Present Situation and Forecast of Bioinformatics in the Field of New Medicine Research and Development

Duan Yibing[1]

[1]School of Life Science, Jilin University, China

Abstract. In the last several centuries, biology has accumulated a large number of data, which are disorganized and hard to be used repeatedly. Bioinformatics, synthesized informatics, statistics and some other subjects, makes them orderly and much more valuable. In drug discovery, Bioinformatics takes the place of some conventional ways because of low cast and high throughput. This article introduces the current situation and application of Bioinformatics in drug discovery and looks forward to the future, hoping to provide a reference for the development of new drugs.

1 Introduction

The research and development process of new medicines is mainly the screening of lead compounds, the discovery of target proteins, the analysis of drug action mechanisms and clinical trials and so on. The average R & D cycle of new medicines is about 10 years, and the cost is about USD 500 million to 1 billion. However, under such high-cost input, the output return is not ideal, and it is in urgent need of methods and tools for cost reduction. Bioinformatics is a cross-discipline that combines biology, mathematics, information technology, computer science, and statistics, and it is an emerging discipline that was born with the Human Genome Project. It aims to collect and manage the current large and messy biological data, and discover the residual value of existing data through deep learning and data mining. Based on the above characteristics, bioinformatics is expected to improve the situation of high drug research and development costs and long research and development cycles. This article reviews the application of bioinformatics in the development of new drugs, and looks ahead to the prospects in the field of bioinformatics for new drugs.

2 Main Research Content of Bioinformatics

The morphological characteristics and genetic information of biology are all contained in the sequence of genes and proteins. To explore pathogens and solve diseases, it is very efficient to use bioinformatics for sequence comparison analysis and protein structure prediction of target proteins.

2.1. Sequence comparison

Sequence comparison is a relatively basic content in bioinformatics. It can infer the function of the target gene, construct and speculate the structure and function of the protein, reveal the relationship of homology in biological evolution, and indicate the conserved regions and different regions between the sequences [1]. Due to the too much manual comparison workload and the uncertainty in the sequence, it is difficult to measure the effect of the comparison, so people often subjectively measure the size of the difference between the sequences after comparison. There are three main mathematical models (objective functions) used to calculate the difference: comparison sum function, consistency function and tree function [2], in which, the comparison sum function is most frequently used. The sum of pairs is usually called the SP value, and many algorithms are dedicated to obtain a higher SP value. Sequence comparison can be divided into dual sequence comparison and multiple sequence comparison. To meet the demand, multiple sequence comparison which is more complicated but has a better effect is usually carried out. [3] Multiple sequence comparison method includes dynamic programming algorithm, iterative alignment algorithm, genetic multiple sequence comparison algorithm, genetic annealing-based comparison algorithm, etc. [4].

2.2. Predication on protein structure

Protein structure prediction refers to guessing its possible secondary structure and tertiary structure through amino acid sequence, so as to understand some functions determined by the predicted protein structure, which is of great significance for the study of unknown proteins. The prediction accuracy of protein secondary structure can reach 80% [5], but the prediction accuracy of tertiary structure is still difficult to reach a satisfactory level today,

so it is often necessary to be verified and improved through experimental means. There are three main prediction methods for the tertiary structure: comparative modeling, folding recognition, and ab initio prediction method [classification modeling and recognition of six protein folding types_Liu Yue (foreign language can be applied)]. Although the ab initio prediction method has a higher accuracy, it is more difficult. This article mainly introduces comparative modeling method and folding recognition method.

2.2.1. Comparative modeling method

If the measured sequence is more than 30% homologous to the template sequence, a comparative modeling method can be used. Therefore, searching the database to obtain sequences with high homology is a crucial step in this method [7]. After obtaining the spatial structure of the sequence with higher homology, the template of the measured sequence can be established based on this, and the final tertiary structure to the measured sequence can be reasonably adjusted by using amino acid deletion.

2.2.2. Folding recognition method

When it is impossible to find a homologous sequence with a homology higher than 30%, it is often necessary to model from another angle. Compared with the complexity and irregularity of protein tertiary structure, the total number of protein folding types is only more than 1,000 [8], so it is easier to predict the tertiary structure from the perspective of protein folding. Compared to sequences, the spatial structure is more conservative in evolution. Therefore, it is allowed to proceed with the folding type and model the measured protein through comparing and learning the spatial structure.

