applied sciences

MDPI

Article

Optimal Paradigms for Quantitative Modeling in Systems Biology Demonstrated for Spinal Motor Neuron Synthesis

Gülbahar Akgün [1,†] and Rza Bashirov [2,*,†]

[1] Department of Computer Engineering, Faculty of Engineering, Başkent University, Fatih Sultan, Bağlıca Campus, Eskişehir Road 18th km, 06790 Ankara, Türkiye; gulbaharakgun@baskent.edu.tr
[2] Department of Mathematics, Faculty of Arts and Sciences, Eastern Mediterranean University, North Cyprus, Mersin-10, 99628 Famagusta, Türkiye
* Correspondence: rza.bashirov@emu.edu.tr
† These authors contributed equally to this work.

Abstract: Since the 1990s, Petri nets have been used in systems biology for quantitative modeling. Despite the increasing number of models developed during this period, doubts remain about their biological relevance. Although biological systems predominantly exhibit intracellular or cellular structures, the models rely largely on deterministic predictions, failing to capture the inherent randomness and uncertainties of such systems. The question arises whether these models accurately describe the dynamic behavior of biological systems. This paper introduces a methodology for selecting the appropriate modeling paradigms in systems biology. Initially, we construct a Petri net model and perform deterministic, stochastic, and fuzzy stochastic simulations. Then we perform various statistical tests to measure the discrepancies between the simulation results. Based on scale-density analysis, we determine the modeling approach that best approximates the biological system. Finally, we compare the results of the statistical tests and the scale-density analysis to identify the optimal modeling approach. We applied the proposed methodology to the synthesis of spinal motor neuron protein from the spinal motor neuron-2 gene. Analysis revealed significant discrepancies between the simulation results of different modeling paradigms. Due to the sparse nature of the underlying drug-disease network, we conclude that the fuzzy stochastic paradigm provides the most biologically relevant results. We predict drug combinations that could lead to an up to 149-fold increase in spinal motor neuron protein levels, indicating a promising treatment for the disease. This methodology has the potential for application to other gene-drug-disease networks and broader biological systems.

Keywords: systems biology; quantitative modeling; Petri nets; descriptive statistics; spinal muscular atrophy

Citation: Akgün, G.; Bashirov, R. Optimal Paradigms for Quantitative Modeling in Systems Biology Demonstrated for Spinal Motor Neuron Synthesis. *Appl. Sci.* **2024**, *14*, 10696. https://doi.org/10.3390/app142210696

Academic Editor: Luca Cattelani

Received: 5 October 2024
Revised: 28 October 2024
Accepted: 15 November 2024
Published: 19 November 2024

1. Introduction

Currently, there is no universally accepted standard for quantitative modeling in systems biology, as it must be highly customized to the specific biological problem at hand. Each model can only offer an approximate representation of nature, and a key challenge is to apply or develop modeling techniques that best address the questions posed while balancing effort and accuracy. Deterministic and stochastic methods are commonly employed in modeling biological systems, leading to the question of what is the most suitable modeling framework for a given biological system. A straightforward analysis reveals the complexity of this decision. A single molecule behaves in a completely stochastic manner, but as the number of molecules increases, stochastic effects diminish due to balancing interactions. The level of stochasticity, represented by the function $\frac{1}{\sqrt{n}}$ of the number of components n in the biological system, decreases as n increases. In the thermodynamic limit, $n \to \infty$, a stochastic formulation effectively reduces to the deterministic one [1]. We

know that sparse biological systems primarily exhibit stochastic behavior, whereas systems with a high density of components tend to follow deterministic patterns. Determining whether a particular biological system is sparse or dense can be challenging. In the following discussion, we will focus on the scale and density characterization of biological systems to determine if a given system is predominantly sparse or dense.

Biological organization spans various spatial scales, with the most notable being intracellular, cellular, intercellular, and population-wide [2,3]. Intracellular processes occur within the cell and involve interactions at the molecular level such as gene regulatory, metabolic, and signal transduction pathways. Intracellular processes take place at the microscopic level. Cellular processes occur at the mesoscopic level but are not limited to a single cell. Cell biology as a whole is fundamentally mesoscopic biology. For example, cell communication involves multiple cells, but takes place at the cellular level. Cellular processes involve interconnected physiological systems that regulate cell function, including biochemical cascades both within and between cells. Intercellular biology focuses on the various structures that cells use to communicate with each other, either directly or through the surrounding biological environment, typically at the tissue level, contributing to organ functionality. Finally, population-wide processes study how populations interact with their environment, such as the spread of epidemiological diseases such as COVID-19 in specific geographical regions [4–6]. Intercellular and population-wide processes both occur at the macroscopic level.

Intracellular and cellular processes form microscopic and mesoscopic systems, respectively, that consist of a relatively low number of biological components and can, therefore, be described as sparse biological systems. Due to their inherent randomness, these processes are best investigated using stochastic methods. In contrast, intercellular and population-wide processes are macroscopic systems represented by a crowd of biological components and, therefore, are characterized by high density. Deterministic models are most suitable for modeling these macroscopic systems. Figure 1 illustrates how the spatial scale and density analysis of biological systems discussed previously affects the selection of modeling methods.

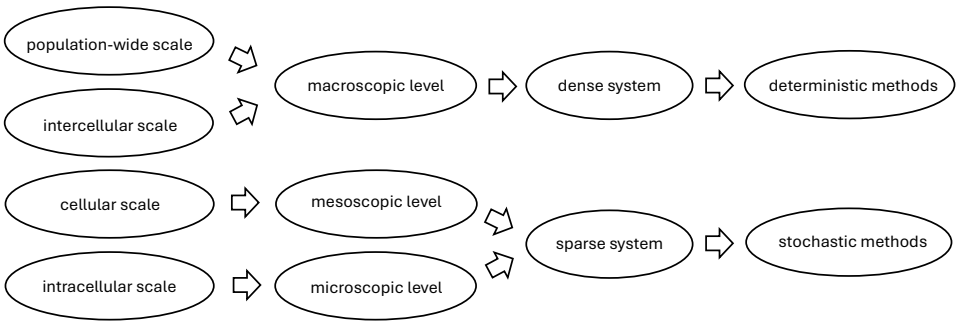

Figure 1. Illustration of how spatial scale and density analysis of biological systems influence the choice of modeling methods.

It is common for a biological phenomenon to take place at varying rates. The reaction rates are typically vague and rather uncertain. This vagueness affects the stoichiometry of the reactions and the quantitative relationships between substances in biochemical reactions. As a result, quantitative kinetic models inherently possess some degree of uncertainty. Fuzzy logic [7] is often used to manage the uncertainty inherent in biological systems.

In line with the preceding discussion, it is clear that deterministic, purely stochastic, and fuzzy stochastic paradigms represent three principal possibilities for modeling in systems biology. The choice of modeling paradigm should be based on the model's scope and specific characteristics. In deterministic models, the outcome remains constant regardless of how many times the model is run, making them straightforward to analyze since they only need to be executed once. However, stochastic models yield different results with

each run, requiring thousands or even tens of thousands of replications to obtain reliable averages. This complexity makes stochastic models more challenging to analyze compared to deterministic ones. The fuzzy stochastic approach adds another layer of complexity by conducting stochastic simulations for each fuzzy number, making the analysis even more difficult.

