Methods 71 (2015) 158–166

Contents lists available at ScienceDirect

Methods

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

Virtual screening in small molecule discovery for epigenetic targets

Guo-Bo Li [a,1], Ling-Ling Yang [b,1], Yiming Yuan [c], Jun Zou [a,*], Yu Cao [d], Sheng-Yong Yang [a], Rong Xiang [e], Mingli Xiang [a,*]

[a] State Key Laboratory of Biotherapy/Collaborative Innovation Center of Biotherapy, West China Hospital, West China Medical School, Sichuan University, Chengdu 610041, China
[b] School of Bioengineering, Xihua University, Chengdu 610039, China
[c] Division of Geriatric Medicine, State Key Laboratory of Biotherapy/Collaborative Innovation Center of Biotherapy, West China Hospital, West China Medical School, Sichuan University, Chengdu 610041, China
[d] Division of Emergency, State Key Laboratory of Biotherapy/Collaborative Innovation Center of Biotherapy, West China Hospital, West China Medical School, Sichuan University, Chengdu 610041, China
[e] School of Medicine/Collaborative Innovation Center of Biotherapy, Nankai University, Tianjin 300071, China

ARTICLE INFO

Article history:
Received 9 July 2014
Received in revised form 27 October 2014
Accepted 14 November 2014
Available online 21 November 2014

Keywords:
Epigenetic targets
Virtual screening
Rescoring methods
Docking pose filtering
Machine learning
Molecular similarity

ABSTRACT

Epigenetic modifications are critical mechanisms that regulate many biological processes and establish normal cellular phenotypes. Aberrant epigenetic modifications are frequently linked to the development and maintenance of several diseases including cancer, inflammation and metabolic diseases and so on. The key proteins that mediate epigenetic modifications have been thus recognized as potential therapeutic targets for these diseases. Consequently, discovery of small molecule inhibitors for epigenetic targets has received considerable attention in recent years. Here, virtual screening methods and their applications in the discovery of epigenetic target inhibitors are the focus of this review. Newly emerging approaches or strategies including rescoring methods, docking pose filtering methods, machine learning methods and 3D molecular similarity methods were also underlined. They are expected to be employed for identifying novel inhibitors targeting epigenetic regulation more efficiently.

1. Introduction

The term "epigenetics" was first coined by Conrad H. Waddington in 1942 to describe the heritable changes that do not result from alterations in the DNA sequence itself [1,2]. Currently, epigenetics is most commonly referred to chromatin-based events that regulate DNA-templated processes [3]. Chromatin is the complex of DNA and proteins that package DNA. Histones, for instance, are the DNA packaging proteins that are abundantly available within chromatin structure [4]. It has been known that nucleosome is the basic functional unit of chromatin, containing 147 base pairs of DNA wrapped around a histone octamer. The octameric core consists of two copies of each of four canonical histone isotypes (H2A, H2B, H3 and H4) [3,5]. At molecular level, covalent modifications of DNA and histone proteins by chromatin-modifying enzymes are usually carried out in a reversible and highly regulated manner [3,5]. These covalent modifications can alter the chromatin structure by affecting the non-covalent interactions between and within nucleosomes. They can also be recognized by specialized chromatin readers, which serve to recruit additional chromatin-modifying enzymes and remodeling enzymes [3,5]. These information provided by the epigenetic modifications play a critical role in the regulation of DNA-templated processes, such as gene transcription, replication and DNA repair.

There exist a large number of enzymes that mediate the epigenetic modifications in humankind, which could roughly be divided into three categories: writers, erasers and readers [5]. Several representatives of epigenetic regulatory enzymes are listed in Table 1. These enzymes have been found to be associated with multiple human diseases. BRD4, a bromodomain containing protein for example, has been considered as a central determinant in various cancers as well as in inflammatory disease [5,6]. It is now a well-recognized therapeutic target receiving much attention in recent years. Many other epigenetic regulatory enzymes, such as HDAC8, EZH2, DOT1L, MLL2, ATAD2 and DNMT3, have also been identified as the main driving force for cancer cell proliferation and growth [3,5,7–12]. Recent studies demonstrated that those enzymes including SIRT1, HDACs and JMJD3 play a critical role in the induction and maintenance of inflammation [13–15]. Considerable evidence show that abnormal activities of SIRT1 and SETD7 as well as their signaling networks exist in metabolic disorders such as type

* Corresponding authors. Fax: +86 28 85164060.
E-mail addresses: zoujunwch@scu.edu.cn (J. Zou), xiang_mingli@scu.edu.cn (M. Xiang).
[1] These people contributed equally to this work.

http://dx.doi.org/10.1016/j.ymeth.2014.11.010
1046-2023/

2 diabetes [16–18]. Additionally, epigenetic regulatory enzymes are often dysregulated in other diseases such as cardiovascular diseases, schizophrenia, and Alzheimer's disease [5,9,19–21]. Consequently, epigenetic regulatory enzymes have been thought as important targets for the treatment of related diseases. Discovery of small molecule inhibitors against these epigenetic targets has become the main focus of today's research in epigenetics.

