APPLICATION OF COMPUTATIONAL MOLECULAR DESIGN TO GENE DELIVERY POLYMERS

Hui Zhao[1], John P. Ralston[2], Russell C. Middaugh[3] and Kyle V. Camarda[1,*]

[1]Department of Chemical and Petroleum Engineering
University of Kansas
Lawrence, Kansas 66045

[2]Department of Physics and Astronomy
University of Kansas
Lawrence, Kansas 66045

[3]Department of Pharmaceutical Chemistry
University of Kansas
Lawrence, Kansas 66045

Abstract

This paper describes the use of optimization techniques to design non-viral gene delivery vectors. In gene therapy, DNA is transferred to cells by first complexing with a delivery vector, such as a cationic lipid or polymer. These synthetic vectors have been largely developed to mimic viral penetration into cells. However, the efficiency of non-viral gene delivery systems is significantly lower than that of viral delivery systems. For successful gene therapy, the non-viral gene delivery vector must satisfy a set of requirements on physical and biological property values. In this work, physical properties are correlated with topological descriptors, which are computed directly from molecular structure. Higher-level biological properties, such as transfection efficiency, are correlated by a phase diagram approach, in which regions of molecular structures are found which have similar values of physical and biochemical properties. The phase diagram provides a set of black box constraints can be included in the molecular design problem. The optimization problem, which includes an objective function minimizing deviation from a set of property targets and constraints defining property prediction algorithms and enforcing basic structural requirements, is a large MINLP problem with black-box constraints. An example of such a design problem is presented here, which has been solved using the stochastic algorithm Tabu Search.

Keywords

Tabu Search, Molecular Design, Gene Delivery

Introduction

A large portion of the research currently performed today in the chemical and biomedical industries is devoted to product design, that is, the search for new materials with specifically tailored properties. In the context of nonviral gene delivery, researchers are searching for more effective polymers or lipids which can effectively transfer DNA into cell nuclei without harmful side effects. This research is quite tedious due to the trial-and-error approach employed in most cases. Due to the high cost of developing novel synthesis procedures, most new designs for such molecules tend to have similar structures to those previously used. In order to discover new products which

are far different from those currently used, an efficient computational screening procedure is required. Two major challenges arise in the development of such a computational molecular design procedure: the ability to predict the physical and chemical properties of a given molecule, and the ability to solve the large optimization problem which is derived from the search for the best molecule for a given application. Research in molecular characterization has reached a point at which many physical and chemical properties can be predicted to a reasonable accuracy for common molecules. However biological properties of large molecules such as polymers or proteins are very difficult to predict. To address this challenge, we have begun to employ a phase diagram approach predict higher biological properties, such as the percentage of a given type of cells transfected with external DNA, which is known as the transfection efficiency. The use of a phase diagram approach to predict complex biological properties causes the inclusion of black-box constraints into the optimization formulation. While most deterministic algorithms require explicit expressions for all constraints, stochastic algorithms such as Tabu search only require function evaluations for these constraints, and thus we have used this algorithm to solve the MINLP problems which result from the molecular design of nonviral gene delivery vectors.

The nonviral gene delivery vectors designed in this work are specifically designed to carry engineered plasmid DNA, which is often used as a anti-cancer therapy. In this case, the DNA reprograms cells to prevent angiogenesis, which starves the tumor of blood supply and hampers tumor growth. Viral gene delivery is clearly an effective technique from the point of view of transfection efficiency, but the side effects, namely a strong immune response, make this approach risky. The use of nonviral gene delivery vectors, such as polymers or lipids, has been shown to generate fewer side effects, but transfection efficiency tends to be low. In this case, the gene delivery vehicle encapsulates the DNA in aqueous solution in an initial complexation step. After injection, the complex enters the cell through the formation of an endosome. The complex then leaves the endosome via a process known as endosomal release, and subsequently the DNA leaves the complex and is transported to the nucleus, where it crosses the nuclear membrane and becomes incorporated into the cell's DNA. The transfection efficiency of such a complex is influenced by physical properties, such as complex size, charge density and solubility, and biological properties, such as the rate of endosomal release. While the rate-limiting step of the process is not yet known, it has been suggested by some authors that the rate of endosomal release is critical, and can be related directly to complex size and charge density.

As a first step into the use of computational molecular design in the area of gene delivery, this work uses structural constraints to generate a polymer known to have a reasonable transfection efficiency for plasmid DNA, and then optimizes to improve values of physical properties such as the critical micelle concentration (which defines the amount of polymer required per unit DNA to induce complexation) and solubility parameter. The target values for each of these properties are prespecified.