This study proposes a methodology to identify the optimal modeling paradigm by combining insights from the analysis of differences between simulations across three paradigms with the scale density analysis of biological systems. If the data sets from two or all three paradigms are sufficiently similar, the simplest paradigm should be preferred. Otherwise, the most appropriate approach is recommended for each biological system. Figure 2 schematically illustrates how the similarities and discrepancies between the results of deterministic, purely stochastic, and fuzzy stochastic simulations affect the selection of a modeling paradigm. The authors are unaware of any other work that addresses this methodological issue in the given context, although there are studies that compare various aspects of deterministic, stochastic, and fuzzy methods [8,9]. In [10], we demonstrated how to select a relevant modeling framework for the p16-mediated pathway. The p16-mediated signaling pathway is a microscopic biological system, which is known to exhibit sparsity and inherent uncertainty. As a result, the p16-mediated pathway is most effectively studied using a fuzzy stochastic approach. In general, it can be stated that almost half of the scientific papers focused on modeling biological systems focus on metabolic networks, signal transduction pathways, and gene regulatory networks. About half of these studies employ a deterministic modeling approach. Each of these networks or pathways, being at the microscopic level, is characterized by high sparsity, making stochastic or fuzzy stochastic methods more suitable for their modeling. This is the primary issue highlighted as a problem in the present work.

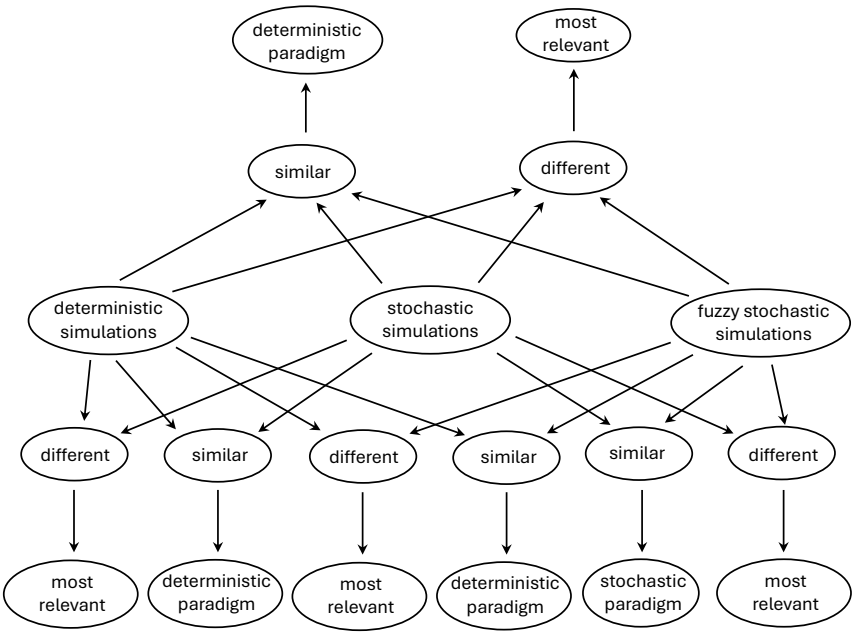

Figure 2. An illustration of how the similarities and differences between results from deterministic, purely stochastic, and fuzzy stochastic simulations affect the selection of a modeling paradigm.

The applicability of this methodology is demonstrated for the synthesis of spinal motor neuron (SMN) protein from spinal motor neuron 2 gene (*SMN2*) [11]. We used descriptive statistics tests to measure the similarity among data sets: the Shapiro–Wilk test to assess normality, the Friedman test to compare all three data sets, and the Dunn–Bonferroni test to pairwise compare the data sets. For the synthesis of SMN, the Shapiro–Wilk test indicated

a non-normal distribution of the data sets, while the Friedman and the Dunn–Bonferroni tests revealed significant differences between the data sets. Due to the fact that the SMN synthesis network operates at the cellular scale, the fuzzy stochastic approach is concluded to provide the most relevant results.

The remainder of the paper is structured as follows: In the next section, we outline the research method, material, and background behind the present work. More specifically, we outline Petri nets, succinctly describe fuzzy logic, review statistical tests to measure similarities among data sets, and introduce associating biological context. Following this, we develop the model, validate the model with known biological data, describe simulations, introduce the clustering algorithm to classify the combinations of drugs into effective and noneffective classes, provide data analysis of simulations, and interpret the statistical tests. The paper concludes with our findings.

2. Materials and Methods

2.1. Petri Nets

We explore Petri net technologies to develop a model of the drug-disease network for the synthesis of the SMN protein from *SMN2* gene. This model incorporates Petri nets, continuous Petri nets (CPNs), hybrid Petri nets (HPNs), and stochastic Petri nets (SPNs). In what follows, we formally define the Petri net, CPN, HPN, and SPN, and for further elaboration on the Petri nets and corresponding extensions, we direct the readers to [12,13].

A Petri net is a 5-tuple $R = (P, T, I, O, M_0)$ such as $P = \{p_1, \ldots, p_n\}$, $T = \{t_1, \ldots, t_m\}$, $I \subseteq P \times T$, $O \subseteq T \times P$, $M_0 : P \to \mathbb{N}$ where P, T, I, O, and M_0 are the sets of places, transitions, input arcs, output arcs, and initial marking, respectively, $P \cap T = \emptyset$, and $\mathbb{N}$ is the set of natural numbers. The weight function $W : (P \times T) \cup (T \times P) \to \mathbb{Z}_0^+$ assigns to each arc a non-negative integer called an arc multiplicity (or weight) where $\mathbb{Z}_0^+$ is the set of non-negative integers. Given transition t and place p, the weight of an arc from t to p, written $w(t, p)$, specifies the number of tokens to be added to p when t occurs. Similarly, the weight of an arc from p to t, written $w(p, t)$, specifies the number of tokens to be removed from p when t occurs. It must be noted that no arc may connect two places or two transitions. In a Petri net, a transition t is said to be enabled if each input place p is marked with at least $w(p, t)$ tokens. Thus, t is enabled if $M(p) \geq w(p, t)$ is valid for all $p \in {}^\circ t$ where ${}^\circ t$ is the set of input places of t. A transition t is said to be disabled in the marking M if there exists an input place p with a number of tokens less than $w(p, t)$. That is, t is disabled if there exists $p \in {}^\circ t$ such that $M(p) < w(p, t)$ holds. An enabled transition may or may not fire or occur, depending on whether or not the event actually takes place. Occurrence of an enabled transition t removes $w(p, t)$ tokens from $p \in {}^\circ t$, and adds $w(t, p)$ tokens to $p \in t^\circ$, where t° represents the set of output places t. More mathematically, the occurrence of the enabled transition t changes the state of the Petri net from M to M' as follows:

