Virtual screening of natural compounds as inhibitors of EGFR 696-1022 T790M associated with non-small cell lung cancer

Mahesha Nand[1], Priyanka Maiti[2], Ragini Pant[1], Madhulata Kumari[3], Subhash Chandra*[,2], Veena Pande[1]

[1]Department of Biotechnology, Kumaun University, Bhimtal Campus Bhimtal, Uttarakhand, India; [2]Department of Botany, Kumaun University, S.S.J Campus, Almora, Uttarakhand, India; [3]Department of Information Technology, Kumaun University, SSJ Campus, Almora, Uttarakhand 263601, India. Mahesha Nand - E-mail: maheshlyf87@gmail.com, Priyanka Maiti – E-mail: priyankamaiti.06@gmail.com, Ragini Pant – E-mail: raginipant1@gmail.com, Subhash Chandra* – E-mail: scjnu@yahoo.co.in, Veena Pande – E-mail: -veena_kumaun@yahoo.co.in; Phone: +91-7351976362; *Corresponding author

Abstract:
Non-small cell lung cancer (NSCLC) is the most dominating and lethal type of lung cancer triggering more than 1.3 million deaths per year. The most effective line of treatment against NSCLC is to target epidermal growth factor receptor (EGFR) activating mutation. The present study aims to identify the novel anti-lung cancer compounds form nature against EGFR 696-1022 T790M by using *in silico* approaches. A library of 419 compounds from several natural resources was subjected to pre-screen through machine learning model using Random Forest classifier resulting 63 screened molecules with active potential. These molecules were further screened by molecular docking against the active site of EGFR 696-1022 T790M protein using AutoDock Vina followed by rescoring using X-Score. As a result 4 compounds were finally screened namely Granulatimide, Danorubicin, Penicinoline and Austocystin D with lowest binding energy which were -6.5 kcal/mol, -6.1 kcal/mol, -6.3 kcal/mol and -7.1 kcal/mol respectively. The drug likeness of the screened compounds was evaluated using FaF-Drug3 server. Finally toxicity of the hit compounds was predicted in cell line using the CLC-Pred server where their cytotoxic ability against various lung cancer cell lines was confirmed. We have shown 4 potential compounds, which could be further exploited as efficient drug candidates against lung cancer.

Key words: lung cancer, natural compounds, EGFR, machine learning, molecular docking, and prediction

Background:
The lung cancer accounts for approximately 1.59 million deaths per year worldwide and imposes major threat to human health. There are two subtypes of lung cancer, small cell lung cancer (SCLC) (occurrence-10-15%) and non-small cell lung cancer (NSCLC) (occurrence-85-90%) **[1]**. NSCLC is mainly caused by mutations like in-frame deletions or amino acid substitutions, clustered in the vicinity with ATP-binding pocket of the tyrosine kinase domain present in the epidermal growth factor receptor (EGFR) **[2]**. The epidermal growth factor receptor family of tyrosine kinases has been defined in four types: EGFR (HER1), ErbB2 (HER2), ErbB3 (HER3) and ErbB4 (HER4) **[3]**. Studies show that mutations in the EGFR gene occur more often in women (37.5%) than in men

(13.0%), in non-smokers (50.8%) than in smokers (9.0%), and in patients of East Asian area (29.1%) than in patients from the United States (9.5%) **[4]**. These statistics prove that targeting EGFR is a promising way of attacking against NSCLC to cure a patient and some Tyrosine kinase inhibitors (TKIs) for instance Gefitinib and Erlotinib are extensively being used in clinical treatment of NSCLC **[5]**. HKI-272 (irreversible inhibitor of the EGFR and HER-2 receptors) is a second-generation TKI that exploits both the strategies of covalent binding and multi-targeting **[6].** Similar other inhibitors in this context are EKB-569 (irreversible inhibitor of EGFR TK), CI-1033 (irreversible inhibitor of EGFR, HER-2 and ErbB-4), ZD6474 (dual-kinase inhibitor of VEGFR-2 and EGFR) etc. Nevertheless, the clinical effectiveness of these inhibitors is

