What to do with Statistical Mechanics when the Figure of Merit Cannot be Calculated: Library Design for High-Throughput Materials Development

Michael W. Deem

Chemical Engineering Department
University of California
Los Angeles, CA 90095-1592, mwdeem@ucla.edu

ABSTRACT

By analogy with Monte Carlo algorithms, we discuss new strategies for design and redesign of libraries in high-throughput experimentation, or combinatorial chemistry. Several Monte Carlo methods are examined, including Metropolis, several types of biased schemes, and composite moves that include swapping or parallel tempering. Among them, the biased Monte Carlo schemes exhibit particularly high efficiency in locating optimal compounds. The Monte Carlo strategies are compared to a genetic algorithm approach. Although the best compounds identified by the genetic algorithm are comparable to those from the better Monte Carlo schemes, the diversity of favorable compounds identified is reduced. Applications to materials discovery, small molecule discovery, and templated materials synthesis are discussed.

Keywords: Monte Carlo, combinatorial chemistry, high-throughput experimentation

1 Materials Discovery

The goal of combinatorial materials discovery is to find compositions of matter that maximize a specific material property [2]–[4], such as superconductivity [5], magnetoresistance [6], luminescence [7]–[9], ligand specificity [10], sensor response [11], or catalytic activity [3], [12]–[17]. This problem can be reformulated as one of searching a multi-dimensional space, with the material composition, impurity levels, and synthesis conditions as variables. The property to be optimized, the figure of merit, is generally an unknown function of the variables and can be measured only experimentally.

Present approaches to combinatorial library design and screening invariably perform a grid search in composition space, followed by a "steepest-ascent" maximization of the figure of merit. This procedure becomes inefficient in high-dimensional spaces or when the figure of merit is not a smooth function of the variables, and its use has limited most combinatorial chemistry experiments to ternary or quaternary compounds.

We have suggested new experimental protocols for searching the space of variables in combinatorial chemistry, exploiting an analogy between combinatorial materials discovery and Monte Carlo computer modeling methods [1]. Several variables can be manipulated in order to seek the material with the optimal figure of merit. Material composition is certainly a variable. But also, film thickness [18] and deposition method [19] are variables for materials made in thin film form. The processing history, such as temperature, pressure, pH, and atmospheric composition, is a variable. The guest composition or impurity level can greatly affect the figure of merit [17]. In addition, the "crystallinity" of the material can affect the observed figure of merit [18]. Finally, the method of nucleation or synthesis may affect the phase or morphology of the material and so affect the figure of merit [20].

The experimental challenges in combinatorial chemistry appear to lie mainly in the screening methods and in the technology for the creation of the libraries. The theoretical challenges, on the other hand, appear to lie mainly in the library design and redesign strategies. We have addressed this second question via an analogy with Monte Carlo computer simulation, and we have introduced the Random Phase Volume Model to compare various strategies. We find the multiple-round, Monte Carlo protocols to be especially effective on the more difficult systems with larger numbers of composition and non-composition variables.

An efficient implementation of the search strategy is feasible with existing library creation technology. Moreover "closing the loop" between library design and redesign is achievable with the same database technology currently used to track and record the data from combinatorial chemistry experiments. These multiple-round protocols, when combined with appropriate robotic controls, should allow the practical application of combinatorial chemistry to more complex and interesting systems.

2 Small-Molecule Discovery

We have also developed new strategies for design and redesign of small molecule libraries in high-throughput experimentation, or combinatorial chemistry [21]. Several Monte Carlo methods were examined, including Metropolis, three types of biased schemes, and composite moves that include swapping or parallel tempering.

Among them, the biased Monte Carlo schemes exhibited particularly high efficiency in locating optimal compounds. The Monte Carlo strategies were compared to a genetic algorithm approach. Although the best compounds identified by the genetic algorithm are comparable to those from the better Monte Carlo schemes, the diversity of favorable compounds identified is reduced by roughly 60%.

Monte Carlo is a natural paradigm for experimental design of multi-round combinatorial chemistry, or high-throughput, experiments. A criticism of high-throughput experimentation has been its mechanical structure and lack of incorporation of *a priori* knowledge. As we have shown, a biased Monte Carlo approach handily allows the incorporation of *a priori* knowledge. Indeed, our results reveal that biased Monte Carlo schemes greatly improve the chances of locating optimal compounds. For moderately complex libraries, the bias can be determined equally well by experimental or theoretical means. Although the compounds identified from a traditional genetic algorithm were comparable to those from the better Monte Carlo schemes, the diversity of identified molecules was dramatically decreased in the genetic approach. Genetic algorithms, therefore, are less suitable when the list of good molecules is further winnowed by a secondary screen, a tertiary screen, patentability considerations, lack of side effects, or other concerns. Presumably, as the complexity of the library is increased, these composite moves will prove more useful for the more challenging figures of merit. Although we often chose the initial library configurations at random, sophisticated initial library design strategies available in the literature can be used, and they would complement the multi-round library redesign strategies presented here.

