Reviews

From Data to Knowledge: Chemical Data Management, Data Mining, and Modeling in Polymer Science

Nico Adams and Ulrich S. Schubert*

Laboratory of Macromolecular Chemistry and Nanoscience, Eindhoven University of Technology and Dutch Polymer Institute (DPI), P.O. Box 513, 5600 MB Eindhoven, The Netherlands

Received June 26, 2003

1. Introduction

In the modern academic and industrial environment, which is characterized by continuously shortening lifetimes of knowledge and products,[1−3] the advent of combinatorial chemistry and high-throughput experimentation has profoundly changed the area of compound discovery and characterization. Both synthesis and analysis can be carried out on much reduced time scales, and sample throughput is usually high.[4−7] One of the consequences of the increase in "discovery activity" is an explosion of experimental data that needs to be organized, administered, stored, and evaluated. This is particularly true in the area of polymer chemistry, where the number of parameters that can be varied during synthesis, formulation, and processing (e.g., monomers, initiators, monomer/initiator ratio, concentrations, temperatures, pressures, additives, stabilizers, etc.) is extremely large.[8] Additionally, there is an extensive amount of characterization and screening data, originating from both classical polymer analysis (T_g, T_m, M_n, M_w, polydispersities) and other materials analytical techniques (conductivity, elasticity, hardness, blend formulations, etc.). The need for computational tools has given rise to the field of "cheminformatics", which has two main functions, namely, the administration of data and the aiding of data comprehension (i.e., data mining and modeling).

Whereas sophisticated cheminformatics systems already exist in the area of medicinal chemistry and pharmacology, similar "matinformatics" tools are only beginning to be developed for materials science. The terms cheminformatics and data handling have somewhat diffuse definitions and encompass aspects ranging from the mechanics of data handling, storage, and searching to data mining and library design tools. This contribution will attempt to address recent developments in all of these areas, with particular consideration of polymer science applications.

2. Data Collection, Administration and Handling

2.1. Databases and Representation of Polymeric Structures. One of the crucial factors determining the success of a combinatorial discovery program is the availability of a database that allows for the storage of structure and reaction information, as well as formulation, characterization, and screening data. A number of companies have developed database solutions for the storage and administration of chemical information (such as the data cartridges Auspyx by Tripos[9] or DayCart by Daylight Chemical Information Systems,[10] as well as the Accord suite by Accelrys,[11] and, on a somewhat lower level, ChemFinder by CambridgeSoft[12]). Such databases are relational in the sense that they allow for the association of structural information with, for example, property as well as characterization and/or screening data. Moreover, there is usually some built-in functionality that allows structure filtering. Whereas methods for the representation of molecular structure have been well-established and documented for small molecules [two representatives are the SMILES strings developed by Weininger[13] and the SYBYL line notation (SLN)[14]], the representation of polymeric structures, as well as the associated database searching, poses special problems. While it is possible to define a small molecule as a unique entity with a clearly defined structure and properties, polymers are rather ill-defined statistical composites, in terms of both structure (e.g., chain length, tacticity, monomer segments) and composition (monomer content, blends, additives), as well as properties. Whereas the 2D structure of benzene, for example, can be exhaustively defined by the string

C[1]H:CH:CH:CH:CH:CH:@1

in the SYBYL line notation, the same is not possible for polymeric structures. In this case, SLN uses a more "categorical" encoding system—an ethylene/propylene copolymer consisting of 40% ethylene and 60% propylene, for example, would be described using the more general string

Ethylene.Propylene<type=polymer;Ratio_1=0.4;
\Ratio_2=0.6>

Ethylene and Propylene are used as "macroatoms"—shorthand notations for groups of atoms that can be expanded into full structure descriptions.[14]

* To whom correspondence should be addressed. Fax: 0031/(0)40/2474186. E-mail: u.s.schubert@tue.nl.

10.1021/cc034021b CCC: $27.50
Published on Web 12/19/2003

DuPont developed a hierarchical classification scheme in which polymers are first divided into three categories: (a) prescribed monomer condensation (PMC), (b) actual starting monomer (ASM), and (c) structural repeating unit (SRU).[15] PMC polymers are subsequently classified as belonging to one of 15 subclasses. To avoid duplicate entries, the polymers are described by stylized monomer component names and formulae. Poly(ethylene terephthalate), for example, can be prepared in a number of ways (from ethylene glycol and terephthalic acid or from ethylene glycol and terephthaloyl chloride, etc). In the DuPont system, however, the polymer is represented as if it had been prepared from ethylene glycol and terephthalic acid. ASM polymers are registered using the name and molecular formulas of the monomers used in the preparation, whereas SRU polymers are registered structurally using a series of atoms that, through repetition, make up all or part of the polymer backbone. Other classification systems, which have been developed by a number of commercial databases, were reviewed some years ago, and the reader is referred to the literature for further information (see, for example, refs 16−18).

2.2. Data Handling. Chemical databases on their own, however, act only as repositories of data. Although greatly simplifying routine procedures in terms of data administration and communication across laboratories, they are "stand-alone" systems in that they are only very loosely integrated into a typical laboratory workflow. However, only data that have been put into context gain added value and can potentially be converted into knowledge.

Particularly from a combinatorial point of view, optimal contextualization and synergisms could be achieved if database content could be combined with design-of-experiment (DoE), design-of-synthesis, and modeling software, which, in turn, could be integrated with both the control of synthetic and analytical equipment and the gathering, archiving, and evaluation of screening and analysis data. However, in a typical present-day laboratory, one encounters a situation in which chemical information is generally contained in a number of often incompatible formats, requiring a collection of conversion programs designed to make incompatible data formats compatible.[19]

An early example of an attempt to integrate disparate data formats and thus to utilize the synergisms that can arise from such integration, was the program suite CACTVS (Chemical Algorithms Construction, Threading and Verification System), reported in 1994 by Ihlenfeldt and co-workers.[19,20] CACTVS is essentially a chemical information handling tool that can handle any kind of arbitrarily complex information "by referring to an open set of descriptions of chemical data and data objects and [using] loadable modules to define and extend its capabilities instead of providing only a fixed set of functions and data it can operate upon."[20] CACTVS is modular and has a graphical worktop: a computational operation is represented by a module icon, and data can be piped from module to module. It is expandable in the sense that modules can be user-defined. Individual modules, containing data format and handling as well as property computation and analysis routines, are stored in a database, allowing the reuse of modules and exchange between different users. The routines are loaded only when they are actually ready to be executed, which allows the core program to remain relatively small and to grow with computational demand.[19] In its most recent incarnation, CACTVS is a distributed client/server system using a network of databases with property descriptors, computational tools, and visualization servers.[21]

Another early attempt was reported by Lindsey et al., who, in the early 1990s, began to develop both the hardware and software pertaining to an automated chemistry workstation.[22−25] By now, the instrumentation and software is in its third generation and includes experimental planning and scheduling modules. The planning functions seem to be similar to what one would expect from the software accompanying commercial synthesizers. However, the software does include modules for the factorial design of experiments, as well as decision-tree, multidirectional, Simplex, and successive-focused grid search algorithms. Moreover, the software allows for adaptive experimentation in some cases, which opens paths toward integrating experimental design and execution.[26−32]

A conceptually very similar but even more closely integrated solution has recently been developed by Avantium Technologies with their software suites VirtualLab (VL) and Data Analysis Package (DAP), which allow for the complete integration of design, execution, analysis, and evaluation of high-throughput experiments (Figure 1).[33] VirtualLab allows for the planning, in work-flow terms, of both classical benchtop experiments and high-throughput designs. Once the workflow for the actual experiment has thus been mapped, it can be combined with a "method editor", allowing for the generation of methods combining manual operations, automated synthesis, and classical as well as parallel characterization techniques into one experimental method. These methods can then be added to scheduling functionality, which, in turn, can be appended to automated equipment. As the experimental methods are being completed, commands can be sent by VirtualLab to the appropriate laboratory equipment to perform the required operations, or if a manual procedure has been defined, VirtualLab instructs users to perform it and to provide data inputs and results as required. The software is furthermore capable of autocollecting analytical data. When combining VirtualLab with DAP, data analysis can also be automated, allowing the chemist to focus on interpretation rather than manipulation.[33] Although VirtualLab was developed for applications in the biosciences, in principle, it is possible to expand the system with the specific requirements of a materials discovery program in mind.[34]

Accelrys developed the software package CombiMat, allowing the user to capture the process of library design, characterization, testing, and analysis.[11] The software is based on an Oracle database and associated application modules. When designing a synthetic procedure, CombiMat allows the user to define arbitrary processing steps with associated arbitrary parameters. These can then be used to create libraries of samples. Recipes for analysis and testing can be defined, and the results can be imported into the CombiMat database (Figure 2).

