Instruction of Molecular Structure Similarity and Scaffolds of Drugs Used in Ebola Virus Treatment by Atom Pair and Scaffold Network Graph Algorithm: a Combination of Favipiravir and Molnupiravir

Veysel Gider (vyslgdr@gmail.com)  
Dicle University

Cafer Budak  
Dicle University

Research Article

Keywords: Drug similarity, Drug repurposing, Atom Pair, Scaffold, Graph, Favipiravir, Molnupiravir

Posted Date: May 11th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1594646/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
INSTRUCTION OF MOLECULAR STRUCTURE SIMILARITY AND SCAFFOLDS OF DRUGS USED IN EBOLA VIRUS TREATMENT BY ATOM PAIR AND SCAFFOLD NETWORK GRAPH ALGORITHM: A COMBINATION OF FAVIPIRAVIR AND MOLNUPIRAVIR

Veysel Gider*, Cafer Budak

Department of Electric-Electronic Engineering, Dicle University, Diyarbakır, Turkey. e-mail: vyslgdr@gmail.com

Department of Computer Engineering, Dicle University, Diyarbakır, Turkey, e-mail: cafer.budak@dicle.edu.tr

ABSTRACT

Drug-drug similarities play a very important role in modern medicine, chemistry, and biology. Having in-depth knowledge of the therapeutic structures of drugs aids wet laboratory work and can greatly improve drug research and development efficiency. Many computational methods have been developed for the analysis of such drug similarities. One of the diseases studied for treatment in this area is the Ebola virus. The Ebola virus is a deadly pathogen. Although many researchers around the world have made various efforts on this deadly pathogen, the mortality rate is quite high. Computational approaches are considered to be very useful for antiviral drug discovery. Developing specific antiviral drugs is expensive and takes a long time. By providing an opportunity for the rapid deployment of effective therapeutics, reusing FDA-approved drugs could provide treatments with known preclinical, pharmacokinetic, pharmacodynamic, and toxicity profiles that can enter clinical trials rapidly. The approach in this study is the idea that the combination of drugs may be more effective for the treatment of Ebola due to their similarities and similarities of drugs approved for Ebola, unlike previous studies. For this reason, similarity analyzes of drugs approved for the treatment of the Ebola virus or investigated in clinical studies were performed with the Atom Pair fingerprint algorithm. Then, molecule and infrastructure molecular scaffolds of the drugs selected according to this similar result were created with Scaffold Network Graph (SNG) algorithms and drug structures were analyzed. Public databases (PubChem, DrugBank) and drugs published in the literature for the treatment of Ebola were used for analysis. The molecular, molecular infrastructure and core structures of the drugs that show the most similarity to the FDA approved drugs, which were analyzed with the Scaffold Network Graph algorithm, are shown graphically. When these results were evaluated, it was seen that a combination of Favipiravir, which has also been used in previous outbreaks, and Molnupiravir, which is currently the first approved oral drug candidate for COVID-19, could be effective in the treatment of Ebola. We also think that knowing the chemical structure similarities of drugs currently used to develop other drugs or drug combinations, their inhibitors, and the core structure or structures in drug molecules that are effective against the Ebola virus will be beneficial for chemists.

Keyword words: Drug similarity, Drug repurposing, Atom Pair, Scaffold, Graph, Favipiravir, Molnupiravir
1. Introduction

Ebola virus (EBOV), which first emerged in the Democratic Republic of the Congo, formerly known as Zaire, is a member of the Filoviridae family, also known as the Zaire Ebola virus. EBOV has been responsible for thousands of deaths from its periodic outbreaks since 1976. According to the World Health Organization (WHO) data, the Ebola virus outbreak is classified as a level 3 emergency due to its high death rates [1]. Although virus-related diseases have been known since 1976 when the danger dimension and global impact of the last epidemic of this virus began to be understood in the years 2014-2016, the need to find effective vaccines and drugs that can treat them emerged. To date, there is a commercial vaccine (Ebanga) against the Ebola virus, several approved drugs, and some that are being investigated in clinical trials. Evaluation of molecular structure and structural similarities of these drugs is important for future studies.

The molecular similarity is a very important concept in drug discovery and development. It is based on the assumption that structurally similar molecules usually have similar effects. Evaluation of similarity between small molecules has a great role in discovering and developing different drugs. In particular, two-dimensional (2D) similarity approaches are very popular due to their simplicity, accuracy and efficiency. Many methods have been developed to describe the molecular shape and determine the structural similarity between molecules [2]. Among them, are the most widely used atomic distance-based methods. Structurally similar states of two molecules are often based on the idea that they share similar physical properties and biological functions. This principle of similarity was widely used in the early stages of drug development to discover molecules that could become new drugs.

Molecular similarity analysis is based on structural representations and quantitative measures of similarity between two structural representations. There are many methods to measure the similarity between these two structural representations. For example Tanimoto coefficient, dice index, cosine coefficient, Euclidean distance etc. Based on the structural representation, molecular similarity approaches can generally be divided into 2D and 3D similarity methods. 2D similarity methods are among the fast, efficient and popular similarity methods based on 2D structural information. Among these methods, there are methods such as fingerprint similarity scanning, infrastructure scanning and 2D identifier. A representation of molecules as bit strings is called molecular fingerprints. Its basic principle is to capture the structural features of a molecular graph and encode them into a bit string for later use in evaluating its similarity with a pair of compounds. Molecular fingerprints have generally been designed, validated and used for small molecule drugs within the confines of classical Lipinski. It is not well suited for identifying larger molecules. Each fingerprint detects the presence of specific circular substructures around each atom in a molecule that predict the biological activities of small organic molecules. It weakly perceives the spherical properties of molecules such as size and shape. All these limitations can be addressed using atom pair fingerprints, which encode molecular shape and are often used for scaffolds. Atom pair fingerprints have excellent molecular shape perception for both large and small molecules and can overcome the limitations mentioned above.