Physical Property Prediction

Bicerano (1996) has correlated a large number of physical properties of polymers with topological indices, which are numerical descriptors based on the electronic structure of the atoms within a molecule, as well as on the interconnectivity of the atoms within that molecule. In this work, Randic's molecular connectivity indices are used to generate structure-property correlations. They provide a quantitative assessment of the degree of branching of molecules, and are based on a set of basic groups, which are defined as functional groups containing one non-hydrogen atom in a specific valence state and a given number of hydrogens. The nth order simple and valence molecular connectivity indices are given by

$$ {}^{n}x = \sum_{(i_0,i_1\ldots i_n)\in\varepsilon_n} \frac{1}{\sqrt{\delta_{i_0}\delta_{i_1}\ldots\delta_{i_n}}} \qquad (1) $$

$$ {}^{n}x^{v} = \sum_{(i_0,i_1\ldots i_n)\in\varepsilon_n} \frac{1}{\sqrt{\delta^{v}_{i_0}\delta^{v}_{i_1}\ldots\delta^{v}_{i_n}}} \qquad (2) $$

where ε_n is the edge set of n consecutive bonds between basic groups in a molecule, $i_0,i_1,\ldots,i_n$ denote the n+1 basic groups forming the n consecutive bonds, δ_{i_k} ,k=0,...,n are the simple atomic connectivity indices for those basic groups (the number of bonds each group can form), and $\delta^{v}_{i_k}$,k=0,…,n are the atomic valency connectivity indices, which are based on the electronic structure of the basic group. Table 1 shows the basic groups employed in the example in this paper, along with their atomic connectivity indices.

These topological indices can then be correlated with various physical properties of polymers, given a set of experimental data for known polymers. In this work, it was determined that complex solubility was closely related to the solubility of the gene delivery polymer itself, and thus the correlation of Bicerano (1996) were used for solubility parameter. This correlation is based on cohesive energy and molar volume, and thus correlations for these two polymer properties were also used from this source. For the critical micelle concentration, a new correlation was developed using data from Rosen (1984). These correlations are included along with structural constraints to ensure that a reasonable molecular structure is generated as the solution to our MINLP molecular design problem.

Table 1: Basic groups and their atomic and valence connectivity indices

	δ	δ^v		δ	δ^v
>C<	4	4	—CH_3	1	1
—CH—	3	3	—OH	1	5
=C—	3	4	—NH_2	1	3
—CH_2	2	2	—NH	2	4
=CH	2	4			

In the correlations of Bicerano (1996) and the new correlation developed in this work, zeroth- and first-order connectivity indices are used within the correlations. The correlations employed to predict the physical properties listed above are as follows:

$$V = 2.29\ {}^0\chi + 17.14\ {}^0\chi^v - 6.80 \quad (3)$$

$$E_{coh} = 9882.5\ {}^1\chi + 25100 \quad (4)$$

$$SC = \sqrt{E_{coh}/V} \quad (5)$$

$$\log(\mathrm{CMC}) = \frac{100}{N}(-1.063\ {}^0\chi + 0.327\ {}^0\chi^v + 1.39\ {}^1\chi - 0.436\ {}^1\chi^v + 0.828) \quad (6)$$

where V is the molar volume (cc/mol), E_{coh} is the cohesive energy density (J/mol), SC $(J/cc)^{0.5}$ is the solubility coefficient and CMC is the critical micelle concentration (molar). The connectivity indices are structural descriptors, which can successfully describe many physical and chemical properties with high accuracy, including solubility and CMC, which are used in our objective function. However, these descriptors alone do not contain sufficient information to predict highly complex biological properties such as transfection efficiency. This problem will be addressed later in this paper.

Problem formulation

In order to search for a nonviral gene delivery vector matching our property target, an optimization problem is constructed. The objective function is to minimize the scaled difference between the physical property values of the candidate molecule and targets:

$$\text{Min} \quad s = \sum_m \frac{1}{P_m^{\text{scale}}} \left| P_m - P_m^{\text{target}} \right| \quad (7)$$

where s is the total scaled property deviation, P_m is the property value of a feasible candidate solution and P_m^{target} is the target value for the mth property. P_m^{scale} is used to control the relative importance of the properties, and is equal to P_m^{target} in the example in this paper. Along with the property prediction equations and the expressions defining the molecular connectivity indices, structural constraints are also included to ensure that a stable, connected molecule is formed. These include valency and uniqueness constraints, which ensure that the valency of each atom is satisfied and two groups may only bond with one type of bond (single or double), as well as connectedness constraints, which guarantee that all the basic groups within the molecule are bonded into one coherent molecule.

In order to store the molecular candidates computationally, a data structure is used which employs binary variables to define whether two basic groups i and j are bonded with a kth multiplicity bond. These binaries form a partitioned adjacency matrix which is structured such that the identity of each basic group is known at each position within the molecule. The constraints and objective function form a nonconvex MINLP problem which when solved to optimality results in a feasible structure which is predicted to match the target properties as closely as possible.

