Automated replication optimization for protocellular information system

Ditlev Hartmann Bornebusch[1,3], Christina Colaluca Sørensen[3], Peter Zingg[2], Gianluca Gazzola[2], Norman Packard[2], and Steen Rasmussen[3,4]

[1]Sino-Danish Center, University of Chinese Academy of Sciences, Beijing, China; [2]Daptics Inc., San Francisco, USA; [3]FLinT Center, University of Southern Denmark, 5230 Denmark; [4]Santa Fe Institute, New Mexico 87501, USA
deltaetabeta@gmail.com[1] & steen@sdu.dk[3]

Introduction

Due to the high cost of experiments, high dimensional complex systems with multiple parameters usually pose grand challenges not only in artificial life but in areas including manufacturing processes, supply chains as well as services in the healthcare sector. We present and verify a fully automated method of reducing the needed experiments to identify optimal operational conditions for complex systems, here tested in simulation on a protocellular information system. The method iteratively becomes better at locating system optima through an adaptive data analysis, which is an advantage over e.g. a Monte Carlo optimization method (2).

Autonomous Exploration and Optimization

We have developed and tested an autonomous system for exploration and optimization of complex systems: an Autonomous Exploration and Optimization Loop (AEOL). It is composed of three principal parts that are connected in a loop: (i) a simulation of the complex system under investigation, (ii) an Artificial Intelligence based Design of Experiment (AI-DoE) algorithm and (iii) a message handling system that sends output from the simulation to the AI-DoE systems that in turn sends new input parameters to the simulation system. The loop can be iterated a predefined number of times or until a desired result is obtained.

Loop component (i) is a Lesion Induced DNA Amplification (LIDA) process (see (1) and (3)), which could function as an informational building block with sufficient replication yield for the protocellular model developed in (6). It is a reaction kinetic equation system of the form $\frac{dx}{dt} = f(x, \alpha)$, where x and α denote the involved physicochemical species concentrations and reaction constants respectively. A DNA template and four shorter complementary DNA oligomers are replicated through template directed ligation. Different sequences and corresponding oligomers yield different hybridization energies and thus different reaction rates. Results from LIDA simulations and lab experiments are shown in Fig. 1.

Loop component (ii) is the AI-DoE prediction algorithm that is based on a neural network that leverages experimental data to predict better experiments (5). The AI-DoE traverses the loop by building predictive models from the inputs of the simulation experiments to their outputs $\alpha_i \rightarrow x_i$, and proposes new experiments α_{i+1} that are most likely to

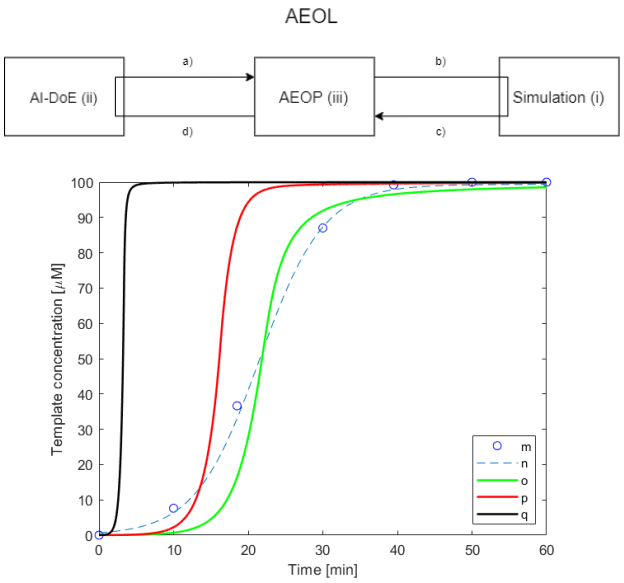

Figure 1: **Top** Four routines (a-d) are executed for every generation in the AEOL: (a) After initial definition of parameter space and regimes a set of proposed experiments (parameter combinations) is collected by AEOP once generated by AI-DoE. (b) AEOP initiates language-specific system simulations of proposed experiments. (c) Simulation results from proposed parameter combinations are collected by AEOP. (d) Corresponding parameter combinations and simulation results are packaged by AEOP and sent to AI-DoE in a single API call. **Bottom** Experimental LIDA data (m), see (1); (n) curve fitted to experimental data; (o) simulation with same parameters as experiment; (p) and (q) are AI-DoE discovered optimal replication yield curves with $k_L \leq 10^{-2}$/s and $k_L \leq 10^{-1}$/s respectively. However, $k_L = 10^{-1}$/s is only of theoretical interest as it is not realizable (too fast).