Recently, drug-drug and drug-protein similarities have been tried to be exploited from the chemical structures and protein sequences of drugs, and studies have been conducted based on the assumption of association that similar drugs may share similar targets or vice versa. When we look at the studies on drug-drug similarity, it is aimed to find drugs with similar pharmacological properties to the drug of interest and the hypothesis is that the effects of similar drugs should be similar. In various fields such as drug repositioning [3], drug-drug interaction prediction [4], drug-target identification [5] and drug side-effects prediction [6] drug-drug similarity, which has a comprehensive application, can be calculated from different sources. Various calculations based on drug properties such as chemical structure characteristics [7], gene expression profiles [8], side effect profiles [9], and biological target [10] drug-drug similarity analytics have been applied. Using a neighbour recommendation method using molecular structure similarity analysis [11] where it is assumed that similar drugs can have nearly similar interactions and a computational framework to extract drug interactions and associated recommendations [12] [13], are other studies on drug-drug interactions.

The main goal of drug design is to find new compounds with desired pharmacological properties. The concept of a chemical scaffold as a common core structure that characterizes a group of individual molecules as a substructure has a long tradition in chemistry [14] Designing molecules that hold specific scaffolds as core structures is an effective way to obtain potential drug candidates [15].
Despite the ubiquity of the concept of scaffolding, defining a scaffold in silico can be subjective when there is uncertainty about which part of the molecule is considered the nucleus. The popular Bemis and Murcko definition [16] ignores these ambiguities by describing a scaffold as a combination of rings and the linking atoms that connect them. To reduce uncertainty, it has been suggested that a molecule can be represented by a tree or network of interrelated scaffolds, defined by the iterative removal of circumferential rings from the Bemis and Murcko structure.

The concept of a chemical scaffold as a common core structure that characterizes a group of individual molecules in which it is a substructure has a long tradition in chemistry. Constructs that share a scaffold can often be assumed to share a common synthetic pathway, and in typical combinatorial libraries, all compounds are based on a common scaffold. From this scaffold, the rings are removed one by one until only one ring remains. Removing a ring means removing the bonds and atoms that are part of the ring, excluding the atoms and bonds that are part of any other ring. In addition, all exocyclic double bonds attached to the removed ring atoms are also removed. If the removed ring was attached to the remaining scaffold by an acyclic (open-chain compound) linker, this linker is now a terminal side chain and is also removed. If removing a ring would result in a disconnected structure, that ring cannot be removed. Aggregating multiple linear graphs create a scaffold tree. The scaffolding network approach [17] deals with the uncertainty of determining the most peripheral ring with a comprehensive enumeration of all possible decompositions. Networks formed from multiple molecules can be aggregated to form a large multi-part directed acyclic graph [18].

Scaffolds are often applied to the structure and explore large screening datasets in small molecule drug discovery. Also, medicinal chemists use scaffolds as an organizational principle in their optimization projects. The scaffold tree approaches [19], [17] apply a hierarchical set of fragmentation rules such that each scaffold is divided into exactly two subparts and a tree-like graph is created. On the other hand, Scaffold meshes such as those proposed by [17] and [20] are created by extensive fragmentation of each scaffold (according to predefined fragmentation rules) into all possible sub-scaffolds. Another category that can be used to distinguish existing approaches is the type of fragmentation rules applied: The HierS algorithm, written by [20], removes side chains and linkers, leaving the ring systems in fragments by the definition of Bemis and Murcko [21].

The reuse of drugs that had previously been FDA-approved as treatments for EBOLA presents an opportunity for the rapid deployment of effective therapeutics in the current epidemic environment where treatment options are greatly limited. In line with the above-mentioned drug-drug similarity and chemical scaffolds approaches, drugs used in the treatment of Ebola were examined. The drugs used in the study were taken from PubChem, DrugBank, and previously published articles suggesting drug candidates for use in the treatment of Ebola. Among these drugs, drug-drug similarity analysis was performed with FDA-approved Favipiravir, which was previously used for Influenza, COVID-19 and is currently used for Ebola, and other approved or drug candidate drugs (Molnupiravir is the first oral drug candidate for COVID-19). This process was performed using Atom Pair similarity analysis. Molecular scaffolds of these analyzed drugs were extracted with the Scaffold Network Graph algorithm in the next step and shown graphically. Considering all these results, the recommendation of a combination of Favipiravir and other drugs examined, especially Molnupiravir, which is the first oral drug for COVID-19, is to assist both clinical and computer analysis processes for use in the treatment of the Ebola virus. It is also to chart a rapid path towards target therapy.
2. Material and method

The molecular similarity is of great importance in the prediction of properties in chemical compounds, in the design of chemicals with predefined properties, and most importantly in the realization of drug discovery studies. It is usually created by scanning large indexes containing the structures of existing or potentially existing chemicals. Within the framework of this logic, a molecular similarity list was obtained with the atom pair algorithm to perform a drug-drug similarity analysis of drugs used or candidates for use in the treatment of Ebola. In the next step, the results obtained by the extraction of molecular scaffolds of drugs with high similarity were evaluated.

2.1. Dataset

The structured format of the drugs used as the data set in this study is the Simplified Molecular-Input Line-Entry System (SMILES). SMILES is a series representation of the 2D structure of a molecule. It maps any molecule to a special string that is usually unique and can be remapped to the 2D structure. In some cases, different molecules may be mapped to the same SMILES sequence, degrading the performance of the model. In this study, drugs from publicly available databases Drugbank, PubChem and drug published articles for Ebola were used as a dataset to extract similarities between selected FDA-approved drugs for the treatment of Ebola. DrugBank is a reliable database containing information about drugs such as drug targets, drug enzymes, drug interactions and drug carriers. PubChem is a database for drug structures. Information on the drugs used is shown in Table 1.

