

Enabling the Discovery of Computational Characteristics of Enzyme Dynamics

Gareth Jones*, Chris Lovell†, Steve Gunn†, Hywel Morgan* and Klaus-Peter Zauner†

*Centre for Hybrid Biodevices

University of Southampton, Southampton, UK

Email: gj07r@ecs.soton.ac.uk

†Electronics and Computer Science

University of Southampton, Southampton, UK

Email: cjl3@ecs.soton.ac.uk

Abstract—Biology demonstrates powerful information processing capabilities. Of particular interest are enzymes, which process information in highly complex dynamic environments. Exploring the information processing characteristics of an enzyme by selectively altering its environment may lead to the discovery of new modes of computation. The physical experiments required to perform such exploration are combinatorial in nature. Thus resource consumption, both time and money, poses major limiting factors on any exploratory work. New tools are required to mitigate these factors. One such tool is lab-on-chip based autonomous experimentation system, where a microfluidic experimentation platform is driven by machine learning algorithms. The lab-on-chip approach provides an automated platform that can perform complex protocols, which is also capable of reducing the resource cost of experimentation. The machine learning algorithms provide intelligent experiment selection that reduces the number of experiments required for discovery. Here we discuss development of the experimentation platform and machine learning software that will lead to fully autonomous characterisation of enzymes.

I. INTRODUCTION

Biological systems exhibit an incredible information processing ability far superior to what can be achieved with conventional computing systems. This capability arises from their molecular components, specifically proteins, and the complex interactions in which they participate [1]. The catalytic proteins or *enzymes* are of particular interest.

Enzymes are capable of recognising specific molecules within a dynamic chemical environment and can be found interacting in large complex networks. Their specificity feature can be utilised to build molecular computers where enzymes are typically employed as switches, or combinations thereof to form simple Boolean logic gates [2], [3], [4], [5]. However given the structural complexity of proteins and their broad range of functionality in nature, the information processing available through individual enzymes is unlikely to be limited to Boolean logic.

Instead by characterising the response behaviour of enzymes available from nature, arguably new modes of information processing could be supported ultimately facilitating the application of enzymatic computers [2]. Fully surveying the response characteristics of an enzyme, in particular in a technical setting not limited to physiological conditions, requires sampling over high-dimensional parameter spaces. Time and materials are

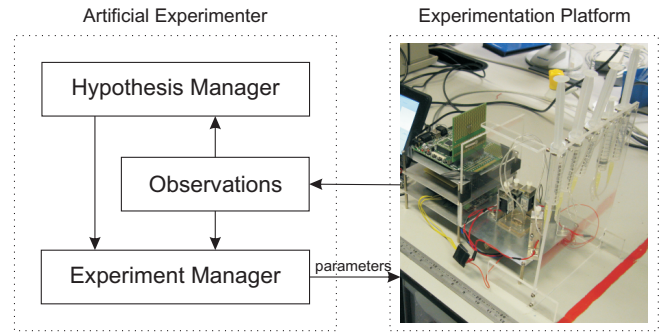


Fig. 1. Overview of lab-on-chip based autonomous experimentation. The artificial experimenter mimics the processes of a successful human experimenter, by analysing the available observations to develop and evaluate a set of working hypotheses, then designing experiments to test these hypotheses. The requested experiment parameters are sent to a lab-on-chip experimentation platform, which uses computer controlled solenoids and on-chip optical measurements to conduct the experiment and return the observation to the artificial experimenter.

typically very limited compared to the large parameter space preventing a detailed investigation of behaviour. New tools are required to address these issues.

A promising tool is autonomous experimentation, where intelligent software drives automated laboratory apparatus in a closed loop manner. The software captures observations from experiments conducted with the apparatus, proposes hypotheses, plans the next set of experiments and then commands the apparatus accordingly (see Fig. 1). Currently there are few autonomous experimentation systems available [6], [7], [8], however within the machine learning community there is increased research activity in developing algorithms capable of efficient active experimental design [9], [10].

Of the examples currently available from both active learning and autonomous experimentation, the issues found in physical experimentation are not fully addressed. On the software side, the resource restrictions are rarely fully appreciated along with errors that can occur within physical experimentation that yield erroneous observations. While on the hardware side, little has been considered in reducing the resource requirements per experiment or in addressing the initial set-up costs of a new

automated hardware platform. For example an autonomous experimentation machine has been developed to allow for high throughput experimentation without need for constant human interaction, using high-cost conventional laboratory automation equipment [8]. Here we discuss components for a new autonomous experimentation system that address these issues.

We present a lab-on-chip experimentation platform designed to conduct enzymatic assays. The lab-on-chip platform reduces the chemical resources required per experiment using microfluidic droplet technology. Furthermore the platform design greatly reduces the requirement for expensive ancillary hardware typically necessary for microfluidic device operation. The device described here also serves to demonstrate that the rapid, low-cost prototyping fabrication techniques we presented in [11] can be extended to create droplet microfluidic devices in a conventional chemistry laboratory. We also present for the first time to the molecular computing community our work on a machine learning component capable of proposing potential hypotheses and selecting the experiments to perform, which is able to minimise the number of experiments required to characterise enzymatic responses [12].

II. ENZYMATIC COMPUTING

The building blocks of an enzyme are amino acid molecules, of which twenty are commonly found in all species. Each amino acid exhibits individual chemical and physical properties. With covalent bonding between the amino acids, polypeptide chains are formed. Individual polypeptide chains fold taking on three-dimensional shapes locally, known as secondary structures, and an overall global three-dimensional shape known as the tertiary structure, as illustrated in Fig. 2. Existing within a tertiary structure are functional domains that can be either catalytic or regulatory. Some enzymes contain multiple tertiary structures and are known as multimeric enzymes. Multimeric enzymes either essentially repeat the functions of a particular tertiary structure, or instead, contain diverse tertiary structures. The structure of an enzyme is dynamic. Ongoing interactions among the amino acids within the enzyme structure cause the conformation of the enzyme to fluctuate among quasi-stable states. Affecting these forms is the physics and chemistry of the environment in which the enzyme resides. Thus it is possible for an enzyme's behaviour to be modified through altering the environment either chemically or physically [13], [14], [15].

