Drug Design Assisted by Computers: New Tools in Bioinformatics

Hafiz Khawar[1], Muhammad Mohsin Zaheer[1], Hamid Khan[2], Umair Akhtar[3], Laiba Bukhari[4], Rahat Zamir[1], Fatima Mushtaq[5], Ayman Naeem[6], Syeda Amber Hameed[7], Fizza Ali[6], Asma Irshad, Ayesha Ameen[9]

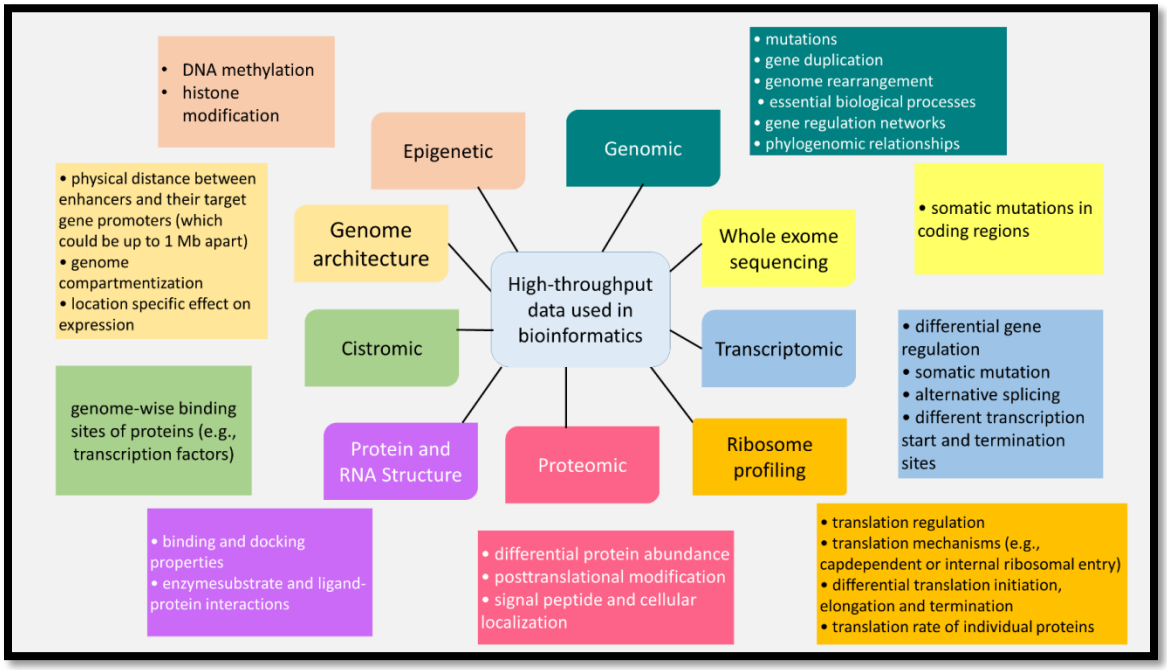
[1]Institute of Industrial Biotechnology, Government College University, Lahore, Pakistan
[2]Department of Biochemistry, Quaid-e-Azam University, Islamabad, Pakistan
[3]Atta-ur-Rahman School of Applied Biosciences (ASAB), National University of Science and Technology, Islamabad, Pakistan
[4] Department of Biochemistry, University of Agriculture, Faisalabad, Pakistan
[5]Service Institute of Medical Sciences, THQ Hospital Sabzazar, Lahore, Pakistan
[6]Department of Life Sciences, University of Management and Technology, Lahore, Pakistan
[7]Department of Microbiology, Ghazi National Institute of Engineering and Sciences, Dera Ghazi Khan, Pakistan
[8]School of Biochemistry and Biotechnology, University of the Punjab, Lahore, Pakistan
[9]Office of Research Innovation and Commercialization, University of Management and Technology, Lahore, Pakistan

Correspondence: ayesha.ameen@umt.edu.pk

Abstract

The utilization of bioinformatics tools can expedite the process of identifying drug targets, examining and modifying drug candidates, and classifying potential side effects and drug resistance. Various high-throughput data, such as genomics, epigenetics, genome architecture, transcriptomics, proteomics, and ribosome profiling, have significant implications for drug repurposing and the development of mechanism-based drugs. Protein-ligand docking experiments and virtual screening are possible due to the creation of protein and RNA structure libraries, advances in structure modeling through homology, and databases of small molecules and metabolites. This article outlines the fundamental framework behind these high-throughput data, summarizes their effectiveness and potential in drug discovery, discusses their limitations, proposes new methods for refining the analysis, and highlights important software and databases commonly used in drug discovery.

Graphical Abstract



Introduction

The effectiveness of a drug for a given disease is largely determined by the symptoms exhibited by the patient during the initial diagnosis. Typically, an effective drug is either a single chemical or a combination of multiple chemicals that work together to alleviate the symptoms with minimal side effects. The drug should be affordable and accessible to even low-income populations, and should also have a low risk of drug resistance and be incapable of reactivation by any bacterial species after human use. This requires not only that the drug be safe and effective for patients, but also that it has minimal adverse impacts on the environment and the larger community.

This study focuses on the role of bioinformatics in drug discovery research. Bioinformatics is an interdisciplinary field that encompasses genomics, transcriptomics, proteomics, population genetics, and molecular phylogenetics. In drug discovery, bioinformatics uses molecular data to compare carriers exhibiting disease symptoms with normal control groups, intending to link disease symptoms to factors modulating gene expression, identifying drug targets that can restore cellular functions and/or induce apoptosis of diseased cells, predicting effective drugs, refining drugs to improve targeting and minimize side effects, and reducing the impact of drugs on the environment and patients.

SEQUENCING OF GENOME AND EXOME

Bioinformatics' contribution to anti-cancer drug development

The first time bioinformatics was used in obtaining the target for the drug was in1983IN identifying the homology sequence between the simian sarcoma virus onc gene and platelet-derived growth factor (Xia, 2017b)(Doolittle et al., 1983). This

comparison was obtained through string matching (Doolittle et al., 1983). Hence, the growth factor was used as a cancer target due to this discovery. The growth factor was used for targeted cancer treatment. This discovery of advantages of bioinformatics led to two schools of thought, which are, by simply changing the transient expression of the growth factor to constitutive expression, the viral transforming factors work (Wooller et al., 2017). This led to the development of anti-cancer drugs by using growth factors as a target. The second thing was that any factors that help or modulate gene expression can have a contributing role in the development of cancer. This type of thinking led to a vast improvement in cancer research and contributed to the new theoretical basis of cancer biology(K. Li et al., 2019).

Database	**Description**	**Link**
COSMIC	Cancer-associated mutations	http://cancer.sanger.ac.uk/cosmic
ClinVar	Clinical significance of variation	http://www.ncbi.nlm.nih.gov/clinvar/
dbGaP	Database of genotypes and phenotypes	s http://www.ncbi.nlm.nih.gov/gap/
dbNSFP	Functional predictions and annotations of nonsynonymous SNPs	https://sites.google.com/site/jpopgen/dbNSFP
dbSNP	Short variation	http://www.ncbi.nlm.nih.gov/projects/SNP/
dbVAR	Structural variation	http://www.ncbi.nlm.nih.gov/dbvar/
Database of Genomic Variants archive (DGVa)	Structural variation	http://www.ebi.ac.uk/dgva
Human Mutation Analysis (HUMA	Comprehensive biological database including variation	https://huma.rubi.ru.ac.za
The Cancer Genome Atlas (TCGA)	Cancer-associated mutations	http://cancergenome.nih.gov/
VnD	Variation and drugs	http://vnd.kobic.re.kr/

Table 1 Databases for Cancer Mutations

Genetic Diseases and The Role of Bioinformatics

Many variations of somatic cells that are associated with inherited genetic diseases were discovered as a result of whole genome sequencing, due to which potential targets for drug delivery have been found (Xia, 2017b)(Ow et al., 2014)(Zhang et al., 2016)(Brown & Tastan Bishop, 2017). Although bioinformatics helped a lot in drug discovery, some difficulty was observed when researching somatic mutations since the identification of the causative mutation is found by observing the genetic differences between the patient and the control (Brücher & Jamall, 2016).

There are three types of somatic mutations; 1. Those that are the reason for the disease and can be potential drug targets; 2. Those that are associated in some way with the mutated gene or its pathway; and 3. Those that are not in any way related to the disease but occur anyway in the patient but not in the normal control. (Wooller et al., 2017)

The second type is mainly used for diagnosis of the disease and not for the drug target. The third type is omitted due to two reasons; reason one being that by increasing the sample-size; if many types of breast cancers have the same somatic mutation then the relevancy of the mutation is high when compared to a somatic mutation that occurs only in one type of breast cancer. The second reason is that by collecting longitudinal data; there is a need to recognize that many diseases have a genetic determinant occurring long before there are any symptoms of the disease.

It is difficult to discern between the first two types of genetic mutational differences that may occur in the patient and the control without knowing the mode of the mechanism of the disease. There can be a loss of function mutation in the coding sequence, in the regulatory motif, or the enhancer (Xia, 2017b). To find out if the mutation has any major impact on the function of the gene professionals in bioinformatics take three approaches; the first approach that if the mutation completely changes the type of amino acid; the second approach is whether the mutation has occurred at a conserved non-coding region; and thirdly, if the mutation has occurred in a cellular machinery's known signal (van Driel & Brunner, 2006)

The third point is done through the widely curated databases of known regulatory motifs (Aliyu Musa et al., 2018; MotifMap: A Human Genome-Wide Map of Candidate Regulatory Motif Sites - PubMed, n.d.), (Daily et al., 2011), (Jasper J. Koehorst, Jesse C. J. van Dam, Edoardo Saccenti, Vitor A. P. Martins dos Santos, 2017; M. J. Li et al., 2015), (Wu et al., 2008). To do this bioinformatic tools are used to scan the genomes like position weight matrix, Gibbs sampler, and support vector machines. It is necessary to find the regulatory motifs since they could be a response from the nuclear receptors and therefore this

identification leads to drug target refinement. This is done through DAMBE which can extract CDS, rRNAs, tRNAs, introns, exons, 5' and 3' splice sites, upstream or downstream sequences of gene features, etc. when specific genome sequence is given (Xia, 2017b). If the loss of function is due to the occurrence of deletion mutation than paralogous gene or an alternative cellular pathway can be identified by bioinformatics.

TOOLS	URL	FUNCTIONS
Condel	http://bg.upf.edu/fannsdb/	Combines FATHMM, mutation assessor etc.
FATHMM	http://fathmm.biocompute.org.uk/about.html	Combines FATHMM, mutation assessor etc.
Mutation assessor	http://mutationassessor.org/r3/	Based on evolutionary conservation of the affected amino acid in protein homologues
Polyphen-2	http://genetics.bwh.harvard.edu/pph2/	Uses straightforward physical and comparative considerations
SIFT	http://sift.bii.a-star.edu.sg/	Based on sequence homology and the physical properties of amino acids
TransFIC	http://bg.upf.edu/transfic/home	Transforms functional impact scores provided by other tools by taking into account the differences in basal tolerance to germline single nucleotide variants (SNVs) of genes that belong to different functional classes
CHASM	http://wiki.chasmsoftware.org/index.php/Main_Page	Probability that the mutation gives the cells a selective survival advantage

Table 2Bioinformatic Tools for Cancer Research

Pathogenic Human Diseases and Bioinformatics

Even for drug targets against pathogenic diseases, well interpreted genomes are necessary. There are four steps through, which bioinformatics is used; first, the potential and essential genes are identified from the pathogen as target for the drug; second, all potential pathways are discovered and annotated. Any minute difference can be used as potential drug target; third, the homologue genes in the host are seen against the essential pathogenic genes. Comparative genomics are done to discover the singular difference between the host and pathogen homology genes that can assist in the designing of the drug; fourth, it is imperative for the drug to be precise in its pathogen so it cannot cause drug resistivity.

Analysis through bioinformatics revealed that there is a glutamate transport system which is only existing in Clostridium perfringens and absent in mammals and birds, due to this the perfect target for the drug could be developed that can protect humans as well as other mammals along with avian species (Xia, 2017b)(Bhatia et al., 2014).

Sequence comparisons between the host and pathogen have shown potential sites for the drugs to target at like a unique insertion in the pathogen. This can also become a significant role in pathogen specific drug that is even sub-specie specific. On the other hand, genomic analysis can also help in the repurposing of already present drugs against other pathogens(Farha & Brown, 2019). This is a very cost-effective process. Complete analysis of the genome also helps in understanding the drug actions.

All significant cell metabolic processes are redundant therefore it is necessary to understand the functional part of the mechanism as it is crucial in developing the most effective drug against the pathogen as possible(Farha & Brown, 2019). Other than this, certain pathogen under stress condition can activate alternative biological pathways to ensure the survival and growth of the pathogen(Hernandez et al., 2019). A drug cannot be efficient against the disease unless all alternative pathways are understood.

