A Look Into Combinatorial Optimization Problems in Pharma

Is There A Market For Combinatorial Optimization Applications In Pharma?

Ashwhin Kaliyaperumal | January 13, 2026

Photo by Jeff Trierweiler on Unsplash

Introduction

Developing a new drug is difficult - it requires significant funding to find potential candidate molecules and prove, through rounds of clinical trials, that they are safe for use and effective [31]. A major bottleneck in this process is research and development (R&D).

Before a drug can even go into trial, one needs to first design that drug. There is an immense amount of research needed beforehand to choose exactly what molecule that drug should be. Historically, this process has required many trial-and-error feedback loops in order to identify effective leads and further supply chains . For instance, one can find a molecule that has a certain property and can tweak it

slightly to give it slightly different benefits, while keeping the core property intact [25].

Unfortunately, clinical trials are both expensive and time-consuming [13] [31]. Between setting up the trial itself - finding ideal patients that both have the condition that a drug aims to cure and would benefit from that drug (versus established medications), developing dosing schedules, and ensuring the trials - and accurately completing piles of regulatory paperwork (and the costs to get an FDA approved trial or similar to ensure efficacy in the first place), running clinical trials for every single potential molecule is not practical. With an expensive feedback loop and a low approval rate, pharmaceutical companies need to ensure that the compounds they take to trial have the highest likelihood of succeeding [31].

If it were possible to simulate everything from initial lead generation to clinical trials in advance and more accurately estimate a drug's likelihood of success, without all of the overhead of running multiple physical clinical trials, the cost to develop new drugs would go down drastically - providing newer opportunities to test drugs to target more illnesses and reducing the costs of these drugs, making them more accessible to more people and for a wider array of ailments.

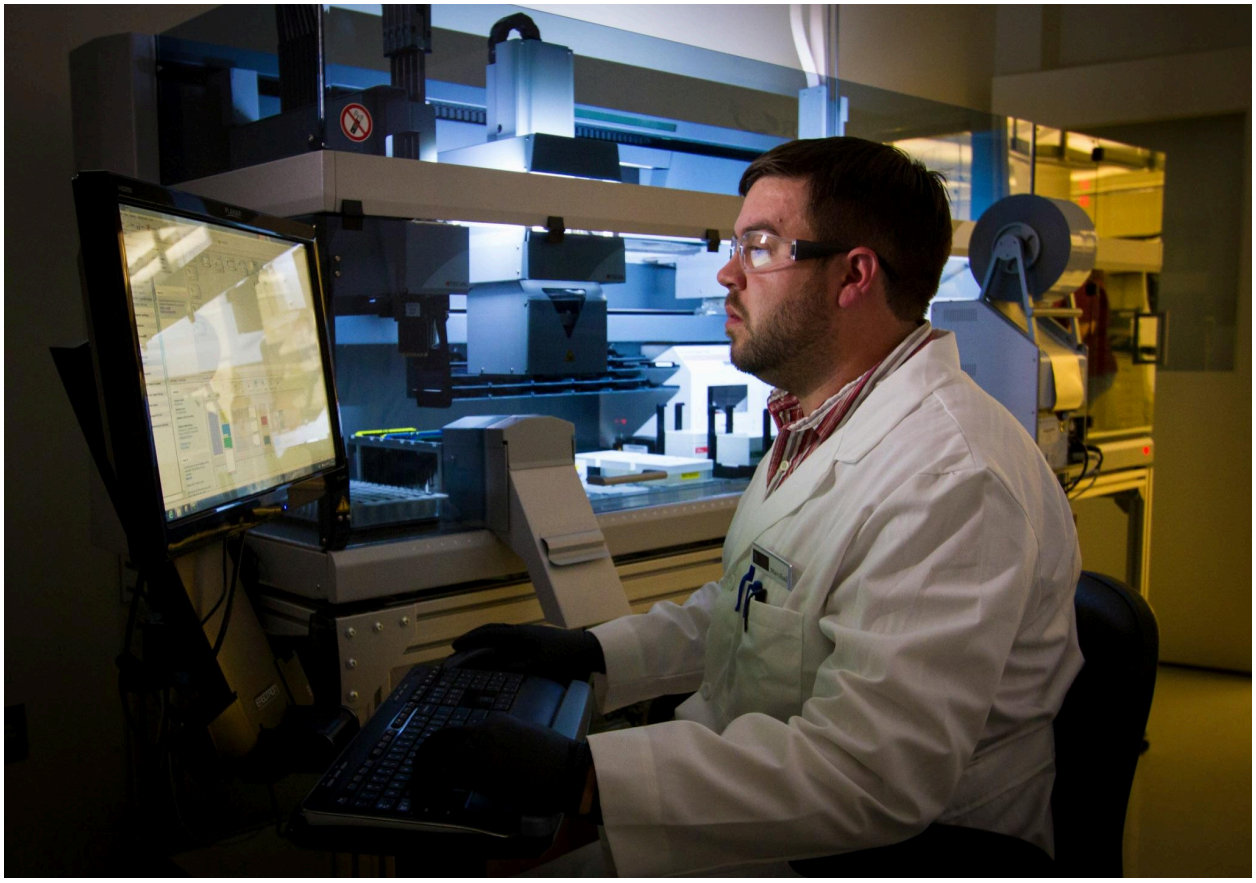


Photo by National Cancer Institute on Unsplash

This is where computational methods help. They provide a digical, *in silico*, way of "testing" various aspects of the drug-discovery pipeline - hopefully, giving us a faster and a dramatically cheaper way to model biological processes as accurately as possible.

Currently, pharmaceutical companies have been working towards what they believe to be AI-powered solutions to optimize their drug discovery, testing, and distribution processes. Quantum technologies, and especially quantum computing, however, have the potential to provide speedups to all of these sectors - especially to problems that can be written as combinatorial optimization formulations.

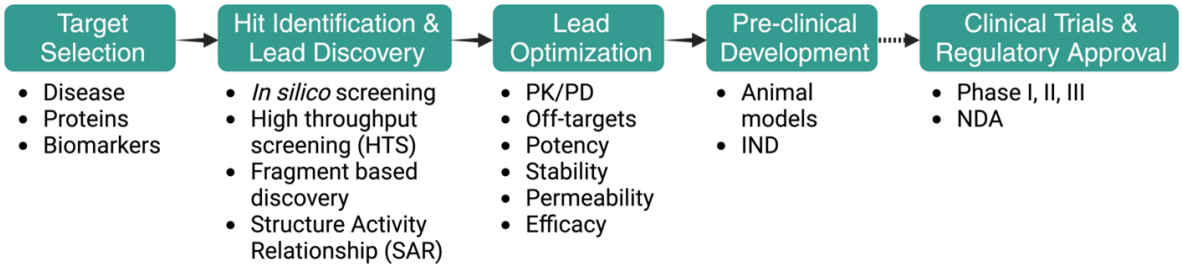
In this article, I will be covering the methodologies and functionalities behind approaches towards solving four relevant combinatorial optimization problems:

molecular docking, pharma-based supply chain optimization, drug synergy and antagonism prediction, and clinical trial design.

