Computational Biology and Chemistry 67 (2017) 48–61

Contents lists available at ScienceDirect

Computational Biology and Chemistry

journal homepage: www.elsevier.com/locate/compbiolchem

Research Article

Role of computational efficiency indices and pose clustering in effective decision making: An example of annulated furanones in *Pf*-DHFR space

Manoj Kumar, Tanpreet Kaur, Anuj Sharma*

Department of Chemistry, Indian Institute of Technology Roorkee, Roorkee, Uttarakhand, 247667, India

ARTICLE INFO

Article history:
Received 20 April 2016
Received in revised form 4 November 2016
Accepted 22 December 2016
Available online 23 December 2016

Keywords:
AutoDock
Docking
Efficiency indices
BEI
SEI
Pose clustering

ABSTRACT

In the present report, the role of computationally estimated efficiency indices and pose clustering has been demonstrated in effective decision making, resource management and chemical prioritization. As an example, 720 annulated furanones from six different scaffold classes were computationally docked against *Pf*-DHFR active site using AutoDock 4.2. Many trends were established by navigating efficiency indices (BEI and SEI) in 2D planes. These trends were then explained by comparing interaction profiles of docked poses with that of known actives/inhibitors. Cases where trends emerged from efficiency plots resonated well with the pattern of a particular cluster diagram were considered as guidelines for optimization purpose. These kind of guidelines can help medicinal chemists in prioritization their work and in effective management of time, energy and chemical resources.

1. Introduction

High throughput screening (HTS) exercises are now become an integral part of any drug discovery project. This screening exercise heavily depends on the rapid and efficient supply of novel scaffolds and chemotypes (Congreve et al., 2005; Crews, 2010; Ghosh et al., 2006; Mayr and Bojanic, 2009). Owing to suitability for automation and miniaturization, certain reaction classes such as multicomponent (MCRs)/click along with others, find important place in combinatorial chemistry, library design and HTS and are quite relevant in drug discovery projects (Teague et al., 1999; Bienaymé et al., 2000).

The fast delivery of novel chemical entities (NCEs) alone does not guarantee success in any drug discovery project. One factor that should also be taken into account is that optimization from hit to lead and then to drug is a very time consuming and expensive process with high attrition rate (Bleicher et al., 2003). As a result, only very limited academic set ups (especially funded by industry) can manage this huge financial burden. Considering this uncertainty, prioritization of available chemical resources seems very crucial for any medicinal chemist.

Till date, several predictive models have been developed by different research groups to speed up the process of library selection and drug optimization (Walters et al., 1999; Caldwell, 2000; Plewczynski et al., 2006; Wang and Ramnarayan, 1999; Pogorelcnik et al., 2015; Greenbaum et al., 2002; Sadowski, 2000; Charifson and Walters, 2000; Auer and Bajorath, 2006; Kumar and Zhang, 2016; Deng et al., 2006). Most of these methods rely on a piece of information gained through computational docking, ligand based pharmacophore or primary chemical attributes of a compound. Some of the previously described models are mentioned below:

(a) Multilevel chemical compatibility (MLCC): This method is based on the systematic comparison of local environment within a compound and those within an existing drug (Wang and Ramnarayan, 1999).
(b) Receiver operating curve (ROC) in combination with molecular docking or ligand based screening: Primary screening is performed using structure base or ligand base methods, followed by the construction of retrospective pharmacophore evaluation by ROC approach (Pogorelcnik et al., 2015).
(c) Molecular affinity fingerprints (MAFs): Cluster algorithm is used to classify closely related targets and this information can be used for inhibitor design and library selection (Greenbaum et al., 2002).

* Corresponding author.
E-mail addresses: anujsfcy@iitr.ernet.in, anujsharma.mcl@gmail.com (A. Sharma).

http://dx.doi.org/10.1016/j.compbiolchem.2016.12.007
1476-9271/

(d) Neural network method (NN): this method uses "atom type discriptors" for defining chemical structures along with a trained feed neural network (Sadowski, 2000).
(e) Filtering based methods: This method removes compounds from a database which are less drug/lead like by applying certain constrains based on molecular weight (MW), polarity etc (Charifson and Walters, 2000).
(f) Emerging chemical patterns (ECP): In this method, key molecular features from a highly active pool of compounds are extracted and this information is then applied to classify the compounds at different potency level (Auer and Bajorath, 2006).
(g) Shape similarity based methods: This method uses information of molecular surface in 3D space and then applied it to classify/screen compound libraries (Kumar and Zhang, 2016).
(h) Structure interaction fingerprint method (SIFt): This method is based on the information of binding interactions between protein and known inhibitors. These interaction profiles are then translate into a filtering constraints to classify compounds (Deng et al., 2006).

