Pharmaceutical Chemistry Department[1], College of Pharmacy, Taif University, Taif, Saudi Arabia; Medicinal Chemistry Department[2], Faculty of Pharmacy, Beni-Suef University, Beni-Suef, Egypt

Drug likeness, targets, molecular docking and ADMET studies for some indolizine derivatives

A. Belal[1, 2]

Received March 29, 2018, accepted August 9, 2018

Amany Belal, Medicinal Chemistry Department, Faculty of Pharmacy, Beni-Suef University, Beni-Suef 62514, Egypt
abilalmoh1@yahoo.com

Pharmazie 73: 635–642 (2018) *doi: 10.1691/ph.2018.8061*

The aim of this work was to investigate the biomolecular targets for a library of indolizines, study their molecular properties, drug likeness, target prediction, performing docking studies and exploring their ADMET profile in search for a lead compound. All compounds appeared to comply with Lipinski's rule without any violation, additionally their solubility is viewed good except for compounds **4a-c** which are anticipated to be reasonably soluble, their Milog P score was 4.18-4.9, proposing that these compounds are the most lipophilic with least water solubility, however this may be helpful as the cannabinoid receptor-1 is the most probable target for these three compounds. The inclusive target for the selected library was tau protein. Structure based studies demonstrated great fitting of indolizines with tau protein, along these lines they are expected to have pharmacological action *in vivo*. This urged us to think about the ADMET properties of this library. These investigations suggested the ability of the selected compounds to pass the blood brain barrier (BBB) (aside from them compounds **2c** and **3c**) and affect tau proteins, which will be valuable for the treatment of Alzheimer's disease, particularly compound **5** which does not require any SAR modifications to attain the BBB.

1. Introduction

Indolizine scaffold **I** is a heterocyclic aromatic system, which is considered as an isomer of indole. The saturated indolizidine analog **II** represents a core for a variety of alkaloids like swainsonine **III** and castanospermine **IV** (Fig. 1). Indolizidines are widely distributed in nature, particularly in plants. Indolizidine alkaloids exert various pharmacological activities (Michael 2002; Takahata et al. 1993) and became a hopeful scaffold for various synthetic studies (Gundersen et al. 2007; Xue et al. 2016).

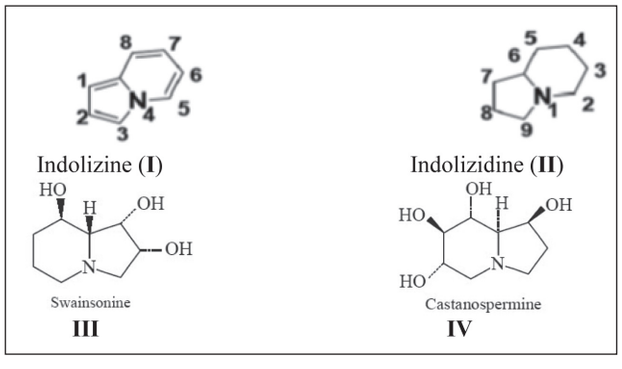

Fig. 1: Indolizine, indolizidine core and indolizidine alkaloids

Tryptophan, the known essential amino acid is the precursor of many important metabolites but it is completely destroyed by acid hydrolysis of proteins. Carbon and Brehm (1961) reported the preparation of β–(1–indolizyl) alanine (**V**) as an analog of tryptophan **VI** (Fig. 2), and as a potential antimetabolite for tryptophan.

Indolizine derivatives show diverse biological activities as herbicidals (Smith et al. 2005), enzyme inhibitors e.g. selective phosphodiestarase 4 (PDE4) inhibitors (Donnell et al. 2010), aromatase and 15-lipoxygenase inhibitors (Sonnet et al. 2000, Gundersen et al. 2003), antileishmanial drugs (Jaisankar et al. 2003), antioxidants (Østby et al. 2000), anticonvulsant/anti-inflammatory drugs (Dawood et al. 2006) and anticancer agents (Moon et al. 2016).

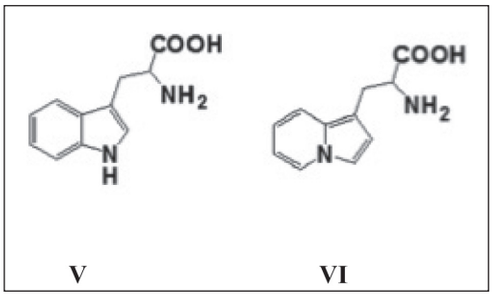

Fig. 2: Tryptophan and its indolizine analog

Most indolizines are exceptionally fluorescent (Borrows and Holland 1948) , and they have been utilized as proficient and adaptable scaffolds in drug labeling (Arvin-Berod et al. 2017). All this demonstrates the significance of indolizines in medicinal chemistry and their capacity to act on different biological targets pushed us forward to investigate a library of indolizines (Belal et al. 2015) (Fig. 3), study their drug likeness properties, using target prediction tools to explore their expected biotargets. Molecular docking is a structure based drug design method useful in determining the possible interactions and binding affinity between the binding site of molecular protein targets and organic molecules. So, molecular docking was used as a computational chemistry tool in this research work. ADME properties describe the four criteria "absorption, distribution, metabolism, and excretion," affecting drug kinetics and pharmacological activities of compounds. Therefore, the ADME profile was studied and discussed in this work. In addition, toxicity potential of the selected compounds was also considered.