Tabu search

Tabu search, developed by Glover (1986), is a stochastic algorithm which uses flexible memory to record recency, frequency, quality and influence of randomly generated candidate solutions. As a stochastic algorithm, Tabu search cannot guarantee the determination of a global optimum, but it has been shown to find near-optimal solutions to molecular design problems which could not be solved using deterministic approaches (Chavali, *et al.*, 2002). The use of Tabu search within the work also allows the inclusion of black box constraints, such as those which arise when a phase diagram approach is used to predict biological properties.

Phase Diagram Approach

It is known from many experiments in gene delivery that numerous factors influence the transfection efficiency of a given nonviral delivery agent. These include both structural factors of the agent itself, as well as formulation conditions and perhaps the specific type of DNA being delivered. Since all of these factors are difficult to combine within a single structure-property relationship, a phase diagram approach has been developed (Keultzo,

2001). These phase diagrams are similar to traditional thermodynamic phase diagrams in that different regions correspond to sets of coherent property values. Rather than being coherent material properties, however, these regions correspond to high or low regions of a given property of the system, in our case transfection efficiency. The parameters defining the regions are the physical properties of the gene delivery agent being evaluated. Thus such a diagram (in numerical form) can be used within a molecular design algorithm by first finding values of physical properties using simpler structure-property relations, and then inputting these properties along with formulation conditions into a numerical phase diagram. The use of a stochastic algorithm to solve the nonconvex MINLPs in this work serves to test the algorithm on a simple gene delivery vehicle design problem, such that the methodology could be employed in conjunction with a phase diagram approach for property prediction.

Results

The methodology was tested using an example set of physical property targets, as shown in Table 2. These property values are typical for an effective gene delivery vehicle.

Table 2: Properties, target values and predicted values for the optimal solution to the example problem

Property	Target value	Best Solution Value
CMC (M)	$9.11*10^{-4}$	$9.51*10^{-4}$
$SC(J/cc)^{0.5}$	27	25.52

Requirements were set on the number of carbons present, so that a polymer repeat unit would be generated which could be useful as a gene delivery vehicle. 100 runs of the Tabu search algorithm were performed, and the CPU time required for this as well as the best objective function value found are listed in Table 3.

Table 3: Statistics for best solution of example problem

Objective function	0.16
Total runs	100
CPU time	22.62 s

A list of near-optimal solutions were found and stored, which could be later evaluated to determine ease of synthesis. One of the top 10 structures designed and stored in the candidate list is presented in Figure 1.

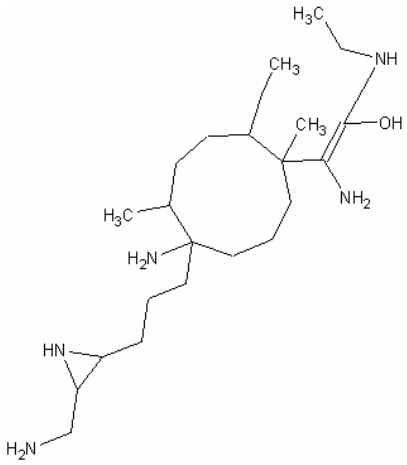

Figure 1: Example monomer unit designed using optimization methodology

Note the methodology in its current form does not provide information on how the monomer unit is to be synthesized or polymerized, but it does give a researcher a short list of potential structures which are likely to match the desired physical property targets. This method may also be used to improve known polymer structures, In this case, one simply fixes certain groups within the molecule, and uses the optimization algorithm to suggest useful modifications of a given structure. The resulting optimization problems have fewer binary variables than a full design problem, and thus a larger number of modifications may be considered.

Conclusions

In this paper, a molecular design problem for nonviral gene delivery vehicles has been recast as an integer optimization problem, and is solved by the Tabu search algorithm. Phase diagrams are introduced as a method to correlate complex biological properties within a computational molecular design framework. An example is given which designs a novel gene delivery polymer matching target values of two physical properties.

References

Bicerano, J. (1996), Prediction of Polymer Properties, Marcel Dekker, New York.

Chavali, S., B. Lin, T.J. Huismann, D. C. Miller and K. V. Camarda (2002), An Optimization Technique for the Design of Transition-Metal Catalysts, AIChE Annual Meeting, Indianapolis, Indiana.

Glover, F. (1986), Future paths for Integer Programming and Links to Artificial Intelligence," *Comp. and Oper. Res.*, **5**, 533-549.

Kueltzo, L. A., (2001), An Investigation into the Role of Molten Globule-Like States in Non-Classical Protein Transport, Ph.D. Dissertation, University of Kansas

Rosen, M. J. (1978), Surfactants and Interfacial Phenomena, John Wiley and Sons, New York.