Original Article

IDENTIFICATION OF POTENT BROMODOMAIN4 (BRD4) INHIBITORS BY ENERGY-PHARMACOPHORE BASED VIRTUAL SCREENING TO TARGET BRD4-NUT MIDLINE CARCINOMA

SAPAM TULESHWORI DEVI, HIMANI TANDON, DINAKARA RAO AMPASALA*

Centre for Bioinformatics, School of Life Sciences, Pondicherry University, Pondicherry-605014, India.
Email: ampasaladr@bicpu.edu.in

ABSTRACT

Objective: Protein-protein interactions (PPI's) have been used as a target in various diseases. One such major role of PPI's comes when the protein Bromodomain4 (BRD4) reads the epigenetic changes in the histones and regulates transcription. This protein has been shown by various research groups to be linked to a rare form of cancer called BRD4-NUT midline carcinoma. The present study incorporates understanding the role of BRD4 in such cancer and as is directed towards finding new lead compounds to target the PPI involved.

Methods: To find potential lead molecules against BRD4 protein, contact based and e-pharmacophore based virtual screening studies approach was adopted with the use of PHASE and E-pharmacophore module of the Schrödinger Maestro tool. Based on the pharmacophore hypothesis developed, virtual screening was performed for 22, 70, 000 Clean Lead-like compounds from the ZINC database by Virtual Screening Workflow of Schrödinger Maestro. Further Molecular dynamics simulations by GROMACS 4.5.5 were performed to study the energetics and stability of the top most docked ligands with BRD4 protein.

Results: Pharmacophore based virtual screening studies results in the retrieval of three potential lead molecules, ZINC68155904, ZINC67910065, and ZINC6710456 from ZINC database which interacts with BRD4 protein with Glide score of-9.98,-8.31,-7.61 kJ/mol respectively. The desired interaction of this ligands with Asn140, Pro82 and Tyr 97 of BRD4 protein showed that the final hits have the potency of forming a stable complex. Molecular dynamics simulations studies also support the stability of the BRD4-ligand docked complex.

Conclusion: The above study shows three compounds obtained viz ZINC68155904, ZINC67910065, and ZINC6710456 may serve as potential lead compounds which can act against BRD4 protein.

Keywords: Bromo-domain, Cancer, Epigenetic, Molecular Docking, Protein-protein interactions, Pharmacophore modelling, Transcription, Virtual screening.

INTRODUCTION

Epigenetic changes are heritable changes that can alter the gene expression at the transcriptional level without any significant changes in the DNA sequence. As new targets are being sought for drug designing, epigenetic markers are being studied extensively as one of the possible new targets especially in cancer. For example, in the case of NUT midline carcinoma which is a rare genetically defined epithelial cancer [1], the interaction of bromodomain epigenetic mark readers (BRD4) fused with NUT was blocked to control the aggressiveness of carcinoma [2]. Bromodomains are the sole protein domains which are known to recognize acetyl-lysine residues on histone proteins [3] and this recognition further helps in chromatin remodeling and gene transcription in normal conditions. Bromodomains containing proteins are of special interest as mostly they are the components of transcriptional factors as well as elements of epigenetic memory [2, 4]. One protein of importance is Brd4 protein belonging to bromodomain and ET-domain (BET) protein family [5, 6]. These domains shared a highly conserved fold composed of a left-handed bundle of four alpha helices (α_Z, α_A, α_B, α_C) connected by a diverse loop regions (ZA and BC loops) that contribute to substrate specificity [2, 7]. All the amino terminal bromodomains of BRD4 family proteins exhibit high levels of sequence conservation whereas the carboxy terminal domains are more divergent [2]. Covalent modifications take place at the amino termini of the core histone proteins in nucleosomes after translation processes such as acetylation, methylation, ubiquitination, SUMOylation and phosphorylation which have an essential role in gene regulation. All of these modifications have their reader proteins present in the cells and one such epigenetic reader protein is the BRD4 protein belonging to BET family of bromodomain [8]. BRD4 protein is one of the important proteins present at the crossroads of gene expression. It is one of the important mediators of transcriptional elongation which functions to recruit the positive transcription elongation factor complex (P-TEFb) [9-11]. The two bromodomains of the BRD4 protein is sufficient enough to interact with histone H3 and H4 proteins [12, 13]. The P-TEFb protein binds with the BRD4 proteins by interacting with the P-TEFb interacting domain (PID) present at the C terminal region of BRD4 protein [14, 15]. Checking the activity of P-TEFb complex has proved to be an emerging strategy for the treatment of chronic lymphocytic leukemia as one of the core component, Cyclin dependent kinase-9 is considered as a validated target in this particular disease [16-21]. BRD4 protein also associates with chromatin and binds to replication factor C indicating its role in cell progression events [22]. Also, expression of BDR4 gene controls G_2-M transition whereas if BRD4 gene is over expressed, it results in the inhibition of G_1-S phase transition indicating its importance in cell proliferation and growth events [12]. Moreover, studies have identified BRD4 as a potential partner in t (15;19)-associated fusion oncogene [23]. The gene encoding BRD4 is present in chromosome number 19 whereas chromosome 15 has the gene for Nuclear protein in Testis (NUT) which expresses only in testis under normal conditions [1]. Thus, by a translocation mechanism of chromosome 15 containing the BRD4 gene, rearrangement of the gene NUT present in chromosome 19 occurs resulting in the formation of BRD4-NUT Fusion Oncogene and this mechanism has been identified in a highly deadly form of carcinoma [1]. Hence BRD4 is considered as the first oncogene from the BET family of bromodomain gene and it proves to be a potential drug target for finding cure against carcinomas and cancers [1]. The role of BRD4 is not only implicated in cancer, but has been shown to play a role in transcriptional regulation of viruses such as HIV [24] and EBV [25] as well as the degradation of HPV [26]. Moreover the search for small molecule inhibitors for BRD4 is at its infancy.

