IN SILICO SCREENING OF MAJOR CANCER DRUG TARGETS (GROWTH FACTOR RECEPTORS) FOR NATURE DERIVED PHYTOCHEMICALS

[1]Tithi S. Trivedi, [2]Sivakumar Prasanth Kumar , [1]Archana U. Mankad*, [1]Saumya K. Patel, [2]Rakesh M. Rawal and [1,2]Himanshu A. Pandya

[1]Department of Botany, Bioinformatics and Climate Change Impacts Management,
[2]Department of Life Science, School of Sciences, Gujarat University, Ahmedabad-380 009.
Corresponding Author Email-ID: aumankad@gujaratuniversity.ac.in

ABSTRACT

Cancer is a group of abnormal cells. The unregulated growth factor receptor tyrosine kinase (GFR-TK) proteins are implicated in the proliferation of more than 60% of all cancer types. Screening of phytochemicals for their anti-angiogenic potential has been a growing area of research in the modern decade. There is a well-known principle that natural compounds are active against several diseases, including various types of cancer. The present research work focuses on known growth factor receptors (GFRs) as an important target for computational studies. In this study, 96 curated anti-cancer compounds were virtually screened against the EGFR, FGFR, IGFR, and HGFR using molecular docking software. For each GFR, we have considered ten top most results as potential hits. Among them, common five results are: Spirosolane, Ginkgetin, Fangchinoline, Theaflavin and Ursolic acid. These compounds have been reported to show anticancer activities in the literature. With the help of different interaction analysis tools, the protein-ligand interaction patterns between the functional groups of these compounds were analyzed. Hydrogen bonding and hydrophobic forces are the main components of the interactions of these hits, similar to those experimental for the known inhibitors. From the maximum number of hits, it could be indicated that compounds Spirosolane, Ginkgetin, Fangchinoline, Theaflavin and Ursolic acid are promiscuous leads in the drug discovery process.

Keywords *: Growth Factor Receptors, Tyrosine Kinase, Virtual screening, Structure-based drug design, Cancer, Phytochemicals*

INTRODUCTION

According to WHO cancer fact sheets, cancer is the second leading cause of death globally and is responsible for an estimated 9.6 million deaths in 2018. Globally, about 1 in 6 deaths is due to cancer (ref link: http://www.who.int/news-room/fact-shects/detail/cancer). The burden of cancer have-been predicted to be increasing globally up-to 15 million by the year 2020, wherein the developing countries would be at a higher risk relating to incidences of cancer morbidity and mortality[1].Growth factor receptors (GFRs) expressed on cell membranes or in the cytoplasm, haveprofound roles in cell growth, survival angiogenesis, and metastasis. They contain three domains: an extracellular ligand (growth-factor) binding domain, a transmembrane domain, a cytoplasmic domain that acts as an enzyme, or forms a complex with another protein that acts as an enzyme[2]. Here we have discussed the "*In-Silico*" screening of major cancer drug targets – growth factor receptors (EGFR, FGFR, IGFR, and HGFR) against nature-derived phytochemicals.There are several efficient anticancer drugs and active inhibitors against several protein targets, growing resistance attached with many side effects

mean that there is a need for new, enhanced treatments[3].Plants are important sources of the active ingredients used in modern medicines. More than half ofthe drugs approved since 1995 are based on natural products[4-6].

