Efficient discovery and optimization of complex high-throughput experiments

James N. Cawse\textsuperscript{a,b,*,} Gianluca Gazzola\textsuperscript{a}, Norman Packard\textsuperscript{a,c}

\textsuperscript{a}ProtoLife Inc., 57 Post St. \#513, San Francisco, CA 94014, USA
\textsuperscript{b}Cawse and Effect LLC, 132 Kittredge Rd., Pittsfield, MA 01201, USA
\textsuperscript{c}European Center for Living Technology, Calle del Clero 2940, 30124 Venice, Italy

\textbf{A B S T R A C T}

As the pace of experimentation in materials science and catalysis has increased, experimental tactics and strategies have had to adapt to meet the demands of goals of experimentalists, and the spaces they explore. This pace has increased from runs/year to runs/day and sometimes to runs/minute in high-throughput experimentation. Although much of this capacity is used to simply speed up conventional experimental designs, the leading-edge application is discovery of low-probability, high-value occurrences (hits) by searching extensive, complex experimental spaces. Conventional design of experiments (DoE) is not capable of dealing with these issues. Instead, more advanced experimental tactics and strategies must be implemented. After introducing the elements that make an experimental campaign complex, here we present a novel statistical model-based evolutionary experimental strategy and apply it to the optimization of a family of artificial complex systems. With our experiments, we show that such a strategy may significantly reduce the experimental effort required for finding the optima compared to other state-of-the-art evolutionary strategies.

\section{1. Introduction}

Since the modern era of high-throughput chemistry and materials science began in the 1990s, experimenters have innovated novel experimental designs to attack the new, larger problems that became accessible with large numbers of experiments \cite{1,2}. These included novelties such as fractal designs \cite{3}, “unpeeled” dense mixture designs \cite{4}, edge-sharing \cite{5} and split-plot designs \cite{6}. Although enthusiasm for the more elaborate of these designs has abated somewhat as the field has matured \cite{7}, problems still appear in which the experimental space is large and complex. These have the issues of:

\begin{itemize}
  \item High-dimensional experimental spaces (many system constituents and experimental parameters);
  \item Complex constraints on the independent variables;
  \item Synergies, or beneficial nonlinear interactions between system constituents;
  \item Unpredictable behavior and the inability to derive experimental results from basic chemical and physical laws.
\end{itemize}

Addressing these complex problems requires a strategic approach to experimentation. Strategy (the overall approach) is distinct from tactics (the conduct of an individual experiment). The designs mentioned above can be tactical elements in the strategic plan, but one-design-at-a-time experimentation should be considered as obsolete as one-factor-at-a-time experimentation.

In this paper we first formulate the problem, defining experimental spaces with both qualitative and quantitative variables, and with a response surface representing experimental measurement. We then examine both tactical and strategic approaches to the problem, including experimental space sampling via a genetic algorithm (GA), and we introduce a novel form of evolutionary design of experiments (Evo-DoE), which combines statistical modeling and stochastic sampling. We finally illustrate Evo-DoE with a numerical case study, comparing it with a GA.
Table 1
Factors for one-step DPC synthesis (1 × 15 × 15 × 15 × 9 × 4 × 3 × 3 × 3 = 3,280,500 potential experiments).

<table>
<thead>
<tr>
<th>Factors in typical screening experiment</th>
<th>No. of levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Principal metal catalyst</td>
<td>1</td>
</tr>
<tr>
<td>Inorganic co-catalysts</td>
<td>15</td>
</tr>
<tr>
<td>Metal ligands</td>
<td>15</td>
</tr>
<tr>
<td>Organic co-catalysts</td>
<td>15</td>
</tr>
<tr>
<td>Anion</td>
<td>9</td>
</tr>
<tr>
<td>Associated cation</td>
<td>4</td>
</tr>
<tr>
<td>Reaction time</td>
<td>3</td>
</tr>
<tr>
<td>Reaction temperature</td>
<td>3</td>
</tr>
<tr>
<td>Reaction pressure</td>
<td>3</td>
</tr>
</tbody>
</table>

experiments, or size of the experimental space, is obtained by multiplying the number of levels for each factor, both qualitative and quantitative. This may be reduced by system constraints (e.g., in mixture experiments). For each experiment, we assume that an experimental response may be measured, thus defining a response surface over the experimental space.