From another point of view, it can be said that the enzyme computes from its environmental context a specific conformation and corresponding output behaviour. The output behaviour can be catalytic, electrical or spectral. For example the change in the catalytic activity of malate dehydrogenase in response to the metal ions Ca^{2+} and Mg^{2+} was used in [2]. The same study will be used as a starting point for the autonomous experimentation system.

Malate dehydrogenase (MDH) catalyses the oxidation of malate to oxalacetate by reducing nicotinamide adenine dinucleotide (NAD⁺) to NADH [16]:

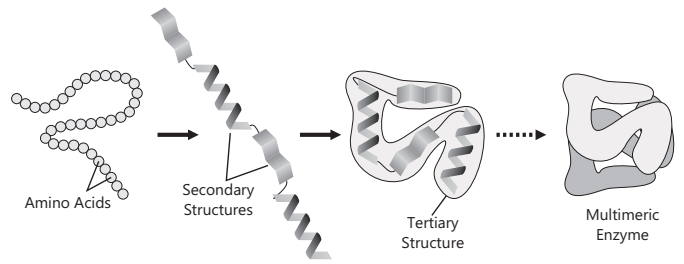
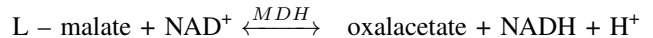


Fig. 2. Structure of an enzyme. The primary structure (left) is formed by the sequence of amino acids, which bond yielding polypeptide chains. With interactions between the individual amino acids, a chain folds into characteristic three-dimensional secondary structures. The secondary structures fold into a functional tertiary structure. An enzyme may consist of multiple subunits that are not necessarily alike.



NADH differs significantly in its absorbance of ultraviolet light at 340 nm when compared to NAD⁺ allowing for convenient monitoring of the enzymatic reaction with semiconductor devices.

In the next sections we consider an automated tool that makes the discovery of computational properties from enzymes more feasible. This achieved through minimising both the chemical resources per experiment and the number of experiments required to characterise a behaviour.

III. LAB-ON-CHIP EXPERIMENTATION PLATFORM

Microfluidic technology enables the manipulation of minute quantities of fluids at the microscale. For example, devices have been made that operate on fluid volumes as low as attolitres (10^{-18} l) [17]. Consequently microfluidic technology is capable of offering substantial chemical resource savings over conventional macroscale laboratory equipment. Additionally, multiple laboratory processes can be incorporated into a single microfluidic device no larger than a few centimetres square. The term *lab-on-chip* is used as a general description of such devices. Automation of laboratory processes on a microfluidic device can be accomplished through utilising microvalves (see [18] for a review). A significant development for automated microfluidic devices was the pneumatically actuated integrated microvalve.

Following Fig. 3, a pneumatically actuated integrated microvalve is constructed as two channels separated by a flexible membrane. One of the channels is used to transmit pressure to the membrane while the other channel is used as a flow channel. Positive pressure drives the membrane into the flow channel therefore restricting flow. Negative pressure pulls the membrane away from the flow channel therefore allowing flow in the case of a normally-closed design.

Early pneumatically actuated microvalves were fabricated primarily in silicon and glass [19], [20]. Photolithography and silicon etching techniques from the microelectronic industry were used to create the silicon and glass microvalve devices. Addressing the need for easier, more-affordable and rapid

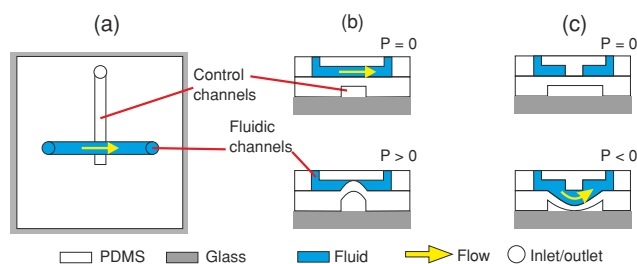


Fig. 3. Two types of pneumatically actuated integrated microvalve. From above two channels, one control and one fluid, intersect each other perpendicularly (a). Cross-section view of a normally open valve which is closed with application of positive pressure (b). Cross-section view of a normally closed valve which is opened with application of negative pressure (c).

prototyping of microfluidic devices, soft lithography was introduced [21]. Consequently there was a move away from silicon and glass materials towards polydimethylsiloxane (PDMS) for constructing microfluidic devices in general.

Pneumatically actuated microvalves too were implemented in PDMS [22], [23] and shown to support construction of on-chip functional components such as peristaltic pumps and rotary mixers [24]. What followed was an unprecedented level of valve integration into microfluidic devices. Microfluidic devices containing hundreds of valves in close proximity were realised [25]. Devices designed to perform complex protocols such as the automatic processing of nucleic acids [26], the generation of complex mixtures for use in on-chip protein experiments [27] and the synthesis of chemical compounds on-chip [28] were demonstrated. Moreover it was shown that device designs containing repeated elements could conduct experiments in parallel [29].

Despite the attractive features of microfluidic technology there are still many problematic issues, particularly within the context of biomolecular computing studies. Ideally we would like to explore interactions within a complex chemical environment. Producing the complex chemical environment would require many chemical inputs, their dilutions and mixing. To achieve this necessitates a high number of associated microfluidic channels, pumps and valves. While theoretically possible, implementing a microfluidic device of complex design with a high valve density may be impracticable. A better solution may lie in utilising droplet microfluidics.