restricted due to emergence of mutations causing drug resistance, including the gatekeeper T790M mutation. There has been a limited success in strategies targeting EGFR T790M with irreversible inhibitors and associated with toxicity because of concurrent inhibition of wild-type EGFR. To extend the finding of better TKI inhibitors, natural compounds are better due to various diverse classes of features. The numerous natural compounds have been reported to have many biological activities as anti-viral, anti-bacterial, and anti-cancer **[7].** The natural products and their derivatives mimic over 50% of all drugs that are being used clinically. The present work aimed *in silico* investigation of natural inhibitors against EGFR696-1022 T790M from diverse natural sources namely plants, fungi, lichen, bacteria, cyanobacteria, vegetables and nutritional food. The hit compounds obtained from this study could play an important role in designing personalized therapy against lung cancer patients and innovative drug discovery against NSCLC.
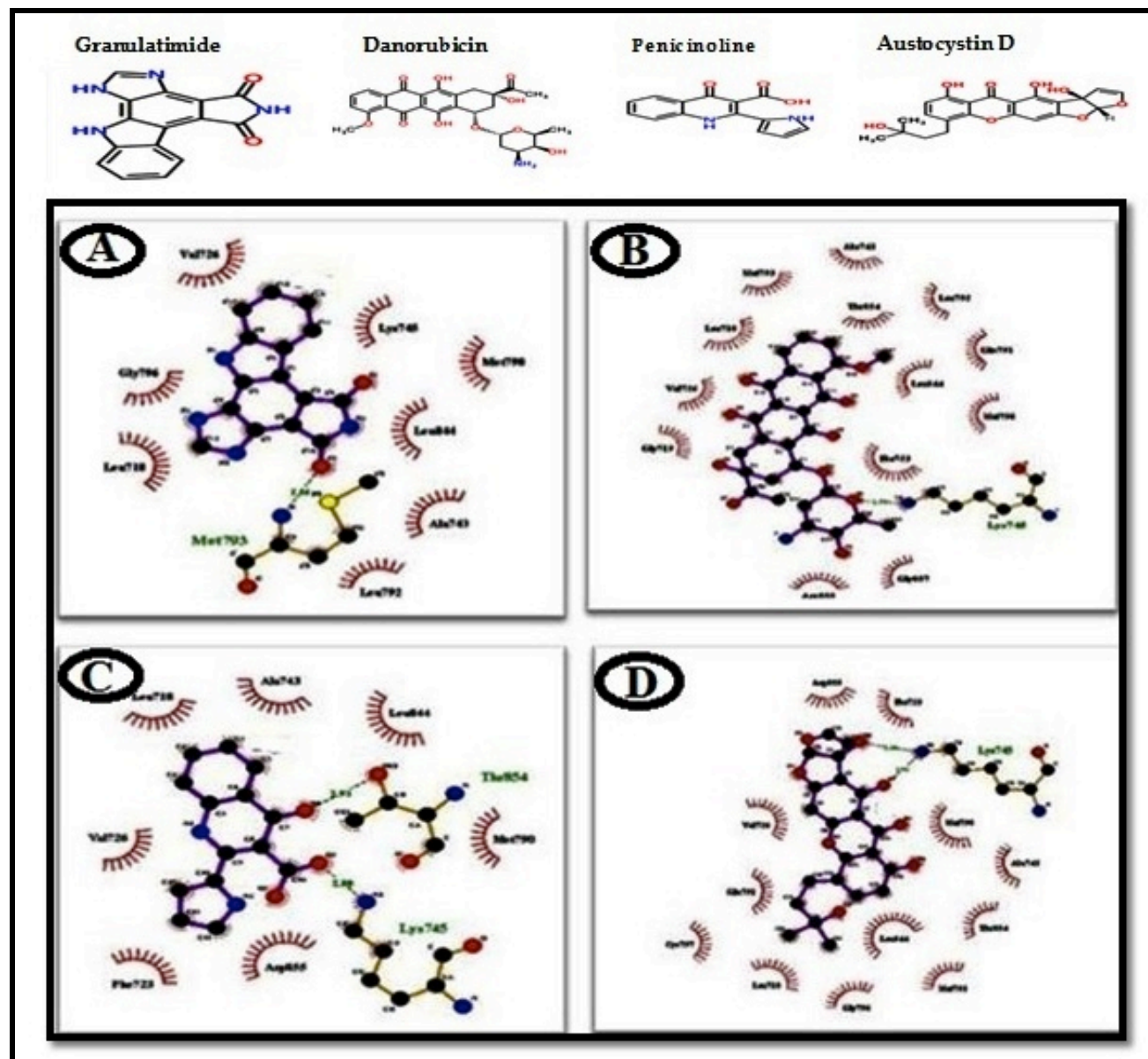
Figure 1: Binding conformation of four hit molecules (A= Granulatimide, B= Daunomycin, C= Penicinoline, D= Austocystin D) with target protein