3 Introduction of mew medicine research and development

The research and development of the medicine can be divided into three stages: the discovery and screening of the new medicine, preclinical research and clinical research. The discovery and screening of the new medicine is mainly drug target confirmation, lead compound screening, structure-activity relationship research, etc .; the preclinical research stage is mainly pharmaceutical, pharmacology and toxicology research; the clinical research stage is mainly the phase-I clinical research, phase-II clinical research and phase-III clinical research.

3.1. The discovery and screening of the new medicine

In the past three decades, the focus of drug discovery has shifted from the chemical structure of drugs to target-based approaches. Target is not limited to protein products, but also includes certain genetic locus. With the development of biological technology, more and more targets have been discovered. The research and development of the new medicine firstly needs to find the most suitable and effective target among many targets and design drugs for it. Target screening methods include cDNA array technology, yeast two-hybrid system and so on. After determining the target, the lead compound can be designed and searched. Lead compound refers to a compound selected by bioassay that can have a certain physical and chemical effect on the target, and its structure and function have certain plasticity [9]. The identification of biological macromolecules and ligands and their interactions are very important. It may have problems such as weak activity, low selectivity, poor absorption, and strong toxicity, and cannot be directly put into use, so it needs to be decorated [10]. There are two ways to find lead compounds, one is to find compounds with new structure from natural compounds, and the other is to rationally design compounds that have an effect on the target. The former relies on medicinal chemists to extract new chemicals from natural organisms and to detect and screen them. This method shows that the lead compound has low efficiency and poor directivity [11]. However, based on the data accumulated over a long period of time, this method still plays an important role. The latter uses more knowledge of molecular biology and structural biology to design drug molecules. After screening targets and lead compounds, the molecular biology and other methods can be used to design various compounds around the lead compound, that is, optimize and modify the lead compound, and finally screen out more effective compounds to enter the preclinical research stage.

3.2. Preclinical research

Medicines must be conducted with the pharmacology and toxicology researches before clinical trials to ensure the effectiveness and safety Safety and effectiveness are the main factors that determine the success of new medicine research and development [12]. Pharmacology is helpful to study the action mechanism of medicines, including the physiological mechanism, pathological mechanism and pharmacological mechanism of drugs, so as to give guiding opinions on clinical medication and improve the effectiveness of clinical application of medicines[13]. Toxicology helps us understand the mechanism of toxic and side effects of new medicines during the research and development process of new drugs, so that we can make a correct assessment of the safety of new drugs. In addition, it helps find the strong toxic and side effects of certain drugs at an early stage, so as to eliminate at as early as possible and avoid causing great economic loss [14]. Early detection also allows us to avoid damaging the health of clinical subjects. However, with the development of technology and the emergence of toxicology, we can now evaluate the safety of drugs throughout the process to further ensure the safety of drugs [15].

3.3.Clinical research

Clinical research is mainly divided into phase I clinical trial, phase II clinical trial, phase III clinical trial and phase IV clinical trial. The phase I clinical trial is to ensure human safety and is based on extensive animal testing. The enrollment of Phase I clinical research is 10 to 100 healthy volunteers (for cancer drugs, the enrollment is usually cancer patients). The phase II clinical trial is to test the safety and effectiveness of the drug for human body. It recruits a certain number of volunteers who are sick. The phase III clinical trial is to further confirm the effectiveness and safety of the drug, and its experimental subjects are often a large number of clinical patients. The phase IV clinical trial is to ensure that the drug will be effective and safe when it is widely used after it is marketed.