Current Drug Discovery Pipeline Process

According to Araujo *et al.*, there are five main stages of the drug discovery pipeline [33]: 1) identification of the drug target, 2) hit identification and lead discovery, 3) lead optimization, 4) pre-clinical development/wet-lab testing, and 5) clinical trials.

The first stage, identification of the drug target, is the stage that outlines the initial definition of the problem. The drug target is the protein or gene associated with an ailment or disease - this helps narrow the search. This step translates a vague goal of preventing some illness to the biological mechanism for how to do so. To assist in measuring the success of potential drug-candidates, the target selection stage also lists out biomarkers, which are biological/chemical/physical measurable signs of a drug's effectiveness that allow clinical trial scientists to quantify the impact of a drug candidate.

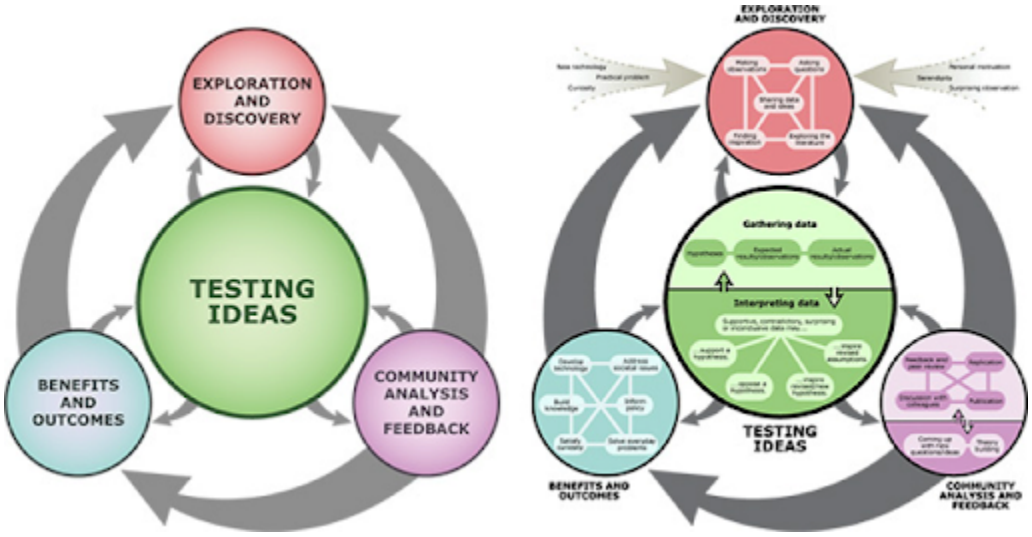


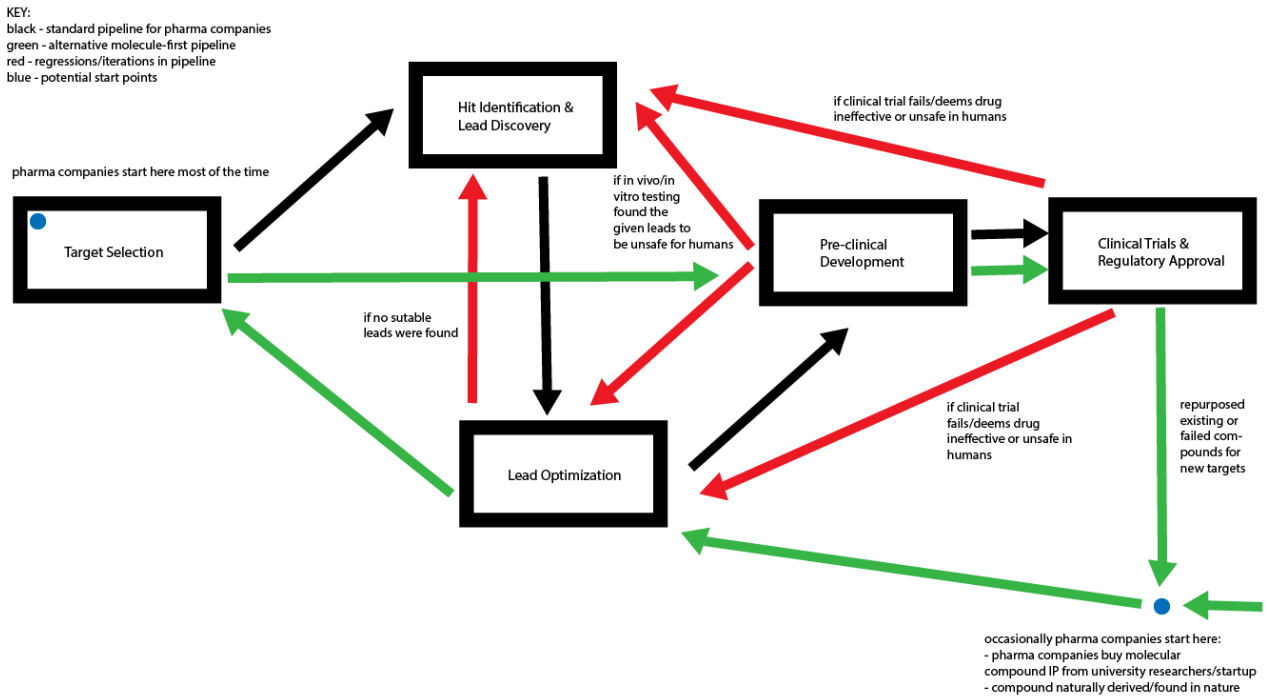
The second stage, hit identification and lead discovery, is where a large number of compounds/trial molecules are vetted and filtered to generate a feasible candidate set of potential drugs. This is the step where the actual drug, the molecules that will eventually make up the final drug, get discovered.

This is done through a variety of techniques, both computational and wet-lab testing, to sort through millions of compounds [34]. Some of these techniques include *in silico* screening, where more accurate and faster computational methods are developed for high throughput screening, fragment based discovery, Structure Activity Relationship, tissue cultures, cell-based arrays, compound management, and array ready plates.

Lead optimization - the third stage of the drug discovery process - takes the filtered list of potential molecules, all that have been confirmed to have some potential relationship with the target, and further tests them to ensure they have the desired effect. Pharmacokinetics (PK) and Pharmacodynamics (PD) are two methods in studying the properties of a drug once absorbed/adsorbed in a patient. Pharmacodynamics focuses on how a drug affects the organism, through chemical binding to receptors and the immediate and post effect that binding has on the organism. Pharmacokinetics, on the other hand, focuses on how an "organism affects the drug", figuring out how the chemical gets digested and transformed, from when the drug enters the body to when it gets eliminated (either completely dissolved or excreted).