ISSN 0973-2063 (online) 0973-8894 (print)

312

Bioinformation 12(6): 311-317 (2016)

©2016

BIOMEDICAL
INFORMATICS

Compounds Name	Pa	Pi	Cell line	Cell line full Name	Tissue	Tumor type	FAF-Drug3
	0.553	0.103	NCI-H226	Non-small cell lung carcinoma cells	Lung	Carcinoma	
	0.586	0.071	L2987	Lung adenocarcinoma cells	Lung	Adenocarcinoma	
Granulatimide	0.610	0.042	Lu1	Lung carcinoma cells	Lung	Carcinoma	Accepted
	0.643	0.050	HOP-62	Non-small cell lung carcinoma cells	Lung	Carcinoma	
	0.847	0.030	NCI-H522	Non-small cell lung carcinoma cells	Lung	Carcinoma	
Danorubicin	0.583	0.073	NCI-H128	Small cell lung cancer	Lung	Carcinoma	
	0.651	0.010	SPC-A4	Lung Adenocarcinoma	Lung	Adenocarcinoma	Accepted
	0.895	0.002	NCI-H157	Non-small cell lung carcinoma cells	Lung	Carcinoma	
	0.926	0.013	NCI-H1975	Bronchoalveolar carcinoma cells	Lung	Carcinoma	
Penicinoline	0.509	0.076	HOP-62	Non-small cell lung carcinoma cells	Lung	Carcinoma	Accepted
	0.568	0.050	NCI-H1975	Bronchoalveolar carcinoma cells	Lung	Carcinoma	
	0.662	0.005	NCI-H157	Non-small cell lung carcinoma cells	Lung	Carcinoma	
Austocystin D	0.688	0.028	NCI-H460	Non-small cell lung carcinoma	Lung	Carcinoma	
	0.798	0.042	NCI-H226	Non-small cell lung carcinoma cells	Lung	Carcinoma	Accepted
	0.811	0.033	NCI-H1299	Non-small cell lung carcinoma	Lung	Carcinoma	
	0.843	0.017	NSCLC	Non-small cell lung carcinoma cells	Lung	Carcinoma	
	0.993	0.007	NCI-H1975	Bronchoalveolar carcinoma cells	Lung	Carcinoma	

Table 1: CLC- Pred (Cell-Line Cytotoxicity Predictor) and binding energy of four hit molecules

Methodology:

Library of compounds
An *in house* library of 419 compounds from various natural resources (Plants-105, fungi-51, lichen- 50, bacteria- 111, cyanobacteria- 52, vegetables- 50 and nutritional food-51) was constructed. The 3D structures of these compounds were retrieved from PubChem, ZINC and Chemspider database. The structures, which were not available in databases, were drawn by using ChemSketch [8] software in mol2 format.

Molecular descriptors and data editing
The molecular descriptors are properties that describe a molecule on the basis of either some physico-chemical property like two-dimensional (2D) fingerprint. The geometrical or 3D descriptors are resulted based on x, y and z coordinates, which provide rich information regarding a molecule's orientation in space. All these calculated descriptors were used as features to build machine-learning model. Total 3500 1d, 2d, 3d and fingerprint descriptors were calculated for each molecule to prepare the data set using PaDEL-Descriptor v2.20 [9], popular software used for calculation of descriptors in chemo informatics. After calculating the descriptors, the WEKA software was used to edit resultant data in .csv format. CfsSubsetEval evaluator was used to remove non-significant features from data [10], which ensures accuracy, completeness, consistency, timeliness, believability and interpretability of the data.

Machine learning models
A data set comprising of 400 compounds were collected from PubChem bioassay database [AID: 977608] with known IC_{50} values. The IC_{50} values (the concentration of an inhibitor required for 50% inhibition of its target) of the highest 100 compounds were used as active to make a training set for current study. Descriptors for training set were calculated by above mention method. Using this data set and WEKA tool kit three machine learning models based on Random Forest, J48, and Bayes Net were built and compared statistically to select the best predictive model.

Molecular docking:

Receptor preparation:
Co-crystal structure of EGFR 696-1022 T790M mutant covalently attached to WZ4002 (PDB NO: 3IKA) was retrieved from the Protein Data Bank [11]. The structure was imported into Chimera software for protein preparation. The water molecules and heat atoms (OUN i.e. WZ4002) were deleted from the protein. Dock Prep function was then used to fix the protein.

