Computational and Structural Biotechnology Journal 14 (2016) 177–184

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

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

Computational approaches in target identification and drug discovery

Theodora Katsila [a,1], Georgios A. Spyroulias [a], George P. Patrinos [a,b], Minos-Timotheos Matsoukas [a,*,1]

[a] University of Patras, School of Health Sciences, Department of Pharmacy, University Campus, Rion, Patras, Greece
[b] Department of Pathology, College of Medicine and Health Sciences, United Arab Emirates University, Al-Ain, United Arab Emirates

ARTICLE INFO

Article history:
Received 31 January 2016
Received in revised form 21 April 2016
Accepted 25 April 2016
Available online 07 May 2016

Keywords:
Data integration
Information technologies
Target identification
Computer-aided drug discovery

ABSTRACT

In the big data era, voluminous datasets are routinely acquired, stored and analyzed with the aim to inform biomedical discoveries and validate hypotheses. No doubt, data volume and diversity have dramatically increased by the advent of new technologies and open data initiatives. Big data are used across the whole drug discovery pipeline from target identification and mechanism of action to identification of novel leads and drug candidates. Such methods are depicted and discussed, with the aim to provide a general view of computational tools and databases available. We feel that big data leveraging needs to be cost-effective and focus on personalized medicine. For this, we propose the interplay of information technologies and (chemo)informatic tools on the basis of their synergy.

Contents

1. Introduction

Current trends in drug discovery focus on disease mechanisms and their understanding, followed by target identification and lead compound discovery. In the era of personalized medicine and better-informed cost-effective public health outcomes, a system of personalized medicine that is based on molecular states (and changes, from DNA to RNA to protein) have become fundamental in drug discovery [1,2]. To build such a system, the molecular characterization of disease is necessary, while environmental influences and the gut microbiome needs to be also considered [3,4]. At the same time, regulatory requirements of safety are increasing [5].

To address the above-mentioned interplay in high-throughput formats, we feel that information technologies and chemoinformatic tools need to be employed on the basis of a synergy that even extends to artificial and human intelligence interplay - humans can detect patterns, which computer algorithms may fail to do so, whereas data-intensive and cognitively complex settings and processes limit human ability [6]. We propose that this synergy will (i) facilitate collaborative data analysis and (ii) guide sense- and decision-making towards rapid and efficient data output. Big and diverse data demand strict filtering and thorough analysis and interpretation. At the same time, biomedicine scientists need to efficiently and effectively collaborate and make decisions. For this, large-scale volumes of complex multi-faceted data need to be meaningfully assembled, mined and analyzed [7]. In such a context, reliable target identification and validation in cooperation

* Corresponding author at: University of Patras, School of Health Sciences, Department of Pharmacy, University Campus, Rion GR-265 04, Patras, Greece. Tel.: +30 2610 962351; fax: +30 2610 997693.
E-mail address: minosmatsoukas@upatras.gr (M.-T. Matsoukas).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.csbj.2016.04.004

Table 1
Drug targets and computational methods used for compound identification and interaction prediction.

Drug target	Computational approach	Reference
p38α MAP kinase	Ligand-based interaction fingerprint (LIFt)	[14]
GPR17	Protein ligand interaction fingerprints (PLIF) method	[15]
Transforming growth factor-b 1 receptor kinase (TGFβ)	Shape-based screening (CatShape, Catalyst)	[49]
T-type calcium channel (CaV)	Bidimensional pharmacophoric fingerprints (ChemAxon and CCG's GpiDAPH3 fingerprints)	[68]
Metabotropic glutamate receptor 5 (mGlu5)	Artificial neural network (ANN) quantitative structure–activity relationship (QSAR)	[72]
prostaglandin D2 receptor 2 (CRTH2)	Proteochemometrics modeling (PCM)	[76]
HUMAN immunodeficiency virus 1 reverse transcriptase (HIV-1 RT)	Molecular mechanics energies combined with the Poisson-Boltzmann surface area (MM-PBS)	[84]
Biotin	Molecular dynamics/free energy perturbation (FEP)	[85]
β-Secretase (BACE)	Linear interaction energy (LIE)	[86]
Chemo-attractant receptor (OXE-R)	Docking virtual screening (PyPx and AutoDock Vina)	[90]
Angiotensin II receptor type 1 (AT1)	Ligand based pharmacophore modeling (Catalyst)	[91]
Pim-1 kinase	Docking virtual screening (Glide)	[104]
Epidermal growth factor receptor (EGFR)/Bromodomain-containing protein 4 (BRD4)	Docking virtual screening (Glide)	[105]
Calcineurin	Structure based pharmacophore virtual screening (Discovery Studio)	[106]

with drug discovery methods will pave the way to more efficient computer aided drug discovery. Moreover, new network-based computational models and systems biology integrate omics databases and optimize combinational regimens of drug development.