The evolutionary perspective and the integration of molecular phylogenetics in bioinformatics contributes a lot in the resolution of multiple controversies seen in the molecular mechanisms(Ebhohimen et al., 2020). DAMBE is used for comprehensive phylogenetic study. In the previous years, studies that involved highly diverged bacterial and viral species had difficulty to obtain reliable multiple sequence alignment, for this, pairwise alignment based phylogenetic method was developed to facilitate the studies that involved highly diverged species (Xia, 2016)(Lakshmi & Ramyachitra, 2020).

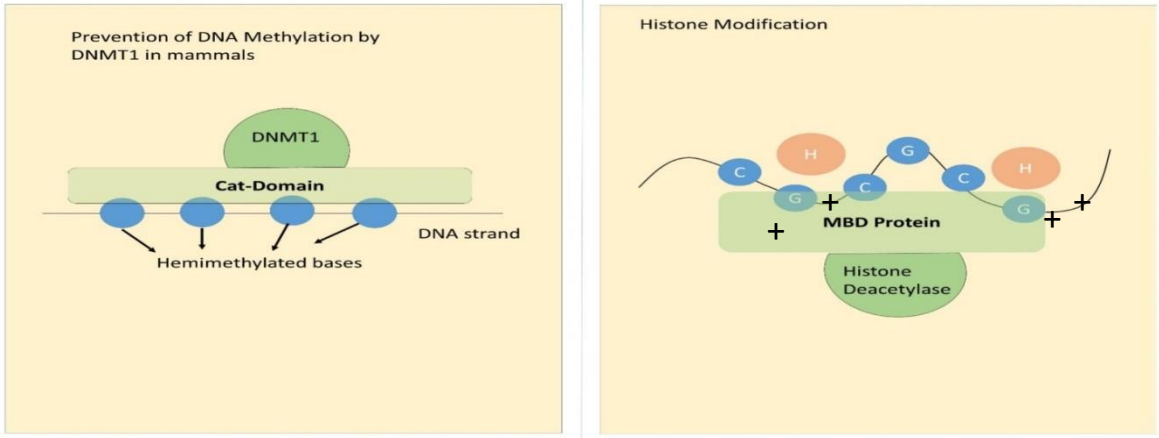
Epigenetics and Genome Architecture in Drug Discovery

The term “epigenetics” was first described in 1942, by Waddington.(Waddington, 1942) According to him, it is the branch of genetics which studies the interactions between genes and their products and the phenotypes they produce. Goldberg et al. have defined this term as the relationship between genotype andits corresponding phenotype.(Goldberg et

al., 2007) It is a phenomenon in which the DNA sequence stays the same but the products produced by a gene, its phenotype, are changed or altered due to methylation or other changes due to environmental factors. Researchers have given the reasoning for identical twins having same genome but different disease susceptibility and phenotypic differences, to be because of epigenetic factors. Case studies of twins have revealed much information about epigenetics and how various environmental factors influence it.(Petronis, 2006; Petronis et al., 2003) Many diseases are related to epigenetics, more than genetics, for example schizophrenia, adrenoleukodystrophy (ALD), cardiovascular diseases and many more. (Korenke et al., 1996; Ordovás & Smith, 2010; Petronis, 2004)Many studies have shown the dependence of disease and propagation, immune response of the body, all depend on epigenetic factors. (Jiang et al., 2004; Zoghbi & Beaudet, 2016)Two major events occur in epigenetics, one is DNA-methylation and the other is modification of histone proteins, to which the DNA is wrapped around. DNA methylation occurs in mammals by three major enzymes, known as derivatives of DNA methyl transferases, often abbreviated as DNMTs. The DNMT1 is involved in the maintenance of methylation, while the enzymes of DNMT3 family methylate the hemimethylated or unmethylated bases. (El-Osta, 2003; Fatemi et al., 2001; B. Jin & Robertson, 2013) This maintenance of methylation is obtained when the CatD domain of DNMT1 binds with the hemimethylated bases and prevents their de-methylation.(Fatemi et al., 2001)In mammals, gene silencing is maintained by tightening the DNA wrapping around the histone proteins, which makes its transcription hard and hence that DNA is known to be silenced. This happens when proteins having an MBD domain (methyl-cpg binding domain) bind to the methylated CpG islands and then bind to histone deacetylase enzyme which removes the acetyl group from histone proteins and makes the lysine residues of histone protein, positive, which causes tight binding of negatively charged DNA. (Song et al., 2012; Wade & Wolffe, 2001)Figure 2

shows a pictorial representation of epigenetic modification. Recent research by Ren et al. has shown that the maintenance of DNA methylation by DNMT1 depends on the interaction between the RFTS domain of DNMT1 and H3K9me3Ub (ubiquitylated H3K9me3).(Ren et al., 2020) H3K9me3 is a modification of histone 3, which occurs by tri-methylation of 9^{th} lysine residue of histone 3.(Huang et al., 2015)

Figure 1: Epigenetic modification by methylation and histone deacetylase.



Common drug targets

When a gene is silenced, the effect is same as that of a mutation that causes loss of function in a gene. Some cancers occur when the proteins involved in the cell apoptosis pathway are not expressed because of the methylation of DNA or the acetylation of histone in the gene which express them.(Insinga, Minucci, et al., 2005; Insinga, Monestiroli, et al., 2005) Therefore, in making of cancer drugs, inhibitors of histone deacetylase enzyme are being used to inactivate the enzyme and activate the genes of the apoptosis pathway, (Bolden et al., 2006) but a problem is associated with this approach, it is that inhibitors of histone acetylase may affect the expression of other genes as well, and this hurdle of unspecific action does not let such drugs enter trials.(Voelter-Mahlknecht,
2016)Methodologies for the editing of the epigenome to treat diseases are being devised, with focus on the specificity of action.(Kungulovski & Jeltsch, 2016)As mentioned before, gene silencing is achieved by DNA methylation and histone deacetylation, but further analysis of it is also required which is gained by high-throughput screening. In high-throughput screening, first a lot of data is required, like the pattern of DNA methylation which is obtained by bisulfite sequencing,(G. Grigg & Clark, 1994; G. W. Grigg, 1996) the binding of DNA and protein and their interactions,(Robertson et al., 2007) which is obtained from Chromatin immunoprecipitation and the genome architecture data obtained by Hi-C.(Lieberman-Aiden et al., 2009) Genome architecture is affected by the DNA and protein binding, which is turn affected by DNA methylation. By getting knowledge about genome architecture data, we can get to know about the interaction between the promoter and the enhancer region of the DNA.(Pazin et al., 1994, 1996, 1998)Translocation experiments have shown that gene expression depends on it location in the genome, its distance from the promoter region and enhancer region. (Muller & Altenburg, 1930) More recent experiments have shown that some interactions between protein and DNA occur which reconfigure the nucleosome, bring the enhancer and

promoter regions in close proximity to eachother.(Pazin et al., 1998) This data gave rise to the gene expression model known as "enhancer-hub model".(Palstra et al., 2003; Tolhuis et al., 2002) This model suggests that the enhancers form a hub and the genes whose promoter lies close to the hub is more efficiently expressed. On the other hand, if the hub is deleted then the gene is silenced, because the expression of gene entirely depends on its location with respect to the hub.

For designation of a drug target, we need to identify the signal of methylation on DNA and how can it be modified to control the gene expression. In some diseases, further research analysis has shown that the patients have correct, un-mutated genes but even then the expression is abnormal. The explanation to this is that protein expression not only depends on the mutations in gene, but also on the mutations which have occurred far away from the gene, especially in its promoter, or enhancers and other up-stream or down-stream regulating elements. Beta-thalassemia is an example of such a disease.(Kioussis et al., 1983)For curing such diseases by formulating drugs, we need high-throughput data about the genome architecture and DNA-protein interactions, and how can we specifically reconfigure the gene, with its hub and associated proteins involved in gene expression. (Deng et al., 2012, 2014) The silenced genes can also be re-activated by altering the methylation pattern. Moreover, by controlling/inhibiting methylation, gene expression can be controlled. In the genomes of mammals, some CpG islands are methylated, while others are not, thus, by using bioinformatics tools, we can check the adjacent upstream and downstream sites of these islands and compare the differences. By this approach, we can check which other factors control methylation and how to control them.(Ma & Xia, 2011)Some experiments of assay of methylation have been done, but not yielded satisfactory results.(Bibikova et al., 2011; Eckhardt et al., 2006) Methylation is also dependent on SAM molecule, which is S-adenosyl L-methionine, and acts as a donor for methyl group, DNMTs take this methyl and attach it with DNA bases for methylation. The enzymes involved in this pathway can also be cause for epigenetic changes and a target for drugs.

Designing of nano-medicine

Nano-medicine can be designed to correct the wrong or irregular methylation patterns, but for this a template of correct methylation should be known. Some experiments have been done on this, and the results have revealed that some RNA's have a role in methylation and epigenetic silencing. (P. Jin et al., 2004) The DNA in zygote undergoes demethylation so that different organs can be formed from cells have the same DNA, and this indicates that epigenetics does not depend on the DNA sequence. (Clark, 2015) Some experiments by Bao et al. have shown that during sperm formation in men, when spermatids differentiate into sperm, the histone proteins of the sperm are converted to protamines.(Bao & Bedford, 2016) Hence this information further suggests that epigenetic code cannot lie in proteins, since they also vary. One thing that does not vary is the RNA. It can be hypothesized that RNA which do not encode for proteins, may have a role in epigenetics, especially non-coding RNAs or lncRNAs. Experimentation and

analytical assays have revealed that RNA is involved in the recruitment of proteins on DNA which take part in methylation.(Chu et al., 2011; Pandey et al., 2008; Rinn et al., 2007) Short RNAs present in sperm(Sendler et al., 2013) may induce epigenetic changes in the baby.(Gapp et al., 2014; Rodgers et al., 2015) For drug discovery, the RNAs participating in epigenetic modification can be a good target.

Transcriptomics and Drug Discovery

Transcriptomics is a study of all the RNAs found in a cell. In other words, transcription product is RNA and hence a compilation of the transcription products is known as transcriptome, and this study is known as transcriptomics. For this, RNA-sequencing technology is used and the gene expression ismeasured.(Z. Wang et al., 2009)The data from RNA-sequencing is used to compare the data of isoforms of RNA which form as a result of alternate splicing, transcript sizes and the regulation of genes, from a healthy group and a group of diseased patients. This way we can identify the problem in the diseased patients' transcriptome. Many transcriptome analyses have led to the identification of core problem, in diseases like cancer.(Arvaniti et al., 2016; Berger et al., 2010) There are two approaches of transcriptomics which contribute in drug design, these are screening of phenotype to design or make drug candidates more potent, while the other is to identify the drug targets, which is usually an abnormal protein.

Phenotypic Screening and Some Recent Case Studies

Phenotypic screening involves the choosing of the best drug amongst many drug candidates, the changing of phenotypes and how this affects the response to drugs, and those drugs are selected which produce the desired effects in organisms. The selected drugs enter further trials where they are further tested and very few enter the market at the end. Phenotypic screening does not focus on the mechanism of action of drugs. However, this approach allows the identification of active ingredients in drugs, like in the case of malaria drug artemisinin.(Miller & Su, 2011)Target-based approach is a good and effective approach in designing drugs for simple diseases like single gene disorders whereas, for complex diseases which occur due to mutations in multigenes, phenotypic screening is more helpful and potent.(D. C. Swinney, 2013; David C. Swinney & Anthony, 2011) Phenotypic screening has proved to be a very effective approach for designing drugs against cancers, since cancerous cells have heterogeneous genetic makeup. (Moffat et al., 2014) Phenotypic screening makes previously discovered drugs more potent and efficient. A study by Laura et al. studied and analysed "Leishmaniasis" drugs. Leishmaniasis is a disease caused by protozoans belonging to the genus Leishmaniasis. It had been observed that the drughad a different effect on different species, therefore, by phenotypic screening Laura et al. identified 51 different active compounds for each species.(Alcântara et al., 2020)Many bacterial species have become anti-biotic resistant and the diseases caused by them are not being cured by normal, regular medications. This problem can also be overcome by phenotypic screening and identification of new molecules or

modification of previously existing drugs to target such microbes.(Mohammed & Alghamdi, 2020) Similarly, in another recent study by Gruber et al. the human sperms were phenotypically screened and many targets for contraceptive medicines were identified

While designing cancer drugs, there are two main approaches, either we can make the gene expression of cancer cell to that of a normal cell, or we can induce apoptosis in the cancer cell. In a recent study by Scott et al. phenotypic screening of cancer cells had led to the discovery of drug target, which was serotonin receptors. Serotonin induces tumor growth and in the study, they discovered modulators which target serotonin receptors of breast cancer cells selectively, and suppress their growth.(Warchal et al., 2020)

$$I_{dd} = \ln\left(\frac{D_{np} - D_{dn}}{D_{nd} + D_{dp} - D_{np}}\right)$$

Index of Drug Desirability:

Gene expression can be used to figure out the desirability and potency of a drug by using bioinformatics tools. There is a formula for the drug desirability, known as index of desirability, denoted by I_{dd}. This formula keeps track of the side effects and therapeutic effects, keeping the phenotypic expression of a healthy control group as a model. The phenotypic expression profile of a patient is designated as "Gp" while that of healthy people is denoted as "Gn" and that of patients treated with the drug is denoted as "Gd". The disctances between these values are calculated which are basically representative of the symptom severity, drug efficiency by reduction in symptom intensity.(Gabrielsson & Green, 2009; Holford, 2011) D_{np} is the difference between Gn and Gp, D_{dp} is the distance between Gd and Gp, and D_{dn} is the distance between Gd and Gn. D_{np} also represents how severe the disease symptoms are. The difference between D_{np} and D_{nd} shows the reduction in symptom severity. When D_{nd} and D_{dp} are added, and the sum is subtracted from D_{np}, then we get a measure of the side effect.(Xia, 2017a) The total formula is as follows:

…… **Equation 1** Index of drug desirability

F**igure 3**, 'A' and 'B' are two drug candidates and their index of desirability is being calculated, which clearly indicates that drug A is more potent than B. In the given situation, the disease is same, but drugs are two different substances. The symptom severity is being kept constant. The side effects of drug 'B' are more than those of drug 'A' hence 'A' is more potent and likely to enter the market.