Experimenters have some flexibility in determining the experimental space, in that they must choose which variables will be varied in a particular experiment. Out of all possible controllable variables, the experimenter may choose to fix some and vary others. In fact, some potentially controllable variables, such as intermediates in a complex reaction, may be unknown at the beginning of an experimental program. The experimenter’s choice of experimental space may depend on his or her perception of the curvature or irregularity of the response surface. We will take as an operational definition of experimental space, for a given experimental program, the space of variables the experimenter chooses to explore, i.e., the factors chosen to vary.

The GE Global Research team determined just such an experimental space in their classic study of the carbonylation of phenol to diphenyl carbonate (DPC) (Table 1) [8]. The immensity of this space and the value of the potential catalyst justified a massive effort exceeding 22,000 experimental runs, subsampling the full experimental space in various non-systematic ways [9].

In approaching experimentation in these spaces, there are two critical questions that must be asked. First, is the space “too large” for the tools at hand? Is the number of factors, levels, and interactions too great, and is the response surface too irregular? If so, the experiment must be pruned using the best chemical knowledge available. Second, if the resources available appear to meet the challenge, what is the best strategy for attacking the project at hand? Some of these questions are addressable only with preliminary experimentation, e.g., to gain knowledge of the response surface.

1.2. Experimental response surface topology

One way to help our thinking in this area is to consider some idealized response surfaces and the consequences of the shapes of these surfaces. There are two extremes in this area. One is a surface over a space consisting of multilevel entirely qualitative factors. The other is a surface over a space with purely quantitative factors.

1.2.1. Qualitative spaces

The ultimate difficulty for a discovery program is the true “needle in the haystack,” illustrated in Fig. 1 as a response surface over a 2-dimensional experimental space consisting of two qualitative factors with many levels each. In such a system, the optimum result appears only when all of the factors are at their precisely correct levels. No experimental design can find this peak more reliably than random search. Mathematically, this system can be represented by a function

\[ Y = \begin{cases} 1 + \varepsilon & \text{if } (\mathbf{x} = \mathbf{c}) \\ \varepsilon & \text{if } (\mathbf{x} \neq \mathbf{c}) \end{cases} \]

where \( \mathbf{x} \) is a point in the space, \( \mathbf{c} \) is the vector of optimal levels, and \( \varepsilon \) a normally distributed random variable with zero mean.

For a rational discovery program of any kind to succeed, there must be lower-dimensional “ridges” which can be followed to the optimum (Fig. 2). The equation then can become

\[ Y = \frac{1}{n} \sum_{i=1}^{n} \delta(x_i) + \varepsilon, \quad \delta(x_i) = \begin{cases} 1 & \text{if } x_i = c_i \\ 0 & \text{if } x_i \neq c_i \end{cases} \]

where \( n \) is the dimensionality of the space, and \( x_i \) and \( c_i \) are the level of the \( i \)-th component (factor) in a point \( \mathbf{x} \) in the space and in the vector \( \mathbf{c} \) of optimal levels, respectively. Here the lower-order ridges
give the clues that identify the factors and levels that participate in the optimum.

We can imagine searching for such optima using full factorial and optimal designs. Full factorial designs suffer greatly from the “curse of dimensionality” (Fig. 3(a)). Optimal designs are far more efficient (Fig. 3(b)), but they require making a reasonably accurate guess as to the level of interactions, and they too will strain the laboratory resources when the number of factors, levels, and interactions gets too high.

Yet another way of considering this problem is to set a reasonable budget of runs for an experiment and examine the size of the spaces that can be examined with different approaches. This is given for a ∼1000-run budget in Fig. 4. This is divided into three classes: systems in which every possible qualitative combination can be tested (“brute force”); those that can be handled by relatively conventional DoE; and those that can be handled with computer-generated optimal designs (D- or I-optimal) which search for second-order or higher interactions.

