Scouting Enzyme Behavior

Naoki Matsumaru\(^\dagger\), Silvano Colombano\(^\dagger\), and Klaus-Peter Zauner\(^\dagger\)

\(^\dagger\)BioComputing Group
Department of Computer Science
Wayne State University
Detroit, MI 48202, USA
biocomputing@cs.wayne.edu

\(^\ddagger\)Evolutionary Biotronics
Computational Sciences Division
NASA-Ames Research Center
Moffett Field, CA 94035, USA
scolombano@mail.arc.nasa.gov

Abstract - Experimental exploration of enzymatic response in a multi-dimensional context faces the challenge of an explosive number of possible milieu conditions. We address this problem with an evolutionary search strategy that scouts the physiochemical milieu space for unanticipated enzyme behavior and rewards the discovery of experimental conditions that yield surprises.

I. EVOLUTIONARY EXPERIMENTATION

Elucidating the marvelous complexity of biological systems requires large amounts of empirical information. Experiments that can be conducted to acquire this information are a limited resource, the effective allocation of which is complicated by the very complexity one wishes to explore. Complexity is crucial to the functioning of biological systems and compressing it into simple mechanistic models risks eliminating vital information [4].

Evolutionary experimentation provides an approach to probing complex systems that is free of preconceived mechanistic models [1,20,21]. It has successfully been used to search large parameter spaces in experimental settings. Both the refinement and the discovery of solutions to engineering problems have been demonstrated [23,27]. Aside from its engineering applications, evolutionary experimentation appears to be well suited to investigating the behavior of biological systems. We report on a small step in this direction: an evolutionary search strategy was used in conjunction with the computer controlled experimental setup shown in Fig. 1 to map out the response of an enzyme to chemical signals.

II. ENZYMES

Proteins are macromolecules consisting of linear chains of typically a few hundred covalently bound amino acids. Under the influence of intra- and intermolecular forces, the chains can curl up into functional units of characteristic three-dimensional structure [9,26]. The spatial shape of the folded protein, which is typically in the size range of 5–10 nm diameter, is flexible and sensitive to the protein’s environment [10]. Enzymes are proteins that serve as very effective and relatively specific catalysts in biological cells. The catalytic activity of an enzyme critically depends on its shape [31]. Thus enzymatic activity is modulated by a multitude of signals from the enzyme’s physiochemical context which are selected and fused through the enzyme’s conformational dynamics [5,8,25]. Consequently complex milieu-factor interactions can arise from the non-linear conformational dynamics of an enzyme [15]. In this case the effect of a concentration change of an individual chemical component in the enzyme’s milieu cannot be predicted without taking the concentrations of other milieu-factors into account.

We chose malate dehydrogenase (MDH) as the test-bed for our scouting method. MDH is an enzyme that occurs in a wide variety of species, including plants. Substantial effort has been expended to clarify its kinetic and regulatory mechanism [17,29]. The complexity that emerged far exceeded expectations, however, and the investigation into factors affecting MDH activity is still ongoing [11,22]. MDH catalyzes the oxidation of malate to oxalacetate by reducing the oxidized form (NAD+) of nicotinamide adenine dinucleotide to the reduced form NADH [7]:

\[
\text{L - malate + NAD}^+ \leftrightarrow_{\text{MDH}} \text{oxalacetate + NADH + H}^+
\]

NADH differs in its absorption of ultraviolet light significantly from NAD\(^{+}\), a property that allows for the convenient monitoring of the reaction by measuring the increase in NADH concentration with a spectrophotometer [14]. At high pH the equilibrium for the reaction is on the right side. The progress of the reaction is affected by the composition of the reaction milieu. The two factors which we varied in the experiments described here were citrate, a known regulator with context dependent activating or inhibiting effect on MDH [11] and MgCl\(_2\), also reported to have both activating and inhibiting effects on MDH [2,6,30]. The factors that could be investigated are not limited to chemical compounds known to have a physiological function, however. The search space is exceedingly large. Rarely are mechanistic models available to guide the exploration of enzyme behavior. We therefore decided to address the mapping of enzyme response with an evolutionary experimentation method.

