Indian Journal of Biochemistry & Biophysics
Vol. 62, June 2025, pp. 596-606
DOI: 10.56042/ijbb.v62i6.15814

Virtual screening and molecular dynamic simulation to identify the potent SOX2-inhibiting drugs

Vagdevi Sai Pathivada[1], Anannya Bandyopadhyay[2] & Ravindresh Chhabra[1]*

[1]Department of Biochemistry, School of Basic Sciences, Central University of Punjab, Bathinda-151 401, Punjab, India

[2]Department of Zoology, University of Delhi, Delhi-110 007, India

Received 18 December 2024; *revised 14 January 2025*

Sex-determining region of Y-box2 (SOX2) is a master regulator of embryonic and induced pluripotent stem cells. SOX2 is also implicated in epithelial mesenchymal transition (EMT) and chemoresistance of cancer cells. Moreover, SOX2 has been described as a biomarker for cancer stem cells in cervical cancer, sarcoma, ovarian cancer, colorectal cancer, head and neck cancer and glioblastoma. The high expression of SOX2 is also negatively correlated with the overall survival of cancer patients which makes it an attractive target for cancer therapy. The current study was intended to identify SOX2 inhibitors with a high binding affinity. Structure based virtual screening was carried out on approved medications against SOX2 with the help of AutoDock VINA, which is included in the PyRx 0.8 package. The compound with the highest affinity was then examined, and structurally comparable compounds were docked to SOX2 protein once again in order to discover a new and more effective inhibitor molecule against SOX2. The docking analysis revealed apatinib as the most efficient anti-SOX2 drug among the known drug molecules. A structural derivative of apatinib, N-(4-Phenoxyphenyl)-2-[(Pyridin-4-ylmethyl) amino] nicotinamide, was identified as an even more effective inhibitor of SOX2 than apatinib.

Keywords: Antitumor agents, Cancer, Metastasis, NMA analysis, Swiss ADME

Cancer patients have a poor prognosis despite significant advances in the treatment strategies, particularly in underdeveloped nations. This is mainly because cancer is mostly detected at advanced stages when it is difficult to treat. Hence, there has been a consistent effort towards discovering novel drug molecules for controlling and treating cancer. Stem cells have the characteristic feature of self-renewal and ability to differentiate into cells of specialized type[1]. These properties connect the embryonic and adult stem cells to cancerous state if they result in unspecified self-renewal activity[1]. According to cancer stem cell (CSC) hypothesis, they are the root cause for the origin of cancer[2]. The CSCs have been implicated in metastasis, angiogenesis and chemoresistance. So identifying the drugs which target CSCs may be helpful in the treatment of cancer patients[3]. Many signaling pathways including the hedgehog signaling, JAK-STAT signaling, Wnt signaling and notch signaling have been observed to regulate the cancer stem cells[4]. Interestingly, many of these aforementioned pathways involve the function of SOX family of transcription factors[5]. Of these, SOX2 is a regulator of cancer stem cells and their characteristic properties[6]. The expression of SOX2 is associated with the tumor formation, metastasis[7], and epithelial mesenchymal transition[8]. Aberrant expression of SOX2 has been reported in many human cancers such as the colon cancer[8], cervical cancer[9], head and neck squamous cell cancer[10], breast cancer[11], ovarian cancer[12], and lung cancer[13]. High levels of SOX2 are found in multiple types of cancer which makes SOX2 a potential anti-cancer target[14]. The traditional process of discovering novel anticancer medications is extremely expensive, laborious and time-consuming. Computer-aided drug design (CADD) methodologies have been identified as a tool to somewhat overcome the problems associated with traditional drug discovery. It leverages *in silico* tools to design and optimize potential drug candidates, significantly reducing the time and cost associated with traditional methods of drug discovery. Among the various CADD techniques, structure-based virtual screening employing molecular docking is shown to be helpful in the identification of new lead compounds for disease therapy and is a particularly efficient tool for drug development[15]. PyRx is a virtual screening application that is free and open

*Correspondence:
Phone: +91-9478723446 (Mob)
E-mail: ravindresh.chhabra@cup.edu.in

source software created in the Python language. It is a combination of different programmes including AutoDock 4.2, Auto Dock Vina, Open Babel, Mayavi, and others. The docking software of PyRx is built on Vina and AutoDock 4.2[16]. ZINC15 docking web server database provides easy access to chemicals for virtual screening, ligand discovery, pharmacophore screening, force field creation, and other cheminformatics applications. ZINC15 combines the biological activity of gene products, pharmaceuticals, and natural chemicals with commercial availability, making it an extremely effective research tool for ligand discovery[17]. DS visualiser is software used to visualize the interactions and depict those interactions in the 2D and 3D format. It helps in preparation of the protein and ligands suitable for the docking and other computer aided drug designing applications. SOX2, a transcription factor, is considered an undruggable target[18]. It is associated with the nucleosome so it is important to know whether the anti-SOX2 drugs function through protein distortion or by change in the protein-nucleic acid interaction[19]. The ligand-protein complexes are often dynamic, involving conformational changes. The normal mode analysis (NMA) method is widely used simulation technique to describe the collective functional movements of such macromolecular complexes[20]. It helps study the dynamics and flexibility of proteins. The NMA approach in internal coordinates can be implemented by the iMod server (iMODS). The obtained conformations that are created can be downloaded and utilized as inputs in more advanced sampling or modeling procedures based on the deformability and eigen values[21].

This study aimed to identify the most effective SOX2 inhibitors based upon their binding affinities using docking and virtual screening. In addition, the simulations of SOX2-ligand interactions are carried out for the most effective anti-SOX2 inhibitors to elucidate the effect of ligand binding on SOX2 protein.

Materials and Methods

Computer and software

The system with windows 11 operating system and i3 processor was used to access the structures. The DS visualiser-modeling software, PyRx and ZINC15, were used with student access. The structure based drug designing was used for the analysis of ligand-drug interactions. The steps used for this approach are described in the following sections.

Accession of structure of SOX2 protein

The structure of SOX2 protein was accessed through the protein data bank (PDB101.rcsb.org). The least resolution structure was chosen for docking (0.5-2.2) as lesser the resolution of protein, better is the docking process. Keeping this in mind, PDB ID 6T7B was chosen. The structural properties were verified with the reported literature present in the database before the structure was retrieved. As there are no crystallographic and other high end structures available, we considered experimental structures generated through electron microscopy. In accordance with the requirement of PyRx software, we used the .pdb format of the protein for docking. The orientation of SOX2 protein for docking was considered based on its Ramachandran plot viewed in DS visualiser so that the required catalytic sites could be highlighted.

