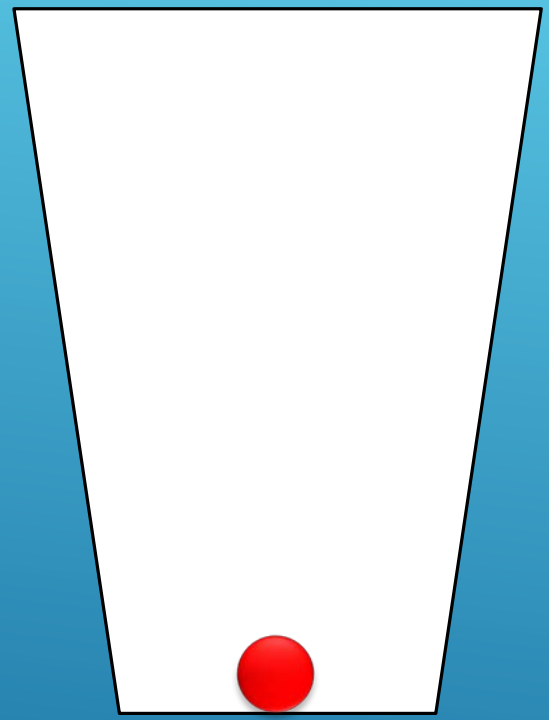
Statistical comparison of modelling approaches demonstrated for biomedical networks

Rza BASHIROV

Eastern Mediterranean University

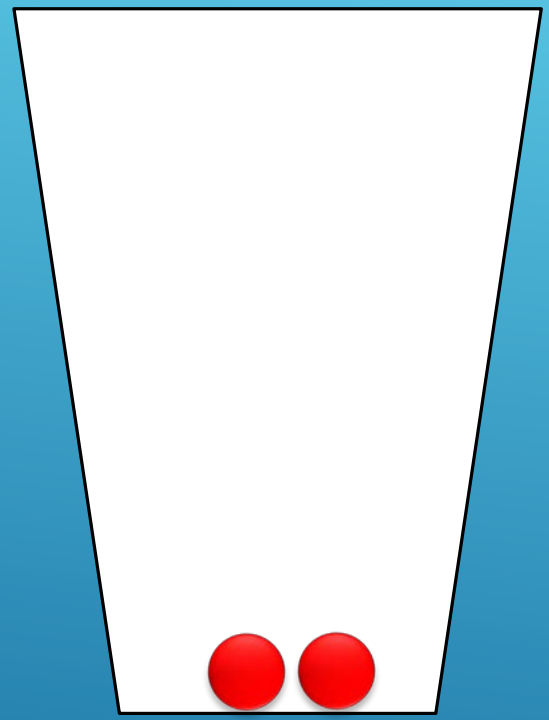
DETERMINISTIC vs STOCHASTIC MODELLING

Dynamic behavior of a single molecule is completely stochastic.



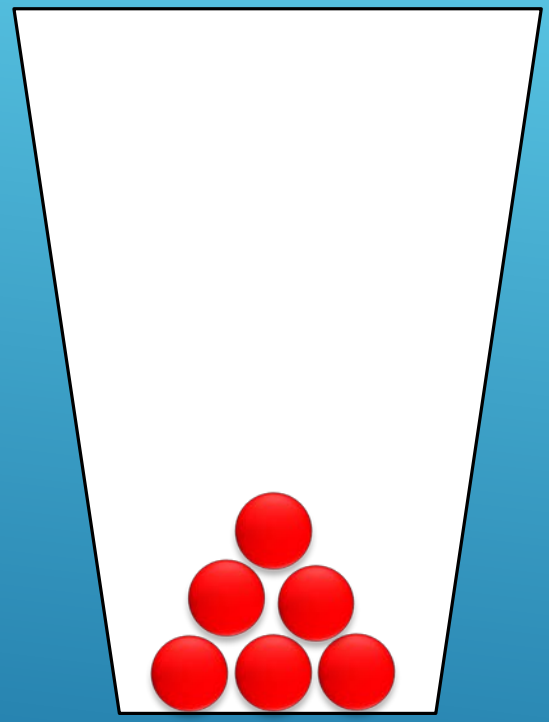
DETERMINISTIC vs STOCHASTIC MODELLING

A system with two molecules is less stochastic since the molecules balance each another.



DETERMINISTIC vs STOCHASTIC MODELLING

The level of stochasticity in a molecular system is inversely proportional to its molecular density so that the system becomes less stochastic and more deterministic with increase of molecular density.



DETERMINISTIC vs STOCHASTIC MODELLING

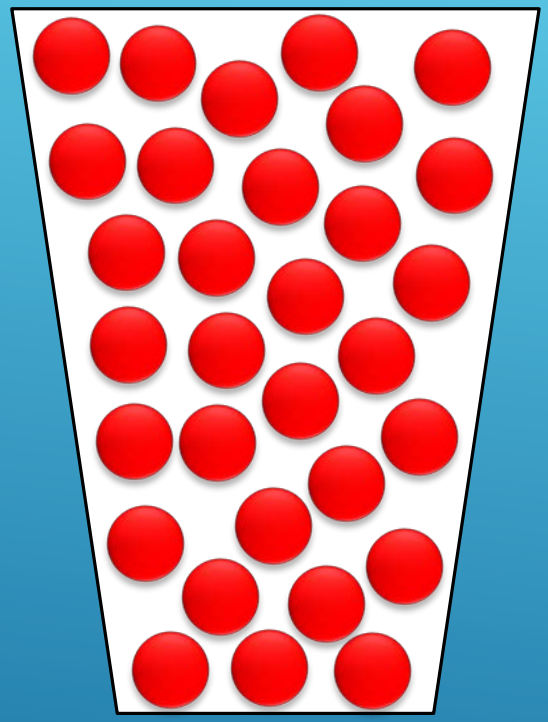
In any molecular system, the level of stochasticity can be expressed as a function

$$\frac{1}{\sqrt{n}}$$

of the number of molecules n. The stochastic formulation reduces to the deterministic one in the thermodynamic limit,

$$\lim_{n\to\infty} \frac{1}{\sqrt{n}} = 0$$

i.e. when number of molecules, n, approaches infinity. The overall behavior of a dense molecular system follows a deterministic pattern while a sparse molecular system remains mainly stochastic.



