Computer Aided Drug Design Based on Artificial Intelligence Algorithm

Wenye Yu\textsuperscript{1,*}, Zhenyu Chen\textsuperscript{2}
\textsuperscript{1}Nanchang Institute of Science and Technology, Nanchang, 330108
\textsuperscript{2}Jiangxi Tourism&Commerce Vocational College, Nanchang, 330100

*Corresponding author e-mail: yuwenye2021@ncist.edu.cn

Abstract. The problems such as high cost and long development time in drug design and development have an important impact on its development, which makes many scholars devote themselves to looking for the auxiliary model of drug design. With the rapid development of computer technology, computer-aided drug molecular research model is more and more mature. This paper aims to study the computer-aided drug system based on artificial intelligence algorithm, so that researchers can speed up the process and reduce the cost when searching for specific protein molecules. In this paper, the principle of complementary matching in the docking process of target molecules and ligands, which is commonly used in drug design, is described, and the functional expression mode and various docking methods of molecular docking are studied. Finally, the research hotspots of molecular docking technology are analyzed, including scoring function, search strategy and flexible protein docking. Ant colony algorithm is introduced into molecular docking platform as a variant of conformation search algorithm, and a new plants algorithm is developed. Finally, the implementation of plants algorithm is analyzed in detail, and the optimized plants system and gold system based on genetic algorithm are simulated, and the relevant experimental data are counted. The simulation results show that the new drug design method based on ant colony algorithm has advantages in docking success rate, docking speed and docking accuracy. The success rate of plants is higher than that of gold, and the docking time is only 1/6 of that of gold.

Keywords: Artificial Intelligence, Computer Aided Drug Design, Molecular Docking, Ant Colony Algorithm

1. Introduction
In the exploration of drug molecular development, a structured drug molecular simulation method molecular docking technology has gradually developed into one of the key areas of innovative medical drug development in China, and is also considered to be an important route for the exploration and development of leading medical compounds in China [1-2]. Molecular docking technology, as an important part of computer-aided drug molecular design engineering, not only requires them to have the ability of high-speed computer operation, but also needs them to be able to effectively combine
various optimization methods and promote their mutual integration with drug development process [3-4]. As an algorithm of artificial intelligence, ant colony algorithm has strong sustainability. The combination of docking with molecules can effectively help the design of drug molecules [5-6].

Many scholars have made some achievements in the research of artificial intelligence algorithm in computer-aided [7-8]. For example, in the past, some scholars in China proposed the application of genetic algorithm in computer-aided drug molecular design, such as two-dimensional quantitative structure-activity relationship and comparative molecular field analysis [9]. Other scholars have proposed the application of immune genetic algorithm in computer-aided design, which optimizes the model parameters and kernel parameters of support vector machine for landscape classification, and further improves the accuracy of landscape classification [10].

This paper aims to optimize the ant colony algorithm, and this improved algorithm is applied to molecular docking. The main working principle is to maintain the original scoring function and flexible strategy, and combine the improved ant colony algorithm with the common docking software autodock. Through the specific detection method of protein molecule and ligand protein molecule binding, the simulation and experiment are carried out. Based on this technology, its characteristics are analyzed. Compared with a typical genetic algorithm, the cellular ant colony algorithm has achieved great advantages and remarkable excellence in some technical fields, on the basis of this technology, some new high-performance protein docking software with strong international influence were innovated and improved [11-12].

2. Computer aided drug design theory based on artificial intelligence algorithm

The process of drug design is the recombination of many macromolecular proteins and ligands, which can be divided into two types: direct drug design and indirect drug design. Molecular docking technology is an important method of direct drug design. The process of molecular docking is to screen a large number of proteins, viruses, polypeptides and other macromolecules by computer simulation, and find out the most stable macromolecules with drug molecules.

2.1. Molecular docking technology

2.1.1. Molecular docking principle.
Molecular docking method is to take the molecules in several major protein 3D structure databases as target molecules, and place the drug molecules under consideration one by one at each active site of the target molecule. According to the chemical environment complementary, energy complementary, geometric complementary and other ways, we can find the most suitable conformation for drug small molecules to match with ligand molecules, thus, the binding mode and affinity of them were predicted theoretically. In order to get the best conformation of drug molecule and ligand molecule, the common method is to construct a scoring function to select the matching method with the lowest energy consumption or the highest affinity to the receptor by comparing the scores.

