

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



# Drug Design for Malaria with Artificial Intelligence (AI)

*Bhaswar Ghosh and Soham Choudhuri*

## Abstract

Malaria is a deadly disease caused by the plasmodium parasites. Approximately 210 million people get affected by malaria every year resulting in half a million deaths. Among several species of the parasite, *Plasmodium falciparum* is the primary cause of severe infection and death. Several drugs are available for malaria treatment in the market but plasmodium parasites have successfully developed resistance against many drugs over the years. This poses a serious threat to efficacy of the treatments and continuing discovery of new drug is necessary to tackle the situation, especially due to failure in designing an effective vaccine. People are now trying to design new drugs for malaria using AI technologies which can substantially reduce the time and cost required in classical drug discovery programs. In this chapter, we provide a comprehensive overview of a road map for several AI based computational techniques which can be implemented in a malaria drugs discovery program. Classical computers has limiting computing power. So, researchers are also trying to harness quantum machine learning to speed up the drug discovery processes.

**Keywords:** Malaria, *Plasmodium falciparum*, machine learning, drug design, Quantum machine learning, Topological data analysis

## 1. Introduction

Malaria is an infectious and dreadful disease caused by the plasmodium parasites. The parasite is transmitted to humans through the bites of infected mosquitoes. Around 210 million people get infected by malaria every year resulting in 440,000 deaths, especially children under the age of five [1]. The *Plasmodium falciparum* is the primary cause of severe infection and death in most cases. So far, numerous drugs are available in the market such as Quinine, Mepacrine, Chloroquine, Mefloquine, Halofantrine, Artemisinin and their derivatives. Unfortunately, malaria parasites, especially the falciparum species, developed resistance against many of these drugs if not all after some time posing a serious threat to the medication's efficacy [2]. As a result, the continued discovery of new drugs against malaria becomes essential to mitigate this threat. During the last decade, various drug design programs focusing on malaria are initiated all over the world. Drug discovery is a time-consuming and complex process that can be broadly divided into four main phases: (i) the target selection and validation, (ii) screening and optimization of lead compounds, (iii) preclinical studies, and (iv) clinical trials. First, the targets associated with specific diseases need to be identified. This

requires an evaluation of cellular and genetic targets, genomics and proteomics analysis, and prediction bioinformatics. The next step was to hit identification. The compound is identified from the library Molec-snake using combinatorial chemistry, high-throughput screening, and virtual screening. Furthermore, in vivo pharmacokinetic studies are performed to conduct toxicity tests on animals. After, the preclinical tests are conducted successfully, clinical trials are performed on the infected patients. Clinical trial is conducted in three phases. The Phase I constitutes drug safety test with few people; In Phase II, the dose amount of the drug necessary to eliminate the infection is determined on few patients and finally Phase III comprises of precisely quantifying the efficacy of the drug on large number of patients. After the drug candidates' safety and efficacy are confirmed in the clinical phases, agencies such as the FDA review this compound for approval and commercialization. The total cost of a conventional drug development pipeline is projected to be USD 2.6 billion, and it can take more than 12 years for a complete traditional workflow.

## 2. Malaria disease overview

Malaria has a wide effect in the subtropical and tropical continents. Sub-Saharan Africa has the most malaria cases as well as significant cases in India, Brazil, Afghanistan, Sri Lanka, Thailand, Indonesia, Vietnam, Cambodia, and China [3, 4]. In many medium climate areas, such as Western Europe and the United States, public health measures and economic development have succeeded in achieving elimination of malaria, apart from occasionally imported cases through international travels. Female Anopheles mosquitoes transmit malaria. There are approximately 400 species of Anopheles in the world. Among them, 30 species are responsible for malaria. Plasmodium species can also infect animals, birds, etc. The four species of malaria parasites namely *Plasmodium falciparum*, *Plasmodium vivax*, *Plasmodium ovale* and *Plasmodium malariae*, can infect humans under natural conditions. *Plasmodium falciparum* is the major killer among these species. In August 1897, Ronald Ross for first time reported that the parasites can infect female mosquitoes and further showed that the parasite completed its development cycle in the female mosquitoes which is the primary source of the spread of infection from one infected patient to another healthy individual.

### 2.1 Life cycle

Malaria infection begins when the anopheles mosquito bites someone and injects plasmodium parasite into the bloodstream in the form of sporozoites. Sporozoites are then passed to the human liver and multiply asexually in liver cells for the next 7 to 10 days, but do not cause any symptom. In animals, parasites are released from liver cells to the vesicles in the form of merozoites which then enter into the heart and in the lungs and finally stays in the lung capillary.