MODELLING OF BIOLOGICAL SYSTEMS

Analysis of biological systems over the last several decades has mainly been performed by continuous deterministic models. In a deterministic model, the initial condition specifies the outcome so that, no matter how many times the model runs, the outcome is always the same.

Harvard Gazette | Harvard University

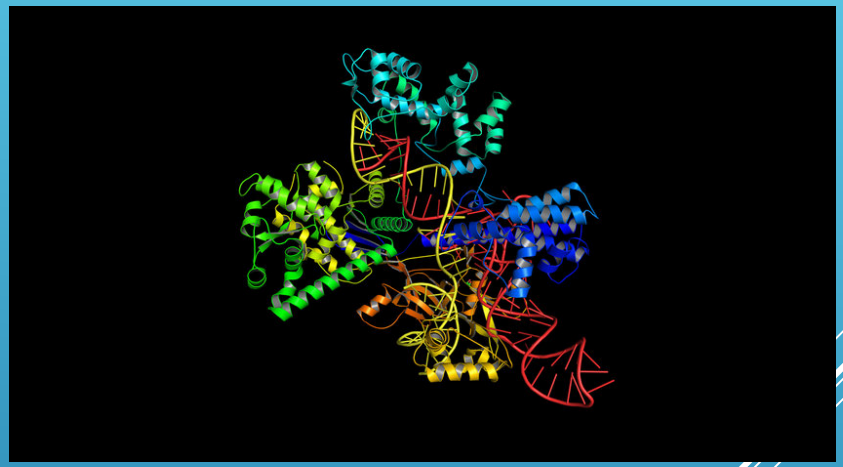
MODELLING OF BIOLOGICAL SYSTEMS

In many cases random fluctuations do not significantly change the behavior of a biological system and the assumption remains correct. There are, however, examples for which this assumption does not remain correct. It has also been shown that stochastic effects are crucial in gene expression, signal transduction and cell replication.

MOPIC | SHUTTERSTOCK

MODELLING OF BIOLOGICAL SYSTEMS

Deterministic models turn out not to be sufficient to adequately describe noise, variability and randomness inherent to biological systems. Stochastic models are being used increasingly in preference to deterministic ones especially when the biological system exhibits unstable behavior.



Genome Editing | Max-Planck-Gesellschaft

MODELLING OF BIOLOGICAL SYSTEMS

Randomness in biological networks may arise due to low molecular density, intrinsic random nature of phenomena, and noise in an experiment. An individual molecular event is subject to stochastic time delays as it takes place whenever the event conditions (availability of substrates, desired level of energy, temperature and pressure, etc.) are present, but not according to a predefined order.

Ancient human DNA taken from African mountain cave | Nature World Report

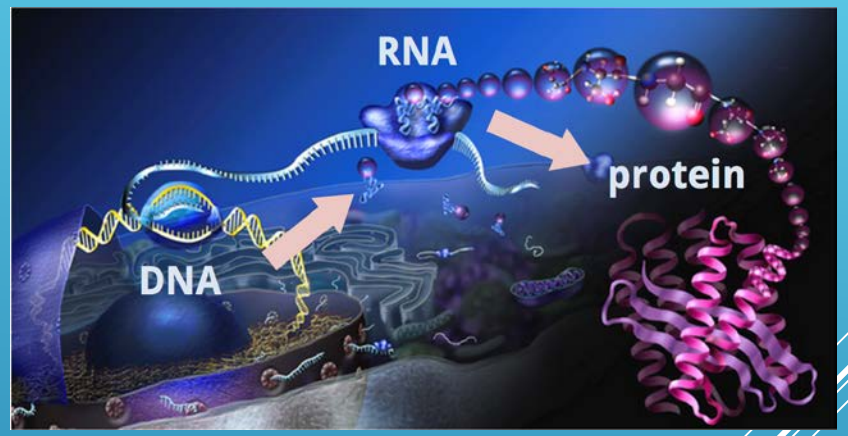
VAGUENESS vs CRISPNESS

It is quite regular that, identical wet lab experiments result in different observations at each time due to the inexactness of measurements and other technical noise. For instance, genetically identical cells even within the same tissue often exhibit different levels of gene expression, protein production, and different rates for biological phenomena.