A real problem of the qualitative type is the discovery of synergetic drug combinations. In [10], e.g., Lehár et al. describe a screening experiment aimed at discovering 2-way synergetic effects in a 90-drug library, for inhibiting the proliferation of HTC116 cancer cells. The response of each of the \( C(90,2) = 4095 \) combinations was an aggregate value extracted from a \( 6 \times 6 \) inhibition matrix of combinations of different amounts of the two drugs, thus requiring a total of \( 4095 \times 36 = 147,420 \) measurements. Now, if we consider the possibility of extending this search to drug triples, the number of possible qualitative combinations that need to be tested becomes \( C(90,3) = 125,580 \), corresponding to 4,520,880 response measurements. If the experimental effort scales linearly with the number of combinations examined, then screening for drug triples would take roughly 30 times as long as screening for drug pairs. If, e.g., screening pairs took one month of effort, then screening for triples would take two and a half years of effort. This would be a daunting task even for the resources of a major pharmaceutical company. However, there are only 4095 possible 2-way combinations, and if the “ridge” hypothesis is correct, three of those could be ridges to a 3-way peak. The key to the problem will be use of a search algorithm that can detect those ridges.

1.2.2. Quantitative spaces

In quantitative systems, the optima will almost always be regions of some width rather than a single point. In these systems, the question will be whether the distance between any two points interrogating the space is on the same scale as the width of the optimum region, which can be modelled as a Gaussian, described by the following equation:

\[
Y = \exp \left( \frac{\sum(x - c)^2}{s} \right)
\]

where \( c \) and \( s \) are the mean and the standard deviation of the Gaussian, respectively. These types of peaks are shown in 2D and 3D in Figs. 5 and 6.

We propose a sense of the needed “resolution” of a search design by considering the volume of that part of the optimum region that has a response greater than the overall system noise. An estimate of the necessary resolution is that volume divided by the total space volume, and an estimate of the number of points needed is the inverse of that ratio. This could be done, e.g., by taking a “range-finding” set of points and considering all experimental points that are 3 standard deviations over the average for the set as candidates directing the search toward an optimum.

---

If the peak is a 3-way A–B–C interaction, then there are three subsidiary 2-way interactions: A–B, A–C, and B–C.
1.3. Experimental space sampling tactics and strategies

When designing complex systems defined on very large, high-dimensional spaces, an evolutionary (or “adaptive”), iterative DoE strategy is usually preferred to a classic DoE tactic. Evolutionary DoE strategies abandon the idea of analyzing the entire space of combinations and all relationships between factors in only one experiment and with only one design. Instead, they iteratively build new designs as a function of the experimental results collected during the exploration. This allows the search to gradually converge on a limited optimal region of experimental space by drawing on clues gathered in the initial stages (or “generations”), sampling for observation only a tiny fraction of all possible points. Choosing the points of the first generation of an iterative program is a tactical issue that then feeds into the general strategy, which determines how points will be chosen in all following generations.

1.3.1. First generation sampling tactics

For sampling quantitative spaces, it is tempting to use a Cartesian design, in which all dimensions are divided equally and points are placed at every intersection. This is very simple to set up, but it has been shown to be far worse than the mathematically optimal method for sampling an unknown experimental region, which is a packing or covering lattice design [14]. This optimality is predicated on the assumption of an infinite space; for practical experimental spaces, lattices have fitting problems at the edges. Cartesian designs will also work in spaces that are composites of quantitative and multilevel qualitative factors. Lattices will not because they typically define points with non-integer coordinates.