EGFR has been a chief target of atomic anti carcinoma treatment. Cellular levels of EGFR do not always associate with the reply to the EGFR TK slow downers, indicating acquired resistance to drugs[7]. Combination of EGFR antagonists with slow downers targeting different signaling mechanism (s) – such as IGFR and VGFR – that share the same downstream mediator may get around or hold – up the development of resistance to EGFR antagonists' resultant in enhanced antitumor activities[7].FGFR over expression, point mutation, or gene fusions is found in 80% of non – muscle – invasive and 15% of muscle-invasive bladder carcinoma[8]. High expression levelsof the IGF-1R have been found in breast and colorectal carcinoma. The most significant connection. Between augmented levels of IGF-1 and the risk of carcinoma, the diagnosis was found for prostate carcinoma, premenopausal breast carcinoma, and colorectal carcinoma[9].The HGF was first recognized as a solvable factor indorsing hepatocyte growth and liver renewal. The HGF receptor (MET) is a prototypal TK receptor. The remark that MET is a proto-oncogene and that its signaling is frequently conquered in carcinoma has permitted deep investigations on its atomic structure and signaling properties that are deliberated. In carcinoma, HGFR has been concerned with cellular expansion, metastasis, and angiogenesis[10].The availability of GFRs structures implies that VS could be used as a tool to search for potential active compounds from medicinal herbs. The structures of the active compounds found in the folk – medicinal herbs have been collected in thePubChem database[11]. We performed a virtual screening protocol using the YASARA server.Theywere applied to screen the 96 curated collection of natural molecules. The resulting hits were analyzed to gain insights into the key structural features required for good protein-ligand interaction.

Materials and methods

Protein preparation

As the selection of target proteins, EGFR, FGFR, IGFR, and HGFR, their proteins have been selected. Protein structures are available for these proteins. Based on low resolution, targets have been selected. Preparation of protein before docking was done by deleting water molecules and clearing the protein in the YASARA structure followed by energy minimization using the same.

Ligand preparation

NPACT database is an important database related to the plant-derived natural anticancer compounds[12]. From this database, 96 natural compounds have been selected, which were further compiled into data-sets and used for the docking purpose. These ligands were cleaned and the hydrogen was added to the ligands for the further procedure. Their 2D structures were converted to 3D using BIOVIA Discovery Studio[13].

Allicin	Gallic Acid	Embelin	flavonol 3-O-glycoside	Combretastatins	Mangiferin
Allyl isothiocyanate	Genistein	Nitidine choride	Damnacanthol	(E)-γ-bisabolene	Quercetin- 3-gluconide
Andrograph aloidD	Gingerol	Protopine	Mallato phenone (1)	Emodin tetraacetate	Costunolide
Apigenin	Gossypol	Psoralen	Podophyllotoxin bromide	Cynaropicrin	Quassinoid
Apiole	Hesperidin	Aloe-emodin	Diterpenes(aconitine)	Dithymoquinone	Ginkgolides B
AntineoplastoneA10	Indole-3-carbinol	Thymoquinone	Lupeol	Psorospermin	RubescensineJ
Capsaicin	Kaempferol	Cicutoxin	Withaferin A	Berberine	Germacrone
	Methyl				

https://iabcd.org

Catechin	allyl trisulfide	Epipodophyllotoxin	Spartine	Chelidonine	Podophyllotoxin
Chlorogenic Acid	Piperine	Elliptinium	Homoharringtonine	Ginkgetin	Spirosolane
Coumarin	Quercetin	Ursolic acid	Cardiac glucosides (Ouabain)	Galangin	Bavachinin
Coumestrol	Resveratrol	Quassin	Bruceantin	Hyperoside	Curdione
Curcumin	Silymarin	PinocembrinTetrandrine	WithanolideD	Rhamnetin	Daidzein
Cyanidin	Theaflavin	Neohesperidin	Oridonin	Psoralidin	Andrographolide
Diallyl sulphide	Zeaxanthin	Ginsenosides C-K	Fangchinoline	Curcumol	Trichothecenes
Ellagic Acid	Naringenin	10-Gingerols	Dehydrocostus lactone	Germacranolide	Ginkgolides A
Epigallocatechin gallate(EGCG)	Bergenin	Rubiadin	Xanthone	LicoagrochalconeA	Ponicidine