With knowledge of considering BRD4 as a potential drug target for cancer, various researches are being motivated for the identification

of inhibitors against BRD4 protein. In such an attempt, a small molecule 'JQ1' a novel thieno-triazolo-1, 4-diazepine compound was reported as a ligand acting against BRD4 protein in which the triazole ring of the ligand forms hydrogen bond with the evolutionary conserved Asp 140 of the protein. The ligand 'JQ1' bind competitively with the chromatin and reduces the cell growth and differentiation in NMC (NUT midline carcinoma) cells lines [2].

Further search for potent ligands against BRD4 proteins lead to the identification of benzodiazepine (BZD) containing compound GW841819X which exclusively binds to acetyl lysine recognition domain of BET family and hence proved to be an efficient lead compound [16, 27]. Since benzodiazepine containing compounds proved to be an effective lead molecule for bromodomains, further work were carried out which led to the identification of BzD (benzodiazepine) and BzT (triazolo-benzotriazepines) derivatives as an effective alternate lead molecules that specifically binds to the acetyl lysine binding site of the bromodomains containing protein with nanomolar potency [16]. Later a researcher group discovered serendipitously a novel class of compounds inhibiting the bromodomain-histone interactions which contains an entirely different backbone core. The compound discovered is dimethylisoxazole derivative, 3, 5-dimethylisoxazole which gave the platform to search for more other novel scaffolds compounds to acts as bromodomain inhibitor [28]. Recently, in 2012, Chun-wa Chung *et al*, reported some fragments targeting the bromodomainAcK pocket by using the approach of fragment based drug discovery and various computational approaches [29]. In continuation of the previous experiments, Bamborough tried to optimize the hits obtained in the first part of the experiments as well as the dimethylisoxazole derivatives by the approach of the fragment based drug discovery and structure-based rational optimization of the resulting fragments. This led to the successful discovery of Phenylisoxazole Sulfonamides as the drug like inhibitors which will provide better opportunities in curbing the various diseases associated with bromo-domains [30].

Henceforth, still with the continuation of finding more potent inhibitors of bromodomain containing BRD4 protein, our study made an attempt to search for improved compounds which can be later recognized as potential lead molecules. The study takes in account all the currently available BRD4 inhibitors and pharmacophore hypothesis model was constructed based on the structure of these ligands by using the approach of pharmacophore studies. The best hypothesis model retrieved was then used to screen compounds from the chemically available ZINC database by virtual screening studies approach. The retrieved compounds were then docked with the crystal structure of bromodomain containing BRD4 protein having PDB ID 3MXF. From the docking results, the best protein-ligand docked complex was further subjected to molecular dynamics studies to calculate the energetics as well as to check the stability of the complex.

MATERIALS AND METHODS

Materials

Seven BRD4 proteins with their respective inhibitors bound to it were retrieved from Protein Data Bank to develop the pharmacophore model (both by energy based as well as contact based pharmacophore). The PDB ID's of the seven BRD4 proteins with the ligands bound to it are 2yel, 3mxf, 3p5o, 3u5l, 3zyu, 4a9i, 4a9m, the structure of which are given in fig. 1.
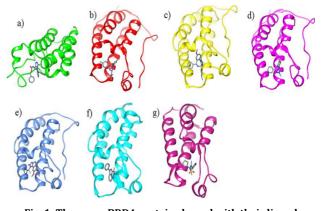
Methods

Protein preparation

Protein Preparation Wizard of Maestro Schrodinger was used to prepare the seven proteins bound with their inhibitors for the use of study of pharmacophore modelling [31]. In protein preparation, bond orders were assigned, if necessary hydrogen atoms are being added, as well as waters which are beyond 5 Å from the hetero groups is deleted. By the use of the PRIME algorithms, missing side chains and loops of the proteins can be filled if needed. Finally the proteins are energy minimized by using the OPLS2005 force field allowing the heavy atoms to converge to RMSD at 0.30 Å.

Ligand docking/refinement of the protein-docked complexes

The ligands bound to each protein are separated and prepared for the docking process by using the LigPrep module of the Maestro Schrodinger [32]. The seven proteins retrieved from the PDB database are re-docked again to obtain the Glide Docking energetics value such as Glide XP score [31]. The Receptor Grid was generated for each protein complex by placing a grid around the ligand with a Van der Waals radius scaling factor of 1.0Å and partial charge cut off of 0.25 Å [33, 34]. After the grid generation, the proteins were re-docked again with their respective ligands by using the Glide XP docking methodology. The ligands were allowed to dock flexibly as well as XP descriptor information was selected to get the detailed energetics information of the docking process which would help to identify the important pharmacophoric features of the ligands.

Fig. 1: The seven BRD4 proteins bound with their ligand retrieved from PDB. a) 2YEL b) 3MXF c) 3P5O d) 3U5L e) 3ZYU f) 4A9I g) 4A9M

Pharmacophore Generation

For the development of pharmacophore generation, two approaches are used for the study. The contact based pharmacophore method and the energy based pharmacophore method will be used to develop the best hypothesis which can be used to screen the compounds which can inhibit the BDR4 proteins with high potency.

a) Contact based pharmacophore modelling

Contact based pharmacophore modeling of the ligands was performed by using the Schrodinger Maestro Phase [35, 36]. Based on its fine-grained conformational sampling and scoring procedures, pharmacophore hypothesis was created which can then be used to screen the databases for potential lead compounds. For our study, four chemical features were selected to generate the pharmacophore which are hydrogen bond acceptor (A), hydrogen bond donor (D), hydrophobe (H) and aromatic ring (R) with the different seven ligands bound to the BRD4 proteins to find a common pharmacophore out of the seven ligands.