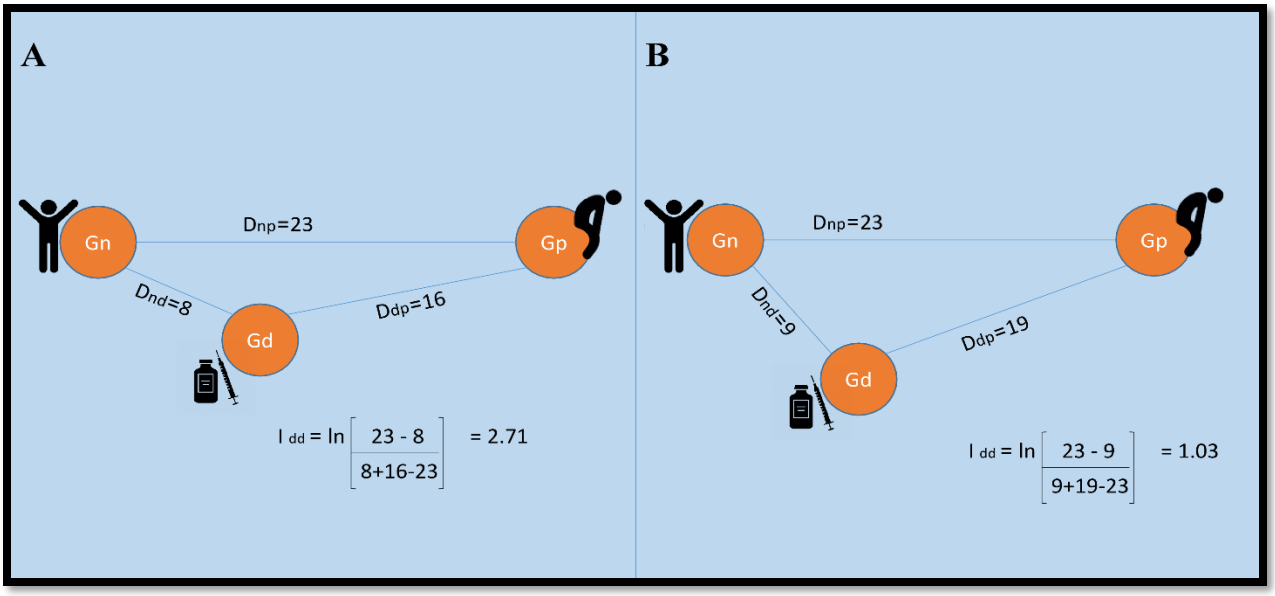


Figure 2 Graphical representation of the equation of Index of desirability. A and B are two drugs, A has higher index of desirability so it can enter further trials, but B has lower index of desirability

Toxicogenomics

Toxicology is the study of harmful effects of drugs on humans or other living organisms. Epigenetics also play a vital role in toxicity of drugs, hence toxicity depends on the genotype and phenotype of the individual. Various transcriptomic approaches are used for the evaluation of toxicogenomics, like DNA microarray, PCR, RNA sequencing and single cell RNA sequencing (also denoted as scRNA).(Kinaret et al., 2020)**Table 3** shows a few databases which are publicly available and used for toxicogenomic analysis of drug.

Database	Description	References
Chemical Effects in Biological Systems (CEBS)	It is a free online web tool, https://tools.niehs.nih.gov//cebs3/ui/ It contains data of microarray and proteomics. It was designed by National Center for Toxicogenomics (NCT)	(Lea et al., 2017; Waters et al., 2008)
Connectivity Map	Also known as CMAP. It contains the gene expression profile of many human genes and it is still a growing	(Lamb, 2007)

	database. It also finds connections between diseases, drugs and physiological processes in the body.	
LINCS 1000	It is an expansion of CMAP. It contains mRNA expression profiles for 77 cell lines.	(Subramanian et al., 2017)
DrugMatrix	It contains the gene expression profile of rats and the response of their bodies towards drugs. It shows the relationship between gene expression, toxicity, drug anddiseases.	(Ganter et al., 2006)
Open TG-GATEs	It contains toxicogenomic data from rats and in vitro human liver cells lines.	(Igarashi et al., 2014)
ArrayExpress	It stores toxicogenomic data from experimental results which are being carried out. Then these results can be used in other studies.	(Kolesnikov et al., 2015)
Gene Expression Omnibus	It contains DNA microarray data and NGS data from experiments. It is a public platform and researchers add data from their experiments. The data is also available for download.	(Clough & Barrett, 2016)

Table 3 Common databases used for toxicogenomic analysis for drug discovery and design.

Proteomic Data and Drug discovery

Proteins, which are integral to the functioning of living cells, can be overproduced in disease. Since a gene's transcription does not always guarantee its translation and different proteins degrade at different rates, transcribed gene data is an unreliable predictor of protein abundance (Xia et al., 2011). Consequently, comparing proteomes between patients and healthy individuals is often more effective than using genomic or transcriptomic data to identify potential targets for drug development (Xia, 2014). The availability of public databases, such as PaxDB, which collect proteomic data from a wide range of model organisms, has facilitated the development of measures to predict translation. This abundance of proteomic data has enabled the creation and refinement of frameworks for translation forecasting. (Prabhakaran et al., 2015).

The bioinformatics approach to reasoning with proteomic data is similar to proteomic data analysis, which involves using proteomic data to discover phenotypes and drug targets. In studies comparing treatment and control animals or patients, the majority of the proteomic data is used, and the usual search is matched (Ujcikova et al., 2016). For example, rats given caffeine exhibit different protein expression compared to control rats. Many links between drugs and protein targets have been reported and stored in databases, enabling the identification of potential interactions between a query drug and specific proteins. (Garg et al., 2016).

Proteomic data can suffer from the same issues as genomic and transcriptomic data when analyzed spontaneously without allowing for sufficient time for a cohort. It is particularly challenging to determine the underlying cause of a disease based solely on differential expression of numerous proteins. The abundance of different proteins varies throughout various stages of the cell cycle (Bitsika et al., 2016). Comparing protein or transcriptomic profiles of patients and normal individuals without taking into account temporal and spatial cell heterogeneity can lead to an excess of false positives with little significance for drug discovery (Heath et al., 2016). Animal models can be used to screen cells over different time periods. To reconstruct a cell-cycle gene expression profile, a single-cell characterization of transcriptomes and proteomes over time would provide much more meaningful results, such as reordering the cell expression profile from phases 3,1,2,4 to phases 1,2,3,4.

Ribosome Profiling and Drug Discovery

The quantity of protein data has limitations because

1. It is often difficult to identify low-concentration proteins, short peptides or transient proteins.
2. Membrane proteins that also act as important components of signal transduction far from isolation, separation and purification.

Transcriptomic data once gave rise to a hope that transcriptomic data would predict proteomic data (Xia, 2014) . The ties between abundance of mRNA and protein abundance are skewed by differential mRNA translation efficiencies and protein degradations efficiencies. Ribosome profiling information, combined with transcriptomic data, should generate good protein production predictions. Information on the abundance of mRNA and the efficiency of translation are given by transcriptome and ribosome profiling. If mRNA values of genes A and B are N_A and N_B from transcriptomic data, and their respective values. The translation effectiveness is R_A and R_B from ribosome-specific profiling results, and then N_A*R_A and N_B*R_B respectively are their relative protein production levels. Differences between the expected volume of protein and the protein observed Protein degradation rate can be measured with abundance. The acquisition of transcriptomic and proteomic data in the same experiment, ideally from one single cell, should facilitate that prediction (Smircich et al., 2015) .

The data for the profiling of ribosome is now almost limited from the depth sequences of mRNA fragments (RPF, ~30 nucleotides) that is protected against ribosome. However,

both approaches are highly consistent with yeast results. The ribosome positions on mRNA can be attached by the sequenced RPFs to protein-coding genes. Ribosomal density can be viewed as an efficiency proxy for translation but ribosome's will gradually travel and get thickly packed along the mRNA for an mRNA with low codon consumption. Of this reason the efficiency of elongation has to be regulated, of example by regressing the density of the ribosome in the elongation index. Ribosome data profiling is appropriate in defining regulatory factors, which are Poly(A) tract which modulates the ability of translation, e.g. short poly(A) at 5' UTR may accelerate the recruiting of it and enhancing transformation aspects but long poly(A), binding and inhibiting protein binding to poly(A) translation. These regulatory objectives may be used for easy-to-use measurable drug targets (Xia et al., 2011) .

Two variables define four key models of the initiation of a translation. Secondly, the scanning cycle of beginning codon begins with the 5-pound end of mRNA or from the internal ribosome entry sites (Jackson et al., 2010). The second is whether the tiny ribosome device scans for the starting codon or whether the scan can also be carried out by a completely developed ribosomeWhile the occurrence of internal ribosomal entering is little discussed, only modern data on the ribosomal profiling has given the fully formed ribosome a good scientific backing for 5'UTR mRNAs (Ingolia et al., 2014).

Unlike the internal ribosomal eukaryotic input sites various viral IRESs have a solid secondary structure whose IRES activity decreases due to secondary structural stability. (IRESs) The Cricket Paralysis Virus (CrPV) has IRES in the intercistronic region, which is able to interact precisely with the ribosome through its complex secondary structure without the need of any translation initiation factors (Boehringer et al., 2005). The Hepatitis C Virus (HCV) has IRES which can imitate translation inputs so that the initations necessary for the cap-dependent translation are not required. The IRES initiation mechanism allows viruses to transmit throughout the host cap-reliant Translation was shut down and viral IRESs, in particular those of fairly rigid secondary and tertiary structure like HCV, were therefore shut down promising drug targets (Schüler et al., 2006).

A cellular mechanism which is essential to react to the extracellular environment is controlled via translation. A couple or so genes are transcribed but not regularly translated for Saccharomyces cerevisiae yeast; they are transferred once surface nutrients are decreased and its component allow yeast cells to burrow down to the medium of the crop to extract development nutrients (Schüler et al., 2006). The translation position of these controlled communications can be revealed by ribosome profiling data and thus guides to know in what way organizations are adapting translation regulations to respond to changes in the environment.

One fundamental problem in the analysis of ribosome profiling data is the assignment of RPFs to paralogous genes when the RPF affiliates multiple genes comparatively well. Transcriptomic and protectomic data share

this problem where protein labeling is usually carried out with fingerprinting of peptide mass and a fragment of peptide can equally well suit several proteins. In most of the programs there are two unsatisfactory options:

1. Sequence reading that coincides with several paralog genes
2. A recent software available at https:/github.com/ratschlab/mmr (MMR: multiple mapper resolution) aims to fix this problem but does not provide a methodological approach Since multicellular eukaryotes have a large number of duplicate genes, irrelevant position of RPFs to paralogous genes leaves all next research devious. Furthermore the method for referring RPFs to three or many paralogous genes in the Tuxedo computer system (Trapnell et al., 2012). Unless the gene family having only two representatives.

To properly assign the sequence of three or many paralogous genes in the family, a phylogenetic tree is required. I demonstrate the basic concept of distribution with the gene family of three paralogous A, B and C genes highlighted in the following work of three segments. The three genes had the same central segment with 23 readings (which automatically matched all three paralogues). Genes B and C share the same first segment, 20 reading Genes A matched to their first segment and four matched reads have been provided. These three genes also have a third segment in which the similar gene A, B and C matched 6 and 12 matched. We then have to assign the three paralogues to the 23 readings assigned by all three and 20 shared by B and C.