A. Droplet Microfluidics

Droplet microfluidic devices employ two or more immiscible fluids in their operation. One fluid (typically an oil) is flowed through a main channel while a second immiscible fluid (typically aqueous) is injected from another channel into the main flow. The second fluid is dragged downstream by the main flow to a point, where it then shears and forms an isolated droplet.

Control over droplets can be achieved through using integrated pneumatic valves. It has since been shown that droplets can be generated on demand in varying size with only a single

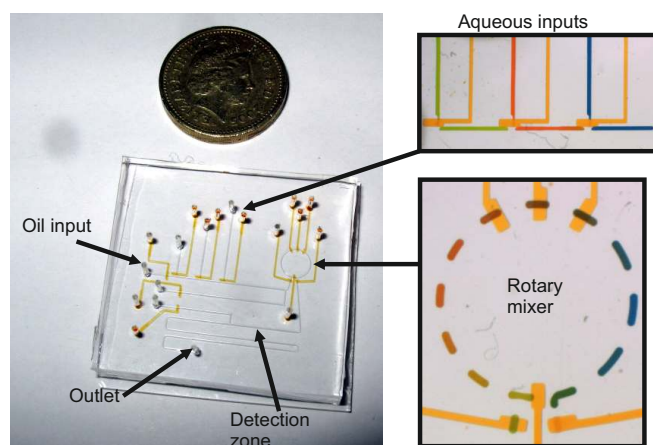


Fig. 4. A prototype droplet microfluidic device pictured alongside a 1 GBP coin. Food dye is used for presentation. Integrated valve channels are filled with yellow food dye. Green, red and blue food dye are used for the three aqueous inlets (top inset). Drops of different composition travel around the loop of the rotary mixer (bottom inset).

valve [30]. Integrated pneumatic valves can be used to create temporary constrictions to facilitate droplet breakup [31]. Similarly controlled fission of droplets has been demonstrated with a pair of valves each acting on a branch of a bifurcated channel [32]. Integrated pneumatic valves have been used to actively trap and fuse separate droplets [33].

There are several additional advantages associated with droplet microfluidics over traditional microfluidics. It is possible to reduce reagent volumes further without resorting to nanoscale channels. Because of reagent confinement to within a droplet, the effect of hydrodynamic (Taylor) dispersion is eliminated therefore reducing measurement error. Samples contained within droplets can be kept separate thus preventing cross-contamination. Mixing of reagents within droplets can be achieved rapidly with simple passive channel designs [34], [35].

Given the advantages discussed above, we consider droplet microfluidics a particularly attractive approach for our experimentation system.

B. A Droplet Microfluidic Experimentation System

Shown in Fig. 4 is a prototype droplet microfluidic device we have developed. The device contains three aqueous input channels that form T-junctions with a main flow channel. Also included in the design is a circular channel containing two isolation valves at the inlet and outlet of the circle, a bypass valve and three additional valves near the top acting together as a peristaltic pump. Just below and beyond the circular channel is a detection region where optical absorbance measurements are recorded. The volume of a droplet measured within the detection zone is 157 nL, which is approximately 4 orders of magnitude smaller than a conventional 10 mm cuvette.

A distinguishing feature of our approach to the device design is a substantial reduction in ancillary macroscopic equipment. Typically droplet microfluidic devices use a single syringe pump per inlet (aqueous and oil) with each costing

easily as much as £2000. Furthermore, control of the overall system is complicated, as the flow rates of each syringe has to be controlled relative to one another and the oil stream.

Here we have used only a single vacuum pump to generate flow within the main channel, which was also sufficient to draw aqueous streams into the main flow when T-junction valves are opened. The vacuum pump was attached to the outlet of the main channel. Pulsing a T-junction valve allowed for droplets of different size to be generated. There was an upper limit to the length of droplet that can be obtained by this approach governed by the physics between the oil and aqueous streams. We have found that by using a valve to stop the main oil stream from flow, then opening up a T-junction valve, very long aqueous drops can be created. The final length of a drop with this approach was governed solely by opening and closing the T-junction valve.

The approach of using a single vacuum pump to drive flow greatly improves the possibility of including numerous chemical inputs. Moreover with drop size now dictated solely by valve actuations, dilutions continuous in range can now be created. With these two features the creation of complex reaction mixtures on-chip should now be realistically achievable.

IV. ALGORITHMS FOR DISCOVERY

The automated lab-on-chip experimentation platform can be complemented by algorithms capable of intelligent experiment selection, which actively seek to minimise the resources used whilst maximising the information gained. Whilst traditional statistical design of experiments techniques such as factorial design may appear as a sensible approach to take, these techniques are often not dynamic with respect to observations obtained, making them ineffective in discovery when the resources are heavily limited due to the potential for repeating experiments in uninteresting regions of the parameter space. Instead we look to machine learning for solutions to this problem and in particular the field of active learning. In active learning the assumption is made that the performance of a learner can be significantly improved if the learner is able to choose the data they learn from [36], [37], [10]. The example often taken in active learning is to consider a classification problem, where the potential experiments are known but the results of those experiments are not, so the learner chooses which experiments to perform. The techniques developed in active learning often provide mathematical guarantees about performance and have been shown to outperform passive learning techniques, however the theoretical assumptions made often do not suitably match those found in physical experimentation. Therefore in designing the algorithms for discovery we ensured that the assumptions we made were compatible to those in physical experimentation. Further to this, in developing the algorithms we took the approach of trying to capture how successful experimenters make decisions within the laboratory, whilst also drawing upon appropriate mathematically sound techniques from machine learning.