$$M'(p) = M(p) - w(p, t) + w(t, p), \forall p \in P.$$

A CPN is a 5-tuple $R = (P, T, Pre, Post, M_0)$, such as P and T are the same as those in the definition of Petri net, and Pre is a function from $P \times T$ to $\mathbb{Q}^+$, $Post$ is a function from $T \times P$ to $\mathbb{Q}^+$, and M_0 is from P to $\mathbb{R}_0^+$ where Pre and $Post$ are weight functions, which associate a positive rational number with the arcs $(p, t) \in P \times T$ and $(t, p) \in T \times P$, respectively, M_0 is a vector of non-negative real numbers, $\mathbb{Q}^+$ is the set of positive rational numbers, and $\mathbb{R}_0^+$ is the set of non-negative real numbers. Let R be a basic Petri net with n places and l transitions, and let $M = (m_{p_1}, \ldots, m_{p_n})$ be any marking reachable from initial marking, where m_{p_i} indicates the number of tokens in p_i, $1 \leq i \leq n$. To create an equivalent CPN, we start by dividing each token into k new tokens, each representing one kth of the original token while preserving the structure of the net. In a new Petri net, markings are expressed in terms of new tokens, e.g., M is transformed into $M' = (m'_{p_1}, \ldots, m'_{p_n})$ where $m_{p_i} = k \cdot m'_{p_i}$ for all i. The firing rule operates on new tokens in a manner similar to the basic case: The occurrence of a transition removes new tokens from input places and adds new tokens to output places. However, as k increases, new tokens become increasingly

fine-grained, leading to smoother firing that resembles continuous flow. When k tends to infinity, the set of reachable markings becomes infinite, and a marking can no longer be expressed in terms of tokens since its components can become infinitely large. Instead, a marking is represented by a vector of real numbers, $m = (\alpha_1, \ldots, \alpha_n)$, to ensure continuity, while rational numbers can still be used for arc weights since they have concrete values. Consequently, the transition t_j is enabled in the marking m if and only if $\alpha_i > 0$ for $p_i \in {}^\circ t_j$ and its degree of enabling is determined by $enab(t_j, m) = \min_{p_i \in {}^\circ t_j} \frac{\alpha_i}{Pre(p_i, t_j)}$. Starting from marking m, firing t_j by an amount β (or β times simultaneously in one go), denoted by $[t_j]^\beta$, where $0 < \beta \leq enab(t_j, m)$, results in a new marking $m' = m + \beta \cdot C[P, t_j]$, where $C[P, t_j]$ represents the column corresponding to t_j in the incidence matrix.

An HPN is a 6-tuple $R = (P, T, h, Pre, Post, M_0)$, where P, the set of places, is composed of subset of D places, P^D, and subset of C places, P^C, and $P \neq \emptyset$, $P = P^D \cup P^C$, $P^D \cap P^C = \emptyset$; T, the set of transitions, consists of subset of D transitions, T^D, and subset of C transitions, T^C, and $T \neq \emptyset$, $T = T^D \cup T^C$, $T^D \cap T^C = \emptyset$; h, hybrid function that discriminates between D and C components, is a rule from $P \cup T$ to $\{D, C\}$, Pre is a function from $P \times T$ to $\mathbb{Q}^+$; $Post$ is a function from $P \times T$ to $\mathbb{N}$; and M_0 is same to that in the definition of CPN, Pre and $Post$, the weight functions, set the weight of the arc (p, t) and (t, p), respectively, to a positive rational number if $p \in P^C$, and to a natural number if $p \in P^D$.

An SPN is a 6-tuple $R = (P, T, I, O, M_0, \Lambda)$ where P, T, I, O, and M_0 are the same as those in the definition of the classical untimed Petri net, and $\Lambda = (\lambda_1, \ldots, \lambda_m)$ is a vector of real numbers, where λ_k represents the firing rate of t_k, $1 \leq k \leq m$. We apply SPNs to model the association between random time delays and the occurrence of transitions. Within an SPN framework, the time interval between the enabling of a transition and its subsequent occurrence is characterized by a random variable following a negative exponential probability distribution function as follows:

$$F(t) = 1 - e^{-\lambda t} \text{ if } t \geq 0 \text{ and } F(t) = 0 \text{ otherwise, } \lambda > 0.$$

In the use of extended Petri nets, we sometimes need to introduce additional arc types beyond the standard arcs defined earlier. Two commonly used nonstandard arcs in this research are the inhibitor arc and the read (or test) arc. An inhibitor arc imposes a condition that a transition can only fire if the connected place is empty. A read arc, on the other hand, models a scenario in which a resource (such as an enzyme in a chemical reaction) is required but not consumed during firing. A read arc is represented as an arc with a solid circle and an inhibitor arc with a hollow circle [14].

2.2. Fuzzy Logic

Fuzzy logic, first introduced by Zadeh [7], provides a systematic approach to handle uncertainty, imprecision, and vagueness. In this framework, a fuzzy set $\widetilde{\zeta}$ on a universal set of real numbers $\mathbb{X}$ is defined by its membership function $\mu_{\widetilde{\zeta}}$ which maps $\mathbb{X}$ to $[0, 1]$. This function assigns to each element $x \in \mathbb{X}$ a real number $\mu_{\widetilde{\zeta}}(x)$ in the interval $[0, 1]$. To enable the use of fuzzy rules, Petri nets with various extensions have been integrated with fuzzy logic. In [13,15], the initial applications of SPNs with fuzzy parameters, the authors demonstrated how continuous stochastic Petri nets with fuzzy parameters can be applied in systems biology, and discrete stochastic Petri nets with fuzzy parameters can model a flexible manufacturing cell, respectively. In [16], it was shown how Petri nets can be applied to identify effective targets for β-globin disorders.

There are different types of fuzzy numbers: triangular, trapezoidal, pentagonal, hexagonal, heptagonal, etc. In our work, we use the triangular fuzzy number (TFN) to represent reaction rates and concentrations. An TFN $\widetilde{\zeta}$ is described by three values $a < b < c$, where $[a, c]$ forms the base of the triangle and b represents its peak.

Here is an explanation of how we determine a TFN for the reaction rates, which is also illustrated in Figure 3. We always have qualitative knowledge on reaction rates at our

disposal, such as translation being almost ten times slower than transcription. First, based on existing qualitative knowledge, we normalize the crisp value of a reaction rate b relative to the transcription rate, which is set to 1. During the validation stage of the model, we fine-tune the reaction rates by adjusting those of the corresponding transcriptions. Once these two stages are completed, the reaction rate $x = b$ is considered acceptable yieldings $\mu(x) = 1$; otherwise, $\mu(x) = 0$. Next, we define the TFN, (a, b, c), of "acceptable rates" for each reaction by assigning a degree of membership to each rate value x, where $\mu(x) \in [0, 1]$. Notably, $\mu(x)$ increases with x in the interval (a, b) and decreases in the interval (b, c). In this research $a = b - \frac{b}{2}$ and $c = b + \frac{b}{2}$.