Selection of the ligand molecule for SOX2

Metastasis is one of the key cancer hallmarks regulated by SOX2, therefore in addition to the anti-SOX2 drugs reported in the literature, anti-metastatic drugs were also considered as ligand molecules in this study. The ligands considered for this study included Paclitaxel (taxol), Actinomycin D, Metformin, Ciclesonide, Zoledronic acid, Apatinib, Tamoxifen, Tretinoin and Selumetinib.

The selected ligands were analyzed for the drug likeness by applying Lipinski's rules[22]. The absorption, distribution, metabolism, excretion and toxicity (ADME) information was extracted from the Swiss ADME software (http://www.swissadme.ch/) which uses *in silico* methods to predict ADME data and create drug-like libraries. It evaluates the drug like properties, blood-brain barrier permeability, skin permeability and intestinal absorption, along with Lipinski's rules. The ADME information helps in understanding the efficacy and safety profile of the drugs. To perform these actions, canonical SMILES notions of the respective drug molecules were used and drugs were analyzed for Lipinski's rule of five which are listed below:

i) The molecular weight of the molecule need not be more than 500 Daltons.
ii) The log P value which is the octanol-water partition coefficient value needs to be less than five.
iii) The hydrogen bond donors in the drug must be 5 or less than 5.

iv) The hydrogen bond acceptors in the drug must be 5 or less than 5.
v) More than one of these rules must not be violated.

SMILES (Simplified Molecular Input Line Entry System) is a way of providing a unique, consistent representation of a molecule, thus eliminating the variations that can occur by drawing the same molecule in different ways. This consistency is essential for database searches and computational analysis.

Preparation of SOX2 protein for virtual screening

The first task for SOX2 protein preparation was to retrieve its 3D microscopic structure (as X-ray crystallographic structure is not available) with pdb id 6T7B from the Protein Data Bank. Biovia DS visualiser software was used for protein analysis. To find the binding site in the protein, the PDB structure of the protein was loaded in the DS visualiser. Protein molecules were prepared by removing the hetatoms (DNA molecules associated with the protein structure) because the presence of DNA molecules may cause hindrance in proper ligand binding site. Next, polar hydrogen groups were removed from the protein structure which helps in maintaining the neutrality and stability of the overall structure. The crystal protein structure, thus obtained, was saved in the pdbqt format in the same folder where the pdbqt file of ligand was saved.

Preparation of ligand (drug) molecules using ZINC15 database

ZINC15 is a public database that was created to provide easy access to chemicals for virtual screening. ZINC15 offers over 210 million lead-like 3D compounds for purchase; all molecules are provided in a physiologically appropriate, ready-to-doc 3D format (https://zinc15.docking.org/). The 2D structures of the selected ligands were extracted in the .sdf format for docking analysis[23].

Screening of the ligand (drug) molecules

The selected drug molecules were screened using PyRx software (https://pyrx.sourceforge.io/). The (.sdf) format of drugs was loaded into the PyRx using the open babel application and the drugs were converted into the pdbqt format and the energy if high was reduced and the reduced energies are indicated along as .uff format for better binding.

Molecular Docking of SOX2 with the ligands

Docking is a computerized method for predicting the non-covalent interactions of a ligand with a target protein. Docking investigations were carried out in the current study using PyRx. The prepared protein and ligand were loaded and the grid was defined blind as no particular active site was known.

Receptor grid generation for SOX2 and ligand interaction

Specific site for binding of the ligand molecules in the protein was defined. Blind docking was used as we did not know the binding site in the protein. To permit looking at the huge conformational space accessible to a ligand around a protein, PyRx utilizes a blind dock based strategy to permit quick assessment of the limiting energy of preliminary conformities followed by further steps for molecular docking. The binding affinity value was utilized to determine how strongly the drug chemical interacts with the target protein. This PyRx molecular docking uses systematic simulation computations and Python based codes for evaluation of the confirmation and the flexible nature of ligands. The binding affinities were represented in kcal/mol and it defines the protein-ligand interaction based energies. Also they mention the RMSD (Root Means Square Deviation) values. The considered ligands were docked with the electron microscopic structure of SOX2 (PDBID: 6T7B) individually and the compounds were chosen for further analysis based on their binding affinities.

Visualization of the final docked structures

The final output files of the protein and ligands were visualized using the Biovia DS visualiser in the 2D or 3D formats for depicting the interactions between the amino acids of the protein and the ligand molecules.

Molecular simulation of SOX2 with ligands

The stability and motion of the protein-ligand complex were evaluated using molecular simulation. For this, the ligand which showed the lowest docking energy was considered and its analogues were used for molecular simulation studies. The molecular simulation was carried out using NMA of the iMOD server (http://imods.chaco nlab.org).

Results and Discussion

Accession of structure of SOX2 protein

The electron microscopic structure of the human SOX2 transcription factor in a complex with a nucleosome was retrieved from protein data bank (Fig. 1). The structure consists of SOX2 bound to a nucleosome of histones and DNA. This structure was

determined using cryo-electron microscopy and highlights how the binding of SOX2 to DNA within the nucleosome enhances the accessibility of DNA for transcription. This interaction is key for the downstream functions of SOX2.

Selection of the ligand molecule for SOX2

The ligands used in this study included Paclitaxel (taxol), Actinomycin D, Metformin, Ciclesonide, and Zoledronic acid, Apatinib, Tamoxifen, Tretinoin and Selumetinib. The selection of ligands is done using Lipinski's rule of five and Metformin, Apatinib, Zoledronic acid, ATRA drug-Selumetinib, ATRA drug-Tretinoin and Tamoxifen were selected for further analysis (Table 1).