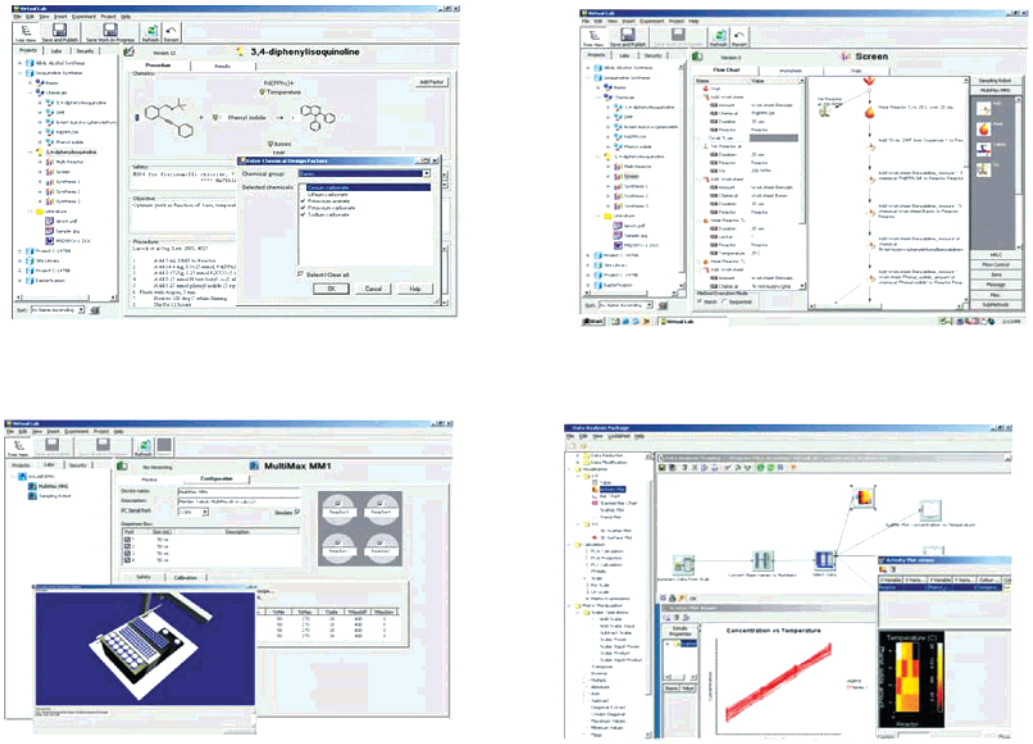

Figure 1. Screenshots of Avantium's VirtualLab and Data Analysis Package software (courtesy of Avantium, http://www.avantium.com).

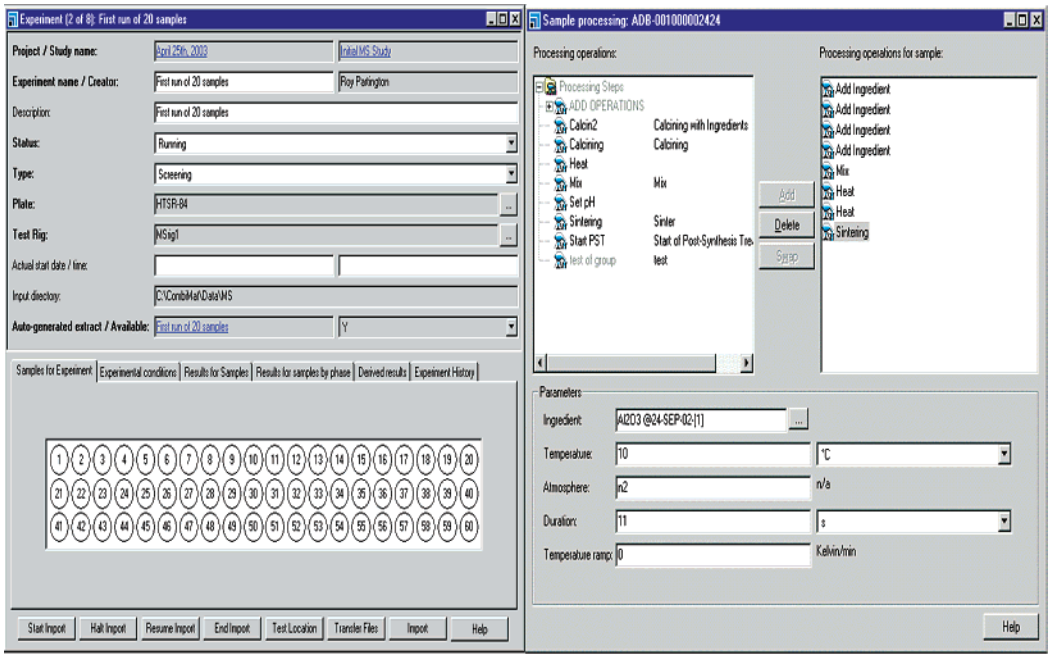

Figure 2. Screenshots of Accelrys' CombiMat software (courtesy of Accelrys Ltd., http://www.accelrys.com).

Creon Lab Control, in collaboration with BASF, recently announced the development of a combinatorial materials research data management system called Q-DIS/DOLPHIN.[35] This system combines modules automating the design of samples and experiments and controls experiment processing, as well as the collection and storage of results. Furthermore, it allows for the evaluation of data and the design of new experiments on the basis of evaluation results (adaptive experimentation). Both workflow and library design are accomplished in a graphical editor.

Apart from these published solutions, companies that are active in the area very often develop their own software to solve matinformatics problems. However, reports concerning the design and functionality of such software are usually sparse. Symyx, for example, has developed a suite of informatics tools, allowing the design of libraries (Library Design Studio); the execution of screening protocols (Impressionist); and the processing, storage, and handling of data in a central database. The company markets the software under the Renaissance trademark.[36]

3. Data Mining

High-throughput experimentation techniques generate large amounts of mostly multivariate data. Although this presents

a significant scientific opportunity, a caveat is in order: complex data sets obtained in this way often tempt researchers to establish relationships using only subsets of the data available, which, in turn, are often fraught with error.[37] Moreover, examples in medicinal chemistry have shown that smaller data sets, provided they are well designed, can yield the same quality of information as large data sets, but with huge gains in efficiency and speed.[38,39] Therefore, before any data mining is undertaken, indeed before any library is synthesized, DoE tools should be used when planning any high-throughput experimentation.[40−43] This will ensure that any information that can be derived from a set of experimentally acquired data is maximized, while keeping the number of necessary experiments to a minimum.

However, if experimentation, even if it has been optimized in this way, leads to large collections of data, the latter needs to be "mined" to derive knowledge. Data mining is an information extraction activity that allows one to uncover the knowledge and information contained in a database. This is usually accomplished using a combination of artificial intelligence, statistical analysis, modeling, and database technology. The process of data mining generally uncovers subtle relationships between sets of data and allows the researcher to make predictions about systems that were not themselves used in the derivation of the relationships. Hand recently published a general and very readable introduction to the field.[44] Modern statistical science has made available a whole armory of techniques that can be used to compute such relationships. When surveying the literature, a number of key methodologies seem to emerge, which will be reviewed briefly in the following section.

3.1. Mining by Visualization. The easiest manner in which data can be mined is by visual inspection, as general trends can be discovered quickly. A number of programs are available for data visualization, including, on a low level, Excel[45] and Origin[46] and, on a more sophisticated level, Spotfire,[47] OpenViz,[49] and Mathematica.[50] Spotfire in this context is very interesting, as both its database connectivity and its query features make it very suitable for data analysis.