Laboratory of Molecular Biology and Genetics | EMU

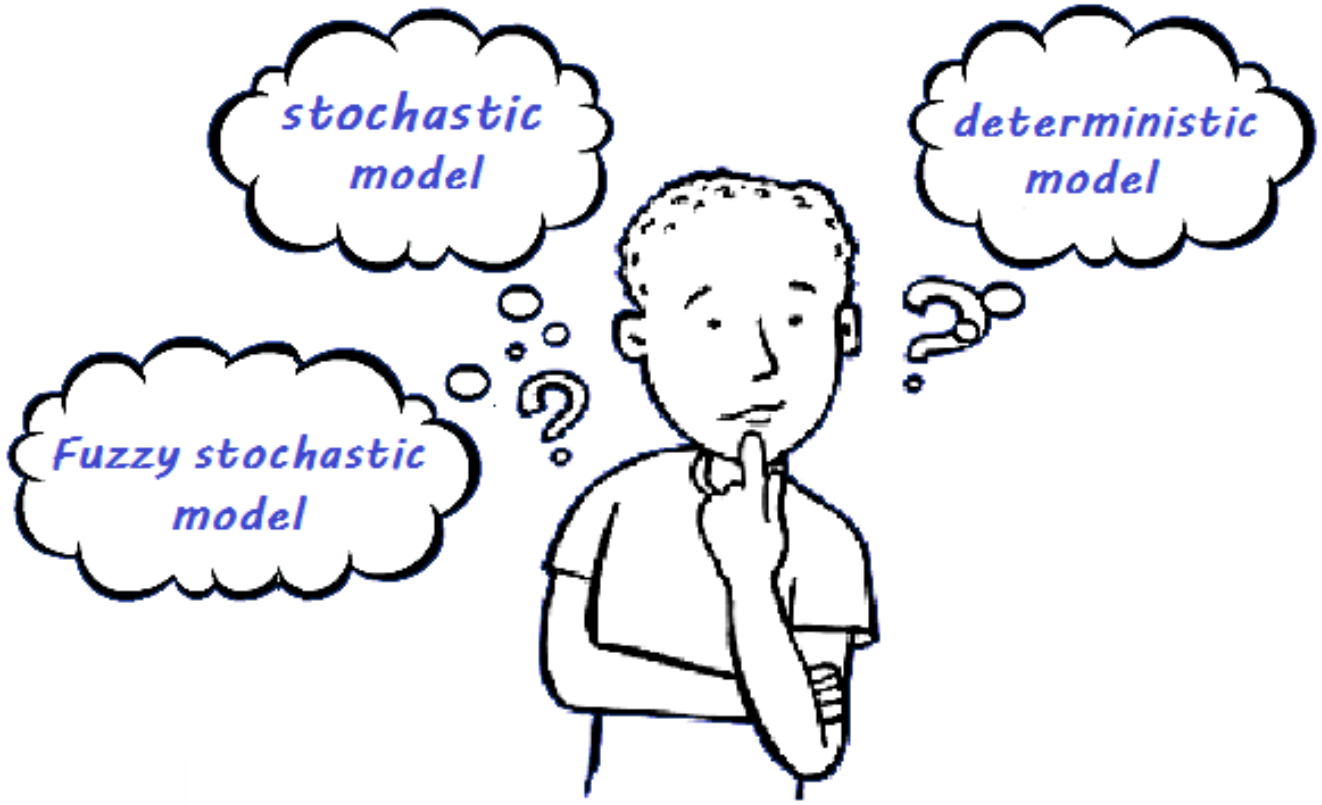
VAGUENESS vs CRISPNESS

Kinetic parameters are usually uncertain and represented by natural language-based qualitative knowledge. We deal with imprecise and incomplete knowledge about reaction rates which are often expressed by qualitative descriptions of parameters such as “is almost disrupted” or “occurs faster than”. Fuzzy logic allows modelling reaction effects which can be derived from qualitative knowledge.



Central dogma of molecular biology | Artlab

Which one is Better?



PREDICTING EFFICIENT DRUG COMBINATIONS FOR SPINAL MUSCULAR ATROPHY

SMA is the leading genetic cause of infant mortality and the second most common fatal autosomal recessive disorder after cystic fibrosis. The disease affects 1 in 6000–10,000 newborns. The disease is caused by the deletion of or mutations in the SMN1 gene.

Newspaper: Medical News Today

PREDICTING EFFICIENT DRUG COMBINATIONS FOR SPINAL MUSCULAR ATROPHY

The existing approaches to treatment of SMA can be classified into four classes:

- Promoting SMN2 transcription through use of histone deacetylase (HDAC) inhibitors,
- Increasing correct splicing of the SMN2 transcript,
- Upregulating promoter activity of SMN2,
- Increasing SMN2 activity through deoxyribonucleic acid (DNA) demethylation.

Newspaper: Medical News Today

PREDICTING EFFICIENT DRUG COMBINATIONS FOR SPINAL MUSCULAR ATROPHY

The chemicals or drug candidates include:

- ValProic Acid (VPA), TrichoStatin A (TSA), Dacinostat, and Resveratrol as HDAC inhibitors
- PTMK-SMA1 for alternative splicing
- Indole for upregulating the promoter activity
- AZA for DNA demethylation

Newspaper: Medical News Today

PREDICTING EFFICIENT DRUG COMBINATIONS FOR SPINAL MUSCULAR ATROPHY

These drug candidates provide 1.3- to 5-fold increase of SMN, which is not sufficient to cure SMA. We explore the intermolecular interactions in SMN production network to predict the most efficient drug combinations which can result in maximum SMN levels from SMN2.The present work combines all these approaches to account for the total efficacy of various combinations of potential drugs with existing qPCR data.

Newspaper: Medical News Today

PREDICTING EFFICIENT DRUG COMBINATIONS FOR SPINAL MUSCULAR ATROPHY

Methods and skills:

- Stochastic Petri nets with fuzzy parameters
- Snoopy software tool
- SPSS Statistics Software Package
- Maple

Newspaper: Medical News Today

PREDICTING EFFICIENT DRUG COMBINATIONS FOR SPINAL MUSCULAR ATROPHY

Model parameters:

- 18 components, 25 processes and 41 interprocess interactions
- validation for each drug candidate
- triangular pattern for a fuzzy number
- stochastic firing rate $h(t,\theta)\colon \mathbb{N}_0^{|\circ t|} \longrightarrow \mathbb{R}^+$ of Gillespie propensity type