The vesicles eventually disintegrate, enabling the merozoites to progress to the blood stage of development. Merozoites enter red blood cells (erythrocytes) in the bloodstream and multiply until the cells burst [5] resulting in released merozoites which further penetrate more erythrocytes. Each time parasites invade the blood cells, this cycle is repeated, causing fever. Some parasites inside the infected blood cells leave the asexual multiplication cycle [6]. Instead of replicating, these parasites in those blood cells develop into sexual forms of parasites, called gametocytes flowing through the bloodstream. When an anopheles mosquito bites an infected human, it swallows the gametocytes, which develop further into mature sex cells





Description	References
DeepMalaria (a graph based model using SMILE)	[9]
Analysis of malaria inhibitors	[10]
Prediction of antimalarial drug target	[12]
inhibitor designing against malarial parasite	[11]

*Note: These are the most significant paper for in-silico malaria research and DeepMalaria is the only paper where authors has used deep neural network for malaria drug design.*

**Table 1.**  
*Different approaches for in silico antimalarial drug discovery.*

computation methods are only optimizing physical or chemical properties but these methods are not able to predict whether the designed drug is biologically relevant for malaria. In the next sections, we will provide a road map to harness recent developments in several genomics and AI technique which can be used in drug design programs (**Table 1**).

### 3. Drug target selection using genomics data

The first step in the drug discovery project is to identify a potential target. Transcriptomics data can be effectively utilized for target identification. Differential gene expression analysis provides information about differences in gene expression between normal and diseased states. For malaria parasites, plasmoDB database [13] provides a large variety of trarencryptomics datasets at different stages of the life cycle or at different times after it infects the RBC's. However, bulk RNA-sequencing data does not have the ability to recognize cell to cell variability within a population as a result of which some essential feature may remain undetected. For example, recent single cell RNA-seq experiment was able to differentiate between sexually and asexually committed scizont from a population of parasites [14]. With the launch of such single cell RNA sequencing (RNA-seq), it is now possible to measure RNA levels on the entire genome scale to gain insights into cellular processes and illuminate the specifics of many important molecular events such as alternative splicing, gene fusion, variation of single nucleotide, and differential genes expression. ScRNA-seq enables analysis of individual cell transcriptomes. ScRNA-seq is generally used to examine transcriptional similarities and variations within the cell population. RNA sequencing technologies continue to advance and provide new ideas for understanding biological processes. Early findings revealed previously unrecognised levels of heterogeneity in embryonic and immune cell population [15, 16]. Thus, the analysis of heterogeneity remains a core reason to embark on scRNA-seq studies. Similarly, assessments of transcriptional variations between individual cells have been used to distinguish unusual cell populations that would otherwise go undetected in pooled cell analysis [17], such as malignant tumor cells within a tumor mass [18], or hyper-responsive immune cells within an otherwise homogeneous group [19]. The scRNA-seq technique is also ideal for the examination of individual cells where each cell is essentially unique, such as individual T lymphocytes expressing highly diverse T-cell receptors [20], brain neurons [21], or early-stage embryo cells [22]. In scenarios such as embryonic development, cancer, myoblast, and lung epithelium differentiation and lymphocyte fate diversification, scRNA-seq is also increasingly used to trace the lineage and developmental relationships between heterogeneous yet related cellular states [22–27].

In addition to solving cellular heterogeneity, scRNA-seq can also provide important information on essential gene expression characteristics which include studying the expression of monoallelic genes [15, 28, 29], splicing patterns [30], as well as noise during transcriptional responses [30–32]. Importantly, the study of gene co-expression patterns at the single-cell level could enable the identification of co-regulated gene modules and even inference of gene-regulatory networks underlying functional heterogeneity and cell-type specifications [32, 33]. Additionally, we can extract many more information such as how many genes can be detected and whether a particular gene of interest is being expressed, or whether there has been differential splicing, depending on the procedure of generating the mRNA data. Several single cell RNA-seq experiments have been performed in the last couple of years for plasmodium parasites which paved a much more sophisticated way to characterise gene expression at different stages of the life cycle [14, 34–36]. Both supervised and unsupervised learning methods can be used to identify differentially expressed genes at different life cycle stages [37, 38]. These information can be harnessed in a gene ontology analysis or a genome-scale metabolic model to identify function of the genes and subsequently potential target for a drug.

## **4. Drug discovery with AI**

Classical approach of drug discovery is a time consuming and complex process. It takes almost 12 years to discover a drug with the cost soaring to billions of dollars. Several pharma-companies are working on drug discovery but 90 percent of all drug discovery programs are failing due to limitations both at the computational as well as clinical phases. We can divide drug discovery into four major steps (1) target selection and validation; (2) compound screening and lead optimization; (3) pre-clinical studies; (4) clinical trials. Bio-pharmaceutical industries are focusing on computational approaches in order to enhance the drug discovery processes as well as to reduce research and development expenses by diminishing failure rates in clinical trials and ultimately generate superior medicines. Different machine learning approaches help to identify drug targets, find suitable molecules from data libraries, suggest chemical modifications, etc. There are several steps for drug discovery and we will discuss how computational approaches help in each step of drug discovery process.