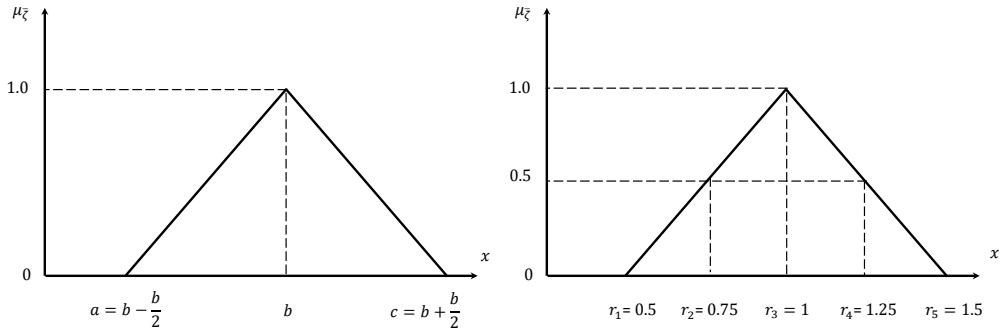

Figure 3. (**a**) The definition of a triangular fuzzy number, and (**b**) its adaptation to the rate of transcription.

Figure 3b illustrates the application of the above principle for the phenomenon of transcription. The TFN is implemented with two α cuts: $\alpha = 0.5$ and $\alpha = 1.0$, yielding the crisp subsets $[0.75, 1.25]$ and $\{1.0\}$, respectively. Finally, the set of fuzzy rates for the transcription phenomenon is defined using these α cuts as $\theta = \{r_1 = 0.0, r_2 = 0.5, r_3 = 1.0, r_4 = 1.25, r_5 = 1.5\}$.

2.3. Statistical Tests

There are various methods available for assessing the similarities between simulation results, such as descriptive statistics, correlation analysis, cluster analysis, and machine learning techniques. In this study, we utilize the methods of descriptive statistics, which are increasingly significant in biomedical research. Following the proposed methodology, we first conduct a normality test, a critical stage in determining the appropriate statistical methods. If the data are normally distributed, we use parametric tests; otherwise, we apply nonparametric methods for comparing the data sets. To evaluate similarities among a group of data sets, we perform two types of independent tests: pairwise comparisons of data sets and simultaneous comparisons of all data sets.

We used the Shapiro–Wilk test to check if the data sets align with a normal distribution. If we observe that the data are normally distributed, we can employ a parametric test to determine significant differences between the data sets; otherwise, we recommend using non-parametric tests: the Friedman test to identify significant differences among data sets in a group and the Dunn–Bonferroni posthoc test for pairwise comparisons. The latest test is designed to pinpoint which specific pairs of means within a data set are significantly different from each other.

2.4. Case Study—Synthesis of Spinal Motor Neuron Protein from Spinal Motor Neuron 2 Gene

A mutation in *SMN1* gene leads to the insufficient production or complete absence of SMN protein, causing spinal muscular atrophy (SMA), a debilitating motor neuron disease. Humans have a second gene copy, *SMN2*, which possesses the necessary genetic information to mitigate the effects of SMA. However, *SMN2* cannot fully compensate for

the loss of *SMN1* because it only produces 10–15% of the full functional SMN protein, the remaining 85–90% being the truncated and dysfunctional SMNΔ7 protein. To address this, various candidate drugs have been proposed to treat SMA or reduce its severity by increasing the production of functional SMN protein from *SMN2*.

As of now, there is no complete cure for SMA. The most effective therapies include Zolgensma [17,18], a gene replacement therapy, and two medications: nusinersen (marketed as Spinraza) and risdiplam (known as Evrysdi) [19,20]. Although these treatments represent major advances in SMA therapy, they are focused primarily on alleviating symptoms and preventing complications rather than offering a complete cure. However, understanding the genetic basis of SMA has allowed the development of targeted treatment options that fall into four main categories:

A. Inhibition of the inhibitor: Targeting pathways that suppress *SMN2* expression.
B. Regulation of pre-mRNA splicing: Enhancing the correct splicing of the *SMN2* transcript.
C. Upregulation of promoter activity: Boost promoter activity to increase *SMN2* expression.
D. Targeting DNA methylation: Modifying DNA methylation patterns to enhance *SMN2* activity.

In the present study, we combine various treatment options to evaluate the comprehensive effectiveness of the following 7 potential drugs or chemicals with existing qPCR data and all possible combinations of these drugs. Thus, we conducted dry lab experiments for 7 drug candidates and their 120 possible combinations (combinations of 2 drugs, combinations of 3 drugs, etc.). A possible SMA treatment strategy, referred to as strategy A, focuses on inhibiting histone deacetylase (HDAC) activity, which suppresses *SMN2* gene expression. Our research has identified Valproic Acid (VPA) [21], Trichostatin A (TSA) [22], Dacinostat, and Resveratrol as the only HDAC inhibitors documented in the biological literature that present qPCR and protein data indicating an increase in SMN protein levels derived from *SMN2*. According to [21], treating fibroblast cultures from SMA patients with 0.5–500 µM of VPA led to a boost in SMN levels up to four times. In [22], it was observed that the application of TSA to an SMA mouse model resulted in a twofold increase in SMN levels in the brain, liver, and spinal cord. Furthermore, in [23], a 1.3-fold increase in SMN levels was reported in cell cultures treated with 100μ resveratrol, compared to the untreated control.

In [24], it was demonstrated that in a type III SMA mouse model, treatment with the tetracycline derivative PTMK-SMA1 significantly increases the inclusion of exon 7 in *SMN2* mRNA during splicing. This is in accordance with strategy B. This leads to a nearly five-fold increase in SMN protein levels compared to untreated mice. Notably, PTMK-SMA1 is the only compound identified so far that directly alters the splicing process to favor the inclusion of exon 7, promoting the production of functional SMN protein.

In [25], according to strategy C, the authors suggest enhancing SMN transcription from *SMN2* by targeting modulation of the *SMN2* promoter's activity. The research demonstrated that Indole treatment in patient-derived cells significantly impacts *SMN2* promoter activity, leading to a three-fold increase in SMN transcription compared to untreated controls.

In line with strategy D, it was discovered that the *SMN2* gene is subject to silencing through DNA methylation [26]. Consequently, inhibiting this methylation process to prevent *SMN2* gene silencing has emerged as a promising strategy for the pharmacologic treatment of SMA. A potential drug, 5-Azacytidine (AZA), works by blocking the methylation of transcription factors associated with the *SMN2* gene, thereby increasing SMN protein production. Their study reported a twofold increase in SMN protein levels in SMA patients treated with AZA.

3. Results

3.1. Creating the Model

We used the Snoopy tool [14] and the biological context detailed in Section 2.4 to develop a Petri net-based model of a drug-disease network for the synthesis of SMN protein from *SMN2*. A Petri net model incorporates discrete, continuous, hybrid, stochastic, and fuzzy parameters and aspects. This model captures all the essential interactions among biological components within the framework of mass action kinetics [27]. We meticulously defined the reaction rates by anchoring them to a baseline transcription rate of 1, and then expressing the rates of other processes relative to this transcription rate. The kinetic rates applied in our study are consistent with those reported in [10].

The Petri net model is illustrated in Figure 4, incorporating features from discrete, continuous, hybrid, and stochastic Petri nets. In this model, discrete places represent the inclusion or exclusion of various drug candidates, while continuous components manage the flow of tokens, represented by real numbers, across the network. The discrete places include Dacinostat, TSA, Resveratrol, VPA, AZA, PTMKSMA1, and Indole, while the other 11 places are continuous, corresponding to genes and gene products whose concentrations change gradually over time. The model features 25 transitions, 2 read arcs, 7 inhibitory arcs, and 32 regular arcs, with no guard function assigned to the transitions.