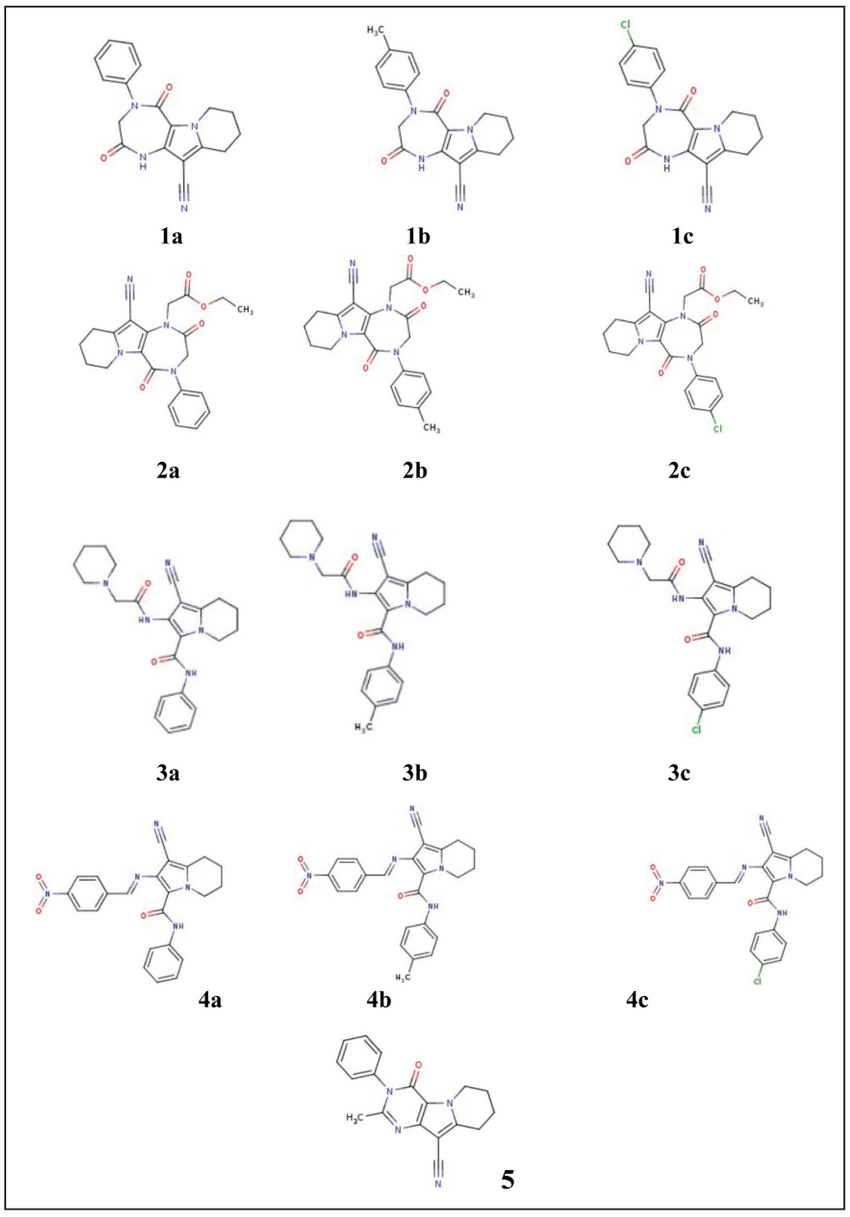

Fig. 3: Selected indolizine library

2. Investigations, results and discussion

2.1. Molecular properties and drug likeness

An *in silico* computational study for a library of indolizine derivatives was performed to detect topological polar surface area (TPSA), the Lipinski's parameters, solubility, volume and drug likeness score (Table 1) using Molsoft (Molsoft 2017). PSA is one of the useful metrics in drug optimization, molecules with polar surface area value greater than 140A^2 won't have the capacity to pass through cell membranes (Pajouhesh and Lenz 2005). Molecules should have PSA less than 90 A^2 to penetrate the blood-brain barrier and attain central nervous system receptors (Hitchcock and Pennington 2006). All selected indolizines showed PSA values less than 140 A^2 and compound **5** revealed the least PSA of all investigated molecules as shown in Table 1. Lipinski's rule of five (RO5) is useful in assessing the bioavailability of orally administered compounds, it states that if the molecule fulfills the accompanying rules: (i) hydrogen bond donors < 5 (OH and NH groups); (ii) hydrogen bond acceptors < 10 (N and O atoms); (iii) molecular weight < 500; (iv) calculated logP < 5, it will have a great potential for oral bioavailability. Absorption of an orally administered compound is more likely if the RO5 is fulfilled (Lipinski et al. 2001). All selected indolizines demonstrated no violations and all were in agreement with RO5 (Table 1), indicating that these compounds have the required cell membrane permeability features. Drug like score was also predicted by Molsoft, all molecules revealed values larger than zero and compounds **4a**, **4b**, in addition to **3a-c** showed values greater than 1. Using more additional filters e.g. Ghose, Veber to detect lead likeness, was also of our interest, swissdock tool (Daina et al. 2017) was used to anticipate increasingly extra parameters for our molecules as number of atoms and aromatic atoms, fraction of sp3 carbons, rotatable bonds and molar refractivity (Table 2). Ghose filter states that, the molar refractivity must be in the vicinity of 40 and 130, and the total number of atoms is in the vicinity of 20 and 70 for compounds to have better drug likeness (Ghose et al. 1999), also Veber filter supports the number of rotatable bonds to be below 10 (Veber et al. 2002). As shown in Table 2, all compounds had a molar refractivity value between 40-130 and total number of atoms from 20-70 and all the compounds were in agreement with these additional filters, suggesting lead similarity of these molecules.

Table 1: Drug-likeness and molecular property prediction

Compd.	MolPSA (A2)	Lipinski´s parameters					Mol LogS in Log (moles /L)	Volume (A3)	Drug like score
		nHBA (NO)	nHBD (OHNH)	LogP	M.W.	No. of viol.			
1a	57.02	3	1	2.06	320.35	0	-4.52	354.45	0.29
1b	57.02	3	1	2.46	334.38	0	-4.97	375.39	0.11
1c	57.02	3	1	2.78	354.09	0	-5.47	371.64	0.38
2a	71.42	5	0	2.36	406.16	0	-3.99	444.73	0.31
2b	71.42	5	0	2.76	420.18	0	-4.45	465.67	0.11
2c	71.42	5	0	3.08	440.13	0	-4.95	461.93	0.55
3a	66.17	4	2	3.18	405.22	0	-4.89	446.50	1.14
3b	66.17	4	2	3.58	419.23	0	-5.34	467.44	1.22
3c	66.17	4	2	3.89	439.18	0	-5.84	463.69	1.67
4a	80.80	5	1	4.18	413.15	0	-7.40	414.27	-0.12
4b	80.80	5	1	4.58	427.16	0	-7.85	435.21	-0.08
4c	80.80	5	1	4.90	447.11	0	-8.35	431.46	0.30
5	39.49	3	0	2.64	304.13	0	-4.45	340.24	0.39

2.2. Target prediction

Compounds that have the ability to bind proteins or macromolecules are considered as bioactive. However, the larger part of these bioactive molecules have more than one target. Computational prediction of their targets based on similarity with known ligands is a promising great approach, helping in the assessment of the number of potential targets and also to predict the off-target effects of these molecules (Gfeller et al. 2013). Swiss target prediction is utilized to distinguish proteins with known ligands similar to our molecules, the obtained results showed the following: Compounds **7a**, **10c** and **11** revealed the highest probability to target muscleblind-like protein-1. Compounds **10a,b** showed the highest probability to target the cannabinoid receptor 1. Compound **8c** had the highest probability to target tyrosyl-DNA phosphodiesterase 1. Compounds **9b** and **9c** demonstrated the highest probability against specific targets like potassium voltage-gated channel subfamily H member 2 and factor X light chain respectively.