Table 1: 96 curated phytochemicals from NPACT database

Virtual screening proceduresMolecular Docking

YASARA Structure (version 20.7.1) has been utilized for the protein-ligand docking urpose[14]. It is based on the Autodock vina algorithm and it utilizes the following formula for calculating the docking score

$$\Delta G = \Delta G_{(vdw)} + \Delta G_{(HBond)} + \Delta G_{(elec)} + \Delta G_{(tor)} + \Delta G_{(desolv)}$$

Wherein, $\Delta G_{(vdw)}$ is the component energy terms related to van der wals bond, $\Delta G_{(HBond)}$ is the component energy term related to Hydrogen bonds, $\Delta G_{(elec)}$ is the component energy terms related to electrostatics, $\Delta G_{(tor)}$ is the component energy term related to the ligand's torsional free energy and $\Delta G_{(desolv)}$ is the component energy term related to the desolvation for the empirical calculation of the docking/binding energy for a protein-ligand complex. The Higher docking score represents the better protein-ligand binding whereas, the negative score represents no binding between the Protein and the ligand. The protein-ligand interactions were further visualized in 3D and 2D using the BIOVIA Discovery Studio Visualizer.
Post – docking analysis

We used several tools to analyze the interactions between the EGFR, FGFR, IGFR, and HGFR andhit compounds. The most important interactions of the protein and ligand compare from the docked results were calculated and the hit compounds were further analyzed using BIOVIA DS[13] the PyMol[15.]
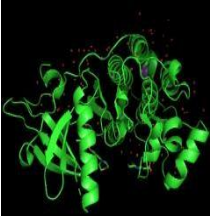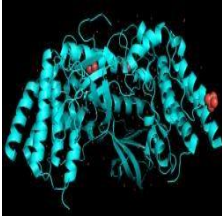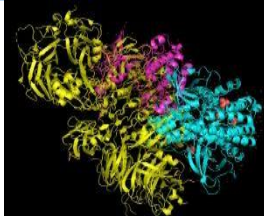Validation of the results through Re-docking

Re-docking is the process of docking in which the receptor has been docked with its natural ligand.By doing this, we could check the behavior of another ligand, whether they are as similar to a natural ligand or not. In case of the absence of natural ligand, then its slow – downer has been selected for re – docking, it is called "Reference – ligand".

RESULTS AND DISCUSSION

Characterization of selective growth factor receptors as potential drug targets

Based on the low-resolution value, the targets for docking purposes have been selected from the PDB database[16]. The detailed information of targets has been given below.

| EGFR	FGFR	IGFR	HGFR in complex with Internalin B
PDB ID:4WKQ Resolution:1.85 Å Chain: ASequence length: 330	PDB ID: 2PSQ Resolution:2.4 Å Chain: A, BSequence length: 370	PDB ID: 5FXQ Resolution:2.3 Å Chain: ASequence length: 308	PDB ID:2UZX Resolution:2.8 Å Chain: B, D Sequence length: 727

Table 2: Growth Factor Receptors as protein targets with detail

Molecular Docking

The molecular docking studies expose the binding affinity of the natural ligands towards the GFRsprotein targets. Different types of interactions such as hydrogen bond interactions, pi – sigma bond,alkyl bond, and pi – alkyl bond were observed in the protein-ligand interaction complex. Key contacting receptor residues were also noted. The top three protein-ligand complexes with the respective binding energies, dissociation constants, and the number of hydrogen bonds for four protein targets are noted in Table – 3, Table – 4, Table – 5, and Table – 6. Better binding of the ligand towards the protein chain is indicated by the higher positive binding energy whereas the negative binding energies indicate no binding[17].

