MDPI

Review

Transforming Drug Design: Innovations in Computer-Aided Discovery for Biosimilar Agents

Shadi Askari [1], Alireza Ghofrani [2] and Hamed Taherdoost [3],*

1   Hamta Group, Hamta Business Corporation, Vancouver, BC V6E 1C9, Canada; shadi@hamta.ca
2   Q Minded, Quark Minded Technology Inc., Vancouver, BC V6E 1C9, Canada; alireza.g@qminded.ca
3   Department of Arts, Communication and Social Sciences, University Canada West,
     Vancouver, BC V6Z 0E5, Canada
*   Correspondence: hamed.taherdoost@gmail.com; Tel.: +1-236-889-5359

Abstract: In pharmaceutical research and development, pursuing novel therapeutics and optimizing existing drugs have been revolutionized by the fusion of cutting-edge technologies and computational methodologies. Over the past few decades, the field of drug design has undergone a remarkable transformation, catalyzed by the rapid advancement of computer-aided discovery techniques and the emergence of biosimilar agents. This dynamic interplay between scientific innovation and technological prowess has expedited the drug discovery process and paved the way for more targeted, effective, and personalized treatment approaches. This review investigates the transformative computer-aided discovery techniques for biosimilar agents in reshaping drug design. It examines how computational methods expedite drug candidate identification and explores the rise of cost-effective biosimilars as alternatives to biologics. Through this analysis, this study highlights the potential of these innovations to enhance the efficiency and accessibility of pharmaceutical development. It represents a pioneering effort to examine how computer-aided discovery is revolutionizing biosimilar agent development, exploring its applications, challenges, and prospects.

Keywords: computational drug design; biosimilar discovery; computer-aided drug development; pharmaceutical innovation; personalized treatment; artificial intelligence

1. Introduction

A drug is a non-natural substance used to treat, diagnose, or prevent disease that influences biological processes [1]. Drugs can be made synthetically or from natural sources. An ideal drug would have a well-defined mechanism of action; be chemically and metabolically stable; be amenable to chemical synthesis; be water-soluble at therapeutic concentrations to prevent precipitation in the bloodstream; be lipid-soluble to facilitate transport across cell membranes; and be a novel molecule [2].

Biopharmaceuticals, particularly therapeutic antibodies, are a class of drugs that are becoming increasingly significant in addition to small molecules, and computational techniques for enhancing the stability, selectivity, and affinity of these protein-based treatments have also made significant strides [3]. Biologics are more expensive than conventional therapy options [4], and these drugs may become less effective with continued usage [5]. Drug survival is the likelihood that patients will continue taking a certain drug. In contrast, therapy cessation can happen for various reasons, the most prevalent of which is treatment ineffectiveness [6].

Various items separated from natural sources or produced using living systems are biological drugs. Drugs made chemically are generally 100–1000 times smaller than biologics, and their molecular structures are more difficult to define [7]. Recombinant DNA technology is used to create several proteins that are used in biologics. Developing a biologic is a multi-step procedure that is technically difficult, confidential, and unique to the producer [8]. As a result, a biosimilar developer needs to independently build a new

production process that can supply a drug that is strikingly comparable to the original through reverse engineering manufacturing [9].

The development of biosimilars has emerged to lower medical expenses and expand patient treatment alternatives [10]. A biosimilar drug is typically described as a biological substance similar to the reference drug. It has no clinically significant deviations in potency, purity, or safety, even though numerous regulatory definitions exist for this term [11]. Biosimilar drugs have a variety of clinical uses, including the treatment of cancer, rheumatic and intestinal illnesses, neutropenia, and psoriasis [12–14]. A fiercely competitive business has developed to create biosimilar drugs due to these treatments' clinical and financial triumphs [15–17].

Drug discovery, preclinical development, and clinical trials are the three key drug research and development phases. A hit molecule is the first step in the drug discovery process. A chemical that causes the required activity in a screening assay is called a hit [18]. Introducing a new medicine to the market takes a lot of money, time, and labor. Drug research and discovery takes 10 to 15 years, costing between USD 800 million [19] and USD 1.8 billion [20]. Because preclinical and clinical data need to be used to support the effectiveness and safety of biosimilars, their development and production are more difficult and expensive than those of synthetic generic drugs. However, it is critical to emphasize that these processes are less expensive than those necessary to create biological drugs [13].

Despite the promising developments in genomics, proteomics, and systems biology, significant scientific and regulatory barriers still prevent the development of effective biologically active agents. As a result, only about 13% of drugs complete the clinical trials stage [21,22]. The time- and money-consuming drug discovery process aims to create new drug candidates [23]. Using computational methods throughout the pre-clinical stage of drug discovery has been one method of accomplishing this [24]. To increase efficiency and widen the feasible window, scientists can design energy harvesting systems and materials thanks to such efforts [25,26].

Computer-aided drug design (CADD) uses computational methods to find, create, and analyze drugs and active compounds with similar biological properties [27,28]. Accordingly, CADD has revolutionized the history of drug discovery, particularly its substantial benefits [29]. These benefits include offering insights into target-drug interactions, utilizing the 3D structure, causing a cost-effective reduction in high-throughput screening failures, inspiring novel drug design concepts, and aiding researchers in predicting targeted proteins and candidate hits [30–32].

This review explores how computer-aided discovery transforms biosimilar agent development by examining its applications, challenges, and future implications. It introduces the biosimilar agents' potential for cost-effective treatments, traces CADD's evolution, and discusses its role in optimizing biosimilar molecules through molecular modeling, virtual screening, quantitative structure-activity relationship (QSAR) modeling, data mining, and bioinformatics. This review also addresses challenges, including validation, AI integration, regulatory considerations, and ethical implications. Through this exploration, this review aims to illuminate the dynamic relationship between computational innovation and biosimilar agent development, shaping the future of pharmaceutical research and accessibility.

2. Biosimilar Agents

Biopharmaceuticals, often biologics, are drugs created using living systems [33]. However, even in industrialized nations, patients still have difficulty accessing biologics despite their therapeutic advantages [34]. Access restrictions include insurance coverage problems, healthcare system reimbursement, formulary inclusion, prescription availability, and patient out-of-pocket expenses [35–37].

The USA Food and Drug Administration (FDA) defines a biosimilar as a biological product that, in terms of safety, purity, and potency, is comparable to and lacks any clinically significant changes from an existing FDA-approved reference product [38]. The amino acid

sequences of biosimilars and reference products are the same, but they differ in protein aggregation, isoform patterns, glycosylation sites, and 3D structure. To demonstrate their similarity, pharmacodynamic and pharmacokinetic investigations are needed [39].

The development of biosimilar agents according to the requirements of the European Medicines Agency (EMA), the US FDA, or the World Health Organization (WHO) has numerous advantages for patients and society [40]. Because biosimilars can be purchased for up to 30% less than their reference drugs, they allow patients to access cheaper treatments while saving healthcare systems much money [41]. Additionally, biosimilars promote business rivalry, which helps to drive down costs [42].

Despite their many advantages, biosimilars also come with several disadvantages, such as the potential for immunogenicity, the idea of interchangeability, the low level of awareness, the lack of acceptance among patients and healthcare professionals regarding their incorporation in clinical practice, the need for clinical trial testing prior to their approval, a strict regulatory framework limiting the anticipated savings, and certification requirements [43]. Immunogenicity is affected by several variables, including the composition of the biosimilar, the patient's features, the method of administration, the dosing regimen, and the formulation [44]. In order to show the safety and effectiveness of biosimilars and boost physicians' confidence in their usage, it is crucial to choose the proper end goals in clinical trials [43].

Comparing the development of biosimilars to that of small-molecule pharmaceuticals involves different difficulties. First of all, the processes involved in developing and approving biosimilars need to be better understood by doctors and patients [45]. The notion of similarity is complicated by the complex nature of biologics, which comprise several molecular variations with comparable amino acid sequences. Although this variability is controlled throughout production and evaluated through testing, misunderstandings exist because the phrase "biosimilar" may imply differences [46]. Conceptually, it cannot be easy to understand how a biosimilar can produce identical clinical results despite structural differences.

Despite being called "abbreviated," the US regulatory process for biosimilars does not compromise the strictness of the approval requirements. This route enables quicker and more cost-effective development using prior FDA findings [47]. The approval of biosimilars is based on the "totality of evidence," in which a foundation is laid by analytical analysis and confirmed by clinical efficacy and safety investigations [48]. Biosimilars streamline their clinical information package compared to original biologics, necessitating separate Phase III studies for each indication [47,48]. The total data package, however, keeps the same level of rigor [47,48].

Another area for improvement is indication extrapolation, which depends on functional and structural similarity data many stakeholders are unfamiliar with [47]. A recognized scientific idea, extrapolation is applied in different regulatory situations [49]. The disparities between "interchangeability" and "similarity" in the legislation's phrasing further complicate matters [50]. The interchangeability standards have different meanings even while the law stipulates that biosimilars need to have "no clinically meaningful differences". Clarity in communication is essential since interchangeability needs to be proven through separate clinical switching investigations [51].

The difficulties in developing biosimilars compared to small molecule pharmaceuticals highlight the necessity of thorough instruction, exact communication, and a nuanced comprehension of the complex regulatory environment. Because of the complicated nature of biosimilar creation, CADD needs to be used to simplify the process and increase its effectiveness. Table 1 provides a comparison between biosimilars and small-molecule pharmaceuticals.

Table 1. Comparison of biosimilars and small molecule pharmaceuticals.

