

Experiment planning for combinatorial materials discovery

L. HARMON

Striatum Incorporated, 8703 Webster Hills Road, Dexter, MI 48130, USA

E-mail: lharmon@striatum.com

The introduction of combinatorial methods into materials discovery and optimization presents new challenges for experiment planning. The need for new or adapted experimental design approaches in combinatorial materials discovery stems from the dramatic expansion in numbers of materials and experimental variables that can be considered using combinatorial approaches. This paper presents an overview of some of the experimental design strategies being developed and used for combinatorial discovery and characterization of materials. Parallels between materials versus drug discovery are drawn and various modes of combinatorial experimentation are outlined: mapping, screening and optimization. Specific methods for incorporation of prior knowledge into experimental design include statistical design of experiments, diversity techniques, hierarchical and hybrid approaches such as neural networks, and search techniques like Monte Carlo optimization and genetic algorithms. © 2003 Kluwer Academic Publishers

1. Introduction

Combinatorial methods have been successfully applied in the pharmaceutical industry to the discovery of small organic molecules, peptides, and proteins. These methods are now being applied to a wide variety of materials discovery and optimization applications [1–3], including polymers, coatings and biomaterials [4–9], heterogeneous catalysts [10–12] and homogeneous catalysts [13, 14].

The U.S. National Institute of Standards and Technology (NIST) held a workshop in July of 2001 to develop a technology roadmap for combinatorial methods [15]. The workshop was part of the broader chemical industry Vision 2020 process [16] and addressed the discovery of new materials and processes, among other applications. The Combinatorial Methods workshop brought together representatives from industries, government labs and universities to identify key needs and challenges in combinatorial methods; experimental design was one of the informatics needs highlighted in the resulting technology road map.

The introduction of combinatorial methods into drug and material discovery presents new challenges for experiment planning. The need for new or adapted experimental design approaches in combinatorial materials discovery stems from the dramatic expansion in numbers of materials and experimental variables that can be considered using combinatorial approaches. The space of possibilities for new small molecule drugs is enormous; conservative estimates suggest that there may be 10^{40} candidate drugs [17] based on simple constraints such as molecular size and constituent elements. The domain of potential materials is even larger, with an effectively infinite number of candidates even for a single material such as a polymeric coating or heterogeneous

catalyst, given a continuum of composition and process variables to consider.

Table I outlines some of the parallels between combinatorial discovery of new drugs and new materials. While the goal of combinatorial drug discovery is primarily to discover or optimize a new molecular entity, combinatorial materials discovery entails discovery and optimization of either a new molecule or composition, frequently associated with a set of process conditions for both synthesis and application. This extension to process parameters, and the broad range of composition and structural variables within some material classes, greatly extends the dimensionality of the experiment space to be considered in materials discovery. Both drug and materials discovery must take into account the possibility of highly nonlinear effects, in which properties may change rapidly or discontinuously as a result of small changes in composition, structure or conditions. Nevertheless, many approaches to the design of combinatorial libraries for pharmaceutical applications assume that similarity of structure leads to similarity of biological behavior [18]. For inorganic materials and polymers, such assumptions break down in the presence of phase changes and other physical phenomena.

A variety of methods are being applied to the design of combinatorial experiments, including statistical design of experiments, diversity methods and search strategies. Each has its place and the selection of method depends upon the goals of the experiment and the intended use of the resulting data. This paper presents an overview of some of the design strategies being developed and used for combinatorial discovery. Different modes of combinatorial experimentation are outlined: mapping, screening and optimization. In each mode, prior knowledge can be used to reduce the experimental space and increase experimental effectiveness.