3.2. Mining by Principal Component Analysis (PCA). Principal component analysis (PCA) is a statistical methodology that allows the dimensionality of information space to be reduced while concurrently keeping the loss of information to a minimum.[51−53] This approach assumes that descriptors are tightly correlated and that one can produce a set of *N* orthogonal descriptors from a set of *N* correlated descriptors. The uncorrelated descriptors are called "principal components" and are essentially a linear combination of the original descriptors. Coefficients (eigenvalues) indicate the weight of these descriptors. The first principal component (the one with the highest eigenvalue) accounts for most of the variance in the system, the second principal component for most of the remaining variance, and so forth. Once all of the eigenvalues have been calculated, only those principal components with variances above a certain threshold are retained.

Within polymer science, PCA methodologies have been used mainly for the characterization and classification of polymeric species. In 1997, Vanden Eynde and Bertrand

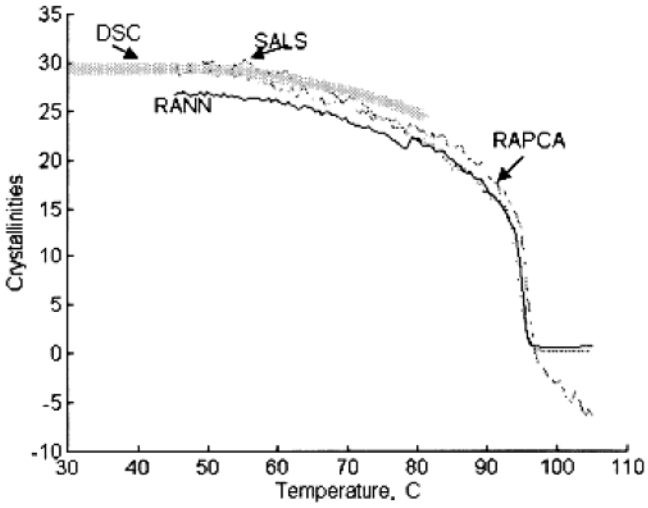

Figure 3. Comparison of crystallinities of polyethylene films determined experimentally by small-angle light scattering (SALS) and differential scanning calorimetry (DSC) and predicted by principal component analysis (RAPCA) and neural networks (RANNs).[55]

reported the use of PCA for the quantification of ToF−SIMS polystyrene spectra.[54] In a prior study, the authors had noticed that the molecular weights of their polystyrene samples had a significant impact on some secondary molecular ion intensities arising from both end groups and the macrochain. By applying PCA to their spectral data, the researchers were able to demonstrate that only one principal component was sufficient to account for the molecular weight variances and a second one allowed samples to be discriminated depending on the type of the butyl end group present. Furthermore, they showed that the correlation between the first principal component and the sample molecular weight allowed for the determination of the polymer molecular weights of an unknown monodisperse polystyrene sample from its secondary ion mass spectrum.[54]

Batur et al. compared the performance of both principal component analysis and artificial neural networks (see below) in the prediction of the crystallinities of low-density polyethylene films.[55] First, a training set was produced by heating a thin polymer film to 120 °C to achieve a completely amorphous state. Subsequently, the polymer sample was slowly cooled in steps of 2 °C, and a Raman spectrum was recorded for each temperature step. The obtained spectra were used as inputs for both principal component and neural network modeling, and the input data were correlated to a crystallinity value (linear regression between factor loadings and crystallinity value in the case of PCA) obtained from small-angle light scattering (SALS) experiments. The SALS methodology, in turn, was calibrated by using the differential scanning calorimetry (DSC) technique. Models developed in this way were subsequently tested on two data sets obtained using cooling rates of 3 and 0.3 °C/min. The authors found that both models performed well in the estimation of crystallinity, although the results estimated for the data set cooled at a rate of 3 °C/min more closely resembled the experimental data (see Figure 3). Furthermore, both models approached the final crystallinity determined after the end of cooling and measured by DSC. However, the researchers

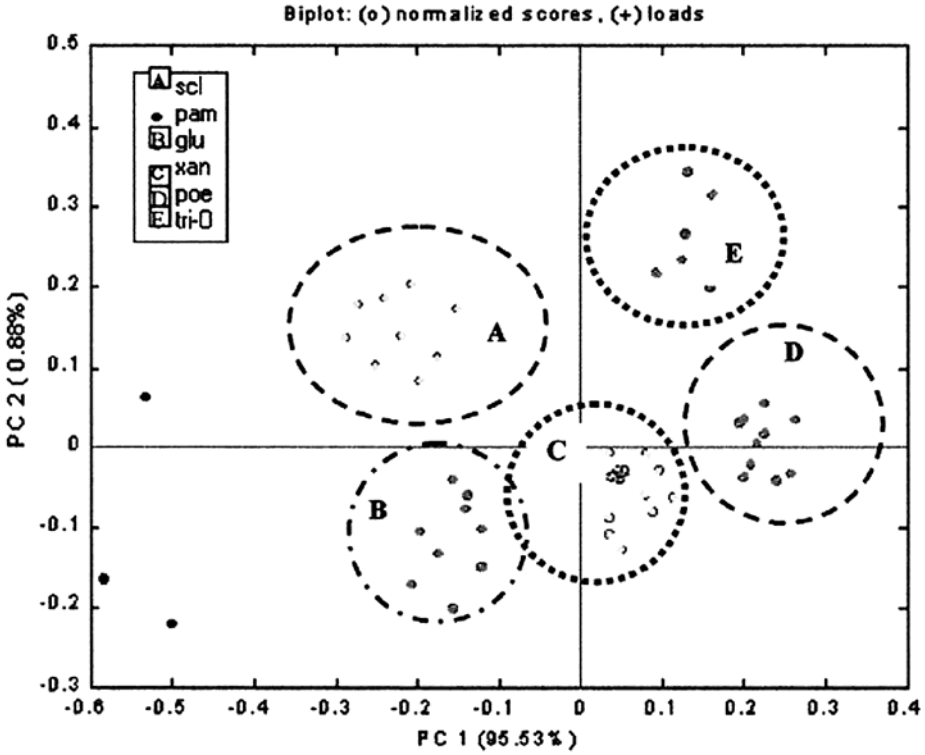

Figure 4. Biplot of principal component scores originating from X-ray fluorescence spectra of (A) scleroglucan, (B) glucomannan, (C) xanthan, (D) poly(ethylene oxide), (E) *ortho*-ethylamylose, and polyacrylamide.[57]

pointed out that, if a new data set were to be applied to the existing model, the neural network would not require further calculations to determine crystallinity, whereas the principal component method would require the redetermination of factor loadings corresponding to the new spectra.[55]

Miranda et al. studied the cross-linking of poly(vinyl alcohol) (PVA) induced by ultraviolet light in the presence of sodium benzoate as a sensitizer through the use of FTIR spectroscopy and PCA.[56] Their study was carried out by casting an aqueous solution of PVA and sodium benzoate onto glass plates and allowing the solvent to evaporate. The resulting polymer films were subsequently irradiated for 1, 2, 3, and 4 h, and IR spectra were recorded. The spectral data were then decomposed by PCA. Analysis of the results helped to clarify the cross-linking mechanism: the authors suggested that a free radical arising from the photolytical decomposition of sodium benzoate abstracts a hydrogen atom from the polymer chain, thus producing a polymeric radical. The latter, in turn, reacts with PVA hydroxyl groups to form ether linkages and therefore cross-links. Furthermore, the authors were able to demonstrate a linear correlation between the second principal component arising from the analysis and the irradiation time. This should allow for the determination of irradiation times of unknown polymer samples after measurement of the corresponding FTIR spectra.[56]

Vazquez et al. used total reflection X-ray fluorescence spectroscopy to develop a taxonomy of a set of synthetic and biopolymers.[57] The authors produced thin films of scleroglucan, xanthan, poly(ethylene oxide), glucomannan, *o*-ethylamylose, and polyacrylamide and recorded the X-ray fluorescence spectra of all samples. The spectra were subsequently subjected to PCA. Analysis of the results revealed that the first two principal components accounted for approximately 96% of the observed variance in the spectra. A biplot of the scores revealed six distinct clusters corresponding to the six polymer classes, indicating that X-ray fluorescence can indeed be used to classify unknown polymer samples (Figure 4).[57]

Within the framework of a high-throughput experimentation discovery program, Tuchbreiter and Mülhaupt used principal component regression (PCR) on data generated by attenuated total reflection Fourier transform infrared (ATR-FTIR) spectroscopy on olefin copolymers to determine polymer compositions.[58] ATR-FTIR spectroscopy allows for the rapid analysis of powders and polymeric solids without the need for sample preparation, as is the case in conventional FTIR techniques, which require the production of KBr pellets. When ATR-FTIR spectroscopy is combined with principal component regression, polymer composition (e.g., comonomer incorporation) can be determined quickly.