For our study the following conditions were considered. The conditions include the following:

Condition A-A1: where the maximum number of pharmacophoric sites are kept as 6 with a minimum number of sites as 4. A2: where the pharmacophore hypothesis must match at least 6 or 7 of the active groups.

Condition B-B1: where the maximum numbers of pharmacophoric sites are kept as 5 and minimum pharmacophoric sites are kept as 5.

B2: where the pharmacophore hypothesis should match at least 5 or 7 active groups.

b) Energy based pharmacophore modelling

It is a hybrid concept whereby pharmacophore modelling and screening of the compound from the databases are coupled with the

Glide XP docking scoring function/energetics to compute the importance of each pharmacophore features and rank them accordingly. This methodology implements the energetics of the Glide XP docking terms to identify the main pharmacophoric features of the docked complexes [37]. The energy based pharmacophore was generated by using the Docking Post-Processing tool from the Scripts menu of the maestro graphical user interface. The pose viewer file of re-docked ligand bound BRD4 protein complexes is used as an input for generating the pharmacophore hypothesis. Here also, the maximum number of features was set to 5, 6, 7 and 8 which resulted in 4 different hypotheses. The pharmacophore was constructed by using the fragment mode of the e-pharmacophore tool.

Database creation

The compounds to be used for the creation of the database were downloaded from ZINC Database maintained at University of California, San Francisco available at (zinc. docking. org) [38]. "Clean Lead-Like" subset of compounds was chosen for our study and around 22, 70, 000 small molecules were downloaded from it. The ligands were divided into 20 subsets each containing 45000 compounds by using the Schrodinger utility sdsubset.

Ligand preparation and ADME test

The ligands which are downloaded need to be prepared for the further use of virtual screening and docking purposes by using the LigPrep module of Maestro Schrodinger [39]. Hence bond orders and the chirality of the ligands need to be checked, conversion of 2D atom structures into 3D all atom structures, if there are any stereoisomers and tautomer's present for the ligands, it need to be created, as well as optimization and minimization of the ligands need to be performed. Thus for our study, by using the OPLS_2005 force field, ionization of the ligands was done at a target pH of 7.0+/-2.0 allowing to generate four stereoisomers and four tautomer's each per ligand as well as the generation of at least 1 low energy ring conformation per ligand. If there are any high energy tautomer states generated for each ligand, it is being removed.

Qik Prop filtering

The ligands which are prepared from the LigPrep are again filtered by using QikProp [40] module to calculate the ADME (Absorption, Distribution, Metabolism and Excretion) properties. ADME test for calculating the pharmacological properties are predicted to check the drug efficacy and toxicity before passing through the clinical trials for developing lead compounds and drugs [39]. Ligands were filtered based on the Lipinski's rule of 5 and the ligands with reactive functional groups are also filtered.

Database creation by phase

The filtered ligands from the previous step are now used to create database in PHASE [35, 36] to search pharmacophore hypothesis obtained from the pharmacophore modelling to search for potent lead compounds. By using the Manage 3D Database tool of PHASE, a database containing the filtered ligands was created which can be used to search against a pharmacophore hypothesis.

Testing the hypothesis

The various pharmacophore hypotheses obtained from the previous step by both contact based and energy based pharmacophore modelling are tested for their ability to retrieve potential compounds from the database. This was done by adding the known ligands to phase database and using the various generated hypothesis for matching the ligands in the database. The results was then retrieved and analyzed further for checking the ability of the hypothesis.

Database searching and clustering

Database searching involves searching the prepared library of compounds with the best retrieved pharmacophore hypothesis after the validation. The database searching was performed in two steps, firstly in the find step, the essential pharmacophoric features of each

conformer of a particular molecule are searched for geometric arrangements of site points that match the hypothesis in both feature types and inter-site distances. If the features of a particular conformer correspond to the hypothesis generated, then the conformer is referred to as a match. In the second step known as the fetch step, the conformer associated with each match is retrieved from the database and aligned to the hypothesis. At this step, the match conformer is now called as the hits which are retrieved from the database. The second fetch step is run automatically after the find step by the PHASE tool.

Clustering and selection of the hit molecules

Clustering is a technique which allows molecules or conformers with similar structural features to group together and processed further. The matched ligands retrieved from the above steps were clustered on the basis of their dendritic fingerprints by using Schrodinger Canvas [41]. Thus the hits obtained were clustered based on certain criteria such that similar hits were grouped together. For our current study, clustering was done based on the fingerprints k clustering method. Various other clustering methods are available in Schrodinger Maestro such as Linear, Dendritic, Radial, MACCS, MOLPRINT2D, Pairwise, Triplet and Torsion are available for performing the fingerprint clustering [41]. A Fingerprint represents the binary strings which encode the presence or the absence of sub structural fragments of a particular conformer molecule to measure inter molecular structural similarity. Hence for our study dendritic fingerprints k clustering approach was used, which code for both linear and branched features [42]. For each subset, 10 clusters were created by using the k-means clustering methodology. Finally from each cluster, the final hits were selected by observing the clustered subset for highest fitness score which result in the retrieval of 230 ligands respectively.