TABLE I Selected characteristics of combinatorial drug and materials discovery

Characteristics	Drug discovery	Materials discovery
Goals	Discover or optimize discrete molecules or proteins	Discover or optimize discrete molecules or compositions and processes
Dimensionality of search	Composition and structural variables	Composition and structural variables synthesis and process parameters
Highly non-linear effects	Single substitutions or small structural changes dramatically alter biological effects	Small changes in composition, structure or treatment can dramatically alter material properties
Measurements	Imprecise and potentially noisy	Complex and potentially noisy

2. Background

2.1. Combinatorial experiment cycle

The combinatorial or high throughput materials discovery process is frequently depicted as a cycle by those practicing in both industry (cf. [9, 12, 19]) and in academia (cf. [20]). An example is shown in Fig. 1. The cycle represents an iterative discovery process, like that developed for high throughput drug discovery [21]. In this model, the cycle is initiated with the design of an experiment, which may consist simply of sample compositions to be synthesized in a library, or may extend further to treatment conditions and process parameters. Samples are fabricated or synthesized in libraries based on the experiment design and their properties of interest are measured in the laboratory. After laboratory data are analyzed and interpreted, qualitative or quantitative models may be generated to provide insight and guidance for the next set of experiments.

2.2. Modes of combinatorial experimentation

There are commonalities among all combinatorial experiments, including: the desire for high throughput, achieved by parallel or rapid serial synthesis and testing, and the systematic investigation of a wide range of parameters in a single experiment. It is nevertheless useful to distinguish among types or *modes* of experiments, classified by their primary objective. Three modes of combinatorial experiments can be characterized as follows:

- *Mapping*. In mapping experiments, the principal goal is to develop quantitative or qualitative knowledge of relationships among material or molecular properties of interest and experimental parameters such as composition, structure and synthesis conditions. These relationships may be obtained without necessarily searching for “hits” or lead compounds or materials. The results of mapping studies can be used as input to guide subsequent screening or optimization experiments.
- *Screening*. The purpose of screening experiments is to identify hits (molecules or materials) for follow-up testing. Screening may also be directed at identifying small regions of materials space with promising properties. While the goal of screening differs from mapping, screening can be used to accumulate knowledge of material properties over time.
- *Optimization*. As the name implies, experiments are designed to refine and optimize material or molecular properties by carrying out intensive studies in the vicinity of a lead compound or material. Though similar in principal to mapping experiments, the scope of optimization experiments is dramatically reduced by varying a subset of parameters and/or by narrowing parameter ranges.

The terms *screening* and *optimization* stem from the original application of combinatorial methods to drug discovery. Analogous terms used in the materials domain include *primary* and *secondary* screening,

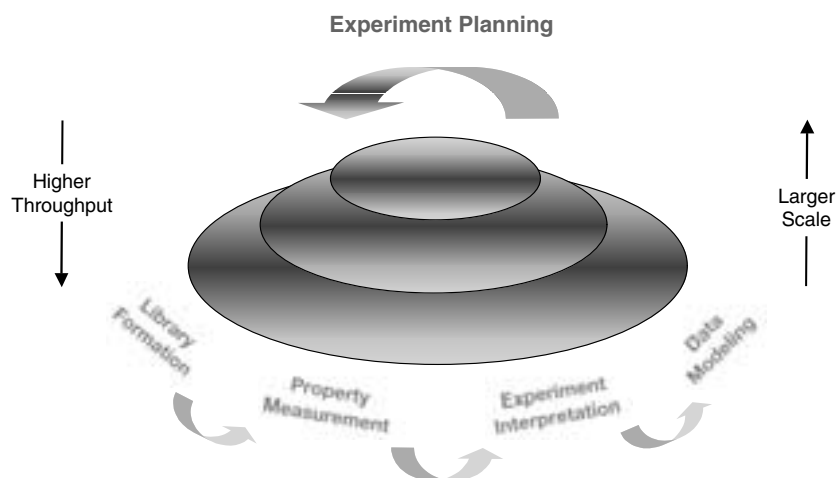


Figure 1 The combinatorial experiment cycle (after [37]).

discovery and *focus* screening, or *Stage 1* and *Stage 2* testing. While screening and optimization are sometimes distinguished by throughput, this obscures the differences in overall experimental objectives.

Mapping studies have received relatively little attention and are often subsumed under screening because both address large search spaces. In drug discovery, GlaxoSmithKline has used the term “progressible hit” to designate hits whose mechanism of action is known and which are accompanied by some limited structure-activity information to guide optimization testing [17]. However, the value of combinatorial methods to systematically probe physical phenomena is becoming increasingly recognized (cf. [6, 7, 12]).

