

ORIGINAL ARTICLES

In-silico designing of hnRNP B1 inhibitors against lung cancer: a computational approach

Neelam Verma, Gurpreet Kaur

Department of Biotechnology, Punjabi University, Punjab, India

Correspondence: Gurpreet Kaur. Address: Department of Biotechnology, Punjabi University, Patiala-147002, Punjab, India. Email: kaur.gurpreetinder@gmail.com

Received: August 31, 2015

Accepted: December 14, 2015

Online Published: December 27, 2015

DOI: 10.5430/jbei.v2n2p47

URL: <http://dx.doi.org/10.5430/jbei.v2n2p47>

Abstract

RNA-binding proteins regulate the maturation of mRNA including pre-mRNA splicing, mRNA transport from the nucleus to cytoplasm, the translocation and longevity of transcripts within the cytoplasm as well as translation of the message. One key RNA-binding protein identified in several aspects of this process is heterogeneous ribonuclear ribonucleoprotein B1 (hnRNP B1) and it plays a role in several cellular functions. hnRNP B1 is overexpressed in several cancers including lung and squamous cell carcinoma of various organs, in addition, its mRNA can be detected in the serum of patients with early stage lung cancer and can therefore be used as a biomarker for cancer detection. Furthermore, it can be used as a target for treatment. In the present study hnRNP B1 was used as a target receptor for *in silico* ligand binding studies using a library of 55 small compounds selected from the current knowledge of lung cancer therapeutic drugs available in the market as well as known inhibitors of hnRNP B1. TSAR studies were carried out for selected compounds from BioMed CAChe to determine docking score and then compared with 7 lactate dehydrogenase inhibitor involved in lung cancer whose IC₅₀ value were known. Results from both software analyses indicated that EGCG was a potent candidate ligand of hnRNP B1 and was selected as a parent compound for pharmacophore modifications to improve the binding affinity and activity against hnRNP B1. A library of 158 compounds was generated after modification of EGCG and molecular modeling and docking identified several analogues with increased binding scores. In addition, pharmacophore studies indicated the significant importance of ring D in EGCG and the formation of H-bonds between the modified ligands and amino acid GLU108 of hnRNPB1. Hence the outcomes of study identify a strategy for the identification of new lead molecules bearing drug-like properties. Future studies are required to validate this approach and measure the efficacy of analogues of compounds identified using *in silico* approaches.

Key words

hnRNP B1, Lung cancer, Drug discovery, Molecular modeling, TSAR, Bioisosterism

1 Introduction

Lung cancer is a disease which is caused by uncontrolled proliferation of cells in tissues of the lung and if left untreated ultimately leads to the spread of tumor cells to adjacent tissues. About 80%-90% of lung cancers, including both small-cell lung carcinoma (SCLC) and non-small-cell lung carcinoma (NSCLC) arises from smoking tobacco^[1-3]. Lung cancer alone has accounts for 1.37 million deaths worldwide and there is an urgent need to identify new molecular markers of lung

cancer^[4]. hnRNP B1 is overexpressed at the early-stage of disease etiology in sputum and has been correlated with the increased risk of lung cancer. These characterizes have attracted the attention of researchers to further understand its role in the development of the disease^[5]. hnRNP B1 is a target for cancer treatment and in this study hnRNP B1 has been targeted using *in silico* docking to identify and improved drug design against lung cancer.

Standard shift has been noticed in preclinical drug discovery with the technique of structure-based drug design (SBDD). The gap between identified hits and their conversion to preclinical candidates can be filled by investigating the interactions between the ligands and their receptors. Combination of various techniques like X-ray structures, bioinformatics and molecular modeling allows the generation of receptor homology models which serve as basis to screen virtual libraries with potential chances of finding hits^[6].

The hnRNP B1 gene belongs to the A/B subfamily of universally expressed heterogeneous nuclear ribonucleoproteins (hnRNPs) and has a molecular weight of approximately 37,000 Daltons^[4]. As an essential component of the nuclear core complex, hnRNP A2/B1 regulates mRNA maturation, including mRNA splicing and mRNA transport from the nucleus to the cytoplasm. Studies reported the detection of high concentrations of hnRNP B1 mRNA in the sera of lung cancer patients, hence it is considered as a biomarker for cancer detection and a target for cancer treatment^[7]. hnRNPs are chromatin-linked RNA-binding proteins which play a significant role in RNA processing. The protein harbors three major RNA-binding motifs including; quasi-RNA Recognition Motifs (RRM), an arginine/glycine-rich box (RGG), and an hnRNP K homology (KH) domain^[8].