3.3. Mining by Quantitative Structure−Property Relationships (QSPRs). The goal of many data mining activities in high-throughput experimentation is ultimately the establishment of quantitative structure−property relationships (QSPRs). Quantitative structure−activity relationships (QSARs) have been used extensively in biology and biomolecular science for many years and have reached a high degree of sophistication.[59] Quantitative structure−property relationships are the materials science equivalent of QSARs. They are multivariant statistical correlations between a property of a system and a number of descriptors of the same system. They generally take the form

$$\text{property} = \text{constant} + (c_1 d_1) + (c_2 d_2) + \cdots + (c_n d_n) \quad (1)$$

where d_n is a property descriptor and the coefficient c_n is

reflective of the influence of that descriptor on the overall property. In the simplest case, one can discern a QSAR/QSPR using simple linear regression analysis on a data set. Linear regression is a standard statistical technique, and the reader is referred to the literature for further information.[60]

Within the area of polymer science, QSPR techniques have been used to address a number of problems. A prominent area is Ziegler−Natta catalysis. Although many theoretical investigations have been carried out to elucidate polymerization and activation mechanisms, as well as factors influencing the activity of this type of catalyst, the very large majority of these studies have employed computationally expensive methodologies, involving ab initio or density functional methods that focused on the structure of only the cationic metal fragment.[61−67] It has been demonstrated, however, that the interaction between catalyst and cocatalyst is of crucial importance in determining the activity of a Ziegler−Natta species.[68]

Yao and co-workers investigated the interaction between [Cp_2ZrMe] and a number of fluorophenylborates [$B(Ph-F_n)_4$] ($n = 0-5$) using computationally "cheap" molecular mechanics (MM) and QSPR methods.[69] To obtain optimized stable structures, the authors employed combined molecular dynamics and mechanics (MD/MM) calculations. The results suggested that, in all of the desired cases, the counterion is located opposite to the methyl group on the metal center. This, in turn, means that one of the aromatic rings on the counterion is oriented in such a way as to shield the vacant site in a "face-capping" manner. A subsequent QSPR analysis revealed that the metal−aryl ring centroid distance and the centroid−methyl group angle are the most pertinent descriptors with respect to catalytic activity.

In a subsequent paper, the same authors used a similar methodology to evaluate the influence of external ethereal donors on classical Ziegler catalyst systems [$TiCl_4$ immobilized on $MgCl_2$ and activated by trimethylaluminum (TMA)].[70] In addition to the interaction energy between the donor and the metal centers, the authors also calculated parameters such as the dipole moment, density, and molecular weight of the donor. A QSPR analysis showed that only the interaction energy and the dipole moment of the donor correlate with the observed activities. Furthermore, it was demonstrated that a correlation exists between the molecular weight distributions of the polymers produced by the catalyst systems and the principal moment of inertia of the external donor: donors with low moments of inertia give rise to polymers with low molecular weight distributions, whereas donors with high moments of inertia produce polymers with broad distributions. The authors reason that smaller molecules are more mobile and can thus move easily to minimize steric interactions with the support, thereby leading to active sites of similar sterics and, consequently, to narrow molecular weight distributions. Heavier external donors cannot achieve this to the same extent, thus maintaining the steric inhomogeneities resulting from the support and, in turn, giving rise to broader molecular weight distributions.[70]

In an earlier paper, Scordamaglia and Barino showed that some conformational features of external donors are strong descriptors for both activity and stereoregularity in the isospecific polymerization of propene.[71] A set of dimethoxysilanes was evaluated by calculating a number of molecular descriptors using a set of most-probable structures derived from a search of the rotational hypersurface of each molecule in the silane set. Of all of the calculated descriptors, two showed significant correlations with respect to donor stereoregulating power: the oxygen-to-oxygen distance and the conformations of the methoxy groups. These results led the authors to consider a new class of potential internal donors in the shape of 1,3-diethers. The compounds were evaluated in the same way as the set of dimethoxysilanes. Subsequent experiments confirmed that those diether compounds that most closely matched the optimal criteria in terms of oxygen−oxygen distance and methyl group conformations indeed gave rise to the most strongly stereoregulated polymers.[71,72]

QSPR methodologies have also found extensive applications in the area of the modeling of polymer properties, particularly the glass transition temperature (T_g). Below the glass transition temperature, polymer strands can both oscillate and vibrate around a fixed position, creating a certain amount of free volume. The size of the motion, and thus the amount of free volume, is dependent on the temperature. The glass transition occurs at the point at which the free volume is sufficient for the polymer backbones to move relative to one another. At this point, the backbone relaxes, and the material makes a transition from the solid to a quasiliquid state.[73]

On a very simple level, van Krevelen used group additive theory to predict T_g.[74] Here, the property under consideration is regarded as the scalar sum of the properties of the corresponding chemical groups. Examining a set of 77 samples of various cross-linked resins, Bicerano used a similar approach to develop a simple QSPR showing that T_g increases with a decreasing average number of "repeat units" between cross-links.[75] Hopfinger and Koehler subsequently extended the group additive theory methodology and combined it with molecular modeling, allowing for the estimation of unknown parameters. In their approach, the authors demonstrated that T_g can be correlated with the intramolecular flexibility of the polymer chain, which is composed of linear contributions of conformational entropies of the repeat units and intermolecular interactions arising mainly from electrostatic phenomena.[76,77]

In 1996, Katritzky and co-workers published a paper on the prediction of glass transition temperatures for low-molecular-weight homopolymers using a quantitative structure−property relationship treatment on a set of 22 polymers.[78] Using the CODESSA suite,[79] 238 different molecular descriptors (constitutional, geometrical, topological, electrostatic, quantum-chemical, and thermodynamic) were calculated in the first step. As considering all possible multiparameter correlations in such a huge descriptor space is practically impossible, the following procedure was used to find the final QSPR equations: (i) After intercorrelations of all 238 descriptors had been calculated, only those pairs of orthogonal descriptors i and j where $R_{ij}^2 < 0.1$ were chosen for the development of a first QSPR model. Such treatment afforded 10 755 orthogonal pairs. (ii) Subsequently, an initial

statistical analysis was carried out using the identified descriptor pairs. Of these, the 400 pairs that gave the highest regression correlation were used for higher-order regression analysis. (iii) Noncollinear descriptor scales were added, and a three-parameter regression was calculated. The calculation was repeated for all noncollinear scales. Of the triplets thus obtained, the 400 with the highest regression correlation were chosen. The process was repeated to give a set of quartets. Subsequently, the descriptor set showing the highest regression correlation was chosen as the final regression model ($R^2 = 0.928$). Analysis of the model showed that the glass transition temperature is strongly influenced by four factors: the difference between the negative and positive partial surface areas normalized by the number of atoms (describing electrostatic properties), the topological Randic index (describing the degree of branching), the number of OH groups present (as a measure of possible hydrogen-bonding interactions), and finally the partial negative surface area (again describing electrostatic properties).

