Computer Aided Molecular Design of DNA:Polymer Complexes using Property Correlations and 3-D Phase Diagrams

Elizabeth Morel¹ (Presenter), Hui Zhao¹, C. Russell Middaugh², and Kyle V. Camarda¹
¹Department of Chemical and Petroleum Engineering, ²Department of Pharmaceutical Chemistry
University of Kansas, Lawrence, Kansas

Background and Definitions

- **Computational Molecular Design** is a methodology to find molecule candidates for particular applications by using optimization techniques. This method speeds up the development of novel molecules and reduces experimental costs. Molecular structure information is used to generate correlations for physical and chemical properties. The methodology has been successfully used to design solvents¹, polymers²,³ and pharmaceuticals⁴.

- **Non-Viral Gene Delivery** is a technology to insert DNA into cells to generate therapeutic effects². Lipids or polymers are preferred versus viral agents to reduce immunogenic issues and toxicity. However, the transfection efficiency of non-viral vectors is often not as high as for viral vectors.

- Transfection Efficiency defines the capability of gene delivery vehicles to transfer plasmid DNA into the nucleus. It is measured using fluorescence techniques.

- Tabu Search⁶ is a stochastic optimization algorithm which guides a local search procedure to explore the solution space and escape local optima. This method can handle non-convex or black-box constraints and is used in this work to solve MINLP problems for gene delivery vector design.

- Phase Diagrams⁷ are used to relate complex biophysical properties to coherent regions in physical parameter space. Transfection efficiency is predicted in this work by using phase diagrams.

Objectives/Scope

- Calculate structural information for current non-viral gene delivery vehicles.
- Correlate relevant physical properties of non-viral gene delivery vehicles with structural descriptors.
- Construct phase diagrams to predict biophysical properties of non-viral gene delivery vehicles.
- Formulate optimization problems with phase diagrams, property correlations, and structural constraints.
- Predict improved gene delivery vector candidates with high transfection efficiency using the Tabu search algorithm.

DNA Complex Chemistry

- Peptoids and lipitoids possess physical, chemical and synthetic characteristics that make them unique candidates for gene delivery: positive charge for binding with DNA, appropriate zeta potentials, weight, etc.
- Experimental data, like FITR peak shifts, polymer length and molar ratios of polymer to DNA (CR), for current non-viral gene delivery vehicles, like cationic lipids and helper lipids, are used to generate structure property correlations.
- Data for a series of peptoid (N-substituted polyglycine) and lipitoid (peptoid-phospholipid conjugate):DNA complexes are used. Neutral lipids can increase transfection efficiency relative to cationic lipids alone. Following are examples of a cationic lipid and a helper lipid

Figure 1: Cationic Lipid and Helper Lipid Example

- Many physical properties can be correlated with topological descriptors, as in simple biophysical properties.

Figure 2: Methodology of Computational Molecular Design

- Molecular structure information is stored numerically using topological descriptors. Topological descriptors characterize a compound using a small set of numbers based on the types of atoms in the molecule and on its interconnectivity. In this work, connectivity indices and Kier’s shape indices are used to store structural information.

- Connectivity indices, Z, provide a quantitative assessment of the degree of branching of molecules.

- Kier’s shape indices, K, are used to characterize the shape of a molecule quantitatively. Index K₁ quantifies the cyclicity of a molecule. Index K₂ quantifies the star-like attributes of a molecule. Index K₃ quantifies the location in the chain where branching occurs.

- If the properties are extensive variables, they are correlated directly with topological descriptors. If the properties are intensive variables, they are correlated with modified topological descriptors which are the values of topological descriptors divided by the number of functional groups.

- In this work, the following properties are compiled with connectivity indices and Kier’s shape indices: ζ and Ethidium Bromide Relative Fluorescence Intensity.

- The Tabu search⁶ algorithm is used to solve the optimization problem formulated above, and provides a list of candidates for non-viral gene delivery vehicles.

- The advantages of this method for solving molecular design MINLPs are that complex non-convex constraints may be included in the formulation, and the algorithm is highly parallelizable.

- This algorithm does not guarantee globally optimal solutions, but since a hard lower bound exist on the solutions, the candidate molecules may be evaluated objectively.

Correlations

- Two correlations were obtained from experimental data for non-viral gene delivery complexes.

\[
\zeta = \text{potential} - 0.57732 - 0.528 \cdot \text{CR} + 0.064 \cdot \text{CR}^2
\]

The data used to obtain the \( \zeta = \text{potential} \) correlation included 15 DNA complexes, constructed from DNA and 7 peptides, 2 lipids, 2 cationic lipids, and 2 helper lipids. Different charge ratios are used for different complexes, leading to 99 data points.