Molecular docking studies

All the retrieved ligands were allowed to dock with the first bromo-domain containing BRD4 protein with PDB ID 3MXF. The previously prepared protein 3MXF was used for our study and receptor grid generation was done by using the receptor grid generation tool of GLIDE Maestro [33, 34]. This tool calculates the existing interactions between a protein and a ligand in terms of Glide energy and Glide docking score which gives the detailed account of the binding energy and affinities between them [33, 34]. Firstly, by using the standard precision algorithm (SP) of Glide tool which uses softer and less stringent functions [33], the retrieved 230 ligands were allowed to dock by using the generated grid of BRD4 protein, 3MXF. After SP docking, few ligands which match a certain docking criteria score were made to proceed for Glide extra precision (XP) which gives more severe penalties to the ligand poses which violate the algorithms for docking procedure [33, 34]. Thus both the docking algorithms were allowed to dock flexibly allowing sample nitrogen inversions and ring conversions as well as Epik state penalties are added to the docking score. To soften the potential for the non-polar parts of the ligand, van der Waals radii were scaled down with a scaling factor of 0.8 and an atomic charge with a cut off value of 0.15 has been set as the docking parameters. For each docking run, 10000 poses and 1 poses per ligand were allowed to retrieve as an output for the docking purposes. Post-docking minimization was performed after docking for at least each 5 poses of a particular ligand.

Molecular dynamics simulation

The ligands with glide score more than the set cut-off value were selected for MD simulation. Total 3 compounds passed the criteria for which the simulation was carried out for 3ns. The GROMACS 4.5.5 [43] molecular dynamics package and GROMOS43 al force field was used to analyze complex stability. PRODRG2.5 [44] server was used to build the GMX topology for the ligand. Generated coordinates were copied and pasted in PDB complex file. Water molecule coordinates were also pasted after pasting ligand coordinates. The complex was solvated with the water model using a 0.80 nm cubic box. Energy minimization was performed by using the

Steepest Descent minimization algorithms. Periodic boundary conditions were applied in all directions, and the system was neutralized by replacing water molecules with 1 chloride counter ion. A twin-range cutoff was applied to long-range electrostatic interactions using the PME method and 1.0 nm for Vander Waals interactions. Equilibration MD for both temperature (300 K) and pressure (1 atm.) were carried out for 100 ps. Potential energy, RMSD, RMSF, H-bond interactions, radius of gyration graph were generated and results were analyzed.

RESULTS AND DISCUSSIONS

Refine docking

The seven BRD4 Bromo domain containing proteins are re-docked again with their respective ligands so as to calculate the energetics and to incorporate the energy value to develop hypothesis pharmacophores by the process of pharmacophore modelling. The value of the docking score as calculated by the Glide tool is tabulated as given in table 1.
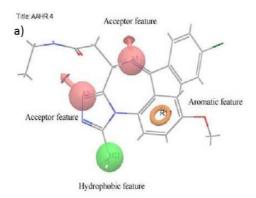
Table 1: Results of the refine docking of the BRD4 proteins bound with their respective ligands

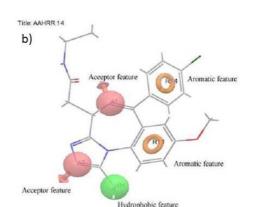| S. No.	Protein-ligand complex	Glide score	Glide energy (kJ/mol)
1.	3MXF	-9.071	-45.428
2.	2YEL	-7.856	-40.357
3.	3P5O	-7.335	-49.892
4.	3ZYU	-7.269	-44.945
5.	3U5L	-6.159	-37.672
6.	4A9M	-6.376	-33.776
7.	4A9I	-6.292	-32.098

Pharmacophore modelling

Contact based and e-pharmacophore based pharmacophore modelling was used to retrieve the common pharmacophores from the seven ligands bound to the retrieved BRD4 protein from PDB. The pharmacophores models are given in fig. 2 and fig. 3.
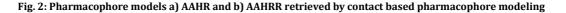
Contact based pharmacophore methodology

Contact based pharmacophore methodology was adopted for the seven ligands by using the PHASE GUI interface of Schrodinger Maestro [35]. In this methodology, only the knowledge of the ligands is needed and any information regarding the structure of the receptor is not needed. The pharmacophore hypothesis generated from phase with the 7 ligands of the 7 bromo-domain containing BRD4 proteins had 4 features when generated using condition A as mentioned in the Methodology section where as by using condition B, the best hypothesis features comprises of five pharmacophores features.

The pharmacophoric features contributing to the hypothesis consists of Hydrogen bond acceptor (A), Hydrophobic feature (H), and Aromatic ring feature (R) resulting in the generation of AAHR and AAHRR as given in fig. 2 and table 2.

Fig. 2: Pharmacophore models a) AAHR and b) AAHRR retrieved by contact based pharmacophore modeling

Moreover, these top scoring hypotheses from both the conditions were put on test for its ability to retrieve the known ligands from the database.

Energy based pharmacophore modelling

The hypothesis generated from energy based pharmacophore modelling utilizes both the features of the receptor proteins as well as the ligands. It also incorporates the energetic terms of the Glide XP docking to the atom centers of the ligands to map the important pharmacophoric features of the ligands which will lead to the generation of the pharmacophore hypothesis [45]. The features finally obtained are optimized based on certain cutoffs which are defined as: sites with less than half of the heavy atoms contributing to the pharmacophore feature are excluded from final hypothesis generated. Consequently, the resulting energy pharmacophore can be directly use for the search of more potent ligands from the database generated by Maestro Schrodinger PHASE GUI [35]. With maximum features of the pharmacophoric sites set to 5, 6, 7, and 8 by excluding the receptor volumes with a van der Waal's scaling 0.5; it results in the generation of 3 different hypotheses. The main pharmacophoric sites contributing to the hypothesis are 1) Hydrogen bond acceptor (H), 2) Aromatic ring feature (R), 3) Hydrophobic feature (H) but no donor feature contributed to the hypothesis. The energy based pharmacophore hypotheses developed are listed below as given in the fig. 3 and table 2.