This is also a good time to note that the drug discovery process, unlike how it's depicted in the diagram above, is not always a linear process. There are many times where the lead optimization stage finds that none of the potential leads found in the lead discovery phase were viable - in which case, a new set of leads needs to be created. This back and forth process of real-world drug creation schedules more closely mimics the scientific process, shown below, rather than the linear diagram shown above.





Caption - Modified drug development pipeline that accounts for the non-linear flow of work and multiple start points of a potential development pipeline

After viable drug candidates are found, there is still one more stage before clinical trials - pre-clinical development.

This stage is crucial to ensure that a drug candidate is safe for human consumption. More accurately, this stage ensures that a drug can't cause serious harm in people. There are two main methods this stage uses: *in vivo*, testing in living organisms, or *in vitro*, testing outside of a living organism.

This is also where the dosage formulation is developed. Pharma companies gather data on how a drug is absorbed, distributed, metabolized, and excreted (ADME) as well as the best way to deliver the drug. PK/PD, as mentioned above, can also be used in the pre-clinical phase to determine the safety of a drug.

The final stage before a drug can be sold is clinical trials and regulatory approval. To ensure that all drugs are safe for human consumption, researchers/pharmaceutical

companies need to go through rigorous testing, with clinical trials and results approved by regulatory bodies (in the US, this is the Food and Drug Administration).

The clinical trial and regulatory approval phase is a pretty interwoven phase, where the clinical trial process and results needs to be approved of by the trial location country's regulatory body - this means that the trials need to be scientifically sound, have a large enough sample set, and that the results need to indicate safety and efficacy.

Clinical efficacy is the measure of how well a drug does what it's supposed to do. This is less of a concern for the FDA (the FDA mainly ensures the safety of the drug when given to humans, and adherence to clinical trial endpoints) [40].

This stage itself has three typical phases. mirroring the three main clinical trials pharmaceutical companies need to complete: Phase I, Phase II, and Phase III.

Each of the phases both builds up in scale and tests different questions [35]. Phase I, just like with the preclinical trial, focuses on ensuring a drug candidate is safe - testing the dosage formulation from the preclinical trial and drug itself on, typically healthy, human volunteers. Phase II focuses on the effectiveness of a drug, testing with sample sizes of a few hundred volunteers with the target disease, to see if the drug candidate actually affects the drug target (and reduces its related disease). Phase III increases the scale of testing, ranging from sample sizes of hundreds to several thousand patients with the target disease. This trial expands on the goal of Phase II, while providing the information needed for FDA approval of the drug.

One of the main concerns that pharma companies have for conducting clinical trials is their cost, especially considering that only around 30% of drugs pass all 3 stages. According to Martin *et. al*, the median trial cost for each respective phase is $3.4 million, $8.6 million, and $21.4 million - in total, this would be $33.4M for a phase III clinical trial for a single drug [31]. Obviously, this makes it infeasible to clinically test hundreds of these drugs, which emphasizes the importance of accurate lead identification and generation.

<u>Combinatorial Optimization</u>

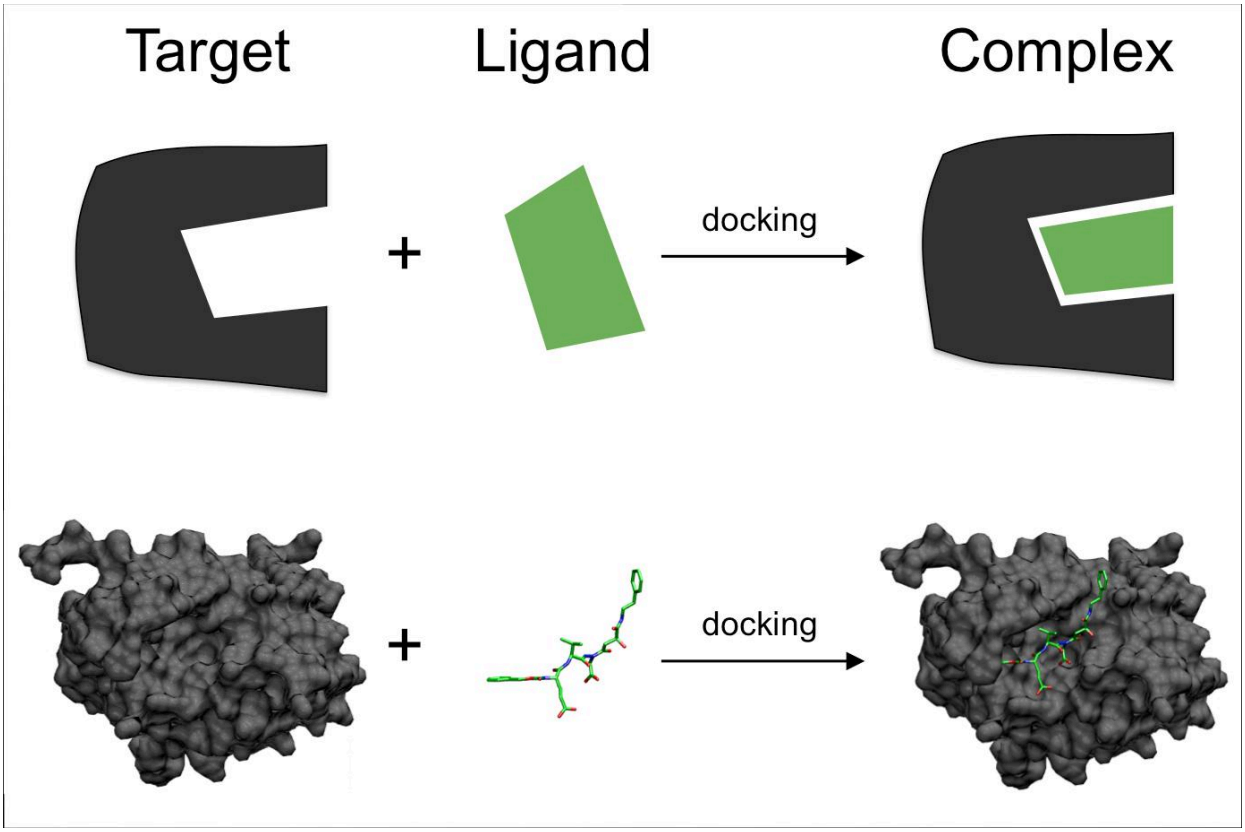
Combinatorial optimization is the method of formalizing a problem to find an optimal solution from a finite set of solutions.

A simple example of a combinatorial optimization problem is, given a list of numbers, finding the sorted list. Our set of solutions would be all possible combinations of numbers in the list and our optimal solution would be the list of sorted numbers.