In a subsequent paper, Katritzky et al. expanded this work considerably, both by using a larger data set (88 polymers) and by using only those descriptors that were calculated solely on the basis of theory.[80] The latter was done to ensure that the determined QSPRs would be applicable to any linear polymeric structure. This treatment afforded five strongly correlated descriptors ($R^2 = 0.946$): the moment of inertia (measuring the mass distribution around the principal axis of rotation), the Kier shape index (relating to the number of skeletal atoms, molecular branching, and the ratio of the atomic radius and the radius of an sp^3-hybridized carbon atom), the most negative atomic charge in the molecule, the descriptor HSA/TFSA (quantifying the ability of a polymer to form hydrogen bonds), and finally the fractional positive partially charged surface area (describing electrostatic interactions between molecules). These findings are certainly important, as they show that structure−property relationships can be established for large sets of polymers with differing chemical structures simply on the basis of calculated molecular descriptors.[80]

Cao and Lin subsequently developed a set of descriptors based on the rotation of the side chain, the bond count of the freely rotating part of the side chain, the substituted backbone electronegativity discrepancy, the polarizability effect index, and a hydrogen-bonding descriptor. These descriptors were subsequently evaluated using the same set of 88 polymers examined by Katritzky et al. QSPR methods showed that there is a good correlation between these descriptors and the glass transition temperature ($R^2 = 0.9056$).[81] Very recently, Shuai et al. developed QSPRs with respect to glass transition temperatures for amorphous low-molecular-weight materials used in the production of organic light-emitting diodes.[82]

Kohn and co-workers made extensive use of QSPRs when investigating a library of biodegradable polyarylates as novel candidates for biomedical applications.[83,84] Having synthesized the library using manual parallel synthesis methodologies, both the contact angles and the glass transition temperatures were determined experimentally (Figure 5). In an initial set of correlations, the researchers were able to show that there was a broad correlation between the length of the aliphatic side chains present in the polymer and the glass transition temperature: as the number of aliphatic carbon atoms in the side chains increased, T_g decreased in an exponential manner. Furthermore, a linear correlation between the air−water contact angle (CA) and the number of carbon atoms was discerned: as the latter increased, so did the angle. Branching could be shown to affect T_g only modestly, whereas oxygen substitutions in both side chains and polymer backbones affected both T_g and the contact angles very markedly. The polymers were subsequently screened against fibroplast proliferation. Again, good linear correlations could be demonstrated: fibroblasts tended to proliferate polymeric materials much more effectively if oxygen substitutions were present in either the backbone or the side chains. Furthermore, proliferation was found to decrease linearly with an increase in polymer surface hydrophobicity, except for those polymers incorporating oxygen in their backbone.[84]

In an elegant piece of work, Reynolds recently demonstrated the use of QSPRs for the design of polymer libraries.[40] In a first step, a subset of 17 members of a 112-membered virtual polymer library was selected on the basis of repeat unit topology and shown to be representative of the property space encompassed by the library as a whole. Subsequently, the subset was used to derive the topology and genetic-algorithm-optimized QSPR equations for T_g and CA. The predictions were tested against experimentally determined values for the residual polymers, and good correlations could be shown to exist: models gave R^2 values of 0.89 for T_g and 0.92 for CA. In a final step, the validated models were used to build focused libraries of polymers having specific values of T_g and CA. Again, it could be shown that the equations correctly identified polymers falling into the preset ranges.[40]

Other quantitative structure−property or structure−reactivity relationships have been developed for determining kinetic chain-transfer constants in polystyrene polymerizations for a set of transfer agents,[85] describing gaseous diffusion in polymers,[86] modeling transport behavior in amorphous polymeric materials,[87] and estimating inelastic mean free paths for polymers and other organic materials.[88] Moreover, an application for a patent concerning a QSAR approach for the prediction of polymer properties was recently filed by Procter & Gamble.[89]

3.4. Mining by Artificial Neural Networks (ANNs). Another increasingly popular approach to data mining and modeling is the use of artificial neural networks (ANNs). Neural networks are designed to mimic the way in which the human brain processes information and are composed of a number of interconnected processing units (neurons) that work in parallel to solve a given problem. In the human brain, a neuron receives information in the form of electric signals from other neurons to which it is connected. The connection happens via an axon, which can split into hundreds of branches. At the end of each branch, a synapse converts signals emanating from axons into electrical effects that can either inhibit or excite another axon connecting the next neuron, which is, in turn, either inhibited or excited.[90]

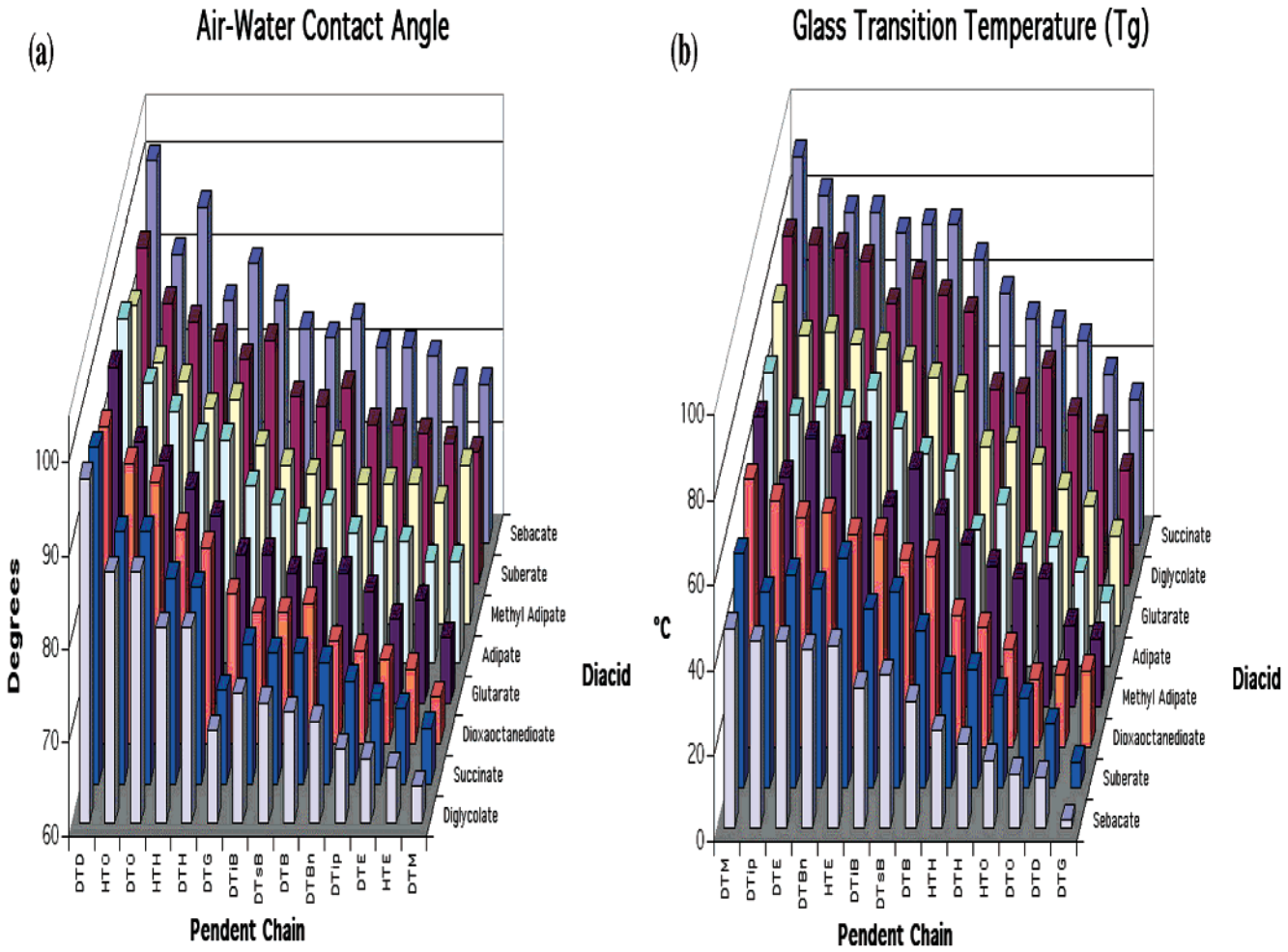

Figure 5. (a) Glass transition temperatures and (b) air/water contact angles for a 112-polymer library as a function of the polymer pendant chain (*x* axis) and backbone structure (*y* axis).[83,84]

Although a vast number of different network architectures exist,[91] all neural networks have some basic elements in common. Sumpter described a computational neural network (CNN) "as a computational system made up of a number of simple but highly connected processing elements that tend to store experimental knowledge by a dynamic state response to external inputs and make the information available for use. Using available data, a typical CNN 'learns' the essential relations between given inputs and outputs by storing information in a weighted distribution of connections. A learning algorithm provides the rule or dynamical equation that changes the distribution of the 'weight' (parameter) space to propagate the learning process."[92] A simple neuron has the microstructure depicted in Figure 6.

