−31.027
Center Z	52.778	29.481	24.569	−823.967

of Discovery Studio.[33] Next, all compounds were refined by the run_mopac routine using the Eigenvector Following (EF) routine.[34] Run_mopac can refine a set of ligands into the fold using the parametric method 7 (PM7)[35] of MOPAC.[34] During the process of refinement, for each molecule, run_mopac warms the uses if the number of alfa electrons are different from beta electrons. This difference suggests that there is a missing atom, for instance, avoiding human errors. Noteworthy that Autodock tools were not able to generate the ligand pdbqt files property. Thus, the pdb files were converted to pdbqt files using Open Babel software.

The receptors of HCT116, PC3, HL60 and SNB19 were obtained from Protein Data Bank (PDB) under code 5QGF,[36] 5LGE,[37] 4HLW,[38] 1 M17.[39] All receptors were prepared by construction of loop regions using MODELLER implemented into MolAr. Next, hydrogen atoms were included, and the structures were optimized by steepest descent and conjugate gradient algorithms using Chimera. Following, all molecular targets pdb files were converted o pdbqt through Autodock Tools. The compounds were docked using Autodock Vina. The exhaustiveness and all values of the grid box were adjusted for 24; the atomic coordinates of the center of grid box were describe according to the Table 4.

At this point, it is worth noting that the therapeutic targets chosen are well established in the literature for treatment of different cancer types:[40-44] The **1M17** (PDB code) is epidermal growth factor receptor (EGFR) which is overexpressed in colorectal cancer cells.[43] The EGFR is a transmembrane receptor that transmits signals from one cell to another to tell to grow and multiply or to die. Therefore, EGFR plays a key role in cell development, proliferation, and differentiation within a variety of tissues. Deregulation of their activity is extremely associated with tumorigeneses, particularly in the lung, breast, ovarian and colorectal cancer. The second target, **4HLW** (PDB code) is a human Androgen Receptor (AR), a member of the nuclear hormone receptor (NHR) family. The BF3 (Binding Function 3 site of the AR) is a ligand-dependent transcription factor with therapeutic relevance in prostate cancer PC3 cells. Therapeutic targeting of the AR BF3 provides an opportunity to develop AR inhibitors with alternative mechanisms of action and the potential to circumvent resistance to conventional antiandrogens.[42] Them, the **5LGE** (PDB code) an isocitrate dehydrogenase (IDH) 1, is a metabolic enzyme having an important role in the citric acid cycle. IDH1 R132H, the most frequent mutation of the IDH1 gene in glioma, results in missense replacement of arginine by histidine, leading to production of the mutant enzyme that catalyzed the synthesis of oncometabolite 2-hydroxyglutarate. IDH1 mutations are early events in the development of gliomas, and thus the majority of diffuse astrocytoma's and secondary glioblastomas.[45] Lastly, **5QGF** (PDB code) is a Protein Tyrosine Phosphatase 1B (PTP1B). PTP1B is a critical regulator of signaling pathways controlling metabolic homeostasis, cell proliferation, and immunity. Consequently, PTP1B inhibitory activity for the treatment of cancer has been a fervent topic of great interest in the pharmacology.[40] Recent studies have shown that myeloid-specific deficiency of PTP1B is sufficient to promote the development of acute myeloid leukemia.[46]

The docking methodology was evaluated by redocking.[47] The run_mopac was implemented into Molecular Architect (MolAr)[48], which is a workflow integrating other software such as MODELLER and Dock 6 (Grid and Amber score), developed to perform structure based-drug design experiments with a friendly interface. Figure 5 shows the redocking of all molecular targets in this study. As can be seen, there was

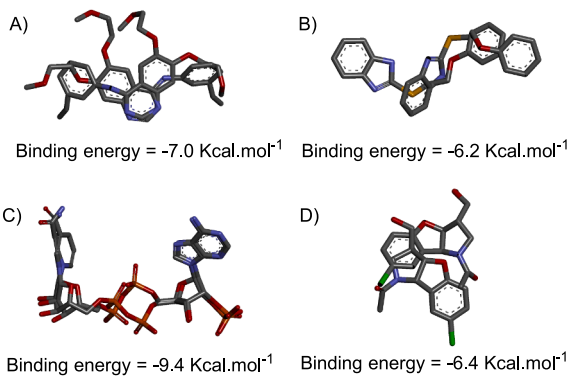


Fig. 6. Redocking results: A, B, C and D correspond to 1 M17, 4HLW, 5LGE and 5QGF, respectively.