4 Application status of bioinformatics in new drug research and development

4.1. Discovery and screening of drug target

Drug targets refer to biological macromolecules such as proteins and nucleic acids that can be acted on by drugs and produce certain pharmacological effects to achieve the effect of treating diseases [16]. The application of bioinformatics technology can greatly shorten the discovery cycle of targets and break through a major bottleneck in the development of new drugs. The development of various detection technologies has resulted the accumulation of information related to drug targets which require further processing, so several databases such as OMIM [17] and COSMIC [18] have been established at home and abroad. These databases collect various information about the targets in the literature and related databases, such as structural information, ligand information and disease information and so on, so that they are systematically presented according to certain standards for human query and target discovery in the process of new drug research and development [19]. In addition to the establishment of related databases of drug targets, there are many methods for bioinformatics to solve the problem of target discovery, such as genomic methods, gene chip methods, proteomics methods, and metabolomics methods. Their research objects include receptor proteins, enzyme proteins, and small molecule transmembrane carriers. The methods of homology search and sequence comparison can be used to compare candidate targets with targets of known characteristics, analyze the nucleotide and amino acid sequences, understand the physical and chemical properties of the target targets such as hydrophobicity and isoelectric point, and preliminary judge the potential of the candidate target as a target [20], so as to avoid spending a large amount of money for verification experiments. At the same time, a lot of research on targets can also promote the new use of old drugs, that is, the redirection of drugs, so as to find new targets for existing drugs. Currently, there are only about seven hundred drug targets that have been used for treatment [21]. But according to genomics and proteomics research, biomolecules that can be used as targets are far more than this number. With the further development of bioinformatics, more biomolecules will be discovered and used.

4.2. Screening of lead compound

The traditional screening methods for lead compounds have high requirements on experimental conditions and funds. However, as the structure of more and more biological macromolecules is determined and included in an ordered database, the methods of using bioinformatics technology to use these data to find lead compounds and design and modify them have been increasingly used. [22] In many methods, the role of molecular docking in the screening of lead compounds is becoming increasingly important. [23] Molecular docking: For the three-dimensional structure or structure-activity relationship of a specific target protein related to disease, it is firstly required to find the compound that can produce an effect from the relevant database through the computer, calculate the binding force of the two and simulate the binding mode of the two, and then obtain the candidate compounds by means of pharmacokinetics, toxicity determination, and activity evaluation. [24] After the outbreak of the new coronavirus, Li Leping et al have used the molecular docking method to discuss the intervention effect of the main ingredients of Sanren Decoction (betulinic acid and gingerolide peroxide) on COVID-19. [25] Liu Leping et al have also studied the affinity of core compounds such as luteolin and quercetin in the components of Maxing Yigan Decoction to COVID-19 through molecular docking. [26] In addition to the molecular docking method, three-dimensional structure search and new drug design are also the specific manifestations of using bioinformatics for lead compound screening. The three-dimensional structure search needs to be matched with the pattern recognition technology of artificial intelligence, so as to compare the structural information of many molecules in the three-dimensional structure database with the structural information of the target protein, thus finally finding the suitable compounds as candidates. The new drug design overcomes the limitations of the previous two methods that require known compounds, and uses LUDI [27 (Applied Literature 18) Application and Prospect of Bioinformatics in Drug Development_Feng Yi], GROW [28 (Applied Literature 19) Application and Prospect of Bioinformatics in Drug Development_Feng Yi] and other software to automatically design new drug molecules matching it according to the structure and chemical properties of the active site of the target protein.

4.3.Discovery of drug action mechanism (prediction of pharmacology and toxicology)

With the advent of the era of big data and the rapid development of bioinformatics, the British pharmacologist Hopkins proposed the concept of “network pharmacology” in 2007 [29]. Network pharmacology has a holistic and systematic characteristic,

which is helpful to comprehensively and systematically analyze the mechanism of action of drugs. Internet pharmacology emphasizes that diseases are caused by long-term and complex imbalances in individual organisms[30], not just single gene and single target problems. It shifts its focus from single target to multiple targets, which is consistent with the research of traditional Chinese medicine.[31] To conduct network pharmacology research, we first need to filter out the target information of drug components from the database, build a multi-dimensional network of drugs, targets and diseases, and perform pharmacological analysis on each node in this network, in order to study the signal transduction pathway of compound action. For example, Zhang Chi et al. used network pharmacology to study the mechanism of Lianhua Qingwen in the treatment of new coronavirus pneumonia [32].