Tau proteins are the products of MAPT (microtubule-associated protein tau) the human gene situated on chromosome 17, they are responsible for stabilizing microtubules, they are abundant in the central nervous system, diseases of the nervous system like Alzheimer's and Parkinson's are associated with imperfect tau proteins that became no longer able to stabilize microtubules appropriately. Hyperphosphorylation of tau proteins will promote the appearance of neurofibrillary tangles causing Alzheimer's disease. Finding new compounds ready to target this receptor and have a good penetration to the BBB or CNS permeability will be promising treatment options of Alzheimer's disease. As shown in Table 3, the selected indolizines can target tau protein and compound **8b** showed the highest probability. Additionally compounds **7a-c** showed similarity to 3D structures of more than 1000 tau protein ligand molecules. These selected indolizines revealed to be multi-target compounds, **7b,c** and **8a,b** showed the highest probability to target tau protein with compound **8b** having the highest potential. Compounds **7a, 8c, 9b, 10a-c** showed the highest 3D similarities with tau ligands, compounds **9a** and **11** showed 3D similarity with more than 100 tau ligands, compound **9c** showed the least similarity, it has 3D similarity to 92 structures of tau ligands. All these findings indicate a high potential of the selected indolizine library to act on microtubule-associated protein tau.

The compounds most probable to target the MBLP-1 receptor are **7a**, **10c** and **11** All other compounds showed 3D similarity with MBLP-1 ligands and expected to target this receptor except **9a-c** that have no probability or any 3D similarity with MBLP-1 receptor ligands. Cannabinoid receptor-1 is the third receptor expected to be affected by the investigated indolizines, compounds **10a,b** showed the highest probability to target the cannabinoid 1 receptor, in addition compounds **8c, 9a,c** and **10c** showed 3D similarity with more than 190 cannabinoid 1 ligands, but compounds **7a-c , 8a,b, 9b** and **11** showed neither probability to target cannabinoid 1 receptor nor similarity with its ligands. Also the selected indolizines showed to target tyrosyl-DNA phosphodiesterase 1, since all of them have a probability and similarity with tyrosyl-DNA phosphodiesterase 1 ligands except compound **9b** only. Moreover, their previously reported anticancer activity might be attributed to their potential to target this enzyme which plays a role in DNA repairing. As we have mentioned that their reported anticancer might be explained by their ability to target tyrosyl-DNA phosphodiesterase 1. It is observed that all the selected library compounds have a potential to target tau protein which plays a crucial role in neurodegenerative diseases especially Alzheimer's disease. These hopeful predictions pushed us forward to perform molecular docking studies to explore the ability of these indolizines to bind to tau protein and study the possible interactions.

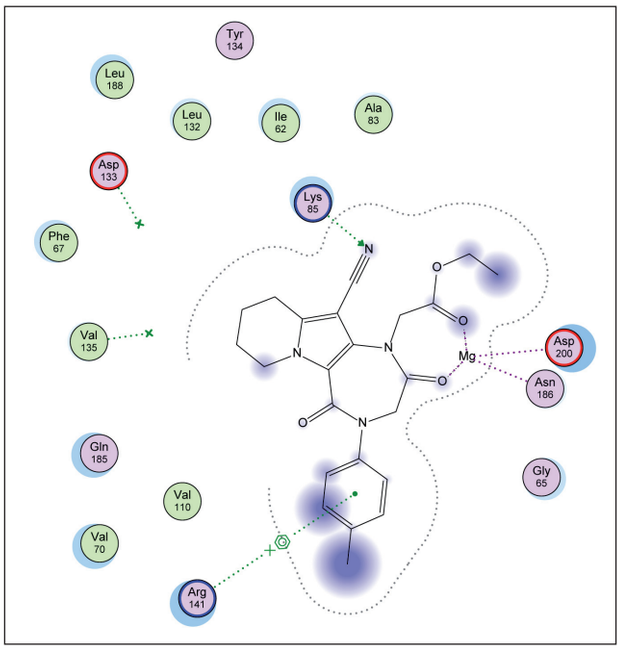

Fig. 4a: 2D interactions between compound **8b** and tau protein binding site

Table 2: Physicochemical properties, obeying Ghose and Veber rules

Compd.	No. of heav. atom	No. of Ar. heav. atom	Csp3	Rot. bonds	Molar ref.	Ghose	Veber	Egan	Muegge	Bioavailab. Score
1a	24	11	0.28	1	95.25	Yes	Yes	Yes	Yes	0.55
1b	25	11	0.32	1	100.22	Yes	Yes	Yes	Yes	0.55
1c	25	11	0.28	1	100.26	Yes	Yes	Yes	Yes	0.55
2a	30	11	0.36	5	115.86	Yes	Yes	Yes	Yes	0.55
2b	31	11	0.39	5	120.83	Yes	Yes	Yes	Yes	0.55
2c	31	11	0.36	5	120.87	Yes	Yes	Yes	Yes	0.55
3a	30	11	0.43	7	120.13	Yes	Yes	Yes	Yes	0.55
3b	31	11	0.46	7	125.10	Yes	Yes	Yes	Yes	0.55
3c	31	11	0.43	7	125.14	Yes	Yes	Yes	Yes	0.55
4a	31	17	0.17	6	119.61	Yes	Yes	Yes	Yes	0.55
4b	32	17	0.21	6	124.57	Yes	Yes	Yes	Yes	0.55
4c	32	17	0.17	6	124.62	Yes	Yes	Yes	Yes	0.55
5	23	15	0.28	1	88.74	Yes	Yes	Yes	Yes	0.55

