

NIH Public Access

Author Manuscript

Published in final edited form as:

Comput Chem Eng. 2009 May 21; 33(5): 954–963. doi:10.1016/j.compchemeng.2008.09.019.

A computational molecular design framework for crosslinked polymer networks

J.C. Eslick^{a,1}, Q. Ye^{b,2}, J. Park^{b,3}, E.M. Topp^{c,4}, P. Spencer^{d,5}, and K.V. Camarda^{a,*}

^a Department of Chemical and Petroleum Engineering, The University of Kansas, 1530 W. 15th St., Lawrence, KS 66045, United States

^b Bioengineering Research Center, The University of Kansas, 1530 W. 15th St., Lawrence, KS 66045, United States

^c The University of Kansas, 2095 Constant Ave., Lawrence, KS 66047, United States

^d Bioengineering Research Center and Department of Mechanical Engineering, The University of Kansas, 1530 W. 15th St., Lawrence, KS 66045, United States

Abstract

Crosslinked polymers are important in a very wide range of applications including dental restorative materials. However, currently used polymeric materials experience limited durability in the clinical oral environment. Researchers in the dental polymer field have generally used a timeconsuming experimental trial-and-error approach to the design of new materials. The application of computational molecular design (CMD) to crosslinked polymer networks has the potential to facilitate development of improved polymethacrylate dental materials. CMD uses quantitative structure property relations (QSPRs) and optimization techniques to design molecules possessing desired properties. This paper describes a mathematical framework which provides tools necessary for the application of CMD to crosslinked polymer systems. The novel parts of the system include the data structures used, which allow for simple calculation of structural descriptors, and the formulation of the optimization problem. A heuristic optimization method, Tabu Search, is used to determine candidate monomers. Use of a heuristic optimization algorithm makes the system more independent of the types of QSPRs used, and more efficient when applied to combinatorial problems. A software package has been created which provides polymer researchers access to the design framework. A complete example of the methodology is provided for polymethacrylate dental materials.

Keywords

Molecular design; Polymer

^{© 2008} Elsevier Ltd. All rights reserved

^{*}Corresponding author. Tel.: +1 785 864 2908; fax: +1 785 864 4967...

E-mail addresses: jceslick@ku.edu (J.C. Eslick), yeq@ku.edu (Q. Ye), bellgu@ku.edu (J. Park), topp@ku.edu (E.M. Topp), pspencer@ku.edu (P. Spencer), camarda@ku.edu (K.V. Camarda). Tel.: +1 785 864 2937; fax: +1 785 864 4967.

²Tel.: +1 785 864 1746; fax: +1 785 864 1742.

³Tel.: +1 785 864 1748; fax: +1 785 864 1742.

⁴Tel.: +1 785 864 3644; fax: +1 785 864 5736.

⁵Tel.: +1 785 864 8140; fax: +1 785 864 1742.

1. Introduction

Polymeric materials, such as dental composites, are increasingly being used to replace mercury-containing dental amalgam. Composite materials experience a relatively short clinical lifetime, so researchers are designing polymers with improved longevity (Murray, Windsor, Smyth, Hafez, & Cox, 2002). Most studies have applied a trial-and-error approach, in which incremental changes to the chemical structure are used to improve specific properties. This requires an expensive and time-consuming set of synthesis and testing steps, and since many properties change with a change in the chemical structure, some properties may be improved in a step, while others may deteriorate. A design method which estimates many properties at the same time and seeks to optimize many property values simultaneously, such as computational molecular design (CMD), can make such research more efficient. However, the crosslinked random copolymer structure of dental polymers presents a challenge to any CMD approach, since the exact molecular structure of these polymers is difficult to determine experimentally or describe in a manner useful for calculation. The goal of this work is to develop a general mathematical framework for CMD capable of handling crosslinked random copolymers.

The first step in any CMD study is to build quantitative structure property relationships (QSPRs) for physical and chemical properties of interest. Experimental property data is collected to develop QSPRs specific to the type of materials to be designed. In the dental polymer example used here, QSPRs are needed for crosslinked polymethacrylates. Polymer properties are often dependent on processing conditions, so it is best to design and conduct a set of consistent experiments to measure important properties, such as tensile strength, elastic modulus, chemical resistance, or glass transition temperature. Once experimental data is collected, the property information and polymer structures are entered into a database. A novel data structure is employed to ease the computation of structural descriptors. The particular descriptors chosen for use in QSPRs depend on the polymer system being studied. These numerical descriptors and property data are then exported to statistical software to develop QSPRs. The QSPRs are used with optimization techniques to design new materials which are likely to have desirable properties. The candidate materials are synthesized, tested experimentally, and the results are used to refine the QSPRs as necessary. After some iterations of the design process, synthesis and characterization of new materials is achieved. A review of CMD research is provided in Section 2.

Background information as well as and details of the design methodology used in this work are provided in the following sections. Section 3 describes how polymer structures are stored. Section 4 presents how this structural information is used to calculate numerical descriptors. Section 5 describes how predictive property models are created in this work employing structural descriptors. Section 6 describes the optimization problem being solved to design new polymers, as well as the Tabu Search algorithm used to solve it. Finally, a small example of the complete design process applied to polymeric dental materials is presented in Section 7.

2. Background

The CMD methodology requires the generation of a predictive model to provide estimated property values for polymer networks. This work focuses on topological descriptors, which describe the identity of atoms and how they are connected, but not the overall network geometry. Calculation of the exact geometry requires complicated molecular simulations; such simulations are far too computationally expensive to be used within an optimization framework for molecular design.