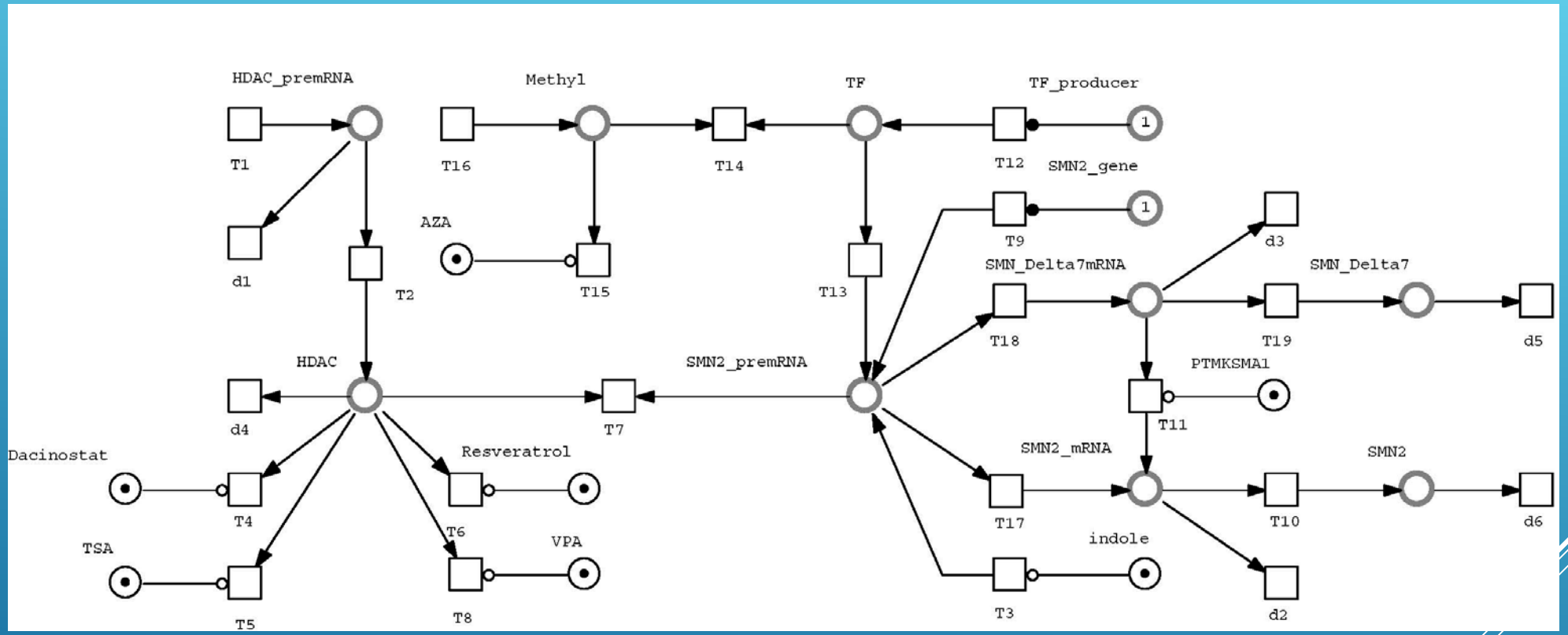
  $$h(t,\theta) = \theta \cdot \#\{reactant\ combinations\}$$

  which is related to random variable Y_t defined by the following exponential probability distribution

  $$F_{Y_t}(\tau) = 1 - e^{-h(\theta,t)\cdot\tau}, \tau > 0$$

Hitching a ride: Target-based drug discovery | Oxford Science Blog

PREDICTING EFFICIENT DRUG COMBINATIONS FOR SPINAL MUSCULAR ATROPHY



The complete model of SMN production network validated for each of the seven drug candidates that inhibit HDAC (TSA, VPA, Dacinostat, and Resviratrol), modulate pre-mRNA splicing (PTK-SMA1), upregulate promoter activity (Indole), and target DNA methylation (AZA). Read arcs and inhibitor arcs are respectively represented by a black dot and hollow dot as arc head. Read arcs are used to ensure continuous expression of SMN2 gene and production of transcription factor, while inhibitory arcs are used to simulate enabling/disabling a drug treatment.

PREDICTING EFFICIENT DRUG COMBINATIONS FOR SPINAL MUSCULAR ATROPHY

Computer simulations:

- 38,000 stochastic replications for each combination of drug candidates, which are further averaged to obtain a reliable estimation of simulation based behavior of SMN production network, allowing to reach a confidence level of 95% with the accuracy of 0,01
- Measurements taken until 1000 pt time for all 3 modelling approaches and for each of 120 drug candidates and their combinations

Dealing with Big Data | Akamai

PREDICTING EFFICIENT DRUG COMBINATIONS FOR SPINAL MUSCULAR ATROPHY

Effective 2- and 3-combinations	Model		
	Deterministic	Stochastic	Fuzzy stochastic
PTKSMA1&Dacinostat	12.716	13.278	(10.800; 18.200)
PTKSMA1&VPA	13.442	14.263	(13.100; 17.500)
PTKSMA1&Indole	15.030	15.367	(14.600; 17.000)
PTKSMA1&AZA	9.926	11.007	(11.700; 12.100)
PTKSMA1&Indole&VPA	35.101	36.522	(30.500; 46.000)
PTKSMA1&Indole&AZA	39.523	39.111	(24.000; 41.600)
PTKSMA1&Indole&TSA	23.936	25.822	(16.100; 32.800)
PTKSMA1&Dacinostat&AZA	23.977	26.689	(23.500; 32.000)
PTKSMA1&Dacinostat&VPA	20.621	23.971	(23.300; 32.400)
PTKSMA1&AZA&VPA	25.217	28.690	(27.300; 31.900)

PREDICTING EFFICIENT DRUG COMBINATIONS FOR SPINAL MUSCULAR ATROPHY

Effective 4- and 5-combinations	Model		
	Deterministic	Stochastic	Fuzzy stochastic
PTKSMA1&Dacinostat&AZA&Indole	84.907	64.907	(40.600,88.900)
PTKSMA1&VPA&Indole&AZA	86.877	67.669	(45.800,87.400)
PTKSMA1&Dacinostat&TSA&Indole	40.452	43.963	(27.200,63.800)
PTKSMA1&Dacinostat&Resviratrol&Indole	36.586	39.217	(24.100,55.100)
PTKSMA1&VPA&Resviratrol&Indole	38.095	41.023	(26.400,53.600)
PTKSMA1&Dacinostat&TSA&AZA	29.515	35.135	(22.800,45.700)
PTKSMA1&VPA&AZA&TSA	30.682	36.712	(25.900,44.800)
PTKSMA1&Dacinostat&Indole&VPA&AZA	104.399	86.020	(57.500,125.600)
PTKSMA1&Dacinostat&Indole&AZA&TSA	93.782	75.767	(48.100,110.700)
PTKSMA1&Indole&VPA&AZA&TSA	95.532	77.820	(52.200,108.800)
PTKSMA1&Dacinostat&Indole&Resviratrol&AZA	39.016	39.016	(44.800,99.500)
PTKSMA1&Dacinostat&Indole&VPA&Resviratrol	52.068	56.240	(35.600,82.800)
PTKSMA1&Indole&AZA&Resviratrol&TSA	77.251	60.768	(37.800,79.100)
PTKSMA1&Dacinostat&Indole&TSA&Resviratrol	43.384	48.289	(31.000,70.800)
PTKSMA1&Indole+VPA&Resviratrol&TSA	44.773	49.285	(31.700,69.600)
PTKSMA1&Dacinostat&VPA&TSA&AZA	41.874	50.427	(38.100,68.400)
PTKSMA1&Dacinostat&VPA&Resviratrol&AZA	39.150	46.747	(32.500,62.000)