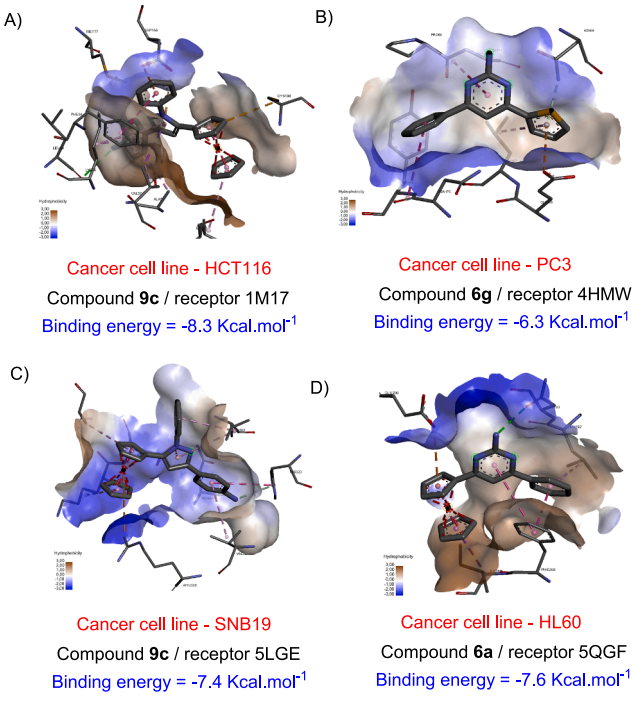


Fig. 7. Hydrophobic surface docked results of the most active compounds with respective cell line. Hydrogen atoms were omitted for a better visualization.

a good fit between docked and crystallographic ligand for the molecular targets 5LGE, reaching root-mean-square deviation (RMSD) value close to zero, and binding energy −9.4 Kcal.mol^{-1} (Fig. 6C). However, there were not fitted between crystallographic ligand and docked of 1 M17, 4HLW and 5QGF (Fig. 6A, B and D). The redocking evaluation showed that the grid box parameters were adjusted addressing the docked ligands correctly for the binding site. In addition, the binding energy of redocked crystallographic ligands were −7.0, −6.2, −9.4 and −6.4 Kcal.mol^{-1} for 1 M17, 4HLW, 5LGE and 5QGF molecular targets, respectively. These binding energies can be used as reference to select the active compounds.

Following, all compounds were submitted against all receptors through docking methodology.[49,50] The docking results showed binding energy from −7.2 to −9.0 Kcal.mol^{-1} for 1 M17; −6.3 to −8.5 Kcal.mol^{-1} for 4HLW; −6.8 to −8.1 Kcal/mol for 5LGE; and −6.3 to −7.9 Kcal/mol for 5QGF. These findings suggest that these ligands can reach the molecular targets. Noteworthy that all compounds had binding energy values less than the crystallographic ligand for 1 M17, 4HLW and 5QGF, with only one except, **9c** for 5QGF. This data suggests that they

Table 5
Data frame applied to build machine learning models.

Compound	Total Mol weight	cLogP	cLogS	Total Surface Area	Relative PSA	Polar Surface Area	Druglikenes	Binding Energy (Kcal/mol)			
								1M17	4HLW	5LGE	5QGF
6a	356.229	2.35	−3.526	190.53	0.19530	51.8	−1.44	−7.9	−6.9	−7.4	−7.6
6b	402.234	2.41	−5.833	265.66	0.25457	97.62	−9.09	−8.1	−6.4	−7.4	−7.9
6c	371.244	1.67	−3.602	199.05	0.26365	77.82	−1.44	−7.9	−6.7	−7.3	−7.5
6d	314.351	2.62	−3.732	245.3	0.26914	82.51	3.03	−8.6	−7.4	−7.4	−7.4
6e	359.348	1.69	−4.192	268.97	0.35855	128.33	−2.34	−9.0	−8.5	−8.1	−7.8
6f	329.366	1.94	−3.808	253.82	0.32027	108.53	3.03	−8.2	−7.3	−7.5	−7.2
6g	253.328	3.31	−4.436	194.76	0.29559	80.04	1.46	−7.2	−6.3	−6.8	−6.4
6h	236.277	2.10	−3.231	188.39	0.27178	67.59	0.52	−7.7	−6.5	−7.0	−6.7
9a	404.293	3.86	−4.272	282.95	0.06305	17.82	1.29	−8.7	−7.1	−7.2	−6.5
9b	449.290	2.94	−4.732	306.62	0.15739	63.64	−3.90	−8.1	−7.0	−7.2	−6.6
9c	419.308	3.18	−4.348	291.47	0.11360	43.84	1.25	−8.3	−6.7	−7.4	−6.3