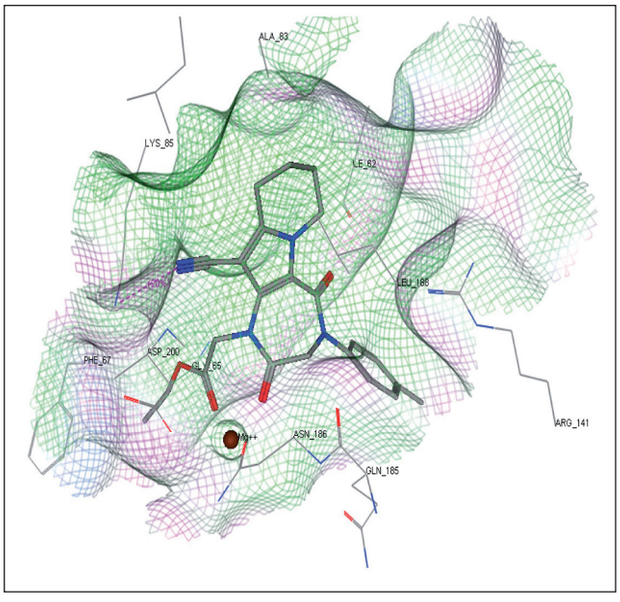

Fig. 4b: Binding mode of compound **8b**inside tau protein binding site

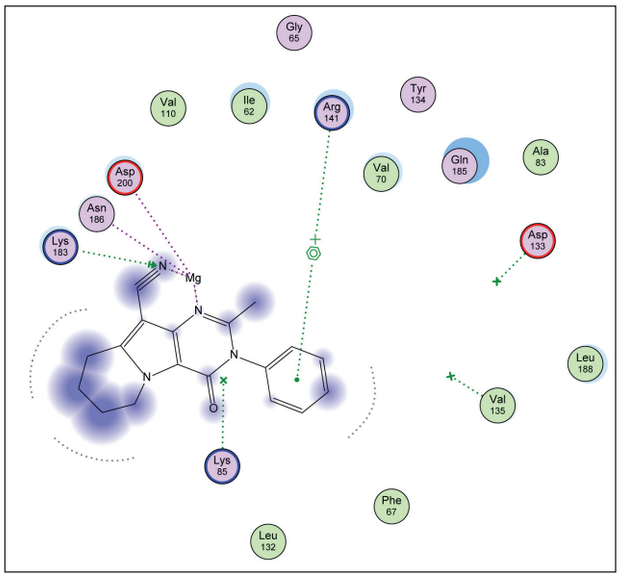

Fig. 5a: 2D interactions between compound **11** and tau protein binding site

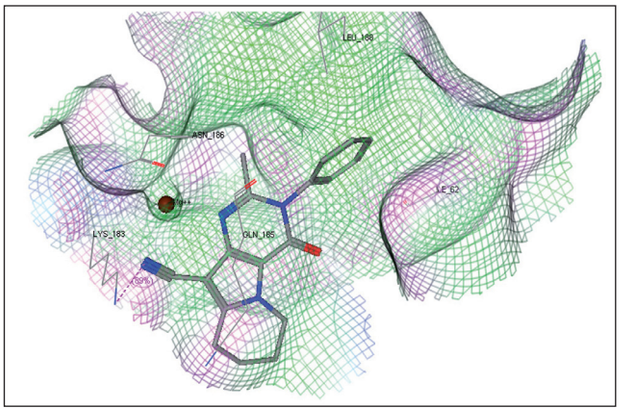

Fig. 5b: Binding mode of compound **11** inside tau protein binding site

2.3. Molecular docking

To further confirm the ability of our indolizine library to target tau protein, docking studies were simulated using the protein data bank file (1J1C), and molecular operating environment program (MOE 2018). A verification process was simulated first after download and refinement of the protein with its co-crystallized ligand, then the ADP (adenosine diphosphate) ligand was removed and redocked to the active binding site in the tau protein. After the verification process the selected indolizines were docked into the active site. 2D and 3D binding study details are available on request from the author.

The binding energy of our indolizines is relatively good, it ranged from -9.15 to -15.80 kcal/mol, which is an indication for affinity and quick fitting of the docked compounds to tau protein binding site. As for binding with the receptor, compounds **8a-c, 9b,c** formed 2 bonds with magnesium as ADP also has 2 bonds with magnesium atom in the binding site. Compounds **7c, 9a, 10a-c** and **11** formed one bond with the magnesium atom of the receptor. Compounds **7c, 8a-c, 9b** and **10a,b** formed a hydrogen bond with lys85 amino acid. Compound **7a** formed a hydrogen bond with val135 amino acid. Some of the compounds formed hydrogen bond with arg141 as **7a,b, 8a,c** and **9a**. Compounds **7a,b, 8b, 9b, 10a** and **11** formed arene-cation interaction with the same amino acid. Compounds **8c** and **10a** formed arene-cation interaction with Lys85. In addition, compound **11** showed more than one site for ligand exposure. Figures 4 and 5 are illustrating the 2D interactions and binding mode of compounds **8b** and **11** with the tau protein binding site. All these findings support the possibility of binding and forming different interactions between the selected indolizines and tau protein and they can act as a potential ligand for this valuable protein target.

Table 3: Compounds targeting microtubule-associated protein tau, 3D/2D similarity

Comp.	Uniprot ID	Gene code	ChEMBL ID	Sim. Comp. (3D / 2D)	Target class
1a	P10636	MAPT	CHEMBL1293224	1234 / 8	Unclassified
1b				1252 / 8	
1c				1405 / 8	
2a				528 / 9	
2b				670 / 10	
2c				290 / 9	
3a				202 / 7	
3b				247 / 10	
3c				92 / 7	
4a				830 / 3	
4b				800 / 3	
4c				968 / 4	
5				105 / 30	

2.4. ADMET profile

Water solubility of any compound has a great effect on its absorption and distribution, compounds with low water solubility will show a bad absorption pattern, It is also known that more than 80% of the marketed drugs have an estimated logS value > -4, our estimated logS values are measured in logmol/liter (Pires et al. 2015). The predicted logS values revealed that all compounds have reasonable solubility with the least soluble compounds **4a-c** (Table 4).