The main goal with formalizing a problem is to use computers to search through the solution space as, with many practical problems, the solution space is often enormous.

In the drug discovery pipeline, the main type of solution space are the molecular compounds that could potentially bind to a target, with each type of formalized problem “ranking” the effectiveness of these compounds to do a certain task. There are many different methods and formalized problems that quantify and optimize for one type of metric, like optimal ligand pose or structure alignment to a known molecule. There are, however, other solution spaces that exist towards the end of the drug development process. For example, formalizing clinical trial schedules encompasses a solution space that must optimize for the cost and efficiency of a trial.

Combinatorial Optimization Problem #1: Molecular Docking



Caption - Diagram from [32], visually showing the "lock-and-key" model of protein-ligand docking interactions. Just like how a key fits perfectly into the grooves of a lock, the ligand fits perfectly into the pocket of the protein.

The Molecular Docking problem, assumed to be an NP-Hard problem [6], focuses on predicting how a ligand, a small molecule, binds to a protein's pocket, the binding site of the protein.

A protein's shape dictates its function. Certain proteins have binding pockets to accomplish the task they were designed for. For example, cationic trypsin, a protease, helps digest food by breaking down proteins into smaller sets of amino acids, which are the building blocks for other proteins [26][27]. Even though cationic trypsin is an enzyme, its job only focuses on facilitating the breakdown of other proteins and peptides [36][37][38], not other molecules (like lipids or inorganic molecules). The way that cationic trypsin does this is by having the protein, that is going to be broken down, bind to trypsin's pocket. Then, the trypsin pocket uses hydrolysis, the addition of a water molecule (H2O) to separate a molecule into two

parts through the donation of a H+ to one part and HO to the other part, to break down the protein.

The key part of this process is the binding pocket and its binding sites - these facilitate the entire process by providing a place for both proteins to interact with each other. If a different molecule binds to cationic trypsin's pocket, it could interfere with the protein's job. By changing the shape and conformal geometry of the protein through site-specific intermolecular electronic contacts during docking, this small molecule changes the function of cationic trypsin. This can make the protease incompetent towards cleaving proteins into a transiently or permanently dysfunctional state, i.e., protease inhibition.

Therefore, this binding site almost acts like a switch for proteins, which is extremely useful in stopping the function of certain proteins - such as the HIV-1 Protease, the protein that aids in HIV reproduction in cells [27][28]. While cationic trypsin is helpful to humans by aiding in digesting food, the HIV-1 Protease protein cleaves Gag and Pol polyproteins, long chains of amino acids, into individual, smaller functional proteins into a form that aids in its reproduction. HIV-1 Protease creates the building block to replicate. If drug companies could find a different molecule that binds to HIV-1 Protease's pocket and stays there, then we eliminate that protein's function. It's like throwing a wrench into the engine of a moving car - we can get the car to stop dead in its tracks.

For pharmaceutical companies, the molecular docking problem is a formulation that allows us to calculate if a small molecule, a ligand, will bind to a protein of interest.

Right now, there are a lot of existing softwares that perform molecular docking between proteins and ligands - AutoDock Vina, rDock, and SEED are the major molecular docking softwares [23].

Two quantum algorithmic approaches to solve the Molecular Docking problem (with some designed to be run on three types of quantum processor units (QPUs)) are 1) the Quantum Approximate Optimization Algorithm (QAOA), largely run on gate-based quantum computers [4] [14], and 2) Quadratic Unconstrained Binary

Optimization (QUBO) formulations, run on quantum annealers [18], and continuously variable quantum devices.

$$\mathcal{H}_{qubo} = AH_{iso} + H_{opt} = A\sum_{i}(1 - \sum_{i'} x_{i,i'})^2 + A \sum_{i,j \in G_{mol}} \sum_{i',j' \notin G_{grid}} x_{i,i'}x_{j,j'} + B \sum_{i,j \in G_{mol}} \sum_{i',j' \in G_{grid}} (w_{i,j} - w_{i',j'})^2 x_{i,i'}x_{j,j'}$$

Caption - Molecular docking objective function formulation, where H_iso is the sub-graph isomorphism term that ensures hard constraints are met to get valid outputs and H_opt is the optimization term the simulated/quantum annealers minimize over. A and B are scaling constants, usually experimentally determined, to ensure hard constraints are met and x are binary variables that represent the measured state of the qubits [18].