Figure 1: Laboratory setup used in the scouting experiments. A peristaltic pump and a reservoir of distilled water (a) serve to flush the cuvette that is installed in the spectrophotometer (b). Servo driven syringe pumps and valves (c) are controlled through a serial interface (e) by the computer that runs the evolutionary search program. The syringe pumps first mix chemical milieus from stock solutions (d) in the photometer cuvette. A reaction is then initiated by injecting enzyme solution into the cuvette and its progress is monitored with the photometer.
III. SCOUTING

Evolutionary search can direct an empirical exploration of a complex system [19]. Fig. 2 illustrates the method we use. An evolution strategy breeds parameter settings for experiments. The genotype specifies a set of factor levels consistent with experimental constraints. Two actions are taken to determine the fitness of a genotype. First, the experience obtained in past experiments is used to make a prediction of what would happen if an experiment with the factor settings of the genotype were performed. This is the expectation regarding the outcome of the specified experiment. Second, the factor levels prescribed by the genotype are forwarded to a computer controlled experimental setup (described in detail below) which autonomously conducts the requested experiment and returns the measured response. The empirically observed response, i.e., the phenotype, is then compared to the expectation derived from the data in the experience database in order to appraise the corresponding genotype. Here we borrow from communication theory the notion that the information content of a message corresponds to its ‘surprise value’ [3,13,16,24].

The key point is that large deviation between expectation and experimental observation is associated with high fitness. From this it follows that the more unexpected the outcome of an experiment is, the more likely similar parameter settings will be investigated in subsequent experiments. At the same time, however, the new observation is added to the experience database and as a consequence the expectations for experiments under similar conditions are likely to be closer to the measured response. Thus a factor combination that leads to a surprising response will initially attract further investigation through the descendants of the genotype that discovered the combination. Then, as more information on the factor combination accumulates in the experience database, the response to such factor combinations will become ‘common knowledge’ and therefore the fitness of genotypes specifying experiments in this region of the factor space will be low. Incidentally this property prevents the scouting method from getting stuck on a local fitness peak—an important issue in the context of evolutionary experimentation, since fitness evaluation is generally costly, necessitating the use of small populations. Let us turn now to the details of the scouting implementation used in the present experiments.

A. Evolution of Experiments

The algorithm starts with an initial population of three genotypes (generation 0) each initialized with one of the extreme factor levels: none of the two milieu components, only milieu component 1, and only milieu component 2 present. It then takes the first genotype with unknown fitness found in the population and evaluates its fitness by conducting an experiment. The information gained in the experiment is added to the experience database and the next genotype is evaluated. When the fitness of all genotypes has been determined, the single genotype with the highest fitness is chosen as the parent of the next generation. In creating offsprings the constraint that the sum of the normalized factor levels in a genotype can not exceed 1 needs to be satisfied.

We applied two different operators to derive the genotypes that form the next generation. Three genotypes are created by changing in turn the volume of one of the three constituents of the milieu (the neutral reaction buffer, the reaction buffer with MgCl₂, and the reaction buffer with citrate) by a random amount consistent with the requirement that the normalized volume has to be within the 0–1 range. The change is then compensated by adjusting the remaining two volumes so that the sum of the normalized factor levels equals one. Two further genotypes are created by first assigning to each milieu constituent a random amount of change and then attempting to scale these proposed negative and positive changes such that the volume is kept constant. If such scaling is not possible, a new set of randomly generated changes is assigned and scaling for constant volume is again attempted.

Lastly, if required, an adjustment for rounding errors is made to insure that the normalized volume contributions have a sum of one. Thus a new generation of five genotypes is obtained. Note that the parent is not preserved. Once evaluated, it has contributed an empirical observation to the experience database and hence would have a low fitness (zero in the absence of measurement noise) in a subsequent evaluation.