3. Planning for combinatorial experiments

The cycle of combinatorial experiments and the large experimental spaces introduce a different perspective on the overall objectives of experiment planning or design. In classical statistical experiment design (DOE) [22], the fundamental objective is *hypothesis testing*. An experiment is designed to generate statistically reliable conclusions to specific questions. This strategy is particularly suited to domains that are known sufficiently well that appropriate questions can be formed. In contrast, combinatorial methods are often employed for the express purpose of exploring novel and unknown domains. Here, the objective might better be expressed as *hypothesis generation*. Combinatorial experiments can be used to gather enough knowledge of some region in experimental space to develop meaningful and testable questions for future, more focused experiments.

3.1. Objectives of combinatorial experiment planning

The purpose of combinatorial experiment planning is to direct a discovery process toward its particular objectives, whether in mapping, screening or optimization modes. This purpose is achieved through the design of individual experiments as well as through the design of successive experiments in a series. The planning must be done in a way that extracts the maximum amount of information from each experiment. Combinatorial experiment planning must inevitably balance two opposing forces: making use of existing knowledge to maximize experimental efficiency and preserving the opportunity for truly novel discoveries.

The miniaturization and parallelization of materials synthesis, novel library formation methods such as composition spreads [7, 23], together with parallel and/or high throughput analytical methods [24, 25] are dramatically increasing the rate at which materials can be created and tested. Nonetheless, experimental capacity is still limited relative to the space of possibilities by such factors as the cost of equipment and other resources, the availability of investigators and technicians, laboratory facilities for follow-up analysis and testing, and allotted calendar time. Therefore,

effective experimental strategies are required to navigate this vast experimental space, including alternatives to statistical design of experiments for experiment planning.

As outlined above, mapping, screening and optimization experiments have different primary goals and therefore different requirements in experiment planning. Because mapping experiments are intended to discover trends and relationships, it is essential that sampling strategies be consistent with the intended analytical tools. These tools in turn must be able to handle the discontinuities and nonlinearities that are likely to arise in materials applications. Some nonlinear multivariate regression and pattern recognition methods are proving effective for this purpose.

In screening experiments, the critical factor is the ability to efficiently search a large parameter space and confidently detect “hits” or active regions. However, the additional need to build up knowledge of how material properties behave within this space means that screening experiments should also, if possible, be designed to support quantitative analysis of the dependence of the measured response to experiment parameters. Finally, optimization studies require fine-grained sampling of a local neighborhood around a specific material or molecule. Within such neighborhoods, data analysis and modeling may be simplified by the presumption of local continuity.

3.2. Prior knowledge in combinatorial experiments

Prior knowledge in many forms is available to guide and constrain combinatorial experiments. At the most basic level, the chemistry and physics of the domain of interest impose limits on compositions and conditions that can be explored. For example, solubility limits, known phase transitions, monomer compatibilities in polymers can all be used to reduce the experimental space. At the other extreme, the realities of downstream limits on processing or manufacturing conditions impose very practical limits on what is realistically worth exploring. Bem *et al.* [26] have also discussed the importance of incorporating laboratory constraints into experiment designs for efficiency.

Prior experiments, whether traditional or combinatorial, provide another class of information that can be exploited. Known analogues or exemplars may provide starting points for new investigations. Predictions from functional relationships, such as quantitative structure-property relations (QSPR) or other empirical data models, may be also be used to identify regions of interest and to identify important variables affecting desired properties. In the development of such relationships and models, the material property (or properties) of interest are modeled as a function of descriptive molecular or material features (*descriptors*), which form the independent variables. An important use of prior knowledge is in selecting or designing the descriptors that go into such models in order to capture as much relevant chemical and physical information about the domain as possible.

4. Methods for combinatorial experiment planning

Methods that are currently in use or under development for combinatorial experiment planning fall into four general classes:

- Traditional statistical DOE approaches such as factorial or fractional factorial designs are intended to generate statistically reliable conclusions from a limited number of experiments.
- “Diversity” methods endeavor to represent or span a space of interest using various measures to characterize ensembles of experimental samples.
- “Search” methods attempt to intelligently navigate through the experiment space in a succession of experiments.
- Hierarchical or hybrid methods combine techniques to develop a series of experiments with increasing focus.