2.1.2. Expression pattern of molecular docking.
How to express the docking process is related to which method is used to optimize the alignment, so the expression pattern of target molecules and drug molecules in the binding process is one of the important contents of drug docking. There are three expression modes of molecular docking: atom based, surface based and grid based. Among them, this expression mode can describe the intermolecular force most accurately, but it is seldom used in large-scale calculation because of its complexity and high cost. The surface based molecular docking model considers the interaction range of macromolecules, which greatly simplifies the calculation and is widely used. The mesh based expression mode needs to mesh the docking molecules first, which can reduce the calculation in scoring molecules.
2.1.3. Scoring function.

In each docking, the algorithm will simply predict the free energy of the system, and the excellent scoring function strategy can balance the docking accuracy and computational efficiency.

1) Scoring method based on molecular force field.

This method is based on the theory of molecular mechanics in physics. By analyzing the binding site of the conformation and estimating the binding free energy, we can score the conformation of the complex. Molecular docking software dock uses the most classic amber molecular force field. The calculation formula of amber's binding free energy is as follows:

\[
E = \sum_{i=1}^{\text{lig}} \sum_{j=1}^{\text{rec}} \left( \frac{A_{ij}}{r_{ij}^2} - \frac{b_{ij}}{r_{ij}^2} + 332 \times \frac{q_i q_j}{\kappa n_{ij}} \right)
\]

(1)

2) Scoring method based on empirical regression parameters.

Based on the experience scoring method, the binding free energy is decomposed into several different and independent energy terms, and the influence of each energy term on the binding free energy is considered. Finally, the binding free energy is calculated by accumulating these energy terms. The calculation formula is as follows:

\[
AG = \sum_i W_i \Delta G_i
\]

(2)

3) Knowledge based scoring method.

Knowledge based scoring method is independent of the force field parameters and training set. It has better balance performance and excellent calculation accuracy on the basis of ensuring a certain calculation speed. The expression is as follows:

\[
A_{ij}(r) = -kT \ln \left( \frac{p_{ij}(r)}{p_{b}^{ij}(r)} \right)
\]

(3)

\[
AG = \sum_{r<r_{cr}} A_{ij}(r)
\]

(4)

4) Scoring method of consistency evaluation.

The most important feature of this method is that multiple scoring functions are used in the scoring process to evaluate the docking complexes in parallel, so as to select the ligand molecules with strong adaptability, thus increasing the versatility.

2.1.4. Typical molecular docking software. Currently, the common molecular docking software includes dock, autodock, affinity and gold.

2.2. Application of molecular docking in computer aided drug design

2.2.1. Ampc β-lactamase inhibitor. In the process of searching for these inhibitors, molecular docking technology has been widely used. As a docking platform, dock software was used to screen out AmpC inhibitors in ACD database β-Drug molecules of lactamases. In the course of the study, the top 500 molecules in each docking process were tested based on structural complementarity. Finally, 56 compounds with the highest binding degree were selected for in vitro experiments.

2.2.2. Ring opening derivatives of corydaline, a new acetylcholinesterase inhibitor. In the design process of this kind of drug molecules, gold software is used as a docking tool for docking and screening PDB crystal structure database.

2.2.3. Ant colony algorithm. Ant colony algorithm is a simulation of ant foraging algorithm, but it is not exactly the same as the real ant, it is a kind of "artificial ant", has the characteristics of information exchange when ant foraging, and adds some other characteristics.
3. Simulation experiment based on artificial intelligence algorithm

3.1. Algorithm implementation and operation environment

This experiment is to improve autodoc4.2 program by ant colony algorithm, replace the original genetic algorithm in the program to search the global. This paper tests based on the protein ligand system model, compares the original genetic algorithm, and then analyzes the convergence rate of computer-aided drug design system based on ant colony algorithm. Autodock 4.2 is a completely open source software. Based on the general public license platform, users can get complete c++ source code, which is also the advantage of autodock as open source software. In this paper, the docking process between macromolecular protein ligands is realized on autodock4.2 platform and combined with the plans optimization algorithm. The results are compared with other genetic algorithms and other software. This experiment is done on a Q8200 workstation with 8-core processor, which uses Linux operating system.