A common way to quantify chemical structure is to use a group contribution method, in which the number of each type of group in a given molecule describes the chemical structure. UNIQUAQ and UNIFAC are group contribution methods that can be used to predict thermodynamic properties in liquid mixtures (Fredenslund, Jones, & Praunsnitz, 1975). Joback and Reid (1987) use a group contribution method to predict several properties of pure organic substances such as: viscosity, boiling point, freezing point, and heat capacity. Group contributions have been used to predict vapor liquid equalibria (Gani, Tzouvaras, Rasmussen, & Fredenslund, 1989). Group contribution methods have also been applied to polymers (Van Krevelen, 1997), and have been used extensively for CMD, as will be described later in this section.

Connectivity indices, first developed by Randi (1975), have also been applied to property prediction. The idea of connectivity indices was expanded and applied to pharmaceutical property prediction by Kier and Hall (1986). More recently, Bicerano (2002) applied connectivity indices to prediction of a wide variety of straight-chained, non-crosslinked, limiting molecular weight polymers. Gani, Harper, and Hostrup (2005) has used connectivity indices to estimate group contributions where a group contribution method exists with insufficient groups to describe certain molecules.

While group contributions and connectivity indices are effective at describing individual repeat units, they do not contain information about polymer crosslinking. While these descriptors are still useful for crosslinked polymers, more information is also needed to accurately predict properties, since crosslinking has a large effect on many polymer properties (Van Krevelen, 1997). For example, crosslink density has a significant effect on the glass transition temperature. The average number of backbone atoms between crosslinks can be used to estimate the glass transition temperature of polymers using a linear relationship (Van Krevelen, 1997). Bicerano, Sammler, Carrier, and Seitz (1996) used the number of rotational degrees of freedom between crosslinks to predict glass transition temperature. Porter (1995) accounts for crosslinking in polymer property prediction using degrees of freedom. Cook, Forsythe, Irawati, Scott, and Xia (2003) studied the effects of crosslinking on the glass transition temperature of dental materials. As seen in dental materials, processing conditions can affect the crosslink structure. Small rings may form that do not contribute to crosslinking, and there are areas in the polymer network of varying crosslink density (Ye, Spencer, Wang, & Misra, 2007).

Most topological descriptors can be calculated from various graphs which represent molecular structure. Graphs consist of a set of vertices and edges. Edges represent connections between vertices (West, 2001). In chemical applications, vertices often represent atoms while edges represent bonds. Hydrogen-suppressed graphs of molecular structure are used to calculate connectivity indices (Bicerano, 2002; Kier & Hall, 1986). Many graph theory algorithms can be employed in the calculation of descriptors. The Ullmann (1976) subgraph isomorphism algorithm can be used to find groups for group contribution methods, and can be useful for other calculations as well. Other algorithms can be used to find other important structures, such as backbone and pendant groups, ring structures, and paths.

Crosslinked random copolymers are more difficult to describe as graphs. The network structure and random arrangement of monomers make it difficult to define a repeat unit. A considerable amount of research has been devoted to describing and predicting polymer network structure. Much of the effort has focused on calculation of the gel point of polymers. The gel point is the point at with an infinite polymer network would form. Flory (1941a, 1941b, 1941c) described a method whereby the degree of conversion is used to calculate the probability that functional groups of a monomer react. As the polymer network

builds, if the expected number of branching points increases, an infinite network will form. This allows the gel point to be calculated. To improve the gel point predictions, additional details were added to the network formation (Dušek, Gordon, & Ross-Murphy, 1978; Gordon, 1962). Methods have been developed which can estimate a more detailed polymer structure, which can be used in the prediction of polymer properties (Stepto, Cail, & Taylor, 2000).

Property estimation methods can be combined with optimization techniques to design new chemical products using CMD methods. The optimization problems are usually formulated as mixed integer linear programming (MILP) or mixed integer non-linear programming problems (MINLP). Gani and Fredenslund (1993) presented a general molecular design method to create chemical products having specific properties. Maranas (1996) describes MILP formulations using group contributions to predict properties of non-crosslinked polymers. Camarda and Maranas (1999) use connectivity indices to formulate a non-crosslinked polymer design problem as an MINLP. Sahinidis, Tawarmalani, and Yu (2003) used a deterministic algorithm to find the solution to a MINLP problem to design new refrigerants.

Deterministic algorithms may not be practical for large combinatorial problems, particularly those including highly non-linear constraints. In such cases, heuristic methods are often useful. Heuristic methods are not guaranteed to find the global optimal solution; however, due to the limited accuracy of QSPRs, near optimal solutions to the optimization problem are often as useful in practice. For example, genetic algorithms have been used in the design of straight-chain polymers using group contribution methods to predict properties (Venkatasubramanian, Chan, & Caruthers, 1994).

Tabu Search is a heuristic algorithm developed by Glover (1990a, 1990b). This method keeps a record of recent solutions in a Tabu list to prevent cycling near local optima, and encourages exploration of the entire search space. Tabu Search has been used in chemical process optimization (Lin & Miller, 2004a, 2004b), planning and scheduling (Dowsland, 1998; Gendreau, Laporte, & Semet, 1998; Kimms, 1996), and molecular design (Chavali, Lin, Miller, & Camarda, 2004; Eslick and Camarda, 2006; Lin, Chavali, Camarda, & Miller, 2005; Zhao, Ralston, Middaugh, & Camarda, 2004). Tabu Search is used in this work because is can handle non-linear combinatorial problems in an efficient manner.