Combinatorial Optimization Problem #2: Pharma-based Supply Chain Optimization

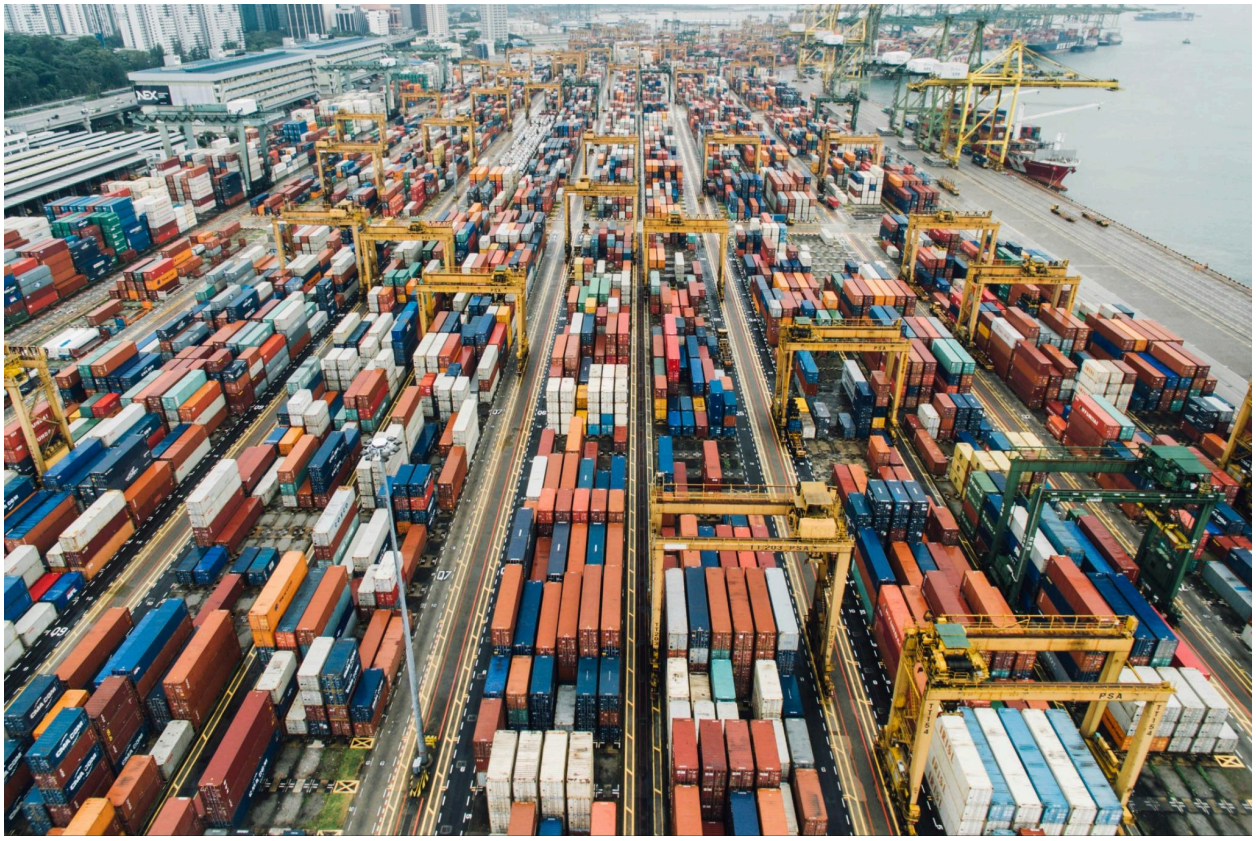


Photo by CHUTTERSNAP on Unsplash

One of the most important stages of creating drugs is getting both the final products and their subcomponents to their final destinations - optimizing pharmaceutical supply chains is crucial in ensuring that drugs get to where they need to do 1) safely and 2) quickly [11].

There are a couple of points of view for supply chain management. In this article, I'll only focus on two of these POVs: the manufacturing side, and the hospital/end-user side.

From the point of view of large, multinational pharmaceutical companies, the ones that manufacture and distribute products, they're focused primarily on answering one main question: How do we make sure the drugs get to the end patient in the shortest amount of time for the lowest cost? These companies also have constraints they're held accountable for - to make these deliveries safely, they need to make sure they're complying with regulations during shipping and, if needed, using

refrigeration units or other specialized transport methodologies required for different therapeutic deliveries to keep their contents from spoiling (which are speciality shipping containers, refrigeration units are a very finite resource because they need to be kept near electricity plugs throughout their journey) [15]. For companies like Sinopharm, they're especially focused on developing a green supply chain - minimizing carbon output throughout their entire supply chain, from product creation to end user use [20].

Hospitals, however, are focused on balancing consumption with cost [1]. They want to answer the following question: how do we balance the inventory needs and storage costs of hospitals? Hospitals, of course, need to make sure they're always stocked on the drugs they need - a shortage could cost someone's life [16]. However, just like in transport, storing these drugs is expensive. They need to be kept in a safe environment and the drugs need to be used before their expiration date; many U.S. hospitals are also not legally allowed to stockpile or independently manufacture drugs to protect suppliers' pricing and distribution schemes, forcing them to resort to overly-fragile Just-In-Time supply management. Many hospitals outsource these responsibilities to third-party storage companies - adding to the many expenses a hospital has.

Photo by RDNE Stock project:
https://www.pexels.com/photo/medical-supplies-on-shelves-6519900/

Large, multinational pharmaceutical companies have already started developing solutions to these problems. Sinopharm, since 2020, has talked about their progress to creating a near zero-carbon supply chain in their yearly Sustainability Reports - they've released guidelines on "Green Supply Chain Construction" and for reducing carbon output in constructing and managing supply chains through optimization techniques in some internal whitepapers [39]. Merck and Co is looking into what they are characterizing as AI solutions for supply chain resilience [8] - making sure that, even in the face of disruptions, drugs can still get to where they need to go on-time, though Merck and Co have not stated how exactly they would use AI or other technologies to accomplish this goal. Johnson and Johnson is working on a solution for supply/demand prediction and for supply chain optimizations to help treatments reach patients faster.

The green supply chain optimization formulation by Umme *et al.* [20] focuses on a case study - the researchers worked with a pharmaceutical company in Bangladesh to reduce their carbon emissions throughout the entire supply chain. Umme *et al.*'s work formulated a model to cover the journey from initial chemical supplier to manufacturer to distributor to retailer. Their work in creating a mixed-integer linear programming formulation provides an interesting gateway into a QUBO formulation of the same problem. It would be interesting to look into the QUBO's ability to keep into account the different hard constraints on the problem, and how to balance the scaling parameters of each hard constraint to make sure the objective function is still being met while generating valid solutions.

$$\begin{aligned}\text{Min } Z1 = &\sum_{m=1}^{M}\sum_{p=1}^{P}\sum_{t=1}^{T} K_{mpt}.d_{mpt} + \sum_{d=1}^{D}\sum_{p=1}^{P}\sum_{t=1}^{T} H_{dpt}.\left(\sum_{\tau=1}^{t}\sum_{m=1}^{M} b_{mdpt} - \sum_{\tau=1}^{t}\sum_{d=1}^{D} c_{drpt}\right) \\ &+ \sum_{s=1}^{S}\sum_{m=1}^{M}\sum_{p=1}^{P}\sum_{t=1}^{T} R_{smpt}.a_{smpt} + \sum_{m=1}^{M}\sum_{d=1}^{D}\sum_{p=1}^{P}\sum_{t=1}^{T} U_{mdpt}.b_{mdpt} \\ &+ \sum_{d=1}^{D}\sum_{r=1}^{R}\sum_{p=1}^{P}\sum_{t=1}^{T} V_{drpt}.c_{drpt} + \sum_{r=1}^{R}\sum_{p=1}^{P}\sum_{t=1}^{T} S_{rpt}.e_{rpt} + \sum_{p=1}^{P}\sum_{t=1}^{T} f_{pt}.A_{p}.\end{aligned} \quad (1)$$

$$\text{Min } Z2 = \sum_{m=1}^{M}\sum_{d=1}^{D}\sum_{p=1}^{P}\sum_{t=1}^{T} CO2_{mdpt}.b_{mdpt} + \sum_{d=1}^{D}\sum_{r=1}^{R}\sum_{p=1}^{P}\sum_{t=1}^{T} CO2'_{drpt}.c_{drpt}. \quad (2)$$

Caption - Z1 optimizes over the total costs throughout the entire supply chain (from supply, production, distribution storage, and all transportation connections between the supply chain). Z2 minimizes the total carbon emissions throughout the supply chain [20].