Testing of the retrieved hypothesis

Testing or validation of the generated hypothesis was done with the known ligands bound to their respective Bromo-domain containing BRD4 proteins. Hence the generated pharmacophore hypothesis are fitted with the known ligands and the pharmacophore which retrieved the maximum ligands is considered as the best hypothesis which is further used to screen the database that was already created with the 'Clean-Lead Like' subsets of compounds

downloaded from ZINC database. In case of contact based pharmacophore hypothesis validation, for both of the pharmacophores features generated, four inhibitors originally bound to the BRD4 containing proteins ie. I-BET, BzT-7, JQ1, GW841819X match both of the pharmacophoric features ie. 'AAHR' and 'AAHRR' pharmacophore hypothesis as given in table 2.
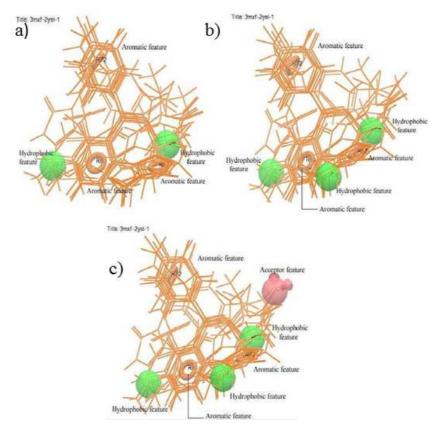

Fig. 3: The pharmacophore models retrieved from energy based pharmacophore modeling a) RRRHH b) RRRHHH c) RRRHHHA

Table 2: Results of the pharmacophore modelling hypothesis testing with known ligands

S. No.	Hypothesis name	Hits retrieved
1.	Contact based pharmacophore modelling	I-BET, BzT-7, JQ1, GW841819.
	AAHR	I-BET, BzT-7. JQ1, GW841819.
	AAHRR	
2.	Energy based pharmacophore modelling	
	RRRHH	Bzt-7, I-BET762, JQ1, 4a9m, I-BET151.
	RRRHHH	JQ1, BzT-7
	RRRHHHA	JQ1, 4a9m, BzT-7, IBET762, GW841819X

The results retrieved showed that both the hypothesis was biased towards one type of ligand backbone. This is because all the compounds retrieved from it belong to similar set of backbone, although the input has ligands from more than one type of backbone, And hence, the two pharmacophore hypothesis generated by contact based pharmacophore was discarded for further studies.

The pharmacophore generated from energy based method utilizes both knowledge of the ligand as well as the structure of the receptor protein and hence from the four hypothesis generated the challenge is to choose the best hypothesis out of the three. In order to validate the hypothesis, the hypotheses were also put on test for their ability to retrieve known compounds.

All of the hypothesis retrieved different number of compounds from the databases but the hypothesis with features RRRHHHA retrieved 5 ligands (when 5 out of 7 features were matched) having compounds from both type of backbones 9 (table 2). If the criteria were set to strictly match 7 out of 7 or 6 out of 7 features, the hypothesis was able to retrieve only two compounds. This was due

to the fact that the ligands had diversified backbones and side chains and all of them could not be retrieved if strict conditions were applied. So, flexible conditions had to be given in order to obtain a compromised hypothesis. The final hypothesis 'RRRHHHA' was thence used for screen the database generated by PHASE for more lead molecules with better potency.

Database searching and clustering of ligands

K-mean fingerprints clustering method by Schrodinger Canvas was used to cluster the hits molecules retrieved after pharmacophore modelling [41]. This algorithm starts to group together the data from a given dataset by taking random number of k mean. Based on these k means, new clusters of ligands were created by associating each cluster to the nearest mean. The centroid of the k cluster becomes the new mean and this process is iterated till the convergence of clusters [46]. Thus after clustering of the ligands as resulted from the screening of ligands by virtual screening through pharmacophore modelling, it resulted in the retrieval of 230 ligands which will be further narrowed down into potential lead like compounds.

Molecular docking

Molecular docking was performed with the unknown ligands retrieved from the pharmacophore modelling into the binding cavity of the 3MXF protein with proper orientation and conformation. The main aim for performing molecular docking is to screen down from the 230 ligands retrieved to a few potentials ligands with good potency leading to the development of new lead molecules which can target against BRD4 protein. First the Glide SP docking was performed followed by the Glide XP docking which gives the results in the form of Glide energy score. Known ligands above a docking glide score (G score) of-6.159 kJ/mol was chosen for the next Glide XP docking run as mentioned in the methodology section. The limit for the docking score was set to-6.159 kJ/mol as it is the lowest docking score obtained after performing Glide SP docking with the known ligands of the BRD4 proteins.
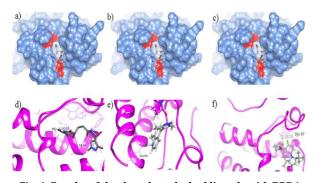

Fig. 4: Results of the three best docked ligands with BRD4 protein. Fig. a-c indicates the surface view of the protein ligand complex with the respective ligand colored as grey and the interacting residues in red color. Fig. d-f shows the protein-ligand interactions, with ligand in the stick form in grey color whereas the particular residue shown in wire form in magenta color with the dashed line as H bond

The 230 ligands were subjected to GLIDE SP docking first resulting in the retrieval of 100 compounds above the docking limit score G score. These 100 compounds are then again re-docked again with more stringent

algorithms of Glide docking XP, resulting in the retrieval of 8 ligands out of the 100 ligands. Out of all the 8 ligands, the top three best ligands with a G score above-7.5kJ/mol along with their hydrogen bond interacting residues as given in table 3 are further subjected to simulation studies. The structure of the docked ligands with the receptor is given in fig. 4.