The first assumption we made was that physical experimentation is heavily resource restricted, meaning that there

will only typically be a handful of experiments available per parameter dimension. The second assumption we made was that physical experimentation is inherently noisy, meaning that all observations will deviate from the true response through some additive noise distribution, for example Gaussian noise. The noise assumption was taken further, in that occasionally errors occur in experimentation that may undetectable, such as chemical contamination, which will cause the experiment performed to not be the one requested. Therefore a third assumption we made that not all observations are valid, where some observations will be unrepresentative of the actual behaviour that should be observed, which we refer to as erroneous observations produced through shock noise. Finally we assumed that the response characteristics of the enzymes under investigation and of potential usefulness in enzymatic computation, will be both monotonic and non-monotonic [38]. Using these assumptions we developed a set of algorithms, collectively named an artificial experimenter, for performing automated discovery. First we consider hypotheses and how they are represented and proposed, then we consider experiment selection.

A. Hypothesis Management

Using the idea from philosophy of science that multiple hypotheses should be considered to ensure different views of potential behaviour [39], we developed an approach that intelligently proposes a set of working hypotheses that are refined over time in response to new observations obtained [40]. To represent our hypotheses we required a way of being able to represent a response curve, whilst also being able to adjust our belief in the accuracy of observations. To achieve this we used a smoothing spline, a regression technique that can be placed within a Bayesian framework, provides error bars on the prediction, does not impose a particular spectral scale, and can be extended to higher dimensions of independent variables through a thin plate spline [41]. The input to a smoothing spline is a set of observations, a weighting of importance on each observation, and a parameter that defines how smooth the curve will be. We found that by adjusting the weightings we could create hypotheses with different beliefs about the validity of observations. With a zero weight, an observation can be declared erroneous and the hypothesis will ignore it in building its predicted response curve. Whilst setting a higher weighting makes the hypothesis consider the observation to be more accurate, which can cause the curve predicted to move closer to those higher weighted observations. This modification of weighting is important within the hypothesis manager as in situations where there is uncertainty about whether an observation is erroneous or not, both cases can be considered through competing hypotheses, with future experiments deciding the most likely hypothesis. To identify potentially erroneous observations we looked for cases where observations fell outside of the 95% error bar for the prediction of a hypothesis. In these situations we can assume that either the observation is erroneous or that the prediction of the hypothesis is incorrect. We take such hypotheses and refine

their parameters into two new hypotheses, where one new hypothesis considers the observation erroneous and the other considers the prediction of the original hypothesis to be incorrect. Both of these hypotheses are kept within a working set of hypotheses that are considered in parallel, along with other periodically randomly generated hypotheses, which allow for as many different opinions of the shape of the behaviour being observed to be made as possible. When obtaining a prediction from the set of hypotheses, the most confident hypothesis is used to provide the prediction.

B. Experiment Selection

The goal for experiment selection techniques is to minimise the resources spent, whilst maximising the information obtained about the behaviour under investigation. By having a set of potential hypotheses, we are provided with a means for creating an experiment selection algorithm motivated by the idea of falsification. The Popper view of experimentation is that learning and discovery can only be made by disproving existing hypotheses, which in turn allows for new refinements of those hypotheses to be made [42]. The hypothesis manager already had part of this philosophy ingrained within it, through the refinement of hypotheses where observations are obtained that disagree with the hypotheses predictions. We require the experiment selection algorithms to actively select those experiments that will have the greatest chance in disproving hypotheses and in particular disprove those hypotheses that appear to be the most likely. To do this we consider placing experiments where the hypotheses maximally disagree in their predictions for potential experiments [43].

However, such a purely exploitation method of experiment selection will not fare well within discovery with limited resources, as there will be no guarantee that a good hypothesis exists within the working hypotheses under consideration, so discriminating between them could result in the best of a bad set of hypotheses. Instead exploration of the parameter space must also occur, to investigate regions where no data has yet been obtained, to increase the likelihood of interesting features of the behaviour being discovered and represented by the hypotheses. In machine learning, deciding on when to explore and when to exploit is a common problem and exists within many theoretical and real-world problems. In particular is the theoretical multi-armed bandit problem, which has been used to develop understanding on sequential experimental design problems [44]. However, the techniques developed often have an assumption of there being far more resources available than would be typically found within physical experimentation. Therefore to address this problem, we examine how human scientists may capture this trade-off between exploration and exploitation. In related work the notion of experimenters reacting to surprising observations has been described [45], [7]. Where surprising observations should be investigated further to understand the surprise, then continue to explore the parameter space after the surprise is sufficiently understood. This matches to the exploration-exploitation trade-off, where exploitation can be thought of

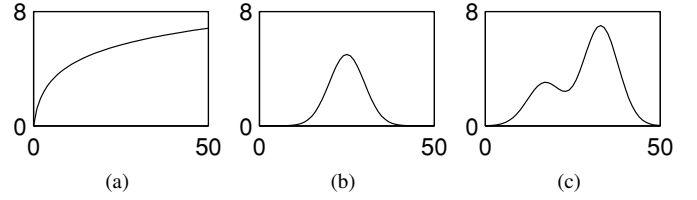


Fig. 5. Underlying behaviours used within a simulated discovery problem with one independent parameter motivated by expected or potential behaviours that would be found within enzyme dynamics. In (a) is a behaviour similar to Michaelis-Menton kinetics. In (b) is a non-monotonic behaviour with a single peak. In (c) is a non-monotonic behaviour with two peaks with different heights on the peaks.