Aspect	Biosimilars	Small Molecule Pharmaceuticals
Development Process	Complex, requires demonstration of similarity in safety, purity, and potency	Relatively simpler, focus on chemical synthesis
Regulatory Approval	Based on "totality of evidence" approach, relies on prior FDA findings	Strict approval requirements, separate Phase III studies for each indication
Cost	Typically 30% cheaper than reference drugs	Cost varies based on manufacturing and development
Advantages	More affordable, potential for business rivalry	Well-established development process, easily understood by doctors and patients
Disadvantages	Potential for immunogenicity, lack of awareness and acceptance	Limited structural understanding, potential for side effects
Challenges	Indication extrapolation, interchangeability confusion	Potential for misunderstanding due to the complex nature
Communication	Clear communication essential for interchangeability and clinical switching investigations	Straightforward communication due to a well-understood development process
Use of CADD	Necessary to simplify the complex development process	Less necessary due to the simpler development process

3. Evolution of CADD

Drug development, discovery, and design are laborious multidisciplinary procedures spanning several research fields [52]. It is well known that traditional drug research and development are time-consuming and expensive, taking, on average, 10 to 15 years to bring a medicine to market and costing, as of 2015, an estimated 58.8 billion dollars [53,54]. Only 200–250 of 10,000 chemical compounds will undergo clinical testing. These 200–250 chemicals will be studied on animals instead of people. According to research carried out by the Tufts Center for the Study of Therapeutic Development between 1995 and 2007, 11.83 percent of therapeutic compounds that go on to Phase I of clinical trials are eventually given the green light to go on the market. Researchers have had to use a new strategy for drug development due to the high cost and failure rates of conventional drug discovery. The development of new drugs has been sped up thanks to CADD.

Since its inception, the discipline of CADD has undergone a remarkable evolution, revolutionizing pharmaceutical research and drug development. Figure 1 depicts the significant turning points and developments in CADD's history. This visualization illustrates the profound impact of computational techniques on the discovery and design of novel pharmaceutical compounds, from its modest beginnings in the 1950s with the introduction of computers in chemistry to the cutting-edge applications of artificial intelligence and quantum computing in the present day.

CADD refers to the use of computational modeling techniques in the process of finding new drugs. The average authorized medicine takes 10 to 15 years to produce and costs between USD 0.8 and 2 billion [55], making drug development an expensive and time-consuming procedure. Nolatrexed, aliskiren, oseltamivir, dorzolamide, and captopril, among many more licensed drugs, were all optimized using CADD [56], and other papers discuss the effective design and discovery of leads/drugs utilizing CADD [57]. CADD's primary objective is to cut these costs and timeframes while maintaining quality [58]. Table 2 compares traditional drug development, characterized by time-consuming steps and high costs, with computer-aided drug development, which offers reduced time and cost investments through computational approaches such as bioinformatics, virtual screening, predictive modeling, etc.
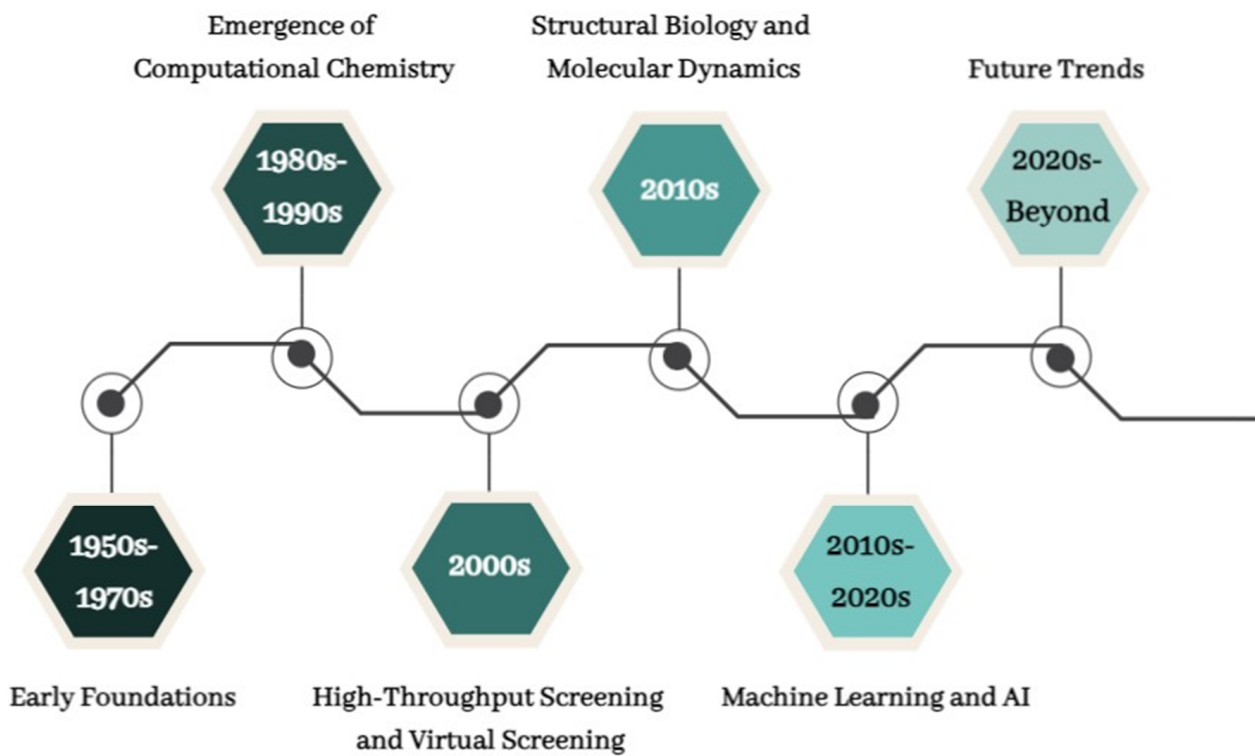
Figure 1. Development of CADD.

Table 2. Comparison of traditional drug development and CADD.

Aspect	Traditional Drug Development	CADD
Steps	Discovery, Target Identification, Lead Compound Identification, Preclinical Testing, Clinical Trials (Phase I, II, III), Regulatory Approval, Post-Marketing Surveillance	Bioinformatics, Molecular Modeling, Virtual Screening, In Silico Testing, Predictive Modeling, Data Analysis
Time Investment	Each step is time-consuming, often taking years	Significantly reduced time per step
Cost Investment	High costs associated with extensive laboratory work, clinical trials, and regulatory processes	Relatively lower costs due to reduced experimentation and reliance on computational methods
Benefits of Approach	Well-established process with proven success, suitable for novel mechanisms with limited data	Faster identification of potential compounds, reduced cost due to in silico testing, streamlined data analysis and prediction
Considerations for Use	Relevant for complex biological systems requiring extensive testing and validation	More suitable for situations with available data and where in silico methods can provide valuable insights
Overall Efficiency	Slower progress due to lengthy experimental phases	Accelerated progress due to computational speed and reduced reliance on physical experiments
Flexibility	Limited flexibility once experiments are initiated	Greater flexibility to adjust and optimize approaches
Risk Management	Higher risk due to resource-intensive nature	Lower risk due to the ability to simulate and predict outcomes
Data Utilization	Heavy reliance on experimental data	Leveraging available data for predictions and insights
Regulatory Approval	Adheres to established regulatory pathways	May require adaptation of regulatory standards for computational methods

To find drugs more quickly and correctly, CADD methods are now widely used. Notably, CADD applies to most drug development stages, including preclinical research, lead discovery, target validation, and optimization. Therefore, it is predicted that CADD could lower the cost of drug development by up to 50% [59,60]. The arsenal of techniques at CADD practitioners' disposal is diverse and powerful. Techniques like molecular dynamics simulations, high throughput, virtual screening, and the comprehensive analysis of absorption, distribution, metabolism, excretion, and toxicity (ADMET) constitute the backbone of this transformative process. The intricate interplay of ligand- and structure-based drug design, binding energy calculations, and quantum studies further amplify the impact. These advanced methodologies heavily rely on intricate hardware and meticulously crafted software frameworks [61]. The implications of these techniques are profound. By employing such strategies, the handling of voluminous databases is made feasible, significantly reducing the reliance on animal models for experimental validation while simultaneously elevating the robustness of investigations [62]. Figure 2 outlines drug design via computer-aided methods, encompassing target selection, compound screening, optimization, validation, clinical development, and regulatory processes.
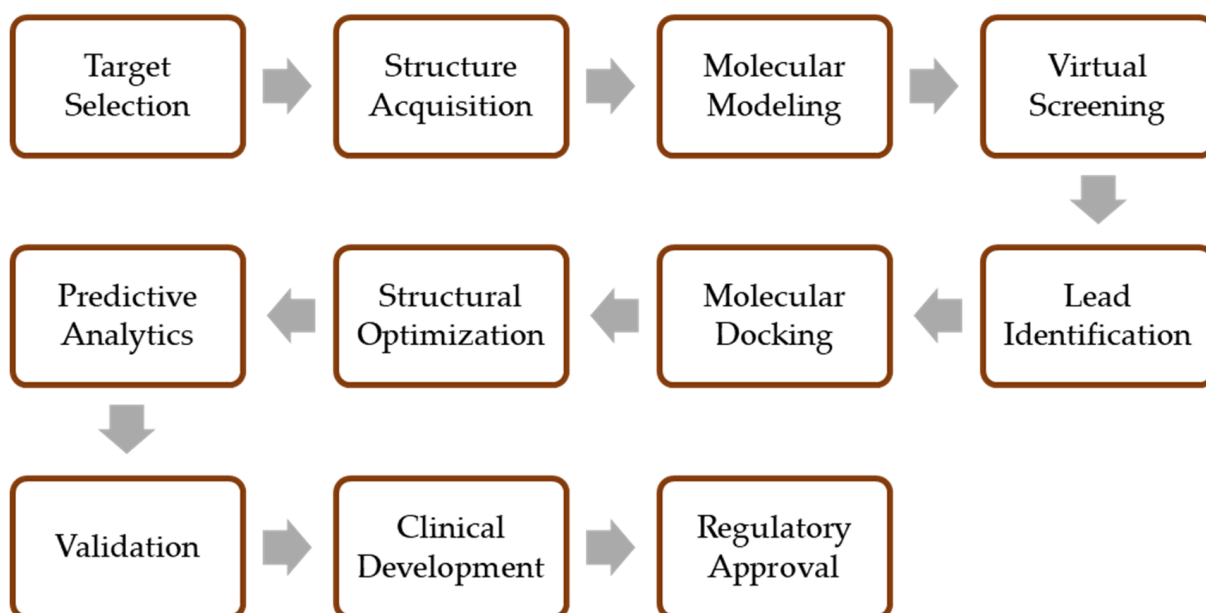

Figure 2. The general principle of drug design using computer-aided methods.