The statistical software SPSS was used to linearly regress \( \zeta = \text{potential} \) data by a stepwise model.

\[
R^2 = 0.941
\]

The data used for the ETBR correlation include 13 DNA complexes, constructed from DNA and 7 peptides, 2 lipids, 2 cationic lipids and 2 helper lipids. Different charge ratios were used for different complexes, leading to 116 data points.

The statistical software SPSS was used to linearly regress ETBR data by using a stepwise model.

\[
R^2 = 0.883
\]

- Helper lipids are neutral lipids, used to improve the transfection efficiency of cationic lipids. The lipids are combined together as gene delivery vehicles.
- In a stepwise model, all the variables must pass the tolerance criterion to be entered in the correlation, regardless of the entry method specified. The default tolerance level is 0.0001. Also, a variable is not entered if it would cause the tolerance of another variable already in the model to drop below the tolerance criterion.
Computer Aided Molecular Design of DNA:Polymer Complexes using Property Correlations and 3-D Phase Diagrams

Elizabeth Morel\(^1\) (Presenter), Hui Zhao\(^1\), C. Russell Middaugh\(^2\), and Kyle V. Camarda\(^1\)

\(^1\)Department of Chemical and Petroleum Engineering, \(^2\)Department of Pharmaceutical Chemistry

University of Kansas, Lawrence, Kansas

**Results & Discussion**

**Phase Diagrams**

- There are two different sets of dimensions for the phase diagram: parameter dimensions and variable dimensions.
- Parameters are used to define and constrain variables in the optimization problem. Variables are the properties that are important in designing effective gene delivery polymers and lipids.
- In this work, we use CR, \(\kappa^2\), and \(N^2\) as parameters to correlate \(\zeta\) - potential and ETBR which are variables. This combination of parameters creates three parameter phase diagrams which are shown in three dimensions below.

![Figure 3. 3D Phase Diagram - Correlated](image)

- From the 3D phase diagram, four coherent regions can be identified: orange, light orange, green, and light green.
- The 3D phase diagram in Fig. 3 shows different slices starting at a CR = 0 and continuing to CR = 1.5 in 0.5 intervals. One can see the distinct regions as CR increases as the green region quickly fades into orange. Also, the green region is more prevalent on the right hand side of the graph (\(\zeta^2 < 0.405\)).
- Transparency in the slices is to help with viewing.

![Figure 4. Phase Diagram - Correlated](image)

- The 3D phase diagram in Fig. 4 shows a different view of the phase diagram with slices of \(\kappa^2 = 0, 0.35, \) and \(0.75\). Once again, there are distinct orange and green regions, but a new shape appears.

**Tabu Search and Phase Diagrams**

- The regions of the phase diagram correspond to similar degrees of transfection ability.
- Each region in the phase diagram has different zeta-potential and ETBR values which are used in the molecular design problem objective function. Different molecules are produced using target values for each region. Two of these molecules are as follows:

![Figure 5. Generated Molecule from Light Green Region of Phase Diagram](image)

![Figure 6. Generated Molecule from Light Orange Region of Phase Diagram](image)

- The molecules in Fig. 5 and Fig. 6 were created using ChemSketch as the results of the optimization, and are not shown in the appropriate geometry.
- Light turquoise atoms are carbon atoms, dark blue atoms are nitrogen atoms, and red atoms are oxygen atoms.
- The objective function value for Figure 5 was 0.44 and was generated in 17 seconds.
- The objective function value for Figure 6 was 0.013 and was generated in 16 seconds.

**Conclusions and Future Work**

- Correlations for several properties of non-viral gene delivery polymers have been created using structural descriptors.
- Phase diagrams are employed to incorporate data from different sources and used as an objective function to compute the value of transfection efficiency for these complexes.
- The Tabu search algorithm has been used to solve the resulting optimization problems to design novel non-viral gene delivery candidates.
- Further transfection experiments in different coherent regions of the phase diagram will be undertaken to label the phase diagrams more accurately.

**References**


**Acknowledgements:**

Dr. Brian Lobo, Dr. Sirirat Choosakoonkriang and Dr. Christopher Wetterhoff provided experimental data on complexes. Chad Braun, Haibing Fan and Dr. Joe Vetro provided consistent assistance for experiments and pharmaceuticals background. Sarah Thompson and Sarah Hordale assisted in performing the experiments and in generating phase diagrams.

Funding provided by the Kimberly-Clark Corporation, the Monsanto Corporation, National Science Foundation and the University of Kansas.