To date, numerous inhibitors against epigenetic targets have been discovered. Some of the well-known epigenetic target inhibitors are shown in Fig. 1. Several inhibitors of histone deacetylation or DNA methylation, including Vorinostat, Romidepsin, Azacitidine and Decitabine (Fig. 1), have been approved for treating haematological malignancies by US Food and Drug Administration, which provides proof of concept for epigenetic therapies [3,5,9]. Many BRD inhibitors, such as GSK525762 (ClinicalTrials.gov Identifier: NCT01587703), OTX015 (ClinicalTrials.gov Identifier: NC T017 13582), and RVX-208 (ClinicalTrials.gov Identifier: NC T01058018, NCT00768274, NCT01863225), are currently being evaluated in clinical trials. More recently, Ciceri et al. reported that several clinical kinase inhibitors, PLK1 inhibitor BI-2536 and JAK2-FLT3 inhibitor TG-101348 (Fig. 1) for instance, could be able to inhibit BRD4 at low nanomolar concentration [22]. There are also many other epigenetic target inhibitors being evaluated in clinical trials, such as Ruxolitinib, Baricitinib, Panzem, EPZ-5676, Olaparib, BMN-673, and E7438 (Fig. 1).

The epigenetic target inhibitors were discovered mainly through high-throughput screening, random screening and biophysical screening approaches. However, these wet experimental methods usually suffered from a high cost and a comparatively low hit rate [23]. Virtual screening as a rapid and economic strategy [23–26] has been widely used in medicinal chemistry area for hit/lead discovery. Here we summarized the commonly used virtual screening methods and their applications in the discovery of epigenetic target inhibitors. A number of newly emerging *in silico* approaches or strategies, which could be expected to identify epigenetic target inhibitors more efficiently, were also the focus of this review.

2. The conventional virtual screening protocols

Virtual screening refers to a diverse combination of computational methods to identify new hit/lead compounds for a biological target from large chemical libraries on basis of known structure information or active ligands. Fig. 2 schematically depicts conventional virtual screening protocols. Generally, chemical libraries are firstly filtered by using Lipinski's Rule of Five and/or physiochemical properties and sometimes ADMET properties. They are then evaluated by employing computational methods. Finally potential hits/leads are selected out from the evaluated molecules for further biological assays by manually checking. Chemical library and computational methods are the two essential components in virtual screening protocol.

Chemical library is the major source of hit/lead compounds in many drug discovery projects. Currently, various kinds of chemical libraries are publicly available, including synthetic compound library, natural product library, combinatorial chemistry library and virtual ligand library. Commercial chemical library, in which compounds could be readily purchased, is the most widely used among all those kinds. Libraries ChemDiv (http://www.chemdiv.com/), Specs (http://www.specs.net), Enamine (http://www.enamine.net/), IBScreen (http://www.ibscreen.com) and Maybridge (http://www.maybridge.com/) are the examples of commercial chemical library. ZINC database (https://zinc.docking.org/), a free database of commercial available compounds, provides a huge number of commercially-available compounds (over 35 million compounds) for virtual screening. Recently, Reymond's group developed several large virtual ligand libraries, GDB-13 [27] and GDB-17 [28] for instance, exceedingly extended the drug-like chemical space. These virtual ligand libraries have been used to identify novel ligands against nicotinic receptor via *in silico* screening [29,30]. In fact, the quality and diversity of chemical libraries has a significant influence on the efficiency of virtual screening. The approaches such as filtering out non-drug-like compounds by using Lipinski's Rule of Five and/or physiochemical properties [31], and removing redundant compounds by similarity comparison [32,33] are often employed to promote the quality of libraries. An alternative way is to establish the focused compound library, which could provide a large number of high-quality lead compounds [34–38]. Several such libraries, G-protein coupled receptor focused library and kinase focused library for instance, have been established and widely applied.

Computational method is another critical element to conduct a virtual screening successfully. Up to now, many computational methods have been developed for virtual screening. They are generally divided into two categories: structure-based and ligand-based methods. Molecular docking is a typical representative of structure-based method. It has been broadly used in a variety of drug discovery projects [39]. Molecular dynamic simulation is another example of structure-based method. Although molecular dynamic simulation has an advantage to fully consider the allosteric effect and flexible fit phenomena during protein–ligand interaction process, its utilization in virtual screening requires specialized computer hardware [40–42]. When the structure information of a target is not available, ligand-based method is usually the only option for virtual screening. Generally, ligand-based method is referred to similarity evaluation method. It can be further divided into two classes: similarity searching and compound classification approach [43,44]. Similarity searching techniques include methods that evaluate global chemical structure similarity and that compare local molecular similarity using key features. Fingerprints (ECFPs) are the examples of the former, while pharmacophore model and quantitative structure–activity relationship (QSAR) model are the representatives of the later. Compound classification approach includes machine learning methods such as support

Table 1
The category of epigenetic targets and their representative enzymes.

Category	Family	Representative enzymes
Writers	DNA methyltransferases	DNMT1, DNMT3A, DNMT3B
	Histone acetyltransferases	KAT3A, KAT3B, KAT6A, KAT6B, HAT1, EP300, CREBBP, MYST1, ELP3
	Histone methyltransferases	EZH2, DOT1L, MLL1, MLL2, MLL3, G9A, SETD2, SETD7, NSD2, NSD3
Readers	Bromodomain-containing proteins	BRD1, BRD3, BRD4, ATAD2, TRIM33, PBRM1
	Methyl-binding proteins	MBD1, MBD2, MBD3,MeCP2, MSH6, ING1, ING4, TRIM33
Erasers	Histone deacetylases	HDAC2, HDAC5, HDAC6, HDAC8, HADC9, SIRT1
	Histone demethylases	JMJD2A, JMJD2B, JMJD3, UTX, LSD1, KDM5A, KDM5C, KDM6A