3. Polymer structure

Numerical descriptors of polymer structure need to be calculated for a large number of candidate structures if they are to be used in QSPRs, so an efficient data structure is needed to store structural information for crosslinked polymers. The data structure should allow the efficient calculation of many types of topological descriptors.

In this work, polymer information including structures and experimental data are stored in a database. Although almost any type of polymer can be managed by the system described, the focus of this work is crosslinked random copolymers. These types of polymers are especially difficult to describe due to their ill-defined structures. The crosslinked structure of the dental polymers is very difficult to determine experimentally, and may vary due to differences in processing (Ye, Spencer, et al., 2007). Dental resins generally contain a photoinitiator, and polymerization is initiated by visible light. Factors such as the presence of solvents, cure time, light intensity, or photoinitiator choice can change the degree of conversion and crosslinking structure. Once the resin has cured, the presence of solvents or unreacted monomer will also affect the properties.

Structural descriptors can be calculated, even though the exact polymer structure is not known, by making a set of assumptions. The assumptions depend on the system being studied; details will be provided for the dental polymer case later in this section. It is generally best to use the simplest possible structure that is adequate to predict properties. Factors such as the exact crosslinking structure may be difficult to predict and may not be necessary to consider depending on the nature of property information desired.

In this work, polymer structures are stored using three types of graphs: monomer, polymer, and full. Monomer graphs are used to store monomer structures. Each vertex represents an atom or connection point between monomers, and each edge represents a chemical bond. Polymer graphs contain a representative section of the overall polymer structure. Each vertex represents a monomer, and each edge represents a bond between monomers. Full graphs describe a representative section of the chemical structure of polymer networks. The full graph is formed by replacing the monomer vertices in the polymer graph with the atoms and bonds of the monomers. Each type of graph is useful for different types of calculations. Monomer graphs are useful in forming the polymer structure, and may be used to predict properties of the unreacted monomers such as viscosity. The polymer and full graphs contain similar information. However, when constructing the polymer structure, it is much easier to deal with monomer vertices than a complete chemical structure. The polymer graph gives a clear picture of the arrangement of monomers. Easy access the to identities of the monomers in the polymer structure are lost in the full graph.

Monomer graphs consist of vertices that represent atoms or connections to another monomer and edges that represent bonds. Connection vertices are labeled for use in polymer and full graphs. Every vertex and edge is labeled with a functional group. Each functional group can have one or more states, which represent ways in which a functional group can react. The functional group system provides an easy way to enter each monomer once, while having several possible structures available. The monomer graphs are well suited to systematic generation of polymer structure due to their ability to exist in several states. When the state of a monomer graph changes, the vertex and edge indices are resorted so that all of the vertices and edges in the current state have a lower index than the ones that are out-of-state. Graph algorithms may then ignore the out-of-state vertices and edges. Fig. 1 shows an example of two monomer graph states for the 2-hydroxyethly methacrylate (HEMA) monomer, and Fig. 2 shows four monomer graph states for the 2,2-bis[4(2-hydroxy-3methacryloyloxy-propyloxy)-phenyl] (bisGMA) monomer. The dummy Xx atoms represent connection points to other monomers.

Polymer graphs show how monomers are bonded. Each vertex represents a monomer, while edges represent bonds between monomer units. For simple polymers such as linear block copolymers, polymer graphs can be constructed manually by arranging a pattern of monomers. In more complex cases such as random crosslinked copolymers, the graphs can be constructed systematically given a set of rules. The rules can be very simple, or more complex based on knowledge of the polymer chemistry. Whether a simple set of rules or a complicated molecular simulation is used to generate the polymer structure, the overall design framework is equally effective. The size of the generated polymer depends on what is needed to adequately represent the structure. An example polymer graph is given in Fig. 3.

Simple polymers have a well-defined regular repeating pattern; however, crosslinked random copolymers form a network with no obvious repeat unit. To solve problems associated with descriptor calculation, a large representative section of polymer is generated. Crosslinked polymer networks are normally treated as infinite, but the polymer graphs must be finite, so the representative sections will have many cut ends. The concept of a core and buffer section is used to minimize the impact of the cut ends on descriptor calculation.

To generate desirable polymer structures in this work, a number of assumptions are made. The exact crosslinked structure of the example network is not known, but the degree of conversion is measured experimentally. As a starting point, it is assumed that all double bonds in the methacrylate monomers have equal reactivity, and that no intramolecular reactions occur during polymerization. While it is known that some level of intermolecular crosslinking does exist in the actual polymer networks, the degree of conversion and estimated crosslink density are used as structural descriptors, which are inexact descriptions of the electronic structure of a polymer network. As long as the intermolecular reactions are all of a similar level in the various experimental systems considered, the statistical nature of the correlations will allow the QSPR models to be reasonably accurate, even with inexact descriptors.

The following example shows how a structure is generated for a polymer consisting of 45 wt % HEMA and 55 wt% bisGMA. The degree of conversion of carbon–carbon double bonds measured experimentally for this case is 76.93%.