Within the realm of CADD, two primary categories reign supreme: structure-based drug design and ligand-based drug design. Gaining a nuanced understanding of both these paradigms and their respective applications is instrumental. The landscape is peppered with many software options, each tailored to specific tasks, arising from a mix of open-access and proprietary solutions. The selection process, however, is a challenging feat. Parameters such as the nature of the application, study prerequisites, budgetary considerations, reproducibility, the interpretability of analyses, and the user-friendliness of the software all weave a complex tapestry of decision making. The absence of a universal "best" program is readily apparent, as the optimal choice hinges on the researcher's inclinations and the unique demands of the study [63,64].

A panoramic view of CADD encompasses a rich tapestry of methodologies and platforms. From sequence-based drug design to the conception of virtual libraries, molecular similarity computations, scoring functions, conformation sampling, docking-oriented virtual screening, and the intricate task of target identification, these components coalesce in a harmonious symphony. The interdependence of these constituents often leads to enhancements in one facet rippling across others, forging a dynamic ecosystem of progress (Figure 3).
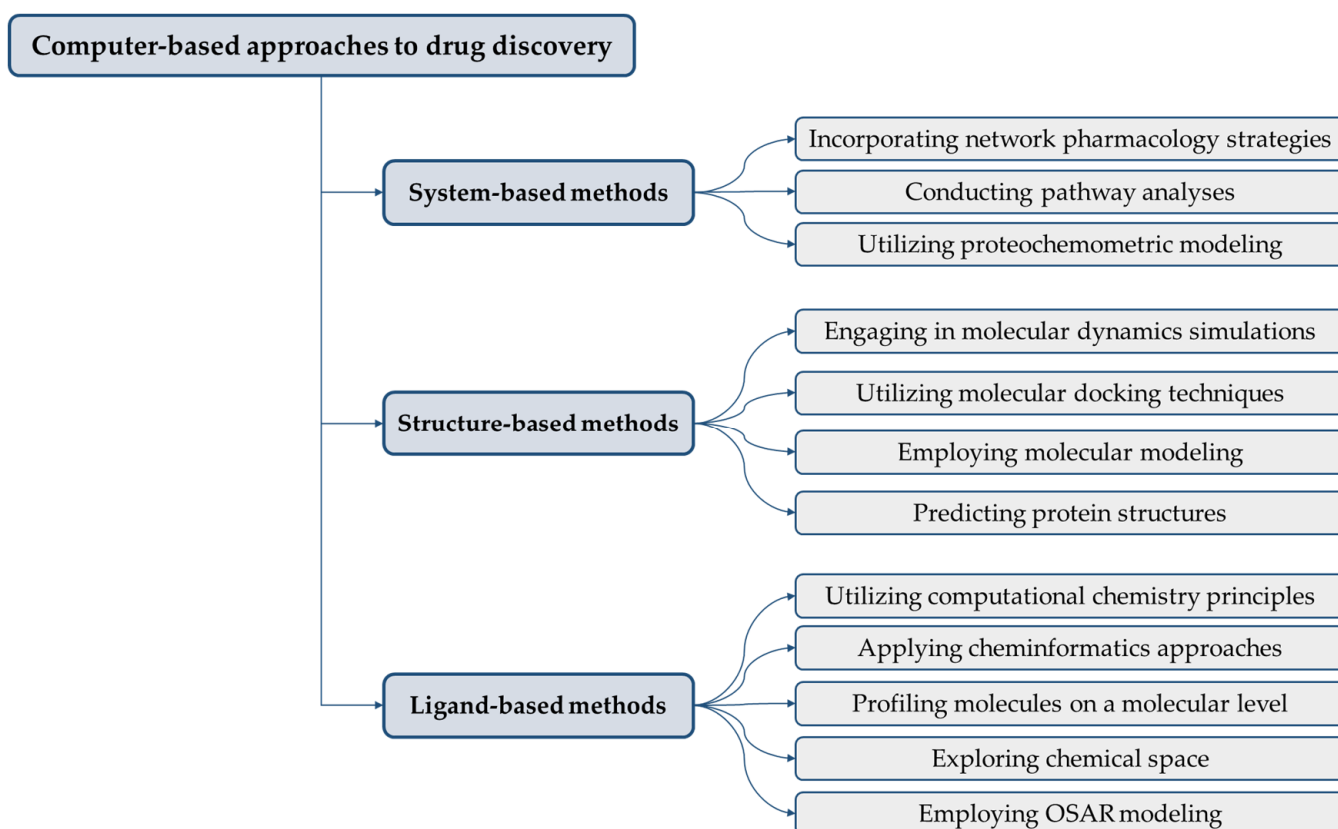
Figure 3. Classifications of various computational instruments used in drug discovery.

Each computational tool or method employed in CADD bears its inherent limitations. This intrinsic constraint is a characteristic of any computational tool relying on predetermined algorithms and scripts [65,66]. The literature has several instances of how these computational techniques have failed, and ADMET systems are not backed by reliable experimental evidence [67].

Recently, new approaches have been applied in CADD, including data mining, artificial intelligence (AI) techniques, deep learning, and machine learning, further improving the speed and accuracy of drug discovery. This is due to advancements in information technology (IT), computational power, and the availability of big data. Future drug development tactics will heavily rely on these cutting-edge IT methods, aiding in feature selection (drug and receptor features), image processing, compound clustering, etc.

Several theoretical fields are included in CADD, including data mining, molecular modeling, chemoinformatics, and bioinformatics [68]. Machine learning and deep learning, which have been used in drug discovery since the 1960s [69], are particularly gaining traction [70]. There are numerous studies on the successful uses of machine learning in CADD, and its significance is well acknowledged [71]. Large data sets are trained using a mathematical framework in the machine learning-based technique before being used to forecast or categorize fresh data sets [72].

In addition to CADD's ongoing contributions, a number of its approaches have entered the hype cycle with high hopes, unrealistic expectations, disillusionments, and useful applications. Fashion, aggravated usage, and a lack of appropriate training to understand the data are common causes of disillusionment [73]. Studies of the QSAR serve as examples. A boom surrounded QSAR experiments a few decades ago, but ill-informed use, unethical behavior, and subpar reporting resulted in inflated hopes and disappointment [74].

Despite the emergence of the above-mentioned ground-breaking methodologies, drug development is still a high-risk endeavor with a poor success rate and large input costs. Additionally, precise experimental data are necessary for CADD's ability to advance fur-

ther [75]. Therefore, it is imperative to develop new experimental or computational tools and scientific methods to find correlations between a drug's nature and structural properties and its safety and efficacy in the human body (pharmacovigilance-based) to identify potentially problematic drug leads at the earliest stages of their development. As a result, the general public will have better health, and medicines will be developed and used in a safe, sensible, and effective way.

The field of computational chemogenomics is expanding, improving drug design with various molecular targets in mind, enhancing the ability of toxicity models and side effects to forecast, and improving collaboration with other fields in order to improve the search for bioactive compounds are just a few of the challenges that CADD continues to face.

CADD conducts a far more tailored search than conventional HTS and combinatorial chemistry, which increases the hit rate of new therapeutic molecules. Its objectives include predicting potential derivatives that might enhance therapeutic efficacy and illuminating the molecular underpinnings of therapeutic action. CADD is typically used to design novel compounds, either orally or topically, direct the optimization of lead compounds, whether to increase its affinity or optimize drug metabolism and pharmacokinetics (DMPK) properties, including ADMET, and filter large compound libraries into smaller sets of predicted active compounds that can be tested experimentally.

While CADD has traditionally been associated with its limitations in accurately predicting complex molecular interactions and binding affinities, its evolution has led to its integration into various stages of biosimilar development. By leveraging computational tools, structural analysis, and molecular modeling, CADD is now proving to be a valuable asset in streamlining the development of biosimilars.

4. Applications of Computer-Aided Discovery in Biosimilar Development

Biosimilar development is a complex and intricate process that demands a comprehensive understanding of the biological molecules' dynamic and structural attributes. In this context, integrating computer-aided techniques has emerged as a pivotal approach to unraveling the mysteries of biosimilar behavior, stability, and relationships. The strategic deployment of computational methods expedites the discovery and optimization of biosimilar candidates and offers insights into their safety, efficacy, and structural alignment with reference molecules. This section delves into the multifaceted applications of computer-aided discovery in biosimilar development, highlighting the different applications' pivotal role in shaping biosimilar innovation's future.

4.1. Molecular Modeling and Simulation

By providing insights into the biosimilar molecules' dynamic and structural features, molecular modeling and simulation methods have evolved into crucial tools in biosimilar development. These computational methods provide scientists with useful data that helps them comprehend these intricate biological entities' behavior, stability, and relationships [76,77].

4.1.1. Homology Modeling for Predicting Biosimilar Structure

A common method used in creating biosimilars is homology modeling, which enables the prediction of a biosimilar molecule's three-dimensional (3D) structure based on the structure of a closely related protein. Computational approaches may provide precise 3D models of the possible conformation of the biosimilar via sequence alignment and comparative analysis. This information is essential for identifying structural similarities and differences, directing further optimization, and verifying the reference molecule's similarity [77,78].