For a real space, it is often preferable to use a stochastic method for generating a sampling set. As spaces grow more complex, fully deterministic designs like lattices become increasingly difficult to generate, because of system constraints and lack of good distance measures. In short, spaces become more irregular. This does not bother stochastic methods. There are several flavors of stochastic methods. A purely random design is by far the easiest to generate, and has the great advantage of being applicable to all types of spaces – quantitative, qualitative, and composite. True randomness, however, is actually somewhat “lumpy” (Abelson’s first law) [15] – points are surprisingly likely to be clustered, leaving relatively large unsampled gaps. Weighted random sampling techniques, in which the probability of a point being sampled is biased toward its distance from previously tested points, should be preferred [16].

There are several other stochastic sampling methods for quantitative factors in the more sophisticated statistics packages such as JMP [17]. The best of these for chemical experimentation appear to be:

- Sphere packing, which emphasizes the spread of points;
- Latin Hypercube, which is a compromise between spread of points and uniform spacing;
- I-optimal, which minimizes the average variance of prediction over the region of the data.

A comparison of some of these methods is given in Fig. 7, showing the distribution of minimum distances of 250 random points in a 4-dimensional square space divided in $10^9$ units in each dimension ($10^9$ points). The best methods are those with the least representation on the right hand (large gap) side, with Latin Hypercube the
Recent literature on design and optimization of experimental systems includes a number of reports in which GAs have been combined with artificial neural networks (ANNs) trained on experimental observations [32]. In some of these, the ANN is trained after several generations of a GA have identified a promising region. This is followed by virtual experimentation, e.g., for optimally adjusting some of GA parameters to the system at hand, using the ANN as artificial response function [33], or for studying how the efficiency of the evolutionary strategy could be improved using the ANN as a response predictor [34]. In other studies, a GA generates a large number of candidate points; then the ANN, trained on all previous experimental observations, selects the most promising points for the next generation of experiments [35]. Our contribution to this research thread is a new strategy of evolutionary design of experiments (Evo-DoE) that combines prediction of fruitful points from nonlinear regression models of previous experimental observations with stochastic exploration of the experimental space based on weighted random sampling. Thus, the ANN or other nonlinear model is, on the one hand, intimately embedded in the evolutionary process at each generation and, on the other hand, a tool for point selection which does not rely on a GA. Elsewhere, we report on the application of this approach to the problem of optimizing the cargo capacity of a complex liposomal drug formulation [36]. Here, we describe the Evo-DoE procedure in detail and assess its performance on a family of artificial problems. We then benchmark such performance with that of a standard GA.

2. Materials and methods

2.1. Artificial response surface and experimental space

To compare the performances of Evo-DoE and the GA, we ran three batches of simulations, in which the two experimental strategies were applied to the problem of optimizing three different complex artificial response surfaces. In previous work [24], we partially resolved the topology of the response landscape of a real high-dimensional mixture amphiphile system. The experimental data provided evidence of a multi-peaked response surface, perturbed by several components of experimental noise. This information was incorporated here in the artificial response surfaces, which were designed by superimposing several stochastically perturbed Gaussian peaks on a simplex-lattice space [37].

The response $F_X$ of a point $x$ in the $q$-dimensional simplex-lattice space was defined as:

$$F_X = \max_k(G(x)), \text{ with } G(x) = a_k \exp \left( \sum_{i=1}^{q} \frac{(x_i - c_{ik})^2}{s_k} \right)$$

where $x_i$ is the level of the $i$th component (factor) of $x$; $k$ is the Gaussian index, varying from 1 to $p$; $c_{ik}$, $a_k$, and $s_k$ are respectively: the coordinate of the mean $c_{ik}$ in the $i$th dimension, the height at $c_{ik}$, and the standard deviation of the $k$th Gaussian. The formula then calculates the heights of the $p$ Gaussians at $x$, and the highest of these is associated to $x$ as response value. A given number $k \leq q$ of dimensions are supposed to be “neutral” with respect to one or more of the $p$ Gaussians, as if in such dimensions the distance of $x$ from $c_{ik}$ were null independently of where $x$ is located. The response of a combination, with respect to each Gaussian, therefore depends on the interaction of $l$ variables.