Charles *et al.* (2010) demonstrated the correlation of Pyruvate kinase 2 (PKM2) expressions with the overexpression of c-Myc, PTB (polypyrimidine tract binding protein, also known as hnRNP I), hnRNP A1 and hnRNP A2 in human gliomas. They also verified the up-regulation of transcription of PTB, hnRNP A1 and hnRNP A2 by oncogenic transcription factor Myc^[9]. Two separate genes, the L and M gene, encode for four isoforms of Pyruvate kinase in mammalian cells. The L gene is spliced to produce the L and R isoform (mainly expressed in liver and red blood cells respectively) while the M gene splices to form the M1 or M2 isoforms of pyruvate kinase. It has been reported that tumor cells overexpress PKM2^[10].

Rapidly proliferating cancer cells and normal cells have different metabolic features as described by Otto Warburg in 1920's. The observed change in the use of glycolytic pathways is now commonly referred to as the Warburg effect. Under normal conditions the basic energy provision of cells (*i.e.* glucose), is metabolized through glycolysis aerobically to pyruvate which is further converted to lactate under anaerobic conditions whereas in cancer cells the increased amount of lactate production is observed even in aerobic conditions^[11]. Pyruvate kinase is involved in the final step of glycolysis for production of pyruvate. Furthermore, lactate dehydrogenase consumes NADH for conversion of pyruvate to lactate^[12]. Hence overproduction of lactate dehydrogenase is a hallmark of cancer cell progression. In this study hnRNP B1 and lactate dehydrogenase were identified and targeted for drug design against lung cancer using BioMed CAChe and TSAR.

2 Importing structures from various databases

Each active site of a protein has a distinct size and shape which is biologically determined and the substrate must complement the size, shape and chemical interactions of a target site. This correlation defines the complementary size, shape and complexity of possible ligands. For docking experiments with hnRNP B1 a target receptor ligand library was built considering currently available lung cancer therapeutic drugs and reported inhibitors of hnRNP B1 and their analogues. The SDF format of 55 small compound structures were downloaded from the PubChem database and the crystal structure of the RRM domain in Heterogeneous nuclear ribonucleoproteins A2/B1 (PDB id-1X4B) was imported from Protein Data Bank (see Figure 1).

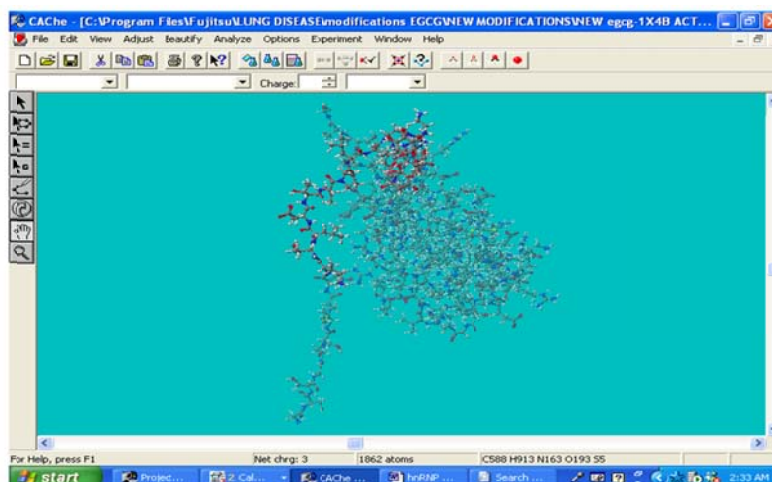


Figure 1. Crystal structure of hnRNP B1 on BioMed CAChe (PDB id-1X4B)