### **4.1 Primary drug screening with AI**

#### *4.1.1 Image processing and usage of AI to sort and classify cells*

AI technology performs well at classifying images that contained various objects or features [39, 40]. Various dimension reduction techniques like principal component analysis (PCA) can be utilized to reduce the features of the images and then we can use AI-based techniques to classify the cells [41]. Least square support vector machine (LS-SVM), which use classification and regression techniques shows the highest accuracy (95.34) during classification. Modern devices like activated cell sorting images (IACS) are used to measure the optical, electrical, and mechanical properties of cells for highly versatile and scalable cell sorting automation. This instrument use neural network algorithms to do decision-making and high-speed digital image processing. AI is recently used to interpret computerized electrocardiography (ECG). This process plays an significant role in the diagnosis/clinical treatment of the workflow.

## 4.2 Secondary drug screening with AI

### 4.2.1 Physical properties predictions

For drug design, features like bioavailability, bioactivity and toxicity are very important defining characteristics of a compound. The Partition coefficient (logP) and melting point affect a drug molecule's bioavailability. The melting point of a drug indicates how easily it dissolves in water, whereas logP quantifies relative solubility between oil and water. logP is used to calculate cellular drug absorption. A molecular fingerprint, SMILE(simplified molecular input line-entry system) string, potential energy measurements (e.g., from ab initio calculations), molecular graphs with varying weights for atoms or bonds, Coulomb matrices, molecular fragments or bonds, and atomic coordinates in 3D are examples of molecular representations used in an AI drug design algorithm that takes these properties into account [42]. These inputs can be utilized in the DNN training phase and can be processed by various DNNs in different stages, including generative and predictive stages. We can also use reinforcement learning (RL). The generative stage of a DNN is trained to generate chemically feasible SMILES strings using SMILES inputs in a typical sample, while the predictive stage is trained to predict molecule properties. Although the two stages are initially trained separately using supervised learning algorithms, different kinds of biases may be introduced by rewarding or penalising specific properties when the two stages are trained simultaneously.

### 4.2.2 Predictions of bioactivity and toxicity

The toxicity and bioactivity profiles are significant properties of a compound. Matched molecular pair (MMP) analysis can be used to explore the local changes of the drug molecule and its significance on the molecular properties as well as bioactivity [43]. MMP is used to study the quantitative structure–activity relationship (QSAR) [43]. Random forest(RF), gradient boosting machine (GBMs), and DNNs, the machine learning techniques previously applied without MMP, are used to gather new transformations, fragments as well as modifications of the core static. When it comes to predicting compound activity, DNN outperforms RF and GBM. MMP with ML has been used to predict many properties of bioactivity such as oral exposure, the distribution coefficient (log D) [44, 45], the intrinsic clearance, the absorption, distribution, metabolism, and excretion (ADME), and mode of action owing to the rapid increase of public databases (such as ChEMBL and PubChem) containing a significant amount of structure–activity relationships (SAR). A few methods for predicting the bioactivity of a drug candidate have recently been created. For example, few researcher used a network coding convolution graph with discrete chemicals to extract the drug target sites' signature into a sustainable space latent vectors (LVS). LVS enables optimization based on molecule gradients in space, allowing predictions to be made based on the model's differential affinity and other binding properties. The DeepTox algorithm is important for toxicity prediction [46].

## 4.3 AI in drug design

### 4.3.1 Prediction of target proteins 3D structure

The 3D structure of a target protein's ligand-binding site is usually used to design new drug molecules [47, 48]. As a result, researchers have used homology

modelling and de novo protein design in the past [49–51]. With the development of AI-based approaches, prediction of the 3D structure of a target protein can be performed more accurately. The AI tool AlphaFold is successfully implemented to predict the 3D structure of a drug target protein in the recent Crucial Assessment of Protein Structure Prediction contest and performed amazingly well. AlphaFold correctly predicted 25 of 43 structures using only primary protein sequences. These results outperformed the second-place finisher, who correctly predicted just three of the 43 test sequences. AlphaFold is based on deep neural networks (DNNs) trained to predict proteins' properties based on their primary sequences. It measures the angles between peptide bonds in close proximity as well as the distances between pairs of amino acids. These two features are then combined to generate a score that can be used to predict the accuracy of a proposed 3D protein structure model. These scoring functions are used by AlphaFold to examine the protein structure landscape and find structures that match predictions.

#### *4.3.2 Predicting drug-protein interactions*

Quantum mechanics (QM) is a very effective tool for predicting protein–ligand (drug) interactions [52, 53]. These methods take the help of quantum effects for the simulated system at the atomic resolution, resulting in substantially higher precision than conventional MM methods. The time–cost of QM-based methods is far higher than that of MM methods since MM methods only use basic energy functions based on atomic coordinates [54, 55]. As a result, applying AI methods to QM calculations necessitates a tradeoff between QM accuracy and molecular mechanics (MM) models' favorable time–cost [56]. AI models have been trained to replicate QM energies from atomic coordinates, and they can outperform MM methods in terms of calculation speed. Deep learning can be used to predict the potential energies of small molecules, thus replacing computationally challenging quantum chemistry calculations with a fast ML method [56]. AI is mainly utilized for atomic simulations and predictions of electrical properties, while DL has been utilized to predict the potential energies of small molecules, thus replacing computationally demanding quantum chemistry calculations with a fast ML method. DFT (density functional theory) potential energies derived from quantum chemistry have been measured and used to train DNNs for large data sets. For example, the accuracy of an ML model improved with increasing sample size in a study of two million elpasolite crystals, reaching 0.1 eV/atom for DFT formation energies trained on 10,000 structures. The model was then used to test different compositional choices for different properties (Table 2) [57].