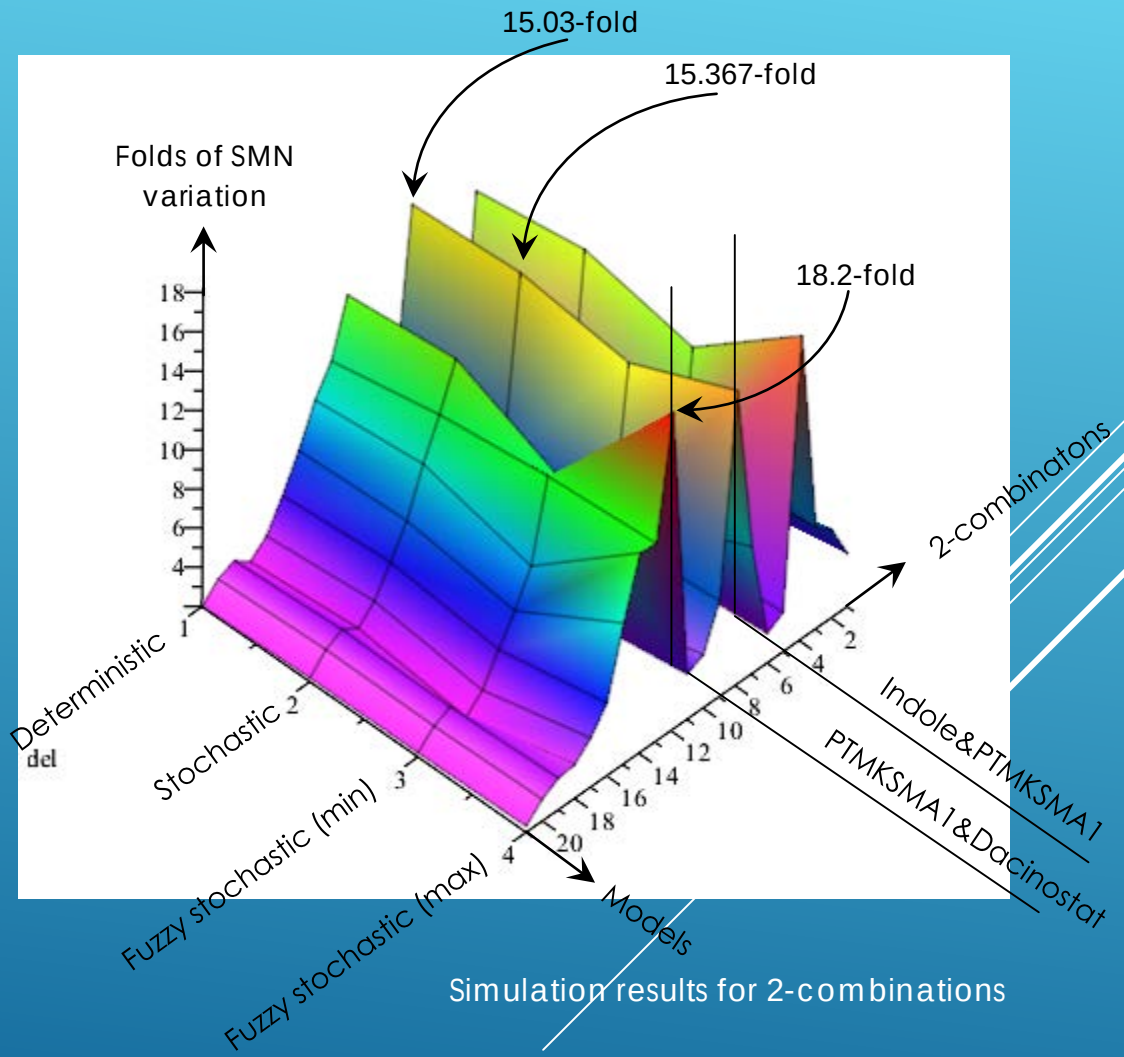
PREDICTING EFFICIENT DRUG COMBINATIONS FOR SPINAL MUSCULAR ATROPHY

Effective 6- and 7-combinations	Model		
	Deterministic	Stochastic	Fuzzy stochastic
PTKSMA1&Dacinostat&Indole&TSA&VPA&AZA	111.086	94.647	(62.900,143.800)
PTKSMA1&Dacinostat&Indole&Resviratrol&VPA&AZA	107.707	90.298	(60.000,133.900)
PTKSMA1&Dacinostat&Indole&Resviratrol&TSA&AZA	97.150	80.664	(51.100,119.400)
PTKSMA1&Indole&Resviratrol&VPA&TSA&AZA	98.842	82.862	(55.500,117.300)
PTKSMA1&Dacinostat&Indole&Resviratrol&VPA&TSA	57.601	63.841	(40.500,97.300)
PTKSMA1&Dacinostat&Resviratrol&VPA&TSA&AZA	43.876	53.701	(37.100,74.400)
7-combination	113.799	98.142	(65.400,149.900)

PREDICTING EFFICIENT DRUG COMBINATIONS FOR SPINAL MUSCULAR ATROPHY

Simulation results:

- Indole&PTK-SMA1 is the most efficent 2-combination in deterministic and stochastic models, respectively resulting in 15.03- and 15.367-fold increase of SMN levels.
- PTK-SMA1&Dacinostat is the most efficient in fuzzy stochastic case, leading to 18.2-fold increase of SMN concentration