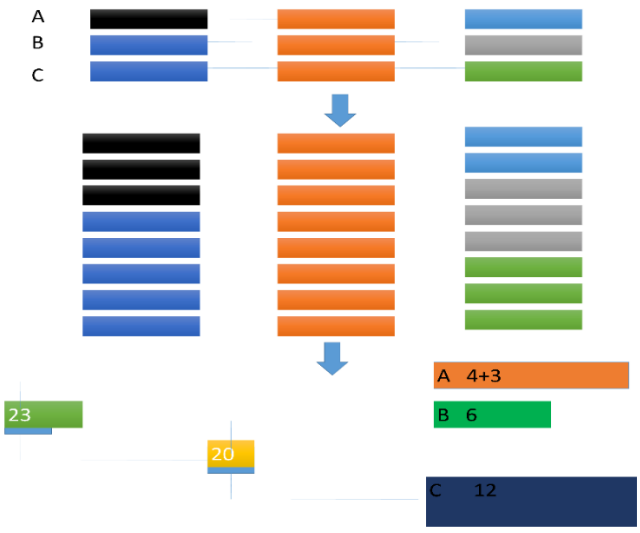


Figure 3**)** Distribution of accessed readings in the gene family of three paralogous genes A, B and C of three idealized sections with a symmetric identical middle segment, a highly homologous first segment which is identical in B and C, and a divergent third segment

A simple counting approach is used in TUXEDO by the following derivation:

$$P_A = \frac{3+4}{3+4+20+6+12} = 0.15556$$

$$P_B = (1-P_A)\frac{6}{6+12} = 0.28148$$

$$P_C = (1-P_A)\frac{12}{6+12} = 0.56296$$

Equation 2simple counting approach is used in TUXEDO

Thus, we assign 23 equally matched readings to the paralogous genes A, B and C according to PA, PB and PC, in both. For the 20 reads that matched B and C equallywell, we allocate 20 * 6/(6 + 12) to B and 20 * 12/(6 + 12) to C. The number of matches to each gene is estimated as follows:

$$N_A = 3+4+23P_A = 10.57778$$

$$N_B = 6+23P_B+20\left(\frac{6}{6+12}\right) = 19.14074$$

$$N_C = 12+23P_C+20\left(\frac{12}{6+12}\right) = 38.28148$$

Equation 3 Final equation

Structural Biology and Drug Discovery

A perfect bioinformatics stage for sedate revelation dependent on auxiliary science ought to permit one to 1 foresee the 3-D structure of a protein. (Xia, 2017a) It has for quite some time been perceived that information on the 3D structures of proteins can quicken drug discovery, yet ongoing improvements in genome sequencing, mechanical technology, and bioinformatics have profoundly changed the chances. Various new protein targets have been perceived from genome examinations and concentrated by X-beam assessment or atomic attractive reverberation NMR spectroscopy. Basic science has been instrumental in coordinating not just lead advancement and target distinguishing proof, where it has entrenched jobs, yet besides lead disclosure, presently that high-throughput techniques for structure resolution can give strong ways to deal with screening.(Congreve et al., 2005) In previous years, macromolecular crystallography has become a standard method utilized by numerous pharmaceutical and biotechnology organizations. This procedure of protein-ligand cooperation's at levels of goals practically unparalleled by some other method and this methodology holds the guarantee of the novel, increasingly powerful, more secure, and less expensive medications. Although crystallography stays a difficult and rather costly strategy, striking advances in structure assurance and structure-based medication plan (SBDD) have been made as of late. This strategy has been upheld by late mechanical turns of events, for instance, high-throughput crystallization, world class synchrotron beamlines, and new strategies in fundamental

bioinformatics and computational science incited by the essential genomics effort.As a result of the expanded accessibility of basic information, the utilization of structure-based data has extended from basic protein-ligand association investigation to incorporate different parts of the medication revelation process like objective choice and beginning lead disclosure that used to be nearly the selective property of biology and chemistry.(Scapin, 2006) As in numerous parts of the medication disclosure process, critical cooperative synergies can be acknowledged in structural biology by the contemporaneous quest for some objective proteins from a structural category and functional category We will audit a portion of those cooperative energies here utilizing the case of the **protein kinases**.(Stout et al., 2005) Protein kinases have an essential job in signal transduction pathways, and atypical kinase action has been seen in numerous illnesses. As of late, kinase restraint has become a significant zone for healing intercession and an assortment of kinase inhibitor pharmacophores has been reporting. This audit represents a portion of the crack and results in the field of structure-based plan of protein kinase inhibitors. The techniques and results talked about here represent the intensity of structure-based plan in lead revelation, for instance using virtual screening and in controlling the enhancement of the pharmacological properties of these particles.(Scapin, 2002) Using some tools to check the structure and Protein-Protein interactions by PDB, Swiss Model, TASSER, and PyMOL.(Xia, 2017a)

Uses of structural biology

Structural biology has been utilized to recognize the instruments of infection, and structure-guided methodologies have shown obviously that they can add to numerous parts of early medication disclosure, both computationally and tentatively. The structure can likewise tell our comprehension regarding the effects of mutations in human hereditary ailments and medication opposition in malignant growths and irresistible illnesses. We talk about the manners in which that structural biology may be valuable in both repurposing off-permit medications and guidethe plan of new atoms that may be less permitting to drug resistance later on.(Pandurangan et al., 2017)

Computer-Aided Drug Design

Drug discovery depends on the investigation of the procedure of the disorder, atomic targets, and ligands, in which cooperation with the objective could guide the standardization of the obsessive procedure. The accessible information on ailments, drugs, pharmacological impacts, atomic targets, and medication like substances, considering the combinatory of the cooperative relations between them, compared to the Big Data. To dissect such information, the use of PC supported medication structure strategies is fundamental. A review of the investigations around there performed by the Laboratory for Structure-Function Based Drug Design of IBMC is introduced. We have built up the ways to deal with distinguishing promising pharmacological targets, breaking down protein-ligand collaborations dependent on amino corrosive arrangements, recognizing sub-atomic instruments of symptoms of

medications, figuring the necessary harmfulness of medications considering their digestion, have been created in the human body.(Poroikov, 2020) Using PyMOL we have grown such a graphical UI consolidating singular scholastic bundles intended for protein arrangement, atomic mechanics applications, and docking and scoring.(Lill & Danielson, 2011)

Semiempirical Quantum-Mechanical Methods

The semiempirical quantum mechanical systems used in the medication configuration are normally parametrized and taken a stab at enlightening files of structures that may not be operator models for sedating biomolecule coordinated efforts to the extent both size and invention association. This is tended to here with another benchmark enlightening record, PLF547, got from protein-ligand structures, containing structures of ligands with protein parts, for instance, amino-corrosive side chains. From these, composite benchmark participation energies are in like manner worked for buildings of the ligand with the absolute unique site of the protein (PLA15 educational collection).[9][10] Quantum mechanical-based techniques have been used starting late for choosing protonation states of ionizable social occasions, evaluating energies, and propelling sub-nuclear structures. For high throughput in silico screening, Quantum mechanical methodologies have been used to decide incredible quantitative structure-development relationship models. Regardless, wide testing of conformational space and treatment of a course of action of macromolecules are so far compelling components for the wide utilization of QM in sedate design. [11][12]

Classification of Computer-Aided Drug Design Methods

We study biological-pharmaceutical nature in computer-aided drug design. Three major groups of computer-aided drug design: structure-based, ligand-based, and hybrid methods.

Structure-Based Methods

Structure-based together strategies depend on respect to the 3D data of the sub-atomic objective. Unmistakable instances of these techniques are docking and sub-atomic elements (MD) Applications of structure-based strategies incorporate portrayal of restricting destinations, clarification of the system of activity of dynamic particles at the sub-atomic level, and evaluation of the energy and thermodynamics engaged with ligand-target acknowledgment.(Prieto-Martínez et al., 2019) Toll-like receptors (TLRs) assume an essential job in different incendiary, immune systems, and neurodegenerative issues, for example, sepsis, rheumatoid joint pain, provocative entrails malady, Alzheimer's ailment, numerous sclerosis, malignant growth, and fundamental lupus erythematosus. TLRs, especially TLR4, have been distinguished as potential medication focuses on the treatment of these maladies, and a few pertinent mixes are under preclinical and clinical assessment. Computational investigations and strategies that have given bits of knowledge into TLR4-focusing on therapeutics. Besides, this article gives an outline of the computational strategies that can profit an expansive crowd in this field and help with the improvement of

novel medications for TLR-related disarranges.(ul Ain et al., 2020)(Mazanetz et al., 2019)

Ligand-Based Methods

The ligand-based method is based on the data on the synthetic structures of a set of ligands with the known organic movement. One of the fundamental objectives of these strategies is recognizing bioactive mixes or better the movement of dynamic particles. Ligand-based techniques are similitude looking and QSAR demonstrating.(Prieto-Martínez et al., 2019) Information is being utilized to advance early drug discovery firms from hit distinguishing proof to applicant determination. Huge improvements in information mining techniques and the availability of devices for research researchers have been instrumental in diminishing medication disclosure courses of events and in improving the probability of a substance element accomplishing drug advancement achievements. KNIME, the Konstanz Information Miner, is a main open-source information examination stage and has upheld medicate disclosure tries for longer than 10 years. KNIME gives a rich palette of apparatuses bolstered by a broad network of supporters of empowering ligand and structure-based drug design.(Mazanetz et al., 2019)

Hybrid Methods

At the point when the structure of the objective is referred to just as the structure of dynamic atoms, it is achievable to utilize transformation or merged techniques, i.e., a blend of structure-based and ligand-based strategies. A model is sure techniques for pharmacophore demonstrating. Different models are in silico ways to deal with anticipate bioactivity dependent on the natural profile of mixes tried versus one or different objectives.(Prieto-Martínez et al., 2019) Peptides assume vital jobs in different physiological and diseased procedures. The examination of peptide-based medications is a feature in the innovative work of new medications. Characteristic peptides are not generally perfect decisions for clinical application because of their set number and now and then cytotoxicity to typical cells. Intending to increase more grounded or explicit or novel organic impacts and defeat the detriments of regular peptides, artificial hybrid peptides have been planned by joining the arrangement of at least two unique peptides with shifted biological function. Contrasted with common peptides, crossbreed peptides have indicated better helpful possibilities against microorganisms, tumors, and metabolic infections.(C. Wang et al., 2019)(Lam et al., 2019)

Applications of Computer-Aided Drug Design

There are two main applications of computer-aided drug design. Hit finding and Lead optimization.

Hit finding

A general way to deal with figure out how to perceive hit mixes is virtual screening. This method can show up diversely about a separating technique: beginning from a usually gigantic number of mixes, structure, ligand, or cross assortment approaches are utilized to pick a lessened number of particles. The working hypothesis is that the diminished number of mixes have stretched out probabilities to be dynamic. Exploratory

support of picked mixes is mandatory. After the test assessment has been facilitated a second round, or more, of disengaging progresses is performed. In the second or more, highlight of the exploratory data of the past cycle ought to be considered in the affirmation of the particles. Virtual screening is in a like way relevant to channel anticipated regular fixations for a given little molecule. The last philosophy is called invert virtual screening or target fishing.(Prieto-Martínez et al., 2019)(Eguida & Rognan, 2020)(Perez et al., 2020)

Lead Optimization

Various structures, ligand, or mixture found techniques can be utilized to better the strength or diminish the symptoms of dynamic atoms. Eminently, issues with assimilation, appropriation, digestion, and discharge properties may hamper the advancement of mixes in the facility. Different machinery is used to detect poisonousness and intensity.(Prieto-Martínez et al., 2019)(Mazanetz et al., 2019)(Pennington et al., 2019)

Using the catalyst for drug design

At the point when a pharmacophore is developed, an important usage of it is a 3D database hoping to recoup novel exasperates that would facilitate the pharmacophore, without in a general sense replicating the topological features of realized dynamic blends (along these lines remain liberated from existing licenses). As the 3D glancing through development has progressed consistently, it has been feasibly used for lead improvement, combinatorial library focusing, similarly as virtual high-throughput screening.(Kurogi & Guner, 2012)(Kotev et al., 2020)

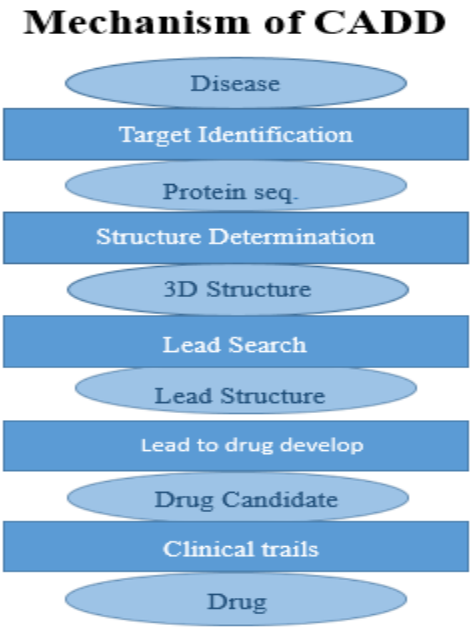


Figure 4 Flow chart of computer aided drug design mechanism

Applications in plant

A focal piece of the objective medication advancement process is the expectation of the intricate structure of a little ligand with a protein, the supposed protein-ligand docking issue, utilized in virtual screening of huge databases and lead improvement. In the work introduced here, we present another docking calculation called PLANTS (Protein-Ligand system), which depends on subterranean insect settlement streamlining. A fake subterranean insect state is utilized to locate a base vitality adaptation of the ligand in the protein's coupling site. We present the adequacy of PLANTS for a few boundary settings just as an immediate correlation with the best in a class program called GOLD, which depends on a hereditary calculation. To wrap things up, results for virtual screening on the protein target factor Xa are introduced. (Korb et al., 2006)(Houssein et al., 2020)

Pseudoreceptor models in drug design:

Pseudoreceptor assumes a job in tranquilize configuration (crossing over ligand-and receptor-based virtual screening). Sound medicine design relies upon unequivocal or undeniable structure-activity relationship models. Routinely, receptor-based or ligand-based frameworks are looked for after, dependent upon the information open about known ligands and the receptor structure. Pseudoreceptor models combine the advantages of these two methods and address a uniting thought for both receptor arranging and ligand planning. They can give an entry point to structure-based exhibiting in sedate divulgence expands that miss the mark on a significant standards structure of the target. Here, we review the field of pseudovector showing techniques close by a continuous hit and lead finding applications, and analyze fundamentals, central focuses, and controls of the various systems.(Mazanetz et al., 2019; G. Subramanian et al., 2020; Tanrikulu & Schneider, 2008)

Table 4: Computer-Aided Drug Design Approaches Based on 3D Structures

Methods	Aim	Example of Recent Approach
Molecular docking	CADD is a promising strategy	Docking suggested that angiotensin II receptor blockers could bind to an active site of kynurenine aminotransferase II. This mechanism action may be advantageous in the treatment of schizophrenia.
Pharmacophore	Approach to identify in 3D the elements required for the receptor-ligand recognition process. If the ligand is an agonist, the recognition	Construction of pharmacophore models of the Mycobacterium structural proteome.

	process can lead to the activation of the receptor upon binding. Pharmacophore-based design can be used to guide the chemical modifications to molecules to improve recognition with the receptor and enhance the biological activity. This can be useful to give some indication of the nature of the functional groups in the receptor responsible for binding to the set drugs.	
De-novo-design	Generation of new molecules with specific and desired properties. In this method it is necessary to use models of the molecular world to produce a trustworthy model that correctly reflects the real world, so it can be used for predicting new molecules that possess the target property reflected in the model.	Designed and tested mini-proteins of 37–43 residues that target influenza hemagglutinin and botulinum neurotoxin B.
Fragment based-screening	This approach enables to rapidly scan many molecular fragments that could have very specific interactions with cavities in a binding pocket. The fragments are later connected with linkers with the final goal of generating small molecules that should be synthetically feasible.	Recreation of liver tumors that could been avatars for high-throughput drug screening.

Bioinformatics Softwares Databases and Drug Discovery:

There are numerous products, databases and web administrations which have a connection with the disclosure of the medication and they are around partitioned into numerous classes. These medications depend on various viewpoints like portrayal of the synthetic structure, atomic mechanics and demonstrating, character model to know the cosmetics of a protein lead by a homologue of the structure that is known as of now, projection of the medication target, plan of the ligand, assessment of the coupling free vitality, tranquilize up-and-comer screening, QSAR and ADME toxicity. There are numerous kinds of programming which are free and simple to utilize so they are supported by numerous well known associations. These fuse databases, for instance, ChEMBL and SwissSidechain. These likewise incorporate programming gadgets, for instance, UCSF Chimera which isn't only a 3D discernment mechanical assembly yet also a phase for programming engineers propelled by assistant science. Swiss Similarity is utilized for virtual screening, SwissBioisostere for ligand plan, SwissSideChain to energize tests that broaden the protein assortment by introducing non-regular amino acids, and Swiss Dock for docking drug competitors for example little particles on proteins.(S0925400505006519 @ Www.Sciencedirect.Com, n.d.)

Successfull Applications of Cadd Grug:

CADD is a mix of a few hypothetical and computational orders, for example, sub-atomic demonstrating, chemo informatics, hypothetical science, among others. There are an wide scope of computational methodologies utilized in CADD that have been utilized for quite a while, for example, sub-atomic docking, elements, QSAR, similarity looking. Like other multidisciplinary approaches, CADD faces a few difficulties that incorporate the refinement of the hypothetical premise, yet in addition its sound application i.e. with the information on the restrictions(B978012816125800002X @ Sci-Hub.Tw, n.d.).High-throughput data, for instance, genomic, epigenetic, genome structure, cistromic, transcriptomic, proteomic, and ribosome profiling data have all made immense pledge to system based medication disclosure and medication repurposing. (Art00005 @ Www.Ingentaconnect.Com, n.d.).

Virtual Screening:

A computational field known as 'virtual screening' (VS) has developed in the previous decades to help test drug revelation concentrates by factually assessing obscure bio-cooperation among compounds and natural biological targets. These strategies utilize the physico-synthetic and auxiliary properties of compounds as well as target proteins alongside the tentatively or experimentally checked bio-collaboration data to produce prescient models(5062947 @ Academic.Oup.Com, n.d.).

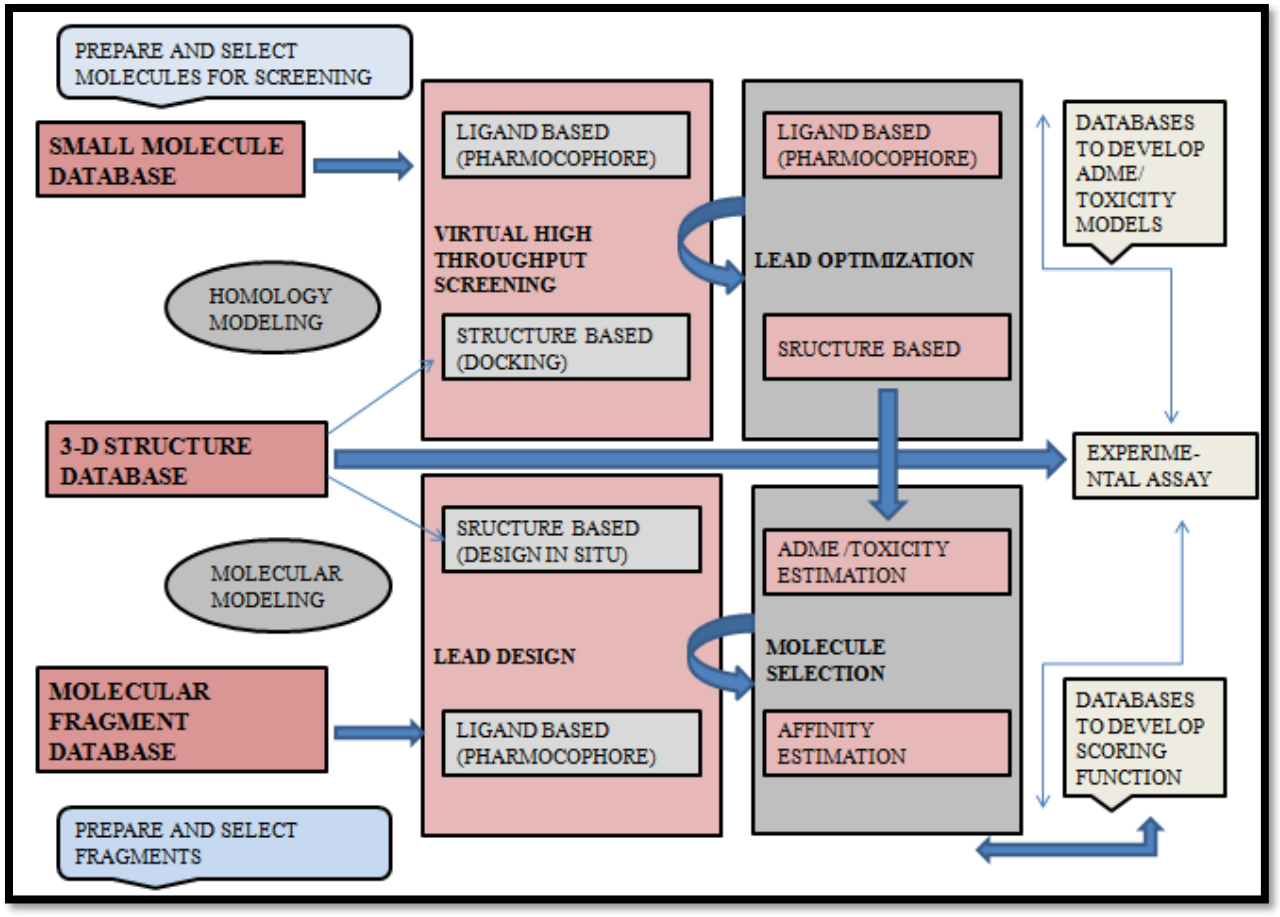


Figure 5 Computer Aided Drug Design Databases

Nonribosomal Antibiotic Peptide Genome Mining :

Bioinformatics gadgets are moreover used in the disclosure of new no ribosomal peptides (NRPs) made by microorganisms. It depends on the ID of the synthetase qualities and the unraveling of the domain engineering of the no ribosomal peptide synthetases (NRPSs). To survey the oddity, the two-dimensional structure of the peptides can be contrasted and the auxiliary patterns of all known NRPs.The investigation of such sequenced genomes may prompt the disclosure of new drugs(978-1-4939-3375-4_14 @ Link.Springer.Com, n.d.).THPdb is also a database of FDA-approved peptide and protein therapeutics. These therapeutic peptides have multiple essential advantages over proteins and antibodies(Katrijn Opstoel , Johan Pion , Marije Elferink-Gemser, Esther Hartman, Bas Willemse, Renaat Philippaerts, Chris Visscher, 2014).

Metabolomics and Transcriptomics:

Metabolomics can be characterized as the whole substance of metabolites in a framework and their jobs and associations in different metabolic pathways mirroring the hereditary data encoded in a genome (978-981-10-7483-7_10 @ Link.Springer.Com, n.d.). There are many sites that assemble tools for metabolomics such as PRIMe is a site that collects devices for metabolomics

and transcriptomics. (Raghavan et al., 2015).Bioinformatics tools also have a part in representation, examination, and understanding of metabolome unique fingerprint information for all out arrangement investigation and quality characteristic examination in potato cultivars.(Raghavan et al., 2015), MetaCyc is a multiorganism database of metabolic pathways and compounds(1133775 @ Doi.Org, n.d.).And furthermore, BRENDA, AMENDA and FRENDA are the compound information system which tells all the properties of the mixes to configuration sedate. (Raghavan et al., 2015). A huge number of infections of the total populace put need on understanding the ordinary functions of the various parts of the body at cell, tissue, and organ level. When the typical and normal functions are known, at that point the topic of tending to the management of the illness among the patient populace can be tended to at various levels from science to drug disclosure to better administration and avoidance(978-3-030-18375-2_1 @ Link.Springer.Com, n.d.).

Almost all the databases, softwares and websites which may have a vital role in drug discovery arelisted below,

DATABASES FOR DRUG DISCOVERY	ZincDatabase, Zinc15Database, ChEMBL, Bingo,JChemforExcel, ChemDiff, ProteinDataBank(PDB), BindingMOAD(MotherOfAllDatabase), PDBbind, TTD.
CHEMICAL STRUCTURE REPRESENTATION	ChemDraw, MarvinSketch, ACD/ChemSketch, jsMolEditor, Marvinmoleculeeditorandview , Ketcher, UCSFChimera, Pymol, OpenStructure, DaylightSMILES, InChI, TriposMol2, OpenBabel, Corina, Indigo, PoseView, PLiP.
MOLECULAR MODELING	CHARMM, GROMACS, Amber, SwissParam, CHARMM-GUI, CHARMMing.
HOMOLOGY MODELING	Modeller, I-TASSER, LOMETS, SWISS-MODEL, SWISS-MODEL Repository, Robetta.
BINDING SITE PREDICTION	MED-SuMo, TRAPP, CAVER, sc-PDB, CASTp, Pocketome, 3DLigandSite, metaPocket , Pock Drug.
DOCKING	Autodock, DOCK, GOLD, SwissDock, DockingServer, 1-ClickDocking.
SCREENING	Pharmer, Catalyst, PharmaGist, SwissSimilarity, Blaster, AnchorQuery.
TARGET PREDICTION	PatchSearch, IXCHEL, CABRAKAN, SwissTargetPrediction, SEA, CSNAP.
LIGAND DESIGN	GANDI, LUDI, BREED, SwissBioisostere, VAMMPIRE, sc-PDB-Frag, e-LEA3D, eDesig , iScreen.

ADME TOXICITY	Prop, VolSurf, GastroPlus, ALOGPS, OSIRISPropertyExplorer, SwissADME, PACT-F, TOXQikNET, LeadscopeToxicityDatabase.

Challenges:

The joining of information from numerous sources is progressively testing a result of expanding volume of information as well as on the grounds that this information are exceptionally different, especially in a medication revelation venture that intrinsically draws on numerous divergent kinds of data. Elevated level data curation/reconciliation and novel computational methodologies have been grown, yet more are required for strong and exact polypharmacology prediction. More trial measures and models are in incredible interest to consider the multi-focusing on and different targeting impacts of mixes for various illnesses.

Author Contributions:

All the authors contributed equally.

Funding:

Not applicable.

Institutional Review Board Statement:

Not applicable.

Informed Consent Statement:

Not applicable.

Conflicts of Interest:

The authors declare no conflict of interest.