#### **4.4 Possibility of drug design using topological data analysis**

If we have a dataset containing information about compounds with respect to their structural features, toxicity, binding affinity to the target then we can use TDA, a mathematical technique to create a similarity network of the compounds by studying shapes preserving high-dimensionality. TDA will enable us to visualize our compound library as a two-dimensional network, with compounds (located in nodes) linked by a series of edges indicating their degree of mutual similarity. As a result, two compounds with identical properties will appear closer together in the network, whereas two compounds with vastly different properties will appear farther apart. This network would make it simple to create subgroups or families of related compounds, which could then be used to choose the best compound.



Tools name	Description	Reference
AlphaFold	Protein 3D structure prediction	
Chemputer	A more structured format for documenting a procedure for chemical synthesis	[58]
DeepChem	A python-based AI platform for different predictions of drug discovery tasks	[59]
DeepNeuralNet-QSAR	Molecular activity predictions	[60]
DeepTox	Toxicity predictions	[46]
DeltaVina	A scoring feature for protein-ligand binding affinity rescoring	[61]
Hit Dexter	ML models for molecule prediction that could react to biochemical assays	[62]
Neural Graph Fingerprints	Prediction of properties for novel molecules	[63]
NNScore	For protein-ligand interactions, neural network-based scoring mechanism	[64]
ODDT	A robust chemoinformatics and molecular modelling toolkit for use	[65]
ORGANIC	An powerful method for molecular generation to build molecules with desired characteristics	[66]
PotentialNet	Ligand-binding prediction of affinity on the basis of a convolution Neural Network (CNN)	[67]
PPB2	Polypharmacology prediction	[68]
QML	A Python toolkit for quantum ML	
REINVENT	Using RNN (recurrent neural network) and RL (reinforcement learning), molecular de novo architecture	[69]
SCScore	A scoring feature for the assessment of a molecule's synthesis complexity	[70]
SIEVE-Score	An improved method of virtual screening based on structure via interaction-energy-based learning	[71]

**Table 2.**  
*List of AI-Based Computational Tools for Drug Discovery.*

**4.5 Possibility drug design using Quantum machine learning**

The Quantum Computer Algorithm (QC) was introduced as a way to speed up classical machine learning algorithm. However, modern computers are not enough to analyze the behavior of the atoms in a molecule. Even the most powerful super-computers at this time can only simulate relatively simple molecules by significantly limiting their ability to predict the interaction of complex molecules and atoms. Classical computing (CC) and big data analysis methods can screen desired properties of molecules. Researchers use Molecular simulator to simulate interactions between electrons of each atom to test how they react with each other in the real world. However, due to CC's relatively limited processing capacity, the computationally designed drug fails to deliver the desired results in 90 percent cases during the first phase of the clinical trial causing loss of billions of dollars every year. When Google announced a successful quantum computing experiment (QC) that has achieved what is called "quantum supremacy" (calculations that are too complex to do using CC), it was praised by many people around the world as a key moment for QC.

It is also an important development for drug discovery. Indeed, QC can change the drug discovery game entirely by packing enough computational muscles to enable molecular analysis at an unprecedented scale which was completely unimaginable with a CC. Additionally, quantum machine learning (QML) which uses quantum algorithms to perform complex algorithms of learning assignments [72] can augment the computational tools of drug discovery programs even further. Classical machine learning has already shown a significant promise in drug discovery. QML allows scientists to translate classical ML algorithms into quantum circuits to run an ultra-strong quantum computer efficiently. This quantum computer will become more efficient in future as the quantum technologies become less prone to mistakes. Scientists from the University of Warwick, the University of Luxembourg, and the University of Berlin recently created a deep ML algorithm that could predict the molecule's quantum state faster than before. "Solve the fundamental equation of quantum mechanics in conventional ways requires high-performance computing resources and computational months," explained the laboratory news. The team said that "the new AI algorithm developed can provide accurate predictions in seconds on a laptop or cellphone." Researchers from Pfizer have made similar progress, using modeling techniques called predictions of the crystal structure (CSP) to map molecular 3D structures - calculations that usually take months to solve. However, there is a long path to travel before we can say that we have achieved a practical quantum advantage. However, QC and QML hold extraordinary promises for the pharmaceutical business - which, when refined, will pave the way towards the development of new drugs as well as safer and cheaper health products for patients in the future.