Parameter	k^+_{O1}, k^+_{O2}	k^-_{O1}, k^-_{O2}	k^+_T	k^-_T	k_L
Low	10^7	10^{-4}	10^6	10^{-10}	10^{-6}
High	10^8	10^{-1}	10^8	10^{-1}	10^{-1}

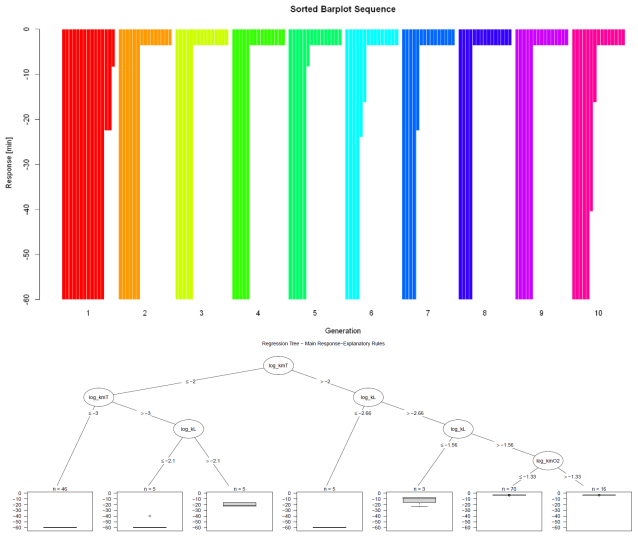

Figure 2: **Top** Parameter intervals for the reaction constants. The algorithm explores parameter combinations within these intervals. **Middle** Sorted barplot sequence for 10 generations shown in different colors (X-axis) with a population size of 10 each. Replication time for the LIDA system, 90% of full replication (Y-axis). Note how very fast replication times, 6:36 min, are discovered already in generation 2. Successive generations explore other regions of parameter space but cannot locate faster replication times. **Bottom** Regression tree with main response explanatory rules discovered by the AI-DoE algorithm. Interestingly, the algorithm firstly discovers how to prevent product inhibition by lowering the hybridization energies for the full strands, which means higher dissociation rate $k^-_m T$. Secondly, the algorithm increases the ligation kinetics k_L. When re-doing the optimization campaign for $k_L < 10^{-2}$, the algorithm reverses the relevant importance of $k^-_m T$ and k_L, so k_L becomes the main rate limiting factor.

give improved results. With each loop iteration, the model is increasingly refined with new accumulated data and its predictive performance improves discovering increasingly better experiments.

Loop component (iii) is the message handling software (AEOP), which is written in Python and currently capable of executing both Python and MATLAB based complex system experiments. AEOP transfers (x_i, α_i) to AI-DoE and α_{i+1} back to $\frac{dx}{dt} = f(x, \alpha)$. The communication between AEOL and AI-DoE is through an Application Interface (API) provided by Daptics (4).

Discussion

Using the AEOL software with the AI-DoE algorithm we achieve fully autonomous optimization of the involved LIDA reaction kinetics. AEOL idenitifies mathematically optimal kinetic parameter combinations for the desired high replication yield, so it may also find parameter combinations that are not physically realizable. Thus a post analysis review of results may be necessary or additional restrictions (expressing physical constraints) need to be imposed in the adaptive search process.

For the optimization process discussed in Fig. 2 the AI-DoE algorithm already identifies an optimal solution in the second generation using a population of 10 in each generation. This means that we could either have asked for fewer generations or used smaller populations in each generation. However, re-doing the optimization with different upper and lower parameter bounds as expected yield different results. If we change the lower bound for the ligation constant to $k_L \leq 10^{-2}$/(mol sec), which is more realistic, and again use a generation population of 10, the algorithm locates optimal solutions after six generations, which is then refined in generation seven. Further, the top parameter in the regression tree with main response explanatory rules (recall Fig. 3, Bottom) becomes k_L, which means that the algorithm views the ligation constant as the main rate limiting reaction.

It should be noted that the same method we use here could be used for automated in vitro experiments instead of simulations. A manual loop utilizing the AI-DoE algorithm is obviously also applicable for exploration and optimization of other complex systems.

References

(1) Alladin-Mustan, B. S., Mitran, C. J., and Gibbs-Davis, J. M. (2015). Achieving room temperature dna amplification by dialling in destabilization. *Chem. Commun.*, 51:9101–9104.

(2) Carothers, J., Goler, J., Juminaga, D., and Keasling, J. (2011). Model-driven engineering of rna devices to quantitatively program gene expression. *Science*, 334:1716–1719.

(3) Engelhardt, J., Andresen, B., and Rasmussen, S. (2020). Thermodynamic and kinetic investigations of of lesion induced dna amplification (lida). *Proceedings of the Artificial Life Conference 2020*.

(4) Packard, N., Zingg, P., and Gazzola, G. *Daptics API*, github repository publicly available at https://github.com/protolife/daptics-api.

(5) Packard, N. and Gazzola, G. Daptics. A white paper describing the fundamentals of daptics' machine-learning methods for experimental discovery and optimization. Publicly available at https://daptics.ai/pdf/White.pdf.

(6) Rasmussen, S., Constantinescu, A., and Svaneborg, C. (2016). Generating minimal living systems from nonliving materials and increasing their evolutionary abilities. *Phil. Trans. Royal Soc. B. Biological Sciences.*, 1701:371.