Machine Learning for Drug Design, Molecular Machines and Evolvable Artificial Cells

Filippo Caschera
FLinT
Institute of Physics and Chemistry
University of Southern Denmark
+45 6550 4438
filippo@ifk.sdu.dk

Martin M Hanczyc
FLinT
Institute of Physics and Chemistry
University of Southern Denmark
+45 6550 4438
martin@ifk.sdu.dk

Steen Rasmussen
FLinT
Institute of Physics and Chemistry
University of Southern Denmark
+45 6550 4438
steen@ifk.sdu.dk

ABSTRACT
An artificial cell is a complex chemical system with many components fabricated and assembled in the laboratory. The molecular components can be designed to interlock in a variety of different way to achieve the emergence of minimal life [1][2]. One experimental design is composed of three modules or sub-systems: lipid vesicles, a metabolic system and a cell-free expression system. Due to the high number of molecular species and their non-trivial interactions in an artificial cell any prediction of the emerging properties in this high dimensional space of compositions is extremely difficult. Previously we have developed and used a machine learning process Evo-DoE (Evolutionary Design of Experiments) coupled with a robotic workstation for liquid handling to optimize a liposomal drug formulation [3] as well as a cell-free expression system for the synthesis of the GFP (green fluorescent protein in vitro) [4]. In addition we have results of vesicle fusion providing a protocol to design a life-cycle for evolvable artificial cells. Now we propose how our technologies could be used to optimize artificial cells.

Categories and Subject Descriptors
Computer methodologies.

General Terms: Algorithms, Design, Experimentation.

Keywords: High throughput screening, fitness function, liposome, cell-free expression system, fusion, life-cycle.

1. INTRODUCTION
The optimization of a liposomal drug formulation [3] and the protein synthesis of a cell-free expression system [4] is driven by a machine learning process that can be engineered to adjusted parameters to obtain targeted properties. The experiments are conducted in iterative cycle, exploiting a neural network type algorithm, and the fitness function value is calculated every time the loop is closed. To start the optimization process, the experimental space is sparsely sampled with a random selection of experiments. Successively the models of the desired response from the experimental data are built, the next sparse sampling of the experimental space is designed - and the process repeats.

Coupling experiments with statistical experimental planning and modeling, predictive algorithms can successful optimize desired chemical behaviors in large high-dimensional experimental spaces. To expand this technology to the design of an artificial cell, an additional level of complexity is necessary. The notion of a life-cycle, which consists of several steps and associated processes. However, we have already developed the key steps in an artificial cell cell-cycle based on vesicles fusion and cell free expression of proteins [5, 6, 7]. But these life-cycle steps are not yet appropriately integrated to account for a living system, which might be mitigated by utilizing the Evo-DoE machine learning process to tune the multiple involved processes.

GECCO'11, July 12-16, 2011, Dublin, Ireland.
ACM 978-1-4503-0690-4/11/07.

2. RESULTS
Iterating our Evo-DoE machine learning algorithm, coupled to high-throughput experiments, a convergence of key variables are observed from generation to generation, and as consequence also the system's fitness value increases. This approach has been verified for liposome drug design [3] and cell free protein synthesis [4].

2.1 Optimization of cell-free expression system for in vitro protein synthesis
As an example this section describes the results of optimization of a cell-free expression system where Evo-DoE is applied.

The cell-free expression system is a commercial *E. Coli* cell extract with defined sets of components also used to express proteins inside the aqueous core of vesicles from DNA encoding molecules [8]. The system contains cellular components required for the transcription/translation of genes (T7 polymerase, ribosome, elongation factors, tRNA, etc.) and these components are adjusted at specific quantitative levels to obtain an efficient yield of synthesized protein *in vitro*. At the current stage of development, the efficiency of the cell-free expression system is low and limited to a narrow range of proteins and genes that can be expressed. The graph shown in figure 1, represents the experimentally measured evolutionary progress of the machine learning algorithm (Evo-DoE). The fitness function is defined as the maximum in fluorescence measured at different time intervals during the expression of the green fluorescence protein (GFP). In conclusion we were with our method able to obtain a 300 % improvement in protein yield, compared to a benchmark recipe. The predictive algorithm (Evo-DoE) indentified the optimal ingredients mixture in the designed experimental space.

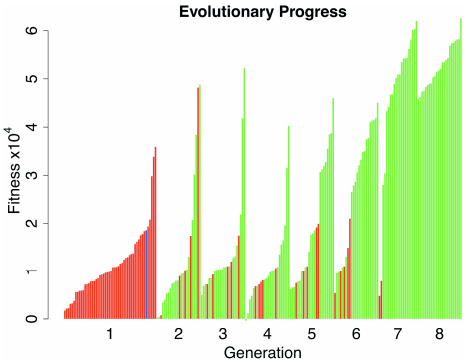

Figure 1. From publication [4]. Progress of the evolutionary algorithm over eight generations is seen for the cell free protein expression system. Experimentally measured fitness. The standard is shown in blue and randomly chosen recipes in red. The light green represents the combinations chosen from the predictive model.