| Databases | Drug properties | URL |
|-----------|-----------------|-----|
| Favipiravir | It is a broad-spectrum antiviral developed by Toyama Chemical Co Ltd. Favipiravir, which was previously recommended for Influenza, Lassa virus, COVID-19 treatments, and currently studied for Ebola treatment, has been shown to have high activity against EBOV in vitro [22], [23]. | [https://go.drugbank.com/](https://go.drugbank.com/) [https://pubchem.ncbi.nlm.nih.gov/](https://pubchem.ncbi.nlm.nih.gov/) |
| Molnupiravir | It is an orally bioavailable isopropyl ester cytidine analogue used to treat COVID-19. Molnupiravir was also granted emergency use authorization by the FDA on December 23, 2021; but not yet fully confirmed [22], [23]. | [https://go.drugbank.com/](https://go.drugbank.com/) |
| Bepridil | It is a long-acting, non-selective calcium channel blocker drug with significant antianginal activity [22], [23]. | [https://go.drugbank.com/](https://go.drugbank.com/) [https://pubchem.ncbi.nlm.nih.gov/](https://pubchem.ncbi.nlm.nih.gov/) |
| Remdesivir | It is a nucleoside analogue used to treat RNA virus infections, including COVID-19. Remdesivir (GS-5734) is an adenosine triphosphate analogue that was first identified in the literature in 2016 as a potential treatment for Ebola. To date, it has shown in vitro activity against the viral families Arenaviridae, Flaviviridae, Filoviridae, Paramyxoviridae, Pneumoviridae and Coronavirusidae. On November 19, 2020, the FDA granted an Emergency Use authorization to Remdesivir, which is used with baricitinib for the treatment of COVID-19 [22], [23]. | [https://go.drugbank.com/](https://go.drugbank.com/) [https://pubchem.ncbi.nlm.nih.gov/](https://pubchem.ncbi.nlm.nih.gov/) |
| Amodiaquine | It is an antimalarial drug. Amodiaquine is at least as effective as chloroquine and can give positive results in different diseases when combined with other drug substances [22], [23]. | [https://go.drugbank.com/](https://go.drugbank.com/) |
| Azithromycin | It is a macrolide antibiotic used to treat a variety of bacterial infections. It was approved by the FDA in 1991 [22], [23]. | [https://go.drugbank.com/](https://go.drugbank.com/) |
| Brincidofovir | It is an oral antiviral drug used in the treatment of human smallpox infections. Developed by Chiremax under the brand name Tembexa, Brincidofovir was approved by the FDA in June 2021 for the treatment of smallpox infection [22], [23]. | [https://go.drugbank.com/](https://go.drugbank.com/) [https://pubchem.ncbi.nlm.nih.gov/](https://pubchem.ncbi.nlm.nih.gov/) |
Drug-drug similarity analysis, which is a strategy of this study, can be considered a valid alternative provided that the drug has been used clinically frequently. In particular, a remarkable number of drugs that have been reconsidered for the treatment of Ebola are used for the treatment of cancer. This makes it possible that drugs that interfere with specific infection pathways may also be effective in defeating viral replication. For this reason, drugs approved for the treatment of Ebola or studied in clinical trials were used in this study to be able to draw up a list of drugs used to treat Ebola symptoms by looking at their chemical structures for drug-drug similarities.

2.2. Proposed method

The fact that treatments for Ebola virus disease are with drugs designed to treat the disease symptoms of previous viruses justifies the reuse of FDA-approved drugs. This study aims to help fight the epidemic as quickly as possible with the use of repurposed drugs for the rapidly spreading Ebola virus by extracting similarities of FDA-approved drugs and analyzing the molecular scaffolds of these drugs. For this purpose, the proposed method for re-determining suitable drug candidates is shown in Figure 1.

![Figure 1](image)

Figure 1. Flowchart of the proposed approach to the analysis of drug-drug similarity and molecular scaffolds for the treatment of Ebola.

All binary chemical similarities were calculated for fingerprint-based similarity calculations. The obtained drugs were compared with other drugs of Favipiravir used in the treatment of Ebola and analyzed by the atomic pair similarity method. Molecular scaffolds were extracted with the Scaffold Network Graph algorithm by choosing the drugs with the highest similarity. Thus, the core structure molecules between the two drug similarities were determined.

2.3. Drug-Drug Similarity Analysis

Clinical drug-drug similarity, associated with chemical similarity and literature-based drug similarity, has many potential applications in assessing drug therapy similarity and disease similarity. Chemical structures are often represented by the inference of molecular fingerprints, in which structural features are converted to bits in a bit vector or numbers in a count vector. This representation allows the calculation of chemical structures to be analyzed and compared efficiently. Using such fingerprints, the similarity between the two molecules is inferred. The molecular Fingerprint structure is shown in Figure 2.
A bit position is associated exactly with a predefined property. The presence of the feature in the molecule is expressed as "1" and the absence of the feature in the molecule as "0". The fact that the drug-drug interaction is assigned to zero (0), which is one of the most critical issues in the field of health and drug development, shows that there is no evidence of their interaction yet. Therefore, they can interact with each other.

\[
J(d_i, d_j) = \frac{|d_i \cap d_j|}{|d_i \cup d_j|} \quad (1)
\]

\[
d_j(d_i, d_j) = 1 - J(d_i, d_j) \quad (2)
\]

In the study, the Jaccard similarity coefficient was used for drug-drug distance measurement. This similarity coefficient shows significant performance in the drug-treatment relationship. To perform this operation, the values are converted to binary bits. Significant input values were set to 1 and non-significant inputs to 0. It can be calculated as the ratio of bits with both drugs having a value of 1 in the same diagnostic code among drugs where at least one drug is 1 stated in Equation 1. In this study, the drug-drug distance was calculated as stated in Equation 2.

**a) Atom Pair Similarity Analysis**

Representation of molecular (chemical) structures as bit strings (1’s and 0’s) is done with molecular fingerprints. The basic logic is to graphically capture the structural features of a molecule and encode it in a bit string for later use in evaluating similarity with a pair of compounds. The major advantage of this process is that it is very useful for storing such a representation and it saves time by making the computation process of comparing two molecular graphs considerably faster than comparing bit strings. Atom Pair (AP) fingerprint, which is relatively easier to implement and performs better, is used [24]. The configuration formula for an atom pair is as shown in Equation 3:

\[
A = s[1 - \sum_{i=1}^{m} \frac{\tau_i \cdot \Delta n_i}{\sum_{i=1}^{m} \tau_i}] \quad (3)
\]

A is the activity concerning the standard drug, S is the similarity between the drug and its analogue, the path length of the \(\tau_i\) atom pair \(\Delta n_i\) is the difference of the standard drug from the analogue of the \(i\).heteroatom pair, and \(n_i\) is the number of the \(i\).heteroatom atom pair in the standard drug. The overall process is outlined in Figure 3.
Figure 3. Atom pair fingerprint structure [24].