Docking:
The computational docking of all molecules and reference ligand were performed by AutoDock Vina program using PyRx graphic user interface (GUI). The binding coordinates were automatically decided by PyRx, which correlated with the co-ordinate of experimentally bound ligand OUN. They are center_x = -10.7299, center_y = 27.9715, center_z = 36.4637, size_x = 25, size_y = 25 and size_z = 25.

Lamarckian Genetic Algorithm (LGA) was used during docking with standard docking protocols. In LGA the orientation/conformation of the ligand with respect to the receptor is represented through a chromosome. The constructed chromosome consists of a number of variables for ligand translation/rotation that was the same for all ligands and a number of variables for ligand flexibility that is specific to each ligand. The sum of the chromosome signifies the genotype of a ligand. Its 3D coordinates represent the phenotype of a ligand after applying all the transformations in the genotype. Fitness of the ligands is

ISSN 0973-2063 (online) 0973-8894 (print)

313

Bioinformation 12(6): 311-317 (2016)

©2016

BIOMEDICAL
INFORMATICS

calculated from its phenotype using any of the standard docking scoring functions. AutoDock uses a modification on the AMBER'95 force field **[12]** with terms empirically determined by linear regression analysis from a set of protein-ligand complexes with already known binding constants **[13].** Free energy (Gibbs, ∆G) is denoted by a master equation that contains six terms to model dispersion/repulsion, internal ligand torsional constraints, electrostatic interactions, deviation from the covalent geometry, hydrogen bonding and desolvation effects: $\Delta G = \Delta G_{vdw} + G_{hbond} + G_{elec} + G_{conform} + G_{tor} + G_{sol}$

Revalidation of docking results
The docking results of screened ligands were revalidated using X-Score a "scoring function" **[14]** to re-score the selected ligands with EGFR 696-1022 T790M protein. FAF-Drug3 **[15]** server was used to check their drug likeness, which is originally, based on the free chemo-informatics toolkit Frowns **[16].**

Pharmacophore elucidation and Cytotoxicity prediction on cell lines:
The common pharmacophores for the hit molecules and nine active molecules (used as reference) obtained from bioassay (AID: 256664) were predicted using PharmaGist server **[17-18].** Finally the hit molecules were annotated to check their cytotoxic effect on tumor and normal cell-lines by using CLC-Pred **[19]** web services, which predicts cytotoxicity of molecules for tumor and normal cell-lines based on the structural formula.

Results and Discussion:
Initially an *in house* library of 419 natural compounds from several diverse sources was constructed [Supplementary data]. The entire library was screened to find compounds against EGFR 696-1022 T790M receptor. Initially machine learning models were built using three classifiers; Random Forest, J48 and Bayes Net and analyzed for their performance by comparing True Positive Rate (TPR) which is ratio of predicted true actives to real number of actives and False Positive Rate (FPR) that is ratio of predicted false actives to actual number of inactive. The accuracy of these models was measured in terms of sensitivity, which predicts the test's ability to identify positive results, and specificity that evaluates the test's ability to identify negative results. Finally Random Forest classifier model (with highest sensitivity and specificity) was applied on the data set of test and library compounds, which resulted 63 compounds with active potential. Subsequent to machine learning approach molecular docking was used for screening using AutoDockVina. Ligand receptor interactions of the docked molecules were visualized by using LigPlot+ software **(Figure 1**) **[20].** Ligplot analysis was done to investigate in-depth interaction pattern (hydrophobic interactions, hydrogen bonding pattern) between docked ligand and active site residues of the protein. Finally four molecules were screened by docking process namely Granulatimide, Daunomycin, Penicinoline, Austocystin D with binding energy -8.8 kcal/mol, -9.4 kcal/mol, -7.7 kcal/mol, -8.6 kcal/mol which showed comparable binding energy with

