Available online at http://pbr.mazums.ac.ir

PBR

Original Article

Pharmaceutical and Biomedical Research

Identification of BK_{Ca} channel openers by molecular field alignment and patent data-driven analysis

Yaseen Gigani[1,2*], Swati Gupta[2,3], Andrew Lynn[4], Kamlesh Asotra[1]

[1]School of Biosciences, Apeejay Stya University, Gurgaon, India
[2]School of Pharmaceutical Sciences, Apeejay Stya University, Gurgaon, India
[3]B.S. Anangpuria institute of Pharmacy, Pt B.D. Sharma University of health sciences,Village-Alampur, Faridabad, India
[4]School of Computational and Integrative Sciences, Jawaharlal Nehru University, New Delhi, India

Received: Sep 4, 2016, Revised: Dec 3, 2016, Accepted: Dec 22, 2016

Abstract

In this work, we present the first comprehensive molecular field analysis of patent structures on how the chemical structure of drugs impacts the biological binding. This task was formulated as searching for drug structures to reveal shared effects of substitutions across a common scaffold and the chemical features that may be responsible. We used the SureChEMBL patent database, which provides search of the patent literature using keyword-based functionality, as a query engine. The extraction of data of the BK_{Ca} channel openers and aligning them for molecular field similarity with newly designed structures did provide a probable validation method with accurate values. Therefore, in an attempt to increase the true positives, we report a procedure that functions on a multiple analyses modeled on molecular field similarity and common sub-structural search with consensus scoring and high confidence values to obtain greater accuracy during conventional virtual screening.

Keywords: BK_{Ca} channel, molecular field alignment, sureChEMBL, chemical curation

Pharm Biomed Res 2016; 2(4): 22-29

Introduction

In Drug discovery, a combinatorial chemical library is a pre-chosen plurality of compounds designed simultaneously to have a common structural scaffold within each structure to represent a unique configuration of substitution at specific positions (1). Therefore, this directed diversity is aimed at pattern comparisons to find related and matching chemical structures explore how chemical structures are associated to various biological processes (2). However, the current techniques behind chemical structure mining applications have mainly focused on the ability of the system to correctly identify the structure name and biological processes in text, while less effort has been spent on the correct identification and matching (3). This is about to change as more and more chemical resources are becoming available and easily accessible.

In recent years, theoretical chemistry and molecular modelling have become increasingly important in both lead finding and optimization. Computational Biologists have been attempting to generate new leads by examination of the common features of existing active compounds or of a single structure itself, of the target protein if it is known. These methodologies assist in the process of lead optimization by predicting what changes to the scaffold are likely to be beneficial since a molecule's affinity to a target is estimated by reference to its similarity to active compounds. It is possible to predict the binding properties of an untested molecule by representing the properties of a molecule which are important in its binding to other molecules, and then assessing the similarity between two such sets, one for the untested molecule and one for a well characterised molecule (4).

In traditional molecular mechanics, the electrostatic properties of a molecule are defined by placing a point charge at the centre of each atom. Many different methods for calculating or estimating the value of such point charges have been described in the literature. The aim of this method is to distribute the point charges in such a way that the resulting electrostatic field is as

*E-mail: giganiyaseen@gmail.com

similar as possible to the true electrostatic field (5). To improve the quality of molecular mechanics models at the molecular surface, extended electron distributions (XEDs) have been developed, which involves replacing the point charge at the centre of some atoms with a set of point charges, one at the centre of the atom and one or more others distributed around that atom a short distance away. In this approach, the 3D conformation of a molecule is represented by a set of field points which measure field strength at a relatively small number of field maxima and minima around the molecule which are relevant to how the molecule is likely to interact with other molecules when it binds to the target molecules (6,7).