The first best scoring ligand with ZINCID number 68155904 binds to the binding cavity of the BRD4 protein in the same way as the known ligands. The ligand forms two hydrogen bonds with residues Asn140 and Pro82 with the BRD4 protein respectively. The oxygen O1 atom of the ligand acts as the hydrogen acceptor whereas the atom HD21 of the Asn140 donates an electron to its acceptor forming a stable hydrogen bond between the ligand and the protein with a bond distance of 1.9Å. Whereas, in the second hydrogen bond formed between the ligand and the Pro82, H24 atom of the ligand acts as the hydrogen donor atom donating its electron to its hydrogen accepting atom O of Pro82 forming hydrogen bond with a distance of 1.9Å. The list of protein residues displaying hydrogen interactions with the ligand is given in table 4 along with the energetic details. The second ligand with ZINC ID 67910065 binds with BRD4 protein at its binding cavity with a G score of-8.31kJ/mol forming a hydrogen bond with atom HD21 of Asn 140 as the hydrogen bond donor donating its electron to the N1 atom of ligand. The bond distance between the protein and ligand is quite stable with a bond distance of 2.01 Å. The third ligand ZINC ID 67510456 interact with BDR4 protein with glide XP docking score of-7.611 kJ/mol forming a hydrogen bond with Tyr 97 at a distance of 2.5Å. The ligands H19 atom acts as the hydrogen bond donor with donating its atom to the hydroxyl group of Tyr 97 of BRD4 protein. Further, to study the energetics and to check the dynamic stability of the protein-ligand complex, the three best docked ligand complexes were subjected to molecular dynamics simulation studies.

Molecular dynamics simulation

Molecular Dynamics Simulation was performed by using Gromacs 4.5.5 tool. Various energetics calculations were studied by using the analytic tools of Gromacs itself and graph were plotted by using the XmGrace tool. The RMSD and the RMSF deviations were also studied extensively to understand the dynamics of the protein. The simulation was carried out for 3ns each for the top most docked ligands with BRD4 protein and is given in fig. 5.
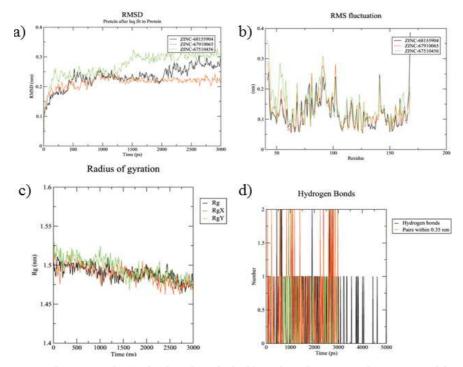

Fig. 5: Molecular Dynamics Simulations results for the three best docked ligands with 3MXF. a) The RMSD graph b) The RMSF graph c) The Radius of gyration graph d) The hydrogen bond graph

Table 3: Results of the Glide XP top docked ligands with 3MXF out of the 100 ligands

S. No.	Ligand number	Glide score (kJ/mol)	Glide E-model energy (kJ/mol)	H bond interactions
1.	ZINC68155904	-9.983	-42.352	Asn 140, Pro 82
2.	ZINC 67910065	-8.310	-43.106	Asn 140
3.	ZINC67510456	-7.611	-42.579	Tyr 97

The energetic values of the molecular dynamics simulations of the three proteins are given in table 4.

Table 4: The various energetic details of the best docked ligands with BRD4 protein by Molecular Dynamics Simulation

Parameters	BRD4-Lig1	BRD4-Lig2	BRD4-Lig3
Grid cell	7.205*7.205*7.205	7.205*7.205*7.205	7.205*7.205*7.205
Number of SOL molecules	11736	11733	11722
Steps of steepest descents EM converged (Fmax<1000)	436	427	537
Steepest descent potential energy (kJ/mol)	-5.8431075e+05	-5.8384175e+05	-5.8462969e+05
Average Potential Energy kJ/mol)	-507161	-5.06862	-506307
Maximum Potential Energy	-5.0493e+05	-5.045e+05	-5.0458e+05
Average total energy	-415903	-415646	-415151
RMSD	0.26 nm	0.22 nm	0.32 nm
Equilibration period	600ps	1300ps	1500ps

Root mean square deviations (RMSD) give an idea of convergence of a structure of a protein towards an equilibrium state. The RMSD values for the first two protein-docked complexes came out to be 0.26 nm, 0.24 nm respectively. Whereas the RMSD plot of the third protein complex was found to be different from the previous two protein-ligand complexes. The RMSD of the BRD4-ZINC67510456 initially increases till 0.26 nm and remain constant till 1400ps. Again the RMSD graph from 0.26 nm to 0.32 nm at 1500 ns and remains constant till the end of the simulation. Analysis of RMSD graph showed the equilibration period for the proteins to be at 500ps, 800ps, and 1500ps as shown in the fig.5a.

Root mean square fluctuations were also studied to study about the mobility of the protein structures. It gives an idea of the flexibility regions of the proteins. The RMSF graph shows the maximum fluctuations from the residues number 85 to 110 as shown in the fig.5b. This region corresponds to the loop regions connecting the helices of the protein. Whereas the particular amino acid residue which binds with the ligand is not flexible indicating that the ligand and the protein are able to form a stable interaction. To check the compactness of the protein, radius of gyration was also analyzed. It gives the overall spread of a protein as well. The Rg of the three protein-ligand docked complexes equilibrate around at 1.5 nm and then the values decrease till the end of the simulation as shown in the fig. 5c. These results show that when ligands bind to the protein, the protein was able to fold properly and has a stable compact structure.