B. Experience and Expectation

Every genotype evaluation causes a computer controlled experiment to be performed. The outcome of the experiment is a time series of absorption measurements reflecting the activity of the enzyme which catalyzes the formation of the measured product. Accordingly in the experiments reported
here, the observed response to the factor levels prescribed by the genotype is a sequence of over 300 measurements. To facilitate the representation of the reaction process in the experiment database we fit the following expression to the measured data:

\[
a(t) = c_1 (1 - e^{-c_2 t}) + c_4 t + c_5 t^2 + c_6 t^3
\]

We need to emphasize that equation 1 should not be construed as a mechanistic model of the reaction process; there is no interpretation for the parameters \(c_1-c_6\). Furthermore, specific values for them cannot be ascertained from the measured data. Nevertheless, in combination the six parameters serve well to represent the measurements, to interpolate for the brief moments between the measurements, and to smooth noise. Note that the content of the experiment database constitutes an empirical model that characterizes the system under investigation.

The entries in the experience database contain the factor levels together with the parameters representing the observation. To obtain the expectation corresponding to a particular factor setting, first \(n\) (or less if not enough entries exist) nearest neighbors in the parameter space and their distances to the factor setting under consideration are determined. Then the expectation \(a_{ex}(t)\) is calculated from the parameter sets of these \(n\) entries averaging with inverse cubic distance weighting \((w_i)\):

\[
a_{ex}(t) = \sum_{i=1}^{n} w_i a_i(t)
\]

In the present experiments we used \(n = 5\). For the first fitness evaluation, with no prior experience available, we arbitrarily set \(a_{ex}(t) = 0\).

C. Surprise as Fitness

The surprise value \((s)\) of an observation is calculated by numerically integrating the absolute difference between the expected absorption \(a_{ex}(t)\) and the observed absorption \(a_{ob}(t)\):

\[
s = \frac{1}{3} \int_{t=0}^{t=300} \text{abs}[a_{ex}(t) - a_{ob}(t)]
\]

The factor \(\frac{1}{3}\) normalizes the absorption values. Accordingly the fitness of a genotype is the surprise value calculated from the expectation and experimental observation at the factor settings represented by it.

IV. RESULTS

The scouting above described was used to conduct an experiment aimed at mapping out the sensitivity of the reaction catalyzed by MDH with respect to two effector substances, MgCl2 and citrate. Both were varied over the range 0–300 mM. The system probed 120 locations in the factor space, evolving for 23 generations (cf. Fig. 3).

Every three generations, i.e., approximately every 2\(\frac{1}{3}\) h, a control assay at a fixed factor combination of 0.25 relative concentration of both MgCl2 and citrate was automatically inserted. Including the eight control assays the experiment was run for 21 h 50 min. No systematic change in reaction speed was detected during this time. The control assays are withheld from the experience database.

A key feature of evolutionary experimentation is that the details of the experimental strategy are not fixed from the outset, but rather each measurement can have an immediate effect on the decision as to which test to perform next. Fig. 3 depicts the progress of these sampling dynamics. On the left the locations in the factor space at which experiments were performed are shown for four time slices. Note that the milieu component axes are scaled as relative concentration, therefore only the milieu compositions represented by the lower left half of the diagrams are possible. Bars in the right graph show the development of the maximum surprise value found within the population; the curve shows its average surprise value. The population size was 5 except in generation 0 which was initialized with three genotypes. At generation 21 we note a high maximum surprise value due to a surprising observation (#113) which fell in a hitherto unexplored area of the factor space (0.29, 0.47) as pointed out by the left arrow in Fig. 3. As can be seen from the bar graph in the figure, peak surprise values have not yet leveled off and consequently further information about the behavior of the enzyme would be gained by taking more than the 120 samples.

The observations collected in the scouting process can be combined into an empirical description of the probed phenomenon. Two snapshots of the time development of the absorption during the reaction are shown in Fig. 4. The contours allow us to visualize the interaction between the two milieu-factors.