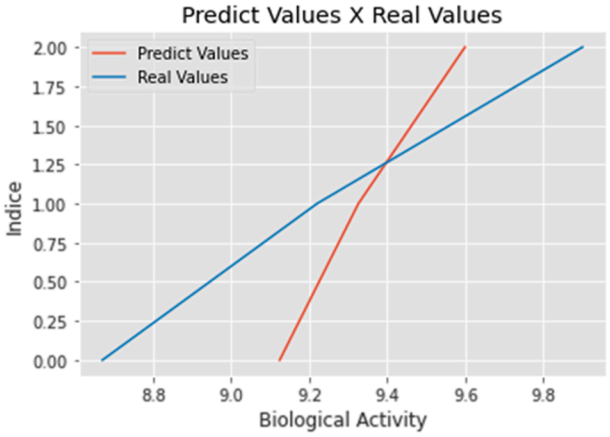


Fig. 8. Correlation between real and predict biological activity for PC3 cell line through linear regression using 4HMW, cLogP and druglines as features.

are more active than crystallographic ligand for the cell line previously described.

Fig. 7 summarize the ligands with the best biological activity. As can be see, the ferrocene **6a** and **9c** compounds showed the best biological activity values for HCT116 (Fig. 7A), SNB19 (Fig. 7C) and HL60 cell lines (Fig. 7D), with binding energy of −8.3, −7.4 and −7.6 Kcal.mol^{-1}, respectively. These values are low than the respective crystallographic ligands (Fig. 6). In addition, as can be seen in the Fig. 6, the ferrocene moiety can be form complex into hydrophobic and hydrophilic cavities of the molecular targets. This feature can explain their biologic activity, which the ferrocene compounds were activities in 3 of the 4 cell lines. In contrast the pyrimidine **6g** was the most active compounds for PC3 cell line with binding energy of −6.3 Kcal.mol^{-1} performing hydrophilic and hydrophobic molecular interaction with 4HLW molecular target.

To Machine Learning (ML) the Autodock Vina[51] binding energy (BE) and physical-chemistry features (descriptors) were obtained by Datawarrior software.[52] The parameters, cLog P, cLog S, Total Molecular Weight, Relative Polar Surface Area (PSA), Drug likeness, Total Surface and Polar Surface Area were used to develop ML models. Thus, supervised, and unsupervised models were generated using K-Nearest Neighbors (KNN) and Linear Regression (LR) through Jupyter Notebook.[53] The data frame was per-processed using pandas and numpy library, which a Log function was applied for the biological values. No missing values were found in the data frame. In addition, the compounds with value equal to 20,000 were classified as inactive and the others as active to build the KNN supervised model, which n value was set to 3. The biological activity was the target to be predicted. In addition, the structures with biological activity more than 2000 were removed from the dataset. The binding energy, together with other descriptors, were used to generate machine learning models, as shown in Table 5.

An exhaustive combination of features was carried out to search the best ML model using and KNN and LR algorithms for each cell line. The significant models were found for SB19 and PC3 cell lines through KNN and LR, respectively. In both models, the best features were cLogP, druglikenes and the binding energy for the respective molecular target (5LGE and 4HLW for SB19 and PC3, respectively). As a result, the KNN model could achieve a precision of 1.00 and 0.67 to predict the active and inactive compounds, respectively, with an accuracy of 0.83 and a cross validation value of 0.67. These results means that KNN model had a good prediction for active compounds (100%), but it had few efficient to predicted inactive compounds. This weakness in the KNN model is due to docking methodology, which generates false-positive results. In addition, LR model had 0.59, 0.95 and 0.32 of regression coefficient, cross validation value and root mean square error, respectively. The −0.48, 0.02, −0.09 coefficient values were obtained for 4HLW, LogP and druglikenes, respectively.