After a brief discussion of statistical DOE, this paper will focus on diversity, search and hierarchical strategies.

4.1. Statistical DOE

Recent reviews address the role of statistical DOE in combinatorial materials discovery [27] and drug discovery [28]. Statistical DOE methods build on a large body of theoretical and practical work and are very powerful tools in the right applications. Some of the limitations of traditional DOE have been summarized in Bem *et al.* [26]. Statistical DOE is motivated by a need to generate statistically reliable conclusions from a minimum number of experiments. As such, it provides plans for systematic sampling and testing that will allow quantitative assessment of the effects of selected independent variables or factors on the dependent variable of interest.

Because they are build on a large body of theoretical and practical research, sophisticated design methods and corresponding data analysis methods have been developed. This leads to powerful tools for certain applications. In general, statistical methods are best suited for problems with relatively small numbers of independent variables and frequently make simplifying assumptions about the domain, such as linearity or low-order polynomial relationships. As a consequence, they are perhaps most useful for combinatorial optimization studies, where such assumptions may be locally valid even if they do not hold over a broad range of parameters.

4.2. Diversity methods

Diversity methods are used to sample an existing collection of molecules or materials, or a pre-defined experimental space. Each method is intended to optimize some property of an *ensemble* of samples, as opposed to a property of any individual sample. All require a measure of similarity or distance within the experimental space, based on molecular or material descriptors, or

experimental and/or process parameters. Prior knowledge may be used to define the experimental space of interest, to select appropriate descriptors, and to select exemplars.

Extensive work has gone into the creation of descriptors of small organic molecules, such as drug candidates, monomers for polymer synthesis and homogeneous catalysts. These descriptors may represent physical-chemical properties, composition, topological indices or 2- and 3-dimensional structural features, among others or may be derived from combinations of other features (cf. [29]). Descriptors may also be based on reagents and other inputs to synthesis or on synthesis conditions themselves. Distance metrics or similarity measures must be matched to the set of descriptors or other representation of the experimental space. With large numbers of descriptors, distance calculations become very computationally intensive. Prior knowledge can be used to select descriptors and distance metrics. Rule- or knowledge-based filters may also be applied to eliminate individual samples or regions from the design (cf. [18]).

Two classes of diversity methods are discussed here: *grid-* or *cell-based* methods and *coverage* methods, as illustrated in Figs 2 through 4. These figures employ a dramatically simplified experimental space, which can be characterized by two descriptors, D1 and D2, with individual samples depicted in the resulting two-dimensional space. Fig. 2 illustrates the use of grid-based methods for sampling from library of discrete possibilities. The method is equally applicable to sampling from a continuum of material candidates described, for example, by composition variables. The two- (or *n*-) dimensional space is divided into cells based on descriptor values. The goal is to design an experiment using samples that are most representative of the resulting cells. The results of two variant strategies are illustrated in cells where they would differ: (1) samples selected to be closest to the center of each cell (blue circles); or (2) samples selected to represent the actual distribution of points within the cell (circles). Grid-based methods are useful in mapping and screening experiments but the number of cells to be sampled can become prohibitively large with large numbers of variables [27] unless sophisticated analyses are performed to transform the space and reduce the dimensionality [29].

Coverage designs identify sample sets that either: (1) represent a set of existing exemplars; or (2) span

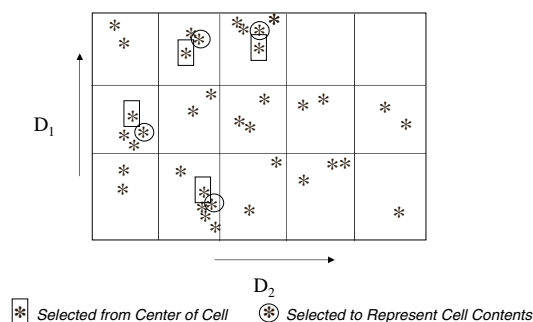


Figure 2 Illustration of grid-based design methods.