Drug treatments are simulated across four scenarios (A to D), detailed in Section 2.4, by introducing an inhibitory arc from a discrete place to a transition. Each drug candidate has a corresponding discrete place connected to its transition via an inhibitor arc. A treatment is activated when the discrete place is empty and is deactivated otherwise. For combination treatments, the related discrete places are kept empty. A Boolean variable is used to monitor the presence or absence of drug treatment for each candidate. For stochastic simulations, stochastic rates are applied as delays for stochastic transitions. We use read arcs to model the constant presence of transcription factor and *SMN2* gene in DNA. For this reason, the input place for the TF_producer transition maintains a constant value of 1 and is connected to the transition via a read arc. Similarly, the input place for the SMN2_gene transition also has a constant value of 1 and is linked to the transition with a read arc. As a result, neither the transcription factor nor *SMN2* is consumed within the system.

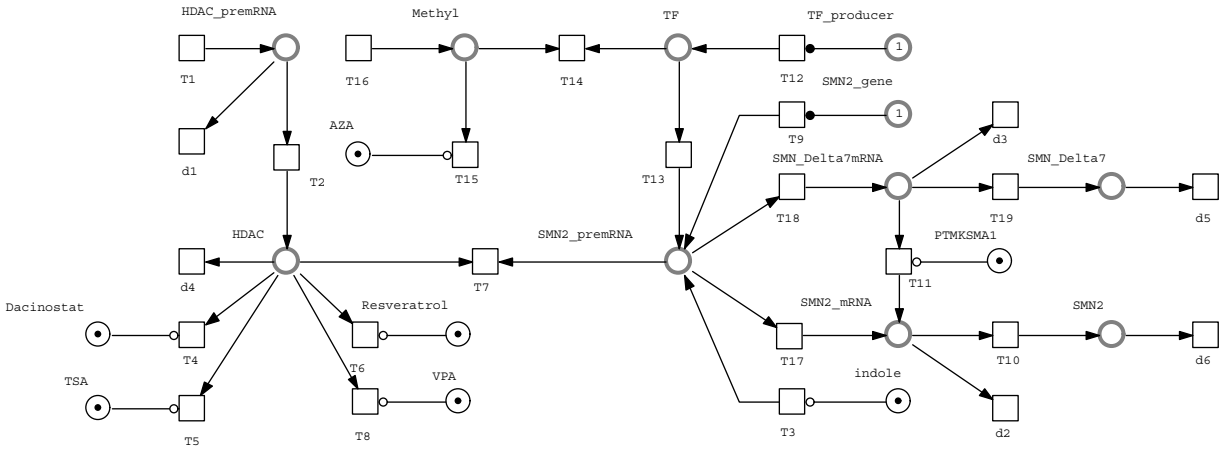

Figure 4. Petri net model of drug-disease network of SMN synthesis from *SMN2*.

3.2. Validation with Existing Biological Data

We employed a two-stage simulative validation of the model using biological knowledge detailed in Section 2.4. In the first stage, we reached the target ratio of protein types produced from *SMN2*, specifically 85% SMNΔ7 and 15% SMN. This is achieved by fine-tuning the rates of transitions *T*7, *T*10, *T*12, *T*13, *T*17, *T*18, and *T*19 until the desired proportion between SMNΔ7 and SMN is obtained. In the second stage, for each of the seven candidate drugs, we individually calibrate the transition rate to known qPCR data [21–26]. The rates of mRNA and protein degradation transitions are aligned with those used in

previous studies [10]. Finally, we fine-tune the reaction rates through stochastic simulations, averaging the results to ensure accuracy.

3.3. Conducting the Simulations

We run the model for each of the 127 drugs and drug combinations up to 1000 Petri time units (pt), a measure of time in the Petri net model, to ensure that the concentration of SMN reaches a steady state. For stochastic simulations, we performed 38,000 replications of the model, averaging the results for each drug or combination of drugs to reliably estimate the behavior of the SMN synthesis network. This approach provides a 95% confidence level and an accuracy of 10^{-2}.

In the fuzzy stochastic simulations, we replaced the crisp kinetic parameter b in the hazard function of each transition with a fuzzy number (a, b, c), where $a = b - \frac{b}{2}$ and $a = b + \frac{b}{2}$. The fuzziness of the model is represented by the rates r_1, r_2, r_3, r_4, and r_5, which correspond to specific α-cuts and are treated according to stochastic rules. We limited the number of α-cuts to two in this study, as increasing the number of cuts would significantly raise the number of replications required for the stochastic simulations. For both deterministic and stochastic simulations, the process rates were set to r_3, the middle component θ.

3.4. Clustering

In this study, we employ equivalence relation-based clustering to classify drug combinations based on their efficacy. The Drug_Clustering algorithm divides the set of combinations of n drugs into clusters of effective and non-effective combinations for $n = 1, \ldots, 7$. The key characteristic of this algorithm is that the lowest effectiveness (fold over control) of any effective combination of n drugs is higher than the highest effectiveness of any non-effective combination of n drugs. This yielded two clusters: (i) one cluster of 35 effective drug combinations that are likely to be beneficial and (ii) a larger cluster of 92 non-effective drug combinations. All effective combinations of drugs are represented in Figures 5–10.

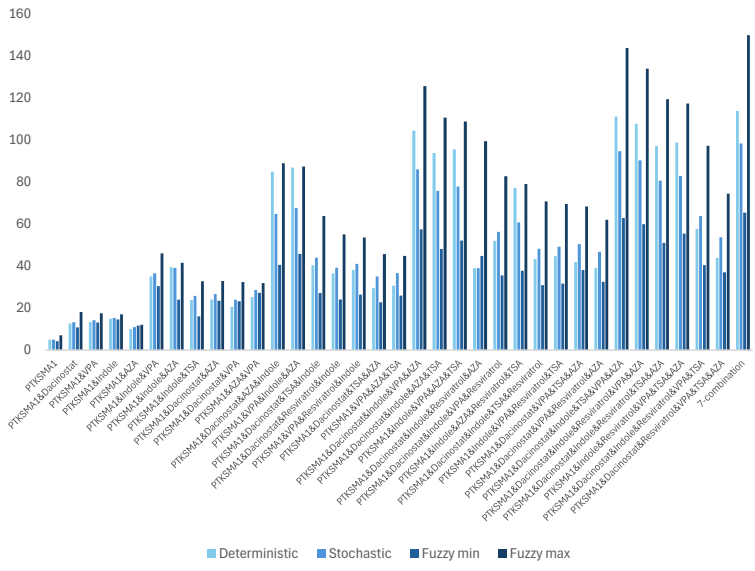

Figure 5. Impact of effective combinations of drugs on the SMN protein folding (expressed in folds).