3 Templated Materials Discovery

Finally, we have addressed how best to design and redesign high-throughput experiments for zeolite synthesis [22]. A model that relates materials function to chemical composition of the zeolite and the structure directing agent was introduced. Using this model, several Monte Carlo-like design protocols were evaluated. Multi-round protocols were found to be effective, and strategies that use *a priori* information about the structure-directing libraries were found to be the best.

High-throughput, or combinatorial, methods allow for simultaneous creation of a large number of structurally diverse and complex compounds, generalizing the traditional techniques of single compound synthesis. Monte Carlo methods have been proposed and shown to be efficient methods for library design and re-design in both material discovery [1] and small molecule design [21]. Material discovery deals with continuous variables, such as composition variables and non-composition variables. Small molecule design deals with discrete variables, such as the identities of template and ligand. For templated zeolite synthesis, we have both the continuous zeolite composition and non-composition variables and the discrete identity variables of the component parts of the organic structure-directing molecules. All of these variables affect the final zeolite material in an interrelated way.

As in previous studies, multi-pass Monte Carlo methods work better than do single-pass protocols. Sophisticated biased Monte Carlo schemes are highly efficient and much better than simple Metropolis Monte Carlo. Parallel tempering is the best method for complicated systems with 5 or more framework chemical compositional variables.

Acknowledgment

This research was supported by the National Science Foundation through grant number CTS–0114263.

REFERENCES

[1] M. Falcioni and M. W. Deem, "Library Design in Combinatorial Chemistry by Monte Carlo Methods," *Phys. Rev. E* **61** 5948-5952, 2000.

[2] M. C. Pirrung, Chem. Rev. **97**, 473 (1997).

[3] W. H. Weinberg, B. Jandeleit, K. Self, and H. Turner, Curr. Opin. Chem. Bio. **3**, 104 (1998).

[4] E. W. McFarland and W. H. Weinberg, TIBTECH **17**, 107 (1999).

[5] X.-D. Xiang, X. Sun, G. Briceño, Y. Lou, K.-A. Wang, H. Chang, W. G. Wallace-Freedman, S.-W. Chang, and P. G. Schultz, Science **268**, 1738 (1995).

[6] G. Briceño, H. Chang, X. Sun, P. G. Schultz, and X.-D. Xiang, Science **270**, 273 (1995).

[7] E. Danielson, J. H. Golden, E. W. McFarland, C. M. Reaves, W. H. Weinberg, and X. D. Wu, Nature **389**, 944 (1997).

[8] E. Danielson, M. Devenney, D. M. Giaquinta, J. H. Golden, R. C. Haushalter, E. W. McFarland, D. M. Poojary, C. M. Reaves, W. H. Weinberg, and X. D. Wu, Science **279**, 837 (1988).

[9] J. Wang, Y. Yoo, C. Gao, I. Takeuchi, X. Sun, H. Chang, X.-D. Xiang, and P. G. Schultz, Science **279**, 1712 (1998).

[10] M. B. Francis, T. F. Jamison, and E. N. Jacobsen, Curr. Opin. Chem. Biol. **2**, 422 (1998).

[11] T. A. Dickinson and D. R. Walt, Anal. Chem. **69**, 3413 (1997).

[12] F. M. Menger, A. V. Eliseev, and V. A. Migulin, J. Org. Chem. **60**, 6666 (1995).

[13] K. Burgess, H.-J. Lim, A. M. Porte, and G. A. Sulikowski, Angew. Chem. Int. Ed. **35**, 220 (1996).

[14] B. M. Cole, K. D. Shimizu, C. A. Krueger, J. P.

A. Harrity, M. L. Snapper, and A. H. Hoveyda, Angew. Chem. Int. Ed. **35**, 1668 (1996).
[15] D. E. Akporiaye, I. M. Dahl, A. Karlsson, and R. Wendelbo, Angew. Chem. Int. Ed. **37**, 609 (1998).
[16] E. Reddington, A. Sapienza, B. Gurau, R. Viswanathan, S. Sarangapani, E. S. Smotkin, and T. E. Mallouk, Science **280**, 1735 (1998).
[17] P. Cong, R. D. Doolen, Q. Fan, D. M. Giaquinta, S. Guan, E. W. McFarland, D. M. Poojary, K. Self, H. W. Turber, and W. H. Weinberg, Angew. Chem. Int. Ed. **38**, 484 (1999).
[18] R. B. van Dover, L. F. Schneemeyer, and R. M. Fleming, Nature **392**, 162 (1998).
[19] T. Novet, D. C. Johnson, and L. Fister, Adv. Chem. Ser. **245**, 425 (1995).
[20] S. I. Zones, Y. Nakagawa, G. S. Lee, C. Y. Chen, and L. T. Yuen, Micropor. Mesopor. Mat. **21**, 199 (1998).
[21] L. Chen and M. W. Deem, "Monte Carlo Methods for Small Molecule High-Throughput Experimentation," *J. Chem. Inf. Comput. Sci.* **41** 950-957, 2001.
[22] L. Chen and M. W. Deem, "Strategies for High-Throughput, Templated Zeolite Synthesis," *Mol. Phys.*, **100**, 2002.