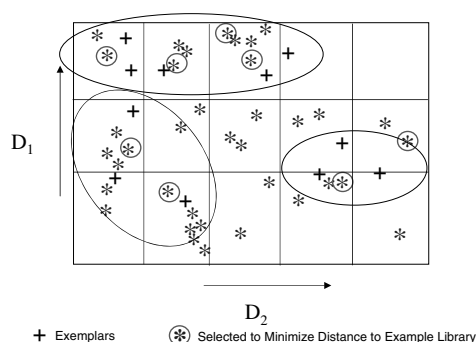


Figure 3 Illustration of coverage approach to represent a set of exemplars.

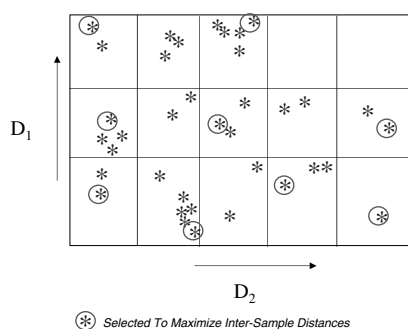


Figure 4 Illustration of coverage approach to span an existing library.

an existing collection or designated experimental space [18]. In Fig. 3, a set of exemplars is shown as crosses while discrete potential candidates are shown as stars. The exemplars may themselves be characterized by an algorithm such as clustering, and then new materials selected to represent each cluster, as indicated by the circles. If the materials can be designed from a continuum, then the cluster structure of the exemplars can be more closely represented in the library.

Fig. 4 illustrates the coverage of the collection as a whole or of the experimental space defined by a set of examples or constraints [30]. Here the goal is to span the space by maximizing inter-sample distances, with the resulting points circled. An optimal coverage algorithm has been applied to heterogeneous catalyst discovery as part of a hierarchical strategy described in Section 4.4 [26].

Although Figs 2 through 4 are notional, they serve to illustrate the differences in design characteristics that result from these different objectives. Working with any of these methods in high-dimensional spaces and with large numbers of descriptors can be computationally expensive for discrete entities (single molecules), but approaches have been developed to optimize or reduce the computations needed [18, 31, 32].

4.3. Search methods

Search approaches to experiment planning start with a set of experiment points and specify an algorithmic strategy for determining the next set of experiments based on results from the preceding set. “Results” may be obtained through *a priori* computational prediction,

from experiments, or from mathematical models derived from experimental data. If the figure of merit can be predicted computationally, then fitness can be evaluated *in silico*. In the materials domain, the complexity of computing material properties means that the search is conducted through a succession of experiments. Two strategies that have been applied to materials discovery are the genetic algorithm and Monte Carlo. An important advantage of these methods is that they make no assumptions about the domain—the response surface may be arbitrarily complex without diminishing their effectiveness. Prior knowledge may be used to select the initial experiment points, or they may be selected at random from within some space of possibilities. Knowledge may also be exploited in defining the method to derive one population from its predecessors.

The genetic algorithm is an optimization algorithm inspired by the operation of natural selection on populations of organisms. An initial set of samples (*population*) is selected, either at random or based on prior information. Each member of the population is described by a set of attributes (analogous to genes) which may be compositional or structural parameters, process and treatment variables. The members of the population are evaluated for their *fitness*, e.g., the figure of merit or material property of interest. Members with high fitness values are selected for *reproduction*. A set of genetic changes is defined which are allowed to occur during breeding between pairs. These changes typically consist of modifications to a single gene (analogous to biological mutations) or exchange of one or more genes between pairs (analogous to crossover events in cell division). The result is a new population that can be evaluated and the cycle is repeated. The genetic algorithm has been applied to the development of heterogeneous catalysts, using the relative stoichiometries of metal components as the “genes” [33, 34].