4.1.2. Molecular Dynamics Simulations to Analyze Stability and Interactions

By replicating the motions and interactions of biosimilar molecules over time, molecular dynamics simulations provide a dynamic view of those molecules [79]. These simulations compute the stresses and movements of individual atoms using precise force

field parameters, allowing microscopic interactions, stability, and complicated conformational changes to be explored [78]. When biosimilars interact with their target proteins or other pertinent biomolecules, molecular dynamics simulations may shed light on the thermodynamic stability, binding kinetics, and structural flexibility of those molecules [77]. Researchers may examine the biosimilars' safety and effectiveness characteristics by using this knowledge to better understand the conformational dynamics of those substances under diverse circumstances.

4.2. Virtual Screening

Small molecule libraries are subjected to a virtual screening to find chemical structures that could bind to a therapeutic target [80,81]. By evaluating and ranking huge chemical libraries, virtual screening is a computer process that expedites the discovery of possible candidates for biosimilar products. By using computer algorithms to estimate the compounds' binding affinities and biological activities, this method enables scientists to choose the molecules that have the best chances of succeeding in further experiments.

4.2.1. High-Throughput Virtual Screening to Identify Potential Biosimilar Candidates

In high-throughput virtual screening, hundreds to millions of chemicals are quickly assessed against a target protein or receptor of interest. Virtual screening makes predictions about the binding affinities and poses of these compounds inside the target's active site by using molecular docking or other approaches that are based on structure [82,83]. Researchers can effectively reduce the number of possible biosimilar candidates for experimental validation by determining how well a molecule fits into the binding site and evaluating its binding strength [84].

4.2.2. Ligand-Based and Structure-Based Approaches for Target Identification

Virtual screening methods based on ligands and structures are complementary approaches to creating biosimilars. Ligand-based methods compare substances with comparable chemical characteristics or binding interactions to known ligands or structural motifs. This method is very helpful when there is a need for precise structural knowledge about the target protein [82]. On the other hand, structure-based techniques employ the 3D structure of the target to forecast interactions with possible molecules. These approaches use molecular docking, molecular dynamics simulations, and other computational techniques to evaluate binding affinities and interactions more precisely.

The time and resources needed for conventional experimental screening are drastically reduced thanks to virtual screening, which speeds up the discovery of possible biosimilar candidates. Researchers may thoroughly assess the binding potential of various compounds by combining computational techniques with experimental validation, which enables prioritizing molecules with the most promising biosimilarity properties and therapeutic potential. Figure 4 provides a comparison between ligand-based and structure-based approaches for target identification.

4.3. QSAR Modeling

To connect a biological response (such as cell viability, enzyme activity, etc.) to the chemical characteristics of a group of molecules, QSAR approaches are used [85,86]. The main advantage of the QSAR technique is that it may be used to pinpoint the properties of novel chemical compounds without the need for their manufacture and testing. Studies have shown connections between the structural features of chemicals, physiological traits, and biological activity [87].

A combination of molecular descriptors, which describe different physicochemical and structural characteristics of molecules, and biological activity data gathered from experimental tests is used to build QSAR models [88,89]. By examining these connections, QSAR models may forecast the activity of new drugs, such as biosimilar candidates, against a particular target or biological endpoint. The capacity of QSAR models to shed light

on the structure–activity connections of biosimilars is what gives them their predictive potential. These models are helpful in the creation and development of biosimilars with increased potency and effectiveness because they may highlight crucial structural elements that contribute to the desired activity. Giving the compounds with the greatest anticipated activity the highest priority and directing further experimental validation, QSAR modeling may also help select prospective lead compounds.
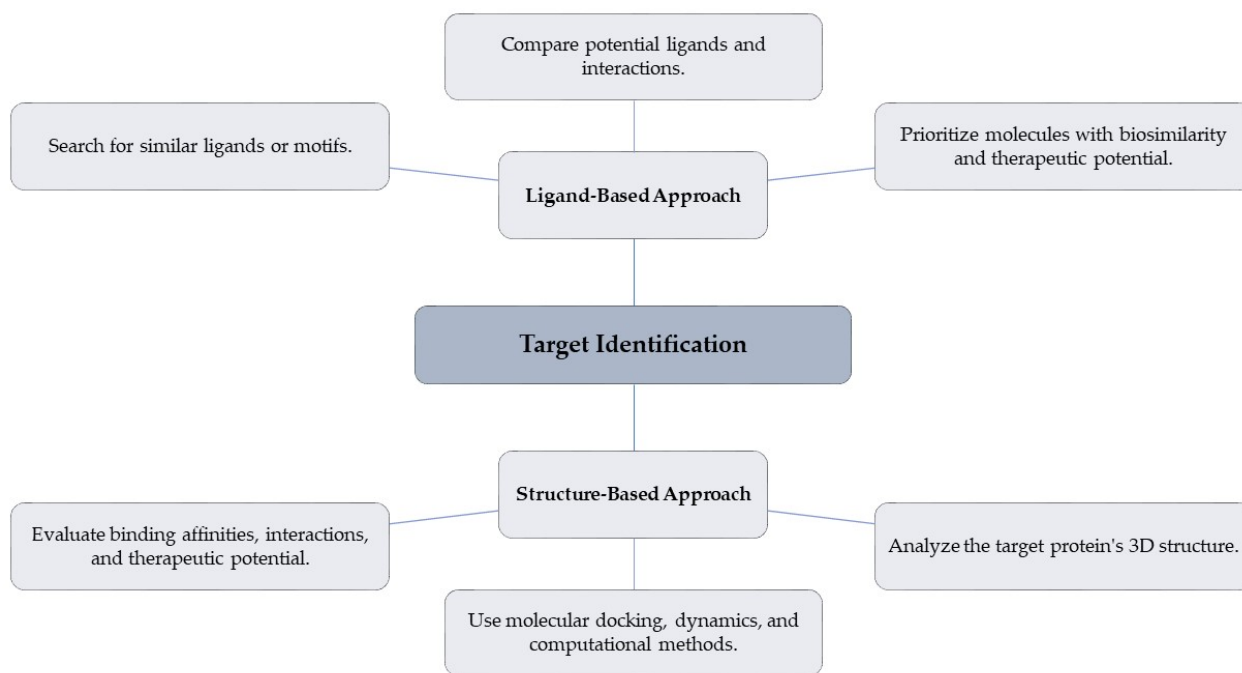

Figure 4. Comparison: ligand-based vs. structure-based approaches for target identification.

Data preparation, descriptor computation, model construction, and model validation are all steps in the QSAR modeling process. To create reliable QSAR models, statistical methods, including multiple linear regression, support vector machines, and neural networks, are often used. Different measures are used to verify these models and evaluate their generalizability [90].

4.4. Data Mining and Bioinformatics

Data mining and bioinformatics are key fields in developing biosimilars that use computer tools to glean insightful information from large biological datasets [91]. These methods are crucial for identifying pertinent biomarkers [92], therapeutic targets [92], and biosimilar candidate optimization [93]. Huge volumes of biological data have been produced due to the development of technologies like next-generation sequencing, omics profiling, and high-throughput screening. These databases extract useful patterns, correlations, and linkages using data mining methods [94].

Bioinformatics tools enable the systematic analysis of molecular and clinical data to identify potential biomarkers and therapeutic targets [95]. Biomarkers are indicators of disease status or treatment response and aid in patient stratification and monitoring [96]. By analyzing molecular profiles of biosimilar candidates and reference molecules, bioinformatics can uncover similarities and differences that contribute to biosimilarity assessment [97]. Furthermore, bioinformatics aids in the identification of potential therapeutic targets by integrating data from various sources, including gene expression [98], protein–protein interactions [99], and pathway analysis [100].

The integration of multiple computing methods, including machine learning, network analysis, and statistical modeling, is common in data mining and bioinformatics. Candidate selection is aided by machine learning techniques, such as clustering and classification,

to categorize biosimilar candidates according to specific characteristics [101]. Network analysis reveals molecular connections, prospective target compounds, and biosimilar intervention paths [102]. Table 3 showcases different approaches of CADD in biosimilar development, along with their use cases, advantages, and disadvantages. It highlights how these methods contribute to discovering, optimizing, and evaluating biosimilar candidates in the biopharmaceutical field.

Table 3. Advantages and disadvantages of CADD approaches in biosimilar development.

Applications	Use Cases	Pros	Cons
Molecular Modeling and Simulation	Predicting 3D structure and dynamic behavior of biosimilars for similarity assessment and optimization	✓ Rapid visualization of 3D structure ✓ Understanding of dynamic behavior	✖ Relies on accurate reference structures ✖ Limited to known protein structures
Virtual Screening	Identifying potential biosimilar candidates with a high binding affinity and biological activity	✓ Accelerates candidate selection ✓ Reduces experimental screening cost	✖ May miss potential candidates ✖ Requires accurate binding affinity predictions
QSAR Modeling	Predicting the activity of biosimilar candidates against specific targets for lead compound selection	✓ Predicts activity without experimental testing ✓ Identifies structural features for activity	✖ Requires reliable experimental data ✖ May not account for complex interactions
Data Mining and Bioinformatics	Identifying biomarkers, therapeutic targets, and optimizing biosimilar candidates from large datasets	✓ Identifies relevant biomarkers and targets ✓ Integrates diverse data sources	✖ Relies on quality and quantity of data ✖ Interpretation of results may be challenging

5. Challenges and Future Directions

In the field of CADD, the accuracy and validation of predictions are becoming increasingly important as computational approaches play a key role in drug discovery. Using CADD predictions can save time and resources in experimental testing, but incorrect predictions can lead to wasted resources and potentially dangerous or useless drug candidates. This section focuses on the challenges and future directions of CADD predictions, particularly in the context of biosimilar development.