References:

Alcântara, L. M., Ferreira, T. C. S., Fontana, V., Chatelain, E., Moraes, C. B., & Freitas-Junior, L. H. (2020). A Multi-Species Phenotypic Screening Assay for Leishmaniasis Drug Discovery Shows That Active Compounds Display a High Degree of Species-Specificity. Molecules, 25(11), 2551. https://doi.org/10.3390/molecules25112551

Aliyu Musa, Laleh Soltan Ghoraie, Shu-Dong Zhang, Galina Glazko, Olli Yli-Harja, Matthias Dehmer, Benjamin Haibe-Kains, & Frank Emmert-Streib. (2018). A review of connectivity map and computational approaches in pharmacogenomics. Briefings in Bioinformatics, 19(3), 506–523. https://doi.org/10.1093/BIB

Art00005 @ Www.Ingentaconnect.Com. (n.d.).

Arvaniti, E., Moulos, P., Vakrakou, A., Chatziantoniou, C., Chadjichristos, C., Kavvadas, P., Charonis, A., & Politis, P. K. (2016). Whole-transcriptome analysis of UUO mouse model of renal fibrosis reveals new molecular players in kidney diseases. Scientific Reports, 6. https://doi.org/10.1038/srep26235

B978012816125800002X @ sci-hub.tw. (n.d.).

Bao, J., & Bedford, M. T. (2016). Epigenetic regulation of the histone-to-protamine transition during spermiogenesis. Reproduction, 151(5), R55–R70. https://doi.org/10.1530/REP-15-0562

Berger, M. F., Levin, J. Z., Vijayendran, K., Sivachenko, A., Adiconis, X., Maguire, J., Johnson, L. A., Robinson, J., Verhaak, R. G.,

Sougnez, C., Onofrio, R. C., Ziaugra, L., Cibulskis, K., Laine, E., Barretina, J., Winckler, W., Fisher, D. E., Getz, G., Meyerson, M., … Garraway, L. A. (2010). Integrative analysis of the melanoma transcriptome. Genome Research, 20(4), 413–427. https://doi.org/10.1101/gr.103697.109

Bhatia, B., Ponia, S. S., Solanki, A. K., Dixit, A., & Garg, L. C. (2014). Identification of glutamate ABC-Transporter component in Clostridium perfringens as a putative drug target. Bioinformation, 10(7), 401–405. https://doi.org/10.6026/97320630010401

Bibikova, M., Barnes, B., Tsan, C., Ho, V., Klotzle, B., Le, J. M., Delano, D., Zhang, L., Schroth, G. P., Gunderson, K. L., Fan, J. B., & Shen, R. (2011). High density DNA methylation array with single CpG site resolution. Genomics, 98(4), 288–295. https://doi.org/10.1016/j.ygeno.2011.07.007

Bitsika, V., Duveau, V., Simon-Areces, J., Mullen, W., Roucard, C., Makridakis, M., Mermelekas, G., Savvopoulos, P., Depaulis, A., & Vlahou, A. (2016). High-throughput LC-MS/MS proteomic analysis of a mouse model of mesiotemporal lobe epilepsy predicts microglial activation underlying disease development. Journal of Proteome Research, 15(5), 1546–1562. https://doi.org/10.1021/acs.jproteome.6b00003

Boehringer, D., Thermann, R., Ostareck-Lederer, A., Lewis, J. D., & Stark, H. (2005). Structure of the hepatitis C virus IRES bound to the human 80S ribosome: Remodeling of the HCV IRES. Structure, 13(11), 1695–1706. https://doi.org/10.1016/j.str.2005.08.008

Bolden, J. E., Peart, M. J., & Johnstone, R. W. (2006). Anticancer activities of histone deacetylase inhibitors. Nature Reviews Drug Discovery, 5(9), 769–784. https://doi.org/10.1038/nrd2133

Boxall, A. B. A., Rudd, M. A., Brooks, B. W., Caldwell, D. J., Choi, K., Hickmann, S., Innes, E., Ostapyk, K., Staveley, J. P., Verslycke, T., Ankley, G. T., Beazley, K. F., Belanger, S. E., Berninger, J. P., Carriquiriborde, P., Coors, A., DeLeo, P. C., Dyer, S. D., Ericson, J. F., … Van Der Kraak, G. (2012). Pharmaceuticals and personal care products in the environment: What are the big questions? In Environmental Health Perspectives (Vol. 120, Issue 9, pp. 1221–1229). Public Health Services, US Dept of Health and Human Services. https://doi.org/10.1289/ehp.1104477

Brown, D. K., & Tastan Bishop, Ö. (2017). Role of Structural Bioinformatics in Drug Discovery by Computational SNP Analysis: Analyzing Variation at the Protein Level. In Global Heart (Vol. 12, Issue 2, pp. 151–161). Elsevier B.V. https://doi.org/10.1016/j.gheart.2017.01.009

Brücher, B. L. D. M., & Jamall, I. S. (2016). Somatic mutation theory - Why it's wrong for most cancers. In Cellular Physiology and Biochemistry (Vol. 38, Issue 5, pp. 1663–1680). S. Karger AG. https://doi.org/10.1159/000443106

Chu, C., Qu, K., Zhong, F. L., Artandi, S. E., & Chang, H. Y. (2011). Genomic Maps of Long Noncoding RNA Occupancy Reveal Principles of RNA-Chromatin Interactions. Molecular Cell, 44(4), 667–678. https://doi.org/10.1016/j.molcel.2011.08.027

Clark, A. T. (2015). DNA methylation remodeling in vitro and in vivo. Current Opinion in Genetics and Development, 34, 82–87. https://doi.org/10.1016/j.gde.2015.09.002

Clough, E., & Barrett, T. (2016). The Gene Expression Omnibus database. In Methods in Molecular Biology (Vol. 1418, pp. 93–110). Humana Press Inc. https://doi.org/10.1007/978-1-4939-3578-9_5

Congreve, M., Murray, C. W., & Blundell, T. L. (2005). Keynote review: Structural biology and drug discovery. In Drug Discovery Today (Vol. 10, Issue 13, pp. 895–907). Elsevier Current Trends. https://doi.org/10.1016/S1359-6446(05)03484-7

Daily, K., Patel, V. R., Rigor, P., Xie, X., & Baldi, P. (2011). MotifMap: Integrative genome-wide maps of regulatory motif sites for model species. BMC Bioinformatics, 12(1). https://doi.org/10.1186/1471-2105-12-495

Davies, J. (1996). Origins and evolution of antibiotic resistance. In Microbiología (Madrid, Spain) (Vol. 12, Issue 1, pp. 9–16). Microbiol Mol Biol Rev. https://doi.org/10.1128/mmbr.00016-10

Deng, W., Lee, J., Wang, H., Miller, J., Reik, A., Gregory, P. D., Dean, A., & Blobel, G. A. (2012). Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. Cell, 149(6), 1233–1244. https://doi.org/10.1016/j.cell.2012.03.051

Deng, W., Rupon, J. W., Krivega, I., Breda, L., Motta, I., Jahn, K. S., Reik, A., Gregory, P. D., Rivella, S., Dean, A., & Blobel, G. A. (2014). Reactivation of developmentally silenced globin genes by forced chromatin looping. Cell, 158(4), 849–860. https://doi.org/10.1016/j.cell.2014.05.050

Doolittle, R. F., Hunkapiller, M. W., Hood, L. E., Devare, S. G., Robbins, K. C., Aaronson, S. A., & Antoniades, H. N. (1983). Simian sarcoma virus onc gene, v-sis, is derived from the gene (or genes) encoding a platelet-derived growth factor. Science, 221(4607), 275–277. https://doi.org/10.1126/science.6304883

Drews, J., & Ryser, S. (1997). The role of innovation in drug development. In Nature Biotechnology (Vol. 15, Issue 13, pp. 1318–1319). Nat Biotechnol. https://doi.org/10.1038/nbt1297-1318

Ebhohimen, I. E., Edemhanria, L., Awojide, S., Onyijen, O. H., & Anywar, G. (2020). Advances in computer-aided drug discovery. In Phytochemicals as Lead Compounds for New Drug Discovery (pp. 25–37). Elsevier. https://doi.org/10.1016/b978-0-12-817890-4.00003-2

Eckhardt, F., Lewin, J., Cortese, R., Rakyan, V. K., Attwood, J., Burger, M., Burton, J., Cox, T. V., Davies, R., Down, T. A., Haefliger, C., Horton, R., Howe, K., Jackson, D. K., Kunde, J., Koenig, C., Liddle, J., Niblett, D., Otto, T., … Beck, S. (2006). DNA methylation profiling of human chromosomes 6, 20 and 22. Nature Genetics, 38(12), 1378–1385. https://doi.org/10.1038/ng1909

Eguida, M., & Rognan, D. (2020). A Computer Vision Approach to Align and Compare Protein Cavities: Application to Fragment-Based Drug Design. Journal of Medicinal Chemistry.

https://doi.org/10.1021/acs.jmedchem.0c00422

El-Osta, A. (2003). DNMT cooperativity?the developing links between methylation, chromatin structure and cancer. BioEssays, 25(11), 1071–1084. https://doi.org/10.1002/bies.10345

Farha, M. A., & Brown, E. D. (2019). Drug repurposing for antimicrobial discovery. In Nature Microbiology (Vol. 4, Issue 4, pp. 565–577). Nature Publishing Group. https://doi.org/10.1038/s41564-019-0357-1

Fatemi, M., Hermann, A., Pradhan, S., & Jeltsch, A. (2001). The activity of the murine DNA methyltransferase Dnmt1 is controlled by interaction of the catalytic domain with the N-terminal part of the enzyme leading to an allosteric activation of the enzyme after binding to methylated DNA. Journal of Molecular Biology, 309(5), 1189–1199. https://doi.org/10.1006/jmbi.2001.4709

Gabrielsson, J., & Green, A. R. (2009). Quantitative pharmacology or pharmacokinetic pharmacodynamic integration should be a vital component in integrative pharmacology. Journal of Pharmacology and Experimental Therapeutics, 331(3), 767–774. https://doi.org/10.1124/jpet.109.157172

Ganter, B., Snyder, R. D., Halbert, D. N., & Lee, M. D. (2006). Toxicogenomics in drug discovery and development: Mechanistic analysis of compound/class-dependent effects using the DrugMatrix® database. Pharmacogenomics, 7(7), 1025–1044. https://doi.org/10.2217/14622416.7.7.1025

Gapp, K., Jawaid, A., Sarkies, P., Bohacek, J., Pelczar, P., Prados, J., Farinelli, L., Miska, E., & Mansuy, I. M. (2014). Implication of sperm RNAs in transgenerational inheritance of the effects of early trauma in mice. Nature Neuroscience, 17(5), 667–669. https://doi.org/10.1038/nn.3695

Garg, N. J., Soman, K. V., Zago, M. P., Koo, S. J., Spratt, H., Stafford, S., Blell, Z. N., Gupta, S., Nuñez Burgos, J., Barrientos, N., Brasier, A. R., & Wiktorowicz, J. E. (2016). Changes in Proteome Profile of Peripheral Blood Mononuclear Cells in Chronic Chagas Disease. PLoS Neglected Tropical Diseases, 10(2). https://doi.org/10.1371/journal.pntd.0004490

Goldberg, A. D., Allis, C. D., & Bernstein, E. (2007). Epigenetics: A Landscape Takes Shape. Cell, 128(4), 635–638. https://doi.org/10.1016/j.cell.2007.02.006

Grigg, G., & Clark, S. (1994). Genes and genomes: Sequencing 5-methylcytosine residues in genomic DNA. BioEssays, 16(6), 431–436. https://doi.org/10.1002/bies.950160612

Grigg, G. W. (1996). Sequencing 5-methylcytosine residues by the bisulphite method. DNA Sequence, 6(4), 189–198. https://doi.org/10.3109/10425179609008443

Heath, J. R., Ribas, A., & Mischel, P. S. (2016). Single-cell analysis tools for drug discovery and development. In Nature Reviews Drug Discovery (Vol. 15, Issue 3, pp. 204–216). Nature Publishing Group. https://doi.org/10.1038/nrd.2015.16

Hernandez, H. W., Soeung, M., Zorn, K. M., Ashoura, N., Mottin, M., Andrade, C. H.,

Caffrey, C. R., de Siqueira-Neto, J. L., & Ekins, S. (2019). High Throughput and Computational Repurposing for Neglected Diseases. In Pharmaceutical Research (Vol. 36, Issue 2). Springer New York LLC. https://doi.org/10.1007/s11095-018-2558-3

Holford, N. (2011). Holford NHG and Sheiner LB "understanding the dose-effect relationship-clinical application of pharmacokinetic-pharmacodynamic models", Clin Pharmacokin 6:429-453 (1981) - The backstory. AAPS Journal, 13(4), 662–664. https://doi.org/10.1208/s12248-011-9306-5

Houssein, E. H., Hosney, M. E., Oliva, D., Mohamed, W. M., & Hassaballah, M. (2020). A novel hybrid Harris hawks optimization and support vector machines for drug design and discovery. Computers and Chemical Engineering, 133, 106656. https://doi.org/10.1016/j.compchemeng.2019.106656

Huang, S., Litt, M. D., & Blakey, C. A. (2015). Epigenetic gene expression and regulation. Academic Press.