2.2 Vesicles fusion dynamic towards the life-cycle of artificial cells

The relevance of a life-cycle to construct an artificial cell was previously described [9]. In particular the life-cycle is important as feeding mechanism to supply the system with fresh resources, to sustain compartmentalized in vitro to apply selection and ultimately evolution. In such contest the fusion of vesicles is a critical mechanism and it was successfully achieved through exploiting oppositely charged vesicle populations [5, 6, 7]. Recently we showed that adding appropriate electric charge to a suspension of pre-formed POPC vesicles induced mixing of their internal contents and activation of gene expression exploiting an encapsulated reconstituted cell-free expression system (data not shown). The internal mixing was evaluated using a fluorescence signal detectable with FACS (fluorescence activated cell sorter).

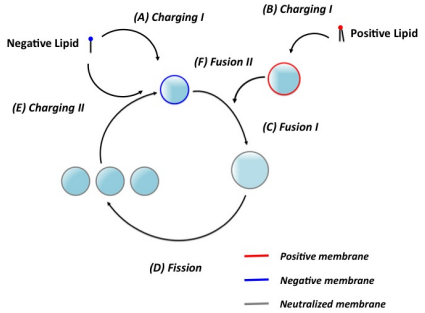

Figure 2. Iterative life-cycle of vesicles fusion and fission is shown based on deposition of electric charge to pre-formed lipid vesicles. After the fusion the electric charge of the membranes **is** neutralized, thereby suitable for a new charging process after the fission.

3. DISCUSSION

So far we have not applied the Evo-DoE method for this cycle, but we aim to integrate our machine learning process to run the vesicle fusion cycle as illustrated in figure 2. The process presented is iterative and therefore suitable for evolutionary studies. We believe that this can be considered a useful strategy for the artificial cells design, by applying it to the turn-over of its building blocks, its up-take of essential molecules to sustains its metabolism as well as to measure evolution.

4. CONCLUSION

In this article we have presented experimental results from a machine learning process (Evo-DoE) coupled to a liquid high throughput robot that optimizes complex chemical systems. It has been demonstrated for optimal drug design [3] and cell free protein synthesis [4]. In addition, we have presented critical steps in an artificial cell cycle through controlled vesicle fusion and gene expression [5, 6, 7]. The involved sub-systems and modules are: the lipid membrane system, the cell-free expression system and the liposome life-cycle system. The last is used to provide resources to an internalized continuous in vitro evolution cycle (CIVE) [10]. In the future, the demonstrated efficiency of robotic workstation for liquid handling coupled with the statistical power of machine learning algorithms will be used for "mixing protocols" and to predict the emerging properties of artificial cells [2] from the very high-dimensional space of compositions.

ACKNOWLEDGMENTS

The teams of scientists at ProtoLife Inc., at the Symbiotic Network Design Laboratory, Osaka University Japan, and at the Center for Fundamental Living Technology, University of Southern Denmark, are gratefully acknowledged for contributing to the development of this research topic.

5. REFERENCES

[1] Rasmussen S., et al., Transitions from nonliving to living matter. *Science* 303: 963-5. 2004

[2] Rasmussen S., et al., Protocells: Bridging nonliving and living matter, MIT Press, Cambridge, 2009

[3] Caschera F., et al., Automated Discovery of Novel Drug Formulations Using Predictive Iterated High Throughput Experimentation. *PLoS ONE* 5(1): e8546. 2010

[4] Caschera F, et al., Coping with complexity: machine learning optimization of cell-free protein synthesis. *Biotechnol. Bioeng.* (In press). 2011

[5] Caschera F., Stano P., Luisi P.L. Reactivity and fusion between cationic vesicles and fatty acids anionic vesicles. *J. Colloid Interf. Sci.* 345: 561-565. 2010

[6] Sunami T*. Caschera F*, et al., Detection of Association and Fusion of Giant Vesicles Using a Fluorescence-Activated Cell Sorter, *Langmuir* 26: 15098–15103. 2010

[7] Caschera F., et. al., Programmed vesicles fusion triggers gene expression *Langmuir* (Submitted). 2011

[8] Hosoda K., et al., Quantitative study of the structure of multilamellar giant liposomes as a container of protein synthesis reaction. *Langmuir* 24: 13540-13548. 2008

[9] Szostak J.W., Bartel D.P., Luisi P.L. Synthesizing life. *Nature* 409:387–390. 2001

[10] Irving R.A., et al., Ribosome display and affinity maturation from antibodies to single V-domains and steps towards cancer therapeutics. *J. Immun. Methods* 248: 31-45. 2001