The steps shown in Figure 3 are as follows. When creating an atom pair fingerprint, the following steps are performed for each heavy atom pair:

1) subtracting the given pair of atoms and the shortest path between them;
2) coding of identifiers (atom type and number of bonds for both atoms and their topological distance);
3) conversion to bit strings;
4) concatenating the bit strings into a number;
5) hashing the number into the index field;
6) is to set the corresponding position in the fingerprint to 1.

The basic logic of this similarity is coloured based on how many of the bits set by the atom are present in the fingerprint. Figure 3 shows normalizing the "weight" of an atom and using the normalized weights to then colour the atoms in a topography-like map; green indicates a positive difference (i.e., the similarity or probability decreases when bits are removed) and pink indicates a negative difference, grey indicates no change. The visualization is shown for atom pairs fingerprint types [24].

b) Scaffold Network Graph Analysis

The Scaffold Network Graph (SNG) is a new command-line tool used in drug discovery and development. SNG is used to extract molecular scaffold meshes and graphs from feature large molecule clusters. This network and graphs can be incorporated into an automated line to visualize and analyze large databases of chemical information to identify scaffolds that can be turned into drugs [25]. Identifying, detecting, and analyzing scaffolds is subjective because there is uncertainty about which part of a molecule’s structure should be considered the core structure and which parts should be considered side chains. The firsts work in this field uses an algorithm to eliminate or ignore such ambiguities and assign each molecule to a single, which may not be ideal, scaffold. For example, Invacacftor's Bemis–Murko pier is shown at the top level in Figure 4.
When **Figure 4** is examined, the isolated molecule in Level 3 is the Bemis-Murcko scaffold of the drug. Sub-scaffolds at Level 1 and Level 2 represent the removal of a peripheral ring and the removal of the resulting side chains from the main scaffolds to which they are attached. The dotted lines show the specific way the Scaffold Tree algorithm parses the Invacaftor drug. SNG can calculate the entire network or just the tree. Molecules with the same scaffold or scaffolds connected in a network or tree tend to have the same biological activity.

Other works in this area were also to more appropriately demonstrate the complexity of removing scaffolds by assigning molecules to a node in a network or a node in a tree of interrelated scaffolds. The first strategy to manage scaffold uncertainty is the Scaffold Tree [14]. This algorithm assigns each molecule to a single Bemis-Murcko scaffold but then parses that scaffold by iteratively stripping the most circumferential ring. Sequential dissociations yield a linear graph of scaffolds, all assigned to the molecule. For example, the linear graph created from Invacaftor is shown with double dotted lines in **Figure 4**. The Scaffold Tree is the sum of linear graphs derived from multiple molecules. Scaffold trees are useful but can introduce new uncertainty in determining which ring is the most 'peripheral'.

The second strategy, Scaffold Networks, decouples each scaffold in all possible ways. This creates a scaffolding network for each molecule. The scaffolding mesh for Invacaftor is shown in **Figure 4**. Networks from multiple molecules can be superimposed to create large multipart graphs [17]. Molecules associated with nodes in this graph (each corresponding to specific scaffolds) often have similar properties. This network of scaffolds is useful for analyzing and visualizing screening data to select the most promising scaffolds to be converted into drugs.

**Scaffolding Tree:** The basic mathematical description of a molecule is the molecular graph, which includes atoms as nodes and chemical bonds as edges. However, molecular graphs are not easy to produce explicitly as graphs due to the presence of rings, relatively large size and chemical validity constraints. For ease of computation, we can see that by following the steps in the work of [26], [27] we can transform a molecular graph into a scaffold tree as a higher-level representation or a substructure tree.

**Step 1 (Substructure):** Substructures can be an atom or single rings. The substructure set is denoted by the $S$ (word set), which includes frequent atoms and single rings in drug-like molecules.

**Step 2 (Scaffolding Tree T):** A scaffolding $\text{tre}T_X$, is a spanning tree whose nodes are substructures. The molecular graph is the high-level representation of $X$. 

![Figure 4](image_url) **Figure 4.** The scaffold network for the drug Invacaftor shows all possible ways in which the drug can be decomposed into composite scaffolds.
$T_X$ is represented by (i) the node indicator matrix, (ii) the adjacency matrix, and (iii) the node weight vector. In $T_X$ we distinguish leaf and non-leaf nodes. Among the $K_1$ nodes in $T_X$ are $K_{leaf}$ leaf nodes (nodes that only connect to an edge) and $K - K_{leaf}$ non-leaf nodes (in other cases). Leaf nodes and non-leaf node clusters are denoted as $V_{leaf}$ and $V_{nonleaf}$ accordingly.

**Step 3:** The node indicator matrix $N$ is decomposed to $N = \begin{pmatrix} N_{nonleaf} \\ N_{leaf} \end{pmatrix} \in \{0, 1\}^{K \times |S|}$, where $N_{nonleaf} \in \{0, 1\}^{(K - K_{leaf}) \times |S|}$ corresponds to non-leaf nodes and $N_{leaf} \in \{0, 1\}^{(K_{leaf}) \times |S|}$ corresponds to leaf nodes. Each N row is a one-hot vector indicating which infrastructure the node belongs to.

**Step 4:** The neighborhood matrix is denoted by $A \in \{0, 1\}^{K \times K}$. $A_{ij} = 1$ indicates that $i_{th}$ node and $j_{th}$ node are connected, 0 indicates not.

**Step 5:** The node weight vector, $w = [1, \cdots, 1]^T \in \mathbb{R}^K$, indicates that the nodes $K$ in the scaffold tree are equally weighted [28].

### 3. Experimental results and discussion

In this study, the Atom Pair fingerprint similarity algorithm and Scaffold Network Graph algorithm method were used. Drug-drug similarity analysis was performed for the reuse of existing drugs useful as Ebola therapeutics. Atom Pair similarity results for Drug-Drug similarity of the proposed method are shown in Table 2.

As stated in Equation 2, drug-drug similarities show how close drug molecules are to each other, as seen in Table 2. Drugs that are very close to each other, almost similar, appear to be completely green in colour.