reference ligand i.e. -8.4 kcal/mol. Rescoring of docked molecules was carried out by estimation of binding free energies of screened molecules with EGFR 696-1022 T790M using X–Score program. All compounds showed acceptable range of negative binding energies with comparable results. After that Faf Drug3 server was used to check their druglikeness. In Faf Drug 3 server more than twenty-three molecular properties, which are important for evaluating drug likeness of a molecule like number of violations of Lipinski's RO5, Veber Rule, Egan Rule, and Topological Polar Surface Area etc., were calculated. All of molecules showed good binding pattern in Ligplot analysis and accepted in terms of Drug likeness by FaF-Drug3 **(Table 1).** PharmaGist server was then used to elucidate pharmacophores for the 4 hit compounds where their common pharmacophore properties were compared with the common pharmacophore properties of 8 potential anti EGFR molecules obtained from PubChem bioassay (AID: 256664). Pharmacophores were detected by extracting common chemical features from 3D structures of the active ligand set that are representative of essential interactions between the ligands and EGRF. The results of pharmacophores analysis predict physico-chemical properties namely, hydrogen-bond acceptor/donor atom; a set of atoms of an aromatic ring and adjacent hydrophobic atoms which are important for specific biological activity. The active compounds showed 3 hydrogen bond acceptor, 2 hydrogen bond donor and 1 hydrophobic group whereas the common pharmacophore from 4 hit molecules showed 3 hydrogen bond acceptor, 1 hydrogen bond donor and 1 hydrophobic group. The final annotation was done using CLC-Pred server where all of the hits were evaluated for their cytotoxicity on cell lines **(Table 1).** The results showed these 4 hit molecules have high potential to inhibit lung cancer by targeting EGFR. All of these molecules where already reported for their anti-cancer properties. Granulatimide is an aromatic alkaloid with G2 checkpoint inhibitory activity **[21].** Daunomycin (daunorubicin) was the primary anthracycline compound to be demonstrated structurally and stereo chemically. Danorubicin is used against acute lymphoblastic and myeloblastic leukaemias **[22].** Penicinoline is a pyrrolyl 4-quinolinone alkaloid with an extraordinary ring system, isolated from a mangrove endophytic fungus. Penicinoline has strong in vitro cytotoxicity toward 95-D and HepG2 cell lines **[23].** Austocystin D is found in *A. pseudoustus* has been reported for inhibiting growth of human colon carcinoma LS174T cells in mice and tumor cell lines that overexpress the multidrug resistance protein. Additionally the compound can inhibit a number of cancer cell lines: SR, U-87, MCF-7, MDA-MB-231, PC-3, SW-620, HCT-15, and MX-2 **[24, 25].**

Conclusion:
Currently the clinical efficacy of TKIs (gefitinib and erlotinib) to treat NSCLC is becoming limited because of the emergence of drug resistance conferred by a mutation: substitution of threonine 790 with methionine (T790M). The amino acid residue of EGFR Threonine 790 holds a key location at the entrance of a hydrophobic pocket in the back of the ATP binding cleft of EGFR protein. When this residue is substituted with a bulky methionine,

Open access

the protein becomes resistant to TKIs, including gefitinib and erlotinib by steric interference. So targeting this mutant EGFR is a promising approach for designing drugs against lung cancer. In this context the findings of present study represent four hit molecules namely Granulatimide, Daunomycin, Penicinoline, Austocystin D, that may have potential as novel EGFR 696-1022 T790M inhibitors. Granulatimide and Daunomycin bind with EGFR by one hydrogen bond where as Penicinoline and Austocystin D by two hydrogen bonds respectively as compare with the reference ligand that makes two hydrogen bonds. The results also showed that these molecules have efficient drug like properties and have also the cytotoxic capacity against lung cancer cell lines. Therefore these molecules have future prospect in drug development with low cost and less side effects (as obtain from nature). Natural compounds constitute a higher percentage of marketed drugs, which indicates a significantly higher rate of return per molecule against NSCLC.

Reference:
 [1] Petersen I. Dtsch Arztebl Int. 2011 108: 531. [PMCID: 3163783].
 [2] Zhang ZF. Nat Genet. 2012 44:852 [PMID: 22751098].
 [3] El-Azab AS *et al.* Eur. J. MedChem. 2010 45:4198[PMID: 20599299].
 [4] Franklin WA *et al.* J. Semin Oncol. 2002 29: 14 [PMID: 11894009].
 [5] Kris MG *et al.* JAMA. 2003, 290:2158 [PMID: 14570950].
 [6] Fukuoka M *et al.* J Clinical Oncology 2003 21:2246 [PMID: 12748244].
 [7] Harvey AL. Drug Discovery Today 2008: 894 [PMID: 18691670].
 [8] http://www.acdlabs.com. [Comment: ACD/Structure Elucidator, Advanced Chemistry Development 2015]
 [9] Liew HY *et al.* Journal of Computer Chem. 2012, 4: 15 [PMID: 23114987].
 [10] Hall M *et al.* ACM SIGKDD Explore News 2009 11: 11. http://dl.acm.org.
 [11] Protein Data Bank http://www.pdb.org Accessed 15th may 2016.
 [12] Morris GM *et al.* Journal of Computer-Aided Mol. Design 1996:304 [PMID: 8877701].
 [13] Kuntz D *et al.* Journal of Molecular Biology 1982 161:288 [PMID: 7154081].
 [14] Wang RJ *et al.* Computer Aided Mol. Des. 2002 16:26 [PMID: 12197663].
 [15] Miteva MA *et al.* J Nucleic Acids Res. 2006 22:744 [PMID: 16845110].
 [16] http://frowns.sourceforge.net
 [17] http://bioinfo3d.cs.tau.ac.il/PharmaGist
 [18] Schneidman-Duhovny *et al.* Nucleic Acids Research 2008 36: W223 [PMID: 18424800]
 [19] http://www.pharmaexpert.ru/PASS online [Comment: CLC-Pred Cell-Line Cytotoxicity Predictor]
 [20] Wallace C *et al.* Protein Engineering1995 8:134[PMID: 7630882].
 [21] Roberto GS *et al.* J. Org. Chem. 1998 63:9856. http://pubs.acs.org.
 [22] Shao B *et al.* Med Chem Lett. 2010 3284: 3286[PMID: 20452770].
 [23] Tanja T *et al.* Molecule. 2013 18:11376
 [24] Yadav, PK *et al* Bioinformation. 2013 9(3): 158-64
 [25] Farmer R *et al.* Bioinformation. 2010 4 (7): 290-294