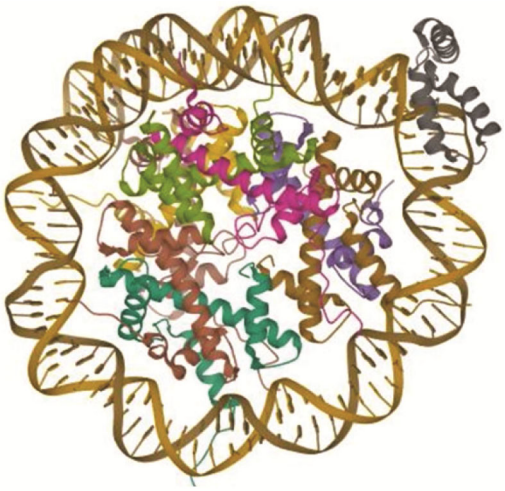


Fig. 1 — The structure of human SOX2 transcription factor in complex with a nucleosome taken from the RSCB PDB (PDB ID: 6T7B). This structure reveals how SOX2 binds to DNA within the nucleosome, causing local distortions at superhelical location 2, a region where the DNA is more accessible. The binding facilitates the detachment of terminal nucleosomal DNA from the histone octamer, aiding in chromatin opening, increasing DNA accessibility and making it more accessible for transcription. This interaction is crucial for the role of SOX2 as a transcription factor in the regulation of stemness properties of the cell.

The functional profile of these selected drugs was extracted from the Pubchem and Drug bank and has been briefly summarized in (Table 2).

The selected ligands were analyzed for the drug likeness properties using Swiss ADME online web server and the drugs, Metformin, Apatinib, ATRA drug-Selumetinib and Tretinoin showed good gastrointestinal absorptivity. SOX2 is not a brain or CNS associated target, therefore, the drugs crossing blood brain barrier had to be excluded. All the drugs except tretinoin do not have permeability to the blood brain barrier. P-glycoproteins (p-gp) are the drug transporters which efflux the drugs. Except apatinib and metformin, remaining drugs act as the p-gp inhibitors and therefore, can be distributed efficiently. The cytochrome P450 enzyme inhibiting drugs generally show good drug interactions. Drugs apatinib and selumetinib show exclusive inhibition towards CYP proteins and increase drug metabolism. LogK_p value which denotes the skin permeability is good for all the drugs as more negative the logK_p, less skin permeable will be the drug. The outcome of this analysis is summarized in (Table 3). Based on pharmokinetics, out of all the drug molecules, selumetinib and apatinib showed better ADME properties.

Preparation of SOX2 protein for virtual screening

The structure of SOX2 downloaded from the PDB databank is in association with a nucleosome (Fig. 2A). This structure contains SOX2 along with the hetatoms such as the associated nucleosome complex and other ligands. This structural information is useful in understanding the interaction pattern of the SOX2 with the ligands present in its environment. In order to avoid the issue of protein-ligand interactions, the hetatoms were removed from the protein structure and downloaded using the Biovia DS visualizer (Fig. 2B). The structure without the

Table 1 — Analysis of Lipinski's rule of five on the drugs using data from Swiss ADME website.
The molecules which violated more than 1 rule of Lipinski's rule of 5 were excluded from further analysis

S. No	Lipinski's rule of five	Paclitaxel (taxol)	Actinomycin D	Metformin	Ciclesonide	Zoledronic acid	Apatinib	Tamoxifen	Tretinoin	Selumetinib
1	Molecular weight (<500g/mol)	853.91	924.1	129.16	540.69	272.09	397.5	371.5	300.4	457.7
2	Hydrogen bond donor(<5)	4	12	3	1	5	2	0	1	3
3	Hydrogen bond acceptor(<5)	14	18	1	7	8	5	2	2	6
4	Log P(<5)	3.52	0	-1.3	4.78	-4.3	4.1	7.1	6.3	3.6
5	Lipinski's rule of five	No	No	Yes	No	Yes	Yes	Yes	Yes	Yes

Table 2 — A brief overview of the drugs selected for molecular docking with SOX2

Drug	Association with SOX2/metastasis	Targeted pathway	Limitations/Drawback	Reported use in the type of cancer	Reference
Metformin	Metastasis controlling drug	P13K/Akt/mTOR pathways main protein	Multi-target inhibitor rather than a specific target	Cervical cancer	[24]
Apatinib	SOX2 inhibition activity	STAT3 signaling pathway	Apoptosis of the cells occurs due to this drug activity which may also affect the other normal cells	Osteosarcoma, Gastric cancer, Lung cancer, Hepatocellular cancer	[25, 26]
Zoledronic acid	Anti-Metastasis activity	Erk1/2 and Akt pathway	Interferes in cell cycle division as it acts by arresting cell cycle	Cervical cancer	[27]
ATRA drug-Selumetinib	Knockdown SOX2 activity	SLUG and SERPINE 1 pathway	Still in the clinical trials	Breast cancer	[28]
ATRA drug-Tretinoin	Knockdown SOX2 activity	SLUG and SERPINE 1 pathway	Still in the clinical trials	Breast cancer	[28]
Tamoxifen	SOX2 pathway	Wnt 2 signaling pathway, Slow cell cycle down streaming intracellular process.	Chances of recurrence	Breast cancer	[29]

Table 3 — The ADME properties of the drugs i.e. gastrointestinal (GI) absorption, blood brain barrier (BBB) absorption, P-gp substrate and CYP activity and LogK_p using Swiss ADME database

Drug name	GI Absorption	BBB per meant	P-gp substrate	CYP1A2 inhibitor	CYP2C19 inhibitor	CYP2C9 inhibitor	CYP2D6 inhibitor	CYP3A4 inhibitor	Log K_p
Metformin	High	No	No	No	No	No	No	No	-7.84
Zoledronic acid	Low	No	No	No	No	No	No	No	-11.02
Apatinib	High	No	Yes	Yes	Yes	Yes	Yes	Yes	-5.84
Tamoxifen	Low	No	Yes	No	Yes	No	Yes	No	-3.50
Tretinoin	High	Yes	No	Yes	Yes	Yes	No	No	-3.66
Selumetinib	High	No	No	Yes	Yes	Yes	Yes	Yes	-6.54