It is also a demonstration of the Atom Pair fingerprints algorithm, how parts should be encoded into directories. The basic idea at this stage is to take into account the structural features (these are bond structure, atomic type, etc.) in molecular fingerprints and get the values of each atom of a particular piece. These are then encoded as a finite number of bits (for example, three bits are sufficient for the number of links) and combined to form a bit representation of the fragment index of the structure. The operation of the basic logic of the AP fingerprint structure is as follows:

1. Remove all-atom pair parts
2. Encode parts to integers (indexes)
3. Create a string of bits of length $n$
4. Add hash directories to a field of the bit string
5. Turn on the corresponding bit for each of the hash indexes,

That is, the bits corresponding to the pairs of atoms in the molecule are turned on and the remaining bits are turned off. Table 2 shows the atom pair similarity results of the obtained drugs.
Table 2. Results of chemical structure similarity of favipiravir drug and other drugs by Atom Pair Analysis

| 2D Reference Drug | 2D Paired Drug | 2D Atom Pair Similarity Result |
|-------------------|----------------|-------------------------------|
| Favipiravir        | Favipiravir    | Favipiravir-Favipiravir       |
| Favipiravir        | Molnupiravir   | Favipiravir-Molnupiravir      |
| Favipiravir        | Bepridil       | Favipiravir-Bepridil          |
| Favipiravir        | Remdesivir     | Favipiravir-Remdesivir        |
| Favipiravir        | Azithromycin   | Favipiravir-Azithromycin      |
These structures are coloured as specified in Section 2.3(b). In this context, the Atom pair similarity results in Table 2:

- Green indicates a positive difference (ie removing bits reduces similarity or probability)
- Pink indicates a negative difference,
- Gray indicates no change.

The colour that indicates how much the Favipiravir drug seen in Table 2 overlaps with other drugs taken from the literature for Ebola is green. In other words, the green colour seen here indicates how similar the two drugs are. Similarity maps are a very useful and easy to understand strategy for atomically visualizing fingerprint similarity between molecular structures. Atomic weights are produced by comparing the similarity resulting from removing the bits of the corresponding atom with the (unmodified) similarity of the previous fingerprint. Similarity maps can be generated for each fingerprint, allowing bits to be traced back to a corresponding atom or substructure.

Used in the second phase, the Scaffold Network Graph is an open-source Python library and command-line tool for the generation and analysis of molecular scaffold networks and trees and is capable of processing large sets of input molecules. Figure 5 shows the scaffolds of the drugs in Table 2, which were created with the Scaffold network graph algorithm.
Figure 5. Representation of the molecular scaffolds of these drugs created by the Scaffold Network Graph algorithm of 7 drugs whose similarity was extracted.

In Figure 5, all possible infrastructure scaffolds of the molecules of 7 drugs, whose similarities were also extracted with the atom pair algorithm in Table 2, were extracted. The scaffolds are
graphically illustrated in Figure 6 to show them more meaningfully and which scaffold structure is linked to which.

![Figure 6](image)

**Figure 6.** In Figure 5, the graphical representation of the nodes of the seven drug molecules whose scaffolds have been extracted, and the interrelationship between them, as an edge.

In Figure 6, the scaffolds of the molecules of the seven drugs are shown graphically. There are 68 nodes in total here. This shows that there are a total of 68 substructures and side chains for seven drugs. The high number of these nodes can be considered a measure of the similarity between drugs. To make this infrastructure and side chains more meaningful, the nodes here are redrawn in Figure 7 by showing the molecular structure images in the node using the Scaffold tree algorithm.

![Figure 7](image)

**Figure 7.** Redrawing the structure tree with the Scaffold Tree algorithm, in which the nodes and the connection relationship between them, as shown in Figure 6, and the structure molecules in the nodes are shown.

This analysis is useful in identifying scaffolds that can be further optimized into drug-like molecules. Until this stage, Favipiravir was used for Influenza, Ebola and COVID-19 was taken as a reference drug. The scaffolding network and tree representations created for Molnupiravir, the first oral drug candidate for the treatment of COVID-19, and Favipiravir, our reference drug, which is thought to be effective in almost all epidemics, are as in Figure 9.
Figure 8. 2D representations of Favipiravir and Molnupiravir, representation of Scaffold scaffolds and side chains of these drugs

Considering the substructures of the molecules between the two drugs, a total of 14 different substructures and side chains were formed. To see which structure is related to which structure among these, it is shown graphically in Figure 9.

Figure 9. Representation of scaffolds of Favipiravir and Molnupiravir in the form of knot graphics
When Figure 9 is examined, a total of 14 nodes and the connections between them will be seen. The reason for this is that, as can be seen in Figure 8, the infrastructure scaffolds of two drug molecules are formed with the Scaffold algorithm and a total of 14 structures are extracted. The drawing of the scaffold tree and scaffold tree is shown in Figure 9. The scaffolding tree can be drawn as a mesh, but the structures of the molecules are not shown in the node. The structured image is important to the chemist, so the mesh showing the structure images in the node is shown in Figure 10 below.

**Figure 10.** Scaffolding tree representation of Favipiravir and Molnupiravir drug molecules.

When Figure 9 and Figure 10 are examined, the same molecules and their substructures are numbered. In Figure 9, the molecules and infrastructure molecules of Favipiravir and Molnupiravir are shown as nodes and edges in the connections between them. Here, molecular structure images are also shown at the nodes, as in Figure 10, so that we can comment on the drugs. It means that the number of nodes connected between the two drugs is high and they have similar structures.