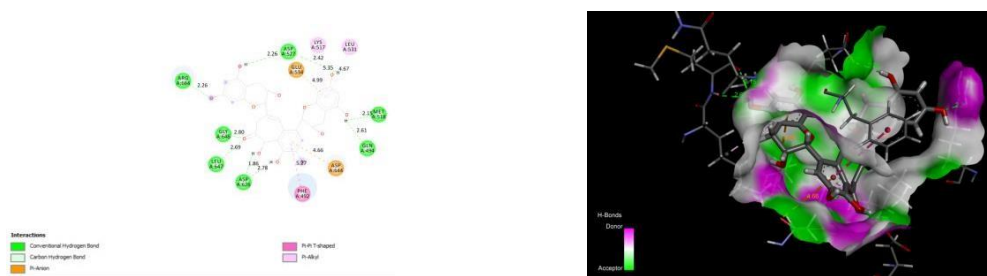| Sr. Num.	Ligand Name	BindingEnergy [kcal/mol]	No. of H-bonds	Contacting Receptor Residues
1	**Fangchinoline**	8.9900	1	LEU 718, GLY 719, SER 720, GLY 724, VAL 726, ALA 743, LYS 745, SER 752, MET 766, THR 790, MET 793, GLY 796, CSX 797, ASP 800, GLU 804,ARG 841, ASN 842, LEU 844, THR 854, ASP 855
2	**Psoralidin**	8.8240	2	LEU 718, VAL 726, ALA 743, LYS 745, THR 790, GLN 791, LEU 792, MET 793, PRO 794, PHE 795 GLY 796, CSX 797, ASP 800, GLU 804,LEU 844, THR 854, ASP 855
3	**Ursolic Acid**	8.8020	2	LEU 718, GLY 719, VAL 726, ALA 743, LYS 745, THR 790, GLN 791, LEU 792, MET 793, GLY 796, CSX 797, LEU 799, ASP 800,ARG 803, ARG 841,ASN 842, LEU 844, THR 854, ASP 855

Table 3: Top 3 Molecular Docking Results of phytochemicals with EGFR
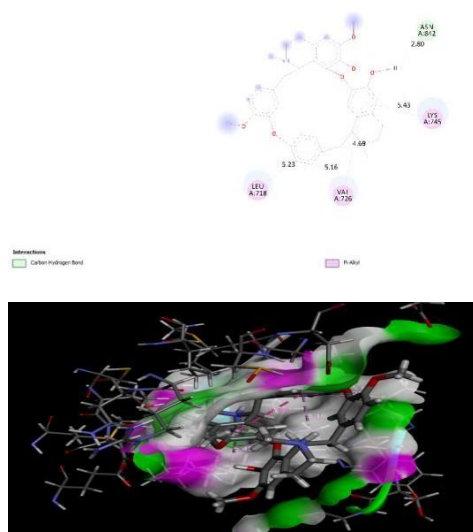
Figure: 1 2D and 3D Protein ligand complex representation of EGFR protein with ligand Fangchinoline

Sr. Num.	Ligand Name	Binding Energy [kcal/mol]	No. of H-bonds	Contacting Receptor Residues
1	Theaflavin	10.8310	7	PHE 492, GLY 493, GLN 494, VAL 495, LYS 517, MET 518, LEU 519, LYS 526, ASP 527, ASP 530, LEU 531, GLU 534, ASP 626, ARG 630, ASN 631, ASP 644, GLY 646, LEU 647, ARG 664
2	Spirosolane	10.3530	1	GLY 488, GLU 489, GLY 490, CYS 491, PHE 492, GLY 493, VAL 495, LYS 517, ASP 527, ASP 530, LEU 531, GLU 534, ASN 571, ARG 630, LEU 633, ASP 644, GLY 646, LEU 647, ARG 664
3	Silymarin	10.0680	5	GLY 488, GLU 489, GLY 490, CYS 491, PHE 492, GLY 493, VAL 495, LYS 517, ASP 527, ASP 530, LEU 531, GLU 534, ASN 571, ARG 630, LEU 633, ASP 644, GLY 646, LEU 647, ARG 664

Table 4: Top 3 Molecular Docking Results of phytochemicals with FGFR

Figure: 2 2D and 3D Protein ligand complex representation of FGFR protein with ligand Theaflavin