Chemical resources such as DrugBank and PubChem have earlier been applied for the identification of drug by chemical names, similar leads and bioassay (8,9). However, these databases identify structures by specific representations such as a connection table, an International Chemical Identification (InChI) string or a simplified molecular input line entry specification (SMILES). Therefore, the challenge of matching structures by the available descriptors differs from the existing one's in the sense that the potential 'bioactive molecules' with different lead/scaffolds structures may also be similar in terms of biological activity and should also be retrieved during the database search. This suggests that retrieving structures which are represented by erroneous or incomplete descriptors should have a detrimental effect on the quality of data derived (10). To achieve accuracy over this and include biological potency during chemical structure search, we focus on chemical identification based on Molecular Field Similarity based XED descriptors, which includes 3D molecular field alignment and mapping, to link the structure components to reference to derive drug structures that been reported in the patents.

Materials and Methods

Data set

The SureChEMBL database because of its availability of chemically annotated or relevant patent documents in its database was used. The structures of the BK_{Ca} openers were extracted from the database using a number of search terms like:

- BK_{Ca} channel, Benzimidazolone, Blood Brain-Tumor Barrier (BTB).
- Searching using a chemical structure or substructure: NS-1619
- Combination of both

The most important entity from the patent documents was the chemical entities and compound names, which were also converted from chemical entities. All duplicates, stereoisomers, and tautomers were filtered out and a list of unique chemical entities from all patents was mapped to form a list of standardized compounds for further work (11).

Molecular field alignment

The field descriptors that we describe here are scalar fields obtained ingeneral from calculating the interaction energy of a 'probe' molecule with the target molecule. These probes define the essential properties of a molecule in terms of a tractable number of field points. The aim of the project is to use these points to compare structurally diverse molecules obtained from patent database to an in-house library of generated structures.

The maximum number of conformations generated for any molecule was limited to 200 in order to have a balance of the quality of alignments and calculation time. Number of high temperature dynamics for flexible rings was set at 5. Gradient cut-off for conformer minimization was 0.5. Coarseness of the sampling of conformational space was controlled by filtering duplicate conformers at RMSD 0.5. Standard scoring function was used based on 50% shape similarity and 50% field similarity to derive overall similarity between the conformations. Finally, we selected the 26 compounds from the patent database and estimated their field similarity to the in-house structures of 46 molecules obtained from an unpublished data (12,7).

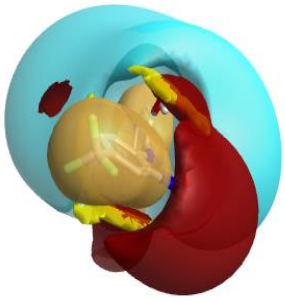
The computational work was carried out under CentOS 6.3 (Linux), The Molecular Field Alignment was done in Forge v10, Cresset Biomolecular Discovery and all chemistry related work was done in JChem module and Marvin Suite of ChemAxon.

Results

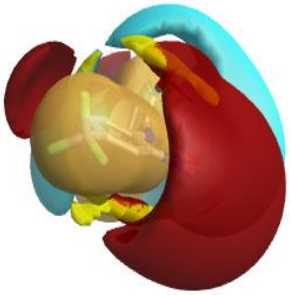
In this work, of the identified patents, the patents dealing structural analogs classified as those activating BK_{Ca} channel were used for Field Alignment with a library of 46 structures. The Patents utilized in the

FieldAlignment screening were US5200422, US5475015, and US7244756 which together contains 26 unique chemical structures as shown in Table 2 (13–15).

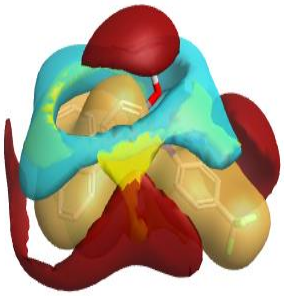
The field point pattern is a sophisticated 'Pharmacophore' which can be used to define an active reported compound for binding at the active site. Herein, each retrieved molecule was overlaid to thelibrary of 46 structures. We used the molecular fields between the two molecules to quantify and convert it to a similarity value. This protocol is based on the assumption that two molecules that bind to the same active site would be expected to have the same field pattern in their bound conformations irrespective of their chemical similarity (12,7,6). This field patterns generated through eXtended Electron Distribution (XED) force field of these known active ligand, *i.e.*, in their bioactive conformation, and was used to virtually screen a database of molecules in field format to find molecules with a similar field. It has been reported, that the molecules which have a high field similarity to a known ligand also have a high probability of being active (7). We also observed that each of the structure in our database was Field similar to every molecule that was retrieved from the database with the similarity ranges of 65% to 94% (Tables 1 and 2).