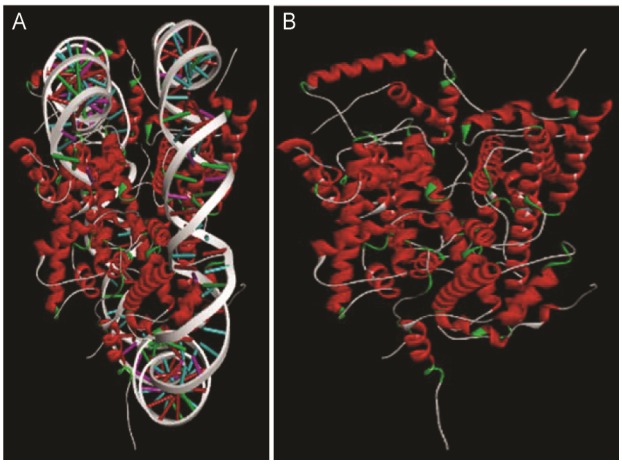


Fig. 2 — The structure of SOX2 protein with the hetatoms (A) and without the hetatoms such as the nucleic acid residues (B) as visualized in the BIOVIA Discovery Studio. The structure with the hetatoms (non-standard atoms) includes water molecules, ions, or other small molecules that are part of the SOX2 protein structure. These hetatoms help us visualize the interactions of the SOX2 in its environment. The structure without the hetatoms focusses solely on the core structure of the SOX2 protein and highlights its intrinsic properties.

hetatoms focusses solely on the core structure of the SOX2 protein and highlights its intrinsic properties.

Preparation of ligand (drug) molecules using ZINC15 database

The structure of the selected ligands based on their ADME properties was extracted from ZINC15 docking online database and is depicted in (Table 4). The ZINC15 database is a comprehensive resource for accessing a vast library of chemical compounds and identifying ligands with favourable ADME properties is crucial in screening for the safest and most effective drug candidates for further research.

Molecular Docking of SOX2 with the ligands

These 6 ligands were subsequently subjected to virtual screening using the PyRx software. PyRx helps in predicting their binding affinities and potential interactions. From the docking analysis, the lowest value of binding affinities, upper RMSD and lower RMSD values of the drugs were considered (Table 5).

Visualization of the final docked structures

From the PyRx results, it was concluded that Apatinib (-7.3 kcal/mol), Selumetinib (-6.7 kcal/mol), Tamoxifen (-5.9 kcal/mol) showed minimum binding affinities compared to the other drug molecules. Apatinib has the best docking score which was linked to the receptor through a amide pi stacked bond which have the large dipole moments with Glutamine at 94 position of E chain and the conventional hydrogen bond with Alanine at 91 position of E chain (Fig. 3A). The drug molecule, Selumetinib, is the second best ligand for anti-SOX2 activity. It was found to interact with the receptor *via* two amide pi bonds with the tyrosine of C chain and isoleucine of D chain at the 50 and 91 positions, respectively. Leucine at 51 position of C chain forms the carbon hydrogen bond and valine at 95 position of C chain shows the conventional hydrogen bonds (Fig. 3B). Tamoxifen with the next better docking results forms two conventional hydrogen bonds with leucine of C side chain and phenylalanine of D side chain at the 51 and the 67 positions, respectively (Fig. 3C). Tretinoin with the better docking score is observed to interact with the receptor as shown in (Fig. 3D) through two conventional hydrogen bonds at threonine and methionine of A side chain at the 107 and 120 positions, respectively. Figure 3E shows the receptor and drug-Zoledronic acid interactions through two conventional hydrogen bonds with alanine at 38, arginine at 45 of B side chain and methionine at 120 position of A side chain. This ligand also showed the unfavorable donor-donor interaction with isoleucine at 119 position of A side chain represented as red dotted lines representing repulsions (Fig. 3E). The amino acid valine at 43 position of B side chain and phenylalanine at 104 position of A side chain forms multiple carbon hydrogen bonds with different oxygen residues of the ligand. Methionine shows the interaction with receptor through four conventional hydrogen bonds with serine, alanine and valine at 18, 21 and 49 positions of side chain C, respectively. It also forms three carbon hydrogen bonds with arginine, phenylalanine and proline at 17, 25 and 48 positions of side chain C, respectively (Fig. 3F).

The docking results established that apatinib has the lowest docking score as well as good drug properties in comparison to the other drugs used in this study. In order to identify novel anti-SOX2 molecules, the drugs similar to apatinib structural

Table 4 — The structure and ZINC IDs of the ligands extracted from the ZINC15 docking software

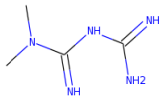
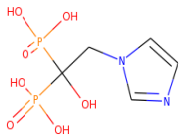
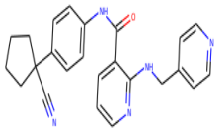
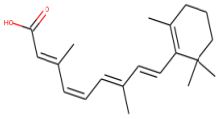
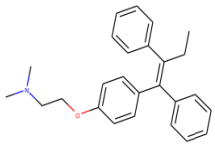
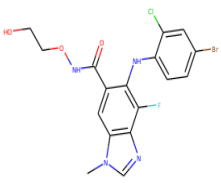
Drug name	ZINC15 ID	2D structure
Metformin	ZINC12859773	
Zoledronic Acid	ZINC3803652	
Apatinib	ZINC70466461	
Tretinoin	ZINC22066351	
Tamoxifen	ZINC1530689	
Selumetinib	ZINC31773258	

Table 5 — The binding affinities of the drugs along with the RMSD values for the SOX2 protein and the individual ligand after docking through PyRx software

Drug name	ZINC 15 ID	Binding affinities(kcal/mol)	Binding energy (kcal/mol)	RMSD Upper bound	RMSD Lower bound
Metformin	ZINC12859773	-4.0	-4.2	19.844	19.404
Zoledronic Acid	ZINC3803652	-4.9	-5.0	3.266	1.472
Apatinib	ZINC70466461	-7.3	-7.5	33.719	30.27
Tretinoin	ZINC22066351	-5.4	-5.5	40.805	38.105
Tamoxifen	ZINC1530689	-5.9	-6.0	37.051	33.901
Selumetinib	ZINC31773258	-6.7	-6.9	36.66	33.804