as investigating using falsification to understand why the observation was surprising to the hypotheses, followed by exploration to locate new surprises. The actual techniques used for quantifying or recognising surprise to manage the exploration-exploitation trade-off in the cited work, were ad-hoc and specific to the problems under investigation. Instead we based our mechanism for managing the trade-off on a technique developed in the machine learning literature to quantify surprise in a form similar to information theory [46]. This quantification, first used by the original authors to identify surprising occurrences within video sequences, allows for a measurement to be taken of the surprise within each hypothesis on obtaining a particular observation from an experiment. We consider an observation surprising where the confidence of a hypothesis before an experiment occurs is greater than the confidence of the same hypothesis after an observation is obtained, due to potential falsification of the hypothesis. By summing over all hypotheses, we obtain a system wide value for the surprise to determine overall if the observation was surprising or not, as shown in the following equation:

$$S = \sum_i^{|H|} C(h_i) \log \frac{C(h_i)}{C'(h_i)} \quad (1)$$

where $C(h_i)$ is the prior confidence of the hypothesis before the last experiment performed and $C'(h_i)$ is the posterior confidence for the same hypothesis using the new observation obtained. The confidence is calculated using the summed error between the prediction of a hypothesis and the actual observations available, within some allowed tolerance to try and prevent overfitting of hypotheses to the data:

$$C(h) = \frac{1}{N} \sum_{n=1}^N \exp \left(- \frac{(\hat{h}(x_n) - y_n)^2}{2\sigma^2} \right) \quad (2)$$

where $\hat{h}(x)$ is the prediction of the hypothesis for an experiment with parameters x_n compared to the actual observation y_n over the N previously performed experiments, and σ is a tolerance parameter kept at constant 1.96. To manage the trade-off we exploit by attempting to falsify the hypotheses when the last experiment was surprising with $S > 0$, and explore when the last experiment was not surprising with $S \leq 0$. In order to allow an initial set of hypotheses to be

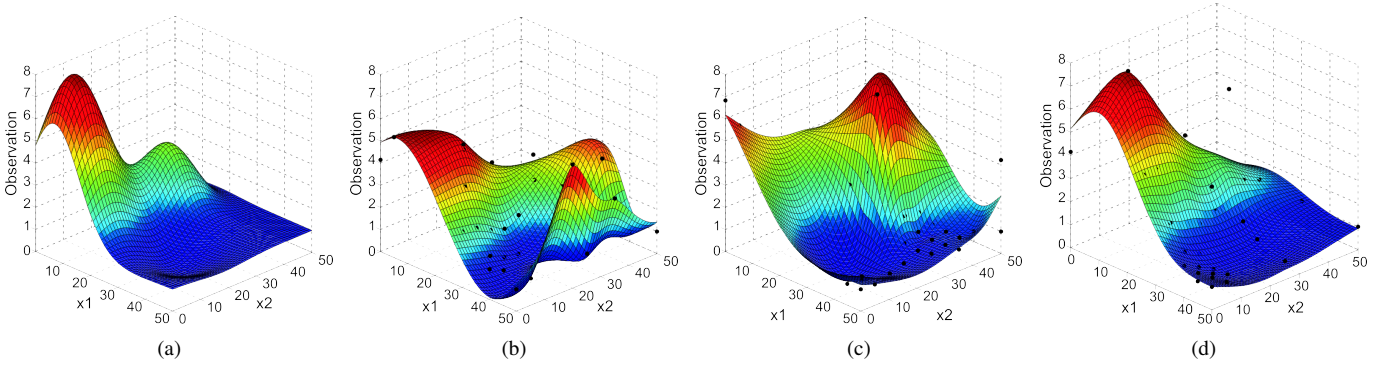


Fig. 6. Illustration of hypotheses proposed by different experiment selection techniques for a 2-dimensional parameter problem after 25 experiments where 12% of the observations obtained were erroneous. In (a) the original underlying behaviour without experimental noise is shown. In (b) the prediction made by a random experiment selection strategy. In (c) the prediction made by an experiment selection strategy that chooses the local maxima in the discrepancy of hypotheses predictions over the parameter space. In (d) the prediction made by the surprise based exploration-exploitation switching strategy. In (b–d) the hypotheses shown have prediction errors representative of the mean error for each over 100 trials of the simulation.

TABLE I

IDENTIFICATION OF STATISTICALLY SIGNIFICANT RESULTS IN THE 1-DIMENSIONAL CASE OVER 15 EXPERIMENTS. THE RESULTS WHERE THERE ARE SIGNIFICANT DIFFERENCES BETWEEN THE MULTIPLE HYPOTHESES SURPRISED BASED EXPERIMENT SELECTION TECHNIQUE AND THE MULTIPLE HYPOTHESES RANDOM EXPERIMENT SELECTION AND MULTIPLE PEAKS EXPERIMENT SELECTION ARE SHOWN. IN ALL CASES, A SIGNIFICANT DIFFERENCE INDICATES THAT THE SURPRISE TECHNIQUE PROVIDES AN IMPROVEMENT OVER THE ALTERNATE TECHNIQUE. THERE ARE NO CASES OF THE SURPRISE TECHNIQUE PERFORMING SIGNIFICANTLY WORSE.

Function	Technique	Active experiments with significant result
Fig. 5(a)	Random	all
	Peaks	12, 14, 15
Fig. 5(b)	Random	10, 11, 13–15
	Peaks	6, 8–15
Fig. 5(c)	Random	3–15
	Peaks	4, 9, 12–15

developed, the first 5 experiments performed are exploration experiments, positioned equidistantly across the parameter space.