5.1. Validation and Accuracy of CADD Predictions

A key component of contemporary drug discovery is using computational approaches to anticipate different features of possible drug candidates in the field of CADD. These tools all work to hasten the discovery and refinement of promising molecules. However, as the use of these prediction models increases, so does the need for careful validation and the evaluation of their efficacy.

Since it directly affects subsequent decision-making steps in drug development pipelines, the accuracy of CADD predictions is crucial. Accurate predictions may save the time and resources needed for experimental testing, speeding up drug development. On the other hand, incorrect or inadequately verified predictions may result in resources being squandered, false leads, and the advancement of potentially dangerous or useless drug candidates to subsequent stages of development.

It is challenging to create reliable prediction models and evaluate their effectiveness across many chemical domains and biological settings. The use of benchmark datasets, blind testing, and comparison with experimental data are thus required. Furthermore, assuring the accuracy and generalizability of CADD predictions becomes more challenging as biological systems' complexity and chemical variety rise.

Researchers and practitioners are investigating solutions, including ensemble methods, cross-validation tactics, and external validation utilizing separate datasets to solve these problems. Additionally gaining momentum are cooperative initiatives to provide uniform standards and evaluation procedures for CADD forecasts. This area also needs to address unique difficulties related to model interpretability and overfitting as AI methods, especially deep learning, become more prevalent.

The confirmation of CADD predictions assumes significant relevance in the context of biosimilar development, where accuracy in predicting the similarity of biological molecules is crucial. Instilling trust in adopting AI-driven methodologies for biosimilar development, robust validation processes can guarantee that the chosen candidates fulfill the strict requirements for similarity to reference biologics.

5.2. Integration of AI and Machine Learning in Biosimilar Discovery

Machine learning and AI have significantly advanced the area of CADD over the years. These developments have great potential for speeding up drug discovery procedures, particularly the creation of biosimilars—biologic medicines that are extremely similar to authorized reference biologics. As CADD techniques develop, they open revolutionary possibilities in compound property prediction, drug candidate optimization, and research pipeline simplification. However, these potentials come with several difficult problems that need our attention. The incorporation of AI in biosimilar discovery is a major topic of discussion in this section as it explores the complex terrain of obstacles and potential future possibilities in CADD.

In recent years, AI has improved the performance of multiple areas [103]. It has made significant contributions to drug target identification, active compound screening, and compound property prediction. Implementing machine learning and deep learning techniques has significantly decreased the costs of developing novel biosimilar drugs and increased the likelihood of success. This is also evident in biosimilar drug synthesis, where AI has a significant impact. The introduction of computer-assisted drug synthesis technologies has simplified and made biosimilar synthesis more accessible. These developments permit the incorporation of newly discovered reactions utilizing efficient methods and inexpensive chemical reactants. In addition, the automation of chemical synthesis using high-throughput automated systems and autonomous devices has ushered in an era of efficiency and safety, alleviating the repetitive, expensive, and dangerous aspects of biosimilar drug synthesis.

Despite these advances, biosimilar drug development still faces obstacles. The availability and integrity of data constitute a significant barrier. Formulating precise routes and predicting optimal reaction conditions are only possible with a comprehensive and reliable database. Nonetheless, ensuring data consistency across multiple sources and periods remains complex. In addition, unobserved errors in response data compromise data integrity, and the absence of essential details such as yields or stereochemistry undermines the credibility of proposed biosimilar drug synthesis routes.

Predicting reactions beyond existing databases and accomplishing complete synthesis of biosimilar products continue to be challenging due to complexities in the chiral center evaluation and strategic management of protective agent addition and removal. Therefore, improving the precision of AI-generated synthesis routes for biosimilars remains a top priority. In addition, the need for more sufficient and diverse data from limited databases hinders the development of effective AI algorithm models. The potential of automated synthesis systems and robotics is also contingent upon the availability of viable and cost-effective chemical reactions tailored to the synthesis of biosimilar drugs.

While significant progress has been made, significant obstacles remain in the biosimilar drug development landscape. The accessibility and integrity of data represent a central challenge. Establishing a comprehensive and dependable database remains crucial for formulating precise synthesis routes and predicting optimal reaction conditions. Deep learning techniques hold immense promise in this regard. Deep learning algorithms can

recognize intricate patterns within immense datasets and could improve the accuracy and dependability of data-driven insights, potentially resolving issues associated with data consistency and integrity. As biosimilar drug development progresses, incorporating deep learning into the current AI framework could catalyze advances in route prediction, reaction optimization, and even the prediction of novel reactions, thereby driving the field toward more efficient and innovative solutions.

Deep learning has shown excellent application potential in the design of therapeutic compounds in recent years, thanks to the ongoing data advances in drug discovery and development. Numerous studies have shown the use of deep learning techniques in ligand-based lead chemical design. Reinforcement learning, variational autoencoders, generative adversarial networks (GANs), and recurrent neural networks (RNNs) are some of the regularly used algorithm model types [104,105].

Although deep learning is not a new technology and has been used for years to analyze language and images, its use in drug development efforts has only recently gained traction [106]. The adoption of Graphics processing units (GPUs) to conduct computationally intensive computations related to deep learning has hastened this [107]. By using several processing layers, referred to as neurons, deep learning goes beyond conventional machine learning techniques to generate predictions based on enormous multidimensional data [108]. Pharmaceutical companies often employ internal experimental study databases [109]. Although there are many other deep learning architectures, convolutional neural networks (CNNs), RNNs, long short-term memory (LSTM), and multi-task learning (MTL) are the ones most often used in drug design and discovery.

RNNs and GANs may be used to create certain bioactive compounds [110–112]. Segler [113], who trained an RNN model and gathered the active compounds of particular targets to fine-tune the RNN, was the first to suggest using the RNN model in molecular design. The RNN model created a focused chemical library of particular targets. RNN-based depth generation models were integrated with data augmentation, migration learning, and temperature sampling in a study by Moret et al. [114] to build novel molecules with particular attributes while working with a small amount of training data.

5.3. Regulatory Considerations for CADD-Generated Biosimilars

To ensure patient safety, effectiveness, and quality, strict regulatory scrutiny is necessary to develop and license biosimilars, which are extremely comparable copies of authorized reference biologic therapies [115]. Regulatory considerations are essential to ensure that CADD-generated biosimilars meet the requirements and expectations set by regulatory agencies, as the integration of CADD methodologies is playing an increasingly significant role in biosimilar discovery [116].

The accuracy and dependability of CADD predictions across various biologics and structural aspects should be shown via thorough validation processes [117]. Transparency in reporting methodology, datasets, and algorithms is required when incorporating CADD into the biosimilar development process [116]. By providing thorough documentation, the agency is certain that it can assess the scientific validity of the predictions made by CADD.

Data integrity and quality are now recognized as major regulatory problems. To ensure accuracy and prevent bias or manipulation, the datasets utilized for CADD predictions need to be representative and well characterized. The experimental data used to create CADD models and the data used for validation are subject to regulatory scrutiny for accuracy. Building confidence in the prediction capacities of CADD-generated biosimilars requires ensuring the traceability and dependability of data sources.

The regulatory environment for biosimilars is constantly changing, and CADD-produced biosimilars need help gaining regulatory clearance. Comprehensive analytical and functional evaluations are necessary to prove comparability between the CADD-produced biosimilar and the reference biologic [115]. Regulatory bodies want proof that the CADD's ability to predict outcomes is supported by actual instances of resemblance in terms of safety, effectiveness, and clinical results.

Engaging with regulatory authorities is recommended as biosimilar developers manage these regulatory issues. Discussions on CADD methodology, validation tactics, and data integrity standards may be had during the first meetings. These conversations may assist in coordinating expectations and guarantee that the regulatory pathway for CADD-produced biosimilars is clear and transparent.

5.4. Ethics of Using AI in Drug Creation

Although combining AI with drug design can significantly advance medicines, it also presents serious ethical questions. Data privacy and consent are the first areas of concern. Using sizable datasets, including private patient information, necessitates open data-gathering processes and protocols for informed consent. The need to combine AI-driven innovation with patient safety is highlighted by ensuring that data anonymization protects patient privacy.

Fairness and prejudice are two other important ethical issues. Healthcare disparities may be maintained by AI algorithms based on skewed data. AI algorithms must be trained on various representative datasets to reduce prejudice and promote fair access to novel treatments to promote ethical drug creation. Addressing algorithmic biases in AI models to reduce inequities and promote moral AI-enabled medicine development is essential.

Accountability and transparency are the focal points of the third ethical component. The complexity of AI algorithms may make the understanding of how they make decisions difficult. To allow academics, authorities, and the general public to examine results, developers need to provide openness in AI-generated predictions. Mechanisms that hold AI developers responsible are necessary for ethical drug development, creating a mutually beneficial partnership between human judgment and AI technologies.

Beyond these factors, it is still crucial to maintain a commitment to patient efficacy and safety. Before entering clinical trials, AI-generated drug candidates should undergo rigorous experimental validation to maintain the proper balance between speeding up drug discovery and patient safety. A significant ethical concern is negotiating the intellectual property environment while promoting free access to AI-generated findings.

6. Concluding Remarks

The incorporation of computer-aided discovery into the development of biosimilar agents represents a transformative path, albeit one fraught with formidable obstacles. The robust validation and accuracy of CADD predictions loom large, necessitating stringent methodologies to ensure the dependability of results. Incorporating AI and machine learning in biosimilar discovery necessitates a delicate balance between ground-breaking innovation and meticulous substantiation, with the need for transparent models and interpretable algorithms serving as a focal point. Regulatory considerations pertinent to CADD-generated biosimilars necessitate collaborative efforts between industry, academia, and regulatory entities to establish all-encompassing guidelines that protect patient safety and therapeutic efficacy. In addition, the ethical considerations entwined with the use of AI in drug development highlight the need for ethical frameworks mandating transparent and principled practices. Envisaging the future of biosimilar development crystallizes the realization that by collectively addressing these challenges, the full potential of computer-assisted discovery can be tapped, reshaping the approach to biosimilars and the broader frontiers of pharmaceutical advancement.