Other supplementary data (available from the authors) Symbols are missing

[1] **Construction of in house library from Plants (105 Compounds)**
The world is rich in medicinal plants. Man cannot survive on this earth for long life without plant because the plant products and their active constituents have always been playing an important role. These include Gamma-Asarone, cis-asarone, Trans-asarone, Isoeugenol methyl ether, Acoramone, asarylaldehyde, Shyobunone, Epishyobunone, Isocalamendiol, Acoragermacrone, Preisocalamendiol, Thujane, Limonene, Myrcene , Cymene-Para, Beta pinenes, Terpinolene, Galangin, Acoradin, -sitosterol, 29-tetrahydroxyolean-12-en-3-O- -D-arabinosyl, 1,3 - -D

Arabinopyranoside, 2,6-diepishyobunone, Calamusenone, Aspargamine A, Racemosol, Racemofuran, Quercitin, Shatavarin V, Sarsasapogenin, Niazimicin, Pterygospermin, Withaferin A, Viscosalactone B, Withanolide D, Lanosterin Anaferine, Epipodophyllotoxins, Etoposide, Teniposide, Daunorubicin, Doxorubicin, Idarubicin, Epirubicin, Anthracyclines, Sabarubicin, Nemorubicin, Rohitukine, flavopiridol, lapachone, Paclitaxel, Cisplatin, Curcumin, Resveratrol, Bisindole alkaloids, Taxanes, Combretastatin A4 phosphate, Stilbenoid, Homoharringtonine, Mepesuccinate, Cephalotaxine A, Homoharringtonine, Benzopyran, Daidzein, Isoflavone,

Phenoxodiol, Ingenol 3-O-angelate, Genistein, Diterpines [Pub chem], p-Sitosterol, a-Amyrin, Lupeol, Kaempferol, Rutin, Hexacosanol, P-Dglucoside, p-Sitosterol, Cleomin, 5,7-Dihydroxychrornone, 5-hydroxy-3,7,4'-trimethoxyflavone, Lupeolin, Gluccapparine, Viscosic acid, flavone-viscosin, P-carotene, Methyl glucosinolate, Glucoiberine, Neoglucobrassicin, 5,7-dihydroxychro1lao~e 5, Hydroxy-3,7,4'-trimethoxy flavone, Lupeolin, Cleogynol (*Cleome gynandra* Linn.), -ocimene, -terpinene, *cis*-sabinene hydrate, -terpinolene , *trans*-sabinene hydrate, Linalool, Camphor, Terpinen-4-ol, Myrtenol, *trans*-caryophyllene, Germacrene D*(Ocimum kilimandscharicum).* In this study 105 natural compounds from different plants were selected through searching scientific literature and deposited in to in houses libraries for further screening.