C. Evaluating the Technique

To evaluate the algorithms developed we performed a number of simulated discovery problems along with a laboratory characterisation of the co-enzyme NADH. In the simulated discovery problems we provided a number of different behaviours that were motivated from either behaviours we may expect to see, or behaviours that would be of use within enzymatic computation. An example of these are given for a 1-dimensional parameter space first considered in Fig. 5, where one behaviour is similar to Michaelis-Menton kinetics [47] and the others provide non-monotonic behaviours that could be of use within enzymatic computation [38]. In the simulation observations were calculated from the behaviour using the requested experiment parameter, then all were adjusted with Gaussian noise ($N(0, 0.5^2)$) and 25% of the observations had shock noise applied to them to provide erroneous observations.

In evaluating the technique we compared the experiment selection against random experiment selection and a technique

that performs experiments at the local maxima of the disagreement between hypotheses over the potential parameters. This last technique is motivated from the belief that hypothesis disproof can occur wherever the disagreement between hypotheses is high and that hypotheses may disagree over different aspects of the behaviour, where the different aspects can be identified by looking for local maxima in disagreement [43]. In the one dimensional problem we found that the surprise technique provided a statistically significant improvement over the alternate techniques, where significance was calculated using a T-Test over 100 runs of the simulation, as shown in Table I. In each case the techniques had a maximum of 15 experiments that could be used and the first 5 were placed equidistantly across the parameter space.

Next we evaluated the algorithms within a 2-dimensional parameter space. This time 25 experiments were available to be used, again with the first 5 being placed equidistant across the parameter space to allow for an initial set of hypotheses to be generated. All observations had random Gaussian noise applied to them and 12% of the observations were erroneous. The simulation showed that the surprise technique provided a significant improvement over the alternate techniques, with the results best shown using the example given in Fig 6, where the surprise technique is the only strategy capable of giving a reasonable prediction of the underlying behaviour.

Finally we performed a laboratory trial to characterise the co-enzyme NADH. As discussed in Section II, NADH is an important part of enzymatic reactions and has been considered within previous work investigating enzymatic computation [2], [7]. Additionally NADH has the nice property that it has a non-linear behaviour where part of the behaviour can be predicted using the Beer-Lambert law [47]. We performed a laboratory trial where the machine learning component controlled the concentration of NADH to use in the experiment and produced potential hypotheses, the process of this procedure is discussed further in Lovell et al. [12]. This trial was performed to validate the surprise technique and the simulated problems used. The prediction provided by the surprise technique for

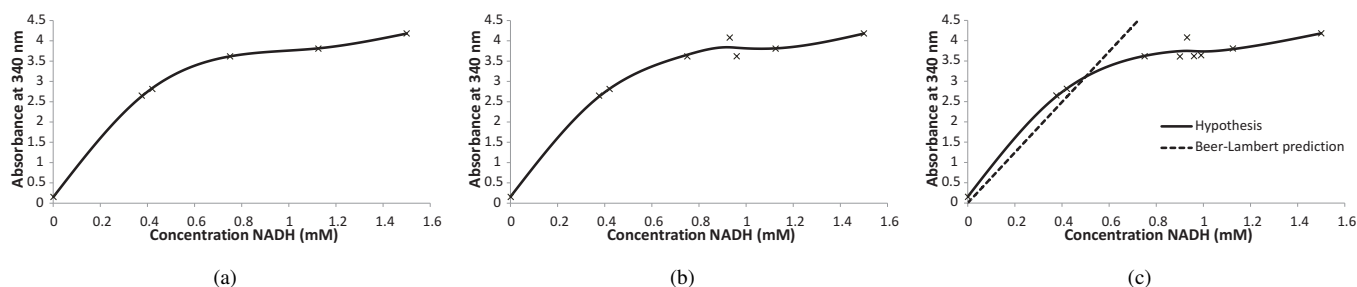


Fig. 7. Illustration of hypothesis development and experiment placement over time in a one parameter laboratory characterisation of NADH. In (a) the most confident hypothesis after 6 experiments is shown, where the first five experiments are placed equidistant across the parameter space and the sixth is actively chosen to investigate the curve of the behaviour at 0.42 mM. In (b) the most confident hypothesis after 8 experiments is shown, where the next two actively chosen experiments obtain noisy observations causing repeat trials. In (c) the most confident hypothesis after 10 experiments is shown, where the remaining experiments further evaluate the erroneous observations. A comparison to the theoretical response is shown in (c), where the prediction closes matches the theoretical value for the linear region.

a characterisation of NADH over 10 experiments is shown in Fig 7. The technique was able to sufficiently match the theoretical prediction, detect the non-linearity of the behaviour and focus a number of experiments on a region where more noisy or potentially erroneous observations were obtained.

V. CONCLUSION

The dynamics and interactions of enzymes may contain repeatable behaviours that can be harnessed for new modes of computation. However, the complexity of the interactions presents large and high dimensional parameter spaces to experiment within, make discovering such interesting repeatable behaviours difficult. Currently the progress of enzymatic computation has been slow. In order to allow enzymatic computation to progress, we believe that new tools are needed to aid the researcher in identifying interesting and repeatable behaviours cost effectively. Tools such as autonomous experimentation allow for this, as they reduce the number of experiments required through statistical machine learning techniques, while microfluidic technology allows for the chemical cost per experiment to be significantly decreased.

The technologies presented also have wider implications for unconventional computing and understanding complex systems. The algorithms presented within the artificial experimenter, although motivated by enzyme characterisation, are generic and can be applied to other discovery problem where there are a large number of possible parameters and a cost restriction. The discovery problems need not be directly related to physical experimentation. For example, in systems biology there is a problem that very large sets of data are available, but the processing time restricts how much data can be analysed, so autonomous experimentation algorithms could be used to help choose the data to process [48]. Further to this, the microfluidic technology presented here has many potential applications within enzymatic computation. The flexibility of the technology, small resource requirements and ability to produce rapid prototypes, make it an ideal platform for hosting enzymatic computational systems or other chemical computing techniques. The new processes we have developed for producing these devices without the need for sophisticated

laboratory equipment, make the entry level requirements for using this technology much more feasible than it once was.