Author Contributions: Conceptualization, S.A.; methodology, H.T.; validation, A.G.; formal analysis, A.G. and S.A.; resources, A.G. and S.A.; data curation, A.G.; writing—original draft preparation, S.A. and H.T.; writing—review and editing, S.A. and A.G.; visualization, H.T. and A.G.; supervision, H.T. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1.   US Food and Drug Administration. *Drugs@FDA Glossary of Terms*; US Food and Drug Administration: Silver Spring, MD, USA, 2017.
2.   Young, D.C. *Computational Drug Design: A Guide for Computational and Medicinal Chemists*; John Wiley & Sons: Hoboken, NJ, USA, 2009.
3.   Luu, K.T.; Kraynov, E.; Kuang, B.; Vicini, P.; Zhong, W.-Z. Modeling, simulation, and translation framework for the preclinical development of monoclonal antibodies. *AAPS J.* **2013**, *15*, 551–558. [CrossRef] [PubMed]
4.   Ahn, C.S.; Gustafson, C.J.; Sandoval, L.F.; Davis, S.A.; Feldman, S.R. Cost effectiveness of biologic therapies for plaque psoriasis. *Am. J. Clin. Dermatol.* **2013**, *14*, 315–326. [CrossRef] [PubMed]
5.   Levin, E.C.; Gupta, R.; Brown, G.; Malakouti, M.; Koo, J. Biologic fatigue in psoriasis. *J. Dermatol. Treat.* **2014**, *25*, 78–82. [CrossRef] [PubMed]
6.   Gniadecki, R.; Bang, B.; Bryld, L.; Iversen, L.; Lasthein, S.; Skov, L. Comparison of long-term drug survival and safety of biologic agents in patients with psoriasis vulgaris. *Br. J. Dermatol.* **2015**, *172*, 244–252. [CrossRef] [PubMed]
7.   Kuhlmann, M.; Covic, A. The protein science of biosimilars. *Nephrol. Dial. Transplant.* **2006**, *21* (Suppl. S5), v4–v8. [CrossRef]
8.   Azevedo, V.; Hassett, B.; Fonseca, J.E.; Atsumi, T.; Coindreau, J.; Jacobs, I.; Mahgoub, E.; O'brien, J.; Singh, E.; Vicik, S.; et al. Differentiating biosimilarity and comparability in biotherapeutics. *Clin. Rheumatol.* **2016**, *35*, 2877–2886. [CrossRef]
9.   Ahmed, I.; Kaspar, B.; Sharma, U. Biosimilars: Impact of biologic product life cycle and European experience on the regulatory trajectory in the United States. *Clin. Ther.* **2012**, *34*, 400–419. [CrossRef]
10. Blauvelt, A.; Cohen, A.; Puig, L.; Vender, R.; van der Walt, J.; Wu, J. Biosimilars for psoriasis: Preclinical analytical assessment to determine similarity. *Br. J. Dermatol.* **2016**, *174*, 282–286. [CrossRef]
11. US Food and Drug Administration. Drugs: Information for Healthcare Professionals (Biosimilars). 2015. Available on-line: https://www.fda.gov/drugs/postmarket-drug-safety-information-patients-and-providers/information-healthcare-professionals-technetium-99m-tc-fanolesomab-marketed-neutrospec-122005 (accessed on 10 October 2023).
12. Olech, E. Biosimilars: Rationale and current regulatory landscape. In *Seminars in Arthritis and Rheumatism*; Elsevier: Amsterdam, The Netherlands, 2016.
13. Weise, M.; Bielsky, M.C.; De Smet, K.; Ehmann, F.; Ekman, N.; Giezen, T.J.; Gravanis, I.; Heim, H.-K.; Heinonen, E.; Ho, K.; et al. Biosimilars: What clinicians should know. *Blood J. Am. Soc. Hematol.* **2012**, *120*, 5111–5117. [CrossRef]
14. McCamish, M.; Woollett, G. The state of the art in the development of biosimilars. *Clin. Pharmacol. Ther.* **2012**, *91*, 405–417. [CrossRef]
15. Kesik-Brodacka, M. Progress in biopharmaceutical development. *Biotechnol. Appl. Biochem.* **2018**, *65*, 306–322. [CrossRef]
16. Gamez-Belmonte, R.; Hernández-Chirlaque, C.; Arredondo-Amador, M.; Aranda, C.J.; González, R.; Martínez-Augustin, O.; de Medina, F.S. Biosimilars: Concepts and controversies. *Pharmacol. Res.* **2018**, *133*, 251–264. [CrossRef]
17. Pierpont, T.M.; Limper, C.B.; Richards, K.L. Past, present, and future of rituximab—The world's first oncology monoclonal antibody therapy. *Front. Oncol.* **2018**, *8*, 163. [CrossRef]
18. Hughes, J.P.; Rees, S.; Kalindjian, S.B.; Philpott, K.L. Principles of early drug discovery. *Br. J. Pharmacol.* **2011**, *162*, 1239–1249. [CrossRef]
19. DiMasi, J.A.; Hansen, R.W.; Grabowski, H.G. The price of innovation: New estimates of drug development costs. *J. Health Econ.* **2003**, *22*, 151–185. [CrossRef]
20. Paul, S.M.; Mytelka, D.S.; Dunwiddie, C.T.; Persinger, C.C.; Munos, B.H.; Lindborg, S.R.; Schacht, A.L. How to improve R&D productivity: The pharmaceutical industry's grand challenge. *Nat. Rev. Drug Discov.* **2010**, *9*, 203–214.
21. Liu, Y.; Li, R.; Liang, F.; Deng, C.; Seidi, F.; Xiao, H. Fluorescent paper-based analytical devices for ultra-sensitive dual-type RNA detections and accurate gastric cancer screening. *Biosens. Bioelectron.* **2022**, *197*, 113781. [CrossRef]
22. Fotis, C.; Antoranz, A.; Hatziavramidis, D.; Sakellaropoulos, T.; Alexopoulos, L.G. Network-based technologies for early drug discovery. *Drug Discov. Today* **2018**, *23*, 626–635. [CrossRef]
23. Salazar, D.E.; Gormley, G. Modern drug discovery and development. In *Clinical and Translational Science*; Elsevier: Amsterdam, The Netherlands, 2017; pp. 719–743.
24. Surabhi, S.; Singh, B. Computer aided drug design: An overview. *J. Drug Deliv. Ther.* **2018**, *8*, 504–509. [CrossRef]
25. Safaei, M.; Sodano, H.A.; Anton, S.R. A review of energy harvesting using piezoelectric materials: State-of-the-art a decade later (2008–2018). *Smart Mater. Struct.* **2019**, *28*, 113001. [CrossRef]
26. Massetti, M.; Jiao, F.; Ferguson, A.J.; Zhao, D.; Wijeratne, K.; Würger, A.; Blackburn, J.L.; Crispin, X.; Fabiano, S. Unconventional thermoelectric materials for energy harvesting and sensing applications. *Chem. Rev.* **2021**, *121*, 12465–12547. [CrossRef]
27. Kore, P.P.; Mutha, M.M.; Antre, R.V.; Oswal, R.J.; Kshirsagar, S.S. Computer-aided drug design: An innovative tool for modeling. *Open J. Med. Chem.* **2012**, *2*, 139–148. [CrossRef]