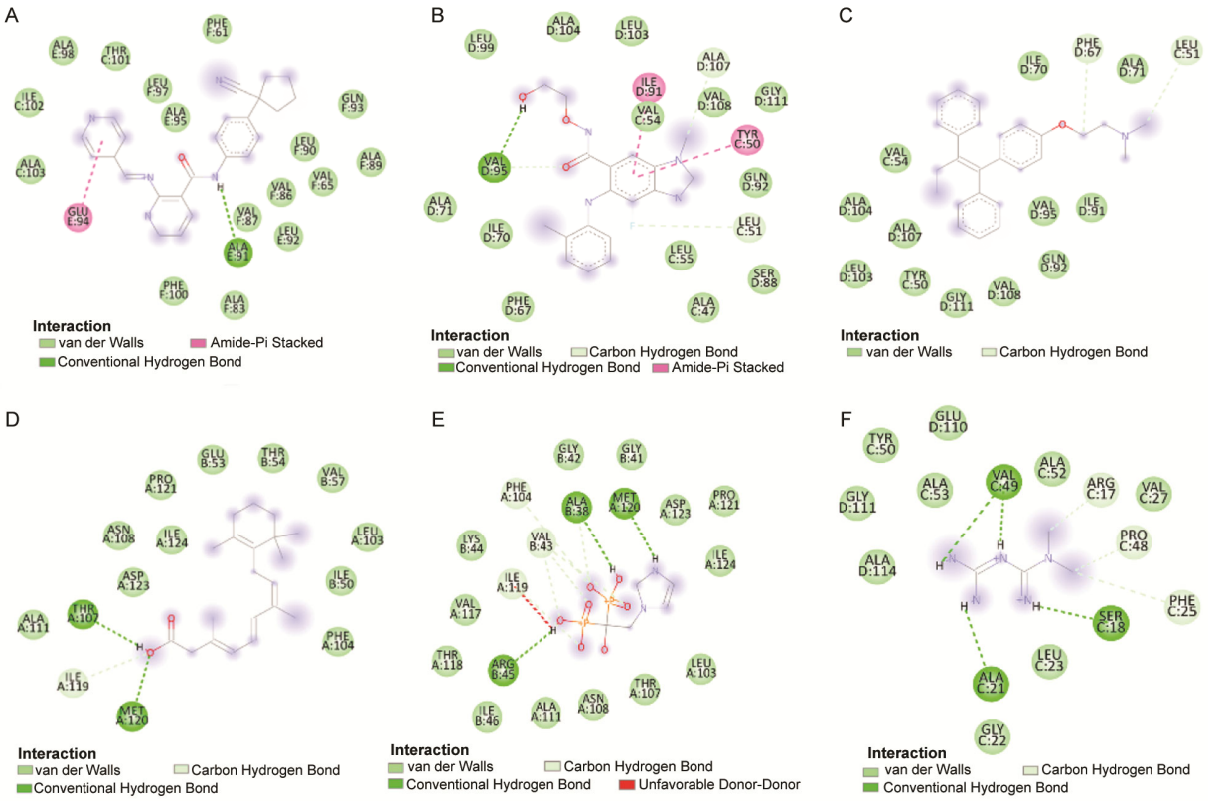


Fig. 3 — The interactions of the ligands with SOX2 protein as visualized in DS visualizer. SOX2 interaction with (A) Apatinib (ZINC70466461); (B) Selumetinib (ZINC31773258); (C) Tamoxifen (ZINC1530689); (D) Tretinoin (ZINC22066351); (E) Zoledronicacid (ZINC3803652); and (F) Metformin (ZINC12859773). These interactions reveal the intricate binding patterns and molecular contact between the ligand and SOX2, thus helping us compare the binding affinity and specificity of the ligands. The visualization only enables us to view how the ligand is binding to SOX2. As per the DS visualizer convention, dark green represents conventional hydrogen bond, light green is used to indicate van der Waals interactions, faint green is used to represent carbon-hydrogen (C-H) interactions and pink color indicates amide π interactions

derivatives were taken from the Drug bank and were analyzed.

Identification of the analogous drug structures for apatinib

The drug apatinib shows the best outcome for the drugs screened based on the binding affinities. The analogous structures for the drug molecule, apatinib, were searched using drug bank database (https://go.drugbank.com/). The structure of two drug molecules, motesanib and N-(4-Phenoxyphenyl)-2-[(Pyridin-4-ylmethyl) amino] nicotinamide were similar to apatinib. These drugs were downloaded in .sdf format and then docked against SOX2 target protein after checking their ADME properties and applying the Lipinski's rule of five. The analogous drugs and their structure along with the drug bank IDs are shown in (Table 6).

Next, the ADME properties and Lipinski's rule of five results for their drug-likeness was carried out for the drug molecules analogous to apatinib (Tables 7 & 8). N-(4-phenoxyphenyl)-2-[(pyridine-4-ylmethyl) amino] nicotinamide showed better ADME properties than motesanib. Also, both the drug molecules follow the Lipinski's rules of five, thus making both of them, orally available.

Table 6 — The 2D structure and Drug Bank ID of the analogous ligands to apatinib

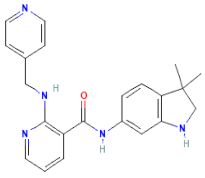
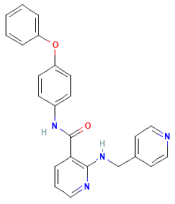
S. No	Drug name	Drug bank ID	2D structure
1.	Motesanib	DB05575	
2.	N-(4-Phenoxyphenyl)-2-[(Pyridin-4-ylmethyl) amino] nicotinamide	DB07183	

Table 7 — The ADME properties of the drugs analogous to apatinib *i.e.* gastrointestinal (GI) absorption, blood brain barrier (BBB) absorption, P-gp substrate and CYP activity using Swiss ADME database

Drug name	GI Absorption	BBB permeant	P-gp substrate	CYP1A2	CYP2C19	CYP2C9	CYP2D6	CYP3A4	Log K_p
Motesanib	High	Yes	Yes	Yes	Yes	Yes	Yes	Yes	-5.83
N-(4-phenoxyphenyl)-2-[(pyridine-4-ylmethyl) amino] nicotinamide	High	No	No	Yes	Yes	Yes	Yes	Yes	-5.52

Table 8 — Analysis of Lipinski's rule of five on the drugs analogous to apatinib obtained from the Swiss ADME website

S. No	Lipinski's rule of five	Motesanib	N-(4-phenoxyphenyl)-2-[(pyridine-4-ylmethyl) amino] nicotinamide
1	Molecular weight(<500 g/mol)	373.46	396.44
2	Hydrogen bond donor(<5)	3	2
3	Hydrogen bond acceptor(<5)	5	4
4	Log P(<5)	4.044	3.09
5	Lipinski's rule of five	Yes	Yes

Table 9 — The binding affinities of the drugs analogous to apatinibalong with RMSD values for SOX2 protein and the individual ligands after docking as revealed by PyRx docking software