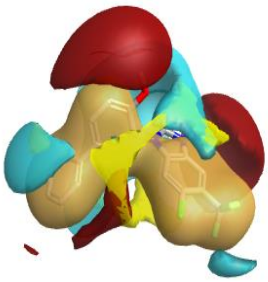
Patent Database Derived Molecule



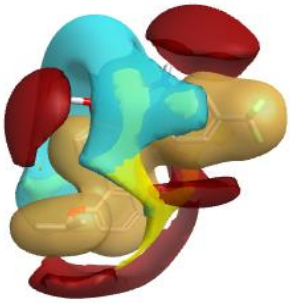
Comlib_120918; Cresset similarity 0.923



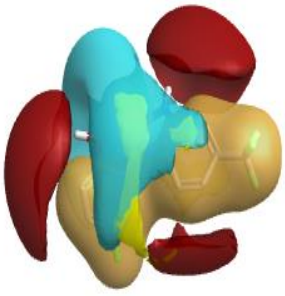
Patent Database Derived Molecule



Comlib_120918; Cresset similarity 0.902



Patent Database Derived Molecule



Comlib_120918; Cresset similarity 0.886

Figure 1 Alignment of BKCa channel activators available in SureCHEMBL patent database. At right, the field alignment of the most potent activator of the series Comlib_120918 similarity in different 3D views.

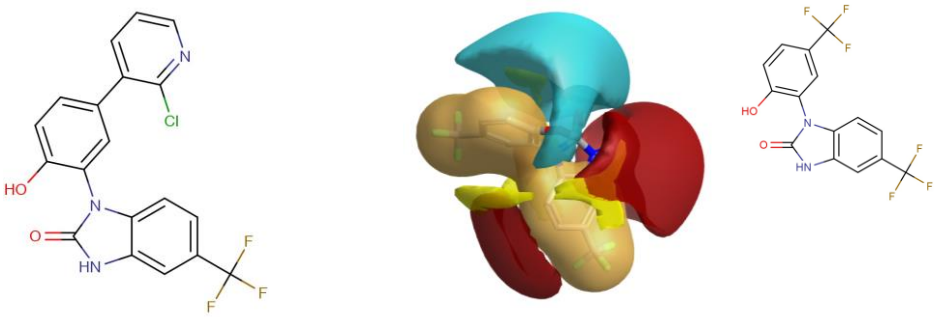


Figure 2 2D structure of Comlib_120918, at right, the field alignment and 2D structure of NS-1619, a known activator of BKCa channel.

Table 1 Field similarity scores of three hits with the available patent database molecules. The patents retrieved are US5200422, US5475015, US7244756 and 26 molecules from them.

Name of Structure	Field Similarity (%) and No. of molecules					
	>65	>70	>75	>80	>85	>90
Comlib_120946	0	2	11	10	3	0
Comlib_045253	2	10	11	1	2	0
Comlib_121660	0	6	12	6	1	1