### 3.1. Discussion

In drug development, the reuse of existing drugs used for the treatment of one disease for disease can be much faster than the discovery of a new drug. If these existing drugs are approved for the treatment of another disease, they can provide new treatments more rapidly and at a higher rate, but there are many considerations when reusing approved drugs for a new treatment. Drug-drug similarity comparisons, which is one of the important fields of study in drug development, help to reveal alternative drug targets and off-target effects of a particular drug. This aims to identify similar modes of action between drugs of a different pharmacological class or chemical structure. In the chemical structure-biological activity comparison, the chemical structure of two different drugs can be compared to investigate the potential shared biological activity between structural similarities [29]. Comparing such drug-drug similarity situations is very useful, but also dangerous since in some cases drugs with similar structures have similar effects. Sometimes the effects of drugs can be very different even if they have very similar skeletons. Many drugs act in different places in the body, that is, they have more than one pharmacological effect. This shows how difficult such comparisons are at times. Various drugs and/or drug candidates for diseases such as Influenza, Ebola, and COVID-19 targeting viruses-associated proteins are under clinical investigation, especially in the world struggling with global epidemics.
recently. There are vaccines available to combat the Ebola and COVID-19 pandemic, but the outbreak is still challenging due to the emergence of mutant strains of the virus, difficulties in producing and distributing vaccines, and more. Even those who have been vaccinated may not be protected against infection and disease, especially with variants that are less susceptible to current vaccines. Also, the frightening speed at which the Influenza, Ebola, and especially the COVID-19 pandemic has spread around the world has only served to reveal how inadequate our current antiviral drug options are. All of the repurposed antiviral drugs have accelerated treatment after rapidly conducting clinical trials. In this context, it is doubtful whether we will ever find a cure for Ebola by chance, but some of the published studies on known drugs may point us in the right direction as to where to look.

Remdesivir (GS-5734), one of the drugs used to combat these epidemics, is currently in clinical development for the treatment of Ebola virus disease (EVD) [30]. Remdesivir, a drug originally developed against the Ebola virus, is now also used for the treatment of COVID-19 [31]. Antiviral activities of Remdesivir on RNA polymerase (RdRp) have been reported against other coronaviruses such as the Ebola virus, MERS-CoV, SARS-CoV, CoV-OC43, CoV-229E and PDCoV [29].

Favipiravir, an orally administered drug with a similar mechanism of action as Remdesivir, has less strong supporting data to support its use, but still emerges as an agent worth considering in mild to moderate cases. Favipiravir, which shows activity against influenza, Ebola and COVID-19 viruses, has been found to have therapeutic activity in cell culture and mouse models of arenavirus, bunyavirus, filovirus, West Nile virus, yellow fever virus, foot and mouth disease virus. Previous studies have already shown that the combination of Favipiravir and Ribavirin may also be synergistic against the Lassa virus, which contains agents that cause viral hemorrhagic fever and encephalitis. Molnupiravir, the first oral antiviral drug candidate for use in the treatment of COVID-19, is active in several preclinical models of SARS-CoV-2, including prophylaxis, treatment, and prevention of transmission. Preclinical and clinical data have shown that Molnupiravir is active against the most common SARS-CoV variants.

Inmazeb is a combination of three monoclonal antibodies, atoltivimab (REGN3470), maftivimab (REGN3479), and odnivosimab (REGN3471), which target the Ebola virus glycoprotein. Based on the results of the PALM study conducted during an Ebola epidemic in the Democratic Republic of the Congo, REGN-EB3 was recently approved by the US FDA as a treatment for Ebola virus infection [32]. Atoltivimab-maftivimab-odesivimab (Inmazeb) is the first FDA-approved treatment for Zaire ebolavirus infection in adult and pediatric patients, including neonates born to a mother with reverse transcription-polymerase chain reaction (RT-PCR). The efficacy of Inmazeb has been established in vivo and has completed a phase I clinical trial in healthy subjects without any drug-related adverse events. In addition, Inmazeb showed a significant reduction in mortality in the PALM (PAmoja tuLindeMaisha) trial [33]. Bepridil, sertraline, and toremifene are orally available drugs that block EBOV and MARV infections in vitro [34].

The combination of Molnupiravir and Favipiravir, two oral drugs with a high barrier to resistance, for which there is the very recent initial evidence of exhibiting antiviral activity in COVID-19 patients, is particularly effective in the treatment of SARS-CoV2 [35]. The main purpose of this study is to emphasize the importance of investigating and examining such combinations of drugs used in the treatment of Ebola and to think that the same combination effect may be effective in the treatment of Ebola. For this, the atom pair algorithm was used for drug-drug similarity and then the Scaffold Network Graph algorithm was used to extract the molecular scaffolds of these drugs.

Another study selected Ulixertinib, a known ERK2 inhibitor, to perform scaffold jumping to discover new scaffolds with similar binding modes [36]. They then performed molecular insertion analysis of the hits with the highest similarity score to determine both the binding mode and affinity at the catalytic domain. As a result, they uncovered new lead molecules with putative ERK2 inhibitory potential that could be further validated through biological evaluation.

From a research perspective, drug discovery is very attractive given that reusing old drugs for the treatment of emerging diseases is more advantageous than de novo, especially in the case of an emerging viral disease. First, the risk of development is substantially lower because candidates for repurposed drugs will have already gone through several stages of clinical development. It will also translate into lower costs and faster development times as it will have safe and established pharmacological profiles. For some drug candidates, it may be possible to skip preclinical trials and enter directly into phase II clinical trials if the safety assessment and development formulation is
currently already completed. Second, the availability of drug screening libraries, and academic and small laboratories can play an important role in drug discovery. This will accelerate the time it takes for a biopharma to review the pharmacopoeia to search for suitable reuse candidates for the purpose [30].

4. Conclusion

On average, 10 years and approximately 2-3 billion dollars are spent to develop a single FDA-approved drug. But to combat sudden global threats like the Ebola epidemic is a very costly and lengthy process. Another way to discover a new drug is to reuse it. To re-use a clinically-trial drug for which the safety risk is already known, it remains only to test and analyze its ability to treat a new disease. In this study, the similarity of drugs in the literature for chemical structure similarities with atom pair algorithm was analyzed and examined to find new potential drug candidates that are already approved for the treatment of Ebola or that have passed the clinical trial phase. Then, the similar infrastructure molecules of these similar drugs and the scaffolds for the core molecule were extracted with the Scaffold Network Graph algorithm. Chemists need to know these beforehand.