However the present structure was not associated with any ligand or compound, so literature based knowledge was used to identify the binding sites of the respective protein. Burd *et al.* in 1989 determined primary structure of hnRNP A2/B1 including their nucleotide and deduced amino acid sequence. According to report, the structures of A2 and B1 are similar except for a 36 nucleotide insert in B1 thus having an insertion of 12 amino acids. The active site was focused on the 12 amino acid insertions near the amino terminus, Lys-Thr-Leu-Glu-Thr-Val-Pro-Leu-Glu-Arg-Lys-Lys which bears substantial resemblance to acknowledge nuclear localization signals^[13] as shown in Figure 2.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
GLY	SER	SER	GLY	SER	SER	GLY	MET	GLU	LYS	THR	LEU	GLU	THR	VAL	PRO	LEU	GLU
19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
ARG	LYS	LYS	ARG	GLU	LYS	GLU	GLN	PHE	ARG	LYS	LEU	PHE	ILE	GLY	GLY	LEU	SER
37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54
PHE	GLU	THR	THR	GLU	GLU	SER	LEU	ARG	ASN	TYR	TYR	GLU	GLN	TRP	GLY	LYS	LEU
55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72
THR	ASP	CYS	VAL	VAL	MET	ARG	ASP	PRO	ALA	SER	LYS	ARG	SER	ARG	GLY	PHE	GLY
73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90
PHE	VAL	THR	PHE	SER	SER	MET	ALA	GLU	VAL	ASP	ALA	ALA	MET	ALA	ALA	ARG	PRO
91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108
HIS	SER	ILE	ASP	GLY	ARG	VAL	VAL	GLU	PRO	LYS	ARG	ALA	VAL	ALA	ARG	GLU	GLU
109	110	111	112	113	114	115	116										
SER	GLY	SER	GLY	PRO	SER	SER	GLY	*									

Figure 2. The sequence of the hnRNP B1 RRM. The highlighted portion is an inserted stretch of amino acid in hnRNP B1 and was considered as the active site. Picture by BioMed CAChe software.

3 Setting measurement properties for docking studies

To screen ligands with high binding affinity the first step was to set the required parameters for the docking experiments using BioMed CAChe software. BioMed CAChe 6.1.12.34 engages a stochastic optimization method to enhance the intra-molecular energy of a ligand by rotation and analyzing various orientations. Properties of the chemical samples were set to “dock ligand into active site” where the software needs a specific active site in the protein and a ligand to dock it into the protein cavity. The procedure used to dock the flexible ligand into the rigid active site used a genetic algorithm for predefined active sites and predefined ligands. The algorithm used instructions for a maximum of 3,000 generations; the mutation rate was set at -0.300000 and Convergence in kcal/mol -1.000000. Results were produced in another chemical file containing small compound (ligands) docked in the active site (hnRNP B1) and simultaneously a map file (see Figure 3) having several lowest energy orientations was also created.

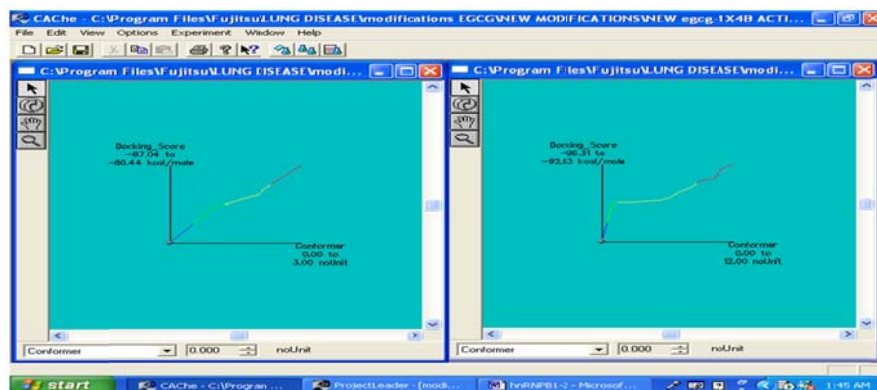


Figure 3. EGCG and compound IV_n map file created after the docking trial on BioMed CAChe software. In Amber PMF (separate file option) ligand was predefined to avoid the entire content of ligand file to be used as a ligand. Furthermore, calculations do not run if more than one ligand is selected.