ACKNOWLEDGMENT

The reported work was supported in part by a Microsoft Research Faculty Fellowship awarded to Klaus-Peter Zauner. This work was supported in part by the IST Programme of the European Community, under the PASCAL2 Network of Excellence, IST-2007-216886. This publication only reflects the authors' views.

REFERENCES

- [1] D. Bray, "Protein molecules as computational elements in living cells," *Nature*, vol. 376, pp. 307–312, 1995.
- [2] K.-P. Zauner and M. Conrad, "Enzymatic computing," *Biotechnol. Prog.*, vol. 17, pp. 553–559, 2001.
- [3] S. Sivan, S. Tuchman, and N. Lotan, "A biochemical logic gate using an enzyme and its inhibitor. part ii: The logic gate," *BioSystems*, vol. 70, pp. 21–33, 2003.
- [4] R. Baron, O. Lioubashevski, E. Katz, T. Niazov, and I. Willner, "Logic gates and elementary computing by enzymes," *J. Phys. Chem. A*, vol. 110, pp. 8548–8553, 2006.
- [5] A. P. de Silva and S. Uchiyama, "Molecular logic and computing," *Nature Nanotechnology*, vol. 2, pp. 399–410, 2007.
- [6] J. M. Żytkow, J. Zhu, and A. Hussam, "Automated discovery in a chemistry laboratory," in *Proceedings of the 8th National Conference on Artificial Intelligence*. Boston, MA: AAAI Press / MIT Press, 1990, pp. 889–894.
- [7] N. Matsumaru, S. Colombano, and K.-P. Zauner, "Scouting enzyme behavior," in *World Congress on Computational Intelligence*. Honolulu, Hawaii: IEEE, 2002, pp. 19–24.
- [8] R. D. King, K. E. Whelan, F. M. Jones, P. G. K. Reiser, C. H. Bryant, S. H. Muggleton, D. B. Kell, and S. G. Oliver, "Functional genomic hypothesis generation and experimentation by a robot scientist," *Nature*, vol. 427, pp. 247–252, 2004.
- [9] I. Guyon, G. Cawley, G. Dror, V. Lemaire, and A. Statnikov, Eds., *Challenges in Causality: Active Learning Challenge*, ser. series name, 2011, vol. 3.
- [10] B. Settles, "Active learning literature survey," University of Wisconsin-Madison, Tech. Rep., 2009.
- [11] G. Jones, C. J. Lovell, H. Morgan, and K.-P. Zauner, "Organising chemical reaction networks in space and time with microfluidics," *International Journal of Nanotechnology and Molecular Computation (IJNMC)*, vol. 3, no. 1, pp. 35–56, 2011.
- [12] C. J. Lovell, G. Jones, S. R. Gunn, and K.-P. Zauner, "An artificial experimenter for enzymatic response characterisation," in *13th International Conference on Discovery Science*, Canberra, Australia, 2010, pp. 42–56.
- [13] M. Gross and R. Jaenicke, "Proteins under pressure," *European Journal of Biochemistry*, vol. 221, no. 2, pp. 617–630, 1994.