Monte Carlo is a stochastic optimization method widely used for computer modeling of physical phenomena. Monte Carlo methods are similar to genetic algorithms in that experimental variables are perturbed by a random process. However, fitness is evaluated using a thermodynamic analogy, in which an “energy” function is compared to an effective “temperature”. This procedure retains some samples that lower the fitness function and, with proper tuning of parameters, can escape local maxima in the response surface. Monte Carlo methods have been applied to the design of small molecule libraries and to simulated material discovery experiments. Using a Random Phase Volume Model to simulate a material with a highly nonlinear response surface, Falcioni and Deem [35] compared Monte Carlo variants to other search methods such as grid-based and random searches. In these simulations, Monte Carlo was shown to be much more effective at maximizing the desired figure of merit (performance) within a fixed number of experimental iterations. In an earlier comparison of Monte Carlo to the genetic algorithm for small molecule library design [36], it was concluded that Monte Carlo methods lead to more diverse libraries, while the genetic algorithm may find the best single candidate.

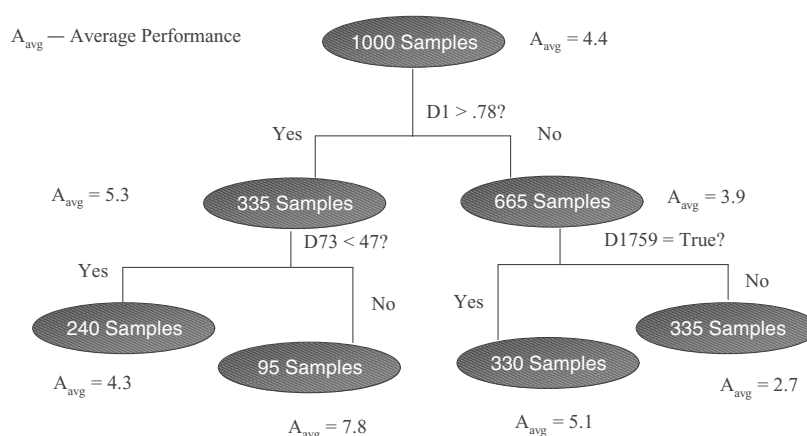


Figure 5 Illustration of a recursive partitioning model. Here, A is an experimentally-measured performance parameter and the Dx are independent variables (descriptors) used to partition data in the model.

4.4. Hierarchical and hybrid approaches

In hierarchical approaches, one method such as a diversity or statistical algorithm, may be employed to design individual experiments, while a sequence of experiments is treated as a search. One approach is to use each round of experiments to build or refine a model of relationships within the domain. The models can then be used to predict outcomes for unmeasured samples and/or conditions, which form the basis for designing the next experiment. Hybrid strategies therefore employ multiple methods in the experiment planning process.

A variety of algorithms may be applied to model combinatorial materials data. To be suitable, methods must be capable of handling the number of independent variables to be considered and the types of nonlinear behavior mentioned above. Examples of methods that meet these criteria are neural networks, ridge regression and recursive partitioning [37]. Hastie *et al.* [40] recently reviewed a variety of statistical learning methods. Of these, neural networks and recursive partitioning are beginning to be applied to high throughput materials research.

4.4.1. Neural networks

Neural networks are nonlinear statistical models that relate some number of input features to one or more output responses. Neural networks can model either quantitative outcomes (regression) or categorical outcomes (classification). A number of methods fall under the umbrella of neural network, but the most common is the single hidden layer back-propagation network. In the single hidden layer neural network, a set of intermediate features (the “hidden layer”) is created from linear combinations of the inputs. The output(s) are then modeled as functions of linear combinations of the intermediate features. The parameters of a neural network model are trained using examples in which both inputs and outputs are known. The network model can then be applied to features of new examples to predict the corresponding outputs.

4.4.2. Recursive partitioning

Recursive partitioning is used to develop decision tree models of the relationships between a set of input features (independent variables) and an output (depend-

ent) variable. Like neural networks, recursive partitioning can be used to perform either regression or classification, depending upon the output variable type. Recursive partitioning iteratively partitions or splits the experimental data into two or more subsets, based on the value of the independent variable and values of the independent variables that provide the best discrimination among the subsets. Splitting is continued until some stopping criterion is met. The result is a tree model of the data, illustrated schematically in Fig. 5. In the model, each node corresponds to a split and the terminal nodes (or leaves) correspond to the final partitioning of the data. Tree models are developed through training on known examples, for instance from a previous experiment. The model only incorporates independent variables that have a significant effect on the dependent variable. Once developed, the model can be applied to unknown cases to predict the output variable. Tree models may be combined by a variety of methods to produce more reliable predictions. Advantages of recursive partitioning over neural networks include the ability to handle larger numbers of independent variables and more interpretable models.