The measured response $F^m_X$ of $x$ is simulated as:

$$F^m_X = F_X + N(0, rF_X) + N(0, t)$$

The “true” response $F_X$ is then perturbed by two error components: one proportional to and one independent from $F_X$. The two quantities are sampled from a normal distribution with zero mean and variance $rF_X$ and $t$, respectively.

---

2 These results will probably vary with sample size and dimensionality.
Table 2

<table>
<thead>
<tr>
<th>Height (a)</th>
<th>Optimal levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment 1</td>
<td>c₁ = 10, c₂ = 10</td>
</tr>
<tr>
<td>3</td>
<td>c₁ = 4, c₂ = 16</td>
</tr>
<tr>
<td>8</td>
<td>c₁ = 15, c₂ = 5</td>
</tr>
<tr>
<td>Experiment 2</td>
<td>c₁ = 4, c₂ = 16</td>
</tr>
<tr>
<td>3</td>
<td>c₁ = 6, c₂ = 1, c₃ = 9</td>
</tr>
<tr>
<td>8</td>
<td>c₁ = 10, c₂ = 3, c₃ = 3, c₄ = 3, c₅ = 1</td>
</tr>
<tr>
<td>Experiment 3</td>
<td>c₁ = 1, c₂ = 1, c₃ = 2, c₄ = 2, c₅ = 2, c₁₀ = 2, c₁₂ = 3, c₁₃ = 3, c₁₄ = 3, c₁₅ = 3, c₁₆ = 2, c₁₇ = 1</td>
</tr>
<tr>
<td>3</td>
<td>c₁ = 5, c₂ = 3, c₁₀ = 2, c₁₁ = 1, c₁₂ = 2, c₁₃ = 1, c₁₄ = 1, c₁₅ = 1, c₁₆ = 1, c₁₇ = 1</td>
</tr>
<tr>
<td>8</td>
<td>c₁ = 4, c₂ = 1, c₁₁ = 2, c₁₈ = 1, c₁₉ = 3, c₂₀ = 2, c₂₁ = 2, c₂₂ = 1, c₂₃ = 1, c₂₄ = 1</td>
</tr>
</tbody>
</table>

The initial random sample, and the random samples chosen at steps 4–6 of every generation, were drawn from a probability distribution biased toward the unsampled regions of the space, in order to favor global exploration of the experimental space. Specifically, the probability of a point being sampled was proportional to the Euclidean distance between that point and the closest already sampled point. The probability distribution was recomputed after sampling every point. A previously sampled point could not be resampled.

The hill-climbing algorithm mentioned at step 4 started with the prediction, based on the model learned at step 3, of the response level of all nearest neighbors of the jth initial point, and the following selection of the neighbor with the highest predicted response. This procedure was then iteratively repeated, selecting the predicted best nearest neighbor of the point selected at the previous iteration, and so on, and it was stopped when either: (a) all nearest neighbors had lower predicted response than the point selected at the previous iteration; (b) the selected point had already been selected in previous hill-climbing runs starting from any of the first J – 1 initial points.

The balance between the model-based and random points was determined in such a way that most of the experimental effort was aimed at intelligent sampling, and only a small fraction of it at pure exploration. The specific figures, however, should be considered arbitrary.

The models used here were feed-forward, single hidden-layer ANNs [38] (learned with back-propagation using net in the R language after standardizing all inputs and normalizing the output to the [0,1] interval), with 100 inputs and 1 output.

Each ANN was constructed with particular metaparameter values (weight decay constant and number of hidden-layer nodes). At step 2 of the Evo-DoE cycle, the model’s metaparameters were selected using a bagging process [39], repeating the model learning on 20 different data sets, each being a different random sample of 80% of the observed points, and 10 times on each data set. Each configuration of metaparameters was then assigned a quality measure, calculated as the mean linear correlation between the remaining 20% observations and the corresponding predictions over all the repeats.