Figure 3: Temporal progress of scouting, see text for details.
V. CONCLUSIONS

We developed a low-cost computer controlled fluidics setup that serves as a test-bed to investigate algorithms for autonomous experimentation. We used this setup to probe the enzyme malate dehydrogenase by means of an evolutionary experimentation algorithm for its response to chemical signals. A search strategy rewarding the discovery of experimental conditions that yield unexpected results was employed. Our system can be used to explore both individual enzyme species and networks of enzymes.

We found that evolutionary experimentation can be employed to empirically map the behavior of a complex biomacromolecule. However, at the current stage no conclusion is possible as to whether scouting will outperform other sampling strategies in high-dimensional factor spaces. Resilience to random errors (e.g., air bubbles in the flow cuvette) was good. On the other hand temporal trends in systematic errors can cause problems since locations in close proximity in the factor space are often also sampled close in time. Further development of the algorithm should address this issue. Nevertheless, in the light of the results obtained so far evolution is a promising candidate for autonomous experimentation. Furthermore, the current growth of laboratory automatization will allow for ready deployment of such algorithms. On a longer perspective the automated construction of empirical enzyme response models could provide complementary information to expression screening and protein structure determination.

Many questions remain open, but it appears that in the future evolutionary experimentation methods may help to empirically unravel complex biological systems that are difficult to approach from a mechanistic viewpoint.

NOTES ON MATERIALS AND METHODS

A. Computer Controlled Fluidics

In order for the evolutionary search program to guide experiments we constructed a fully automated system that has the ability to compose chemical milieus from a variety of components and to perform enzyme assays in these prepared milieus. This setup provides computer control over 6 disposable 3 ml syringes, 6 t-valves (Kontes 3-way stopcock, VWR Scientific Products, West Chester, PA), 3 on-off-valves (t-valves with less servo movement), and a peristaltic pump (Mini-Pump III, Control Company, Friendswood, TX). These elements can be combined manually with nylon Luer lock fittings (Small Parts Inc., Miami Lakes, FL) and flexible tubing with an inner diameter of 1.6 mm (Tygon, S-50-HL, Fisher Scientific, Pittsburgh, PA) to form various fluidics circuits. Figure 5 illustrates the circuit used in the experiments described here.

Each of the six syringes (labeled S1–S6) is configured with a check valve inlet (acrylic supravalves, Small Parts Inc.) and a servo-controlled outlet valve (T1–T3 and L1–L3) to act as piston pump. We calibrated these pumps by pumping water onto an analytic scale; the typical total error (systematic plus random) is within 50 μl, with the best circuit (here used for the enzyme solution) approaching half this value. A peristaltic pump together with valves T4–T6 allowed for the automatic draining and washing of the cuvette after each reaction. Washing of the cuvette comprises three cleaning cycles of 2.33 ml distilled water that remains for 5 s in the cuvette. The valves and syringes are driven by positive pulselwidth modulated servos (available for use in radio controlled airplane models) with a nominal torque of 19.8 kg/cm at 48 V (CS-80 2BB, Hobbico, Champaign, IL). We supplied the servos with 5.0 V and used two RS232 to eight channel pulse-proportional serial servo controllers (Mini SSC II, Scott Edwards Electronics, Sierra Vista, AZ) at 9600 baud for the control signals. One of the servo controllers was configured for high resolution/short movement (servos 0–7 in Fig. 5), the other for maximum movement (servos 8–15). To measure the absorption of NADH a flow cuvette (Fig. 6) was built and installed in an Ultraspex II Model 4050 spectrophotometer (LKB Biochrome, Cambridge, UK). Six capillaries (C1–C6 in Fig. 5) are used to compose the chemical milieu in the cuvette and to inject the enzyme solution. Capillary C7 serves as air outlet during injection and as water inlet during the cleaning cycles; C8 is used to drain fluid from the
enzyme solution was made by adding 150 mg NAD+ (β-nicotinamide adenine dinucleotide, lyophilized free acid, 93–96%; ICN Biomedicals, Costa Mesa, CA) and 180 μl of a suspension of mitochondrial malate dehydrogenase from porcine heart (in 70% saturated ammonium sulfate; ICN Biomedicals) to 50 ml MOPS buffer. During the experiments all solutions were kept at room temperature (≈18°C). The protocol was derived from one originally developed for a prototype enzymatic signal-processor [33, 34].