Caco2 permeability is used to predict the absorption of orally administered drugs. According to the PkCSM predicative model (Pires et al. 2015), a compound is considered to have high Caco2 permeability if its predictive value is > 0.90. As presented in Table 4, our selected compounds showed good permeability, however compounds **4a-c** are expected to be moderately permeable. An intestinal absorption model is used to predict the absorbed percentage of an orally administered drug through human small intestine, PkCSM predicative model states that a molecule of absorbance less than 30% will be considered as a poorly absorbed drug. Our compounds, however, showed very good and promising absorbance behavior through human small intestine.

Skin permeability is useful for development of transdermal drug delivery systems. Our compounds showed log Kp values > -2.5 (Table 4), hence all are considered to be skin permeable. P-glycoprotein (P-gp), an ATP-dependent protein, acts as a defense mechanism against harmful substances as it pumps out toxins and foreign substances away from the cells. Expression of P-gp is decreased in Alzheimer's disease but increased in some cancer cells, which can, thus, acquire multidrug resistant characteristics. According to the PkCSM predictive model, our compounds **1b**, **3a-c** and **4a-c** are predicted to be P-gp substrates. Compounds **4a-c** only are expected to be inhibitors for both P-gp I/II and this may increase their potential to be useful in treatment of multi drug resistant cancer cells with over-expressed P-gp.

The higher the volume of distribution, the more amount of the drug reaches the tissues, log VDss values less than -0.15 indicate a low volume of distribution and values > 0.45 indicate a high volume of distributions. The predicted VDss values of our compounds indicate that compounds **3a-c** have a high VDss, compounds **2a-c** have low VDss; the other compounds gave moderate values.

A free unbound portion of a drug is important for traversing cell membranes and distributions into the tissues. Our compounds all showed to have an unbound fraction except compounds **4a-c**, which were that predicted to be totally bound to plasma proteins. Compounds **2c** and **3c** have a log BB value less than -1 indicating that they will be poorly distributed into the brain, compound **5** has a log BB value > 0.3 which indicates that it can pass the BBB. For the other compounds, we predicted values in-between.

Injecting drugs directly into the carotid artery will overcome systemic distribution which affects brain penetration, hence CNS permeability will be a useful direct measurement. The predicted CNS permeability for our selected library showed that compounds **4a-c** and **5** have log PS values > -2 so they are expected to penetrate the CNS, compounds **2b,c** with values < -3 should poorly penetrate the CNS, the other compounds showed moderate values.

As most of the drugs are metabolized by cytochrome P450 enzymes, it is of interest to know if the drug is a substrate for this enzymes or not. Two main isoforms are responsible for metabolism, 2D6 and 3A4. None of the tested compounds was predicted to be a 2D6 substrate but all of them are expected to be substrates for 3A4 except **9b**. It is important to explore if the drugs are inhibitors for cytochromeP450 enzyme isoforms or not, as this will affect the pharmacokinetics of other drugs, According to the PkCSM prediction model (Pires et al. 2015) none of the selected indolizines inhibited all isoforms of the enzyme as shown in Table 5.

Table 4 : ADMET: Absorption and distribution predictions

Compd.	Absorption							Distribution			
	Wat. sol. (log mol/L)	Caco2 perm. (log Papp in 10 cm/s)	Intest. abso. (hum.) (% Absorbed	Skin Per. (log Kp)	P-gP sub.	P-gP I inh.	P-gP II inh.	VDss (human) (log L/kg)	Frac. unb. (hum.) (Fu)	BBB perm. (log BB)	CNS perm. (log PS)
1a	-3.984	0.961	95.805	-3.276	No	No	No	0.047	0.182	-0.228	-2.118
1b	-4.346	1.006	96.275	-3.465	Yes	No	No	0.08	0.174	-0.244	-2.821
1c	-4.754	1.055	94.307	-3.32	No	No	No	0.026	0.176	-0.441	-2.837
2a	-3.998	1.192	94.958	-3.003	No	No	No	-0.205	0.073	-0.842	-2.694
2b	-4.274	0.726	95.401	-3.04	No	No	No	-0.164	0.094	-0.855	-3.026
2c	-4.6	0.718	96.109	-3.071	No	No	No	-0.226	0.077	-1.03	-3.049
3a	-3.767	0.861	93.51	-2.907	Yes	No	No	0.746	0.232	-0.887	-2.327
3b	-4.459	0.959	91.919	-2.862	Yes	No	No	1.022	0.286	-0.755	-2.915
3c	-4.244	0.998	92.129	-2.896	Yes	No	No	0.719	0.229	-1.074	-2.923
4a	-5.54	0.608	100	-2.73	Yes	Yes	Yes	0.054	0	-0.338	-1.81
4b	-5.653	0.496	100	-2.723	Yes	Yes	Yes	0.353	0	-0.606	-1.816
4c	-6.004	0.535	100	-2.735	Yes	Yes	Yes	0.079	0	-0.808	-1.732
5	-3.183	1.222	99.875	-2.643	No	No	No	0.229	0.242	0.41	-1.996

Table 5: ADMET properties: Metabolism and excretion

Compd.	Metabolism							Excretion	
	CYP2D6 subst.	CYP3A4 subst.	CYP1A2 inhibit.	CYP2C19 inhibit.	CYP2C9 inhibit.	CYP2D6 inhibit.	CYP3A4 inhibit.	Total Clear. (log ml/min/kg)	Renal OCT2 subst.
1a	No	Yes	No	Yes	No	No	No	0.149	No
1b	No	Yes	Yes	No	No	No	Yes	0.137	No
1c	No	Yes	No	Yes	No	No	No	-0.183	No
2a	No	Yes	No	Yes	No	No	No	0.28	No
2b	No	Yes	No	Yes	No	No	Yes	0.282	No
2c	No	Yes	No	Yes	Yes	No	Yes	-0.069	No
3a	No	Yes	No	No	No	No	Yes	0.97	No
3b	No	No	No	No	No	No	Yes	0.952	No
3c	No	Yes	No	No	No	No	No	0.946	No
4a	No	Yes	No	Yes	Yes	No	Yes	0.306	No
4b	No	Yes	No	Yes	Yes	No	Yes	0.323	No
4c	No	Yes	No	Yes	Yes	No	Yes	-0.264	No
5	No	Yes	Yes	No	No	No	No	0.9	Yes