Igarashi, Y., Nakatsu, N., Yamashita, T., Ono, A., Ohno, Y., Urushidani, T., & Yamada, H. (2014). Open TG-GATEs: a large-scale toxicogenomics database. Nucleic Acids Research, 43(D1), D921–D927. https://doi.org/10.1093/nar/gku955

Ingolia, N. T., Brar, G. A., Stern-Ginossar, N., Harris, M. S., Talhouarne, G. J. S., Jackson, S. E., Wills, M. R., & Weissman, J. S. (2014). Ribosome Profiling Reveals Pervasive Translation Outside of Annotated Protein-Coding Genes. Cell Reports, 8(5), 1365–1379. https://doi.org/10.1016/j.celrep.2014.07.045

Insinga, A., Minucci, S., & Pelicci, P. G. (2005). Mechanisms of selective anticancer action of histone deacetylase inhibitors. Cell Cycle, 4(6), 741–743. https://doi.org/10.4161/cc.4.6.1717

Insinga, A., Monestiroli, S., Ronzoni, S., Gelmetti, V., Marchesi, F., Viale, A., Altucci, L., Nervi, C., Minucci, S., & Pelicci, P. G. (2005). Inhibitors of histone deacetylases induce tumor-selective apoptosis through activation of the death receptor pathway. Nature Medicine, 11(1), 71–76. https://doi.org/10.1038/nm1160

Jackson, R. J., Hellen, C. U. T., & Pestova, T. V. (2010). The mechanism of eukaryotic translation initiation and principles of its regulation. In Nature Reviews Molecular Cell Biology (Vol. 11, Issue 2, pp. 113–127). Nat Rev Mol Cell Biol. https://doi.org/10.1038/nrm2838

Jasper J. Koehorst, Jesse C. J. van Dam, Edoardo Saccenti, Vitor A. P. Martins dos Santos, M. S.-D. and P. J. S. (2017). GISAID Global Initiative on Sharing All Influenza Data. Phylogeny of SARS-like betacoronaviruses including novel coronavirus (nCoV). Oxford, 34(8), 1401–1403. https://doi.org/10.1093/BIOINFORMATICS

Jiang, Y., Bressler, J., & Beaudet, A. L. (2004). EPIGENETICS AND HUMAN DISEASE. Annual Review of Genomics and Human Genetics, 5(1), 479–510. https://doi.org/10.1146/annurev.genom.5.061903.180014

Jin, B., & Robertson, K. D. (2013). DNA Methyltransferases, DNA Damage Repair, and Cancer. In Advances in experimental medicine and biology (Vol. 754, pp. 3–29).

NIH Public Access. https://doi.org/10.1007/978-1-4419-9967-2_1

Jin, P., Alisch, R. S., & Warren, S. T. (2004). RNA and microRNAs in fragile X mental retardation. Nature Cell Biology, 6(11), 1048–1053. https://doi.org/10.1038/ncb1104-1048

Katrijn Opstoel , Johan Pion , Marije Elferink-Gemser, Esther Hartman, Bas Willemse, Renaat Philippaerts, Chris Visscher, M. L. (2014). Article @ Journals.Plos.Org. In Plos One (Vol. 9, Issue 4).

Kinaret, P. A. S., Serra, A., Federico, A., Kohonen, P., Nymark, P., Liampa, I., Ha, M. K., Choi, J. S., Jagiello, K., Sanabria, N., Melagraki, G., Cattelani, L., Fratello, M., Sarimveis, H., Afantitis, A., Yoon, T. H., Gulumian, M., Grafström, R., Puzyn, T., & Greco, D. (2020). Transcriptomics in toxicogenomics, part i: Experimental design, technologies, publicly available data, and regulatory aspects. Nanomaterials, 10(4), 750. https://doi.org/10.3390/nano10040750

Kioussis, D., Vanin, E., Delange, T., Flavell, R. A., & Grosveld, F. G. (1983). β-Globin gene inactivation by DNA translocation in γβ-thalassaemi. Nature, 306(5944), 662–666. https://doi.org/10.1038/306662a0

Kolesnikov, N., Hastings, E., Keays, M., Melnichuk, O., Tang, Y. A., Williams, E., Dylag, M., Kurbatova, N., Brandizi, M., Burdett, T., Megy, K., Pilicheva, E., Rustici, G., Tikhonov, A., Parkinson, H., Petryszak, R., Sarkans, U., & Brazma, A. (2015). ArrayExpress update--simplifying data submissions. Nucleic Acids Research, 43(Database issue), D1113-6. https://doi.org/10.1093/nar/gku1057

Korb, O., Stützle, T., & Exner, T. E. (2006). PLANTS: Application of ant colony optimization to structure-based drug design. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 4150 LNCS, 247–258. https://doi.org/10.1007/11839088_22

Korenke, G. C., Wilichowski, E., Hunneman, D. H., Hanefeld, F., Fuchs, S., Krasemann, E., & Doerr, H. G. (1996). Cerebral adrenoleukodystrophy (ALD) in only one of monozygotic twins with an identical ALD genotype. Annals of Neurology, 40(2), 254–257. https://doi.org/10.1002/ana.410400221

Kotev, M., Sarrat, L., & Gonzalez, C. D. (2020). User-Friendly Quantum Mechanics: Applications for Drug Discovery. In Methods in Molecular Biology (Vol. 2114, pp. 231–255). Humana Press Inc. https://doi.org/10.1007/978-1-0716-0282-9_15

Kungulovski, G., & Jeltsch, A. (2016). Epigenome Editing: State of the Art, Concepts, and Perspectives. Trends in Genetics, 32(2), 101–113. https://doi.org/10.1016/j.tig.2015.12.001

Kurogi, Y., & Guner, O. (2012). Pharmacophore Modeling and Three-dimensional Database Searching for Drug Design Using Catalyst. Current Medicinal Chemistry, 8(9), 1035–1055. https://doi.org/10.2174/0929867013372481

Lakshmi, P., & Ramyachitra, D. (2020). Review About Bioinformatics, Databases, Sequence Alignment, Docking, and Drug Discovery

(pp. 11–23). Springer, Singapore. https://doi.org/10.1007/978-981-15-2445-5_2

Lam, P. C. H., Abagyan, R., & Totrov, M. (2019). Hybrid receptor structure/ligand-based docking and activity prediction in ICM: development and evaluation in D3R Grand Challenge 3. Journal of Computer-Aided Molecular Design, 33(1), 35–46. https://doi.org/10.1007/s10822-018-0139-5

Lamb, J. (2007). The Connectivity Map: A new tool for biomedical research. Nature Reviews Cancer, 7(1), 54–60. https://doi.org/10.1038/nrc2044

Lea, I. A., Gong, H., Paleja, A., Rashid, A., & Fostel, J. (2017). CEBS: a comprehensive annotated database of toxicological data. Nucleic Acids Research, 45(D1), D964–D971. https://doi.org/10.1093/nar/gkw1077

Li, K., Du, Y., Li, L., & Wei, D.-Q. (2019). Bioinformatics Approaches for Anti-cancer Drug Discovery. Current Drug Targets, 21(1), 3–17. https://doi.org/10.2174/1389450120666190923162203

Li, M. J., Yan, B., Sham, P. C., & Wang, J. (2015). Exploring the function of genetic variants in the non-coding genomic regions: approaches for identifying human regulatory variants affecting gene expression. Briefings in Bioinformatics, 16(3), 393–412. https://doi.org/10.1093/bib/bbu018

Lieberman-Aiden, E., Van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B. R., Sabo, P. J., Dorschner, M. O., Sandstrom, R., Bernstein, B., Bender, M. A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., Mirny, L. A., Lander, E. S., & Dekker, J. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science, 326(5950), 289–293. https://doi.org/10.1126/science.1181369

Lill, M. A., & Danielson, M. L. (2011). Computer-aided drug design platform using PyMOL. Journal of Computer-Aided Molecular Design, 25(1), 13–19. https://doi.org/10.1007/s10822-010-9395-8

Loging, W., Harland, L., & Williams-Jones, B. (2007). High-throughput electronic biology: Mining information for drug discovery. In Nature Reviews Drug Discovery (Vol. 6, Issue 3, pp. 220–230). Nature Publishing Group. https://doi.org/10.1038/nrd2265

Ma, P., & Xia, X. (2011). Factors affecting splicing strength of yeast genes. Comparative and Functional Genomics, 2011.

Mazanetz, M. P., Goode, C. H. F., & Chudyk, E. I. (2019). Ligand- and Structure-Based Drug Design and Optimization using KNIME. Current Medicinal Chemistry, 26. https://doi.org/10.2174/0929867326666190409141016

Miller, L. H., & Su, X. (2011). Artemisinin: Discovery from the Chinese herbal garden. Cell, 146(6), 855–858. https://doi.org/10.1016/j.cell.2011.08.024

Moffat, J. G., Rudolph, J., & Bailey, D. (2014). Phenotypic screening in cancer drug discovery-past, present and future. Nature Reviews Drug Discovery, 13(8), 588–602. https://doi.org/10.1038/nrd4366

Mohammed, A., & Alghamdi, O. A. (2020).

Phenotype screenings of drugs for combination therapy against multidrug resistance. In Combination Therapy Against Multidrug Resistance (pp. 207–220). Elsevier. https://doi.org/10.1016/b978-0-12-820576-1.00011-4

MotifMap: A Human Genome-Wide Map of Candidate Regulatory Motif Sites - PubMed. (n.d.).

Muller, H. J., & Altenburg, E. (1930). The Frequency of Translocations Produced by X-Rays in Drosophila. Genetics, 15(4), 283–311.

nrd1608 @ www.nature.com. (n.d.).

Ordovás, J. M., & Smith, C. E. (2010). Epigenetics and cardiovascular disease. Nature Reviews Cardiology, 7(9), 510–519. https://doi.org/10.1038/nrcardio.2010.104

Ow, G. S., Ivshina, A. V., Fuentes, G., & Kuznetsov, V. A. (2014). Identification of two poorly prognosed ovarian carcinoma subtypes associated with CHEK2 germline mutation and non-CHEK2 somatic mutation gene signatures. Cell Cycle, 13(14), 2262–2280. https://doi.org/10.4161/cc.29271

Palstra, R. J., Tolhuis, B., Splinter, E., Nijmeijer, R., Grosveld, F., & De Laat, W. (2003). The β-globin nuclear compartment in development and erythroid differentiation. Nature Genetics, 35(2), 190–194. https://doi.org/10.1038/ng1244

Pandey, R. R., Mondal, T., Mohammad, F., Enroth, S., Redrup, L., Komorowski, J., Nagano, T., Mancini-DiNardo, D., & Kanduri, C. (2008). Kcnq1ot1 Antisense Noncoding RNA Mediates Lineage-Specific Transcriptional Silencing through Chromatin-Level Regulation. Molecular Cell, 32(2), 232–246. https://doi.org/10.1016/j.molcel.2008.08.022

Pandurangan, A. P., Ascher, D. B., Thomas, S. E., & Blundell, T. L. (2017). Genomes, structural biology and drug discovery: Combating the impacts of mutations in genetic disease and antibiotic resistance. In Biochemical Society Transactions (Vol. 45, Issue 2, pp. 303–311). Portland Press Ltd. https://doi.org/10.1042/BST20160422

Pazin, M. J., Hermann, J. W., & Kadonaga, J. T. (1998). Promoter Structure and Transcriptional Activation with Chromatin Templates Assembled In Vitro. Journal of Biological Chemistry, 273(51), 34653–34660. https://doi.org/10.1074/jbc.273.51.34653

Pazin, M. J., Kamakaka, R. T., & Kadonaga, J. T. (1994). ATP-dependent nucleosome reconfiguration and transcriptional activation from preassembled chromatin templates. Science, 266(5193), 2007–2011. https://doi.org/10.1126/science.7801129

Pazin, M. J., Sheridan, P. L., Cannon, K., Cao, Z., Keck, J. G., Kadonaga, J. T., & Jones, K. A. (1996). NF-κB-mediated chromatin reconfiguration and transcriptional activation of the HIV-1 enhancer in vitro. Genes and Development, 10(1), 37–49. https://doi.org/10.1101/gad.10.1.37

Pennington, L. D., Aquila, B. M., Choi, Y., Valiulin, R. A., & Muegge, I. (2019). Positional Analogue Scanning: An Effective Strategy for Multiparameter Optimization in Drug Design. In Journal of Medicinal Chemistry. American Chemical Society. https://doi.org/10.1021/acs.jmedchem.9b020

92

Perez, C., Soler, D., Soliva, R., & Guallar, V. (2020). FragPELE: Dynamic Ligand Growing within a Binding Site. A Novel Tool for Hit-To-Lead Drug Design. Journal of Chemical Information and Modeling. https://doi.org/10.1021/acs.jcim.9b00938

Petronis, A. (2004). The origin of schizophrenia: Genetic thesis, epigenetic antithesis, and resolving synthesis. Biological Psychiatry, 55(10), 965–970. https://doi.org/10.1016/j.biopsych.2004.02.005

Petronis, A. (2006). Epigenetics and twins: three variations on the theme. Trends in Genetics, 22(7), 347–350. https://doi.org/10.1016/j.tig.2006.04.010