Favipiravir is an effective drug used to treat many viral epidemics. Results were obtained by taking chemical structure similarities with other drugs as a reference in drug-drug similarity analysis procedures and extraction of scaffolds. According to these results, it was observed that the drug with the highest similarity to Favipiravir is Molnupiravir. The drug with the lowest molecular structure similarity is considered to be Azithromycin. There are practically no structural similarities. A total of 14 molecules, including the molecular and infrastructure molecular scaffolds of Favipiravir and Molnupiravir, were separated when the results were analyzed using the scaffold network graph algorithm. It was found that 11 of them have a similar structure and are related to each other. When we show these 14 molecular structures graphically, a large number of common nodes (11) indicate the similarity between drugs. Drugs with similar structures may show similar effects. This suggests that the drug-drug similarity is quite useful. As a result, the combination of Molnupiravir and Favipiravir, drugs and/or drug candidates currently being studied in clinical trials for the treatment of COVID-19, may form the basis for the design of clinical trials to test the effectiveness of the Ebola virus. Such a combination for the treatment of Ebola can also lead to the adjustment of appropriate dosages and the reduction of the amount of doses administered. The same combinations can be considered for other drugs on the list. The study offers a different approach to Ebola treatment in this aspect.

The identification of safe drugs or drug candidates to study and mitigate the effects of Ebola and other global epidemic crises and future outbreaks of pathogenic diseases has a crucial role in determining timely and effective therapeutic approaches. Evaluation and knowledge of chemical structure similarities or interrelated infrastructure molecules between these drugs are crucial for identifying those that have the potential to decimate Ebola virus replication and virulence. Computational approaches are an attractive method for identifying potential drug reuse opportunities, especially when in vitro or in vivo screening is difficult or impossible. Here, inferring similarities between U.S. FDA-approved and experimental drugs that have the potential to inhibit the proliferation and virulence of Ebola can be used as a stepping stone to confirm Ebola activity. This could open up new opportunities for designing new therapeutics for Ebola.

Favipiravir is an effective drug used in the treatment of many viral outbreaks. In the drug-drug similarity analysis procedures and the extraction of scaffolds, results were obtained by taking chemical structure similarities with other drugs as a reference. According to these results, it was observed that Molnupiravir was the drug with the highest similarity to Favipiravir. The drug with the lowest molecular structure similarity is seen as Azithromycin. There is almost no structural similarity. When the results made with the scaffold network graph algorithm were examined, a total of 14 molecules, including the molecule and infrastructure molecular scaffolds of Favipiravir and Molnupiravir, were separated. It was observed that 11 of them had a similar structure and were related to each other. When we show these 14 molecular structures graphically, the high number of common nodes (11) shows the similarity between the drugs. Drugs with similar structures may show similar effects. This shows that drug-drug similarity is quite useful. In conclusion, the combination of Molnupiravir and Favipiravir, which are drugs and/or drug candidates currently being investigated in clinical trials for the treatment of COVID-19, may form the basis for the design of clinical trials to test the efficacy of the Ebola virus. Such a combination for the treatment of Ebola may also result in the adjustment of appropriate dosages, and the lowering of
dose amounts administered. The same combinations can be considered for other drugs on the list. The study offers a different approach to Ebola treatment in this aspect.

Identifying safe drugs or drug candidates to study and mitigate the effects of the Ebola and other global epidemic crises and future outbreaks of pathogenic disease has a crucial role in determining timely and effective therapeutic approaches. Evaluation and knowledge of the chemical structure similarities or interrelated infrastructure molecules between these drugs are crucial to identify those that have the potential to inhibit Ebola virus replication and virulence. Computational approaches are an attractive method to identify potential drug reuse opportunities, especially when in vitro or in vivo screening is difficult or impossible. Here, extracting similarities of US FDA-approved and experimental drugs that have the potential to inhibit the proliferation and virulence of Ebola can be used as a stepping stone to confirm Ebola activity. This may present new opportunities for designing new therapeutics for Ebola.

Data Availability Statements:

Data analyzed in this study were a re-analysis of existing data, which are openly available at locations cited in the reference section ([22], [23]). Further documentation about data is available at https://go.drugbank.com/drugs and https://pubchem.ncbi.nlm.nih.gov/

References
[1] World Health Organization (WHO) Online, Retrieved Feb, 2019, from https://www.who.int/
[2] Xue, H., Li, J., Xie, H., & Wang, Y. (2018). Review of drug repositioning approaches and resources. International journal of biological sciences, 14(10), 1232.
[3] Bibi Nousheen, Farid Ayesha, Gul Sana, Ali Johar, Amin Farhat, Kalthiya Umesh, Hupp Ted, 2021, Drug repositioning against COVID-19: a first line treatment, Journal of Biomolecular Structure and Dynamics, 1, 15, 0739-1102, doi: 10.1080/07391102.2021.1977698
[4] Shtar, G., Rokach, L., & Shapira, B. (2019). Detecting drug-drug interactions using artificial neural networks and classic graph similarity measures. PloS one, 14(8), e0219796.
[5] Rifaigulu, A. S., Atas, H., Martin, M. J., Cetin-Atalay, R., Atalay, V., & Doğan, T. (2019). Recent applications of deep learning and machine intelligence on in silico drug discovery: methods, tools and databases. Briefings in bioinformatics, 20(5), 1878-1912.
[6] Zhao, X., Chen, L., Guo, Z. H., & Liu, T. (2019). Predicting drug side effects with compact integration of heterogeneous networks. Current Bioinformatics, 14(8), 709-720.
[7] Lo, Y. C., Rensi, S. E., Torng, W., & Altman, R. B. (2018). Machine learning in chemoinformatics and drug discovery. Drug discovery today, 23(8), 1538-1546.
[8] Chiu, Y. C., Chen, H. I. H., Zhang, T., Zhang, S., Gorthi, A., Wang, L. J., ... & Chen, Y. (2019). Predicting drug response of tumors from integrated genomic profiles by deep neural networks. BMC medical genomics, 12(1), 143-155.
[9] Cakir, A., Tuncer, M., Taymaz-Nikerel, H., & Ulucan, O. (2021). Side effect prediction based on drug-induced gene expression profiles and random forest with iterative feature selection. The Pharmacogenomics Journal, 21(6), 673-681.
[10] Zhao, Y., Zheng, K., Guan, B., Guo, M., Song, L., Gao, J., ... & Shi, D. (2020). DLDTI: A learning-based framework for identification of drug-target interaction using neural networks and network representation. bioRxiv.
[11] Vilar, S., Harpaz, R., Uriarte, E., Santana, L., Rabadan, R., & Friedman, C. (2012). Drug—drug interaction through molecular structure similarity analysis. Journal of the American Medical Informatics Association, 19(6), 1066-1074.
[12] Ibrahim, H., El Kerdawy, A. M., Abdo, A., & Eldin, A. S. (2021). Similarity-based machine learning framework for predicting safety signals of adverse drug–drug interactions. *Informatics in Medicine Unlocked*, 26, 100699.