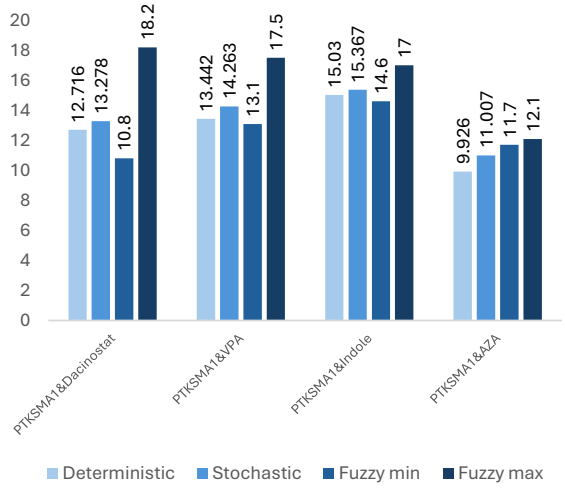

Figure 6. Impact of effective combinations of two drugs on the SMN protein folding (expressed in folds).

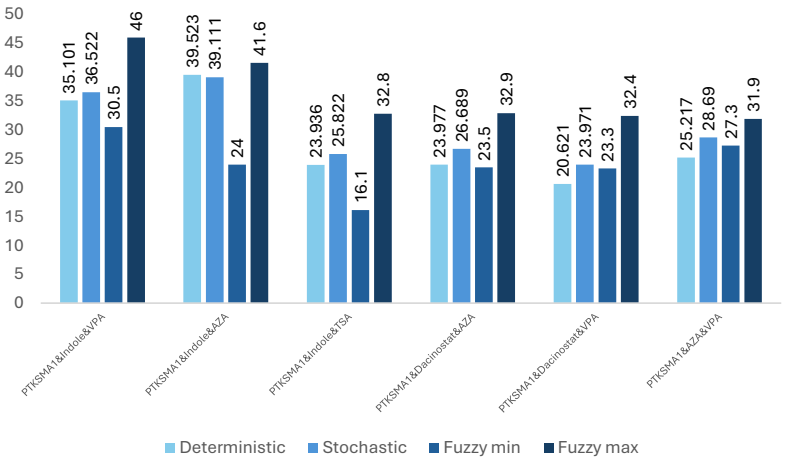

Figure 7. Impact of effective combinations of three drugs on the SMN protein folding (expressed in folds).

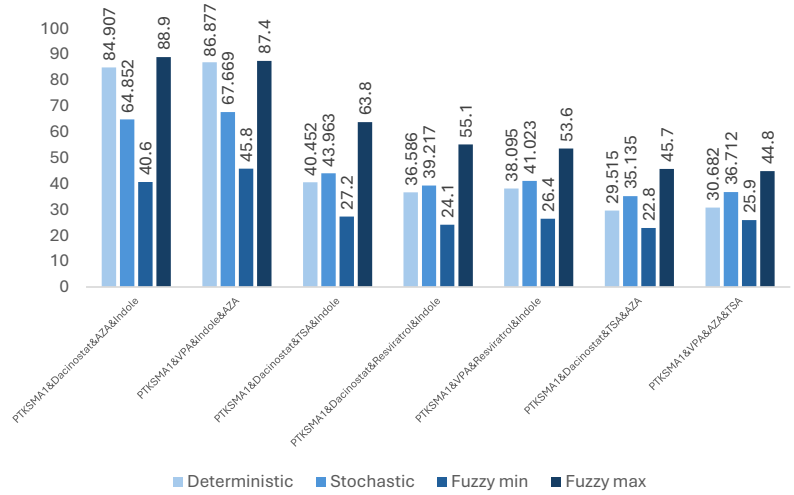

Figure 8. Impact of effective combinations of four drugs on the SMN protein folding (expressed in folds).

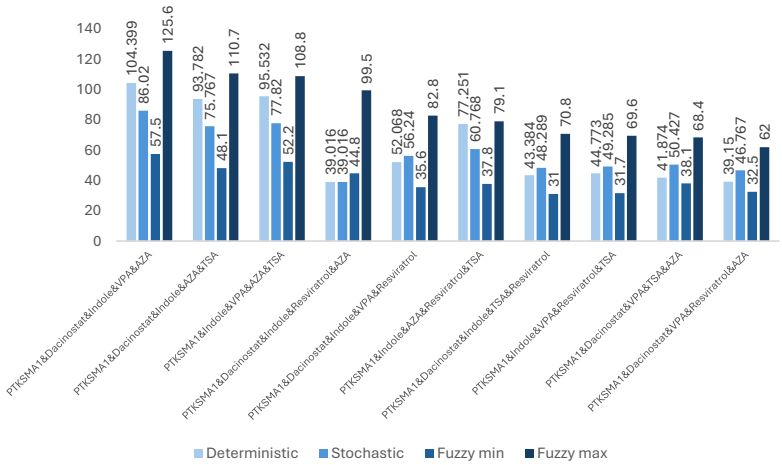

Figure 9. Impact of effective combinations of five drugs on the SMN protein folding (expressed in folds).

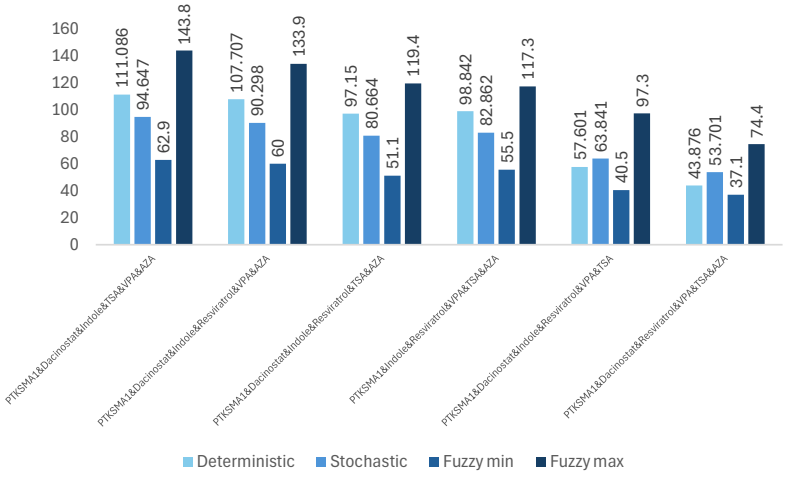

Figure 10. Impact of effective combinations of six drugs on the SMN protein folding (expressed in folds).

3.5. Data Analysis of Simulation Results

Based on the simulation results, combining all seven candidate drugs results in the most significant increase in SMN levels obtained from *SMN2*, achieving a 149.9-fold increase in SMN levels. This is a substantial breakthrough compared to the five-fold increase in the case of PTK-SMA1 [24]. However, despite their compatibility, using all seven drugs together can lead to severe side effects. Even if some combinations result in unavoidable side effects, there might still be many effective combinations worth testing. Figures 5–10 show that the worst result achieved with an effective combination of two drugs leads to a 9.926-fold increase in protein levels (see Figure 6). This is almost twice as high as reported in [24] for the best drug candidate.