4.4. Clinical trial stage

In the three clinical stages of new drug approval, it is necessary to recruit a certain number of patients. Qualified subjects play a vital role in the success of clinical trials [33]. Traditional clinical recruitment needs to check the adaptability of volunteers one by one, which is a relatively long process. With the help of bioinformatics, the health information provided by the volunteers can be analyzed, and the volunteers who are not suitable for the experiment can be quickly excluded, which greatly improves the matching efficiency and the success rate of the experiment, and saves the cost of doing more experiments. At the same time, analyzing the existing huge clinical data can also optimize the clinical trial process, further saving time and money costs.

5 Prospect

The development of bioinformatics in the 21st century is very rapid. Although there are still many imperfections, it has been able to help people use the huge data accumulated in the development of new drugs and save a lot of manpower and material resources for drug research and development units. Thus, to a certain extent, the problems of current high drug research and development cost, low research and development success rate, and low market return rate are alleviated. We can see from the past development that the direction of bioinformatics development is more precise, more reliable and more multidimensional. Over time, bioinformatics will penetrate into all aspects of new drug development and become an indispensable and important force in drug development.

References

1. Guo Zhiyun, Zhang Huaiyu, Liang Long, Advances in Bioinformatics Technology [J], Biotechnology Communications, 2004 (03): 313-317.
2. Zou Quan, Guo Maozu, Han Yingpeng, Li Wenbin. Research Progress of Multiple Sequence Comparison Algorithm [J], Bioinformatics, 2010,8 (04): 311-315.
3. Zhang Min, Research Status and Prospect of Biological Sequence Comparison Algorithm [J], Journal of Dalian University, 2004 (04): 75-78 + 82.
4. Li Meiman, Research Progress of Sequence Comparison Technology and Algorithm in Bioinformatics [J], Modern Computer (Professional Edition), 2012 (26): 18-21.
5. Chen Jiale, Wang Shuning, Application of Bioinformatics in Protein Research [J], Science and Technology Innovation Herald, 2018, 15 (26): 250-252.
6. Li Yue, Li Xiaoqin, Xu Haisong, Qiao Hui, Classification Modeling and Recognition of Protein Folding Types [J], Journal of Physical Chemistry, 2009, 25 (12): 2558-2564.
7. Guo Zhiyun, Zhang Huaiyu, Liang Long, Advances in Bioinformatics Technology [J], Biotechnology Communications, 2004 (03): 313-317.
8. [8]Guo Zhiyun, Zhang Huaiyu, Liang Long, Advances in Bioinformatics Technology [J], Biotechnology Communications, 2004 (03): 313-317.
9. Wang Danni, Liu Guannan, Discovery and Activity Screening of Lead Compounds [J], Enterprise Herald, 2012 (20): 278.
10. Ye Deju, Luo Xiaomin, Shen Jianhua, Zhu Weiliang, Shen Xu, Jiang Hualiang, Liu Hong. Discovery of Lead Compounds——Integrating Computer Virtual Screening, Chemical Synthesis and Biological Test Method [J], Progress in Chemistry, 2007 (12): 1939-1946.
11. Fu Qingjie, Liu Shuwen, Wu Shuguang, Selection and Verification of Drug Targets [J], Life Science Research, 2001 (S1): 206-209.
12. Du Guanhua, Pharmacology Development Promotes New Drug Research and Development [J], Chinese Journal of Pharmacology and Toxicology, 2016, 30 (12): 1243-1249.
13. Du Guanhua, Pharmacology Development Promotes New Drug Research and Development [J], Chinese Journal of Pharmacology and Toxicology, 2016, 30 (12): 1243-1249.
14. Wang Quanjun, Wu Chunqi, Liao Mingyang, New Progress in Drug Toxicology Research [J], China Journal of New Drugs, 2007 (03): 177-181.
15. Liao Mingyang, Status and Prospect of Drug Toxicology Research in China [J], Chinese Journal of Pharmacology and Toxicology, 2015,29 (05): 727-728.
16. Liu Wei, Xie Hongwei, Discovery of Potential Drug Targets Based on Bioinformatics Methods

[J], Advances in Biochemistry and Biophysics, 2011,38 (01): 11-19.