Drug name	DRUG BANK ID	Binding affinities (kcal/mol)	Binding energy	RMSD Upper bound	RMSD Lower bound
Motesanib	DB05575	-8.4	-8.6	4.418	1.994
N-(4-Phenoxyphenyl)-2-[(Pyridin-4-ylmethyl) amino] nicotinamide	DB07183	-7.8	-8.0	4.319	2.545

The docking of SOX2 protein (6T7B) was carried out with drug molecules analogous to apatinib (Table 9) which is considered as a reference standard for these drugs. The docking analysis showed that the binding affinities of motesanib and N-(4-Phenoxyphenyl)-2-[(Pyridin-4-ylmethyl) amino] nicotinamide were -8.4 kcal/mol and -7.8 kcal/mol, respectively which is lesser than the binding affinity of apatinib *i.e.* -7.3 kcal/mol. Their interactions with the target protein, SOX2, showed that they are forming the carbon hydrogen bonds between the certain amino acids of protein and O residues of the drug molecules. These carbon-hydrogen bonds are the covalent bonds that makes the interaction stable as the carbon shares its outer valence electrons to attain stability. Interactions were visualized using the DS visualiser as shown in (Fig. 4A & B).

Normal Mode Analysis (NMA) of SOX2 with ligands

The NMA analysis was carried out for the docked complexes of SOX2 with ligands-Motesanib and N-(4-Phenoxyphenyl)-2-[(Pyridin-4-ylmethyl) amino] nicotinamide (Fig. 5A). Both the ligands showed similar values after binding with SOX2. The peaks in the deformability graph (Fig. 5B) indicate the protein regions which are showing flexibility and assists in deformation. B-factor/mobility graph tells the rate at which the main chain is getting deformed (Fig. 5C). The grey colored portion of the graph is the experimental value taken from the respective PDB field which is the average of RMS values. The red colored portion of the graph is the NMA calculated value obtained by multiplying NMA mobility and the $8pi^2$. The difference in the B-factor for the PDB and NMA values (Fig. 5C) is because of the change in protein structure to attain a particular conformation for ligand binding during docking. The eigen value depicts the stiffness of the protein in various modes (Fig. 5D). If the eigen value is less, it indicates easy deformation. The eigen value for SOX2 and ligand complex is 9.52e-07 which indicates the stability of protein-ligand interaction. The variance plot in Figure 5E shows the variances with each normal mode. The variance is inversely proportional to the respective eigen value. The cumulative variances are shown in green and individual variances are shown in blue. The covariance map depicts the correlation among the different residues of SOX2 which indicates if the protein residues exhibit correlated (red), uncorrelated (white) or anti-correlated motion (blue). Most of the paired residues of SOX2 move in a correlated motion, thereby implying a stable ligand-SOX2 complex (Fig. 5F). The elastic network model for ligand-SOX2 interactions shows limited flexibility in SOX2 protein (Fig. 5G). Each grey dot in the graph represents the stiffer springs between the pair of atoms of SOX2.

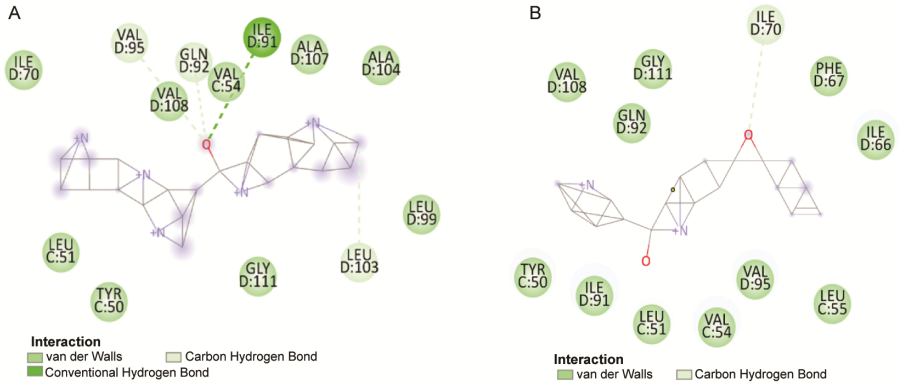


Fig. 4 — The interactions of ligands with SOX2 protein as visualized in DS visualizer. (A) Motesanib makes the carbon-hydrogen bond with glutamine at 92 position, valine at 95 position, leucine at 103 position and isoleucine at 91 position of D chain of SOX2 protein. But when viewed for the ligand receptor attachment, it is found not being attached to any of the chains of the protein other than forming bonds; and (B) N-(4-Phenoxyphenyl)-2-[(Pyridin-4-ylmethyl) amino] nicotinamide makes the carbon-hydrogen bonds with isoleucine at 70 position of D chain of SOX2 protein