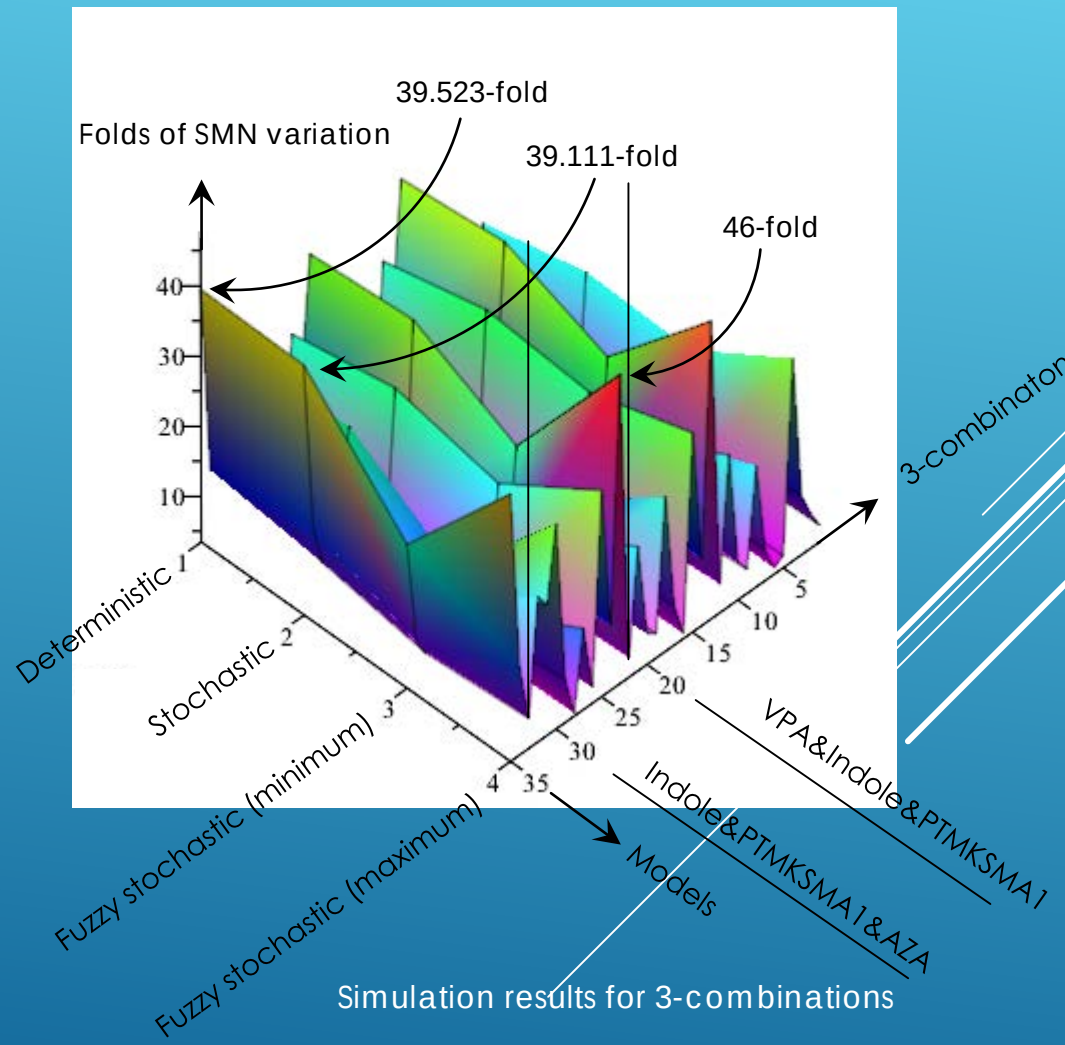


Simulation results for 2-combinations

PREDICTING EFFICIENT DRUG COMBINATIONS FOR SPINAL MUSCULAR ATROPHY

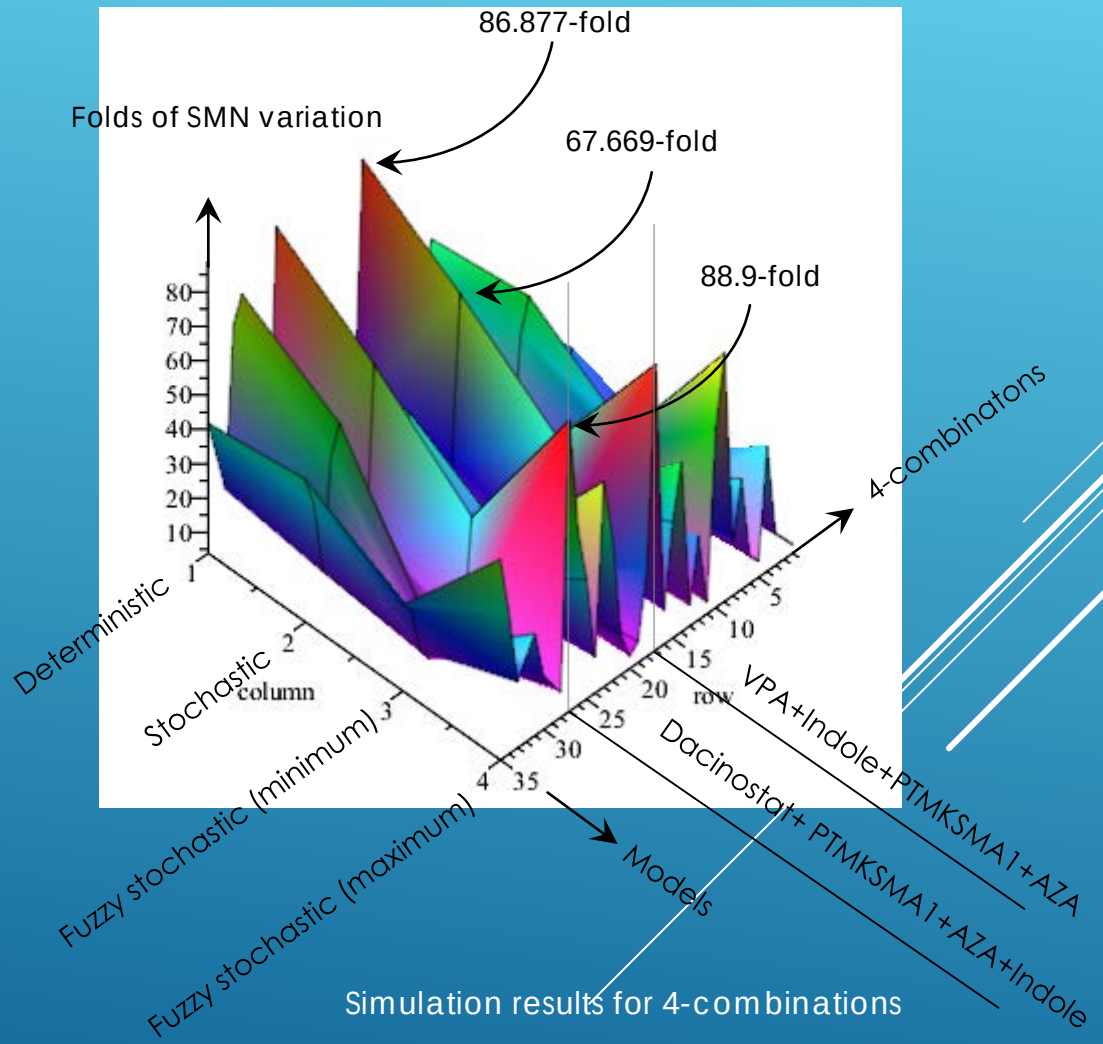
Simulation results:

- Deterministic and stochastic models agree that Indole&PTK-SMA1&AZA is the most efficent 3-combination, respectively resulting in 39,523- and 39.111-fold increase of SMN levels
- VPA&Indole&PTK-SMA1 with 46-fold over the control group is the most efficent in fuzzy stochastic case



PREDICTING EFFICIENT DRUG COMBINATIONS FOR SPINAL MUSCULAR ATROPHY

Simulation results:
For all n-combinations, both deterministic and stochastic simulations result in the same most effective drug combination while fuzzy stochastic case in general demonstrates different behavior.



STATISTICAL COMPARISON OF MODELLING ENVIRONMENTS

- We performed statistical analysis on SPSS Statistics Software Package to measure more accurately how much deterministic, stochastic, and fuzzy stochastic models agree or differ
- We conducted normality tests for all data sets and observed that neither of these data sets is normally distributed
- We applied a nonparametric statistical test to pairwise compare corresponding data sets picked up from distinct approaches

STATISTICAL COMPARISON OF MODELLING ENVIRONMENTS

Appropriate tests are based on the following hypotheses:

$$H_0: \text{Median}(x) = \text{Median}(y)$$
$$H_1: \text{Median}(x) \neq \text{Median}(y)$$

where $x \neq y$ and x and y stand for variables created for data sets obtained in deterministic, stochastic, and fuzzy stochastic approaches.

STATISTICAL COMPARISON OF MODELLING ENVIRONMENTS

Friedman test is conducted to compare the data sets. For all drug combinations, Friedman test resulted in the rejection of the null hypotheses, H_0 , with a $P < 0.001$, which indicates that there is an essential difference between the medians of data sets determined by the deterministic, stochastic, and fuzzy stochastic models.

STATISTICAL COMPARISON OF MODELLING ENVIRONMENTS

Then we performed a paired difference test called the Wilcoxon Signed Rank Test which compares two related data sets on a single sample to assess whether data sets have the same distribution. This test either yielded a P-value of $< 0{,}001$, which leads us to reject the hypothesis H_0 .

STATISTICAL COMPARISON OF MODELLING ENVIRONMENTS

Statistical analysis reveals that there is substential difference between distribution of related data sets in deterministic, stochastic, and fuzzy stochastic models which, when coupled with the fact that fuzzy stochastic model provides the closest approximation of SMN protein production network, successfully coping not only with randomness but also uncertainty, suggests that fuzzy stochastic model is the most appropriate choice for the present case study.

DISCUSSION

A question of practical interest may be the issue of determining precise fuzzy numbers to be assigned to kinetic parameters with uncertain or unknown experimental values. Several authors adopted the following scheme for fuzzy parameter estimation: A fuzzy number is initially represented as a union of its α-cuts. The α-cut for each output is obtained by decomposing all fuzzy parameters into their α-cuts and then running stochastic simulations at each level. Following this step, the membership function for each output is obtained by composing all the α-cuts. Unfortunately, there are some complications preventing the applicability of this approach in the present work. Firstly, this approach increases the number of simulation runs by the number of levels. In the present study, we perform 127 simulation runs for the drugs and their combinations. Application of the above scheme for even 10 levels would require 1270 simulation runs and any further decrease in the step size would result in substantial increase in the number of the simulation runs. Next, the approach suggests the step size of the levels be determined carefully according to the nature of the problem, it is not quite clear, however, how to determine the step size based on the nature of the current case study.

FURTHER WORK

- We are aware that medications may have side effects and that, if a medication possesses side effects, its release, when not required, poses an extra burden on the metabolic system. A combination of multiple medications may even complicate the situation in the sense that it may cause unexpected side effects. As further work, in collaboration with pharmacogenetics groups, we propose the in vitro analysis of the current results to determine the practical applicability of the in silico models in established disease model tissues.
- The present research proposes a methodology which can be adapted to other biological, biomedical and drug-disease pathways. Our bioinformatics research group is presently working on adaptation of this approach to (i) fetal-to-adult hemoglobin switch network, which is important for identifying molecular targets of β-globin disorders; and (ii) p16-mediated pathway, which plays substential role in investigation of human cancer.