Sr. Num.	Ligand Name	Binding Energy [kcal/mol]	No. of H-bonds	Contacting Receptor Residues
1	Fangchinoline	9.2850	4	GLU 1046, ASN 1049, GLU 1050, VAL 1053, PHE 1131, VAL 1132, HIS 1133, ARG 1134, ASP 1135, PHE 1154, GLY 1155, ARG 1158, LEU 1173, LEU 1174
2	Theaflavin	8.8080	4	LEU 1005, GLY 1006, GLN 1007, VAL 1013, ALA 1031, LYS 1033, VAL 1063, MET 1079, GLU 1080, LEU 1081, MET 1082, THR 1083, ARG 1084, GLY 1085, ASP 1086, SER 1089, MET 1142, ASP 1153, MET 1156, THR 1157, ILE 1160
3	Ellagic Acid	8.6380	3	LEU 1005, GLY 1006, GLN 1007, GLY 1008, VAL 1013, ALA 1031, LYS 1033, VAL 1063, MET 1079, GLU 1080, LEU 1081, MET 1082, GLY 1085, MET 1142, ASP 1153, MET 1156, ILE 1160

Table 5: Top 3 Molecular Docking Results of phytochemicals with IGFR

Figure: 3 2D and 3D Protein ligand complex representation of IGFR protein with ligand Fangchinoline

Sr. Num.	Ligand Name	BindingEnergy [kcal/mol]	No. of H-bonds	Contacting Receptor Residues
1	Theaflavin	10.0580	10	ASN 54, VAL55, ILE 56, LEU 57, HIS 58, MET 118, ALA 119, LEU 120, VAL 121, ALA 182, LYS 183, VAL 184, HIS 251, PHE 253, ALA 320, TYR 321, VAL 322, PRO 488, GLU 489, VAL 490, ILE 491, VAL 492, GLU 493
2	Withanolide	9.2110	3	GLN 53, ASN 54, ILE 116, ASN 117, MET 118, GLY 181, ALA 182, LYS 183, LYS 248, TYR 249, VAL 250, VAL 264, ILE 316, LEU 317, GLN 318, PRO 356, THR 440, PRO 488, GLU 489
3	*Fangchinoline*	9.1470	5	ASN 54, VAL 55, ILE 56, MET 118, ALA 119, LEU 120, VAL 121, VAL 122, VAL 184, LEU 185, SER 186 ,SER 187, HIS 251, ALA 252, PHE 253, GLU 254, TYR 321, VAL 322, PRO 488, GLU 489, VAL 490, ILE 491, VAL 492

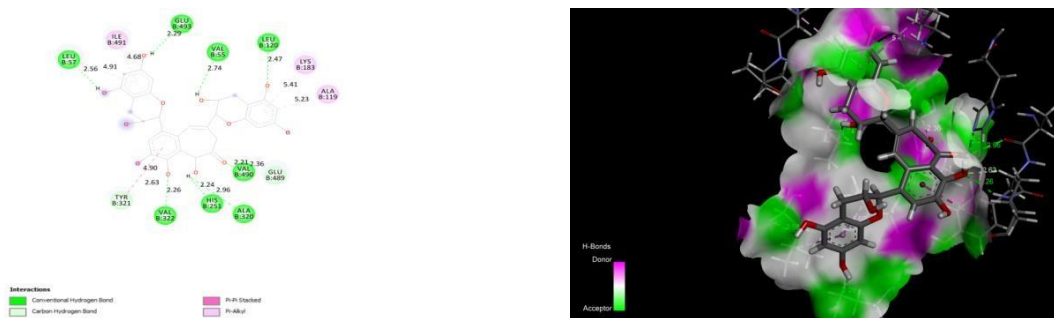**Table 6: Top 3 Molecular Docking Results of phytochemicals with HGFR**

Figure: 4 2D and 3D Protein ligand complex representation of HGFR protein withligand Theaflavin