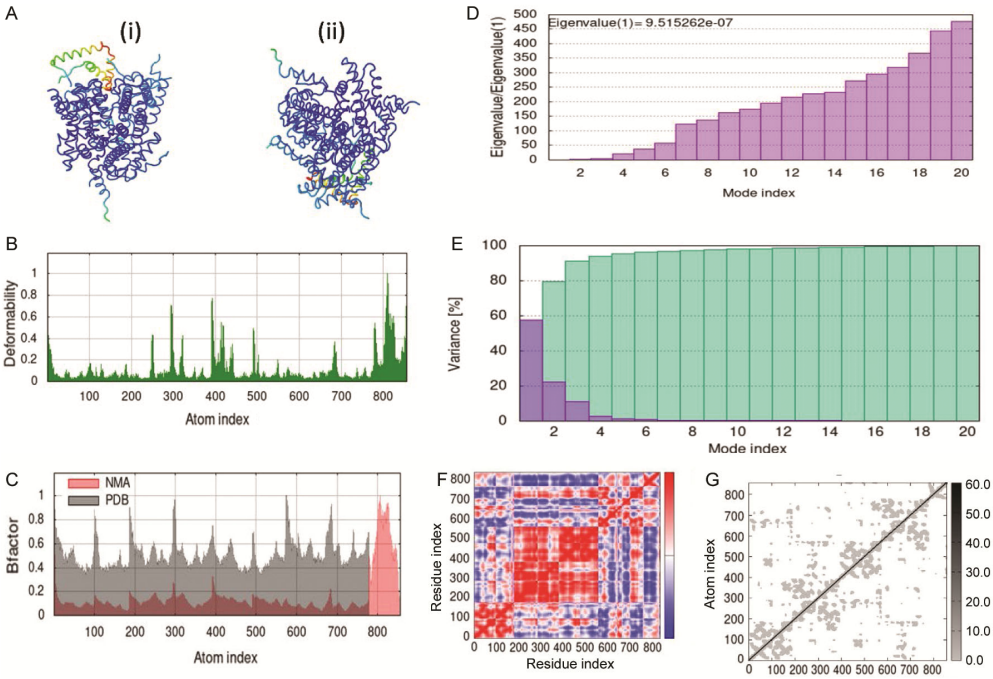


Fig. 5 — NMA analysis of SOX2 with ligands in iMODS. (A) The docked complex of SOX2 with i) Motesanib ii) N-(4-Phenoxyphenyl)-2-[(Pyridin-4-ylmethyl) amino] nicotinamide; (B) Deformability graph- It shows how different atoms within a protein-ligand complex respond to mechanical stress. Atom index represents the atoms in a sequential manner in a protein ligand complex and deformability indicates how much the said atom moves or deforms upon experiencing stress. The atoms with high deformability are deemed to be more flexible; (C) B-factor/motion plot- It is a measure of mobility of atoms within a protein-ligand complex. This means that atoms with lower B-factors at the binding site would imply stronger and more stable interactions with the ligand; (D) Eigen value plot- The eigen value plot highlights the importance of each mode of motion in the protein-ligand complex. The mode with low eigen values shows the rigid or small movements in the complex and high eigen values shows the bigger movements which may be responsible for its biological function. Mode index on the x-axis shows the order of modes (pattern of motion like collective motion or localized motion of atoms), starting from lowest (low energy) to the highest (high energy) mode; (E) Variance graph- The variance graph shows how much variance each mode contributes to the overall motion of the complex; and (F) Covariance map- This map shows the extent of coordinated movements of atoms in the complex. The atoms showing positive covariance indicate movement in the same direction and the ones showing negative covariance indicate movement in the opposite direction. The zero covariance highlights no correlation in the movement of these atoms. In the graph, red indicates positive covariance, blue indicates negative covariance and white indicates no covariance; and (G) Elastic network model- This model depicts the simplified view of the complex dynamics of protein-ligand interaction. The protein ligand complex is represented by the atoms connected by springs. Each dot shows one spring between the respective atoms and the colour intensity of the dots highlight stiffness of the spring.

Conclusion

Out of nine compounds retrieved from the drug bank based on their reported anti-SOX2 or anti-metastatic activity using literature search, only six drugs satisfied Lipinski's rule of five which were selected for further analysis. Upon docking, the first three modes of all the ligand structures were used in order to compare their binding affinities. The overall most effective binding of SOX2 protein was shown by apatinib (-7.3 kcal/mol), selumetinib (-6.7 kcal/mol) and tamoxifen (-5.9 kcal/mol). From this analysis, it was concluded that apatinib seems to be the most effective inhibitor of SOX2 among the considered drugs. Apatinib was thus considered as a standard drug molecule for anti-SOX2 activity and drug molecules with structure similar to apatinib were searched in drug data bank to identify novel SOX2-inhibitors. This search identified two such drug molecules, Motesanib and N-(4-Phenoxyphenyl)-2-[(Pyridin-4-ylmethyl) amino] nicotinamide. Both these drugs displayed better binding affinities with SOX2 than apatinib and between the two, N-(4-Phenoxyphenyl)-2-[(Pyridin-4-ylmethyl) amino] nicotinamide showed better docking with SOX2 than motesanib. And after docking, when visualized for the interactions, the ligand, motesanib, was not bound to any of the chains which limits its protein-ligand binding properties, even though it shows good docking affinity. In comparison to apatinib for the ADME properties, N-(4-Phenoxyphenyl)-2-[(Pyridin-4-ylmethyl) amino] nicotinamide acts as the inhibitor of p-gp and also, the $\log K_p$ is less for this drug molecule. This implies that N-(4-Phenoxyphenyl)-2-[(Pyridin-4-ylmethyl) amino] nicotinamide could be better at inhibiting SOX2 protein than apatinib. Interestingly, the visuals of SOX2-ligand interactions shows that most of the anti-SOX2 ligands were bound to amino acids of the D chain, which indicates the significance of D chain as a crucial docking site for SOX2 inhibition. The NMA analysis further showed that the binding of anti-SOX2 ligands alters the conformation of SOX2 which would, perhaps, prevent the interaction of SOX2 with other proteins or nucleic acids. In conclusion, this study has shown that among the existing anti-cancer drugs, apatinib seems to be the most potent anti-SOX2 inhibitor. Further screening for molecules with analogous structure to apatinib shows that N-(4-Phenoxyphenyl)-2-[(Pyridin-4-ylmethyl) amino] nicotinamide could be an even better inhibitor of SOX2 protein. However, experimental validation of its anti-SOX2 effect is warranted in future studies.

Acknowledgement

The authors acknowledge the financial support received from Department of Science and Technology, Govt. of India, New Delhi (DST-FIST-SR/FST/LS-I/2018/125©) under the DST-FIST program to Department of Biochemistry, Central University of Punjab, Bathinda. We also acknowledge the critical inputs from Dr Sanjiv Kumar for this study.

Conflict of interest

All authors declare no conflict of interest.

References

1 Yu Z, Pestell TG, Lisanti MP & Pestell RG, Cancer stem cells. *Int J Biochem Cell Biol*, 44 (2012) 2144.

2 Lawson JC, Blatch GL & Edkins AL, Cancer stem cells in breast cancer and metastasis. *Breast Cancer Res Treat*, 118 (2009) 241.

3 Santos P & Almeida F, Role of exosomal miRNAs and the tumor microenvironment in drug resistance. *Cells*, 9 (2020) 1450.

4 Matsui WH, Cancer stem cell signaling pathways. *Medicine*, 95 (2016) S8.

5 Lefebvre V, Dumitriu B, Penzo-Méndez A, Han Y & Pallavi B, Control of cell fate and differentiation by Sry-related high-mobility-group box (Sox) transcription factors. *The Int J Biochem Cell Biol*, 39 (2007) 2195.