For a reaction first the chemical milieu is mixed by pumping 2.0 ml of reaction buffer into the cuvette. The mixture is composed according to the specification provided by the scouting algorithm from reaction buffer containing MgCl₂, reaction buffer containing citrate, and plain reaction buffer. The latter serves to compensate the volume differences associated with varying amounts of MgCl₂ or citrate in the reaction milieu to a constant volume of 2.0 ml. Absorption measurements are taken every 1.5 s. After a delay of 25 s that allows air bubbles to escape, the measurement frequency is increased to 2.5 Hz and the reaction is initiated by injecting 0.3 ml of enzyme/co-enzyme solution into the cuvette. After 90 s the measurement frequency is reduced to 0.5 Hz. Progress of the reaction is recorded for a total of 360 s. Subsequently the cuvette is drained and cleaning cycles clear the cuvette for the next reaction. With the above parameters a full assay including clean up takes about 10½ minutes.

C. Implementation

All parts of the software, i.e., computer control of the fluidics, the evolutionary algorithm and the user interface, were implemented in Tcl/Tk [18]. Tcl/Tk is convenient for programming the concurrent operation of the servo action and polling of the photometer. With the experiments being the dominant time factor, performance on a 400 MHz Linux PC was found to be sufficient. The software currently enables three different types of experiments. Single experiments take a specification for a milieu mixture and perform the specified reaction, displaying and recording its progress. Systematic experiments measure reactions on a regular grid in the concentration space and are used as controls. Finally, the scouting mode explores the concentration space through evolutionary search.

Figure 5: Configuration of the computer controlled fluidics. See text for details; from [32]. © Zauner.

cuvette. All absorption measurements were made at a wavelength of λ = 339 nm; the photometer was controlled by our software through an RS232 port.

B. Protocol

Reaction medium is mixed from a buffer containing the substrate (L-malate) and milieu components (MgCl₂ and citrate), and a buffered enzyme/co-enzyme solution used to initiate the reaction. The reaction buffer contains 144 mM glycine and 6 mM L-malate, adjusted with NaOH to pH 9.8. Citrate solution is prepared by adding to an aliquote of reaction buffer 350 mM citrate and adjusting the pH with NaOH to 9.8. MgCl₂ solution is prepared by adding 350 mM MgCl₂ to an aliquote of reaction buffer and readjusting to pH 9.8. For the enzyme/co-enzyme solution, first a stock solution of 30 mM MOPS (3-[N-Morpholino]propanesulfonic acid) adjusted with NaOH to pH 6.9 was prepared. The enzyme/co-

Figure 6: Flow cuvette. Eight glass capillaries were positioned with hot glue and sealed with silicon aquarium sealant (A). This head was then attached to a polystyrene cuvette of 1 cm light path (B). From [32]. © Zauner.
D. Data Analysis

On-line visualization during the experiments was implemented by interfacing the scouting software with gnuplot [28]. Fitting of the time development of absorption measurements for representation in the experience database was also performed through gnuplot. For this fitting the measured absorption curves were aligned by first subtracting the base line (calculated by averaging the 5 measurements taken before starting the enzyme injection), and second, taking the end of the enzyme injection as the starting point (t = 0) of the reaction. To fit a new series of measurements, the parameters (cₜ₋₀ of equation 1) of the closest observation available in the experience database are used to seed the curve fitting algorithm. Off-line analysis and visualization of the enzymatic response was performed with the Ferret program [12]. To render the time and concentration dependence of the absorption from the observations collected by the scouting algorithm, we first used linear interpolation to fill the gaps between measured milieu combinations and then applied a Welch filter to reduce noise.

ACKNOWLEDGMENTS

The reported material is based upon work supported by NASA under Grant No. NCC2-1189.

REFERENCES