Different families of growth factors and growth factor receptors are involved in the autonomous growth of cancer cells[18]. Fangchinoline, Psoralidin, and Ursolic acid were the three phytochemicals that achieved better binding with EGFR. They developed a single, two, and two hydrogen bounds respectively. It has been observed to be interacting with the protein EGFR with the binding energies of 8.990 kcal/mol, 8.824 kcal/mol, and 8.802 kcal/mol, respectively. The protein-ligand complex of EGFR with the topmost ligand Fangchinoline has shown the hydrogenbond interacting with the protein residue ASN842. For FGFR, Theaflavin, Spirosolane, and Silymarin ligands were achieved better binding. They showed binding energies of 10.891 kcal/mol,10.3530 kcal/mol, and 10.068 kcal/mol, respectively with the protein FGFR. Theaflavin (the topmost ligand) has shown seven hydrogen bonds, they are ARG664, GLY646, LEU647, ASP626,GLN494, MET518 and ASP527. Again, Fangchinoline, Theaflavin, and Ellagic acid were achieved better binding with IGFR. They developed a single,

four and four hydrogen bonds, respectively. For Fangchinoline, the single hydrogen bond is ARG1158. It has been observed to be interacting with the protein IGFR with the binding energies of 9.285 kcal/mol, 8.808 kcal/mol,and 8.638 kcal/mol respectively. For HGFR, Theaflavin, Withanolide, and Fangchinoline were achieved better binding. They showed binding energies of 10.058 kcal/mol, 9.211 kcal/mol, and

9.147 kcal/mol respectively, with the protein HGFR. The topmost ligand, Theaflavin has shown ten hydrogen bond interactions. They are LEU57, GLU493, VAL55, LEU120, TYR321, VAL322,HIS251, ALA320, VAL490, and GLU489. Spirosolane was interacting with the protein EGFR with the binding energy of 8.712 kcal/mol, with FGFR protein with the binding energy of 10.353kcal/mol, with IGFR protein with the binding energy of 8.476 kcal/mol and with HGFR protein with the binding energy of 8.5470 kcal/mol. Ginkgetin was also showing the interaction with all four proteins- EGFR, FGFR, IGFR and HGFR with different binding energies 8.717 kcal/mol, 9.842 kcal/mol, 8.617 kcal/mol, 8.986 kcal.mol respectively. Ursolic acid was also showing the interaction with proteins FGFR and HGFR with the binding energy of 9.729 kcal/mol and 8.69 kcal/mol respectively.

CONCLUSION

Our results indicated that the five phytochemicals, obtained as hits using YASARA screening method could all be inhibitors of selective growth factor receptors. This suggests that YASARA molecular docking procedure can be effectively used to predict the binding modes of NPACT compounds. These small molecules were obtained from fruits and vegetables that are accessible locally, and they have been used as components of primitive medicinal recipes. Anticancer activities were reported for all five of the hit compounds, which mean they could be potential leadsin the drug discovery in the future. Additionally, the molecular hits identified in the study can be further tested in vitro before claiming its inhibitory potential.