Files were visualized using BioMed CAChe software. Scores of the calculated binding affinity formed the basis for ranking the molecules. Finally for further studies the 5 top molecules gefitinib (Pubchem CID 123631; compound I), Diferuloylmethane/curcumin (Pubchem CID 969516; compound II), Methotrexate (Pubchem CID 126941; compound III), Epigallocatechin-3-gallate(EGCG) (PubChem CID 65064; compound IV), curcumin diglucoside (Pubchem CID 46173989; compound V), with binding scores of -106.546 kcal/mol, -98.557 kcal/mol, -94.553 kcal/mol, -86.618 kcal/mol, -81.474 kcal/mol respectively were selected. Structures have been illustrated in Figure 4.

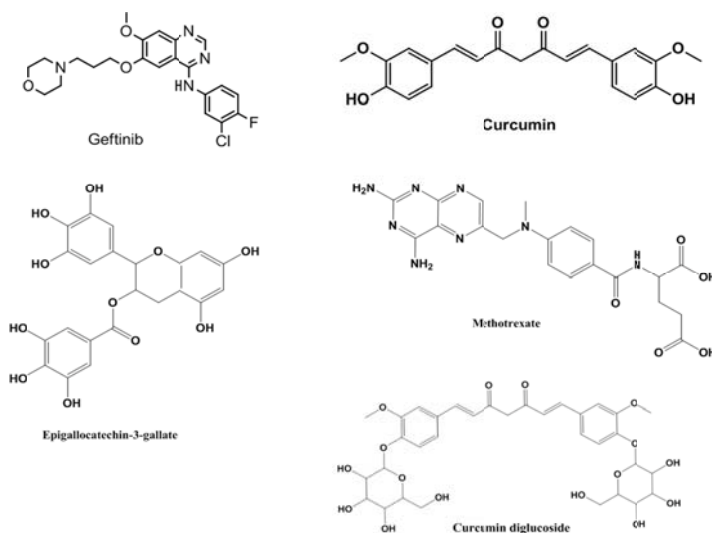


Figure 4. Structures of starting compounds

4 Performing structure-activity relationship calculations

Compounds which could be used for additional further modification analysis were identified from the docking studies. To develop analogues to the identified compounds structure-activity relationship were used to predict the activity of compounds in relation to their structures. It was hypothesized that compounds with similar structures could exhibit similar functional activity. Therefore, the top 5 screened compounds, determined by BioMed CAChe, were taken and compared with 7 lactate dehydrogenase inhibitors having IC₅₀ values (see Table 1) retrieved from the BRENDA database.

Table 1. Training set of lactate dehydrogenase inhibitors having IC50 values for structure activity relationship study

Lactate dehydrogenase inhibitors	IC50 (mM)
8-(phenylamino) naphthalene-1-sulfonic acid	1.1
Naphthalene-2,6-dicarboxylic acid	1.4
6,6'-disulfanediylidipyridine-3-carboxylic acid	4.6
3,7-dihydroxynaphthalene-2-carboxylic acid	5
4,7-diromo-3-hydroxynaphthalene-2-carboxylic acid	5.9
Naphthalene-2,6-disulfonic acid	9.8
3,5-dihydroxynaphthalene-2-carboxylic acid	150

Docking scores of 7 lactate dehydrogenase inhibitors were also evaluated by BioMed CAChe because lactate dehydrogenase is considered a major checkpoint in cancer cell progression, whereas hnRNP B1 has been implicated in for lung cancer progression. Therefore, the known affinity of the lactate dehydrogenase inhibitors was used as a baseline and guide to interrogate potential ligands for hnRNP B1. The correlation between putative hnRNP B1 inhibitors and the 7 lactate dehydrogenase inhibitors with calculated IC50 by TSAR was used to create Partial Least Square equations for prediction.