3.2. Plants parameter optimization

The compounds selected in this experiment are all from CCDC / Astex database. Among the 212 complexes, 11 contain covalent ligands, which are excluded here. The number of ligand rotatable bonds of the remaining complexes fluctuates from 0 to 28. In all experiments, the spherical binding sites of each pair of protein ligand complexes determine the search space of rotational degrees of freedom of ligands. Before docking, the translation, inversion and bending degrees of freedom of the ligand structure are random. These random structures are then passed to the plants algorithm to prevent parameter setting errors. We selected 33 compounds with rotatable bond in the range of 0-10, which is to reduce the calculation time in the test process. Three to four different values are considered for each parameter, so there will be 144 different combinations in parameter configuration. For each compound, the plants algorithm is run 10 times independently. Finally, the average success rate, operation time and the number of scoring functions of each configuration are calculated. The calculation time is calculated by computer. The calculation time needs to exclude the preparation time of protein molecules and ligands.

3.3. Virtual screening

Virtual screening of large-scale compound databases is also one of the main application fields of molecular docking tools, so it is necessary to test the ability of plants to distinguish the biological activity of ligands from that of flying organisms.

4. Analysis of simulation results

4.1. Analysis of simulation results of optimization algorithm based on multiple ant colony

Considering that the actual situation is generally complex and changeable, a series of simulation experiments are carried out to verify the optimization ability of acomac algorithm. The specific parameter configuration is: Here's is the length of the path generated by the nearest neighbor, and N is the number of city nodes. After 50 times of simulation, the weight of local range is w = 0.6, and the external weight () = 0.4. Table 1 compares and analyzes the performance of acomac algorithm and Dorigo's ACS algorithm in solving various traveling salesman problems.

| TSP problem | ACS       | ACOMAC    |
|------------|-----------|-----------|
| Ei151      | 412.928   | 399.344   |
| Ei176      | 597.483   | 596.235   |
| Kroa100    | 22341.234 | 20345.212 |
| D198       | 9798.24   | 9567.79   |
As shown in Figure 1, the improved acomac algorithm has better convergence speed than ACS algorithm in solving TSP problem, and the final value is obviously closer to the global optimal solution.

4.2. Analysis of simulation results based on ant colony algorithm

The plants algorithm is tested under different settings for the list of compounds in the whole database. Table 2 lists the configuration parameters, the success rate of ranking first, third and tenth respectively, the average docking time, and the evaluation times of scoring function.

![Figure 1. Data comparison of acomac algorithm and ACS algorithm in solving different TSP problems](image)

**Table 2. Molecular docking simulation results of plants algorithm**

| ants | Success rate(%) up to rank | time | eval |
|------|---------------------------|------|------|
|      | 1 | 3 | 10 |      |      |
| 0.25 | 64.89 | 75.25 | 80.54 | 27.09 | 0.95 |
| 0.25 | 64.56 | 73.78 | 78.90 | 23.78 | 0.83 |
| 0.50 | 67.90 | 78.56 | 82.12 | 52.87 | 1.75 |
| 0.50 | 68.88 | 79.47 | 83.25 | 48.23 | 1.68 |
| 1.00 | 71.34 | 81.45 | 87.89 | 87.89 | 2.97 |
| 1.00 | 74.45 | 83.87 | 88.45 | 98.23 | 3.34 |
| 3.00 | 75.99 | 88.02 | 93.12 | 289.99 | 9.95 |
Figure 2. Molecular docking simulation results of plants algorithm

As can be seen from Figure 2, for each compound, among the highest ranked solutions, when $\sigma = 25$, the success rate is about 63% $\sigma = 3$, the success rate increased to 75%, but the docking time was extended to 290s. Therefore, in different practical applications, it is necessary to choose the value of parameter to find the best balance between success rate and docking time.

This paper also compares plants with genetic optimization for liquid docking (gold), which is also a docking technology often used in drug design. Gold version 3.0.1 is used in the test described in this section. The maximum value of genetic algorithm running on each ligand is set to 10, and the mechanism of early termination and void detection is adopted. For plants and gold, different parameter settings will compare the time cost and success rate. Chemplp scoring function is used in plants, and gold scoring function is used in gold. The detailed experimental results of gold are shown in Table 3.