Our observations indicate that adding more drugs does not always lead to higher levels of SMN. There are many such examples where fewer drugs achieve better results. For example, a three-drug combination of PTK-SMA1, Indole, and VPA results in up to a 46-fold increase in SMN levels (see Figure 7), which exceeds that of some four- or even five-drug combinations (see Figures 8 and 9). Similarly, a four-drug combination of PTK-SMA1, Dacinostat, AZA, and Indole results in up to an 88.9-fold increase in SMN levels (see Figure 8), which exceeds that of most five- and even some six-drug combinations (see

Figures 9 and 10). Furthermore, combinations involving PTK-SMA1 often outperform larger combinations that do not include it.

3.6. Interpretation of Statistical Tests

To precisely assess the level of agreement or discrepancy among the data sets obtained from simulations for the best effective combinations of drugs, we utilized Statistical Package for the Social Sciences Release 13.0 (SPSS Release 13.0) Package for a detailed analysis. Initially, we conducted normality tests across all data sets, which revealed that none of the data sets followed a normal distribution. Consequently, we applied nonparametric statistical tests to make pairwise comparisons between the data sets derived from the different models.

First, we conducted the Friedman test. The probability of obtaining the observed results by chance referred to as p-value, less than 0.05, is typically considered statistically significant, indicating that the hypothesis that there are no differences between the groups is rejected. For the case study of SMN synthesis, this value was $p < 0.0001$. The direct implication of this observation is that for the seven best effective combinations of drugs and their combinations shown in Figure 11 there is a significant difference between the medians of the simulation results from three paradigms.

Best effective combinations of drugs		PTKSMA1 (best effective candidate drug)	PTKSMA1 & Dacinostat (best effective combination of two drugs)	PTKSMA1 & Indole & VPA (best effective combination of three drugs)	PTKSMA1 & Dacinostat & AZA & Indole (best effective combination of four drugs)	PTKSMA1 & Dacinostat & Indole & VPA & AZA (best effective combination of five drugs)	PTKSMA1 & Dacinostat & Indole & TSA & VPA & AZA (best effective combination of six drugs)	PTKSMA1 & Dacinostat & Indole & TSA & VPA & AZA & Resviratrol (best effective combination of seven drugs)
Median of SMN concentration	Deterministic	0.45	0.85	2.19	5.01	6.11	6.49	6.64
	Stochastic	0.60	1.16	2.91	4.95	6.44	7.09	7.35
	Fuzzy stochastic	2.10	3.44	7.37	10.71	11.79	12.26	12.48
Friedman test	χ_r^2	1287.6	1287.6	1264.9	913.1	1253.8	1251.8	1253.8
	df	2	2	2	2	2	2	2
	p	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
Dunn-Bonferroni test		< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001

Figure 11. Test statistics collected from Friedman and Dun–Bonferroni tests for the best effective combinations of drugs.

Then, we performed the Dunn–Bonferroni test to compare a pair of related data sets on a single sample to assess whether the data sets have the same distribution. The pairwise comparison of data sets yielded $p < 0.001$. Hence, we concluded that there is a substantial

difference between the distribution of related values in deterministic, stochastic, and fuzzy stochastic simulations. Moreover, statistical analysis reveals that values in the stochastic cases are significantly higher than corresponding values in the deterministic cases and that values in the fuzzy stochastic models are substantially higher compared to related values in the stochastic models. The results of the comparison of the simulation results in line with three paradigms are summarized in Figure 11.

4. Conclusions and Prospects

In conclusion, this study demonstrates the critical importance of selecting the appropriate modeling paradigm to accurately approximate biological systems. By integrating a comparison of simulation results and spatial scale-density analysis of a biological system, the proposed methodology offers a robust framework for selecting the most suitable modeling paradigm, thereby enhancing the reliability and biological relevance of computational models in systems biology. This methodology is versatile and independent of the research methods used for developing a model and simulations, and comparison of simulation results. Although we use Petri net technologies for model development and simulation, the methodology is compatible with other approaches, such as differential equations. Similarly, although we employ descriptive statistics to compare simulation results, other techniques, such as machine learning and cluster analysis, are equally applicable. This methodology is demonstrated through a specific example of SMN synthesis but can be generalized to other gene-drug-disease networks and broader biological systems.

Using deterministic, purely stochastic, and fuzzy stochastic simulations, we evaluated their performance in modeling the synthesis of SMN from *SMN2*, a key factor in the treatment of SMA. Statistical analyses confirmed significant discrepancies between the results obtained from these modeling frameworks, highlighting that the choice of paradigm significantly affects the precision of the biological representation.

Our results indicate that the fuzzy stochastic paradigm provides the most biologically relevant and accurate quantitative results for the drug-disease network under study due to its ability to capture the inherent uncertainties and randomness of a biological system, which is characterized by being at the cellular scale and sparse in density. Specifically, fuzzy stochastic modeling predicted combinations of drug candidates that could lead to a 149.9-fold increase in SMN levels, which itself has promising therapeutic implications.

We recognize that medications can have side effects and that unnecessary use can place additional strain on the metabolic system. When multiple medications are combined, the risk of unexpected side effects can increase. As a next step, in collaboration with pharmacogenetic teams, we propose conducting in vitro analyses of the current findings to assess the practical relevance of the in silico models using established disease model tissues.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/app142210696/s1.

Author Contributions: Conceptualization, R.B.; Methodology, R.B.; Software, G.A.; Validation, G.A.; Formal analysis, R.B.; Resources, G.A.; Writing—original draft, R.B.; Project administration, R.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The original data presented in the study are openly available in PrePrints.org at https://doi.org/10.20944/preprints202410.1483.v1 and the Supporting Information.

Acknowledgments: We are thankful to İlke Nimet Akçay and Mani Mehraei, who have helped with the computer simulations. The authors thank Adil Şeytanoğlu for his assistance in handling the biological context behind the Spinal Muscular Atrophy.

Conflicts of Interest: The data from this study can be obtained upon request from the corresponding author. The data sets, and the clustering algorithm is not publicly accessible at this time because their adoption is a key aspect of the innovation in this paper. Initially, the data sets and the clustering algorithm will not be open-sourced, as further possibilities will be explored in ongoing experimental research. Once this exploration is complete, the data sets and the clustering algorithm, along with the relevant code, will be made publicly available.