## 5. Conclusions

Many pharmaceutical companies are using AI tools in drug discovery process. Costs and time remain big challenges in drug discovery programs. There are usually about 1 million compounds in a standard high-throughput screening library, where designing each compound typically costs 50–100 USD. As a result, an initial screening phase would cost several million dollars and take months to complete. AI or ML techniques help to do optimization of lead compound. It takes only few days to find the lead compounds by AI, when classical approach takes several years. AI helps to predict bio-activity, toxicity, physical properties, structure prediction of potential drug. There are few companies like Merck, Novartis etc. who are using AI technologies to design drug. Classical computer has some limiting computing power. So, Researcher are trying to make quantum computer or using quantum machine learning algorithm in Classical computer to do computation in a faster way. Since, malaria is one of the major health burdens in the developing world, AI based drug design programs will be immensely helpful in aiding WHO's goal to reduce cases of malaria by 90 percent by 2030. The inefficacy of vaccination strategies further impose all the burdens on continuous discovery of new drugs. We strongly suggest through this review that AI based drug program would substantially benefit in tackling this debilitating disease with respect to saving human life at lower amount of time and cost.

## Acknowledgements

Authors thank Department of Biotechnology (No. BT/RLF/Re-entry/32/2017), Government of India for funding this project.

The authorship criteria are listed in our Authorship Policy:

<https://www.intechopen.com/page/authorship-policy>.

This section of your manuscript may also include funding information.

IntechOpen

## Author details

Bhaswar Ghosh<sup>†</sup> and Soham Choudhuri<sup>\*†</sup>

Center for Computational Natural Sciences and Bioinformatics, International  
Institute of Information Technology, Hyderabad, India

\*Address all correspondence to: [choudhurisoham@gmail.com](mailto:choudhurisoham@gmail.com)

<sup>†</sup> These authors contributed equally.

## IntechOpen

© 2021 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

## References

- [1] World Malaria Report 2020. <https://www.who.int/news-room/fact-sheets/detail/malaria> (2020)
- [2] White, Nicholas J. "Antimalarial drug resistance." *The Journal of clinical investigation* vol. 113,8 (2004): 1084–92. doi:10.1172/JCI21682
- [3] Snow RW, Craig M, Deichmann U Marsh K: Estimating mortality, morbidity and disability due to malaria among Africa's non-pregnant population. 1999, *Bull WHO* 77, 624–640. <https://pubmed.ncbi.nlm.nih.gov/10516785/> [Accessed: 21 January 2021]
- [4] Breman JG, Egan A Keusch GT, The intolerable burden of malaria: a new look at the numbers. 2001, *J Trop Med Hyg* 64, (Suppl. 1–2), iv–vii [https://core.ac.uk/reader/13114159?utm\\_source=linkout](https://core.ac.uk/reader/13114159?utm_source=linkout) [Accessed: 21 January 2021]
- [5] Yamauchi LM, Coppi A, Snounou G Sinnis P (2007) *Plasmodium sporozoites* trickle out of the injection site. *Cell Microbiol* [Epub ahead of print].
- [6] Mota MM, Pradel G, Vanderberg JP, Hafalla JC, Frevert U, Nussenzweig RS, Nussenzweig V Rodriguez A (2001) Migration of *Plasmodium sporozoites* through cells before infection. *Science* 291, 141–144.
- [7] Frevert U, Sinnis P, Cerami C, Shreffler W, Takacs B Nussenzweig V (1993) Malaria circumsporozoite protein binds to heparan sulfate proteoglycans associated with the surface membrane of hepatocytes. *J Exp Med* 177, 1287–1298.
- [8] Sturm A, Amino R, van de Sand C, Regen T, Retzlaff S, Rennenberg A, Krueger A, Pollok JM, Menard R Heussler VT (2006) Manipulation of host hepatocytes by the malaria parasite for delivery into liver sinusoids. *Science* 313, 1287–1290.
- [9] Arash Keshavarzi Arshadi, Milad Salem, Jennifer Collins, Jiann Shiun Yuan and Debopam Chakrabarti : DeepMalaria: Artificial Intelligence Driven Discovery of Potent Antiplasmodials. *Frontiers in pharmacology*. 15th January 2020;doi: 10.3389/fphar.2019.01526. <https://www.frontiersin.org/articles/10.3389/fphar.2019.01526/full> [Accessed: 26th February 2021]
- [10] Monika Samant, Nidhi Chadha, Anjani K. Tiwari, and Yasha Hasija: In Silico Designing and Analysis of Inhibitors against Target Protein Identified through Host-Pathogen Protein Interactions in Malaria. 17 November 2015. *International Journal of Medicinal Chemistry Volume 2016*, Article ID 2741038, 13 pages <http://dx.doi.org/10.1155/2016/2741038> <https://www.hindawi.com/journals/ijmc/2016/2741038/> [Accessed: 27th February 2021]
- [11] Manila Kashyap, Vipin Kumar Sohpal and Parul Mahajan: In silico approaches for inhibitor designing against Plasmepsin-II of malarial parasite, *Plasmodium malariae*. *Biotechnological Communication. Biosci. Biotech. Res. Comm.* 9(1): 25–31 (2016) [https://www.researchgate.net/publication/333822594\\_In\\_silico\\_approaches\\_for\\_inhibitor\\_designing\\_against\\_Plasmepsin-II\\_of\\_malarial\\_parasite\\_Plasmodium\\_malariae](https://www.researchgate.net/publication/333822594_In_silico_approaches_for_inhibitor_designing_against_Plasmepsin-II_of_malarial_parasite_Plasmodium_malariae) [Accessed: 27th February 2021]
- [12] Philipp Ludin, Ben Woodcroft, Stuart A Ralph, Pascal Mäser: In silico prediction of antimalarial drug target candidates. 2012 Jul , *Int J Parasitol Drugs Drug Resist* 2:191–9. doi: 10.1016/j.ijpddr.2012.07.002 <https://pubmed.ncbi.nlm.nih.gov/24533280/> [Accessed: 27th February 2021]
- [13] Bahl, A., Brunk, B., Crabtree, J., Fraunholz, M. J., Gajria, B., Grant, G. R.,