2. Target identification

Chemoinformatic tools present a tremendous potential to advance *in silico* drug design and discovery, as they serve the integration of information in several levels to enhance the reliability of data outcomes. To name a few, chemical structure similarity searching [8], data mining/machine learning [9], panel docking [10], and bioactivity spectra based algorithms [11] have been routinely and successfully implemented [12,13]. Some examples are the ligand-based interaction fingerprint (LIFt) approach [14] in predicting potential targets for small-molecule drugs using physics-based docking and sampling methods and the protein ligand interaction fingerprints (PLIF) method [15] for summarizing interactions between ligands and proteins using a fingerprint scheme. In both cases, compounds were identified for the p38α MAP kinase and GPR17, respectively (Table 1).

Target identification can also be studied through network-based drug discovery, a field integrating different levels of information in drug-protein and protein-disease networks. This approach involves a highly collaborative scheme between databases and correlations across genomics, transcriptomics, proteomics, metabolomics, microbiome, pharmacogenomics, which highly depends on the development of relevant computational and systems biology tools for such data interpretation [16,17]. Such approaches, for example relating pharmacological and genomic spaces can be used to develop computational frameworks for drug target identification [18]. Another recent network-based application was the integration of large-scale structural genomics and disease association studies, to generate three-dimensional human interactome, that resulted in the identification of candidate genes for unknown disease-to-gene associations with proposed molecular mechanisms [19].

To facilitate gaining in-depth knowledge of disease mechanisms and/or phenotypes information technologies are greatly needed today more than ever [20]. Indeed, the study of disease mechanisms and/or phenotypes has turned from investigating a particular gene or protein into the analysis of entire sets of biomolecules [21]. The advent of omics technologies further complicates storing, visualizing and analyzing voluminous biological data. For this, information technologies provide the means towards extensive data processing and interpretation. Tools such as the human metabolome database [22] and MetaboAnalyst [23] support integrative omics pathway analysis. The human metabolome database contains metabolite entries linked with chemical, clinical, and molecular biology data, that can assist applications in metabolomics, clinical chemistry and biomarker discovery. Metaboanalyst is a web-based analytical pipeline for high-throughput metabolomics studies, which offers a variety of procedures for metabolomic data processing and integrates biomarker and pathway analysis. MAGENTA (http://www.broadinstitute.org/mpg/magenta/) and Ingenuity (http://www.ingenuity.com/) users can further exploit several curated biological pathways. Databases play a key role and no doubt, an extremely rich repertoire is available today (Table 2). When kinome is of interest, a computational platform ReKINect has been recently reported to identify network-attacking mutations and validated with the interpretation of exomes and quantitative proteomes of ovarian cancer cell lines and the global cancer genome repository [24]. Another useful approach helping to identify functional connections between diseases, genes and drugs is the Connectivity Map [25]. Connectivity Map is a collection of genome-wide transcriptional expression data from cultured human cells treated with bioactive small molecules and simple pattern-matching algorithms that together enable the discovery of functional connections between drugs, genes and diseases through the transitory feature of common gene-expression changes [26]. Other computational methods have been also applied to reconstruct biological networks and extract information from them, such as Bayesian [27] and Boolean networks [28] and graph based models [29].

Furthermore, applications and web services, enable sharing of data and resources for visualization and analysis purposes. The Biological General Repository for Interaction Datasets (BioGRID) [30] is an interaction repository with compiled biological data freely available in standardized formats, linked with software platforms for visualization of complex interaction networks such as Osprey [31] and Cytoscape [32]. BioMart, is a community-driven project, which call for scientists to

Table 2
Web-accessible databases for drug target identification.

Utility	Url
Human metabolome data	http://www.hmdb.ca
In silico target identification	http://www.dddc.ac.cn/pdtd/
	http://www.genome.jp/kegg/
	http://www.geneontology.org
Pathway analysis	http://www.reactome.org
	http://www.pantherdb.org
	http://www.biocarta.com
	http://www.ingenuity.com/
Chemogenomic data	http://www.ebi.ac.uk/chembldb
	http://pubchem.ncbi.nlm.nih.gov
Drug target database	http://www.drugbank.ca
Protein data bank	http://www.pdb.org
Disease specific target database	http://thomsonreuters.com/metacore
Pharmacogenomic data	http://www.pharmgkb.org
Multi-level drug data	http://r2d2drug.org/DMC.aspx
Comparative toxicogenomic database	http://ctdbase.org
Target-toxin database	http://www.t3db.org
Protein expression information	http://www.proteinatlas.org
Therapeutics target database	http://bidd.nus.edu.sg/group/cjttd/