[13] Zeng, X., Zhu, S., Lu, W., Liu, Z., Huang, J., Zhou, Y., ... & Cheng, F. (2020). Target identification among known drugs by deep learning from heterogeneous networks. *Chemical Science*, 11(7), 1775-1797.

[14] Schuffenhauer, A., Ertl, P., Roggo, S., Wetzel, S., Koch, M. A., & Waldmann, H. (2007). The scaffold tree—visualization of the scaffold universe by hierarchical scaffold classification. *Journal of chemical information and modeling*, 47(1), 47-58.

[15] Li, Y., Hu, J., Wang, Y., Zhou, J., Zhang, L., & Liu, Z. (2019). DeepScaffold: a comprehensive tool for scaffold-based de novo drug discovery using deep learning. *Journal of chemical information and modeling*, 60(1), 77-91.

[16] Bemis, G. W., & Murcko, M. A. (1999). Properties of known drugs. 2. Side chains. *Journal of medicinal chemistry*, 42(25), 5095-5099.

[17] Varin, T., Schuffenhauer, A., Ertl, P., & Renner, S. (2011). Mining for bioactive scaffolds with scaffold networks: improved compound set enrichment from primary screening data. *Journal of chemical information and modeling*, 51(7), 1528-1538.

[18] Scott, O. B., & Edith Chan, A. W. (2020). ScaffoldGraph: an open-source library for the generation and analysis of molecular scaffold networks and scaffold trees. *Bioinformatics*, 36(12), 3930-3931.

[19] Horrocks, P., Ullah, I., Wetzel, D., Sharma, R., Biagini, G., & Mete, A. (2019). The relative rate of kill of the MMV Malaria Box compounds provide links to the mode of antimalarial action and highlight scaffolds of medicinal chemistry interest. *Journal of Antimicrobial Chemotherapy*, 75(2), 362-370.

[20] Wilkens, S. J., Janes, J., & Su, A. I. (2005). HierS: hierarchical scaffold clustering using topological chemical graphs. *Journal of medicinal chemistry*, 48(9), 3182-3193.

[21] Kruger, F., Stiefl, N., & Landrum, G. A. (2020). rdScaffoldNetwork: the scaffold network implementation in RDKit. *Journal of Chemical Information and Modeling*, 60(7), 3331-3335.

[22] DrugBank Online. (2022). Retrieved Feb, 2022, from https://go.drugbank.com/drugs

[23] PubChem Online. (2022). Retrieved Feb, 2022, from https://pubchem.ncbi.nlm.nih.gov/

[24] Budak, C., Mençik, V., & Gider, V. (2021). Determining similarities of COVID-19–lung cancer drugs and affinity binding mode analysis by graph neural network-based GEFA method. *Journal of Biomolecular Structure and Dynamics*, 1-13.

[25] Matlock, M. K., Zaretzki, J. M., & Swamidass, S. J. (2013). Scaffold network generator: a tool for mining molecular structures. *Bioinformatics*, 29(20), 2655-2656.

[26] Wengong Jin, Regina Barzilay, and Tommi Jaakkola. (2018). Junction tree variational autoencoder for the molecular graph generation. ICML.

[27] Wengong Jin, Kevin Yang, Regina Barzilay, and Tommi Jaakkola. (2019). Learning multimodal graph-to-graph translation for molecular optimization. ICLR

[28] Fu, T., Gao, W., Xiao, C., Yasonik, J., Coley, C. W., & Sun, J. (2021). Differentiable scaffolding tree for molecular optimization. *arXiv preprint arXiv:2109.10469*.

[29] Ng, Y. L., Salim, C. K., & Chu, J. J. H. (2021). Drug repurposing for COVID-19: Challenges, promises and promising candidates. *Pharmacology & Therapeutics*, 107930.

[30] Tchesnokov, E. P., Feng, J. Y., Porter, D. P., & Götte, M. (2019). Mechanism of inhibition of Ebola virus RNA-dependent RNA polymerase by remdesivir. *Viruses*, 11(4), 326.

[31] Deb, S., & Reeves, A. A. (2021). Simulation of Remdesivir Pharmacokinetics and Its Drug Interactions. *Journal of Pharmacy & Pharmaceutical Sciences*, 24, 277-291.

[32] Markham, A. (2021). REGN-EB3: first approval. *Drugs*, 1-4.
[33] Saxena D, Kaul G, Dasgupta A, Chopra S. Atoltivimab/maftivimab/odesivimab (Inmazeb) combination to treat infection caused by Zaire ebolavirus. Drugs Today (Barc). 2021 Aug;57(8):483-490. doi: 10.1358/dot.2021.57.8.3280599. PMID: 34405205.

[34] Finch, C. L., Dyall, J., Xu, S., Nelson, E. A., Postnikova, E., Liang, J. Y., ... & White, J. M. (2021). Formulation, Stability, Pharmacokinetic, and Modeling Studies for Tests of Synergistic Combinations of Orally Available Approved Drugs against Ebola Virus In Vivo. Microorganisms, 9(3), 566.

[35] Abdelnabi, R., Foo, C. S., Kaptein, S. J., Zhang, X., Do, T. N. D., Langendries, L., ... & Neyts, J. (2021). The combined treatment of Molnupiravir and Favipiravir results in a potentiation of antiviral efficacy in a SARS-CoV-2 hamster infection model. EBioMedicine, 72, 103595.

[36] Pathania, S., Singh, P. K., Narang, R. K., & Rawal, R. K. (2021). Identifying novel putative ERK1/2 inhibitors via hybrid scaffold hopping–FBDD approach. Journal of Biomolecular Structure and Dynamics, 1-16.