Table 2. Property calculations of independent descriptors

Independent descriptors	Their Significance
Molecular mass	This descriptor is calculated supposing that the various atomic isotopes occur in their common proportions
Molecular surface area	The calculations use a probe radius of 1.4Å
Molecular volume	It reports the volume within the surface area defined by the van der Waals radii of the atoms
Verloop parameters	Describe the steric influence of substituents in the interaction of organic compounds with macromolecules or drug receptors This factor calculation assumes that all atoms have Van der Waals radii and use these to define the substituent's space necessities
Verloop L (Length parameter)	Defined as the maximum length of the substituent along the axis of the bond between the first atom of the substituent and the parent molecule
Verloop B1 (Width parameter)	Notes the smallest width of the substituent in any direction perpendicular to L
Verloop B3-5	Are analyzed by measuring the width of the substituent, as follows: • in the direction opposite to the axis defined by B1 • in the two directions perpendicular to this axis and the original bond axis
verloop B5 (Final parameter)	Consider the maximum width of the substituent in any direction perpendicular to the bond axis
Kappa 1 index	Demonstrate the degree of complexity of a bonding pattern
Kappa 2 index	Directs the degree of linearity or star-likeness of bonding patterns
Ellipsoid volume	The moments of inertia and principal axes of inertia for a molecule are calculated using the inertia tensor, by following standard methods of calculation (Riley, 1983). These results are reported as Moment 1 Size, Moment 1 Length, <i>etc.</i> The volume specified by these values is calculated and reported as the Ellipsoid Volume The ellipsoid's principal axes are adjusted with the axes of the inertia tensor
Randic topological index	It is the sum of the weighted edges
Ipso atom E-state index	Depicts the state of the generic atom to which the choosen substituent is attached

Structures were loaded into the TSAR workspace sheet making two sets of data, the first was the training set (compounds having known IC50 values) and the second was the test set (compounds under prediction study) using the statistical tool of Partial Least Square (PLS) dimension equations. For quantitative structure-activity relationship analysis IC50 values (mentioned above) and docking scores obtained through BioMed CAChe software were treated as dependent variables. While independent variables contained various factors of substituents such as; molecular mass, molecular surface area,

molecular volume, verloop L, verloop B1, verloop B2, verloop B3, verloop B4, verloop B5, Kappa 1 index, Kappa 2 index, ellipsoid volume, Randic topological index and Ipso atom E-state index. Some descriptors with their significance are described in Table 2.

Results were based on predicted classes and compounds with IC50 values and docking scores. Class distribution showed that compounds in higher class could exhibit improved activity against the hnRNP B1. Based on the data analysis EGCG was found to be potent candidate for further modifications (see Figure 5).

Row ID	Whole Molecule	Subst. 1	Subst. 2	Subst. 3	Name (Whole Molecule)	PUBCHEM_CID	K2 50 (mM)	Docking scores (kcal/mol)	Molecular Mass (Subst. 1)	Molecular Mass (Subst. 2)	Molecular Mass (Subst. 3)	Molecular Surface Area (Subst. 1)	Molecular Surface Area (Subst. 2)	Molecular Surface Area (Subst. 3)
1					naphthalene-2,6-dicarboxylic acid	14357	1.4	-49.286	171.18	171.18	16	32.814	32.814	179.31
2					naphthalene-2,6-difluoroic acid	11390	9.9	-72.225	276.29	276.29	276.29	241.23	241.23	240.91
3					3,5-dihydroxynaphthalene-2-carboxylic acid	68637	190	-49.698	192.18	192.18	192.18	194.52	194.52	194.21
4					3,7-dihydroxy-8-methoxy-1-naphthol		6	-49.11	171.18	199.17	16	32.335	32.335	192.41
5					6-methoxy-1-naphthol		1.1	-19.021	206.28	81.01	206.28	238.47	238.47	79.081
6					naphthalene	128941		-84.513	204.25	204.25	204.25	206.67	206.67	240.9
7					gallic acid	123631		-108.15	144.22	290.12	290.12	167.81	167.81	282.5
8					curcumin diglucoside	48173889		-81.414	179.17	561.55	561.55	189.39	189.39	514.61
9					epigallocatechin gallate	81054		-85.815	445.39	445.39	445.39	379.75	379.75	580.81
10					curcumin	96916		-85.537	111.13	245.27	111.13	136.43	136.43	257.11

Figure 5. Structure activity relationship on TSAR software

5 Analyzing the pharmacophore properties of EGCG

To address the observation that both software packages identified EGCG as having higher affinity making it potent candidate for modification studies of the pharmacophore properties of respective compound were analysed again using BioMed CAChe software. The optimal results suggested that when EGCG docked with hnRNP B1 it was surrounded by 12 amino acids and that it formed a hydrogen bond with 1 amino acid. The model showed a strong interaction of 1.735 angstrom between the carboxylic group present in side chain of Glu108 and hydroxyl group of ring-D (see Figure 6).