4.4.3. Hybrid approach

An example of a hierarchical hybrid approach is described in Bem *et al.* [26]. The method, termed *optimal coverage*, distributes points within a region of interest described by constraints supplied by the researcher. The points are selected so as to maximize diversity in the space of independent experimental variables. Regions of interest for the next experiment are automatically derived from the results of the previous experiment, using either a greedy algorithm or by modeling the response surface using multivariate nonlinear regression. The approach was shown to converge rapidly on the global performance maximum, using a complex simulated response surface with several local maxima.

5. Conclusions

Combinatorial experiment planning requires new approaches to experiment design. The selection of method depends upon the mode of the experiment and any underlying assumptions about the domain. The

experiment design strategy must be compatible with the intended method of experimental data analysis and modeling to support the cycle of experiments illustrated in Fig. 1. Prior knowledge can be incorporated into experiment planning in a variety of ways to reduce the problem space and to leverage experience. At the same time, the dynamic tension between exploiting that knowledge and seeking truly novel discoveries will remain a constant feature of combinatorial experimentation.

References

1. R. DAGANI, *Chem. Eng. News* **77** (1999) 51.
2. *Idem.*, *ibid.* **78** (2000) 66.
3. R. MALHOTRA (ed.), "Combinatorial Materials Development," ACS Symposium Series 814 (American Chemical Society Publications, Washington, D.C., 2002).
4. S. BROCCINI, K. JAMES, V. TANGPASUTHADOL and J. KOHN, *J. Amer. Chem. Soc.* **119** (1997) 4553.
5. *Idem.*, *J. Biomed. Mater. Res.* **42** (1998) 66.
6. J. C. MEREDITH, A. KARIM and E. J. AMIS, *Macromolecules* **33** (2000) 5760.
7. J. C. MEREDITH, A. P. SMITH, A. KARIM and E. J. AMIS, *ibid.* **33** (2000) 9747.
8. R. A. POTYRAILO, D. R. OLSON, G. MEDFORD and M. J. BRENNAN, *Anal. Chem.* **74** (2002) 5676.
9. H. BACH, Society of Plastics Engineers ANTEC Presentation in New Technology Forum, May 7, 2002.
10. D. E. AKPORIAYE, I. M. DAHL, A. KARLLSON and R. WENDELBO, *Angew. Chem.* **110** (1999) 629.
11. J. SCHEIDTMANN, P. A. WEISS and W. F. MAIER, *Appl. Catal. A: Gen.* **222** (2001) 79.
12. J. S. HOLMGREN, D. BEM, R. GILLESPIE, M. BRICKER, G. LEWIS, R. MURRAY, A. SACTLER, R. WILLIS, D. AKPORIAYE, A. KARLSSON, M. PLASSEN and R. WENDELBO, "COMBI2002—Combinatorial Approaches for New Materials Discovery" (The Knowledge Foundation, Boston, MA, 2002).
13. J. A. LOCH and R. H. CRABTREE, *Pure Appl. Chem.* **73** (2001) 119.
14. A. HAGEMeyer, B. JANDELEIT, Y. LIU, M. P. DAMODARA, H. W. TURNER, A. F. VOLPE and W. H. WEINBERG, *Appl. Catal. A: Gen.* **221** (2001) 23.
15. NIST, Technology Roadmap for Combinatorial Methods. National Institute of Standards and Technology. September, 2001.