The structure commonly employed for learning and modeling tasks is that of the feed-forward network. This type of network allows a signal to travel in one direction only, thus associating inputs with outputs. Between the input and output layers are one (or more) hidden layers of neurons. Artificial neural networks are somewhat different from the other data mining tools described so far in the sense that they are theory-poor. Whereas linear regression, for example, assumes that two variables are related in a linear fashion, i.e., that the relationship can be described by the formula $y = a + bx$, where a is the intercept and b the slope, neural networks do not require a similar sort of theoretical underpinning; relationships between variables are stored in the weight matrix of the network. Furthermore, neural networks are capable of modeling any continuous function, whether linear or not. It also appears that factors that would normally present a serious obstacle to traditional modeling techniques such as multimodal distributions, data fuzziness, outliers, or partial nonavailability of data are less of an issue in NN modeling.[93] A number of general reviews have appeared discussing the use of ANNs in drug discovery[94] and materials science[92] in greater detail.

As was the case with QSPRs, neural networks, too, have been used to model glass transition temperatures. In a 1995 paper, Osguthorpe and co-workers described the use of neural networks for the prediction of physical and mechanical properties of linear homopolymers solely on the basis of their monomer structures. Using a number of network architectures and training procedures, they established that the best networks are capable of predicting T_g with an rms error of 35 K, thus demonstrating that information about the properties of the overall polymer is contained in the small monomer molecules.[95]

Mattioni and Jurs used a combination of linear and nonlinear modeling procedures to predict glass transition temperatures.[96] In the first part of their study, the authors evaluated the use of descriptors derived exclusively from the monomer units for the prediction of the glass transition

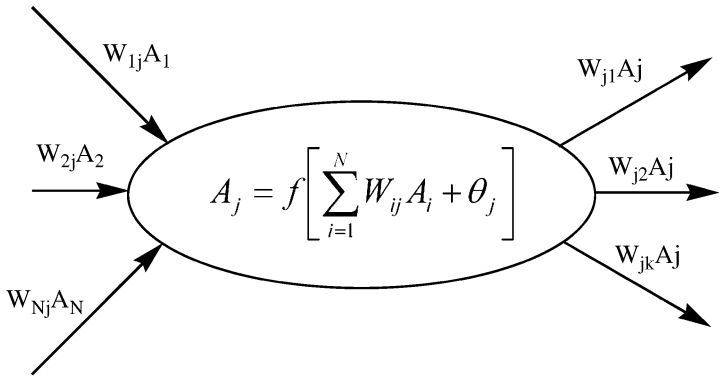

Figure 6. Neuron microstructure.

temperature. In a first attempt, a number of topological, geometric, and electronic descriptors were calculated, and linear modeling was used to identify the most effective subset of descriptors. Evaluation of the linear model showed that the rms error in the training set was 25.87 K ($R^2 = 0.869$) and 26.58 K ($R^2 = 0.870$) for the test set. The most effective descriptors from linear modeling were subsequently fed into computational neural networks of various architectures. The best net subsequently generated models with an rms error of 15.67 K for the training set ($R^2 = 0.952$) and 21.76 ($R^2 = 0.919$) for the prediction set. Use of either simulated annealing[97] or genetic algorithms to select the best descriptors did not lead to a significant improvement over the model generated using the linear approach.[96] In the second part of the study, descriptors were derived from the structure of the repeat unit, rather than the monomer, as this was thought to represent the properties of the polymer more accurately. However, despite the larger range and greater diversity of the data set, the prediction accuracy did not improve significantly: the best neural network prediction showed rms errors of 21.14 K ($R^2 = 0.958$) for the training set and 21.94 K ($R^2 = 0.962$) for the test set.

Zhang et al. used experimentally determined values of glass transition temperatures, entanglement molecular weights, and melt densities to train a simple three-layer feed-forward network with error back-propagation to model polymer chain dimensions, namely, the characteristic ratio C_∞. The characteristic ratio is the ratio of the mean-square end-to-end distance, $\langle r^2 \rangle_0$, of a linear polymer chain in the theta state to the product NL^2, where N is the number of rigid sections in the main chain, each of length L. In the case of $N \rightarrow \infty$, the symbol for the characteristic ratio becomes C_∞,[98] Data for 19 polymeric species were used in the model development. The authors were able to demonstrate that, using these three parameters, the characteristic ratio could be modeled with satisfactory accuracy.[99]

In 1998, Smith et al. reported the use of a combination of neural networks in the modeling process: rather than using one neural network to model all polymer properties, the authors used neural network modules, termed "local property experts".[100] Each module was designed to model one property only and contained an ensemble of neural networks. The overall output of the model "consists of a committee of the local experts (Figure 7) for which the result is taken as the ensemble average over all the networks comprising each local expert module". The local property experts have different architectures and topologies and have undergone different training methods. By combining several experts, the authors hoped both to optimize the accuracy of prediction and to minimize any overfitting. The trained networks were subsequently used in the design of new homopolymers not included in the training set. Of particular interest to the researchers was the design of bisphenol A polycarbonate (BPAPC) with improved impact resistance. An evaluation of nine BPAPC derivatives using the trained neural networks delivered three lead compounds. Unfortunately, the authors did not provide any experimental confirmation in their paper as to whether the lead structures were indeed superior with respect to already known polymers. However, in a subsequent patent, they claimed that these materials do indeed show improved impact resistance.[101]

Other papers reporting the use of neural network modeling are concerned with the prediction of electronic properties of polymers,[102,103] as well as the prediction of heat capacities, tensile strength, tensile modulus, compressive strength, and elongation.[93] Furthermore, neural networks have been employed in the optimization of polymer processing[104] and in the inferential estimation of polymer quality during polymerization.[105]

3.5. Other Methods. The above overview does not present a comprehensive list of all available data mining techniques, but rather focuses on those methods most commonly used in combinatorial polymer science to date. Methodologies such as recursive partitioning[106−109] are becoming increasingly commonplace in the biomedical field but have, thus far, not been applied to polymer science problems. Other methods have been used only very sporadically. Debska et al. reported the use of cluster analysis to improve the water resistance of acrylamide-modified melamine resins,[110] and Sun et al. used fuzzy-set theory to investigate the relationship between polymer structure and the glass transition temper-

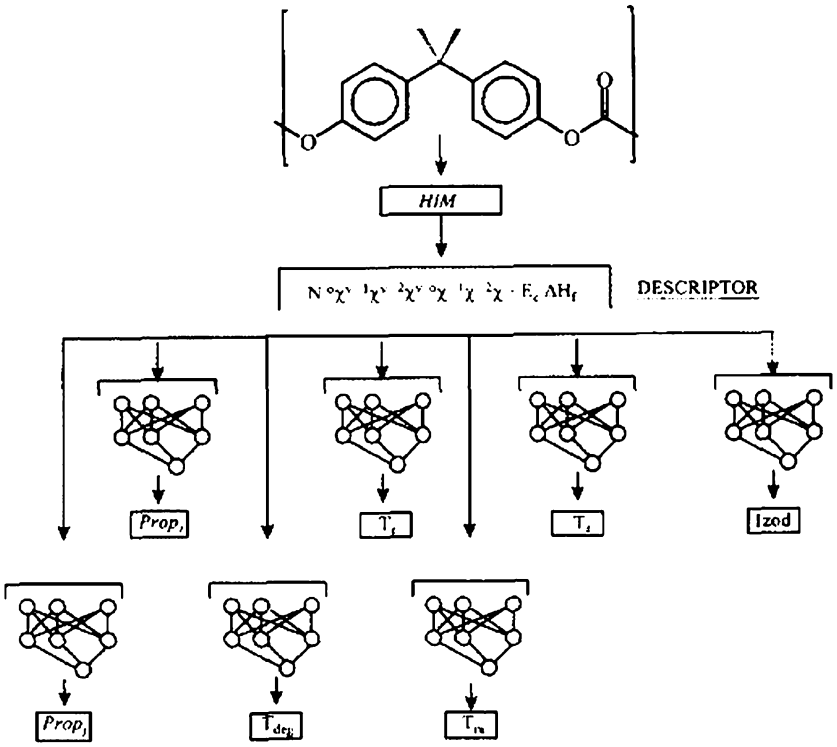

Figure 7. Neural network system composed of multiple expert modules (one for each property).[100]

ature.[111] However, these are isolated examples that, so far, are not in common use.