REFERENCES

1) Li, N., Deng, Y., Zhou, L., Tian, T., Yang, S., Wu, Y., ... & Zhang, D. (2019). Global burden of breast cancer and attributable risk factors in 195 countries and territories, from 1990 to2017: results from the Global Burden of Disease Study 2017. *Journal of hematology & oncology*, *12*(1), 140.
2) Wee, P., & Wang, Z. (2017). Epidermal growth factor receptor cell proliferation signaling pathways. Cancers, 9(5), 1–45. https://doi.org/10.3390/cancers9050052
3) Shi, Z., Chen, J., Guo, X., Cheng, L., Guo, X., & Yu, T. (2018). In silico identification of potent small molecule inhibitors targeting epidermal growth factor receptor 1. *Journal of Cancer Research and Therapeutics*, *14*(1), 18.
4) Seca, A. M., & Pinto, D. C. (2018). Plant secondary metabolites as anticancer agents: successes in clinical trials and therapeutic application. *International journal of molecular sciences*, *19*(1), 263.
5) Newman, D. J., & Cragg, G. M. (2020). Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. *Journal of Natural Products*, *83*(3), 770-803.
6) Thomford, N. E., Senthebane, D. A., Rowe, A., Munro, D., Seele, P., Maroyi, A., & Dzobo,
7) K. (2018). Natural products for drug discovery in the 21st century: innovations for novel drug discovery. *International journal of molecular sciences*, *19*(6), 1578.
8) Vigo- Morales, P. R., Borrero- Garcia, L. D., Maldonado, M. D. M., & Dharmawardhane,
9) S. (2019). Sensitization of Anti- EGFR/HER2 Targeted Therapy Resistance in Breast Cancer Cells by MiRNA Manipulation. *The FASEB Journal*, *33*(S1), 647-47.
10) Pettitt, G. A., McPherson, H. R., Hurst, C. D., Burns, J. E., Alder, O. A., Dunning, M. C., & Knowles, M. A. (2017). Mechanisms of resistance to FGFR-targeted therapy in bladder cancer.
11) Arshad, S., Naveed, M., Ullia, M., Javed, K., Butt, A., Khawar, M., & Amjad, F. (2020). Targeting STAT-3 signaling pathway in cancer for development of novel drugs: Advancements and challenges. *Genetics and Molecular Biology*, *43*(1).
12) Zhang, Y., Xia, M., Jin, K., Wang, S., Wei, H., Fan, C., ... & Zeng, Z. (2018). Function of the c-Met receptor tyrosine kinase in carcinogenesis and associated therapeutic opportunities. *Molecular cancer*, *17*(1), 1-14.
13) Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, Li Q, Shoemaker BA, Thiessen PA, YuB, Zaslavsky L, Zhang J, Bolton EE. PubChem 2019 update: improved access to chemical

https://iabcd.org

data. Nucleic Acids Res. 2019 Jan 8; 47(D1):D1102-1109. doi:10.1093/nar/gky1033.

14) Mangal M, Sagar P, Singh H, Raghava GP, Agarwal SM. NPACT: naturally occurring plant-based anti-cancer compound-activity-target database. Nucleic acids research. 2012; 41(D1):D1124-9.

15) BIOVIA, D. S. (2017). BIOVIA Discovery Studio 2017 R2: A comprehensive predictive science application for the Life Sciences. *San Diego, CA, USA http://accelrys. com/products/collaborative-science/biovia-discovery-studio.*

16) Hu, B., Zhou, R., Li, Z., Ouyang, S., Li, Z., Hu, W., ... & Jiao, B. (2019). Study of the binding mechanism of aptamer to palytoxin by docking and molecular simulation. *Scientific reports, 9*(1), 1-11.

17) De Lano, W. L. California U. PyMOL: An Open-Source Molecular Graphics Tool.

18) Burley, S. K., Berman, H. M., Kleywegt, G. J., Markley, J. L., Nakamura, H., & Velankar,

19) S. (2017). Protein Data Bank (PDB): the single global macromolecular structure archive. In *Protein Crystallography* (pp. 627-641). Humana Press, New York, NY.

20) Pandya, P. N., Mankad, A. U., Rawal, R. M., & Prasanth, K. S. (2019). Screening of diverse phytochemicals with Aurora Kinase C protein: An In Silico approach. *Journal of Drug Delivery and Therapeutics, 9*(1-s), 67-74.

21) Keller, S., Kneissl, J., Grabher-Meier, V., Heindl, S., Hasenauer, J., Maier, D., ... & Luber,

22) B. (2017). Evaluation of epidermal growth factor receptor signaling effects in gastric cancer cell lines by detailed motility-focused phenotypic characterization linked with molecular analysis. *BMC cancer, 17*(1), 845.