The hydrogen bonds formed between the ligand and the protein were also analyzed to study the interactions between them. The hydrogen bonds formed with the three ligands were found to be stable through the course of simulation except for the third ligand docked with 3MXF. The hydrogen bond for the third ligand did not form at the start of the simulation but was observed to have a stable bond from 800ps till the end of the simulation.

CONCLUSION

A top down approach for selecting the potential hits for the target BRD4 was performed by choosing the known ligands and extracting a reliable energy based pharmacophore to be used as reference for screening the compounds from ZINC databases. The results obtained were further cut down to few hundreds of compounds by clustering based on fingerprints. To put more confidence in the hits obtained, a molecular docking study was also performed for the hits and the results obtained were quite encouraging. The desired interaction of ligands ZINC68155904, ZINC67910065 and ZINC 67510456 with conserved Asparagine (N) 140, as well as with Pro 82 and Tyr 97 showed that the final hits have the potency of forming stable complex. This stability was checked with MD simulations which resulted in less fluctuating RMSD and RMSF, stable potential energy,

required number and position of hydrogen bonds for two hits. So the above study concludes that the 3 hits obtained viz. ZINC 68155904, ZINC 67910065 and ZINC 67510456 may serve as potential leads with a different backbone as compared to known ligands for BRD4.

ABBREVIATIONS

PPI: protein-protein interaction

AML: Acute myeloid leukemia

BRD4: Bromodomain containing protein 4

NUT: Nuclear protein in Testis

BET: Bromodomain and extraterminal domain family

BZD: Benzodiazepine

BZT: Triazole-benzotriazepines

XP: Extra Precision

ADME: Absorption, Distribution, Metabolism, Excretion

SP: Standard Precision

CONFLICTS OF INTERESTS

The above article contains no conflicts of interest

REFERENCES

1. French CA, Miyoshi I, Kubonishi I, Grier HE, Perez-Atayde AR, Fletcher JA. BRD4-NUT fusion oncogene a novel mechanism in aggressive carcinoma. Cancer Res 2003;63(2):304-7.
2. Filippakopoulos P, Qi J, Picaud S, Shen Y, Smith WB, Fedorov O, *et al.* Selective inhibition of BET bromodomains. Nature 2010;468(7327):1067-73.
3. Sanchez R, Zhou M-M. The role of human bromodomains in chromatin biology and gene transcription. Curr Opin Drug Discovery Dev 2009;12(5):659.
4. Dey A, Nishiyama A, Karpova T, McNally J, Ozato K. Brd4 marks select genes on mitotic chromatin and directs postmitotic transcription. Mol Biol Cell 2009;20(23):4899-909.
5. Chiang C-M. Brd4 engagement from chromatin targeting to transcriptional regulation: selective contact with acetylated histone H3 and H4. F1000 biology reports; 2009. p. 1.

6. Wu S-Y, Chiang C-M. The double bromodomain-containing chromatin adaptor Brd4 and transcriptional regulation. J Biol Chem 2007;282(18):13141-5.

7. Owen DJ, Ornaghi P, Yang JC, Lowe N, Evans PR, Ballario P, *et al*. The structural basis for the recognition of acetylated histone H4 by the bromodomain of histone acetyltransferase gcn5p. EMBO J 2000;19(22):6141-9.

8. Berger SL. Histone modifications in transcriptional regulation. Curr Opin Genet Dev 2002;12(2):142-8.

9. Yang X-J. Multisite protein modification and intramolecular signaling. Oncogene 2004;24(10):1653-62.

10. Yang Z, Yik JH, Chen R, He N, Jang MK, Ozato K, et al. Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4. Mol Cell 2005;19(4):535-45.

11. Ai N, Hu X, Ding F, Yu B, Wang H, Lu X, *et al*. Signal-induced Brd4 release from chromatin is essential for its role transition from chromatin targeting to transcriptional regulation. Nucleic Acids Res 2011;39(22):9592-604.

12. Dey A, Ellenberg J, Farina A, Coleman AE, Maruyama T, Sciortino S, *et al*. A bromodomain protein, MCAP, associates with mitotic chromosomes and affects G2-to-M transition. Mol Cell Biol 2000;20(17):6537-49.

13. Dey A, Chitsaz F, Abbasi A, Misteli T, Ozato K. The double bromodomain protein Brd4 binds to acetylated chromatin during interphase and mitosis. Proc Natl Acad Sci 2003;100(15):8758-63.

14. Bisgrove DA, Mahmoudi T, Henklein P, Verdin E. Conserved P-TEFb-interacting domain of BRD4 inhibits HIV transcription. Proc Natl Acad Sci 2007;104(34):13690-5.

15. Krueger BJ, Varzavand K, Cooper JJ, Price DH. The mechanism of release of P-TEFb and HEXIM1 from the 7SK snRNP by viral and cellular activators includes a conformational change in 7SK. PloS One 2010;5(8):e12335.

16. Filippakopoulos P, Picaud S, Fedorov O, Keller M, Wrobel M, Morgenstern O, *et al*. Benzodiazepines and benzotriazepines as protein interaction inhibitors targeting bromodomains of the BET family. Bioorg Med Chem 2012;20(6):1878-86.

17. Peng J, Zhu Y, Milton JT, Price DH. Identification of multiple cyclin subunits of human P-TEFb. Genes Dev 1998;12(5):755-62.