The organic cation transporter2 (OCT2) is a renal uptake transporter. It plays a key role in renal clearance of ionized forms of drugs and endogenous compounds as it extracts substances from the blood into the renal tubular cell as the first step in the elimination process (Busch et al. 1998). It is of great value to know if a compound is a substrate for OCT2 or not, to predict its elimination pattern. FDA recommended the evaluation of OCT2 liabilities for the drugs. None of the selected indolizines is predicted to be a substrate for OCT2 except compound **5**. As total clearance is referred to renal clearance and hepatic clearance, it is important to adjust the doses of the drugs until obtaining a steady state concentration. The predicted values for the selected derivatives are expressed in Table 5.

Toxicity prediction (Pires et al. 2015) of the compounds is very important to give an idea about possible hazards and to help in detecting the safest doses of the given compound. Effects of toxins depend on their doses, and Table 6 illustrates the expected toxicity of the selected indolizines. However, the AMES toxicity test has its limitations as *S. typhi* is a prokaryote, and hence it is not an ideal model for humans. However, this test is widely used to test whether a chemical can cause mutations in the DNA of the tested bacteria and positive results indicate that a compound is mutagenic and therefore may act as a carcinogen (Mortelmans and Zeiger 2000). Compounds **7c**, **8a-c**, **9a-c** are expected not to cause damage or injury to *S. typhi* DNA. The maximal tolerated dose is a tool useful in detecting the toxic dose threshold of the chemical compounds, maximum recommended tolerated doses (MRTD) are expressed in log mg/kg/day (Table 6). This will help in choosing the maximal recommended starting dose for pharmaceuticals in clinical trials.

Potassium channel encoded by human ether-a-go-go-related gene hERG inhibition will lead to long QT syndrome that will result in a risk of sudden death. As a side-effect of these inhibitors, hERG is an important antitarget that has to be avoided in the drug development process (Sanguinetti and Tristani-Firouzi 2006). Predicting the hERG inhibition effect for the selected indolizine library in Table 6 revealed that none of these compounds have a tendency to inhibit hERG1. Only compounds **9a-c** and **10a-c** may have a tendency to inhibit hERG2.

The acute lethal dose in rats is expressed as LD_{50} which is the dose given at once and caused death for 50% of the tested animals. Prediction values are in mol/kg and most of the compounds were

Table 6: ADMET properties: Toxicity

Compd.	AMES tox.	Max. toler. dose (hum.) (log mg/kg/day)	hERG I inhibi.	hERG II inhibi.	Oral Rat Acu. Tox. (LD_{50}) (mol/kg)	Oral Rat Chr. Tox. (LOAEL) (log mg/kg_bw/day)	Hepat Toxic.	Skin Sens.	T. Pyri. Tox. (log ug/L)	Minn. Tox. (log mM)
1a	Yes	-0.61	No	No	2.588	1.354	Yes	No	0.885	2.242
1b	Yes	-0.455	No	No	2.634	1.47	Yes	No	0.943	2.102
1c	No	-0.439	No	No	2.546	1.298	Yes	No	0.934	1.666
2a	No	-0.279	No	No	3.038	1.211	Yes	No	0.421	0.58
2b	No	-0.194	No	No	2.876	1.277	Yes	No	0.421	0.221
2c	No	-0.21	No	No	2.979	1.18	Yes	No	0.412	0.003
3a	No	-0.195	No	Yes	2.255	0.35	Yes	No	0.317	1.904
3b	No	-0.32	No	Yes	2.579	0.751	Yes	No	0.302	1.914
3c	No	-0.091	No	Yes	2.277	0.289	Yes	No	0.317	1.952
4a	Yes	-0.438	No	Yes	2.704	1.451	Yes	No	0.476	0.108
4b	Yes	-0.196	No	Yes	3.092	1.405	Yes	No	0.541	1.616
4c	Yes	-0.466	No	Yes	2.962	1.34	Yes	No	0.497	-0.352
5	Yes	-0.452	No	No	2.64	0.904	Yes	No	0.426	1.387

considered safe as to attain the LD_{50} value, one have to use more than 2.5 mol of the compound per kilogram of the tested animal as shown in Table 6.

Chronic toxicity studies aim to detect the lowest doses of a compound that will cause LOAEL (the lowest observed adverse effects) and the treatment period and exposure time to the compound also have to be considered. Log LOAEL predicted values for our compounds indicate that a lager dose of each compound have to be used to induce adverse effects and this is hopeful and indicates an expected large safety margin for these compounds.

None of the selected indolizines is expected to cause any skin sensitization.

Any compound expected to disrupt only one of pathological or physiological liver functions is considered to be hepatotoxic according to the PkCSM model (Pires et al. 2015), so that monitoring and dose adjusting is necessary to avoid liver injury. *Tetrahymena pyriformis* is a protozoic bacterium, predicted doses of our compounds that expected to inhibit 50% of its growth are expressed in log µg/L. The lethal concentration value which is the concentration needed to cause 50% death of flathead Minnow fish, according to the used predicting tool, log the lethal concentration (LC_{50}) values < -0.3 will be considered as highly toxic. As shown in Table 6, all compounds are much far > -0.3 except compound **10c** that showed log an LC_{50} value of -0.352 mM. Moreover, the ADMET-SAR prediction tool (ADMET-SAR 2018), showed that all the selected indolizines are non-carcinogenic with a probability value more than 72%.

2.5. Conclusion

Thirteen indolizine derivatives were selected for investigation e.g. drug likeness properties and bioactivity prediction. The selected indolizine library is considered to contain potential candidates for further research and development as their bioactivity prediction indicates that they can act on various drug targets, and their anticancer activity might be attributed to their ability to target tyrosyl-DNA phosphodiesterase 1. Compounds 4a-c are expected to target the cannabinoid receptor-1, all of them are expected to target Muscleblind-like protein 1 except **3a-c**, compounds **9b** and **9c** are expected to target potassium voltage-gated channel subfamily H member 2 and Factor X light chain, respectively. The universal target for all of the selected indolizines was the microtubule-associated protein tau. Thus structure based drug design e.g. molecular docking was simulated and revealed their ability of binding and interacting with this useful target. According to ADMET-SAR none of these compounds is carcinogenic. All these studies suggest the ability of the selected library to target tau protein and compound **5** has a great potential to penetrate the BBB, hence it can be considered as a promising indolizine derivative for treatment of neurodegenerative diseases associated with tau proteins.