First, the percentage of each monomer state contained in the polymer network is calculated. The fraction of unreacted monomer is also calculated, but unreacted monomer is not part of the polymer network. Unreacted monomer is removed from the overall composition for the purpose of generating a polymer structure. The composition, in mole fraction, for the example above is 0.7236 HEMA in state B (Fig. 1), 0.0518 bisGMA in state B (Fig. 2), 0.0518 bisGMA in state C and 0.1728 bisGMA in state D. Probabilities based on the composition are used to generate a polymer graph describing the connectivity of the monomers. The graph is a tree because it was assumed that no intramolecular reactions occur; however, polymer graphs are not necessarily trees. The information included in the polymer graph, along with the degree of conversion, is then used to compute the estimated crosslink density and the other structural descriptors. The method for computation of the monomer is known, shown, and the weight fraction of each monomer in the mixture is known (and the weight fractions sum to one), the crosslink density may be computed via

$$CD = DC \sum_{i} \frac{w_i (nv_i - 1)}{MW_i} \quad (1)$$

where CD is the crosslink density, DC is the degree of conversion, w_i is the weight fraction of monomer *i*, nv_i is the number of vinyl groups in monomer *i*,MW_i is the molecular weight of monomer *i*, and *i* runs over all of the monomers present in the polymer network. The assumptions used here lead to the simplest method of generating polymer structure, while still providing sufficient information for the prediction of many important physical properties. Fig. 3 shows a small graph for this example.

Full graphs are large graphs which show the detailed chemical structure of a representative section of polymer. The full graph is obtained by combining the information in the monomer and polymer graphs, and can be used to calculate most topological structure descriptors in a straightforward way. While this graph provides a detailed chemical structure, information about the monomer connectivity provided in the polymer graph is difficult to extract. Fig. 4 shows a small example of a full graph for poly(HEMA).

The simplicity of using a fixed representative section of polymer greatly accelerates computation of structural descriptors needed for property prediction.

4. Calculation of structural descriptors

While there are many options in terms of numerical descriptors of chemical structure to be used to predict the properties of polymeric systems, we have chosen to focus on connectivity indices in this work. Connectivity indices have been used successfully in QSPRs and require relatively little computational effort (Kier & Hall, 1986; Bicerano, 2002). Of course, other descriptors are need for crosslinked polymers, some of which will be described in Section 7. Most topological descriptors can be easily calculated from the polymer structure information described in the previous section.

A number of graph theoretic algorithms have been implemented to aid in the calculation of topological descriptors. A subgraph isomorphism algorithm (Ullmann, 1976) is used to identify particular functional groups. Group identification is needed to calculate some types of topological descriptors such as group contributions. A block finding algorithm (Gibbons, 1985) is used (slightly modified) to distinguish the backbone from pendant groups. Backbone and pendant groups may have different effects on properties so it is often important to distinguish between them (Bicerano, 2002).

The data structures used in this work for polymer storage make descriptor calculation relatively easy. Since a representative section of polymer is generated, descriptors can be calculated as they would for any simple molecule. A feature of this system is the use of core and buffer sections of the polymer network. Calculation of descriptors often requires information about neighboring polymer sections; the buffer provides this information while calculating descriptors for the core section. The accuracy and repeatability of the descriptor calculations for random copolymers depend on the size of the representative section. Larger sizes cause the calculations to take longer. The time required for descriptor calculation in analysis of experimental polymer data is generally insignificant, since the number of polymers for which properties are measured is generally less than 500. However, descriptor calculation time may a be concern when using optimization to find new structures, since an optimization routine may evaluate thousands of polymer structures while searching for those likely to have desired properties.

The system for descriptor calculation employed in this work saves a significant amount of time in preliminary analysis of structural descriptors and property models. Many different descriptors can be evaluated without the need to deal with complex probability calculations. This allows for much flexibility in terns of the structural descriptors and property models to be used. Two simple examples of structural descriptor calculation are presented. The first example is the calculation of the zeroth-order connectivity index in which the core and buffer concept is not needed. The second example describes the calculation of the second-order connectivity index, which demonstrates the use of the core and buffer sections.

To calculate connectivity indices, each non-hydrogen atom is assigned a delta value based on its hybridization and the number of hydrogen atoms attached (Bicerano, 2002; Kier & Hall, 1986). The order of a connectivity index indicates the path length used to calculate the connectivity index. For example, a second-order connectivity index is calculated from all unique paths in the hydrogen-suppressed graph of length two. The equation for calculating the zeroth-order connectivity index is given by Eq. (2), and Eq. (3) gives the second-order equation. The indices i, j, and k label the three atoms in any path of length two.

$$\chi^0 = \sum_{i \in \text{vertices}} \frac{1}{\sqrt{\delta_i}} \quad (2)$$

$$\chi^2 = \sum_{ijk \in \text{ paths of length } 2} \frac{1}{\sqrt{\delta_i \delta_j \delta_k}} \quad (3)$$

The obvious problem when trying to calculate higher order connectivity indices of polymers is that many paths can extend beyond the defined repeat unit, so information on neighboring monomers is needed. The problem becomes even more complicated when calculating connectivity indices for a random polymer network.

To calculate the zeroth-order connectivity indices, only the structures of the monomers are needed. To simplify the example, we will consider a poly(HEMA) homopolymer. Fig. 4 shows the structure of poly(HEMA). Only the core is needed for the zeroth-order calculation. The zeroth-order connectivity index is 6.906.

The core and buffer concept can be used to calculate the second order connectivity index for poly(HEMA). Of course, this is more useful with a more complex structure, but this example is intended to illustrate the concept. Paths that extend into the buffer area can be included, but these paths are not entirely in the core. Paths that are only partly in the core are multiplied by the fraction of atoms in the core before being added to the sum. For example, the contribution of a path that has two atoms in the core and one in the buffer is multiplied by two-thirds when added to the sum as shown in Eq. (4). The second order connectivity index for poly(HEMA) is 3.874.

$$\chi^{2} = \sum_{ijk \in \text{ paths of length } 2} \frac{n_{ijk,\text{core}}}{3\sqrt{\delta_{i}\delta_{j}\delta_{k}}} \quad (4)$$

It is clear from the calculations the magnitudes of the connectivity indices depend on the size of the polymer section that is used to calculate them. To determine a connectivity index that is independent of size, the values are usually scaled by the number of non-hydrogen atoms in the section. Size-dependent connectivity indices are useful for polymers that have a unique repeat unit, when they can be used to predict extensive properties such as molar volume or cohesive energy (Bicerano, 2002). In this work, values corresponding to size-independent connectivity indices will be used to predict polymer properties such as tensile strength, modulus of elasticity, and glass transition temperature.