2.2.2. Benchmark genetic algorithm

The genetic algorithm used here is a variation of the one described in [24]. It started with a tactic consisting of a set of 384 randomly selected points. The nth generation of the genetic algorithm consisted of the following steps:

(1) Measure the experimental response of the nth set of points;
(2) Choose 168 parent points via tournament selection as described below;
(3) Generate 1 mutant child from every parent, as described below;
(4) Divide parent points in pairs and generate 2 child points from every pair by crossover, as explained below;
(5) Add 48 more randomly selected untried points to the nth set of points.

---

3 The exact number of points is arbitrary and will vary with the experimental circumstances. High-throughput experimentation commonly makes use of 384-well plates. One such plate would then fit one entire generation of the evolutionary strategy.

4 If less than 336 points fall within the top decile, the top 336 points should be selected.
Fig. 8. Expected value of mean and maximum response for Evo-DoE (circles) and for the GA (triangles), in experiments 1 (a, b), 2 (c, d), 3 (e, f). Both statistics were calculated at every generation on all observed points (which also include those observed at the current generation). The time series stop at the generation at which the global optimum was found in all replicates. By definition, the maximum response would have remained the same if the simulations were run further (dotted line), while it is not possible to accurately predict what the dynamics of the mean response would have been. Error bars show 95% confidence intervals, estimated on 10 replicates. Confidence intervals smaller than the size of the symbols are not shown.

The initial random sample, and the random sample chosen at step 5 of every generation, were drawn from a uniform probability distribution.

Parent points were selected with the so-called “tournament” criterion [40], as follows. Let \( b \) the number of already observed points, \( d \) the number of parent points to select and \( e \) an integer such that \( 1 < e < b - d + 2 \). The tournament criterion requires sampling without replacement for \( d \) times of \( e \) of the \( b \) points, and the selection as parents of those with the highest observed response in the \( d \) samples. In the experiments we set \( d = e = 168 \).

Mutant child points were generated following this scheme: one integer number \( u \) was randomly sampled from the values 1, 2, 3, 4, 5, with probability 5/15, 4/15, 3/15, 2/15, and 1/15 respectively; \( u \) genes (factors) were sampled with replacement among those present in positive quantity; from those factors, the \( u \) previously sampled units were removed and added to \( u \) other factors, sampled with replacement among the remaining ones. Any mutant point coincident with a previously sampled point was discarded and regenerated.

Parent points were coupled following the order in which they were sampled (the first with the second, the third with the fourth, etc.), and each couple generated 2 crossed child points following this scheme:

(a) One integer number \( v \) was sampled from a uniform distribution in the [1,19] interval: \( v \) corresponded to the number of volume units, indexed with the factor they referred to, provided by the first parent. The remaining \( w \) units (with \( w = 20 - v \)) were provided by the second parent. The combination of these \( v + w \) units corresponded to the first child created with crossover.
for all other (neutral) factors. If a factor is relevant for more than one Gaussian, its time series is shown with the type of line associated to the highest Gaussian. E.g., the time of a factor can vary from 0 (in all points the level is equal to 0) to 336 (in all points the level is >0). The time series are shown: with diamonds for the factors relevant to the composed by the 2-, 3- and 5-dimensional Gaussians (experi-
cations). The dynamics of the mean response is, as expected, well correlated with that of the maximum response.

3. Results and discussion

The simulations were stopped at the generation at which the best point, corresponding to the mean of the highest Gaussian, was found. Every experiment was repeated 10 times. To evaluate the behavior of the two evolutionary experimental strategies we used two different performance measures. The first consists in the computation of the maximum response reached on average in the different replicates, after testing a given number of points. We can interpret this measure as the expected value of the maximum response, for a given level of experimental effort. Analogously, the second measure calculates the expected value of the mean response as a function of the number of observed points. Both measures are calculated at every generation on all observed points and not only on those observed at the current generation.