3. Experimental

3.1. Molecular property prediction and drug likeness

Molecular property predictions are computed using Molsoft tool, LogP (octanol/water) :13K compounds were used as database to discover the Partial Least Squares (PLS) regression model, Q2 (cross-validated squared correlation coefficient) value for the best model was found to be = 0.92 and root-mean-square-error (Rmse) value = 0.56. Concerning LogS (water solubility) predictions: 5K compounds were used to find a PLS regression model, Q2 value for the best model is = 0.82 and Rmse = 0.87. As for Molecular Polar Surface Area (PSA) and Volume: 6K compounds were utilized to find the PLS regression model, values of Q2 and Rmse for the best model are 0.99 and 1.56 respectively. The overall drug-likeness score was expected using Molsoft's chemical fingerprints using 10K of non-drug compounds and 5K of marketed drugs. Number of heavy atoms, number of aromatic heavy atoms, Csp3 fraction, rotatable bonds, molar refractivity, obeying Ghose and Veber rules were evaluated using Swiss ADME, 2D of the target compounds were drawn and their smiles were generated, then admitted to Swiss ADME tool, the desired feature predictions were obtained.

3.2. Target prediction

Smiles of the selected library were submitted one by one at the predefined position in Swiss Target Prediction, this enabled us to predict the targets of our molecules, by comparing their 2D and 3D similarities, Swiss Target Prediction perform a comparison between the submitted molecule and 280'000 active compounds against more than 2000 targets in Homo sapiens.

3.3. Molecular docking

The co-crystallized structure of ADP bound at tau protein (PDB: ID: 1J1C) was obtained from the protein data bank (PDB). To study the binding affinity and amino acid interactions, ligand redocking was carried out. RMSD value = 2.60 Å. Docking was performed using London dG force and refinement of the results was done using force field energy. The selected indolizines were prepared for docking via their 3D structure built by Molecular Operating Environment (MOE, Version 2008.09, Chemical Computing Group Inc., Montreal, Quebec, Canada). Protonation of the 3D structures, energy minimizations and docking simulation was performed applying the same docking protocol used with the ligand.

3.4. ADMET profile

ADMET-related properties were calculated using the pkCSM tool. The Caco-2 cell line was used for absorption prediction. This model uses 674 drug like compounds of known permeability. A compound is considered as highly permeable, if its predicted value is > 0.90. The pkCSM also predicts compound water solubility at 25 °C, this model is using values of more than 1700 compounds. The absorbed proportion of the tested compounds through the small intestine of the human can also be predicted. If the absorbance percent is less than 30% , this compound will be considered as poorly absorbed. 332 compounds are being used in predicting the ability of the compound to be transported by P-gp, this will help us to know if the compound is probably going to be a substrate of P-gp or not. P-gp I and II inhibitors: These models are using 1273 and 1275 as reference compounds that showed the ability to inhibit P-gp-I and P-gp-II transport, respectively. Skin permeability is of interest for the advancement of topical medications, this tool uses 211 reference compounds, with an estimated skin permeability in humans, if the logKp value > -2.5, so this compound will be considered to have a moderately low skin permeability.

Volume of distribution (Human VDss), this model uses the calculated VDss for 670 drugs, the predicted VDss of a compound is considered to be low if its value is less than -0.15 and considered of high value if it exceeded 0.45. The Unbound Fraction (Human), this predictive model is using 552 reference compounds. BBB penetration: 320 compounds are used in this predictive model, the compound that show a logBB predictive value > 0.3 is considered to be able to readily cross the blood-brain barrier, if the logBB < −1, this indicates poor distribution of this compound into the brain. CNS permeability: 153 compounds with an experimentally evaluated logPS have been used in the prediction process, a logPS value > -2 gives indication about the ability of the examined compound to penetrate the CNS, while logPS < -3 indicates that this compound will not be able to penetrate the CNS.

Models of different cytochrome P450 isoforms and 14000 to 18000 compounds with the ability to inhibit cytochrome P450 were utilized, inhibitors of cytochrome P450 must cause 50% inhibition of the enzyme with a concentration less than 10 µM. CYP2D6/CYP3A4 substrate, these models are using data base of 671 compounds, their metabolism by each cytochrome P450 isoform has been previously evaluated, the predictor will help us to know if the examined molecule can be metabolized by cytochrome P450 or not. Excretion: OCT2 renal substrate: 906 compounds are being used in this model and 398 compounds are used for predicting the total clearance. As for LD_{50}, the model uses more than 10000 compounds. AMES toxicity method might help in assessment of compounds mutagenicity. This predictive model is using a data base of more than 8000 compounds. *T. pyriformis* toxicity: 1571 compounds are used in this predictive model, the pIGC50 is the logarithm of the required concentration to inhibit 50% of microbial growth, if the value is more than -0.5, this compound will be considered as a toxic. Minnow toxicity: this predictive model uses the LC_{50} values for 554 reference compounds, if the LC_{50} value for the tested compound is less than -0.3, it indicates the high acute toxicity of this compound. Maximum Tolerated Dose (MRTD) gives an estimate of the toxic dose in humans, the model predicts the logarithm of the MRTD for the examined compound, if the value is less than or equal 0.477, it will be considered low dose, and will be considered as high if the value is greater than 0.477. Identifying both LOAEL and NOAEL, which are the lowest dose of a compound that causes adverse effect and the highest dose at which no adverse effects are observed, this predictor uses 445 reference compounds. A hepatotoxic compound is one that interferes with one or more of the normal liver functions. Skin sensitization: This model uses 254 reference compounds which are of known ability to induce skin sensitization. hERG I and II inhibitors: Inhibition of the potassium channels (hERG) will lead to prompting deadly ventricular arrhythmia, these predictors are using 368 and 806 reference compounds, respectively, the predictor will give an idea about the ability of the compounds to inhibit hERG I/II or not.