5. Quantitative structure property relations

To create a predictive model, numerical descriptors of structure are related to physical or chemical properties of the polymeric systems. While phenomenological models are more accurate in terms of property prediction, QSPRs based on structural descriptors allow rapid estimation of properties directly from the chemical structure, and are straightforward to include within the optimization framework for the design of novel polymer networks.

This work focuses mainly on linear regression as a means to develop property models due to the limited experimental data collected on the systems being studied. Nothing in the framework, however, prevents use of other types of models. Non-linear regression can be used to estimate parameters if there is reason to believe a property model has a particular

functional form that is non-linear with respect to the parameters. Partial least squares regression (PLS) is an effective method when a large number of descriptors are to be used (Oprea, 2005). This section provides a brief description of the method used to develop preliminary QSPRs for the example dental resin design problem explored in this work.

A small preliminary set of descriptors is selected including connectivity indices, degree of conversion, and crosslink density. Information about the particular descriptors will be provided in Section 7. After collecting experimental property data for a set of dental polymers, the structural descriptors are calculated.

Multiple linear regression is used to develop property models, and no transformations of the data are considered. It is possible to use linear regression to find polynomial, logarithmic or any other models where the parameters have a linear relationship to each other. To find the best sets of descriptors, the LEAPS package (Lumley, 2004) for the R statistical software (R Development Core Team, 2007) is used. Combinations of descriptors are obtained which provide the best correlation coefficient for QSPRs containing different numbers of descriptors. Once these models are found, it is still necessary to select those with the most appropriate number of descriptors.

There are several methods to determine the best number of descriptors to use in a model. Measures which weigh the quality of fit against the degrees of freedom, such as adjusted correlation coefficient, are often employed; these values tend to overestimate the number of descriptors that are useful. It is also valid to consider the statistical significance of the model and of the parameters. The most effective means of determining model size is to use cross-validation, in which part of the experimental data is left out when making the model (Efron and Tibshirani, 1993). Error is then measured for the predicted values of data not used to create the model. If too many descriptors are used, large errors occur in the predictions. The QSPRs used in this work will be described in Section 7.

6. Formulation and solution of the MILP problem

The goal of this CMD study is to determine new polymer structures which are likely to possess desirable properties for use in dental materials. To find these new polymers, the QSPRs are included in an optimization framework. Structural requirements, such as valency restrictions, are combined with the QSPR model to form the constraint set for this problem. The objective function seeks to minimize variation in the predicted properties from the user-specified target values. Eq. (5) shows a typical CMD optimization formulation (Siddhaye, Camarda, Topp, & Southard, 2000). *M* is the set of properties of interest, and *m* is a property in the set *M*. *P*_m is the predicted value, P_m^{scale} is a scaling factor, and P_m^{target} is the target for property *m*. The vector **y** is a set of structural descriptor values, and *f*_m is a QSPR for property *m*. The functions that define the structural descriptors are the vector **g**. The identity of group *i* is stored in *w*_{*i*}. The binary variable *a*_{*ijk*} indicates whether a bond of type *k* connects groups *i* and *j*. The functions *h*_c are constraints which ensure a feasible structure.

$$\min s = \sum_{m} \frac{1}{P_{m}^{\text{scale}}} |P_{m} - P_{m}^{\text{target}}|$$

$$P_{m} = f_{m}(y)$$

$$y = g(a_{ijk}, w_{i})$$

$$h_{c}(a_{ijk}, w_{i}) \ge 0$$
(5)

If the functions in Eq. (5) are linear, the optimization problem. Then becomes an MILP. Otherwise an MINLP results. Some instances of the CMD optimization problem have been

solved using deterministic approaches (Maranas, 1996; Sahinidis et al., 2003). In this work, a framework is described which can accept non-linear property prediction constraints and a non-linear objective function. Thus the problem is reformulated and solved via a heuristic approach (Tabu Search).

As discussed previously, many approaches have been used to solve specific instances of the CMD optimization problem. This work focuses on the Tabu Search method to allow for flexibility in the types of property models used. Tabu Search is a heuristic method, and does not necessarily find the global optimum to a problem. Due to the limited accuracy of QSPRs, a list of near optimal structures may be as practically useful as the global optimum.

Tabu Search methods vary in implementation, but they generally operate by generating a list of neighbor solutions similar to a current solution (Glover, 1990a, 1990b). A solution is Tabu if it is too close to a solution in a Tabu list, which contains some number of recent solutions. Comparing solutions will be discussed later in this section. The Tabu list prevents the solution from cycling around local optima, and encourages new areas to be explored. The best neighbor that is not Tabu becomes the new current solution. Tabu Search can also be repeated with different starting points to try to get a more diverse solution set. A record is kept of good solutions which are structures estimated to have acceptable properties. The list of good solutions can be used to determine which new structures to synthesize and test experimentally, thus accelerating the design process.