Petronis, A., Gottesman, I. I., Kan, P., Kennedy, J. L., Basile, V. S., Paterson, A. D., & Popendikyte, V. (2003). Monozygotic twins exhibit numerous epigenetic differences: Clues to twin discordance? Schizophrenia Bulletin, 29(1), 169–178. https://doi.org/10.1093/oxfordjournals.schbul.a006988

Poroikov, V. V. (2020). Computer-aided drug design: From discovery of novel pharmaceutical agents to systems pharmacology. Biomeditsinskaya Khimiya, 66(1), 30–41. https://doi.org/10.18097/PBMC20206601030

Prabhakaran, R., Chithambaram, S., & Xia, X. (2015). Escherichia coli and Staphylococcus phages: Effect of translation initiation efficiency on differential codon adaptation mediated by virulent and temperate lifestyles. Journal of General Virology, 96(5), 1169–1179. https://doi.org/10.1099/vir.0.000050

Prieto-Martínez, F. D., López-López, E., Eurídice Juárez-Mercado, K., & Medina-Franco, J. L. (2019). Computational Drug Design Methods—Current and Future Perspectives. In In Silico Drug Design (pp. 19–44). Elsevier. https://doi.org/10.1016/b978-0-12-816125-8.00002-x

Raghavan, M., Steinrücken, M., Harris, K., Schiffels, S., Rasmussen, S., DeGiorgio, M., Albrechtsen, A., Valdiosera, C., Ávila-Arcos, M. C., Malaspinas, A.-S., Eriksson, A., Moltke, I., Metspalu, M., Homburger, J. R., Wall, J., Cornejo, O. E., Moreno-Mayar, J. V., Korneliussen, T. S., Pierre, T., … Tumen, D. (2015). Scholar_Lookup @ Scholar.Google.Com. In Science (Vol. 349, Issue 6250, p. aab3884). https://doi.org/10.1126/science.aab3884

Ren, W., Fan, H., Grimm, S. A., Guo, Y., Kim, J. J., Li, L., Petell, C. J., Tan, X.-F., Zhang, Z.-M., Coan, J. P., Yin, J., Gao, L., Cai, L., Detrick, B., Çetin, B., Patel, D. J., Wang, Y., Cui, Q., Strahl, B. D., … Song, J. (2020). Direct readout of heterochromatic H3K9me3 regulates DNMT1-mediated maintenance DNA methylation. BioRxiv, 2020.04.27.064493. https://doi.org/10.1101/2020.04.27.064493

Rinn, J. L., Kertesz, M., Wang, J. K., Squazzo, S. L., Xu, X., Brugmann, S. A., Goodnough, L. H., Helms, J. A., Farnham, P. J., Segal, E., & Chang, H. Y. (2007). Functional Demarcation of Active and Silent Chromatin Domains in Human HOX Loci by Noncoding RNAs. Cell, 129(7), 1311–1323. https://doi.org/10.1016/j.cell.2007.05.022

Robertson, G., Hirst, M., Bainbridge, M., Bilenky, M., Zhao, Y., Zeng, T., Euskirchen, G., Bernier, B., Varhol, R., Delaney, A., Thiessen, N., Griffith, O. L., He, A., Marra, M., Snyder, M., & Jones, S. (2007). Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. Nature Methods, 4(8), 651–657. https://doi.org/10.1038/nmeth1068

Rodgers, A. B., Morgan, C. P., Leu, N. A., & Bale, T. L. (2015). Transgenerational epigenetic programming via sperm microRNA recapitulates effects of paternal stress. Proceedings of the National Academy of Sciences of the United States of America, 112(44), 13699–13704. https://doi.org/10.1073/pnas.1508347112

S0925400505006519 @ www.sciencedirect.com. (n.d.).

S1359644619303897 @ www.sciencedirect.com. (n.d.).

Scapin, G. (2002). Structural biology in drug design: Selective protein kinase inhibitors. In Drug Discovery Today (Vol. 7, Issue 11, pp. 601–611). Elsevier Current Trends. https://doi.org/10.1016/S1359-6446(02)02290-0

Scapin, G. (2006). Structural Biology and Drug Discovery. Current Pharmaceutical Design, 12(17), 2087–2097. https://doi.org/10.2174/138161206777585201

Schüler, M., Connell, S. R., Lescoute, A., Giesebrecht, J., Dabrowski, M., Schroeer, B., Mielke, T., Penczek, P. A., Westhof, E., & Spahn, C. M. T. (2006). Structure of the ribosome-bound cricket paralysis virus IRES RNA. Nature Structural and Molecular Biology, 13(12), 1092–1096. https://doi.org/10.1038/nsmb1177

Sendler, E., Johnson, G. D., Mao, S., Goodrich, R. J., Diamond, M. P., Hauser, R., & Krawetz, S. A. (2013). Stability, delivery and functions of human sperm RNAs at fertilization. Nucleic Acids Research, 41(7), 4104–4117. https://doi.org/10.1093/nar/gkt132

Smircich, P., Eastman, G., Bispo, S., Duhagon, M. A., Guerra-Slompo, E. P., Garat, B., Goldenberg, S., Munroe, D. J., Dallagiovanna, B., Holetz, F., & Sotelo-Silveira, J. R. (2015). Ribosome profiling reveals translation control as a key mechanism generating differential gene expression in Trypanosoma cruzi. BMC Genomics, 16(1). https://doi.org/10.1186/s12864-015-1563-8

Song, J., Teplova, M., Ishibe-Murakami, S., & Patel, D. J. (2012). Structure-based mechanistic insights into DNMT1-mediated maintenance DNA methylation. Science, 335(6069), 709–712. https://doi.org/10.1126/science.1214453

Stout, T., Foster, P., & Matthews, D. (2005). High-Throughput Structural Biology in Drug Discovery: Protein Kinases. Current Pharmaceutical Design, 10(10), 1069–1082. https://doi.org/10.2174/1381612043452695

Subramanian, A., Narayan, R., Corsello, S. M., Peck, D. D., Natoli, T. E., Lu, X., Gould, J., Davis, J. F., Tubelli, A. A., Asiedu, J. K., Lahr, D. L., Hirschman, J. E., Liu, Z., Donahue, M., Julian, B., Khan, M., Wadden, D., Smith, I. C., Lam, D., … Golub, T. R. (2017). A Next Generation Connectivity

Map: L1000 Platform and the First 1,000,000 Profiles. Cell, 171(6), 1437-1452.e17. https://doi.org/10.1016/j.cell.2017.10.049

Subramanian, G., Vairagoundar, R., Bowen, S. J., Roush, N., Zachary, T., Javens, C., Williams, T., Janssen, A., & Gonzales, A. (2020). Synthetic inhibitor leads of human tropomyosin receptor kinase A (: H TrkA). RSC Medicinal Chemistry, 11(3), 370–377. https://doi.org/10.1039/c9md00554d

Swinney, D. C. (2013). Phenotypic vs. Target-based drug discovery for first-in-class medicines. Clinical Pharmacology and Therapeutics, 93(4), 299–301. https://doi.org/10.1038/clpt.2012.236

Swinney, David C., & Anthony, J. (2011). How were new medicines discovered? Nature Reviews Drug Discovery, 10(7), 507–519. https://doi.org/10.1038/nrd3480

Tanrikulu, Y., & Schneider, G. (2008). Pseudoreceptor models in drug design: Bridging ligand- and receptor-based virtual screening. In Nature Reviews Drug Discovery (Vol. 7, Issue 8, pp. 667–677). Nature Publishing Group. https://doi.org/10.1038/nrd2615

Tolhuis, B., Palstra, R. J., Splinter, E., Grosveld, F., & De Laat, W. (2002). Looping and interaction between hypersensitive sites in the active β-globin locus. Molecular Cell, 10(6), 1453–1465. https://doi.org/10.1016/S1097-2765(02)00781-5

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., Pimentel, H., Salzberg, S. L., Rinn, J. L., & Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature Protocols, 7(3), 562–578. https://doi.org/10.1038/nprot.2012.016

Ujcikova, H., Vosahlikova, M., Roubalova, L., & Svoboda, P. (2016). Proteomic analysis of protein composition of rat forebrain cortex exposed to morphine for 10 days; comparison with animals exposed to morphine and subsequently nurtured for 20 days in the absence of this drug. Journal of Proteomics, 145, 11–23. https://doi.org/10.1016/j.jprot.2016.02.019

ul Ain, Q., Batool, M., & Choi, S. (2020). TLR4-targeting therapeutics: Structural basis and computer-aided drug discovery approaches. In Molecules (Vol. 25, Issue 3, p. 627). MDPI AG. https://doi.org/10.3390/molecules25030627

van Driel, M. A., & Brunner, H. G. (2006). Bioinformatics methods for identifying candidate disease genes. In Human Genomics (Vol. 2, Issue 6, pp. 429–432). BioMed Central Ltd. https://doi.org/10.1186/1479-7364-2-6-429

Voelter-Mahlknecht, S. (2016). Epigenetic associations in relation to cardiovascular prevention and therapeutics. Clinical Epigenetics, 8(1), 1–17. https://doi.org/10.1186/s13148-016-0170-0

Waddington, C. H. (1942). The epigenotype. Endeavour, 1, 18–20.

Wade, P. A., & Wolffe, A. P. (2001). Recognizing methylated DNA. Nature Structural Biology, 8(7), 575–577. https://doi.org/10.1038/89593

Wang, C., Yang, C., Chen, Y. chen, Ma, L., & Huang, K. (2019). Rational Design of Hybrid

Peptides: A Novel Drug Design Approach. Current Medical Science, 39(3), 349–355. https://doi.org/10.1007/s11596-019-2042-2

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: A revolutionary tool for transcriptomics. Nature Reviews Genetics, 10(1), 57–63. https://doi.org/10.1038/nrg2484

Warchal, S. J., Dawson, J. C., Shepherd, E., Munro, A. F., Hughes, R. E., Makda, A., & Carragher, N. O. (2020). High content phenotypic screening identifies serotonin receptor modulators with selective activity upon breast cancer cell cycle and cytokine signaling pathways. Bioorganic and Medicinal Chemistry, 28(1), 115209. https://doi.org/10.1016/j.bmc.2019.115209

Waters, M., Stasiewicz, S., Merrick, B. A., Tomer, K., Bushel, P., Paules, R., Stegman, N., Nehls, G., Yost, K. J., Johnson, C. H., Gustafson, S. F., Xirasagar, S., Xiao, N., Huang, C.-C., Boyer, P., Chan, D. D., Pan, Q., Gong, H., Taylor, J., … Fostel, J. (2008). CEBS--Chemical Effects in Biological Systems: a public data repository integrating study design and toxicity data with microarray and proteomics data. Nucleic Acids Research, 36(Database issue), D892-900. https://doi.org/10.1093/nar/gkm755

Wooller, S. K., Benstead-Hume, G., Chen, X., Ali, Y., & Pearl, F. M. G. (2017). Bioinformatics in translational drug discovery. In Bioscience Reports (Vol. 37, Issue 4). Portland Press Ltd. https://doi.org/10.1042/BSR20160180

Wu, J., Sieglaff, D. H., Gervin, J., & Xie, X. S. (2008). Discovering regulatory motifs in the Plasmodium genome using comparative genomics. Bioinformatics (Oxford, England), 24(17), 1843–1849. https://doi.org/10.1093/bioinformatics/btn348

Xia, X. (2014). A major controversy in codon-Anticodon adaptation resolved by a new codon usage index. Genetics, 199(2), 573–579. https://doi.org/10.1534/genetics.114.172106

Xia, X. (2016). PhyPA: Phylogenetic method with pairwise sequence alignment outperforms likelihood methods in phylogenetics involving highly diverged sequences. Molecular Phylogenetics and Evolution, 102, 331–343. https://doi.org/10.1016/j.ympev.2016.07.001

Xia, X. (2017a). The international journal for in-depth reviews on Current Topics in Medicinal Chemistry Impact Factor: 2.9. Current Topics in Medicinal Chemistry, 17, 15. https://doi.org/10.2174/1568026617666161116143440

Xia, X. (2017b). Bioinformatics and Drug Discovery. Current Topics in Medicinal Chemistry, 17(15), 1709–1726. https://doi.org/10.2174/1568026617666161116143440

Xia, X., Mackay, V., Yao, X., Wu, J., Miura, F., Ito, T., & Morris, D. R. (2011). Translation initiation: A regulatory role for poly(A) tracts in front of the AUG codon in saccharomyces cerevisiae. Genetics, 189(2), 469–478. https://doi.org/10.1534/genetics.111.132068

Zhang, W., Tan, A. Y., Blumenfeld, J., Liu, G., Michaeel, A., Zhang, T., Robinson, B. D., Salvatore, S. P., Kapur, S., Donahue, S., Bobb, W. O., & Rennert, H. (2016). Papillary

renal cell carcinoma with a somatic mutation in MET in a patient with autosomal dominant polycystic kidney disease. Cancer Genetics, 209(1–2), 11–20. https://doi.org/10.1016/j.cancergen.2015.11.002

Zoghbi, H. Y., & Beaudet, A. L. (2016). Epigenetics and human disease. Cold Spring Harbor Perspectives in Biology, 8(2), 1–28. https://doi.org/10.1101/cshperspect.a019497