[2] Compounds from vegetables (51Compounds)

Glucoerucin, ERUCIN, Sulphoraphen, Docetaxel, Paclitaxel, Gemcitabine, Depocyt, Cytarabine, [Pub Chem], Oxalic acid, Glucoraphanin, Folic acid, Phytic acid, Chlorogenic acid, Epigallocatechin 3-gallate, Resveratrol, Lycopene, Flavonols, Sulforaphane, Caffeic acid, Ferulic Acid, Xylitol, Sorbitol, Mannitol, Lactitol, Daidzein, Genistein, Dithiolthiones, omega-3 fatty acids, Glucosinolates, Kaempferol, Quercetin, Isorhamnetin, 4-Methoxyindol-3-ylmethyl, 1-Methoxyindol-3-ylmethyl, Indol-3-ylmethyl, 4-Hydroxy-3-indolylmethyl, 3-Butenyl, 5-Methylthio pentyl, 4- Methylsulphinyl butyl, Glucoberteroin, Glucoputranjivin isopropyl, Glucochlearin, Glucojiabutin, Desulpho-glucosinolates, Quercetin-*O*-sophorotrioside-7-*O*-sophoroside, Tetradecanoic acid, Octadecanoic acid, Eicosanoic acid, Kaempferol-*O*-tetraglucoside-7-*O*-sophoroside, Kaempferol 3-*O*-sophoroside-7-*O*-(caffeoyl)-rhamnoside, Isorhamnetin3-sophorotrioside-7-sophoroside, Hydroxycinnamic acids, 3-caffeoyl quinic acid Decanoic acid.

[3] Construction of in house library from Lichen (50 Compounds)

Lichens are considered one of the rich sources of bio medically useful compounds likes Chloratranorin, D-Arabinitol, Lichesterol, Methyl haematommate, Atranorin, Glutinol, Vulpinic acid (2), 6 -acetoxyhopane-16 , 22-diol, 16 -acetoxyhopane-6 , 22-diolzeorin, Connorstictic acid, Salazinic acid, Norstictic acid, Chloroatranoric acid, Constipatic acid, Atranorin, Chloroatranorin (*Heterodermia appendiculata*), Stictic acid, Constictic acid, Consalazinic acid, Atranorin (*Bulbothrix isidiza*), Confumarprotocetraric acid, Protocetraric acid, fumarprotocetraric acid, Sekikaic acid, Homosekikaic acid, Atranorin, Hyperhomosekikaic acid (*Cladonia adspersa*), -alectoronic acid, -collatolic acid, Skyrin (*Parmotrema rampoddense*), Isonephrosterinic acid, Protolichesterinic acid, Lichesterinic acid, (*Hypotrachyna ikomae*), Norobtusatic acid, 4-*O*-demethylbarbatic acid, Obtusatic acid, Barbatic acid (*Hypotrachyna imbricatula)* (3) Diffractaic acid, Protocetric acid, Lecanoric acid, Orsellinic acid (*Ramalina lacera*), (8) Gyrophoric acid, Usnic acid (17) flavo-obscurin A-B, oxolobaric acid, Lobaric acid.

[4] Bacteria derived compounds for in house library (60 Compounds)

Around 23,000 bioactive secondary metabolites from microorganisms have been reported and over 10,000 of these are from actinomycetes, representing 45% of all bioactive microbial metabolites and submit in house library for further screening. Actinomycin D, Bleomycin, l-asparaginase, Ecteinascidin 743, Aureolic acid, Dactinomycin, Idarubicin, Laxaphycins A, Laxaphycins B (Streptomyces spp.), (Streptomyces verticillus), Daunomycin (Streptomyces coeruleorubidus), Epirubicin (Streptomyces pneuceticus), Mitomycin C (Streptomyces caespitosus), Geldanamycin, Rapamicin (Streptomyces hygroscopicus), Wortamannin (Talaromyces wortmanni (9), Dolastatin 10, Dolastatin 15, Macrocyclic polyethers, Halichondrin B, Didemnin B, Cyclic Depsipeptide, Aplidine, Bryostatins, Synthadotin, Kahalalide F, Squalamine, Dehydrodidemnin B, Cemadotin, Didemnin, Soblidotin, Discodermolide, Bengamide B, Curacin A, Salinosporamide A, Laulimalide, Vitilevuamide, Diazonamide, Eleutherobin, Sarcodictyin, Peloruside A, Salicylihalimides A and B, Thiocoraline, Ascididemin, Variolins, Lamellarin D, Dictyodendrins, Psammaplin A-F, Bromotyrosine disulfide, Biprasin, Laxaphycins A and B, Epirubicin, Daunomycin, Aeroplysinin.