The Fig. 8 shows the correlation between real and predict biological activity values. These findings suggest a good model able to biological activity. For instance, the real log of biological active values was 9.22, 8.67 and 9.90; whereas the predicted values were 9.12, 9.33, 9.60, respectively, for **6a**, **6f** and **9b** compounds. Hence, two good ML models were obtained, which can be used to predict the biological activity of SB19 and PC3 cell lines, respectively.

In conclusion, we developed a simple, fast, and efficient methodology for the Atwal reaction under microwave irradiation to synthesize eight 2-amino-4-phenylpyrimidine substituted at carbon C6 with heterocycles and ferrocene and three pyrazole derivatives with phenyl and ferrocene in yield ranging from good to excellent – 52–80%. Furthermore, eight crystal structures were determined successfully.

All hybrids screened for their *in vitro* anti-proliferative activities against four cancer cell lines where it was possible to correlate the anticancer activity with the structures of the synthesized hybrids. The compounds **6g** and **9c** stand out showed predominant cytotoxic potential in all tested cell lines. Furthermore, the biological activity of PC3 cell line could be estimated by docking simulation, whereas the biological activity for HL60 and HCT116 could be estimated by ML model using LR and RF, respectively. These calculations can be used to design new compounds with anticancer activity.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors acknowledge the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq 305117/2017-3 and 2020), the Coordenadoria de Aperfeiçoamento de Pessoal do Nível Superior

(CAPES), Fundação de Amparo à Pesquisa do Espírito Santo (FAPES), Fundação de Amparo à Pesquisa de Minas Gerais (FAPEMIG APQ-02742-17), Federal University of São João del-Rei (UFSJ/PPGBiotec/PPBE/PPGMQ-MG) and Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico (Funcap) for their financial support. A.G. T. is grateful to Mineriando Dados team.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bmcl.2021.128240.