The implementation of Tabu Search in this work is somewhat different to what has been used previously for molecular design. The polymers are stored in the data structures described in a previous section. Neighboring structures are created by modifying monomer structures, or changing composition. For optimization, the monomers are built from several small groups of atoms. Structures of the monomers are modified by adding, deleting, or replacing groups. Sufficient groups are provided such that nearly any reasonable monomer can be made. Monomers are modified in such a way that only feasible structures will result. The method of structure modification eliminates the need for most formal mathematical constraints. Some constraints may still be required to maintain a valid resin composition.

To determine whether a monomer matches one on the Tabu list, the structural descriptors of a solution can be compared to solutions in the Tabu list. The structures cannot be compared just using the groups they contain. A given set of chemical groups or atoms can be arranged to form many different structural isomers. A certain amount of difference is allowed, so structures do not have to exactly match those on the list to be considered the same as a previous structure. The amount of difference allowed is an adjustable parameter. Solutions can be refined by allowing them to be close to those in the Tabu list, and diversification can be encouraged by requiring a large difference.

7. Example: design of polymethacrylate dental adhesives

Over two-thirds of restorative dentistry involves replacement of failed restorations (Murray et al., 2002). Clearly restorative dental materials with improved durability are needed. This example is preliminary work to describe our methodology for the design of novel monomers to improve the durability of dentin adhesives.

Properties of a small set of polymethacrylates were measured experimentally, and QSPRs were created by relating this data to topological indices and formulation. The properties include tensile strength, modulus of elasticity, glass transition temperature, initial polymerization rate, and degree of conversion. The details of the experimental methods used to measure these properties are described by Ye, Spencer, et al. (2007) and Ye, Wang, Williams, and Spencer (2007). Polymer samples were prepared in a consistent way. The

polymers were composed of 45 wt% HEMA, 30 wt% bisGMA and 25 wt% of a third monomer. Seven monomers were used as the third monomer: bisGMA, triethyleneglycol dimethacrylate (TEGDMA), urethane dimethacrylate (UDMA), bisphenol A polyethylene glycol diether dimethacrylate (bisEMA), polyethylene glycol dimethacrylate (PEGDMA), trimethylol-propane mono allyl ether dimethacrylate (TMPEDMA), and 1,1,1-tri-[4- (methacryloxyethylaminocarbonyloxy)-phenyl]ethane (MPE). TMPEDMA and MPE are newly synthesized monomers (Park et al., 2007). Figs. 5–10 show the structures of these monomers. Resins samples were cured without water and with 11 wt% water, so 14 total samples were used.

A short list of structural descriptors is compiled for use in development of QSPRs. The degree of conversion (DC) is used as a descriptor even though it is experimentally measured, since it is needed to characterize the crosslink density. Future work will attempt to predict DC based on the chemical structure of the monomers and curing procedure. The maximum crosslink density (CD_{max}) and estimated crosslink density (CD) as described by Cook et al. (2003) are also used. The zeroth- and first-order valence and simple connectivity ($\xi^0, \xi^0_{\nu}, \xi^1$, and ξ^1_{ν}) indices were used in their size independent form (Bicerano, 2002), since there are no obvious repeat units for the polymers in this example, and the properties of interest are intensive.

Using LEAPS (Lumley, 2004), the best combinations of descriptors were found. The appropriate number of descriptors to include in each model is selected by evaluating statistical significance. Since this is preliminary work and more data must be collected cross-validation was only performed with the tensile strength model, and this evaluation confirmed the model size choice. The property models are shown in Table 1.

These property models are then used to create an objective function for optimization. Target properties are selected which are expected to yield a more durable dental polymer. In cases where a property was to be maximized, a target near the upper limit of applicability of the property model was selected. The first row of Table 2 shows the target properties, and Eq. (6) provides the objective function.

$$\min s = \sum_{m} \left(\frac{P_m^{\text{target}} - P_m}{P_m^{\text{target}}} \right)^2 \quad (6)$$

The goal of the optimization problem was to design an improved monomer, while holding composition used in the experiments constant. In this example, solutions to the optimization problem are monomer structures. The composition and structure of the first two monomers is fixed. The monomer structures are constructed from small groups of atoms. The groups include: a methacrylate functional group, ring structures such as aromatic rings, and small basic groups. Sufficient groups are supplied such that almost any methacrylate monomer of interest can be generated. Groups can be added, deleted, or replaced to generate neighbor solutions to a current solution. These changes can only lead to new solutions which are feasible. Other CMD methods may also consider infeasible solutions. The number of functional groups is fixed by the starting solution, and variations are included by using starting points with different numbers of functional groups.

Several starting points are selected for Tabu Search, including currently used monomers such as bisGMA, as well as randomly generated monomers with different numbers of functional groups. The Tabu Search algorithm is run until a certain number of non-improving iterations are reached, in this case 300. A list of the best solutions found was compiled. Figs. 11–15 show five of the resulting structures. Table 2 shows objective

function values of the candidate monomers as compared to bisGMA, which is typically used in commercial dental adhesives. The solutions shown all have a better objective function than bisGMA. The results provide numerous structures, which can be further evaluated for use in polymethacrylate dentin adhesives.

The Tabu Search algorithm did not always yield the same solution, so numerous candidate monomers where found. While there is no guarantee that the globally optimal solution was found, each of these strong candidate monomers was generated in about 30 s on average. The Tabu Search algorithm was terminated after 300 non-improving iterations. The computation time depends on a number of Tabu Search parameters, including the length of the Tabu list, and further refinement of the algorithm parameters is expected to improve solution time and solution quality.