Acknowledgement: This work was financially supported by Deanship of Scientific Research, Taif University, Taif, Kingdom of Saudi Arabia, Grant No. 1-438-5517.

Conflicts of interest: None declared.

References

ADMET-SAR prediction tool: http://lmmd.ecust.edu.cn/admetsar1, last accessed 2 Jan (2018).

Arvin-Berod M, Desroches-Castan A, Bonte S, Brugière S, Couté Y, Guyon L, Feige JJ, Baussanne I, Demeunynck M (2017) Indolizine-based scaffolds as efficient and versatile tools: application to the synthesis of biotin-tagged antiangiogenic drugs. ACS Omega 2: 9221–9230.

Belal A, Gouda AM, Ahmed AS, Abdel Gawad NM (2015) Synthesis of novel indolizine, diazepinoindolizine and Pyrimidoindolizine derivatives as potent and selective anticancer agents. Res Chem Intern 41: 9687–9701.

Borrows ET, Holland DO (1948) The chemistry of the pyrrocolines and the octahydropyrrocolines. Chem Rev 42: 611–643.
Busch AE, Karbach U, Miska D, Gorboulev V, Akhoundova A, Volk C, Arndt P, Ulzheimer JC, Sonders MS, Baumann C, Waldegger S, Lang F, Koepsell H. (1998) Human neurons express the polyspecificcation transporter hOCT2, which translocates monoamine neurotransmitters, amantadine, and memantine. Mol Pharmacol 54: 342-352.
Carbon JA, Brehm S (1961) 1–Indolizinealanine–A possible tryptophan antimetabolite. J Org Chem 26: 3376–3379.
Daina A, Michielin O, Zoete V. SwissADME (2017) a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. Sci Rep 3: 42717.
Dawood K, Abdel-Gawad H, Ellithey M, Mohamed HA, Hegazi B (2006) Synthesis, anticonvulsant, and anti-inflammatory activities of some new benzofuran-based heterocycles. Arch Pharm 339: 133-140.
Donnell AF, Dollings PJ, Butera JA, Dietrich AJ, Lipinski KK, Ghavami A, Hirst WD (2010) Identification of pyridazino[4,5-b]indolizines as selective PDE4B inhibitors Bioorg Med Chem 20: 2163-2167.
Gfeller D, Michielin O, Zoete V (2013) Shaping the interaction landscape of bioactive molecules. Bioinformatics 29: 3073–3079.
Ghose AK, Viswanadhan VN, Wendolowski JJ (1999) 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.
Gundersen LL, Charnock C, Negussie AH, Rise F, Teklu S (2007) Synthesis of indolizine derivatives with selective antibacterial activity against Mycobacterium tuberculosis, Eur J Pharm Sci 30: 26-35.
Gundersen LL, Malterud KE, Negussie AH, Rise F, Teklu S, Østby OB (2003) Indolizines as novel potent inhibitors of 15-lipoxygenase. Bioorg Med Chem 11: 5409-5415.
Hitchcock SA, Pennington LD (2006) Structure-brain exposure relationships. J Med Chem 49: 7559–7583.
Jaisankar P, Pal B, Manna RK, Pradhan PK, Medda S, Basu MK, Giri VS (2003) Synthesis of antileishmanial (5R)-(-)-5-carbomethoxy-3-formyl-5, 6-dihydroindolo-[2, 3-a]-indolizine. Arkivoc 150-157.
Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 46: 3-26.
Michael JP (2002) Indolizidine and quinolizidine alkaloids. Nat Prod Rep. 19(6):719-741.
MOE Software (last accessed 20 Jan 2018): https://www.chemcomp.com/MOE-Molecular_Operating_Environment.htm,.
Molsoft: https://www.molsoft.com/ , last accessed 30 Dec 2017.
Moon SH, Jung Y, Kim SH, Kim I (2016) Synthesis, characterization and biological evaluation of anti-cancer indolizine derivatives via inhibiting β-catenin activity and activating p53. Bioorg Med Chem Lett 26:110-113.
Mortelmans K, Zeiger E (2000) The Ames Salmonella/microsome mutagenicity assay. Mutat Res 455: 29–60.
Østby OB, Dalhus B, Gundersen LL, Rise F, Bast A, Haenen GRMM (2000) Synthesis of 1-substituted 7-cyano-2,3-diphenylindolizines and evaluation of antioxidant properties. Eur J Org Chem 22: 3763-3769.
Pajouhesh H, Lenz GR (2005) Medicinal chemical properties of successful central nervous system drugs. NeuroRx 2: 541–553.
Pires DE, Blundell TL, Ascher DB (2015) PkCSM: predicting small-molecule pharmacokinetic properties using graph-based signatures. J Med Chem 58: 4066–4072.
Sanguinetti MC, Tristani-Firouzi M (2006) hERG potassium channels and cardiac arrhythmia. Nature 440: 463–469.
Smith SC, Clarke ED, Ridley SM, Bartlett D, Greenhow DT, Glithro H, Klong AY, Mitchell G, Mullier GW (2005) Herbicidal indolizine-5,8-diones: photosystem I redox mediators. Pest Manag Sci 61: 16-24.
Sonnet P, Dallemagne P, Guillon J, Enguehard C, Stiebing S, Tanguy J, Bureau R, Rault S, Auvray P, Moslemi S, Sourdainem P, Séralini G-E (2000) New aromatase inhibitors. Synthesis and biological activity of aryl-substituted pyrrolizine and indolizine derivatives. Bioorg Med Chem 8: 945-955.
Takahata H, Momose T (1993) Simple indolizidine alkaloids. Alkaloids Chem Pharmacol 44: 189–256.
Veber DF, Johnson SR, Cheng HY, Smith BR, Ward KW, Kopple KD (2002) Molecular properties that influence the oral bioavailability of drug candidates. J Med Chem 45: 2615-2623.
Xue Y, Tang J, Ma X, Li Q, Xie B, Hao Y, Jin H, Wang K, Zhang G, Zhang L, Zhang L (2016) Synthesis and biological activities of indolizine derivatives as alpha-7 nAChR agonists. Eur J Med Chem 115: 94-108.