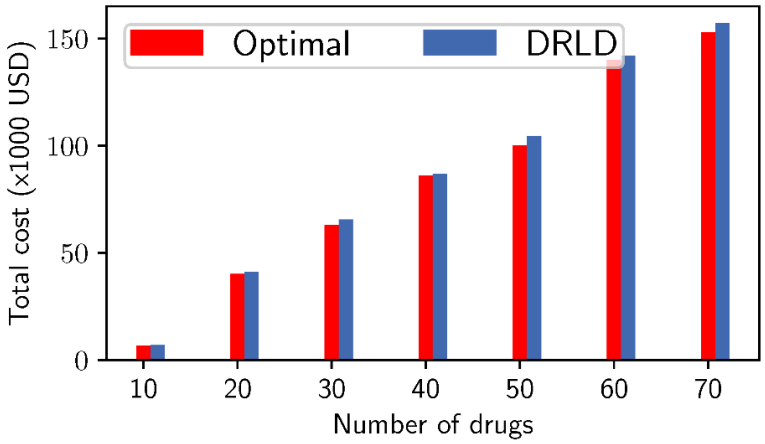
The formalized problem for hospital supply chain optimization is the Dynamic Filling Drug Optimization problem. The objective function for this problem has three main parts: the 1) medicine cost, mainly the cost of buying medicine weighted by the priority of the type of drug being bought, the 2) storage cost (the cost of storing medicine safely), and a 3) penalty cost to prevent supply shortages [1].

$$\begin{aligned}
\min_{\mathbf{x}} \quad & \sum_{t=1}^{T}\sum_{i\in\mathcal{I}} \alpha_i p_i(t)\, x_i(t) + R_i(v_i(t), r_i(t)) + \psi(v_i(t)) \\
\text{s.t} \quad & \underline{C}_i(t) \le v_i(t) = v_i(t-1) - \lambda_i(t) - e_i(t) + x_i(t) \le C_i(t), \forall i \in \mathcal{I}, t = 1,2,\ldots,T, \\
& \sum_{t=1}^{T}\sum_{i\in\mathcal{I}} x_i(t)\, p_i(t) \le \mathbb{B}, \\
& x_i(t)\, p_i(t) \le B_i(t), \forall i \in \mathcal{I}, \\
& \underline{\rho}_i(t) \le x_i(t) \le \overline{\rho}_i(t), \forall i \in \mathcal{I}, t = 1,2,\ldots,T.
\end{aligned}$$

Caption - Tarek et al. formulation of the Dynamic Filling Drug Optimization problem with the medicine, storage, and shortage penalty costs added to the minimization function (the first line), and the lowest mandatory drug, budget, and timeslot constraints imposed on the problem listed in the lines below [1].

Tarek *et al.* relate the Dynamic Filling Drug Optimization problem with the formalized Ski-Rental Problem. We have a person who wants to ski, but doesn't know how long the snow season will last; should one buy or rent skis to get the most for their money? The Dynamic Filling Drug Optimization problem is especially challenging because we don't know the user's demand in advance, making shortage prediction more difficult

For these reasons, DFDO appears to be embeddable onto the Ski-Rental Problem because both can be represented by the more famous Knapsack problem; given an algorithm that perfectly finds solutions to the Ski-Rental problem (given that you have the inputs to the ski-rental problem in the right format), you can do some quick (polynomial-time) pre-processing for inputs of the DFDO problem to transform DFDO inputs into Ski-Rental formatted inputs, run the Ski-Rental algorithm, and then do some trivial post-processing to turn those Ski-Rental solutions into DFDO solutions



(c) Evaluation of the total cost.

Caption - Comparison of optimal solution's cost with the Deep Reinforcement Learning (DRLD) model's output cost from [1].

The solution by Tarek *et al.* was to use a deep reinforcement learning approach to solve this problem to prevent these shortages. After experimental testing to determine the learning rate, their algorithm converged on a solution that was very close to the optimal solution, though they did not provide precise bounds between the reinforcement approach's output and the optimal solution.

Combinatorial Optimization Problem #3: Predicting Drug Synergy and Antagonism

Photo by Mary Borozdina on Unsplash

Drugs, when combined, don't always act in the same way as the sum of their parts could lead one to believe. Certain drugs can have strong synergistic reactions, where their effects are more than the sum of their parts, or they can have antagonistic effects, canceling each other's effects out (i.e., they're non-linear).

Predicting drug synergy is the key to combining drugs for many medications, done with pharmacokinetics/pharmacodynamics. It could allow two existing drugs to provide a greater reaction, while limiting the side effects that large doses of those drugs could have individually. The combination achieves the same end result in a safer manner [10]. Drug antagonist is the opposite reaction - where one drug inhibits, or stops, the effect of another drug. Paying attention to these antagonistic relationships ensures that we do not have two drugs that should be targeting different ailments targeting each other [21].

Finding these synergies and antagonisms has historically been through trial and error. AstraZeneca's DREAM challenge [12], established in 2014, created a community effort to computationally predict the effects of drug combinations; the hope was to create an accurate method to drastically reduce the number of combinations that needed expensive and time-consuming wet lab testing to prove these drug synergies and antagonisms.

The challenge was split up into two subchallenges, with the first objective function was the average weighted Pearson correlation of the predicted vs observed synergy scores across each drug combination. The second objective function was optimizing the separation of predicted synergy and non-synergy scores through the ANOVA score - which has a linear regression formulation that accounts for the drug combination, cell line, and the binary synergy prediction vector [3].

$$\rho_w = \frac{\sum_{i=1}^{N} \sqrt{n_i - 1}\,\rho_i}{\sum_{i=1}^{N} \sqrt{n_i - 1}}, \tag{1}$$

$$y \sim \beta_1 \cdot \mathrm{dc} + \beta_2 \cdot \mathrm{cl} + \beta_3 \cdot x \tag{3}$$

Caption - Standard form and linear regression formulation of the ANOVA formula [12]

The top scoring team in all sub-challenges, Guan *et al*. from the University of Michigan, was a novel network propagation [9], that inputted the information from the individual drug's effects and simulated molecular data, and used random walks to transfer information throughout the algorithm's saved network. They were able to accurately predict the best synergistic drug pairs in 30% of the cell lines and were able to correctly place the best synergistic drug pair in the top five predicted pairs in 73% of test cell lines.

Combinatorial Optimization Problem #4: Clinical Trial Design Optimization

Photo by CDC on Unsplash

Clinical trial design is the key step in ensuring that the potential drug candidates pharma companies create are both safe for human use and effective at their proposed jobs in humans. There are many places to optimize clinical trial design, but this article will only focus on three problems that have already been formalized: optimizing the trial site [7], cohort identification, and drug dose scheduling [19].

Trial site optimization chooses the right site, with specific characteristics, for a clinical trial. Cohort identification finds the right people for the trial - the goal is to pick patients that are 1) relevant to the trial (they have the condition you want to target with your drug) and 2) that the trial would benefit (especially if there isn't any other effective treatment alternative for the specific condition the drug is targeting). Drug dose scheduling personalizes the trial for each patient, making sure the trial gives patients the right dose and at the right times to properly understand the effects of the drug.

A few pharmaceutical companies are looking at using AI models to optimize these problems. Johnson and Johnson is using AI to improve targeted clinical trial recruitment - it's finding patients that could benefit from medicine that Johnson and Johnson is making. Further description of what these AI methodologies are and if they include combinatorial optimization have not been publicly disclosed as of now.