Table 3. Molecular docking simulation results of gold system

| Parameter value | Success rate(%) up to rank 1 | Success rate(%) up to rank 3 | Success rate(%) up to rank 10 | time | eval |
|-----------------|------------------------------|------------------------------|------------------------------|------|------|
| $\sigma = 0.2$  | 75.25                        | 73.78                        | 78.23                        | 43.23| N/A  |
| $\sigma = 0.3$  | 69.34                        | 75.45                        | 81.14                        | 114.27| N/A  |
| $\sigma = 1.0$  | 73.87                        | 78.21                        | 82.56                        | 307.68| N/A  |
Figure 3. Molecular docking simulation results of gold system

As can be seen from Figure 3, except that the parameter autoscale in gold is set to 0.1, the results of plants and gold almost follow the same rule: the higher the success rate, the longer the docking time. Therefore, the docking time should be controlled in an acceptable range on the premise that the success rate is high enough. Considering the success rate to level 3 or level 10, we find that the success rate of plants is higher than that of gold, and the docking time is only one sixth of that of gold.

5. Conclusion
As a kind of swarm intelligence algorithm and strategy, ant colony algorithm has strong robustness in solving performance. Therefore, the model of ant colony algorithm can be applied to other problems with a little modification. Moreover, ant colony algorithm is a population-based evolutionary algorithm, which is easy to realize in parallel. In this paper, ant colony algorithm is introduced into computer-aided drug design. The ant colony algorithm is improved and replaced by genetic algorithm used in the global search process of autodock, and its performance and advantages are analyzed and studied. In this paper, the significance and background of the research work in this paper are described, the development history, current situation and future development prospect of CAD are described. The application of algorithm optimization theory in practice is introduced. Then, the application of molecular docking technology in drug design is analyzed. The principle of molecular docking and the key technologies in the current research are introduced in detail, including the conception of search strategy, scoring function and flexible docking. Finally, several common molecular docking software are introduced and the advantages and disadvantages of each software are pointed out.

References
[1] Sari S. Molecular Modelling and Computer-Aided Drug Design: The Skill Set Every Scientist in Drug Research Needs and Can Easily Get [J]. Hacettepe University Journal of the Faculty of Pharmacy, 2020, 40(1):34-47.
[2] Kitchen, Douglas B. Computer-aided drug discovery research at a global contract research organization [J]. Journal of Computer-Aided Molecular Design, 2017, 31(3):1-10.
[3] Acuna V V, Hopper R M, Yoder R J. Computer-Aided Drug Design for the Organic Chemistry Laboratory Using Accessible Molecular Modeling Tools [J]. Journal of chemical education, 2020, 97(3):760-763.
[4] Computer-Aided Ligand Discovery for Estrogen Receptor Alpha [J]. International Journal of Molecular Sciences, 2020, 21(12):4193.
[5] Scotti L, Francisco J.B.M. Júnior **, Ishiki H M, et al. Computer-Aided Drug Design Studies in Food Chemistry [J]. Natural and Artificial Flavoring Agents and Food Dyes, 2018:261-297.

[6] Rasheed M A, Iqbal M N, Saddick S, et al. Identification of Lead Compounds against Scm (fms10) in Enterococcus faecium Using Computer Aided Drug Designing [J]. Life, 2021, 11(2):77.

[7] Yang W, Wang J, Wang R. Research and Application of a Novel Hybrid Model Based on Data Selection and Artificial Intelligence Algorithm for Short Term Load Forecasting [J]. Entropy, 2017, 19(2):52.

[8] Liu Peijun, Wang Yining, Yu Min, et al. Impact of artificial intelligence based optimization algorithm on image quality of low dose coronary CT angiography in big size patients [J]. Radiology practice, 2019, 034 (007): 760-766

[9] Kuppusamy P G. An Artificial Intelligence Formulation and the Investigation of Glaucoma in Color Fundus Images by Using BAT Algorithm [J]. Journal of Computational & Theoretical Nanoscience, 2017, 14(4):1-5.

[10] Vivekanadam B. Artificial Intelligence Algorithm with SVM Classification using Dermoscopic Images for Melanoma Diagnosis [J]. Journal of Artificial Intelligence and Capsule Networks, 2021, 3(1):34-42.

[11] Abdollahi J, Keshanbehghian A, Gardaneh M, et al. Accurate Detection of Breast Cancer Metastasis Using a Hybrid Model of Artificial Intelligence Algorithm [J]. Archives of Breast Cancer, 2020:18-24.

[12] Zhang Y, Zhou Y M, Liao Z H, et al. Artificial Intelligence-Guided Subspace Clustering Algorithm for Glioma Images [J]. Journal of Healthcare Engineering, 2021, 2021(2):1-9.