A CMD methodology for complex polymer systems has been implemented. With the framework in place, preliminary work for the design of dentin adhesives can progress rapidly as more experimental data is collected. With additional data, further evaluation of new structural descriptors and QSPRs, and the addition of more properties, the newly designed polymers are more and more likely to have practical value.

8. Conclusions and future work

A software tool, which implements the methodology described in this work, has been developed. The system employed for storing polymer structure can accommodate practically any type of polymer, and is particularly useful for random polymer networks. The structure and the core and buffer concept allow most topological structure descriptors to be calculated relatively easily.

The Tabu Search implementation provides considerable flexibility in the types of QSPRs and formulations that can be used with CMD. Tabu Search and the formulation used in this work also help manage the combinatorial increase in problem size when designing large molecules.

In future work with polymethacrylate dentin adhesives, a larger set of property data will be obtained. The expanded data set will provide several improvements to the QSPRs. A larger set of structural descriptors will also be examined. Non-linear QSPRs will be developed when needed using transformations of the descriptors. The larger variety of polymers will allow the QSPRs to be applicable to a wider range of polymers, and increase confidence in property predictions. Synthesis and testing of the monomers designed in the CMD process will provide valuable new data which will improve the accuracy of the QSPRs resulting in improved candidate monomers.

Work has also started in calculating several higher order connectivity indices, and treating backbone and pendant groups independently. This works leads to a large number of descriptors, and PLS regression will be used to create QSPRs based on these more accurate structural descriptors. Also new experiments will help to quantify the crosslinking structure of the polymers. With this information, an improved model for polymer property prediction will be devised. The inclusion of this model in within our optimization framework will lead to the creation of more realistic structures.

Acknowledgments

The authors gratefully acknowledge NIH/NIDCR grant DE14392 (PI: Spencer) and Honeywell Corporation for financial support. We would like to thank Sarah Shulda, Janine Einhellig, Natalia Davydova, and Harrison Davis for helpful advice and experimental assistance.

References

- Bicerano J, Sammler RL, Carrier CJ, Seitz JT. Correlation between glass transition temperature and chain structure for randomly crosslinked high polymers. Journal of Polymer Science. 1996; 34:2247.
- Bicerano, J. Prediction of polymer properties. 3rd ed.. Marcel Dekker; New York: 2002.
- Camarda KV, Maranas CD. Optimization in polymer design using connectivity indices. Industrial & Engineering Chemistry Research. 1999; 38:1884.
- Chavali S, Lin B, Miller DC, Camarda KV. Environmentally-benign transition-metal catalyst design using optimization techniques. Computers & Chemical Engineering. 2004; 28:605.
- Cook WD, Forsythe JS, Irawati N, Scott TF, Xia WZ. Cure kinetics and thermomechanical properties of thermally stable photopolymerized dimethacrylates. Journal of Applied Polymer Science. 2003; 90:3753.
- Dowsland KA. Nurse scheduling with Tabu Search and strategic oscillation. European Journal of Operational Research. 1998; 106:393.
- Dušek K, Gordon M, Ross-Murphy SB. Graphlike state of matter. 10. Cyclization and concentration of elastically active network chains in polymer networks. Macromolecules. 1978; 11:236.
- Eslick, JC.; Camarda, KV. Polyurethane design using stochastic optimization. In: Marquart, W.; Pantelides, C., editors. Proceedings of 9th international symposium on process systems engineering; Amsterdam: Elsevier; 2006. p. 769-774.
- Efron, B.; Tibshirani, RJ. An introduction to the bootstrap. Chapman & Hall; New York: 1993.
- Flory PL. Molecular size distribution in three dimensional polymers. I. Gelation. Journal of American Chemical Society. 1941a; 63:3083.
- Flory PL. Molecular size distribution in three dimensional polymers. II. Branching units. Journal of American Chemical Society. 1941b; 63:3091.
- Flory PL. Molecular size distribution in three dimensional polymers. III. Tetrafunctional branching units. Journal of American Chemical Society. 1941c; 63:3096.
- Fredenslund A, Jones RL, Praunsnitz JM. Group-contribution estimation of activity coefficients in nonideal liquid mixtures. AIChE Journal. 1975; 21:1086.
- Gani R, Tzouvaras N, Rasmussen P, Fredenslund A. Prediction of gas solubility and vapor-liquid equilibria by group contribution. Fluid Phase Equilibrium. 1989; 47:133.
- Gani R, Harper PM, Hostrup M. Automatic generation of missing groups through connectivity index for pure component property prediction. Industrial & Engineering Chemistry Research. 2005; 44:7262.
- Gani R, Fredenslund A. Computer-aided molecular design with specific property constraints. Fluid Phase Equilibrium. 1993; 82:39.
- Gendreau M, Laporte G, Semet F. A Tabu Search heuristic for the undirected selective traveling salesman problem. European Journal of Operational Research. 1998; 106:539.
- Gibbons, A. Algorithmic graph theory. Cambridge University Press; Cambridge: 1985.
- Glover F. Artificial intelligence, heuristic frameworks and Tabu Search. Managerial and Decision Economics. 1990a; 11:365.
- Glover F. Tabu Search: A tutorial. Interfaces. 1990b; 20:74.
- Gordon M. Good's theory of cascade processes applied to the statistics of polymer distributions. Proceeding of the Royal Society of London. 1962; 268:240.
- Joback KG, Reid RC. Estimation of pure-component properties from group contributions. Chemical Engineering Communications. 1987; 57:233.
- Kier, LB.; Hall, LH. Molecular connectivity in structure-activity analysis. Research Studies Press; Letchworth, England: 1986.
- Kimms A. Competitive methods for multi-level lot sizing and scheduling: Tabu Search and randomized regrets. International Journal of Production Research. 1996; 34:2279.
- Lin B, Miller DC. Tabu search algorithm for chemical process optimization. Computers & Chemical Engineering. 2004a; 28:2287.