- Ginsburg, H., Gupta, D., Kissinger, J. C., Labo, P., Li, L., Mailman, M. D., Milgram, A. J., Pearson, D. S., Roos, D. S., Schug, J., Stoeckert, C. J., Jr, Whetzel, P. (2003). PlasmoDB: the Plasmodium genome resource. A database integrating experimental and computational data. *Nucleic acids research*, 31(1), 212–215. <https://doi.org/10.1093/nar/gkg081>.
- [14] Ruberto, A.A., Bourke, C., Merienne, N. et al. Single-cell RNA sequencing reveals developmental heterogeneity among Plasmodium berghei sporozoites. *Sci Rep* 11, 4127 (2021). <https://doi.org/10.1038/s41598-021-82914-w>
- [15] Deng Q, Ramskold D, Reinius B, Sandberg R. *Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells*. *Science*. 2014;343:193–6.
- [16] Jaitin DA, Kenigsberg E, Keren-Shaul H, Elefant N, Paul F, Zaretsky I, et al. *Massively parallel single-cell RNAseq for marker-free decomposition of tissues into cell types*. *Science*. 2014;343:776–9.
- [17] Miyamoto DT, Zheng Y, Wittner BS, Lee RJ, Zhu H, Broderick KT, et al. *RNA-Seq of single prostate CTCs implicates noncanonical Wnt signaling in antiandrogen resistance*. *Science*. 2015;349:1351–6.
- [18] Tirosh I, Izar B, Prakadan SM, Wadsworth MH, Treacy D, Trombetta JJ, et al. *Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq*. *Science*. 2016;352:189–96.
- [19] Shalek AK, Satija R, Shuga J, Trombetta JJ, Gennert D, Lu D, et al. *Single-cell RNA-seq reveals dynamic paracrine control of cellular variation*. *Nature*. 2014;510:363–9.
- [20] Stubbington MJ, Lonnberg T, Proserpio V, Clare S, Speak AO, Dougan G, et al. *T cell fate and clonality inference from single-cell transcriptomes*. *Nat Methods*. 2016;13:329–32.
- [21] Zeisel A, Munoz-Manchado AB, Codeluppi S, Lönnerberg P, La Manno G, Juréus A, et al. *Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq*. *Science*. 2015;347:1138–42.
- [22] Blakeley P, Fogarty NM, Del Valle I, Wamaitha SE, Hu TX, Elder K, et al. *Defining the three cell lineages of the human blastocyst by single-cell RNA-seq*. *Development*. 2015;142:3613.
- [23] Treutlein B, Brownfield DG, Wu AR, Neff NF, Mantalas GL, Espinoza FH, et al. *Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq*. *Nature*. 2014;509:371–5.
- [24] Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, et al. *The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells*. *Nat Biotechnol*. 2014;32:381–6.
- [25] Petropoulos S, Edsgard D, Reinius B, Deng Q, Panulä SP, Codeluppi S, et al. *Single-cell RNA-seq reveals lineage and X chromosome dynamics in human preimplantation embryos*. *Cell*. 2016;167:285.
- [26] Lonnberg T, Svensson V, James KR, Fernandez-Ruiz D, Sebina I, Montandon R, et al. *Single-cell RNA-seq and computational analysis using temporal mixture modelling resolves Th1/Tfh fate bifurcation in malaria*. *Sci Immunol*. 2017;2:eaal2192.
- [27] Venteicher AS, Tirosh I, Hebert C, Yizhak K, Neftel C, Filbin MG, et al. *Decoupling genetics, lineages, and microenvironment in IDH-mutant gliomas by single-cell RNAseq*. *Science*. 2017;355:eaai8478.