18. Marshall NF, Price DH. Purification of P-TEFb, a transcription factor required for the transition into productive elongation. J Biol Chem 1995;270(21):12335-8.

19. Marshall NF, Peng J, Xie Z, Price DH. Control of RNA polymerase II elongation potential by a novel carboxyl-terminal domain kinase. J Biol Chem 1996;271(43):27176-83.

20. Phelps MA, Lin TS, Johnson AJ, Hurh E, Rozewski DM, Farley KL, *et al*. Clinical response and pharmacokinetics from a phase 1 study of an active dosing schedule of flavopiridol in relapsed chronic lymphocytic leukemia. Blood 2009;113(12):2637-45.

21. Yang Z, He N, Zhou Q. Brd4 recruits P-TEFb to chromosomes at late mitosis to promote G1 gene expression and cell cycle progression. Mol Cell Biol 2008;28(3):967-76.

22. Maruyama T, Farina A, Dey A, Cheong J, Bermudez VP, Tamura T, *et al*. A mammalian bromodomain protein, Brd4, interacts with replication factor C and inhibits progression to S phase. Mol Cell Biol 2002;22(18):6509-20.

23. French CA, Miyoshi I, Aster JC, Kubonishi I, Kroll TG, Dal Cin P, et al. BRD4 bromodomain gene rearrangement in aggressive carcinoma with translocation t (15;19). Am J Pathol 2001;159(6):1987-92.

24. Zhou M, Huang K, Jung K-J, Cho W-K, Klase Z, Kashanchi F, et al. Bromodomain protein Brd4 regulates human immunodeficiency virus transcription through phosphorylation of CDK9 at threonine 29. J Virol 2009;83(2):1036-44.

25. Lin A, Wang S, Nguyen T, Shire K, Frappier L. The EBNA1 protein of Epstein-Barr virus functionally interacts with Brd4. J Virol 2008;82(24):12009-19.

26. Gagnon D, Joubert S, Sénéchal H, Fradet-Turcotte A, Torre S, Archambault J. Proteasomal degradation of the papillomavirus E2 protein is inhibited by overexpression of bromodomain-containing protein 4. J Virol 2009;83(9):4127-39.

27. Chung C-w, Coste H, White JH, Mirguet O, Wilde J, Gosmini RL, *et al*. Discovery and characterization of small molecule inhibitors of the BET family bromodomains. J Med Chem 2011;54(11):3827-38.

28. Hewings DS, Wang M, Philpott M, Fedorov O, Uttarkar S, Filippakopoulos P, *et al*. 3, 5-Dimethylisoxazoles act as acetyl-lysine-mimetic bromodomain ligands. J Med Chem 2011;54(19):6761-70.

29. Chung C-w, Dean AW, Woolven JM, Bamborough P. Fragment-based discovery of bromodomain inhibitors part 1: inhibitor binding modes and implications for lead discovery. J Med Chem 2012;55(2):576-86.

30. Bamborough P, Diallo H, Goodacre JD, Gordon L, Lewis A, Seal JT, *et al*. Fragment-based discovery of bromodomain inhibitors part 2:optimization of phenylisoxazole sulfonamides. J Med Chem 2012;55(2):587-96.

31. SS, Protein Preparation Wizard; Epik version 2.3, Schrödinger L, New York, NY, 2012;, Impact version 5.8 S, LLC,, New York N, 2012;Prime version 3.1, Schrödinger,, et al.

32. S Lig Prep v, Schrödinger L: New York N; 2012.

33. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, *et al*. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J Med Chem 2004;47(7):1739-49.

34. Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT, *et al*. Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. J Med Chem 2004;47(7):1750-9.

35. Dixon SL, Smondyrev AM, Knoll EH, Rao SN, Shaw DE, Friesner RA. PHASE: a new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. Methodology and preliminary results. J Comput-Aided Mol Des 2006;20(10-11):647-71.

36. Dixon SL, Smondyrev AM, Rao SN. PHASE: a novel approach to pharmacophore modeling and 3D database searching. Chem Biol Drug Design 2006;67(5):370-2.

37. Salam NK, Nuti R, Sherman W. Novel method for generating structure-based pharmacophores using energetic analysis. J Chem Inf Model 2009;49(10):2356-68.

38. Irwin JJ, Sterling T, Mysinger MM, Bolstad ES, Coleman RG. ZINC: a free tool to discover chemistry for biology. J Chem Inf Model 2012;52(7):1757-68.

39. Wan H. What ADME tests should be conducted for preclinical studies? ADMET DMPK 2013;1(3):19-28.

40. S, QikProp v, Schrödinger L, New York N; 2012.

41. S, Canvas v, Schrödinger L, New York N; 2012.

42. Duan J, Dixon SL, Lowrie JF, Sherman W. Analysis and comparison of 2D fingerprints: Insights into database screening performance using eight fingerprint methods. J Mol Graphics Modell 2010;29(2):157-70.

43. Pronk S, Páll S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, *et al*. GROMACS 4.5:a high-throughput and highly parallel open source molecular simulation toolkit. Bioinf 2013;29(7):845-54.

44. Schuttelkopf AW, Van Aalten DM. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. Acta Crystallogr Sect D: Biol Crystallogr 2004;60(8):1355-63.

45. Kalyaanamoorthy S, Chen Y-PP. Energy based pharmacophore mapping of HDAC inhibitors against class I HDAC enzymes. Biochim Biophys Acta Proteins Proteomics 2013;1834(1):317-28.

46. Pham DT, Dimov SS, Nguyen C. Selection of K in K-means clustering. Proc Inst Mech Eng Part C 2005;219(1):103-19.