Merck and Co has also partnered with McKinsey [8] and its AI arm QuantumBlack to automate Clinical Study Reports w/ LLMs - reducing the time to finish the paperwork behind running clinical trials.

For the algorithmic research on drug dose scheduling, Tsuchiwata *et al.* from Pfizer [19] created a genetic algorithm to optimize the blood sample schedule (a similar problem to the drug dose scheduling problem) for patients in a bioequivalence study in pediatrics.

$$\text{Fitness} = \left(\frac{\text{Number\_of\_samples}}{7}\right)^2 + \text{MAPE}_{C_{\max}} + \text{MAPE}_{\text{AUC}_t} + \text{RMSPE}_{C_{\max}} + \text{RMSPE}_{\text{AUC}_t}$$

Caption - the fitness function for the blood scheduling genetic algorithm as a sum of the mean absolutely percentage error (MAPE) and the root mean square percentage error (RMSPE) [19]

Wang *et al*. developed a three-step optimization strategy [21] to optimize the combination of drug therapies. Their work focuses on both the drug interactions (as talked about in the prior section) but also the individual drug doses, and balancing both to bring out the synergies between the two drugs - though they note that this is only the first step in proving the efficiency and accuracy of combinatorial optimization problems for combinatorial drug therapy.

$$X_{opt} = \arg\max_X E = \arg\max_X f(X) \quad (1)$$

where X is the drug combination input; E is the efficacy output, which can be any measurable and quantifiable parameter; f is the function relation between drug doses and efficacy; and X_{opt} is the optimal combination that we need. Thus, the 'phenotype-driven medicine concept associates combinatorial drug therapy with systems engineering and optimization theories.

Caption - Wang *et al. drug combination optimization formulation based on optimizing phenotype combinations.* [21]

Conclusion

Right now, molecular docking software has had the greatest success in the applicability of combinatorial optimization to pharmaceutical development - with well used and well characterized software, like AutoDesk Vina, providing molecular docking capabilities, capable of running complex protein-ligand complexes from PDBBind+ on home computers in reasonable times.

With regards to better solving these combinatorial optimization problems using quantum algorithms or quantum devices most algorithms, especially around the QAOA/VQE algorithms and QUBO formulations, have not been tested beyond small case studies. For reference, the Molecular Docking QUBO paper [18] by Triuzzi *et al.* tested their cost function on pockets with 15 atoms - resulting in QUBOs with fewer than 100 variables. By comparison, most protein-ligand complexes on PDBBind would have a QUBO size of around 50,000 to 150,000 variables, with the maximum size being closer to 1,000,000 variables.

There has also been research on quantum algorithms for clinical trial design optimization. Doga *et al.* mentions how optimizing trial site selection is extremely similar to portfolio optimization, a problem that has both classical and quantum algorithms for the finance industry. There is also a potential for quantum machine learning for clinical trial simulations, but there are still difficulties in fully simulating human biology - and assumptions and data taken from static measurements (like x-ray crystallography) or animal based models can prove to be too inaccurate for human clinical trials.

While quantum technology literature has been focused on the time and space scaling of their algorithms - on if their algorithms are theoretically faster than classical algorithm and if the number of qubits needed is feasible for future quantum computers, there has been a lack of up-to-date algorithmic testing on quantum hardware, which is quickly growing past the scale of small-scale test cases. Papers on practical quantum algorithms have been shying away from using the growth of

quantum processing units (QPUs) to, in turn, scale the problem test sets on their quantum algorithms. This makes it difficult to both empirically test scaling of quantum algorithms and to compare outputs of quantum algorithms with their state-of-the-art classical algorithm counterparts. But, current improvements in combinatorial optimization, as covered in this post, combined with improvements in quantum algorithms and various aspects of QPU development mean that this is an area that should be greater emphasized in future work, including our own

Thank you for reading my blog post! I'd love to hear if you're working on a biopharma related combinatorial optimization formulation that I didn't mention or if you're working on a better algorithm than the ones I covered!

Works Cited

1. Abu Zwaida, Tarek, et al. "Optimization of Inventory Management to Prevent Drug Shortages in the Hospital Supply Chain." *Applied Sciences*, vol. 11, no. 6, 18 Mar. 2021, p. 2726, https://doi.org/10.3390/app11062726.
2. Alharby, Tareq Nafea, and Bader Huwaimel. "Machine Learning Analysis of Pharmaceutical Cocrystals Solubility Parameters in Enhancing the Drug Properties for Advanced Pharmaceutical Manufacturing." *Scientific Reports*, vol. 15, no. 1, 15 Aug. 2025, pmc.ncbi.nlm.nih.gov/articles/PMC12356857/, https://doi.org/10.1038/s41598-025-12886-8. Accessed 2 Oct. 2025.
3. Chen, Jinghong, et al. "Computational Frameworks Transform Antagonism to Synergy in Optimizing Combination Therapies." *Npj Digital Medicine*, vol. 8, no. 1, 19 Jan. 2025, www.nature.com/articles/s41746-025-01435-2, https://doi.org/10.1038/s41746-025-01435-2. Accessed 4 Aug. 2025.

4. Ding, Qi-Ming, et al. “Molecular Docking via Quantum Approximate Optimization Algorithm.” *ArXiv.org*, 2023, arxiv.org/abs/2308.04098. Accessed 19 Oct. 2025.
5. Dong, Jie, et al. “FormulationAI: A Novel Web-Based Platform for Drug Formulation Design Driven by Artificial Intelligence.” *Briefings in Bioinformatics*, vol. 25, no. 1, 22 Nov. 2023, academic.oup.com/bib/article/25/1/bbad419/7441064, https://doi.org/10.1093/bib/bbad419.
6. García-Godoy, María, et al. “Solving Molecular Docking Problems with Multi-Objective Metaheuristics.” *Molecules*, vol. 20, no. 6, 2 June 2015, pp. 10154–10183, https://doi.org/10.3390/molecules200610154. Accessed 2 June 2022.
7. Hakan Doga, et al. “How Can Quantum Computing Be Applied in Clinical Trial Design and Optimization?” *Trends in Pharmacological Sciences*, 1 Sept. 2024, https://doi.org/10.1016/j.tips.2024.08.005. Accessed 1 Oct. 2024.
8. Kitishian, Dany. “Merck’s AI Strategy: Analysis of Dominating Biopharmaceutical.” *Klover.AI*, 17 July 2025, www.klover.ai/merck-ai-strategy-analysis-of-dominating-biopharmaceutical/. Accessed 19 Oct. 2025.