28. Veselovsky, A.; Ivanov, A. Strategy of computer-aided drug design. *Curr. Drug Targets-Infect. Disord.* **2003**, *3*, 33–40. [CrossRef]
29. Schneider, G.; Fechner, U. Computer-based de novo design of drug-like molecules. *Nat. Rev. Drug Discov.* **2005**, *4*, 649–663. [CrossRef]
30. Sliwoski, G.; Kothiwale, S.; Meiler, J.; Lowe, E.W., Jr. Computational methods in drug discovery. *Pharmacol. Rev.* **2014**, *66*, 334–395. [CrossRef]
31. Bajorath, J. Integration of virtual and high-throughput screening. *Nat. Rev. Drug Discov.* **2002**, *1*, 882–894. [CrossRef]
32. Bleicher, K.H.; Böhm, H.-J.; Müller, K.; Alanine, A.I. Hit and lead generation: Beyond high-throughput screening. *Nat. Rev. Drug Discov.* **2003**, *2*, 369–378. [CrossRef]
33. Rader, R.A. (Re) defining biopharmaceutical. *Nat. Biotechnol.* **2008**, *26*, 743–751. [CrossRef]
34. Socinski, M.A.; Curigliano, G.; Jacobs, I.; Gumbiner, B.; MacDonald, J.; Thomas, D. Clinical considerations for the development of biosimilars in oncology. In *MAbs*; Taylor & Francis: Abingdon, UK, 2015.
35. Cherny, N.; Sullivan, R.; Torode, J.; Saar, M.; Eniu, A. ESMO European Consortium Study on the availability, out-of-pocket costs and accessibility of antineoplastic medicines in Europe. *Ann. Oncol.* **2016**, *27*, 1423–1443. [CrossRef]
36. Monk, B.J.; Lammers, P.E.; Cartwright, T.; Jacobs, I. Barriers to the access of bevacizumab in patients with solid tumors and the potential impact of biosimilars: A physician survey. *Pharmaceuticals* **2017**, *10*, 19. [CrossRef]
37. Baer, W.H.; Maini, A.; Jacobs, I. Barriers to the access and use of rituximab in patients with non-Hodgkin's lymphoma and chronic lymphocytic leukemia: A physician survey. *Pharmaceuticals* **2014**, *7*, 530–544. [CrossRef]
38. Kaida-Yip, F.; Deshpande, K.; Saran, T.; Vyas, D. Biosimilars: Review of current applications, obstacles, and their future in medicine. *World J. Clin. Cases* **2018**, *6*, 161. [CrossRef]
39. Lyman, G.H. Emerging opportunities and challenges of biosimilars in oncology practice. *J. Oncol. Pract.* **2017**, *13* (Suppl. S9), 7s–9s. [CrossRef]
40. Dutta, B.; Huys, I.; Vulto, A.G.; Simoens, S. Identifying key benefits in European off-patent biologics and biosimilar markets: It is not only about price! *BioDrugs* **2020**, *34*, 159–170. [CrossRef]
41. Lucio, S.D.; Stevenson, J.G.; Hoffman, J.M. Biosimilars: Implications for health-system pharmacists. *Am. J. Health-Syst. Pharm.* **2013**, *70*, 2004–2017. [CrossRef]
42. Zelenetz, A.D. Biosimilars in oncology. *Oncol. Hematol. Rev.* **2016**, *12*, 22–28. [CrossRef]
43. Chopra, R.; Lopes, G. Improving access to cancer treatments: The role of biosimilars. *J. Glob. Oncol.* **2017**, *3*, 596–610. [CrossRef]
44. Farhat, F.; Torres, A.; Park, W.; Lopes, G.d.L.; Mudad, R.; Ikpeazu, C.; Aad, S.A. The concept of biosimilars: From characterization to evolution—A narrative review. *Oncologist* **2018**, *23*, 346–352. [CrossRef]
45. Peyrin-Biroulet, L.; Lönnfors, S.; Avedano, L.; Danese, S. Changes in inflammatory bowel disease patients' perspectives on biosimilars: A follow-up survey. *United Eur. Gastroenterol. J.* **2019**, *7*, 1345–1352. [CrossRef]
46. Cohen, H.; Beydoun, D.; Chien, D.; Lessor, T.; McCabe, D.; Muenzberg, M.; Popovian, R.; Uy, J. Awareness, knowledge, and perceptions of biosimilars among specialty physicians. *Adv. Ther.* **2016**, *33*, 2160–2172. [CrossRef]
47. Christl, L.; Lim, S. Biosimilar and Interchangeable Products in the United States: Scientific Concepts, Clinical Use, and Practical Considerations. 2020. Available online: https://www.fda.gov/media/122832/download (accessed on 10 October 2023).
48. Cohen, H.P.; Lamanna, W.C.; Schiestl, M. Totality of evidence and the role of clinical studies in establishing biosimilarity. In *Biosimilars: Regulatory, Clinical, and Biopharmaceutical Development*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 601–628.
49. Krendyukov, A.; Schiestl, M. Extrapolation concept at work with biosimilar: A decade of experience in oncology. *ESMO Open* **2018**, *3*, E000319. [CrossRef]
50. Cohen, H.P.; McCabe, D. The importance of countering biosimilar disparagement and misinformation. *BioDrugs* **2020**, *34*, 407–414. [CrossRef] [PubMed]
51. Food and Drug Administration. Considerations in Demonstrating Interchangeability with a Reference Product: Guidance for Industry. Available online: https://www.fda.gov/media/124907/download (accessed on 10 October 2023).
52. EFPIA MID3 Workgroup; Marshall, S.; Cosson, V.; Cheung, S.Y.A.; Chenel, M.; DellaPasqua, O.; Frey, N.; Hamrén, B.; Harnisch, L.; Ivanow, F.; et al. Good practices in model-informed drug discovery and development: Practice, application, and documentation. *CPT Pharmacomet. Syst. Pharmacol.* **2016**, *5*, 93–122.
53. Mullard, A. Biotech R&D spend jumps by more than 15%. *Nat. Rev. Drug Discov.* **2016**, *15*, 447–448. [PubMed]
54. Ghosh, B.; Choudhuri, S. Drug Design for Malaria with Artificial Intelligence (AI). In *Plasmodium Species and Drug Resistance*; IntechOpen: London, UK, 2021.
55. Am Ende, D.J.; am Ende, M.T. Chemical engineering in the pharmaceutical industry: An introduction. In *Chemical Engineering in the Pharmaceutical Industry: Drug Product Design, Development, and Modeling*; Wiley: Hoboken, NJ, USA, 2019; pp. 1–17.
56. Talele, T.T.; Khedkar, S.A.; Rigby, A.C. Successful applications of computer aided drug discovery: Moving drugs from concept to the clinic. *Curr. Top. Med. Chem.* **2010**, *10*, 127–141. [CrossRef] [PubMed]
57. Lu, W.; Zhang, R.; Jiang, H.; Zhang, H.; Luo, C. Computer-aided drug design in epigenetics. *Front. Chem.* **2018**, *6*, 57. [CrossRef] [PubMed]
58. Kapetanovic, I. Computer-aided drug discovery and development (CADDD): In silico-chemico-biological approach. *Chem. Biol. Interact.* **2008**, *171*, 165–176. [CrossRef] [PubMed]

59. Macalino, S.J.Y.; Gosu, V.; Hong, S.; Choi, S. Role of computer-aided drug design in modern drug discovery. *Arch. Pharmacal Res.* **2015**, *38*, 1686–1701. [CrossRef]

60. Xiang, M.; Cao, Y.; Fan, W.; Chen, L.; Mo, Y. Computer-aided drug design: Lead discovery and optimization. *Comb. Chem. High Throughput Screen.* **2012**, *15*, 328–337. [CrossRef]

61. Hodos, R.A.; Kidd, B.A.; Shameer, K.; Readhead, B.P.; Dudley, J.T. In silico methods for drug repurposing and pharmacology. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **2016**, *8*, 186–210. [CrossRef]

62. Yu, W.; MacKerell, A.D. Computer-aided drug design methods. In *Antibiotics: Methods and protocols*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 85–106.

63. Wang, A.; Durrant, J.D. Open-Source Browser-Based Tools for Structure-Based Computer-Aided Drug Discovery. *Molecules* **2022**, *27*, 4623. [CrossRef] [PubMed]

64. Wójcikowski, M.; Zielenkiewicz, P.; Siedlecki, P. Open Drug Discovery Toolkit (ODDT): A new open-source player in the drug discovery field. *J. Cheminformatics* **2015**, *7*, 1–6. [CrossRef] [PubMed]

65. Korb, O.; Olsson, T.S.G.; Bowden, S.J.; Hall, R.J.; Verdonk, M.L.; Liebeschuetz, J.W.; Cole, J.C. Potential and limitations of ensemble docking. *J. Chem. Inf. Model.* **2012**, *52*, 1262–1274. [CrossRef] [PubMed]

66. Ferreira, L.G.; Dos Santos, R.N.; Oliva, G.; Andricopulo, A.D. Molecular docking and structure-based drug design strategies. *Molecules* **2015**, *20*, 13384–13421. [CrossRef]

67. Blomme, E.A.; Will, Y. Toxicology strategies for drug discovery: Present and future. *Chem. Res. Toxicol.* **2016**, *29*, 473–504. [CrossRef] [PubMed]

68. López-López, E.; Bajorath, J.; Medina-Franco, J.L. Informatics for chemistry, biology, and biomedical sciences. *J. Chem. Inf. Model.* **2020**, *61*, 26–35. [CrossRef]

69. Gasteiger, J. Chemistry in times of artificial intelligence. *ChemPhysChem* **2020**, *21*, 2233–2242. [CrossRef]

70. Bajorath, J. State-of-the-art of artificial intelligence in medicinal chemistry. *Future Sci.* **2021**, *7*, FSO702. [CrossRef]

71. Vamathevan, J.; Clark, D.; Czodrowski, P.; Dunham, I.; Ferran, E.; Lee, G.; Li, B.; Madabhushi, A.; Shah, P.; Spitzer, M.; et al. Applications of machine learning in drug discovery and development. *Nat. Rev. Drug Discov.* **2019**, *18*, 463–477. [CrossRef]

72. Deo, R.C. Machine learning in medicine. *Circulation* **2015**, *132*, 1920–1930. [CrossRef]

73. Medina-Franco, J.L.; Martinez-Mayorga, K.; Gortari, E.F.-D.; Kirchmair, J.; Bajorath, J. Rationality over fashion and hype in drug design. *F1000Research* **2021**, *10*. [CrossRef] [PubMed]

74. Johnson, S.R. The trouble with QSAR (or how I learned to stop worrying and embrace fallacy). *J. Chem. Inf. Model.* **2008**, *48*, 25–26. [CrossRef] [PubMed]

75. Bharath, E.; Manjula, S.; Vijaychand, A. In silico drug design tool for overcoming the innovation deficit in the drug discovery process. *Int. J. Pharm. Pharm. Sci.* **2011**, *3*, 8–12.

76. Chu, X.; Wang, Y.; Tian, P.; Li, W.; Mercadante, D. Advanced Sampling and Modeling in Molecular Simulations for Slow and Large-Scale Biomolecular Dynamics. *Front. Mol. Biosci.* **2021**, *8*, 795991. [CrossRef]

77. Badar, M.S.; Shamsi, S.; Ahmed, J.; Alam, M.A. Molecular dynamics simulations: Concept, methods, and applications. In *Transdisciplinarity*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 131–151.

78. Schlick, T. *Molecular Modeling and Simulation: An Interdisciplinary Guide*; Springer: Berlin/Heidelberg, Germany, 2010; Volume 2.