- Lin B, Miller DC. Solving heat exchanger network synthesis problems with Tabu Search. Computers & Chemical Engineering. 2004b; 28:1451.
- Lin B, Chavali S, Camarda K, Miller DC. Computer-aided molecular design using Tabu Search. Computers & Chemical Engineering. 2005; 29:337.
- Lumley, T. The leaps package. 2004. cran.r-project.org/doc/packages/leaps.pdf
- Maranas CD. Optimal computer-aided molecular design: a polymer design case study. Industrial & Engineering Chemistry Research. 1996; 35:3403.
- Murray PE, Windsor LJ, Smyth TW, Hafez AA, Cox CF. Analysis of pulpal reactions to restorative procedures, materials, pulp capping, and future therapies. Critical Reviews in Oral Biology & Medicine. 2002; 13:509. [PubMed: 12499243]
- Oprea, TL, editor. Chemoinformatics in drug discovery. Wiley-VCH; Weinheim: 2005.
- Park J, Ye Q, Topp EM, Kostoryz EL, Wang Y, Kieweg SL, Spencer P. Preparation and properties of novel dentin adhesives with esterase resistance. Journal of Applied Polymer Science. 2007; 107:3588. [PubMed: 22919119]
- Porter, D. Group interaction modeling of polymer properties. Marcel Dekker; New York: 1995.
- R Development Core Team. R: A language and environment for statistical computing. 2007. http:// www.R-project.org
- Randi M. On the characterization of molecular branching. Journal of American Chemical Society. 1975; 97:6609.
- Sahinidis NV, Tawarmalani M, Yu M. Design on alternative refrigerants via global optimization. AIChE Journal. 2003; 49:1761.
- Siddhaye S, Camarda KV, Topp E, Southard MZ. Design of novel pharmaceutical products via combinatorial optimization. Computers & Chemical Engineering. 2000; 24:701.
- Stepto RFT, Cail JI, Taylor DJR. Predicting the formation, structure and elastomeric properties of endlinked polymer networks. Macromolecular Symposia. 2000; 159:163.
- Ullmann JR. An algorithm for subgraph isomorphism. Journal of the Association of Computer Machinery. 1976; 23:31.
- Van Krevelen, DW. Properties of polymers. Elsevier; Amsterdam: 1997.
- Venkatasubramanian V, Chan K, Caruthers JM. Computer aided molecular design using genetic algorithms. Computers & Chemical Engineering. 1994; 18:833.
- West, DB. Introduction to graph theory. Prentice Hall; Upper Saddle River: 2001.
- Ye Q, Spencer P, Wang Y, Misra A. Relationship of solvent to the photopolymerization process, properties, and structure in model dentin adhesives. Journal of Biomedical Materials Research Part A. 2007; 80:342. [PubMed: 17001655]
- Ye Q, Wang Y, Williams, Spencer P. Characterization of photopolymerization of dentin adhesives as a function of light source and irradiance. Journal of Biomedical Materials Research Part B. 2007; 80:440.
- Zhao, H.; Ralston, JP.; Middaugh, RC.; Camarda, KV. Application of computational molecular design to gene delivery polymers. Proceedings of foundations of computer-aided process design; Austin, Texas: Computer Aids for Chemical Engineering Education; 2004. p. 415-418.



Fig. 1. HEMA monomer graph states.











Fig. 4. Full graph for poly(HEMA).





Fig. 5. Structure of TEGDMA monomer.



Fig. 6. Structure of UDMA monomer.







Fig. 8. Structure of PEGDMA monomer.







Fig. 10. Structure of MPE monomer.





Fig. 11. Candidate monomer 1.



Fig. 12. Candidate monomer 2.



Fig. 13. Candidate monomer 3.



Fig. 14. Candidate monomer 4.



Fig. 15. Candidate monomer 5.

Table 1

QSPRs.

Property model	R ²
Tensile strength, $\sigma = 1406.6 - 7484.5\ ^1\xi + 6611.6\ ^1\xi^v + 78231.7 CD_{max} - 149268.6 CD$	0.94
Modulus, $E = 257.362 - 135.89 \ {}^{0}\xi - 276.37 \ {}^{1}\xi - 78.24 \ {}^{1}\xi^{v} + 0.02336DC - 0.03146WC$	0.97
Glass transition temperature, $T_{\rm g} = 11664.9 - 14036.5 {}^{1}\xi - 15286.9 {}^{1}\xi^{\rm V} + 94671.6 {\rm CD}$	0.82
Initial polymerization rate, IPR =–4028.55 + 6510.10 $^0\xi-2394.13$ $^1\xi^{\rm v}$	0.82

Table 2

Objective function values.

Monomer	σ (MPa)	E (GPa)	IPR (mol/L/s)	$T_{\rm g}(^{\circ}{ m C})$	Objective
Target	80	3.0	130	120	0.000
bisGMA (exp.)	70	1.2	96	119	0.497
bisGMA (pred.)	68	1.0	107	115	0.444
Candidate 1	79	1.4	98	132	0.372
Candidate 2	77	1.3	105	134	0.365
Candidate 3	88	1.7	119	148	0.262
Candidate 4	85	1.3	116	143	0.367
Candidate 5	86	1.4	121	134	0.299