References

1 Mo C, Zhang Z, Guise CP, et al. *ACS Med Chem Lett*. 2017;8:543–548.
2 Alam R, Alam A, Panda AK, Rahisuddin. *Med Chem Res*. 2017:560–570.
3 Choi MJ, Roh EJ, Hur W, et al. *Bioorg Med Chem Lett*. 2018;28:3761–3765.
4 Del Bello F, Farande A, Giannella M, et al. *Bioorg Med Chem*. 2015;23:5725–5733.
5 Kim S, Jung JK, Lee HS, et al. *Bioorg Med Chem Lett*. 2011;21:3002–3006.
6 Kim J, Hong S, Hong S. *Bioorg Med Chem Lett*. 2011;21:6977–6981.
7 Fukuda T, Ishiyama T, Katagiri T, et al. *Bioorg Med Chem Lett*. 2018;28:3333–3337.
8 Sutherlin DP, Bao L, Berry M, et al. *J Med Chem*. 2011;54:7579–7587.
9 Koroleva EV, Ignatovich ZI, Sinyutich YV, Gusak KN. *Russ J Org Chem*. 2016;52:139–177.
10 van Staveren DR, Metzler-Nolte N. *Chem Rev*. 2004;104:5931–5985.
11 Ornelas C, New J. *Chem*. 2011;35:1973–1985.
12 Fouda MFR, Abd-EIzaher MM, Abdelsamaia RA, Labib AA. *Appl Organomet Chem*. 2007;21:613–625.
13 Wang R, Chen H, Yan W, Zheng M, Zhang T, Zhang Y. *Eur J Med Chem*. 2020;190,112109.
14 Fiorina VJ, Dubois RJ, Brynes S. *J Med Chem*. 1978;21:393–395.
15 Guo Y, Wang SQ, Ding ZQ, Zhou J, Ruan BF. *J Organomet Chem*. 2017;851:150–159.
16 Skiba J, Karpowicz R, Szabó I, Therrien B, Kowalski K. *J Organomet Chem*. 2015;794:216–222.
17 Hafez TS, Osman SA, Yosef HAA, et al. *Sci Pharm*. 2013;81:339–357.
18 Xie YS, Zhao HL, Su H, et al. *Eur J Med Chem*. 2010;45:210–218.
19 Claisen L, Claparbde A. *Eur J Inorg Chem*. 1881:2460–2468.
20 Atwal KS, Rovnyak GC, O'Reilly BC, Schwartz J. *J Org Chem*. 1989;54:5898–5907.
21 Nielsen AT, Houlihan WJ. *The Aldol Condesation*. 2011.
22 Ducki S, Rennison D, Woo M, et al. *Bioorg Med Chem*. 2009;17:7698–7710.
23 Tadeu D, Gonzaga G, Braga L, et al. *Eur J Med Chem*. 2017;139:698–717.
24 Eicher T, Hauptmann S, Speicher A. The Chemistry of Heterocycles Structure, Reactions, Synthesis, and Applications, Third Edit., 2012.
25 Joule JA, Keith Mills. Heterocyclic Chemistry, Fifth Edit., 2010.
26 Chebanov VA, Gurley TW, Desenko SM. Azaheterocycles Based on , α,β-Unsaturated Carbonyls, 2008.
27 Clayden J, Greeves N, Wothers P, Warren S. Organic Chemistry, New York, First Edit., 2001.
28 Rashinkar GS, Pore SB, Mote KB, Salunkhe RS. *Indian J Chem*. 2009;48:606–610.
29 Mariella RP, Zelko JJ. *Master Sci*. 1960;513:647–648.
30 El-Rayyes NR. *J Heterocycl Chem*. 1982;19:415–419.
31 Katritzky A. Advances in Heterocyclic Chemistry, First edit., 2016, vol. 119.
32 Zsoldos-Mády V, Ozohanics O, Csámpai A, Kudar V, Frigyes D, Sohár P. *J Organomet Chem*. 2009;694:4185–4195.
33 Dassault Systèmes BIOVIA. 2016.
34 Stewart JJP. 2016, http://openmopac.net/MOPAC2016.html.
35 Dutra JDL, Filho MAM, Rocha GB, Freire RO, Simas AM, Stewart JJP. *J Chem Theory Comput*. 2013;9:3333–3341.
36 Keedy DA, Hill ZB, Biel JT, et al. *Elife*. 2018;7:1–36.
37 Pusch S, Krausert S, Fischer V, et al. *Acta Neuropathol*. 2017;133:629–644.
38 Lack NA, Axerio-Cilies P, Tavassoli P, et al. *J Med Chem*. 2011;54:8563–8573.
39 Al-Suwaidan IA, Abdel-Aziz NI, El-Azab AS, et al. *J Enzyme Inhib Med Chem*. 2015;30:679–687.
40 Tian Zhao B, Hung Nguyen D, Mi Lee B, et al. *Med Chem Res*. 2017;26:2870–2878.
41 Huang GH, Du L, Li N, et al. *Cancer Lett*. 2018;432:93–102.
42 Munuganti RSN, Hassona MDH, Leblanc E, et al. *Chem Biol*. 2014;21:1476–1485.
43 Sever B, Altıntop MD, Özdemir A, et al. *Molecules*. 2020:25.
44 Khan I, Garikapati KR, Setti A, et al. *Eur J Med Chem*. 2019;163:636–648.
45 Jakovļevs A, Vanags A, Gardovskis J, Štrumfa I. *Proc Latv Acad Sci Sect B Nat Exact Appl Sci*. 2020;74:318–324.
46 Murray GI, Mody N, Neel BG, Bence KK, Wilson HM, Delibegovic M. Cancer Res. 2018; 78: 75–87.
47 Balius TE, Mukherjee S, Rizzo RC. *J Comput Chem*. 2011;32:2273–2289.
48 Maia EHB, Medaglia LR, Da Silva AM, Taranto AG. *ACS Omega*. 2020;5:6628–6640.
49 Nunes RR, Da Fonseca AL, Pinto ACDS, et al. *Mem Inst Oswaldo Cruz*. 2019;114:1–10.
50 Carregal AP, Maciel FV, Carregal JB, dos Reis SB, da Silva AM, Taranto AG. *J Mol Model*. 2017;23:1–9.
51 Trott O, Olson AJ. *J Comput Chem*. 2009;31:455–461.
52 Sander T, Freyss J, Von Korff M, Rufener C. *J Chem Inf Model*. 2015;55:460–473.
53 Fernando Perez, Brian E. Granger. 2018, Computing in Science and Engineering.