79. Karplus, M.; McCammon, J.A. Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* **2002**, *9*, 646–652. [CrossRef] [PubMed]

80. Zerroug, A.; Belaidi, S.; BenBrahim, I.; Sinha, L.; Chtita, S. Virtual screening in drug-likeness and structure/activity relationship of pyridazine derivatives as Anti-Alzheimer drugs. *J. King Saud Univ. Sci.* **2019**, *31*, 595–601. [CrossRef]

81. Nunes, R.R.; da Fonseca, A.L.; Pinto, A.C.d.S.; Maia, E.H.B.; da Silva, A.M.; Varotti, F.d.P.; Taranto, A.G. Brazilian malaria molecular targets (BraMMT): Selected receptors for virtual high-throughput screening experiments. *Memórias Do Inst. Oswaldo Cruz* **2019**, *114*. [CrossRef]

82. Neves, B.J.; Braga, R.C.; Melo-Filho, C.C.; Moreira-Filho, J.T.; Muratov, E.N.; Andrade, C.H. QSAR-based virtual screening: Advances and applications in drug discovery. *Front. Pharmacol.* **2018**, *9*, 1275. [CrossRef]

83. Yang, J.; Yu, S.; Yang, Z.; Yan, Y.; Chen, Y.; Zeng, H.; Ma, F.; Shi, Y.; Shi, Y.; Zhang, Z.; et al. Efficacy and safety of anti-cancer biosimilars compared to reference biologics in oncology: A systematic review and meta-analysis of randomized controlled trials. *BioDrugs* **2019**, *33*, 357–371. [CrossRef]

84. Bloomfield, D.; D'Andrea, E.; Nagar, S.; Kesselheim, A. Characteristics of clinical trials evaluating biosimilars in the treatment of cancer: A systematic review and meta-analysis. *JAMA Oncol.* **2022**, *8*, 537–545. [CrossRef]

85. Gbeddy, G.; Egodawatta, P.; Goonetilleke, A.; Ayoko, G.; Chen, L. Application of quantitative structure-activity relationship (QSAR) model in comprehensive human health risk assessment of PAHs, and alkyl-, nitro-, carbonyl-, and hydroxyl-PAHs laden in urban road dust. *J. Hazard. Mater.* **2020**, *383*, 121154. [CrossRef] [PubMed]

86. Ray, R. Understanding the Structural Importance of the Non-Binding and Binding Parts of Bedaquiline and Its Analogues with ATP Synthase Subunit C Using Molecular Docking, Molecular Dynamics Simulation and 3D-QSAR Techniques. In Proceedings of the International Conference on Drug Discovery (ICDD), Hyderabad, India, 29 February–2 March 2020.

87. Padole, S.S.; Asnani, A.J.; Chaple, D.R.; Katre, S.G. A review of approaches in computer-aided drug design in drug discovery. *GSC Biol. Pharm. Sci.* **2022**, *19*, 075–083. [CrossRef]

88. Kwon, S.; Bae, H.; Jo, J.; Yoon, S. Comprehensive ensemble in QSAR prediction for drug discovery. *BMC Bioinform.* **2019**, *20*, 1–12. [CrossRef] [PubMed]

89. Golbraikh, A.; Wang, X.S.; Zhu, H.; Tropsha, A. Predictive QSAR modeling: Methods and applications in drug discovery and chemical risk assessment. *Handb. Comput. Chem.* **2012**, 1309–1342.

90. Kausar, S.; Falcao, A.O. An automated framework for QSAR model building. *J. Cheminformatics* **2018**, *10*, 1–23. [CrossRef] [PubMed]

91. Singh, P.; Singh, N. Role of data mining techniques in bioinformatics. *Int. J. Appl. Res. Bioinform.* **2021**, *11*, 51–60. [CrossRef]

92. Raza, K. Application of data mining in bioinformatics. *arXiv* **2012**, arXiv:1205.1125.

93. Avila-Portillo, L.M.; Aristizabal, F.; Riveros, A.; Abba, M.C.; Correa, D. Modulation of adipose-derived mesenchymal stem/stromal cell transcriptome by G-CSF stimulation. *Stem Cells Int.* **2020**, *2020*, 5045124. [CrossRef]

94. Vázquez, J.; López, M.; Gibert, E.; Herrero, E.; Luque, F.J. Merging ligand-based and structure-based methods in drug discovery: An overview of combined virtual screening approaches. *Molecules* **2020**, *25*, 4723. [CrossRef]

95. Bhunia, S.S.; Saxena, M.; Saxena, A.K. Ligand-and structure-based virtual screening in drug discovery. In *Biophysical and Computational Tools in Drug Discovery*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 281–339.

96. Florian, J.; Sun, Q.; Schrieber, S.J.; White, R.; Shubow, S.; Johnson-Williams, B.E.; Sheikhy, M.; Harrison, N.R.; Parker, V.J.; Wang, Y.; et al. Pharmacodynamic biomarkers for biosimilar development and approval: A workshop summary. *Clin. Pharmacol. Ther.* **2023**, *113*, 1030–1035. [CrossRef]

97. Han, Y.; Klinger, K.; Rajpal, D.K.; Zhu, C.; Teeple, E. Empowering the discovery of novel target-disease associations via machine learning approaches in the open targets platform. *BMC Bioinform.* **2022**, *23*, 1–19. [CrossRef]

98. Hosseini, M.; Hammami, B.; Kazemi, M. Identification of potential diagnostic biomarkers and therapeutic targets for endometriosis based on bioinformatics and machine learning analysis. *J. Assist. Reprod. Genet.* **2023**, *40*, 1–13. [CrossRef] [PubMed]

99. Adinew, G.M.; Messeha, S.; Taka, E.; Ahmed, S.A.; Soliman, K.F. The Role of Apoptotic Genes and Protein-Protein Interactions in Triple-negative Breast Cancer. *Cancer Genom. Proteom.* **2023**, *20*, 247–272. [CrossRef] [PubMed]

100. Chujan, S.; Kitkumthorn, N.; Satayavivad, J. Identification of potential molecular mechanisms and prognostic markers for oral squamous cell carcinoma: A bioinformatics analysis. *J. Int. Soc. Prev. Community Dent.* **2023**, *13*, 237. [CrossRef]

101. Namba, S.; Iwata, M.; Yamanishi, Y. From drug repositioning to target repositioning: Prediction of therapeutic targets using genetically perturbed transcriptomic signatures. *Bioinformatics* **2022**, *38* (Suppl. S1), i68–i76. [CrossRef] [PubMed]

102. Zhang, X.; Wu, F.; Yang, N.; Zhan, X.; Liao, J.; Mai, S.; Huang, Z. In silico methods for identification of potential therapeutic targets. *Interdiscip. Sci. Comput. Life Sci.* **2022**, *14*, 285–310. [CrossRef]

103. Taherdoost, H.; Madanchian, M. Artificial intelligence and sentiment analysis: A review in competitive research. *Computers* **2023**, *12*, 37. [CrossRef]

104. Liu, X.; Zhou, F.; Hu, G.; Shao, S.; He, H.; Zhang, W.; Zhang, X.; Li, L. Dynamic contribution of microbial residues to soil organic matter accumulation influenced by maize straw mulching. *Geoderma* **2019**, *333*, 35–42. [CrossRef]

105. Olivecrona, M.; Blaschke, T.; Engkvist, O.; Chen, H. Molecular de-novo design through deep reinforcement learning. *J. Cheminformatics* **2017**, *9*, 1–14. [CrossRef]

106. Ma, J.; Sheridan, R.P.; Liaw, A.; Dahl, G.E.; Svetnik, V. Deep neural nets as a method for quantitative structure–activity relationships. *J. Chem. Inf. Model.* **2015**, *55*, 263–274. [CrossRef]

107. Raina, R.; Madhavan, A.; Ng, A.Y. Large-Scale Deep Unsupervised Learning Using Graphics Processors. In Proceedings of the 26th Annual International Conference on Machine Learning, Montreal, QC, Canada, 14–18 June 2009.

108. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef]

109. Smalley, E. AI-powered drug discovery captures pharma interest. *Nat. Biotechnol.* **2017**, *35*, 604–606. [CrossRef]

110. Kadurin, A.; Nikolenko, S.; Khrabrov, K.; Aliper, A.; Zhavoronkov, A. druGAN: An advanced generative adversarial autoencoder model for de novo generation of new molecules with desired molecular properties in silico. *Mol. Pharm.* **2017**, *14*, 3098–3104. [CrossRef]

111. Bian, Y.; Wang, J.; Jun, J.J.; Xie, X.-Q. Deep convolutional generative adversarial network (dcGAN) models for screening and design of small molecules targeting cannabinoid receptors. *Mol. Pharm.* **2019**, *16*, 4451–4460. [CrossRef] [PubMed]

112. Gupta, A.; Müller, A.T.; Huisman, B.J.H.; Fuchs, J.A.; Schneider, P.; Schneider, G. Generative recurrent networks for de novo drug design. *Mol. Inform.* **2018**, *37*, 1700111. [CrossRef] [PubMed]

113. Segler, M.H.; Kogej, T.; Tyrchan, C.; Waller, M.P. Generating focused molecule libraries for drug discovery with recurrent neural networks. *ACS Cent. Sci.* **2018**, *4*, 120–131. [CrossRef] [PubMed]

114. Moret, M.; Friedrich, L.; Grisoni, F.; Merk, D.; Schneider, G. Generative molecular design in low data regimes. *Nat. Mach. Intell.* **2020**, *2*, 171–180. [CrossRef]

115. Rahalkar, H.; Sheppard, A.; Santos, G.M.L.; Dasgupta, C.; Perez-Tapia, S.M.; Lopez-Morales, C.A.; Salek, S. Current regulatory requirements for biosimilars in six member countries of BRICS-TM: Challenges and opportunities. *Front. Med.* **2021**, *8*, 726660. [CrossRef]

116. Gundersen, T.; Bærøe, K. The future ethics of artificial intelligence in medicine: Making sense of collaborative models. *Sci. Eng. Ethics* **2022**, *28*, 17. [CrossRef]
117. Nupur, N.; Joshi, S.; Gulliarme, D.; Rathore, A.S. Analytical similarity assessment of biosimilars: Global regulatory landscape, recent studies and major advancements in orthogonal platforms. *Front. Bioeng. Biotechnol.* **2022**, *10*, 832059. [CrossRef]