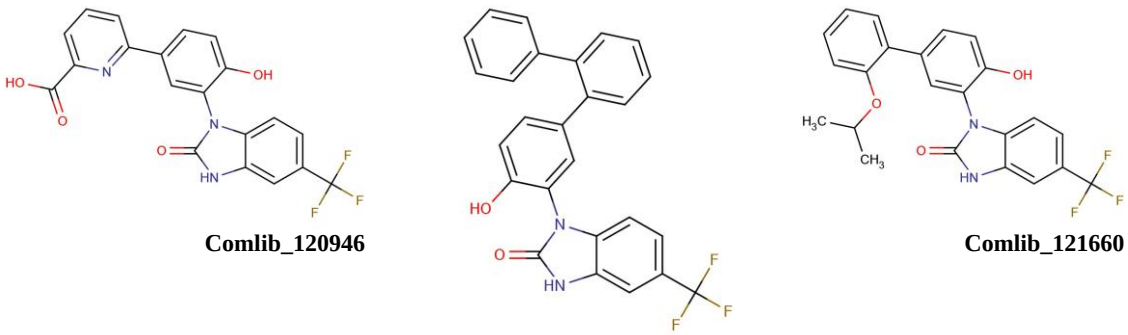


Figure 4 Field similarity scores of three hits with the available patent database molecules. The patents retrieved are US5200422, US5475015, US7244756 and 26 molecules from them**.**

Table 2 Molecular field alignment and cresset similarity scores of patent structures.

Structure Name	1	2	3	4	5	6	7	8	9	10	11	12	13
Comlib_042074	0.8	0.9	0.7	0.7	0.7	0.7	0.8	0.8	0.8	0.7	0.8	0.8	0.8
Comlib_042331	0.8	0.8	0.6	0.6	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.8	0.8
Comlib_045253	0.7	0.8	0.6	0.6	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.8
Comlib_045319	0.8	0.9	0.7	0.7	0.7	0.8	0.8	0.8	0.8	0.7	0.8	0.8	0.8
Comlib_045346	0.8	0.8	0.7	0.7	0.7	0.7	0.7	0.7	0.8	0.7	0.7	0.7	0.8
Comlib_048214	0.7	0.7	0.6	0.6	0.7	0.7	0.7	0.7	0.7	0.6	0.7	0.7	0.6
Comlib_048858	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
Comlib_051890	0.8	0.9	0.7	0.7	0.7	0.7	0.8	0.8	0.8	0.7	0.8	0.8	0.8
Comlib_051906	0.8	0.9	0.7	0.7	0.7	0.7	0.8	0.7	0.8	0.7	0.8	0.8	0.8
Comlib_051923	0.8	0.9	0.7	0.7	0.7	0.7	0.8	0.7	0.8	0.7	0.8	0.8	0.8
Comlib_052252	0.8	0.8	0.7	0.7	0.7	0.7	0.7	0.7	0.8	0.6	0.7	0.8	0.7
Comlib_052677	0.8	0.9	0.7	0.7	0.7	0.7	0.8	0.7	0.8	0.7	0.8	0.8	0.8
Comlib_053302	0.8	0.7	0.7	0.6	0.7	0.7	0.7	0.7	0.8	0.6	0.7	0.8	0.7
Comlib_113498	0.8	0.8	0.7	0.7	0.7	0.7	0.7	0.7	0.8	0.7	0.7	0.8	0.8
Comlib_113554	0.8	0.8	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.8	0.8
Comlib_113840	0.8	0.8	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.8	0.7
Comlib_114033	0.8	0.8	0.6	0.6	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
Comlib_114156	0.8	0.8	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
Comlib_114179	0.8	0.8	0.6	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.8
Comlib_114217	0.8	0.8	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.8	0.8