[5] Ligand library from Fungus (51 Compounds)

Natural compounds produced by fungus have been great potential and useful in safety and human health. There is significant demand of fungal compounds in drug industry for synthetic products due to economic and time reducing reasons. Brefeldin A, Irofulven, Lodamin A, Aculeatins A, Aculeatins B, Lodamin B, Aculeatols A-E, Cytotoxic dioxadispiroketal, Silvestrol, Amomol A, Episilvestrol, Methyl rocaglate, Camptothecin, Gliotoxin, Illudin S, Leptomycins, Palmarumycin, Terrecyclic acid, Wortmannin, Lovastatin, Pravastatin, Griseofulvin, Compactin, Simvastatin, Tricitrinol A-C, Terrein, Brefeldin A, Asperlin, Griseofulvin, Sequoiamonascin, Phaeosphaerin B, Hypocrellin, Calphostin C, Diketopiperazine, Fumitremorgin C, Asperphenamate, Notoamide A, Oxo tryprostatin A, Stephacidin A, Xanthocillin, Phenylahistin, Aurantiamine, Leucinostatin A, Taxol, Trichothecene A, 3-Hydroxyroridin E, Caffeic acid, p-coumaric acid, Hyperoside, Ellagic acid, Quercetin of fungus through searching scientific literature and submit in house library for further screening.

[6] Nutritional and Functional Food for preparation of in house library (51 Compounds)

Anticancer activities from many functional food sources have been reported. Some dietary factors now known to promote cancer development like polished grain, foods and low intake of

BIOMEDICAL INFORMATICS

fresh vegetables. Lycopene, Beta-glucan, Conjugated linoleic acid (CLA), Flavones, Catechins, Anthocyanidins, Lignans, Daidzein, Ganistein, Balsamita, Parsley, Parsnips, Epigallocatechin 3-gal, - linolenic acid, Docosapentaenoic acid, Epigallocatechin gallate, Phytosteryl oleates, Docosahexeanates, Phytosteryl caffeates, Ferulates, Sinapates, Phytosteryl phenolates, Malvidin, Cyanidin, Delphinidin, Pelargoindin, Peonidin, Peltunidin, Apigenin, Luteolin, Myricetin, Quercetin, Catechin, Epicatechin,, Gallocatechin, Epigallocatechin, Daidzein, Genistein, Gallic acid, Phydroxybenzoic, Vanillic acid, Genistein, Caffeic acid, Ferulic acid, *p*-coumaric acid, Sinapic acid, Pinoresinol, Podophyllotoxin, Steganacin, Castalin, Pentagalloyl glucose, Procyanidins, Phytic acid, Secoisolariciresinol diglucoside (SDG), Enterodiol, Enterolactone, Vanillin, Syringic acid, Tocotrienols (*T. durum*). In this study we extract 50-50 chemical compounds from 6 different foods and drink (brown rice, barley, wheat, green tea and coffee) as well some vegetables (bitter melon, garlic, onions, broccoli, and cabbage) through searching scientific literature and deposited in to in houses libraries for further screening.

[7] Compounds from marine and terrestrial cyanobacteria (52 Compounds)

The majority of species are found in the oceans, which constitute 70% of the world's surface. Marine organisms are a rich source of chemical products. Overall, more than 3000 new substances have been identified from marine organisms that demonstrate the great potential as a source of novel chemical classes. Grenadamied B-C, Curacin A, Grenadadiene, Lyngbyabellin A - B, Tanikolide, Kalkitoxin, Isomalyngamide A-B, Barbamide (Lyngbya majuscula), Guineamide G, Semiplenamide A-G, Pahayokolide A (*Lyngbya semiplena)*, Phormidolide (*Phormidium sp.*), Comnostin A-E (*Nostoc commune*), Cyanolide A (*Lyngbya bouillonii*), Aerucyclamide A- B (*Microcystis aeruginosa*), Oscillapeptin J (*Planktothrix rubescens*) (4 Rif), Caylobolide A- B, Desmethoxy majusculamide C, Apratoxins A-C, Cocosamides A, Curacin A, Grassypeptolides F -G, PitipeptolidesA-B, Hectochlorin, Jamaicamides A-C, Palmyramide A, Dolabellin, Tanikolide dimer (Lyngbya majuscule), Palau'imide, Bisebromoamide, Lyngbyacyclamides A-D, Ulongamides A-F, Ulongapeptin (*Lyngbya* sp.), Guamamide, Tasipeptins, Apramides A-G, Dolastatin, Belamide A, Symplostatin, Tasiamide B, Micromide (*Symploca* sp.) (6) Apratoxin D-E, Aurilide B-C, Biselyngbyaside, Coibamide A, Cryptophycin 1, Dolastatin 10, Ethyl Tumonoate A, Hoiamide A, Dragonamide, Kalkitoxin, Largazole, Malyngolide dimmer

©2016

BIOMEDICAL INFORMATICS