4. Conclusion

Although, when compared to bioinformatics, the field of materials informatics is small, it is nevertheless evolving rapidly. This development is fueled by the increasing application of combinatorial and high-throughput experimentation techniques in the materials sciences and the development of sophisticated data handling and statistical technology, as well as by the easy and inexpensive availability of brute computing power. Particularly in the area of polymer science, these developments present a huge window of opportunity. On one hand, research will simply be accelerated, i.e., it will be possible to carry out more polymerizations, to produce more polymer blends, and to evaluate more process conditions in continuously shortening time spans. On the other hand, and more importantly, research will become more intelligent, provided that the data produced are of sufficient quality and structure to allow for the development of quantitative structure−property relationships. If this is the case, these developments will allow both adaptive experimentation and materials design on the basis of predictable properties, rather than as a result of what is essentially serendipity. If, for example, a polymer with a certain glass transition temperature is required, modeling should enable the design of polymer architectures that are known will fall in the desired property range even before they have been physically prepared in the laboratory. The data mining examples discussed in the previous sections illustrate that the technology necessary to achieve this is, in principle, in place. In this context, it is also becoming increasingly clear that the crucial factor to success in any combinatorial endeavor is not simply "high-throughput" experimentation, generating huge amounts of possibly superfluous data, but "smart high-throughput" experimentation that includes both design-of-experiment methodologies and data mining techniques in every step of the combinatorial discovery process, making the latter adaptive. The further integration of all aspects of compound discovery and evaluation, from library design to data mining, will lead to the most effective use of possible synergisms, thus making smart high-throughput experimentation even smarter.

Acknowledgment. The authors wish to thank Avantium Technologies and Accelrys Ltd. for the information provided and the Dutch Polymer Institute (DPI) for financial support of this work.

References and Notes

(1) Terrett, N. K. *Combinatorial Chemistry*; Oxford University Press: Oxford, U.K., 1998.
(2) Terrett, N. K.; Gardner, M.; Gordon, D. W.; Kobylecki, R. J.; Steele, J. *Tetrahedron* **1995**, *51*, 8135.
(3) Terrett, N. K.; Gardner, M.; Gordon, D. W.; Kobylecki, R. J.; Steele, J. *Chem. Eur. J.* **1997**, *3*, 1917.
(4) Hagemeyer, A.; Jandeleit, B.; Liu, Y.; Poojary, D. M.; Turner, H. W.; Volpe, J., A. F.; Weinberg, W. H. *Appl. Catal. A: Gen.* **2001**, *221*, 23.
(5) Jandeleit, B.; Turner, H. W.; Uno, T.; van Beek, J. A. M.; Weinberg, W. H. *Cattech* **1998**, *2*, 101.
(6) Jandeleit, B.; Schaefer, D. J.; Powers, T. S.; Turner, H. W.; Weinberg, W. H. *Angew. Chem., Int. Ed. Engl.* **1999**, *38*, 2492.
(7) Jandeleit, B.; Schaefer, D. J.; Powers, T. S.; Turner, H. W.; Weinberg, W. H. *Angew. Chem.* **1999**, *111*, 2648.
(8) Hoogenboom, R.; Meier, M. A. R.; Schubert, U. S. *Macromol. Rapid Commun.* **2003**, *24*, 16.
(9) http://www.tripos.com.
(10) http://www.daylight.com.
(11) http://www.accelrys.com.
(12) http://www.cambridgesoft.com.