Comlib_114662	0.8	0.8	0.7	0.6	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.8
Comlib_114678	0.8	0.8	0.6	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.8	0.8
Comlib_114709	0.8	0.9	0.6	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.8	0.8
Comlib_115648	0.8	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.8	0.7	0.7	0.7	0.7
Comlib_116143	0.7	0.7	0.6	0.6	0.6	0.6	0.7	0.7	0.7	0.6	0.7	0.7	0.6
Comlib_116937	0.8	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.6	0.7	0.7	0.7
Comlib_117053	0.8	0.7	0.7	0.6	0.7	0.7	0.7	0.7	0.8	0.6	0.7	0.7	0.6
Comlib_117275	0.7	0.7	0.7	0.6	0.7	0.7	0.7	0.7	0.7	0.6	0.7	0.7	0.6
Comlib_117747	0.8	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.8	0.8	0.7
Comlib_118081	0.8	0.8	0.7	0.7	0.7	0.7	0.7	0.7	0.8	0.7	0.7	0.8	0.7
Comlib_120478	0.8	0.8	0.6	0.6	0.7	0.7	0.7	0.7	0.8	0.7	0.7	0.8	0.7
Comlib_120730	0.9	0.9	0.7	0.7	0.8	0.8	0.8	0.8	0.8	0.7	0.8	0.8	0.8
Comlib_120773	0.8	0.9	0.7	0.7	0.7	0.8	0.8	0.8	0.8	0.7	0.8	0.8	0.8
Comlib_120786	0.8	0.9	0.7	0.7	0.7	0.8	0.7	0.8	0.8	0.7	0.8	0.8	0.8
Comlib_120792	0.8	0.9	0.7	0.7	0.7	0.7	0.8	0.7	0.8	0.7	0.8	0.8	0.8
Comlib_120804	0.8	0.8	0.6	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.8
Comlib_120918	0.8	0.9	0.7	0.7	0.7	0.8	0.8	0.8	0.8	0.7	0.8	0.8	0.8
Comlib_120946	0.8	0.8	0.7	0.7	0.7	0.8	0.7	0.8	0.8	0.7	0.8	0.8	0.8
Comlib_121016	0.8	0.8	0.7	0.7	0.7	0.7	0.7	0.7	0.8	0.7	0.8	0.8	0.8
Comlib_121231	0.8	0.8	0.7	0.7	0.7	0.8	0.7	0.7	0.8	0.7	0.8	0.8	0.8
Comlib_121339	0.8	0.8	0.7	0.7	0.7	0.7	0.7	0.7	0.8	0.7	0.7	0.8	0.7
Comlib_121480	0.8	0.9	0.7	0.7	0.7	0.8	0.8	0.8	0.8	0.7	0.8	0.8	0.8
Comlib_121609	0.8	0.9	0.7	0.7	0.7	0.7	0.8	0.8	0.8	0.7	0.8	0.8	0.8
Comlib_121660	0.8	0.9	0.7	0.7	0.7	0.7	0.7	0.7	0.8	0.7	0.8	0.8	0.8
Comlib_121908	0.8	0.8	0.7	0.7	0.7	0.7	0.7	0.7	0.8	0.7	0.7	0.8	0.8
Comlib_121910	0.8	0.9	0.7	0.7	0.7	0.8	0.7	0.8	0.8	0.7	0.8	0.8	0.8