Most of the above mentioned models require complicate statistical treatment, expensive computational resources, knowledge of programming or large number of empirical data. Considering all these problems, we wish to propose herein, a very simple predictive model for routine decision making and for compound classification. This method uses computationally derived efficiency indices in conjugation with comparative interaction profile of a compound with that of a reference. Although the usefulness of efficiency indices have been recently challenged (Kenny et al., 2014), but still a fairly large amount of literature support their utility specially at the initial stage of drug discovery (Cortes-Ciriano, 2016; Ponte-Sucre et al., 2015; Shultz, 2013; Abad-Zapatero and Metz, 2005; Schultes et al., 2010; Abad-Zapatero, 2007; Abad-Zapatero and Blasi, 2011; García-Sosa et al., 2011, 2008, 2010). Some of the facts in support of efficiency indices are as under:

(i) Recently, It was shown that QSAR models based on efficiency indices have higher predictive power than models based on direct use of potency (or such as IC_{50}, MIC or K_i). This result is based on extensive investigation of 11 ligand efficiency indices (including BEI and SEI) across four algorithms (gradient boosting machine, partial least square, random forest and support vector machine) (Cortes-Ciriano, 2016).
(ii) Although drug optimization is a multivariable process and many parameters such as selectivity, toxicity, metabolic stability, cellular activity, permeability etc are important for successful optimization. But surprisingly, only two molecular properties molecular weight (MW) and total polar surface area (TPSA) directly or indirectly govern all these variables and thus dominate in drug discovery process. Using composite parameters such as BEI and SEI; we can correlate potency, MW and TPSA (Abad-Zapatero, 2007; Abad-Zapatero and Blasi, 2011).
(iii) Use of composite parameters (such as BEI and SEI) greatly reduced the total number of variables and hence these indices are easy to represent and understand (Abad-Zapatero, 2007; Abad-Zapatero and Blasi, 2011).
(iv) Both the efficiency indices (BEI and SEI) have been successfully employed on many drug discovery ventures and are very useful in drug discovery especially at the initial level (Ponte-Sucre et al., 2015; García-Sosa et al., 2011, 2008, 2010).

By comparing the location of established drugs or known hits with the results of a particular high throughput virtual screening exercise (HTVS) in combined BEI-SEI plane (Abad-Zapatero, 2007) and in cluster dendrograms (Bouvier et al., 2010; Mantsyzov et al., 2012), several important trends can be noticed. Several important questions such as (1) whether our hits are lead/drug like or not, (2) in comparison of an established drugs what are the efficiencies and interaction profiles of our hits, (3) which kind of changes (such as introduction of polar group, change in the position of substitution etc) should be done to navigate to a particular direction in optimization plane, (4) which scaffold classes should be priorities *in wet* synthesis than others, can be answered (Abad-Zapatero, 2007; Abad-Zapatero and Blasi, 2011). In a way this type of exploration resembles SPR (structure property relationship) studies. Inclusion of several other indices and parameters related to physiology and toxicity (ADMET) can provide multidimensional framework to this study.

The above mention strategy can be easily applied to explore the chemico-biological space of several efficient reactions such as MCRs in structure base drug discovery (SBDD) efforts. Most of these reactions if not all, are quite efficient, atom and step economical, diversity oriented and suitable for automation (Teague et al., 1999; Ruijter et al., 2011). As a result, very large pool of diverse and biologically important scaffolds can be generated in a very shorter span of time. *Wet lab* synthesis and actual screening of all these compounds by conventional methods can be an expensive exercise, particularly when there is no previous history of these scaffolds against a known or novel target/(s). Computer based screening and insightful use of resultant data by methods, such as efficiency indices and pose clustering can be very useful in such instances and can save a lot of precious chemical, time and energy.

As an illustration we have navigated chemico-biological space of several annulated furanones against *Pf*-DHFR using the above mentioned idea of combined efficiency indices and pose clustering. There were primarily two reasons for the selection of annulated furanones:

(i) Author's personal experience of their synthesis (Kumar et al., 2015a,b,c). These functionalized furanone derivatives are easily accessible through multicomponent condensation of a C-H acid, aldehyde and isonitrile.
(ii) Secondly, annulated furanones form structure backbone of myriads of natural products and display a wide spectrum of medicinal activities such as antibacterial, anti-cancer, ichtyotoxic etc (Jacobi and Walker, 1981; Igoli et al., 2012; Ojida et al., 1994; Perdih et al., 2015, 2009). Acetymonrifoline, an alkaloid extracted from bark/root of *Teceanobilis* plant display a significant antimalarial activity with IC_{50} = 56 μg/ml (Yenesew and Dagne, 1998). In the present study, we have shown the use of SBDD (folate space of the *Plasmodium falciparum*) driven efficiency indices and pose clustering approaches in chemical prioritization of chemical resources.

2. Tools and techniques

2.1. Molecular docking

2.1.1. Protein structure

Crystal structure of wild type ***Pf*-DHFR-TS** (resolution 2.33 Å) in complex with the third generation inhibitor **WR99210** was retrieved from Protein Data Bank (PDB ID – 1J3I) (Yuvaniyama et al., 2003). The inhibitor, cofactors ***d*UMP** and all the associated water molecules were removed. Protein structure was dissected at **Asn231**, which correspond to DHFR portion of ***Pf*-DHFR-TS**. DHFR domain of this bi-functional enzyme consists of two isomeric side chains (**A** and **B**) with missing residues. Chain **A** has missing residues from 86 to 95 (total 10), while chain **B** has five missing