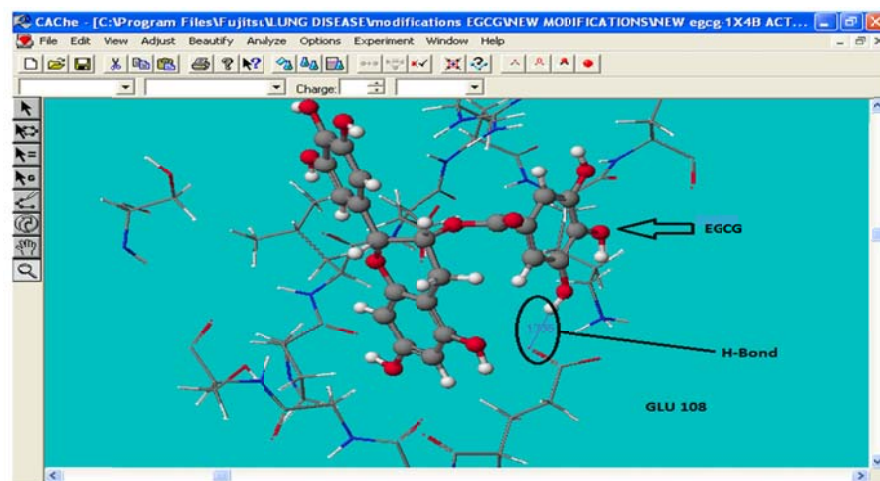


Figure 6. The optimal candidate chosen after docking experiments with H-bond interaction between hydroxyl group of ring D and carboxylic end of amino acid GLU108

Subsequently pharmacophore studies also identified hydrophilic and hydrophobic regions (see Figure 7). These regions were distinguished by color codes where red area focuses on H-bond acceptors and blue region emphasis H-bond donors. Additionally, in the structure, cream regions at some places highlight hydrophobic regions. Regarding hnRNP B1, the region around ring-A and D highlighted a H-bond donor region while H-bond acceptor deficiency was featured at ring-B (For reference see Table 3).

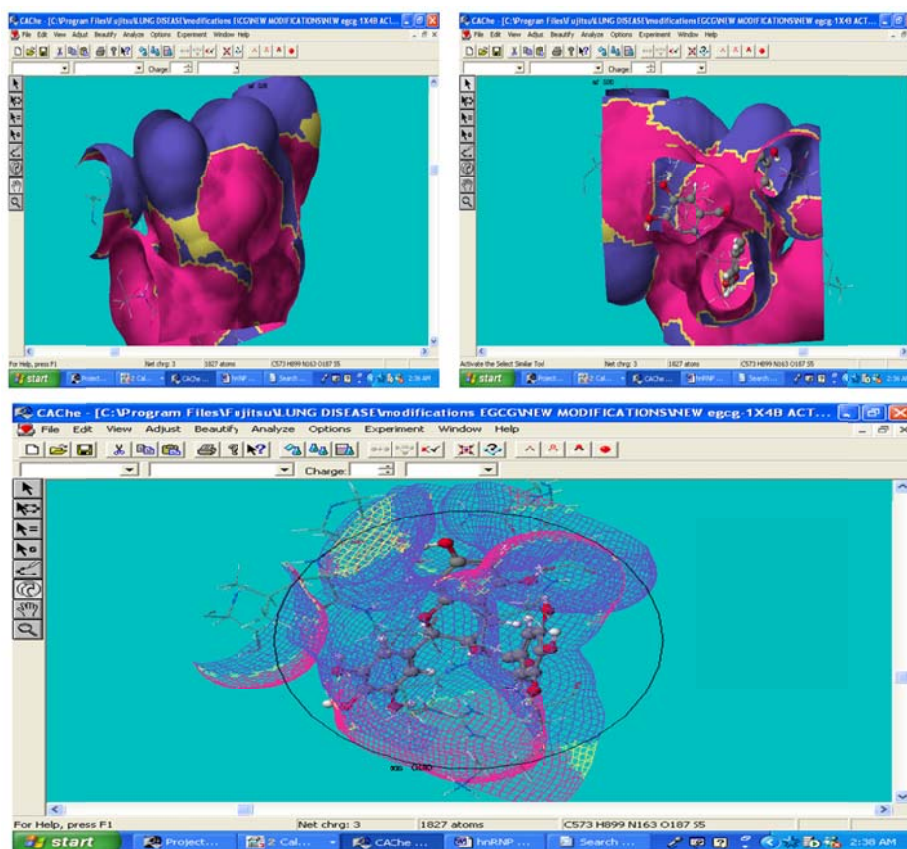
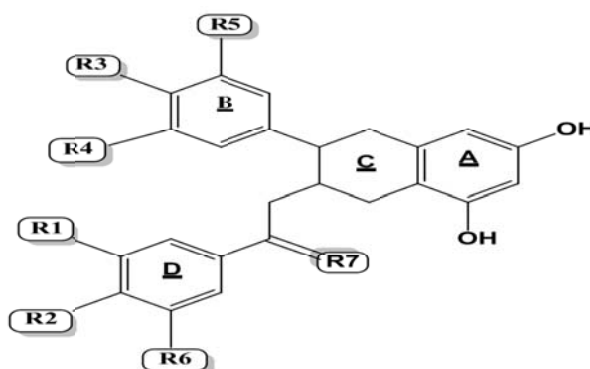