continue…

Structure	14	15	16	17	18	19	20	21	22	23	24	25	26
Comlib_042074	0.7	0.7	0.7	0.7	0.8	0.8	0.8	0.7	0.7	0.8	0.8	0.8	0.8
Comlib_042331	0.6	0.7	0.7	0.7	0.8	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
Comlib_045253	0.7	0.7	0.7	0.7	0.8	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
Comlib_045319	0.7	0.7	0.8	0.7	0.8	0.8	0.8	0.8	0.7	0.8	0.7	0.8	0.8
Comlib_045346	0.7	0.7	0.7	0.7	0.8	0.7	0.7	0.7	0.7	0.7	0.7	0.8	0.7
Comlib_048214	0.6	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
Comlib_048858	0.7	0.7	0.7	0.6	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
Comlib_051890	0.7	0.7	0.7	0.7	0.8	0.8	0.8	0.7	0.7	0.8	0.8	0.8	0.8
Comlib_051906	0.7	0.7	0.8	0.7	0.8	0.8	0.8	0.7	0.7	0.8	0.8	0.8	0.8
Comlib_051923	0.7	0.7	0.8	0.7	0.8	0.8	0.7	0.8	0.7	0.8	0.8	0.8	0.8
Comlib_052252	0.6	0.7	0.7	0.7	0.8	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
Comlib_052677	0.7	0.7	0.8	0.7	0.8	0.7	0.7	0.8	0.7	0.7	0.8	0.8	0.8
Comlib_053302	0.6	0.7	0.7	0.7	0.8	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
Comlib_113498	0.7	0.7	0.8	0.8	0.8	0.8	0.7	0.7	0.7	0.9	0.7	0.8	0.8
Comlib_113554	0.7	0.7	0.8	0.8	0.8	0.8	0.7	0.7	0.7	0.8	0.7	0.8	0.7
Comlib_113840	0.7	0.7	0.8	0.8	0.8	0.8	0.7	0.7	0.7	0.8	0.7	0.7	0.7
Comlib_114033	0.7	0.7	0.8	0.8	0.7	0.8	0.7	0.7	0.7	0.8	0.7	0.7	0.7
Comlib_114156	0.7	0.7	0.8	0.8	0.8	0.8	0.7	0.7	0.7	0.8	0.7	0.7	0.7
Comlib_114179	0.7	0.7	0.8	0.8	0.8	0.8	0.7	0.7	0.7	0.8	0.7	0.7	0.7
Comlib_114217	0.7	0.7	0.8	0.8	0.8	0.8	0.7	0.7	0.7	0.8	0.7	0.7	0.7
Comlib_114662	0.7	0.7	0.8	0.7	0.8	0.8	0.7	0.7	0.7	0.8	0.7	0.7	0.7
Comlib_114678	0.7	0.7	0.8	0.7	0.8	0.8	0.7	0.7	0.7	0.8	0.7	0.7	0.7
Comlib_114709	0.7	0.7	0.8	0.8	0.8	0.8	0.7	0.7	0.7	0.8	0.7	0.7	0.7
Comlib_115648	0.6	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.8	0.7
Comlib_116143	0.6	0.6	0.6	0.6	0.7	0.7	0.7	0.6	0.6	0.7	0.7	0.7	0.7
Comlib_116937	0.6	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
Comlib_117053	0.6	0.7	0.7	0.6	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
Comlib_117275	0.6	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.8	0.7	0.7
Comlib_117747	0.6	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.8	0.7
Comlib_118081	0.6	0.7	0.7	0.7	0.8	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
Comlib_120478	0.6	0.7	0.7	0.7	0.8	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
Comlib_120730	0.7	0.8	0.8	0.8	0.9	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
Comlib_120773	0.7	0.7	0.8	0.7	0.8	0.8	0.8	0.8	0.7	0.8	0.7	0.8	0.8
Comlib_120786	0.7	0.7	0.7	0.7	0.9	0.8	0.8	0.8	0.7	0.8	0.8	0.8	0.8
Comlib_120792	0.7	0.7	0.8	0.7	0.8	0.7	0.7	0.8	0.7	0.8	0.7	0.7	0.8
Comlib_120804	0.7	0.7	0.7	0.7	0.8	0.7	0.7	0.7	0.7	0.7	0.7	0.8	0.7
Comlib_120918	0.7	0.8	0.8	0.7	0.9	0.8	0.8	0.8	0.7	0.8	0.8	0.8	0.8
Comlib_120946	0.7	0.7	0.7	0.7	0.8	0.8	0.7	0.8	0.7	0.7	0.8	0.8	0.8
Comlib_121016	0.7	0.7	0.7	0.7	0.8	0.8	0.7	0.7	0.7	0.8	0.7	0.8	0.8
Comlib_121231	0.7	0.7	0.7	0.7	0.8	0.8	0.7	0.7	0.7	0.7	0.7	0.8	0.8
Comlib_121339	0.7	0.7	0.7	0.7	0.8	0.7	0.7	0.8	0.7	0.7	0.7	0.8	0.7
Comlib_121480	0.7	0.8	0.8	0.7	0.8	0.8	0.8	0.8	0.7	0.8	0.8	0.8	0.8
Comlib_121609	0.7	0.7	0.7	0.7	0.9	0.8	0.7	0.8	0.7	0.8	0.8	0.8	0.8
Comlib_121660	0.7	0.7	0.7	0.7	0.8	0.7	0.7	0.8	0.7	0.7	0.7	0.7	0.7
Comlib_121908	0.7	0.7	0.7	0.7	0.8	0.7	0.7	0.8	0.7	0.8	0.7	0.8	0.8
Comlib_121910	0.7	0.7	0.8	0.7	0.8	0.8	0.7	0.8	0.7	0.8	0.7	0.8	0.8