9. Li, Hongyang, et al. *Network Propagation Predicts Drug Synergy in Cancers*. Vol. 78, no. 18, 15 Sept. 2018, pp. 5446–5457, https://doi.org/10.1158/0008-5472.can-18-0740. Accessed 24 May 2023.
10. Lieberman, H. R., et al. “The Effects of Caffeine and Aspirin on Mood and Performance.” *Journal of Clinical Psychopharmacology*, vol. 7, no. 5, 1 Oct. 1987, pp. 315–320, pubmed.ncbi.nlm.nih.gov/3680601/.
11. Marciano, Michele. “Pharmaceutical Supply Chain Optimization Is Crucial for Ensuring the Efficiency and Reliability of Delivering Medicines to Patients. In This Guide, We Will Explore the Key Strategies and Best Practices for Optimizing the Supply Chain in the Pharmaceutical Industry.” *Linkedin.com*, 2 Jan. 2024, www.linkedin.com/pulse/guide-supply-chain-optimization-pharmaceutical-michele-marciano-ocfcc/.
12. Menden, Michael P., *et al*. “Community Assessment to Advance Computational Prediction of Cancer Drug Combinations in a Pharmacogenomic Screen.” *Nature Communications*, vol. 10, no. 1, 17 June 2019, p. 2674, www.nature.com/articles/s41467-019-09799-2, https://doi.org/10.1038/s41467-019-09799-2. Accessed 17 Dec. 2021.
13. “Performance Improvement in Pharmaceuticals Operations.” *Chemanager-Online.com*, 17 Sept. 2025,

chemanager-online.com/en/topics/performance-improvement-in-pharmaceuticals-operations. Accessed 19 Oct. 2025.

14. “Quantum Approximate Optimization Algorithms for Molecular Docking.” *Arxiv.org*, 2024, arxiv.org/html/2503.04239v1. Accessed 19 Oct. 2025.
15. Settanni, Ettore, et al. “Pharmaceutical Supply Chain Models: A Synthesis from a Systems View of Operations Research.” *Operations Research Perspectives*, vol. 4, no. 1, 2017, pp. 74–95. *Sciencedirect*, www.sciencedirect.com/science/article/pii/S2214716016301105, https://doi.org/10.1016/j.orp.2017.05.002.
16. Slimstock. “Pharmaceutical Supply Chain: Optimizing Operations in the MEA Region.” *Slimstock*, 23 Oct. 2024, www.slimstock.com/blog/pharmaceutical-supply-chain-management-in-the-mea-region/.
17. Thomsen, René. “Flexible Ligand Docking Using Evolutionary Algorithms: Investigating the Effects of Variation Operators and Local Search Hybrids.” *Biosystems*, vol. 72, no. 1-2, Nov. 2003, pp. 57–73, https://doi.org/10.1016/s0303-2647(03)00135-7. Accessed 1 Aug. 2019.
18. Triuzzi, Emanuele, et al. “Molecular Docking via Weighted Subgraph Isomorphism on Quantum Annealers.” *ArXiv.org*, 2024, arxiv.org/abs/2405.06657. Accessed 19 Oct. 2025.

19. Tsuchiwata, Shinichi, and Yasuhiro Tsuji. “Computational Design of Clinical Trials Using a Combination of Simulation and the Genetic Algorithm.” *CPT: Pharmacometrics & Systems Pharmacology*, vol. 12, no. 4, 5 Mar. 2023, pp. 522–531, pmc.ncbi.nlm.nih.gov/articles/PMC10088085/, https://doi.org/10.1002/psp4.12944. Accessed 19 Oct. 2025.
20. Umme Habiba, et al. “Optimization of a Green Supply Chain Network: A Case Study in a Pharmaceutical Industry.” *DOAJ (DOAJ: Directory of Open Access Journals)*, 1 Sept. 2021, https://doi.org/10.22105/riej.2021.297117.1236.
21. Wang, Boqian, et al. “The Optimization of Combinatorial Drug Therapies: Strategies and Laboratorial Platforms.” *Drug Discovery Today*, vol. 26, no. 11, 28 July 2021, pp. 2646–2659, www.sciencedirect.com/science/article/abs/pii/S1359644621003263, https://doi.org/10.1016/j.drudis.2021.07.023.
22. Wikipedia Contributors. “List of Protein-Ligand Docking Software.” *Wikipedia*, Wikimedia Foundation, 10 Aug. 2025, en.wikipedia.org/wiki/List_of_protein-ligand_docking_software.
23. wwPDB.org. “WwPDB: Pdb_00003atl.” *Wwpdb.org*, 2024, www.wwpdb.org/pdb?id=pdb_00003atl. Accessed 19 Oct. 2025.
24. Yamane, Junji, et al. “In-Crystal Affinity Ranking of Fragment Hit Compounds Reveals a Relationship with Their Inhibitory Activities.” *Journal of Applied*

Crystallography, vol. 44, no. 4, 7 June 2011, pp. 798–804, https://doi.org/10.1107/s0021889811017717. Accessed 28 Jan. 2025.

25. Wang, Wei, et al. “Computational Pharmaceutics - a New Paradigm of Drug Delivery.” *Journal of Controlled Release*, vol. 338, Oct. 2021, pp. 119–136, https://doi.org/10.1016/j.jconrel.2021.08.030. Accessed 21 Feb. 2022.
26. https://doi.org/10.2210/pdb3atl/pdb
    a. Cationic Trypsin PDB Citation
27. https://journals.iucr.org/paper?S0021889811017717
    a. Cationic Trypsin Primary Publication Citation
28. https://doi.org/10.2210/pdb1hvr/pdb
    a. HIV-1 Protease PDB Citation
29. https://pubmed.ncbi.nlm.nih.gov/8278812/
    a. HIV-1 Protease Primary Publication Citation
30. https://pmc.ncbi.nlm.nih.gov/articles/PMC10088085/
    a. Computational design of clinical trials using a combination of simulation and the genetic algorithm
31. https://www.nature.com/articles/nrd.2017.70
    a. How much do clinical trials cost?
32. https://en.wikipedia.org/wiki/Docking_(molecular)
33. https://ecampusontario.pressbooks.pub/medicinalchemistry/chapter/how-are-drugs-discovered/

34. https://www.mdpi.com/1422-0067/20/18/4331

    a. Molecular Docking's utility for drug discovery

35. https://med.uc.edu/depart/psychiatry/research/clinical-research/crm/trial-phases-1-2-3-defined

36. *https://www.ncbi.nlm.nih.gov/books/NBK548893/*
37. *https://en.wikipedia.org/wiki/Cysteine_protease*
38. *https://my.clevelandclinic.org/health/treatments/24937-protease-inhibitors*
39. http://ir.sinopharmgroup.com.cn/pdf/2024sr_en.pdf
40. https://www.irishnews.com/news/uk/alzheimers-drugs-rejected-for-nhs-because-benefits-too-small-to-justify-cost-3EN7L3H57BNH3PK3EE6U3A3K6Y/