References

1. Gillespie, D.T. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.* **1976**, *22*, 403–434. [CrossRef]
2. Herman, M.A.; Aiello, B.R.; DeLong, J.D.; Garcia-Ruiz, H.; González, A.L.; Hwang, W.; McBeth, C.; Stojković, E.A.; Trakselis, M.A.; Yakoby, N. A unifying framework for understanding biological structures and functions across levels of biological organization. *Integr. Comp. Biol.* **2021**, *61*, 2038–2047. [CrossRef] [PubMed]
3. Schamberger, B.; Ziege, R.; Anselme, K.; Ben Amar, M.; Bykowski, M.; Castro, A.P.; Cipitria, A.; Coles, R.A.; Dimova, R.; Eder, M.; et al. Curvature in biological Systems: Its quantification, emergence, and implications across the scales. *Adv. Mater.* **2023**, *35*, 2206110. [CrossRef] [PubMed]
4. Bashirov, R. Prediction of COVID-19 pandemic spreading in North Cyprus. *J. Mod. Technol. Eng.* **2023**, *8*, 73–83.
5. Andrianou, X.D.; Konstantinou, C.; Rodríguez-Flores, M.A.; Papadopoulos, F.; Makris, K.C. Population-wide measures due to the COVID-19 pandemic and exposome changes in the general population of Cyprus in March–May 2020. *BMC Public Health* **2022**, *22*, 2279. [CrossRef] [PubMed]
6. Chams, N.; Chams, S.; Badran, R.; Shams, A.; Araji, A.; Raad, M.; Mukhopadhyay, S.; Stroberg, E.; Duval, E.J.; Barton, L.M.; et al. COVID-19: A multidisciplinary review. *Front. Public Health* **2020**, *8*, 1–20 . [CrossRef]
7. Zadeh, L.A. Fuzzy sets. *Inf. Control* **1965**, *8*, 338–353. [CrossRef]
8. Simoni, G.; Vo H.T.; Priami, C.; Marchetti, L. A comparison of deterministic and stochastic approaches for sensitivity analysis in computational systems biology. *Briefings Bioinform.* **2020**, *21*, 527–540. [CrossRef]
9. Liu, F., Heiner, M., Gilbert, D. Hybrid modelling of biological systems: Current progress and future prospects. *Briefings Bioinform.* **2022**, *23*, bbac081.
10. Akçay, N.; Bashirov, R. Comparison of modeling approaches demonstrated for the p16-mediated signaling pathway in higher eukaryotes. *BioSystems* **2021**, *210*, 104562. [CrossRef]
11. Bashirov, R.; Duranay, R.; Şeytanoğlu, A.; Mehraei, M.; Akcay, N. Exploiting stochastic Petri nets with fuzzy parameters to predict efficient drug components for Spinal Muscular Atrophy. *Turk. J. Electr. Eng. Comput. Sci.* **2019**, *27*, 4009–4022. [CrossRef]
12. David, R.; Alla, H. *Discrete, Continuous, and Hybrid Petri Nets*; Springer: Berlin/Heidelberg, Germany, 2005.
13. Liu, F.; Heiner, M.; Yang, Y. Fuzzy stochastic Petri nets for modeling biological systems with uncertain kinetic parameters. *PLoS ONE* **2016**, *11*, e0149674. [CrossRef] [PubMed]
14. Heiner, M.; Herajy, M.; Liu, F.; Rohr, C.; Schwarick, M. Snoopy—A unifying Petri net tool. In Proceedings of the International Conference on Application and Theory of Petri Nets, Hamburg, Germany, 25–29 June 2012; LNCS; Springer: Berlin/Heidelberg, Germany, 2012; Volume 7347, pp. 398–407.
15. Tüysüz, F.; Kahraman, C. Modeling a flexible manufacturing cell using stochastic Petri nets with fuzzy parameters. *Expert Syst. Appl.* **2009**, *37*, 3910–3920. [CrossRef]
16. Mehraei, M.; Bashirov, E. Petri net-based simulation of effective targets for β-globin disorders. *J. Mod. Technol. Eng.* **2019** *4*, 139–152.
17. Blair, H.A. Onasemnogene abeparvovec: A review in Spinal Muscular Atrophy. *CNS Drugs* **2022**, *36*, 995–1005. [CrossRef] [PubMed]
18. Ogbonmide, T.; Rathore, R.; Rangrej, S.B.; Hutchinson, S.; Lewis, M.; Ojilere, S.; Carvalho, V.; Kelly, I. Gene therapy for spinal muscular atrophy (SMA): A review of current challenges and safety considerations for onasemnogene abeparvovec (Zolgensma). *Cureus* **2023**, *15*, e36197. [CrossRef]
19. Kakazu, J.; Walker, N.L.; Babin, K.C.; Trettin, K.A.; Lee, C.; Sutker, P.B.; Kaye, A.M.; Kaye, A.D. Risdiplam for the use of Spinal Muscular Atrophy. *Orthop. Rev.* **2021**, *13*, 25579. [CrossRef]
20. Cornell, N.; Childs, A.M.; Wraige, E.; Munot, P.; Ambegaonkar, G.; Chow, G.; Hughes, I.; Illingworth, M.; Majumdar, A.; Marini-Bettolo, C.; et al. Risdiplam in Spinal Muscular Atrophy: Safety Profile and Use Through The Early Access to Medicine Scheme for the Paediatric Cohort in Great Britain. *J. Neuromuscul. Dis.* **2024**, *11*, 361–-368. [CrossRef]
21. Brichta, L.; Hofmann, Y.; Hahnen, E.; Siebzehnrubl, F.A.; Raschke, H.; Blumcke, I.; Eyupoglu, I.Y.; Wirth, B. Valproic acid increases the SMN2 protein level: A well-known drug as a potential therapy for Spinal muscular atrophy. *Hum. Mol. Genet.* **2003**, *12*, 2481–2489. [CrossRef]
22. Avila, A.M.; Burnett, B.G.; Taye, A.A.; Gabanella, F.; Knight, M.A.; Hartenstein, P.; Cizman, Z.; Di Prospero, N.A.; Pellizzoni, L.; Fischbeck, K.H.; et al. Trichostatin A increases SMN expression and survival in a mouse model of Spinal muscular atrophy. *J. Clin. Investig.* **2007**, *117*, 659–670. [CrossRef]

23. Dayangac-Erden, D.; Bora, G.; Ayhan, P.; Kocaefe, C.; Dalkara, S.; Yelekci, K.; Demir, A.S.; Erdem-Yurter, H. Histone deacetylase inhibition activity and molecular docking of (E)-Resviratrol: Its therapeutic potl in spin DTaluscular atrophy. *Chem. Biol. Drug Des.* **2009**, *73*, 355–364. [CrossRef] [PubMed]
24. Hastings, M.L.; Berniac, J.; Liu, Y.H.; Abato, P.; Jodelka, F.M.; Barthel, L.; Kumar, S.; Dudley, C.; Nelson, M.; Larson, K.; et al. Tetracyclines that promote SMN2 exon 7 splicing as therapeutics for Spinal Muscular Atrophy. *Sci. Transl. Med.* **2009**, *1*, 5ra12. [CrossRef] [PubMed]
25. Jarecki, J.; Chen, X.; Bernardino, A.; Coovert, D.D.; Whitney, M.; Burghes, A.; Stack, J.; Pollok, B.A. Diverse small-molecule modulators of SMN expression found by high-throughput compound screening: Early leads towards a therapeutic for Spinal Muscular Atrophy. *Hum. Mol. Genet.* **2005**, *14*, 2003–2018. [CrossRef] [PubMed]
26. Hauke, J.; Riessl, M.; Lunke, S.; Eyüpoglu, I.Y.; Blümcke, I.; El-Osta, A.; Wirth, B.; Hahnen, E. Survival motor neuron gene 2 silencing by DNA methylation correlates with Spinal muscular atrophy disease severity and can be bypassed by histone deacetylase inhibition. *Hum. Mol. Genet.* **2009**, *18*, 304–317. [CrossRef] [PubMed]
27. Ackermann, J.; Koch, I. Quantitative analysis. In *Modeling in Systems Biology—Petri Net Approach*; Koch, I., Reisig, W., Schreiber, F., Eds.; Springer: London, UK, 2011; pp. 153–179.