***Darkness signifies level of similarity**

Discussion

Patents are a valuable and unique source of information in scientific research. It is estimated that only a small fraction of the new science and technology first reported in patents is subsequently disclosed in scientific literature sources. Moreover, the first public disclosure of new chemical entities, scaffolds and series typically takes place in patent applications prior to their publication in scientific journals (11,16). In addition to novel composition of matter, chemical patent documents contain information on reactions and synthetic pathways, biomarkers, assays, experimental conditions, active ingredients and catalysts, as well biological target, mechanisms of action, bioactivity data and disease indications that are not present in other sources, such as commercial or publicly available bioactivity databases or peer-reviewed scientific literature (17,18). Therefore, extracting information from filed patents, these days forms an important aspect of literature search prior to a research protocol. Retrieving accurate chemical information for probable

structures with different lead/scaffolds with similar bioactivity has been a different task. Herein, we report a method that combines a Molecular Field Similarity along with Maximum Substructure (MCS) for the identification of Large Conductance Calcium Activated Potassium (BK_{Ca}) Channel Openers by chemical curation with molecules, which are proprietary and available only through patent database SureChembl (12,6).

The basic assumption underlying the field point approach is that two molecules which have similar sets of field points should have similar interactions with other molecules and hence should have similar biological activities. With the field point approach, the similarity between conformations of two molecules is calculated according to differences between the field point positions and energy values of the field points in the two field point sets. The result is a scalar quantity referred to as the field similarity value.Instead of relying on the 2D structure alone, we used the fields around molecules to assess their likely activity and properties, regardless of structural similarity. The most important regions on the molecular field, where interactions with another molecule are strongest can then be identified and that portion of the field surface substituted with a Field Point (5). As we present in Figures 1 and 2, Field Points provide a highly condensed and accurate representation of the nature, size and location of a critical property required for binding and instigating a specific therapeutic effect when compared with a known structure. This pattern contains no information about the bonds and angles that generated it, and in fact many different structures could potentially generate a similar pattern. Crucially, any molecule that can present that same configuration of Field Points is likely to have the same activity. Therefore, we believe that Fields can be used to discover novel bioisosteres with diverse chemotypes to provide a logical explanationto their common biological properties.

Finally, we believe that the initial extraction of bioactivity data of the BK_{Ca} channel from the scientific patent and aligning them for molecular field similarity did provide a probable validation method (19). This research work was executed to identify the validity of the simulation studies undertaken earlier. Most encouraging is the finding that even in cases of non-availability of negative controls, these methods perform very strongly compared to only Virtual Screening method (20). It also suggests that knowledge about a single active ligand for a drug target can be as valuable as a crystal structure for obtaining novel scaffolds required for the development of Biosimiliars.

Conclusion

Even with stringent molecular modeling techniques about 40% of the false positives are still obtained. In an attempt to increase the true positives, we developed a Molecular field based descriptor that functions on multiple analyses modeled on molecular field similarity and structural similarity with high confidence values. We considered a number of factors to provide confidence for the data generated and a solid basis to predict the highest possibility in terms of field based empirical optimization. We conclude, optimistically, that the high accuracy achieved at this modest computational cost implies that molecular Field based descriptors may be poised toward generating high accuracy based scoring functions that may replace the rather uncontrolled approximations that are common in empirical searches.

Finally, though we present a general approach to iterative computation of maximum-likelihood estimates when the observations can be viewed, further experiments in terms of laboratory activity are still requiredto provide solid evidence of the effectiveness of the method developed.