REFERENCES

1. Ross SM. Introduction to Probability Models. San Diego, CA, USA: Academic Press, 2014.
2. Meng TC, Somani S, Dhar P. Modeling and simulation of biological systems with stochasticity. In Silico Biology 2004; 4(3): 293-309.
3. Gibson MA, Bruck J. Efficient exact stochastic simulation of chemical systems with many species and many channels. Journal of Physical Chemistry A 2000; 104: 1876-1889. doi: 10.1021/jp993732q
4. Phillips A, Cardelli L. Efficient, correct simulation of biological processes in the stochastic Pi-calculus. In: 5th International Conference on Computational Methods in Systems Biology; Edinburg, UK; 2007. pp. 184-199.
5. Gillespie DT. Exact stochastic simulation of coupled chemical reactions. Journal of Physical Chemistry 1977; 81: 2340-2361. doi: 10.1021/j100540a008
6. Yang X, Han R, Guo Y, Bradley J, Cox B et al. Modelling and performance analysis of clinical pathways using the stochastic process algebra PEPA. BMC Bioinformatics 2011; 13: 1-17. doi: 10.1186/1471-2105-13-s14-s4
7. Heiner M, Gilbert D, Donaldson R. Petri nets for systems and synthetic biology. In: 8th International Conference on Formal Methods for Computational Systems Biology; Bertinogo, Italy; 2008. pp. 215-264.
8. Goss P, Peccoud J. Quantitative modeling of stochastic systems in molecular biology by using stochastic Petri nets. Proceedings of the National Academy of the Sciences of the United Stated of America 1998; 95: 6750-6755. doi: 10.1073/pnas.95.12.6750
9. Srivastava R, Peterson MS, Bentley WE. Stochastic kinetic analysis of the Escherichia coli stress circuit using σ targeted antisense. Biotechnology and Bioengineering 2001; 75(1): 120-129. doi: 10.1002/bit.1171
10. Bahi-Jaber N, Pontier D. Modeling transmission of directly transmitted infectious diseases using colored stochastic Petri nets. Mathematical Biosciences 2003; 185(1): 1-13. doi: 10.1016/S0025-5564(03)00088-9

REFERENCES

11. Marwan W, Sujatha A, Starostzik C. Reconstructing the regulatory network controlling commitment and sporulation in physarum polycephalum based on hierarchical Petri net modelling and simulation. Journal of Theoretical Biology 2005; 236(4): 349-365. doi: 10.1016/j.jtbi.2005.03.018
12. Mura I, Csiksz-Nagy A. Stochastic Petri net extension of a yeast cell cycle model. Journal of Theoretical Biology 2008; 254(4): 850-860. doi: 10.1016/j.jtbi.2008.07.019
13. Lamprecht R, Smith GD, Kemper P. Stochastic Petri net models of Ca2+ signaling complexes and their analysis. Natural Computing 2011; 10: 1045-1075. doi: 10.1007/s11047-009-9143-y
14. Marwan W, Rohr C, Heiner M. Petri nets in Snoopy: a unifying framework for the graphical display, computational modelling, and simulation of bacterial regulatory networks. In: Helden J, Toussaint A, Thieffry D (editors). Methods in Molecular Biology. Vol. 804. Berlin Heidelberg, Germany: Humana Press, 2012, pp. 409-437.
15. Castaldi D, Maccagnola D, Mari D, Archetti F. Stochastic simulation of the coagulation cascade: a Petri net based approach. In: Euro-Par 2012 Workshops; Rhodes, Greece; 2013. pp. 248-262.
16. Liu F, Heiner M, Yang M. Fuzzy stochastic Petri nets for modeling biological systems with uncertain kinetic parameters. PLoS ONE 2016; 11(2). doi: 10.1371/journal.pone.0149674
17. Bashirov R, Akçay Nİ. Stochastic simulation-based prediction of the behavior of the p16-meadiated signaling pathway. Fundamenta Informaticae 2018; 160: 167-179. doi: 10.3233/FI-2018-1679
18. Heiner M, Herajy M, Liu F, Rohr C. Snoopy a unifying Petri net tool. In: 33rd International Conference on Application and Theory of Petri Nets; Hamburg, Germany; 2012. pp. 398-407.
19. Clark G, Courtney T, Daly D, Deavours D, Derisavi S et al. The Möbius modeling tool. In: International Workshop on Petri Nets and Performance Models (PNPM'01); Aachen, Germany; 2001. pp. 241-250.
20. Chiola G. A software package for analysis of generalized stochastic Petri net models. In: International Workshop on Timed Petri Nets; Torino, Italy; 1985. pp. 136-143.
21. Sokhansanj B, Fitch J, Quong J, Quong A. Linear fuzzy gene network models obtained from microarray data by exhaustive search. BMC Bioinformatics 2004; 5(108). doi: 10.1186/1471-2105-5-108

REFERENCES

22. Gintrowski A. Modeling gene networks using fuzzy logic. In: 6th Doctoral Workshop on Mathematical and Engineering Methods in Computer Science; Mikulov, Czech Republic; 2010. pp. 32-39.
23. Hamed R. Quantitative modeling of gene networks of biological systems using fuzzy Petri nets and fuzzy sets. Journal of King Saud University - Science 2018; 30: 112-119. doi: 10.1016/j.jksus.2017.01.005
24. Mehraei M. Identifying emotion regulation altering targets as depressive mood disorder treatments using fuzzy stochastic hybrid Petri nets. IAFOR Journal of Psychology and the Behavioral Sciences 2018; 4(1). doi: 10.22492/ijpbs.4.1.04
25. Bordon J, Moskon M, Zimic N, Mraz M. Semi-quantitative modelling of gene regulatory processes with unknown parameter values using fuzzy logic and Petri nets. Fundamenta Informaticae 2018; 160: 81-100. doi: 10.3233/FI-2016-0000
26. Liu F, Chen S. Colored fuzzy Petri nets for dealing with genetic regulatory networks. Fundamenta Informaticae 2018; 160: 101-118. doi: 10.3233/FI-2018-1676
27. Zadeh LA. Fuzzy Sets. Information and Control 1965; 8: 338-353
28. Avila AM, Burnett B, Taye AA, Gabanella F, Knight MA et al. Trichostatin A increases SMN expression and survival in a mouse model of spinal muscular atrophy. Journal of Clinical Investment 2007; 117(3): 659-670. doi: 10.1172/JCI29562
29. Dayangaç-Erden D, Bora G, Ayhan P, Kocaefe Ç, Dalkara S et al. Histone deacetylase inhibition activity and molecular docking of (E)-Resviratrol: its therapeutic potential in spinal muscular atrophy. Chemical Biology &Drug Design 2009; 73(3): 355-364. doi: 10.1111/j.1747-0285.2009.00781.x
30. Hastings ML, Berniac J, Liu YH, Abato P, Jodelka FM et al. Tetracyclines that promote SMN2 exon 7 splicingas therapeutics for Spinal Muscular Atrophy. Science Translational Medicine 2010; 1(5): 5ra12. doi: 10.1126/scitranslmed. 3000208