- [28] Tang F, Barbacioru C, Nordman E, Bao S, Lee C, Wang X, et al. *Deterministic and stochastic allele specific gene expression in single mouse blastomeres*. PLoS One.2011;6:e21208.
- [29] Reinius B, Mold JE, Ramskold D, Deng Q, Johnsson P, Michaelsson J, et al. *Analysis of allelic expression patterns in clonal somatic cells by single-cell RNA-seq*. Nat Genet. 2016;48:1430–5.
- [30] Shalek AK, Satija R, Adiconis X, Gertner RS, Gaublot JM, Raychowdhury R, et al. *Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells*. Nature. 2013;498:236–40.
- [31] Kim JK, Kolodziejczyk AA, Illicic T, Illicic T, Teichmann SA, Marioni JC. *Characterizing noise structure in singlecell RNA-seq distinguishes genuine from technical stochastic allelic expression*. Nat Commun. 2015;6:8687.
- [32] Kar G, Kim JK, Kolodziejczyk AA, Natarajan KN, Torlai Triglia E, Mifsud B, et al. *Flipping between Polycomb repressed and active transcriptional states introduces noise in gene expression*. Nat Commun. 2017;8:36.
- [33] Liu S, Trapnell C. *Single-cell transcriptome sequencing: recent advances and remaining challenges*. F1000Res. 2016;5:182.
- [34] Katelyn A. Walzer, Hélène Fradin, Liane Y. Emerson, David L. Corcoran, Jen-Tsan Chi. *Latent transcriptional variations of individual Plasmodium falciparum uncovered by single-cell RNA-seq and fluorescence imaging*. December 19, 2019, <https://doi.org/10.1371/journal.pgen.1008506>.
- [35] Virginia M. Howick, Andrew J. C. Russell, Tallulah Andrews, Haynes Heaton, Adam J. Reid, Kedar, The Malaria Cell Atlas: Single parasite transcriptomes across the complete Plasmodium life cycle, Science 23 Aug 2019: Vol. 365, Issue 6455, eaaw2619, DOI: 10.1126/science.aaw2619.
- [36] Reid AJ, Talman AM, Bennett HM, et al. *Single-cell RNA-seq reveals hidden transcriptional variation in malaria parasites*. Elife. 2018;7:e33105. Published 2018 Mar 27. doi:10.7554/eLife.33105.
- [37] Xishuang Dong, Shanta Chowdhury, Uboho Victor, Xiangfang Li, Lijun Qian, *Cell Type Identification from Single-Cell Transcriptomic Data via Semi-supervised Learning*, 6 May 2020 in arXiv. url: <https://arxiv.org/abs/2005.03994>.
- [38] Jian Hu, Xiangjie Li, Gang Hu, Yafei Lyu, Katalin Susztak, Mingyao Li, *Iterative transfer learning with neural network for clustering and cell type classification in single-cell RNA-seq analysis*, February 03, 2020. doi: <https://doi.org/10.1101/2020.02.02.931139> Nature Machine Intelligence doi: 10.1038/s42256-020-00233-7
- [39] Zhou, L.Q. et al. (2019) *Artificial intelligence in medical imaging of the liver*. World J. Gastroenterol. 25, 672–682
- [40] Ho, C.W.L. et al. (2019) *Governance of automated image analysis and artificial intelligence analytics in healthcare*. Clin. Radiol. 74, 329–337
- [41] Samui, P. and Kothari, D.P. (2011) *Utilization of a least square support vector machine (LSSVM) for slope stability analysis*. Sci. Iran. 18, 53–58
- [42] Sanchez-Lengeling, B. and Aspuru-Guzik, A. (2018) *Inverse molecular design using machine learning: generative models for matter engineering*. Science 361, 360–365
- [43] Tyrchan, C. and Evertsson, E. (2017) *Matched molecular pair analysis in short: algorithms, applications and limitations*. Comput. Struct. Biotechnol. J. 15, 86–90