17. Li Qun, Application of OMIM Database in the Practice of Medical Genetics [J], Chinese Journal of Eugenics and Genetics, 2010, 18 (06): 7.
18. Wu Meng, Li Jiao, Kang Hongyu, Hou Li, Research on Classification and Integration of Gene Mutation Data for Precision Medicine [J], Chinese Journal of Medical Library and Information Science, 2018, 27 (11): 16-22.
19. Pang Xiaocong, Liu Ailin, Du Guanhua, Application Progress of Drug Target Database [J], Chinese Journal of Pharmacy, 2014, 49 (22): 1969-1972.
20. [20] Ferrari Stefania,Losasso Valeria,Costi Maria Paola. Sequence-based identification of specific drug target regions in the thymidylate synthase enzyme family[J]. ChemMedChem,2008,3(3).
21. Lafferty-Whyte Kyle, Mormeneo David, Del Fresno Marimon Montse. Trial watch: Opportunities and challenges of the 2016 target landscape[J]. Nature reviews. Drug discovery,2017,16(1).
22. Feng Yi, Jing Linhai, Wang Bin, Application and Prospect of Bioinformatics in Drug Research and Development [J], Western Medicine, 2007 (05): 971-973.
23. Westbrook John, Feng Zukang,Chen Li,Yang Huanwang,Berman Helen M. The Protein Data Bank and structural genomics[J]. Nucleic acids research,2003,31(1).
24. Ye Deju, Luo Xiaomin, Shen Jianhua, Zhu Weiliang, Shen Xu, Jiang Hualiang, Liu Hong. Discovery of Lead Compounds——Integrating Computer Virtual Screening, Chemical Synthesis and Biological Test Method [J], Progress in Chemistry, 2007 (12): 1939-1946.
25. Li Jiali, Yang Liangjun, Zhou Hengli, Lin Kunyang, Liang Qiting, He Wei, Zhao Ziming, Pan Huafeng, Using Network Pharmacology and Molecular Docking Method to Explore the Mechanism of Sanren Decoction's Main Ingredients on the New Coronavirus Pneumonia (COVID-19), Chinese Herbal Medicine, 2020: 1-10.
26. Liu Leping, Long Qian, Cao Xueshuai, Xu Xinyi, Luo Yanwei, Gui Rong, Explore the Active Compounds of Maxing Yigan Decoction for the Treatment of New Coronavirus Pneumonia (COVID-19) Based on Network Pharmacology and Molecular Docking Method, Chinese Herbal Medicine, 2020: 1-9.
27. Bohm HJ. The computer prog ram LUDI: A new method for the de novo design of enzyme inhibitors [J]. Journal of Computer-Aided Molecular Design,1992,Vol.6(1),pp.61-78.
28. Xie Jing, Gao Shan, Li Lin, Xu Yilan, Gao Shuming, Yu Chunquan, Research Progress and Application Strategies of Network Pharmacology in the Field of Traditional Chinese Medicine [J], Chinese Herbal Medicine, 2019, 50 (10): 2257-2265.
29. Xie Jing, Gao Shan, Li Lin, Xu Yilan, Gao Shuming, Yu Chunquan. Research progress and application strategies of network pharmacology in the field of traditional Chinese medicine [J]. Chinese Herbal Medicine, 2019, 50 (10): 2257-2265.
30. Zhang Yanqiong, Li Shao, Some Progresses in Network Pharmacology and Modern Research in Traditional Chinese Medicine [J]. Chinese Journal of Pharmacology and Toxicology, 2015,29 (06): 883-892.
31. Wang Lin, Yang Zhihua, Zhang Haoran, Yu Hangxing, Yang Kang, Fu Baohui, Yang Hongtao, Network Pharmacology Research and Preliminary Evidence of Lianhua Qingwen in Treating New Coronavirus (2019-nCoV) Pneumonia [J / OL], Chinese Medicinal Materials, 2020 (03): 772-778.
32. Zhao Junpeng, Chen Xuesong, Chang Tianying, Liu Lang, Shi Xiaodan, Li Yuanying, Liu Fanghan, Yang Haimiao, Establishment and Practice of the Recruitment Management System for Subjects in Phase I Clinical Trial [J], Electronic Journal of Clinical Medical Literature, 2019, 6 (32): 7-8.