6 Pouremamali F, Vahedian V, Hassani N, Mirzaei S, Pouremamali A, Kazemzadeh H, Faridvand Y, Jafari-gharabaghlou D, Nouri M & Maroufi NF, The role of SOX family in cancer stem cell maintenance: With a focus on SOX2. *Pathol Res Pract*, 231 (2022) 153783.

7 Lundberg IV, Lofgren Burstrom A, Edin S, Eklof V, Oberg A, Stenling R, Palmqvist R & Wikberg ML, SOX2 expression is regulated by BRAF and contributes to poor patient prognosis in colorectal cancer. *PLoS One*, 9 (2014) e101957.

8 Lundberg IV, Edin S, Eklof V, Oberg A, Palmqvist R & Wikberg ML, SOX2 expression is associated with a cancer stem cell state and down-regulation of CDX2 in colorectal cancer. *BMC Cancer*, 16 (2016) 471.

9 Liu XF, Yang WT, Xu R, Liu JT & Zheng PS, Cervical cancer cells with positive Sox2 expression exhibit the properties of cancer stem cells. *PLoS One*, 9 (2014) e87092.

10 Chung JH, Jung HR, Jung AR, Lee YC, Kong M, Lee JS & Eun YG, SOX2 activation predicts prognosis in patients with head and neck squamous cell carcinoma. *Sci Rep*, 8 (2018) 1677.

11 Leis O, Eguiara A, Lopez-Arribillaga E, Alberdi MJ, Hernandez-Garcia S, Elorriaga K, Pandiella A, Rezola R & Martin AG, Sox2 expression in breast tumours and activation in breast cancer stem cells. *Oncogene*, 31 (2012) 1354.

12 Bareiss PM, Paczulla A, Wang H, Schairer R, Wiehr S, Kohlhofer U, Rothfuss OC, Fischer A, Perner S, Staebler A,

Wallwiener D, Fend F, Fehm T, Pichler B, Kanz L, Quintanilla-Martinez L, Schulze-Osthoff K, Essmann F & Lengerke C, SOX2expression associates with stem cell state in human ovarian carcinoma. *Cancer Res*, 73 (2013) 5544.

13 Voigt E, Wallenburg M & Wollenzien H, Thompson E, Kumar K, Feiner J, McNally M, Friesen H, Mukherjee M, Afeworki Y & Kareta MS, Sox2 is an oncogenic driver of small-cell lung cancer and promotes the classic neuroendocrine subtype. *Mol Cancer Res*, 19 (2021)2015.

14 Mamun MA, Mannoor K, Cao J, Qadri F & Song X, SOX2 in cancer stemness: tumor malignancy and therapeutic potentials. *J Mol Cell Biol*, 12 (2020) 85.

15 Sharma S & Shukla VJ, Molecular docking study of six pyrimidine derivatives as EGFR (Epidermal Growth Factor Receptor) and Ca IX (Carbonic Anhydrase IX) inhibitor. *Int J Pharm Pharm Sci*, 66 (2019) 71.

16 Dallakyan S & Olson AJ, Small-molecule library screening by docking with PyRx. In: *Chemical Biology* (Ed. by Hempel JE, Williams CH & Hong CC; Springer New York, New York, USA), 2015, 243.

17 Irwin JJ & Shoichet BK, ZINC − A free database of commercially available compounds for virtual screening. *J Chem Inf Model*, 45 (2005) 177.

18 Pietrobono S, De Paolo R, Mangiameli D, Marranci A, Battisti I, Franchin C, Arrigoni G, Melisi D, Poliseno L & Stecca B, p38 MAPK-dependent phosphorylation of transcription factor SOX2 promotes an adaptive response to BRAF inhibitors in melanoma cells. *Journal of Biological Chemistry*, 298 (2022) 102353.

19 Coleman N & Rodon J, Taking aim at the undruggable, in *American Society of Clinical Oncology Educational Book*, 2021, e145.

20 Dykeman EC & Sankey OF, Normal mode analysis and applications in biological physics. *J Phys: Condens Matter*, 22 (2010) 423202.

21 Lopez-Blanco JR, Aliaga JI, Quintana-Ortí ES & Chacon P, iMODS: internal coordinates normal mode analysis server. *Nucleic Acids Res*, 42 (2014) W271.

22 Ghazwani MY, Bakheit AH, Hakami AR, Alkahtani HM & Almehizia AA, Virtual screening and molecular docking studies for discovery of potential RNA-dependent RNA polymerase inhibitors. *Crystals*, 11 (2021) 471.

23 de Sousa ACC, Combrinck JM, Maepa K & Egan TJ, Virtual screening as a tool to discover new β-haematin inhibitors with activity against malaria parasites. *Sci Rep*, 10 (2020) 3374.

24 Xia C, Liu C, He Z, Cai Y & Chen J, Metformin inhibits cervical cancer cell proliferation by modulating PI3K/Akt-induced major histocompatibility complex class I-related chain A gene expression. *J Exp Clin Cancer Res*, 39 (2020) 127.

25 Tian ZC, Wang JQ & Ge H, Apatinib ameliorates doxorubicin-induced migration and cancer stemness of osteosarcoma cells by inhibiting Sox2 *via* STAT3 signalling. *J Orthop Translat*, 22 (2020) 132.

26 Zhang XH, Cao MQ, Li XX & Zhang T, Apatinib as an alternative therapy for advanced hepatocellular carcinoma. *World J Hepatol*, 12 (2020) 766.

27 Wang L, Liu Y, Zhou Y, Wang J, Tu L, Sun Z, Wang X & Luo F, Zoledronic acid inhibits the growth of cancer stem cell derived from cervical cancer cell by attenuating their stemness phenotype and inducing apoptosis and cell cycle arrest through the Erk1/2 and Akt pathways. *J Exp Clin Cancer Res*, 38 (2019) 93.

28 Bhat-Nakshatri P, Goswami CP, Badve S, Sledge GW & Nakshatri H, Identification of FDA-approved drugs targeting breast cancer stem cells along with biomarkers of sensitivity. *Sci Rep*, 3 (2013) 2530.

29 Piva M, Domenici G, Iriondo O, Rabano M, Simoes BM, Comaills V, Barredo I, Lopez-Ruiz JA, Zabalza I, Kypta R & Vivanco MdM, Sox2 promotes tamoxifen resistance in breast cancer cells. *EMBO Mol Med*, 6 (2014) 66.