(13) Weininger, D. J. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31.
(14) Ash, S.; Cline, M. A.; Homer, W. R.; Hurst, T.; Smith, G. B. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 71.
(15) Patterson, J. A.; Schultz, J. L.; Wilks, E. S. *J. Chem. Inf. Comput. Sci*. **1995**, *35*, 8.
(16) Gankin, Y.; Tripathy, S.; Aronov, V.; Votano, J.; Bilibin, A.; Mazurin, O.; Shilov, V. *Synth. Met.* **2001**, *119*, 387.
(17) Herz, M. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 469.
(18) http://www.dtwassociates.com.
(19) Ihlenfeldt, W.-D.; Takahashi, Y.; Abe, H.; Sasaki, M. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 109.
(20) Ihlenfeldt, W.-D.; Voigt, J. H.; Bienfait, B.; Oellien, F.; Nicklaus, M. C. *J. Chem. Inf. Comput. Sci.* **2002**, *2002*, 46.
(21) http://www2.ccc.uni-erlangen.de/software/cactvs.
(22) Corkan, L. A.; Lindsey, J. S. *Chemom. Intell. Lab. Syst.* **1992**, *17*, 47.
(23) Plouvier, J.-C.; Corkan, L. A.; Lindsey, J. S. *Chemom. Intell. Lab. Syst.* **1992**, *17*, 75.
(24) Corkan, L. A.; Plouvier, J.-C.; Lindsey, J. S. *Chemom. Intell. Lab. Syst.* **1992**, *17*, 95.
(25) Lindsey, J. S.; Corkan, L. A. *Chemom. Intell. Lab. Syst.* **1993**, *21*, 139.
(26) Du, H.; Corkan, L. A.; Yang, K.; Kuo, P. Y.; Lindsey, J. S. *Chemom. Intell. Lab. Syst.* **1999**, *48*, 181.
(27) Du, H.; Shen, W.; Kuo, P. Y.; Lindsey, J. S. *Chemom. Intell. Lab. Syst.* **1999**, *48*, 205.
(28) Du, H.; Jindal, S.; Lindsey, J. S. *Chemom. Intell. Lab. Syst.* **1999**, *48*, 235.
(29) Du, H.; Lindsey, J. S. *Chemom. Intell. Lab. Syst.* **2002**, *62*, 159.
(30) Matsumoto, T.; Du, H.; Lindsey, J. S. *Chemom. Intell. Lab. Syst.* **2002**, *62*, 149.
(31) Matsumoto, T.; Du, H.; Lindsey, J. S. *Chemom. Intell. Lab. Syst.* **2002**, *62*, 129.
(32) Kuo, P. Y.; Du, H.; Corkan, L. A.; Yang, K.; Lindsey, J. S. *Chemom. Intell. Lab. Syst.* **1999**, *48*, 219.
(33) http://www.avantium.com.
(34) Gruter, G. J. M. Avantium Technologies, Amsterdam, The Netherlands. Personal communication, 2003.
(35) http://www.creonlabcontrol.com.
(36) Murphy, V.; Bei, X.; Boussie, T. R.; Bruemmer, O.; Diamond, G.; Goh, C.; Hall, K. A.; LaPointe, A. M.; Leclerc, M.; Longmire, J. M.; Shoemaker, J. A. W.; Turner, H. W.; Weinberg, W. H. *Chem. Rec.* **2002**, *2*, 278.
(37) Wold, S.; Berglund, A.; Kettaneh, A. *J. Chemom.* **2002**, *16*, 377.
(38) Linusson, A.; Gottfries, J.; Lindgren, F.; Wold, S. *J. Med. Chem.* **2000**, *43*, 1320.
(39) Andersson, P. M.; Lundstedt, T. *J. Chemom.* **2002**, *16*, 490.
(40) Reynolds, C. H. *J. Comb. Chem.* **1999**, *1*, 297.
(41) Cawse, J. N. *Acc. Chem. Res.* **2001**, *34*, 213.
(42) Gruter, G. J. M.; Graham, A.; McKay, B.; Gilardoni, F. *Macromol. Rapid Commun.* **2003**, *24*, 73.
(43) Iden, R.; Schorf, W.; Hadeler, J.; Lehmann, S. *Macromol. Rapid Commun.* **2003**, *24*, 63.
(44) Hand, D. J.; Blunt, G.; Kelly, M. G.; Adams, N. A. *Stat. Sci.* **2000**, *15*, 111.
(45) http://www.microsoft.com.
(46) http://www.microcal.com.
(47) http://www.spotfire.com.
(48) http://www.omniviz.com.
(49) http://www.avs.com.
(50) http://www.wolfram.com.
(51) Suh, C.; Rajagopalan, A.; Li, X.; Rajan, K. *Data Sci. J.* **2002**, *1*, 19.
(52) Bajorath, J. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 233.
(53) Wold, S.; Esbensen, K.; Geladi, P. *Chemom. Intell. Lab. Syst.* **1987**, *2*, 37.
(54) Vanden Eynde, X.; Bertrand, P. *Surf. Interface Anal.* **1997**, *25*, 878.
(55) Batur, C.; Vhora, M. H.; Cakmak, M.; Serhatkulu, T. *ISA Trans.* **1999**, *38*, 139.
(56) Miranda, T. M. R.; Goncalves, A. R.; Amorim, M. T. P. *Polym. Int.* **2001**, *50*, 1068.
(57) Vazquez, C.; Boeykens, S.; Bonadeo, H. *Talanta* **2002**, *57*, 1113.
(58) Tuchbreiter, A.; Marquardt, J.; Zimmermann, J.; Walter, P.; Mülhaupt, R.; Kappler, B.; Faller, D.; Roths, T.; Honerkamp, J. *J. Comb. Chem.* **2001**, *3*, 598.
(59) Hansch, C.; Hoekman, D.; Weininger, D. J.; Selassie, C. D. *Chem. Rev.* **2002**, *102*, 783.
(60) Philips, J. L. *How to Think about Statistics*; 6th ed.; W.H. Freeman and Company: New York, 2000.
(61) Bierwagen, E. P.; Bercaw, J. E.; Goddard, W. E., III. *J. Am. Chem. Soc.* **1994**, *116*, 1481.
(62) Woo, T. K.; Fan, L.; Ziegler, T. *Organometallics* **1994**, *13*, 2252.
(63) Yoshida, T.; Koga, N.; Morokuma, K. *Organometallics* **1995**, *14*, 746.
(64) Axe, F. U.; Coffin, J. M. *J. Phys. Chem.* **1994**, *98*, 2567.
(65) Lohrenz, J. C. W.; Woo, T. K.; Ziegler, T. *J. Am. Chem. Soc.* **1995**, *117*, 12793.
(66) Maiti, A.; Sierka, M.; Andzelm, J.; Golab, J.; Sauer, J. *J. Phys. Chem. A* **2000**, *104*, 10932.
(67) Weiss, H.; Boero, M.; Parrinello, M. *Macrmol. Symp.* **2001**, *2001*, 137.
(68) Murtuza, S.; Casagrande, O. L.; Jordan, R. F. *Organometallics* **2002**, *21*, 1882.
(69) Yao, S.; Shoji, T.; Iwamoto, Y.; Kamei, E. *Comput. Theor. Polym. Sci.* **1999**, *9*, 41.
(70) Yao, S.; Tanaka, Y. *Macromol. Theory Simul.* **2001**, *10*, 850.
(71) Scordamaglia, R.; Barino, L. *Macromol. Theory Simul.* **1998**, *7*, 399.
(72) Barino, L.; Scordamaglia, R. *Macromol. Theory Simul.* **1998**, *7*, 407.
(73) Stevens, M. P. *Polymer Chemistry. An Introduction*; Oxford University Press: Oxford, U.K., 1990.
(74) van Krevelen, D. W. *Properties of Polymers: Their Correlation with Chemical Structure, Their Numerical Estimation and Prediction from Additive Group Contributions*; 3rd ed.; Elsevier: Amsterdam, 1990.
(75) Bicerano, J.; Sammler, R. L.; Carriere, C. J.; Seitz, J. T. *J. Polym. Phys. B* **1996**, *34*, 2247.
(76) Hopfinger, A. J.; Koehler, M. G.; Pearlstein, R. A. *J. Polym. Sci. B* **1988**, *26*, 2007.
(77) Koehler, M. G.; Hopfinger, A. J. *Polymer* **1989**, *30*, 116.
(78) Katritzky, A. R.; Rachwal, P.; Law, K. W.; Karelson, M.; Lobanov, V. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 879.
(79) Ivanciuc, O. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 405.
(80) Katritzky, A. R.; Sild, S.; Lobanov, V.; Karelson, M. *J. Chem. Inf. Comput. Sci*. **1998**, *38*, 300.
(81) Cao, C.; Lin, Y. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 643.
(82) Yin, S.; Shuai, Z.; Wang, Y. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 970.
(83) Brocchini, S.; James, K.; Tangpasuthadol, V.; Kohn, J. *J. Am. Chem. Soc.* **1997**, *119*, 4553.
(84) Brocchini, S.; James, K.; Tangpasuthadol, V.; Kohn, J. *J. Biomed. Mater. Res.* **1998**, *42*, 66.
(85) Ignatz-Hoover, F.; Petrukhin, R.; Karelson, M.; Katritzky, A. R. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 295.
(86) Patel, H. C.; Tokarski, J. S.; Hopfinger, A. J. *Pharmaceutical Research* **1997**, *14*, 1349.
(87) Tokarski, J. S.; Hopfinger, A. J.; Hobbs, J. D.; Ford, D. M.; Faulon, J.-L. M. *Comput. Theor. Polym. Sci.* **1997**, *7*, 199.
(88) Cumpson, P. J. *Surf. Interface Anal.* **2001**, *31*, 23.
(89) Schneiderman, E.; Stanton, D.; Trinh, T.; Laidig, W. D.; Kramer, M. L.; Gosselink, E. P. Predictive Method for Polymers. International Patent WO 02/44686 A2, 2002.
(90) Tortora, G. J.; Grabowski, S. R. *Introduction to the Human Body. The Essentials of Anatomy and Physiology*; John Wiley & Sons: New York, 2001.

(91) Sumpter, B. G.; Getino, C.; Noid, D. I. *Annu. Rev. Phys. Chem.* **1994**, *45*, 439.
(92) Sumpter, B. G.; Noid, D. I. *Annu. Rev. Mater. Sci.* **1998**, *26*, 233.
(93) Sumpter, B. G.; Noid, D. I. *J. Therm. Anal.* **1996**, *46*, 833.
(94) Gasteiger, J.; Teckentrup, A.; Terfloth, L.; Spycher, S. *J. Phys. Org. Chem.* **2003**, *16*, 232.
(95) Joyce, S. J.; Osguthorpe, D. J.; Padgett, J. A.; Price, G. J. *J. Chem. Soc., Faraday Trans.* **1995**, *91*, 249.
(96) Mattioni, B. E.; Jurs, P. C. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 233.
(97) Sutter, J. M.; Dixon, S. L.; Jurs, P. C. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 77.
(98) Gold, V.; Loening, K. L.; McNaught, A. D. *Compendium of Chemical Terminology: IUPAC Recommendations*; Blackwell Scientific: London, 1987.
(99) Zhang, L.-X.; Xia, A.; Zhao, D.-L. *J. Polym. Sci. B* **2000**, *38*, 3163.
(100) Ulmer, C. W., II; Smith, D. A.; Sumpter, B. G.; Noid, D. I. *Comput. Theor. Polym. Sci.* **1998**, *8*, 311.
(101) Smith, D. A.; Ulmer, C. W., II. Impact Resistant Polymers. PCT International Patent Application 98/37118, 1998.
(102) Luo, Q.; Darsey, J. A.; Compadre, C. M. *Polym. Prepr.* **1997**, *38*, 243.
(103) Taylor, K. K.; Darsey, J. A. *Polym. Prepr.* **2000**, *41*, 331.
(104) Allan, G.; Yang, R.; Fotheringham, R. M. *J. Mater. Sci.* **2001**, *36*, 3113.
(105) Zhang, J.; Martin, E. B.; Morris, A. J.; Kiparissides *Comput. Chem. Eng.* **1997**, *21*, 1025.
(106) Chen, X.; Rusinko, A., III; Young, S. S. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 1054.
(107) Rusinko, A., III; Farmen, M. W.; Lambert, C. G.; Brown, P. L.; Young, S. S. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 1017.
(108) Rusinko, A., III; Young, S. S.; Drewry, D. H.; Gerritz, S. W. *Comb. Chem. High Throughput Screening* **2002**, *5*, 125.
(109) Adams, N.; Schubert, U. S. *Macromol. Rapid Commun.* **2004**, *25*, in press.
(110) Debska, B.; Wianowska, E. *Polym. Test.* **2002**, *21*, 43.
(111) Sun, H.; Tang, Y.; Wu, G.; Zhang, F. *J. Polym. Sci. B* **2001**, *40*, 454.

CC034021B