Figure 7. Hydrophobic and hydrophilic regions analyzed during pharmacophore studies

Pharmacophore results obtained during the preliminary analysis guided the process of modifying the best lead compound to optimize the ligand receptor binding.

Table 3. Compounds with improved binding affinity prepared as derivatives of parent compound (EGCG) through computational approach



Compound	R1	R2	R3	R4	R5	R6	R7	Binding affinity (kcal/mol)
IV								-87.042 (EGCG)
IV _a								-91.654
IV _b								-92.050
IV _c								-90.423
IV _d								-89.976
IV _e								-90.357
IV _f								-96.167
IV _g								-89.453
IV _h								-90.653
IV _i								-95.746
IV _j								-88.272
IV _k								-88.644
IV _l								-87.816
IV _m								-87.001
IV _n								-96.307
IV _o								-95.169
IV _p								-94.247

(Table 3 continued on page 55)

Table 3. (continued.)

Compound	R1	R2	R3	R4	R5	R6	R7	Binding affinity (kcal/mol)
IV _q								-93.869
IV _r								-98.665
IV _s								-97.448
IV _t								-95.356
IV _u								-93.798

6 Executing modifications for improvement in the binding of the ligand with hnRNP B1

Further *in silico* modifications of the drug were performed to create a compound with improved binding affinity as compared to parent compound (EGCG). A library of 158 compounds was generated based on bioisosterism and a requirement for hydrophilic and hydrophobic regions. Based on docking scores of compounds IV_n and IV_u results revealed the formation of 3 Hydrogen bonds and 2 Hydrogen bonds respectively between the ligand and the surrounding amino acids in active site of hnRNP B1, as illustrated in Table 4. Further changes in hydrophobic and hydrophilic regions was also seen in improved structures. The listed compounds with higher binding affinities than parent compound are shown in Table 3.

Although computational methods gave new and improved analogues of EGCG, namely compounds IV_n and IV_u the next step in this research is to organically synthesis the compounds according to bio-isosterism whose synthesis is close to the real molecule which will be followed by validation experiments for their inhibitory effects on cell lines.

Table 4. Pharmacophore study of some improved derivatives

Compound	Distance (in Angstroms)	Interaction with amino acids	No. of H-bonds
IV	1.735	Glu108	1
IV _r	2.107	Glu108	2
	2.090	Lys20	
IV _i	1.682	Glu108	1
IV _n	1.869	Glu108	3
	2.074	Arg19	
	1.948	Lys20	
IV _o	2.010	Glu108	1
IV _r	1.704	Glu108	1
IV _t	1.976	Glu108	1
IV _u	2.142	Glu108	2
	1.883	Arg19	

7 Conclusions

The results attained in this paper showed that a slight modification of lead molecules can enhance binding affinity to generate new compound from parent compound. The modeling studies indicated that replacement of the hydroxyl group around ring B gave compounds with improved scores and preserved the precise activity as compared to changes in hydroxyl groups of ring A, C or D in EGCG against hnRNP B1. Further a strong bond between ring D and GLU108 proved to be effective in these *in silico* inhibition studies. It was observed that any disruption in this linkage decreases the calculated affinity of compound. Furthermore, the increase seen in the docking score, such as that seen for the modified compounds such as IV_n and IV_u formed additional hydrogen bonds with ARG 19 and LYS 20.