16. American Chemical Society (1996). Technology Vision 2020: The U.S. Chemical Industry. Available at <http://www.acs.org>.
17. J. J. VALLER and D. GREEN, *Drug Discovery Today* **5** (2000) 286.
18. R. E. HIGGS, K. G. BEMIS, I. A. WATSON and J. H. WIKEL, *J. Chem. Inf. Comput. Sci.* **37** (1997) 861.
19. J. NEWSAM, "COMBI2002—Combinatorial Approaches for New Materials Discovery" (The Knowledge Foundation, Boston, MA, 2002).
20. A. HOLZWARTH, P. DENTON, H. ZANTHOFF and C. MIRADATOS, *Catal. Today* **67** (2001) 309.
21. D. K. AGRAFIOTIS, R. F. BONE, F. R. SALEMME and R. M. SOLL US6421612. Assigned to 3-Dimensional Pharmaceuticals Inc. (2002).
22. R. A. FISHER, "The Design of Experiments" (Oliver and Boyd, Edinburgh, 1935).
23. L. F. SCHNEEMEYER, R. B. VAN DOVER, C. K. MADSEN and C. L. CLAYPOOL, in "Combinatorial Materials Development, edited by R. Malhotra (ACS Symposium Series 814 (American Chemical Society Publications, Washington, D.C., 2002).
24. C. M. SNIVELY, G. OSKARSDOTTIR and J. LAUTERBACH, *Catal. Today* **67** (2001) 357.
25. R. A. POTYRAILO and R. J. MAY, *Rev. Sci. Instrum.* **73** (2002) 1277.
26. D. BEM, E. J. ERLANDSON, R. D. GILLESPIE, L. A. HARMON, S. G. SCHLOSSER and A. J. VAYDA, in "Experimental Design for Combinatorial and High Throughput Materials Development," edited by J. N. Cawse (John Wiley & Sons, 2003).
27. J. N. CAWSE, *Acc. Chem. Res.* **34** (2001) 213.
28. S. ROSE, *Drug Discovery Today* **7** (2002) 133.
29. D. J. CUMMINS, C. W. ANDREWS, J. A. BENTLEY and M. CORY, *J. Chem. Inf. Comput. Sci.* **36** (1996) 750.
30. R. TOBIAS, Space-Filling Experimental Designs for Combinatorial Chemistry, "COMBI2000—Experimental Strategy Workshop" (The Knowledge Foundation, Boston, MA, 2000).
31. D. K. AGRAFIOTIS and V. S. LOBANOV, *J. Chem. Inf. Comput. Sci.* **39** (1999) 51.
32. D. K. AGRAFIOTIS, *ibid.* **41** (2001) 159.
33. M. BAERNS, "COMBI2000—Combinatorial Approaches for New Materials Discovery" (The Knowledge Foundation, Boston, MA, 2000).
34. D. WOLF, O. V. BUYEVSKAYA and M. BAERNS, *Appl. Catal. A: Gen.* **200** (2000) 63.
35. M. FALCIONI and M. W. DEEM, *Phys. Rev. E* **61** (2000) 5948.
36. L. CHEN and M. W. DEEM, *J. Chem. Inf. Comput. Sci.* **41** (2001) 950.
37. L. A. HARMON, S. G. SCHLOSSER and A. J. VAYDA, in "Combinatorial Materials Development, edited by R. Malhotra, ACS Symposium Series 814 (American Chemical Society Publications, Washington, D.C., 2002).
38. D. BEM, C. BRATU, R. BROACH, G. LEWIS, C. MCGONEGAL, M. MILLER, J. MOSCOSO, R. MURRAY, A. RAICH, D. WU, D. AKPORIAYE, A. KARLSSON, M. PLASSEN and R. WENDELBO, "COMBI2000—Combinatorial Approaches for New Materials Discovery" (The Knowledge Foundation, Boston, MA, 2000).
39. M. BRICKER, K. VANDEN BUSSCHE, C. MCGONEGAL, A. KARLSSON, D. AKPORIAYE, I. DAHL and M. PLASSEN, "COMBI2000—Combinatorial Approaches for New Materials Discovery" (The Knowledge Foundation, Boston, MA, 2000).
40. T. HASTIE, R. TIBSHIRANI and J. FRIEDMAN, "The Elements of Statistical Learning: Data Mining, Inference and Prediction" (Springer-Verlag, New York, 2002).