The results of the first set of experiments show that both strategies optimize the response surface very quickly, considering the vastness of the experimental space (Fig. 8). The exploration operated by the Evo-DoE approach, however, is significantly more efficient, for any level of experimental effort. After only one generation the maximum response determined by Evo-DoE reaches a level of roughly 4.5, which implies that the strategy has already found the highest Gaussian and has selected points that are located in the neighborhood of its center (by construction, no combination that does not belong to the counter-image of this Gaussian can have a response higher than 3). To reach the same level of maximum response, the GA requires an experimental effort roughly 5 times larger (note that in the second generation the maximum response is almost unvaried and close to 0). Evo-DoE identifies the best combination within 8 generations (corresponding to 3072 observations), as opposed to 18 required by the GA (6012 observations). The dynamics of the mean response is, as expected, well correlated with that of the maximum response.

The relative performances of the two algorithms are qualita-
tively confirmed in the two other experiments. However, it is particularly interesting to note that the GA is affected much more significantly by the increase in the complexity level of the problem, compared to Evo-DoE. The optimization of the response surface composed by the 2-, 3- and 5-dimensional Gaussians (experiment 2) by the GA requires 35 generations (corresponding to 13,440 observations), as opposed to 10 required by Evo-DoE (3840 observations). As for the response surface composed by the 10-dimensional Gaussians (experiment 3), the necessary experimental effort increases to 64 generations (corresponding to 24,576 observations) for the GA and only 11 for Evo-DoE (4224).

Fig. 9 allows us to visualize approximately the path followed by the two algorithms in the exploration of the experimental space, up to the selection of the best combination. The figure shows, for every generation and for each of the 100 factors, the time series of the number of combinations that contain such factors at a level greater than zero, for a representative replicate of experiment 3. The 48 random points are excluded from this calculation in order to isolate the behavior of the only “intelligent” components of the strategies (ANN/parent selection, crossover and mutation). The path followed by the GA shows rather gradual changes of direction, with relatively small differences in the population of points between successive generations. The path followed by Evo-DoE, on the other hand, changes direction much more rapidly, suggesting that the topology of the response surface predicted by the ANNs tends to vary substantially from one generation to another (at least in terms of where the best decile of hill-climbed points is located).

Fig. 9(a) shows that the Evo-DoE has identified the six most important of the ten participating factors by generation 6, and the additional four appear by generation 9. The GA (Fig. 9(b)) had identified only three by generation 6, three more by generation 14, and the next four by generation 35. It must also be emphasized that the GA erroneously selects points with positive levels in the non-relevant factors much more frequently than Evo-DoE.

4. Conclusion

Understanding the structure of the experimental space is criti-
cal to the planning process of a high-throughput experiment. Only when the experimental space is small enough to search exhaustively, or when interactions among system components are weak enough that the space can be simplified, are traditional DoE meth-
ods an appropriate optimization tool. In the presence of complex, synergistic systems defined on large, high-dimensional experi-
mental spaces, more sophisticated optimization techniques are required. The project must be viewed with a holistic mindset. There should be a lively discussion of the interrelationships among factors constrains exploration of the relevant ones.
• The factorial space (defined as all possible combinations of the controllable factors, with constraints included);
• The response surface (the anticipated degrees of interaction or irregularity);
• The understanding of the underlying chemistry and physics and;
• The high-throughput experimental system (its capacity and limitations).

Only when all these have been considered (and reconsidered as the experiment progresses) should the tactical elements be addressed. The tactics also should be considered as part of an iterative strategy. This is an update of classical DoE thinking, as in Box's famous dictum to limit the first experiment to no more than 25% of the planned effort [41].

Tactical designs such as sphere packing will serve as tools, but evolutionary strategies are required to find optima in truly complex systems. Stochastic methods such as GAs are effective, but we have demonstrated in our simulations that incorporation of nonlinear modeling and predictive power (Evo-DoE) can enhance the efficiency with which the global system optimum is successfully detected and climbed. In a live experimental context, one may never be sure of reaching a global maximum of the response without exhaustive sampling of the experimental space, but Evo-DoE is the most effective available strategy for finding a good result with a given experimental budget.

Acknowledgements

Thanks to the Prodolife team, especially Mark Bedau, Martin Hanczyc, and Andrew Buchanian, for valuable discussion, and Emily Parke for thorough and sensitive editing.

References