Acknowledgment

We would like to express our sincere thanks and appreciation to Dr. Derek Kennedy, Eskitis Institute for Cell and Molecular Therapies, Griffith University, Brisbane, Australia for positive comments and corrections of this manuscript.

References

- [1] Horn L, Pao W, Johnson DH. Neoplasm of the lung. In Harrison's Principles of Internal Medicine. 18th edn; Longo DL, Fauci A, Kasper D, *et al.* 2012.
- [2] Lung Carcinoma: Tumors of the Lungs. Merck Manual Professional.
http://www.merckmanuals.com/professional/pulmonary_disorders/tumors_of_the_lungs/lung_carcinoma.html
- [3] Thun MJ, Hannan LM, Adams-Campbell LL, *et al.* Lung cancer occurrence in never-smokers: an analysis of 13 cohorts and 22 cancer registry studies. *PLoS Medicine*. 2008; 5 (9). PMID:18788891. <http://dx.doi.org/10.1371/journal.pmed.0050185>
- [4] Sueoka E, Sueoka N, Goto Y, *et al.* Heterogeneous Nuclear Ribonucleoprotein B1 as Early Cancer Biomarker for Occult Cancer of Human Lungs and Bronchial Dysplasia. *Cancer Res*. 2001; 61(5): 1896-1902. PMID:11280744.
- [5] Greenberg AK, Lee MS. Biomarkers for Lung Cancer: Clinical Uses. *Curr Opin Pulm Med*. 2007; 13(4): 249-55. PMID:17534168. <http://dx.doi.org/10.1097/MCP.0b013e32819f8f06>
- [6] Lounnas V, Ritschel T, Kelder, J, *et al.* Current progress in Structure-Based Rational Drug Design marks a new mindset in drug discovery. *Comput Struct Biotechnol J*. 2013; 5. <http://dx.doi.org/10.5936/csbj.201302011>
- [7] Sueoka E, Sueoka N, Iwanaga K, *et al.* Detection of plasma hnRNP B1 mRNA, a new cancer biomarker, in lung cancer patients by quantitative real-time polymerase chain reaction. *Lung Cancer*. 2005; 48(1): 77-83. PMID:15777973. <http://dx.doi.org/10.1016/j.lungcan.2004.10.007>
- [8] Shin KH, Kang MK, Park NH. Heterogeneous nuclear ribonucleoprotein G, nitric oxide, and oral carcinogenesis. *Nitric Oxide*. 2008; 19(2): 125-32. PMID:18474262. <http://dx.doi.org/10.1016/j.niox.2008.04.012>
- [9] David CJ, Chen M, Assanah M, *et al.* HnRNP proteins controlled by c-Myc deregulate pyruvate kinase mRNA splicing in cancer. *Nature*. 2010; 463(7279): 364-8. PMID:20010808. <http://dx.doi.org/10.1038/nature08697>
- [10] Vander Heiden MG, Christofk HR, Schuman E, *et al.* Identification of small molecule inhibitors of pyruvate kinase M2. *Biochem Pharmacol*. 2010; 79(8): 1118-24. PMID:20005212. <http://dx.doi.org/10.1016/j.bcp.2009.12.003>
- [11] Auld D, Shen M, Skoumbourdis AP, *et al.* Identification of activators for the M2 isoform of human pyruvate kinase. Probe Reports from the NIH Molecular Libraries Program, Bethesda (MD): National Center for Biotechnology Information (US). 2010. <http://www.ncbi.nlm.nih.gov/books/NBK56225/>
- [12] Madern D. Molecular evolution within the L-malate and L-lactate dehydrogenase super-family. *J Mol Evol*. 2002; 54 (6): 825-40. PMID:12029364. <http://dx.doi.org/10.1007/s00239-001-0088-8>
- [13] Burd CG, Swanson MS, Görlach M, *et al.* Primary structures of the heterogeneous nuclear ribonucleoprotein A2, B1, and C2 proteins: A diversity of RNA binding proteins is generated by small peptide inserts. *Proc. Natl. Acad. Sci*. 1989; 86(24): 9788-92. PMID:2557628. <http://dx.doi.org/10.1073/pnas.86.24.9788>