Acknowledgement

YG is thankful to ASU for Teaching Assistantship during his PhD, the authors also acknowledge SCIS, JNU for providing computational facilities for carrying out this work.

Conflict of Interest

The authors declared no conflict of interest.

References

1. Pandeya SN, Thakkar D. Combinatorial Chemistry: A novel method in drug discovery and its application. Indian J Chem 2005;44:335–48.
2. Seoane JA, Aguiar-Pulido V, Munteanu CR, Rivero D, Rabunal JR, Dorado J, et al. Biomedical data integration in computational drug design and bioinformatics. Curr Comput Aided Drug Des 2013;9:108–17.

3. Hettne KM, Williams AJ, van Mulligen EM, Kleinjans J, Tkachenko V, Kors JA. Automatic vs. manual curation of a multi-source chemical dictionary: the impact on text mining. J Cheminform 2010;2:1-7
4. Cleves AE, Jain AN. Extrapolative prediction using physically-based QSAR. J Comput Aided Mol Des 2016;30 :127–52.
5. Cheeseright T, Mackey MD, Vinter JG. Comparison of molecules using field points. United States patent US 7,805,257. 2010 Sep 28.
6. Cheeseright TJ, Mackey MD, Melville JL, Vinter JG. FieldScreen: virtual screening using molecular fields. Application to the DUD data set. J Chem Inf Model 2008;48:2108–17.
7. Cheeseright T, Mackey M, Rose S, Vinter A. Molecular field extrema as descriptors of biological activity: definition and validation. J Chem Inf Model 2006;46:665–76.
8. Cheng T, Pan Y, Hao M, Wang Y, Bryant SH. PubChem applications in drug discovery: a bibliometric analysis. Drug Discov Today 2014;19:1751–6.
9. Liao C, Sitzmann M, Pugliese A, Nicklaus MC. Software and resources for computational medicinal chemistry. Future Med Chem 2011;3:1057–85.
10. Taylor KR, Gledhill RJ, Essex JW, Frey JG, Harris SW, De Roure DC. Bringing chemical data onto the semantic web. J Chem Inf Model 2006;46:939–52.
11. Papadatos G, Davies M, Dedman N, Chambers J, Gaulton A, Siddle J, et al. SureChEMBL: a large-scale, chemically annotated patent document database. Nucleic Acids Research. 2016;44:220-8.
12. Cheeseright T, Mackey M, Rose S, Vinter A. Molecular field technology applied to virtual screening and finding the bioactive conformation. Expert Opin Drug Discov 2007; 2:131–44.
13. Olesen SP, Watjen F. Benzimidazole derivatives, their preparation and use. United States patent US 5,200,422. 1993 Apr 6.
14. Olesen SP, Jensen LH, Moldt P. Benzimidazole derivatives, their preparation and pharmaceutical use. United States patent US 5,475,015. 1995 Dec 12.
15. Madsen LS, Bæk C, Lauridsen A, Olesen SP. Benzimidazol-2-one derivatives and their use United States patent US 7,244,756. 2007 Jul 17.
16. Bali A. Molecular-Field-Based Three-Dimensional Similarity Studies on Quinoline-Based CNS Active Agents. ISRN Pharm 2011; ID 186943.
17. Bousfield D, McEntyre J, Velankar S, Papadatos G, Bateman A, Cochrane G, et al. Patterns of database citation in articles and patents indicate long-term scientific and industry value of biological data resources. F1000Research. Faculty of 1000 Ltd; 2016;5.
18. Papadatos G, Gaulton A, Hersey A, Overington JP. Activity, assay and target data curation and quality in the ChEMBL database. J Comput Aided Mol Des. 2015;29:885–96.
19. Senger S, Bartek L, Papadatos G, Gaulton A. Managing expectations: assessment of chemistry databases generated by automated extraction of chemical structures from patents. J Cheminform 2015;7:1–12.
20. Oprea TI, Matter H. Integrating virtual screening in lead discovery. Curr Opin Chem Biol 2004;8:349–58.