- [14] S. Y. Ho, G. S. Mittal, and J. D. Cross, "Effects of high field electric pulses on the activity of selected enzymes," *Journal of Food Engineering*, vol. 31, pp. 69–84, 1997.
- [15] A. C. Møller and L. F. Olsen, "Effect of magnetic fields on an oscillating enzyme reaction," *Journal of the American Chemical Society*, vol. 121, pp. 6351–6354, 1999.
- [16] S. England and L. Siegel, "Mitochondrial l-malate dehydrogenase of beef heart," in *Citric Acid Cycle*, ser. Methods in Enzymology, J. M. Lowenstein, Ed. Academic Press, New York, 1969, vol. XIII.
- [17] G. M. Whitesides, "The origins and future of microfluidics," *Nature*, vol. 442, pp. 368–373, July 27th 2006.
- [18] K. W. Oh and C. H. Ahn, "A review of microvalves," *Journal of Micromechanics and Microengineering*, vol. 16, pp. R13–R39, 2006.
- [19] M. Esashi, "Silicon micromachining for integrated microsystems," *Vacuum*, vol. 47, no. 6–8, pp. 469–474, 1996.
- [20] T. Otori, S. Shoji, K. Miura, and A. Yotsumoto, "Partly disposable three-way microvalve for micro total analysis system (μ TAS)," *Sensors and Actuators A*, vol. 64, pp. 57–62, 1998.
- [21] D. C. Duffy, J. C. McDonald, O. J. A. Schueller, and G. M. Whitesides, "Rapid prototyping of microfluidic systems in poly(dimethylsiloxane)," *Analytical Chemistry*, vol. 70, no. 23, pp. 4974–4984, 1998.
- [22] M. A. Unger, H.-P. Chou, T. Thorsen, A. Scherer, and S. R. Quake, "Monolithic microfabricated valves and pumps by multilayer soft lithography," *Science*, vol. 288, pp. 113–116, April 7th 2000.
- [23] K. Hosokawa and R. Maeda, "A pneumatically-actuated three-way microvalve fabricated with polydimethylsiloxane using the membrane transfer technique," *Journal of Micromechanics and Microengineering*, vol. 10, pp. 415–420, 2000.
- [24] H.-P. Chou, M. A. Unger, and S. R. Quake, "A microfabricated rotary pump," *Biomedical Microdevices*, vol. 3, no. 4, pp. 323–330, 2001.
- [25] T. Thorsen, S. J. Maerkl, and S. R. Quake, "Microfluidic large-scale integration," *Science*, vol. 298, pp. 580–584, 2002.
- [26] J. W. Hong, V. Studer, G. Hang, W. F. Anderson, and S. R. Quake, "A nanoliter-scale nucleic acid processor with parallel architecture," *Nature Biotechnology*, vol. 22, no. 4, pp. 435–439, April 2004.
- [27] C. L. Hansen, M. O. A. Sommer, and S. R. Quake, "Systematic investigation of protein phase behaviour with a microfluidic formulator," *PNAS*, vol. 101, no. 40, pp. 14 431–14 436, October 5th 2004.
- [28] C.-C. Lee, G. Sui, A. Elizarov, C. J. Shu, Y.-S. Shin, A. N. Dooley, J. Huang, A. Daridon, P. Wyatt, D. Stout, H. C. Kolb, O. N. Witte, N. Satyamurthy, J. R. Heath, M. E. Phelps, S. R. Quake, and H.-R. Tseng, "Multistep synthesis of a radiolabeled imaging probe using integrated microfluidics," *Science*, vol. 310, pp. 1793–1796, 2005.
- [29] S. Jambovane, E. C. Duin, S.-K. Kim, and J. W. Hong, "Determination of kinetic parameters, k_m and k_{cat} , with a single experiment on a chip," *Analytical Chemistry*, vol. 81, no. 9, pp. 3239–3245, 2009.
- [30] S. Zeng, B. Li, X. Su, J. Qin, and B. Lin, "Microvalve-actuated precise control of individual droplets in microfluidic devices," *Lab on a Chip*, vol. 9, pp. 1340–1343, 2009.
- [31] J.-H. Choi, S.-K. Lee, J.-M. Lim, S.-M. Yang, and G.-R. Yi, "Designed pneumatic valve actuators for controlled droplet breakup and generation," *Lab on a Chip*, vol. 10, pp. 456–461, 2010.
- [32] H.-W. Zhu, N.-G. Zhang, R.-X. He, S.-Z. Li, S.-S. Guo, W. Liu, and X.-Z. Zhao, "Controllable fission of droplets and bubbles by pneumatic valve," *Microfluid Nanofluid*, vol. 10, pp. 1343–1349, 2011.
- [33] B.-C. Lin and Y.-C. Su, "On-demand liquid-in-liquid droplet metering and fusion utilizing pneumatically actuated membrane valves," *Journal of Micromechanics and Microengineering*, vol. 18, pp. 1–10, 2008.
- [34] H. Song, J. D. Tice, and R. F. Ismagilov, "A microfluidic system for controlling reaction networks in time," *Angew. Chem. Int. Ed.*, vol. 42, no. 7, pp. 768–772, 2003.
- [35] H. Song, M. R. Bringer, J. D. Tice, C. J. Gerdt, and R. F. Ismagilov, "Experimental test of scaling by mixing by chaotic advection in droplets moving through microfluidic channels," *Applied Physics Letters*, vol. 83, pp. 4664–4666, 2003.
- [36] D. Cohn, L. Atlas, and R. Ladner, "Improving generalization with active learning," *Machine Learning*, vol. 15, pp. 201–221, 1994.
- [37] D. J. C. MacKay, "Information-based objective functions for active data selection," *Neural Computation*, vol. 4, pp. 589–603, 1992.
- [38] K.-P. Zauner, "Molecular information technology," *Critical Reviews in Solid State and Material Sciences*, vol. 30, no. 1, pp. 33–69, 2005.
- [39] T. C. Chamberlin, "The method of multiple working hypotheses," *Science (old series)*, vol. 15, pp. 92–96, 1890, reprinted in: *Science*, v. 148, p. 754–759, May 1965.
- [40] C. J. Lovell, G. Jones, S. R. Gunn, and K.-P. Zauner, "Characterising enzymes for information processing: Towards an artificial experimenter," in *9th International Conference on Unconventional Computation*, C. S. C. et al., Ed., vol. 6079, Tokyo, Japan, 2010, pp. 81–92.
- [41] G. Wahba, *Spline Models for Observational Data*, ser. CBMS-NSF Regional Conference series in applied mathematics. Philadelphia, PA: Society for Industrial and Applied Mathematics, 1990, vol. 59.
- [42] K. Popper, *The Logic of Scientific Discovery*, 2nd ed. New York: Harper & Rowe, 1968.
- [43] C. J. Lovell, G. Jones, S. R. Gunn, and K.-P. Zauner, "Autonomous experimentation: Active learning for enzyme response characterisation," *JMLR: Workshop and Conference Proceedings*, vol. 16, pp. 141–154, 2011.
- [44] H. Robbins, "Some aspects of the sequential design of experiments," *Bulletin on the American Mathematical Society*, vol. 58, no. 5, pp. 527–535, 1952.
- [45] D. Kulkarni and H. A. Simon, "Experimentation in machine discovery," in *Computational Models of Scientific Discovery and Theory Formation*, J. Shrager and P. Langley, Eds. San Mateo, CA: Morgan Kaufmann Publishers, 1990, pp. 255–273.
- [46] L. Itti and P. Baldi, "Bayesian surprise attracts human attention," *Vision Research*, vol. 49, pp. 1295–1306, 2009.
- [47] D. L. Nelson and M. M. Cox, *Lehninger Principles of Biochemistry*, 5th ed. New York, USA: W. H. Freeman and Company, 2008.
- [48] N. Matsumaru, F. Centler, K.-P. Zauner, and P. Dittrich, "Self-adaptive scouting—autonomous experimentation for systems biology," in *Applications of Evolutionary Computing: EvoWorkshops*, ser. LNAI, vol. 3005. Springer, 2004, pp. 52–62.