- [44] Warner, D.J. et al. (2010) WizePairZ: a novel algorithm to identify, encode, and exploit matched molecular pairs with unspecified cores in medicinal chemistry. *J. Chem. Inf. Model.* 50, 1350–1357
- [45] Lapins, M. et al. (2018) A confidence predictor for logD using conformal regression and a support-vector machine. *J. Cheminform.* 10, 17.
- [46] Mayr, A. et al. (2016) DeepTox: toxicity prediction using deep learning. *Front. Environ. Sci.* 3, 80.
- [47] Chan, H.C.S. et al. (2019) New binding sites, new opportunities for GPCR drug discovery. *Trends Biochem. Sci.* 44, 312–330.
- [48] Chan, H.C.S. et al. (2018) Exploring a new ligand binding site of G protein-coupled receptors. *Chem. Sci.* 9, 6480–6489.
- [49] Kufareva, I. et al. (2014) Advances in GPCR modeling evaluated by the GPCR Dock 2013 assessment: meeting new challenges. *Structure* 22, 1120–1139.
- [50] Yang, Z. et al. (2012) UCSF Chimera, MODELLER, and IMP: an integrated modeling system. *J. Struct. Biol.* 179, 269–278.
- [51] Cavasotto, C.N. and Phatak, S.S. (2009) Homology modeling in drug discovery: current trends and applications. *Drug Discov. Today* 14, 676–683.
- [52] Wang, M. et al. (2018) Predicting relative binding affinity using nonequilibrium QM/MM simulations. *J. Chem. Theory Comput.* 14, 6613–6622
- [53] Hayik, S.A. et al. (2010) A mixed QM/MM scoring function to predict protein–ligand binding affinity. *J. Chem. Theory Comput.* 6, 3079–3091
- [54] Ryde, U. (2016) QM/MM calculations on proteins. *Methods Enzymol.* 577, 119–158
- [55] Smith, J.S. et al. (2017) ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost. *Chem. Sci.* 8, 3192–3203
- [56] Zhang, Y.J. et al. (2018) The potential for machine learning in hybrid QM/MM calculations. *J. Chem. Phys.* 148, 241740.
- [57] Faber, F.A. et al. (2016) Machine learning energies of 2 million elpasolite (ABC2D6) crystals. *Phys. Rev. Lett.* 117, 135502.
- [58] Steiner, S. et al.: Organic synthesis in a modular robotic system driven by a chemical programming language. 2019, *Science* 363, eaav2211 <https://science.sciencemag.org/content/363/6423/eaav2211#:~:text=CONCLUSION,robotic%20platform%20for%20organic%20synthesis>. [Accessed: 21th February 2021]
- [59] Ramsundar, B. et al.: Deep Learning for the Life Sciences, 2019, O'Reilly Media. <https://www.oreilly.com/library/view/deep-learning-for/9781492039822/> [Accessed: 16 November 2016] [Accessed: 15th February 2021]
- [60] Xu, Y. et al.: Demystifying multitask deep neural networks for quantitative structure–activity relationships. 2017, *J. Chem. Inf. Model.* 57, 2490–2504. <https://pubs.acs.org/doi/10.1021/acs.jcim.7b00087> [Accessed: 15th February 2021]
- [61] Wang, C. and Zhang, Y.: Improving scoring-docking-screening powers of protein–ligand scoring functions using random forest., 2017, *J. Comput. Chem.* 38, 169–177 <https://pubmed.ncbi.nlm.nih.gov/27859414/> [Accessed: 15th February 2021].

- [62] Stork, C. et al., Hit Dexter 2.0: machine-learning models for the prediction of frequent hitters., 2019, J. Chem. Inf. Model. 59, 1030–1043. <https://pubs.acs.org/doi/10.1021/acs.jcim.8b00677> [Accessed: 15th February 2021].
- [63] Duvenaud, D.K. et al.: Convolutional networks on graphs for learning molecular fingerprints., 2015, In Advances in Neural Information Processing Systems (Vol. 28) (Cortes, C., et al., eds), pp. 2224–2232, NIPS Foundation. <https://arxiv.org/abs/1509.09292> [Accessed: 10th February 2021].
- [64] Durrant, J.D. and McCammon, J.A.: NNScore 2.0: a neural-network receptor–ligand scoring function., 2011, J. Chem. Inf. Model. 51, 2897–2903.
- [65] Wojcikowski, M. et al. (2015) Open Drug Discovery Toolkit (ODDT): a new open-source player in the drug discovery field. J. Cheminform. 7, 26.
- [66] Benjamin, S-L. et al. (2017) Optimizing distributions over molecular space. An objective-reinforced generative adversarial network for inverse-design chemistry (ORGANIC). ChemRxiv Published online August, 17, 2017. <https://chemrxiv.org/articles/ORGANIC<sub>1p</sub>df/5309668>.
- [67] Feinberg, E.N. et al. (2018) PotentialNet for molecular property prediction. ACS Cent. Sci. 4, 1520–1530
- [68] Awale, M. and Reymond, J.L. (2019) Polypharmacology browser PPB2: target prediction combining nearest neighbors with machine learning. J. Chem. Inf. Model. 59, 10–17
- [69] Olivecrona, M. et al. (2017) Molecular de-novo design through deep reinforcement learning. J. Cheminform. 9, 48
- [70] Coley, C.W. et al. (2018) SCScore: synthetic complexity learned from a reaction corpus. J. Chem. Inf. Model. 58, 252–261
- [71] Yasuo, N. and Sekijima, M. (2019) Improved method of structure-based virtual screening via interaction-energy-based learning. J. Chem. Inf. Model. 59, 1050–1061
- [72] Batra, Kushal; Zorn, Kimberley M.; Foil, Daniel H.; Minerali, Eni; Gawriljuk, Victor O.; Lane, Thomas R.; et al. (2020): Quantum Machine Learning for Drug Discovery. ChemRxiv. Preprint. <https://doi.org/10.26434/chemrxiv.12781232.v1>.