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

Background and Definitions

- Computational Molecular Design is a methodology to find molecular 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 solvents[1], polymers[2,3] and pharmaceuticals[4].
- Non-Viral Gene Delivery is a technology to insert DNA into cells to generate therapeutic effects[5]. Lipids or polymers are preferred versus viral agents to reduce immunogenic issues and toxicity. However, the transfection efficiency of non-viral vectors are 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 Search[6] 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 Diagrams[7] 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.

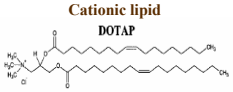
Objectives/Scope

- 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 FTIR peak shifts, polymer length and molar ratios of polymer to DNA (CR), for current non-viral gene delivery vehicles candidates, like cationic lipids and helper lipids, are used to generate structure property correlations.
- Data for a series of peptoid (N-substituted polyglycine):DNA 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

Cationic lipid
DOTAP



Helper lipid
Cholesterol

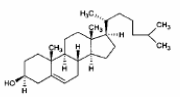


Figure 1: Cationic Lipid and Helper Lipid Example

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

Optimization Approach

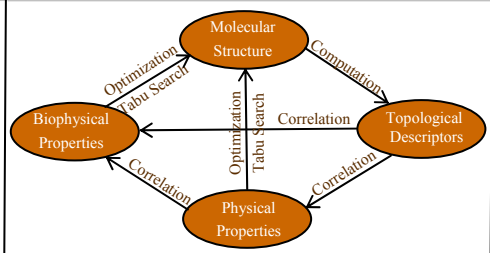


Figure 2: Methodology of Computational Molecular Design

- 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, χ, provide a quantitative assessment of the degree of branching of molecules[8].
- Kier's shape indices, κ, are used to characterize the shape of a molecule quantitatively. Index ${}^{1}\kappa$ quantifies the cyclicity of a molecule. Index ${}^{2}\kappa$ quantifies the star-like attributes of a molecule. Index ${}^{3}\kappa$ quantifies the location in the chain where branching occurs[9].
- 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 correlated with connectivity indices and Kier's shape indices. ${}^{2}\chi/{}^{3}\kappa_N$

ζ – potential	The **ζ – potential** is the potential difference between the shear radius and the bulk solution. It is often used as a measure of colloidal stability since the greater the **ζ – potential** of a population of particles, the less likely they are to aggregate because of increased repulsive forces.[10]
Ethidium Bromide Relative Fluorescence Intensity	Assays monitoring the displacement of ETBR or other intercalating dyes from DNA by various cationic delivery systems are routinely used to verify the condensed state of DNA in these complexes[5].

- Maximizing transfection efficiency is the ultimate goal for gene delivery vector design. Unfortunately, it is nearly impossible to correlate transfection efficiency directly with topological descriptors and simple physical properties.
- The phase diagram[7] method is used here to correlate data from a wide variety of methods, including techniques such as circular dichroism (CD), fluorescence, FTIR and Raman analyses, as well as calorimetric and hydrodynamic techniques.
- By "phase diagram" we refer to an empirical determination of regions of parameters, such as pH and temperature, in which the structural data are substantially uniform and coherent.
- In this work, charge ratio (CR) and ${}^{2}\chi/{}^{3}\kappa_N$ (a modified connectivity index) are used as parameters for phase diagrams. These phase diagrams are constructed using two critical variables, **ζ – potential** and Ethidium Bromide (ETBR) Relative Florescence Intensity, and they show different coherent regions which correspond to different levels of transfection efficiency, determined by using experimental data.
- After creating phase diagrams to predict transfection efficiency, the optimization problem is formulated as follows.

Optimization Approach

Objective Function	$\text{Min } s = \sum_m \frac{1}{P_m^{\text{scale}}} \left\| P_m - P_m^{\text{target}} \right\|$
Property Constraints	**ζ – potential** and Ethidium Bromide Relative Fluorescence Intensity
Connectivity Constraints	All the atoms of candidate molecules must be connected with each other
Valency Constraints	Valency for all atoms must be satisfied

- The Tabu search[11] 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.

ζ – potential	$\zeta - \text{potential} = -197.732 - 124.520 \cdot CR + 114.037 \cdot CR^2 - 22.78 \cdot CR^3 + 497.348 \cdot {}^{2}\chi/{}^{3}\kappa_N - 32.287 \cdot {}^{3}\kappa/{}^{3}\kappa_N$
	The data used to obtain the **ζ – potential** correlation included 15 DNA complexes, constructed from DNA and 7 peptoids, 2 lipitoids, 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 **ζ – potential** data by using a stepwise model. $R^2 = 0.941$
Ethidium Bromide Relative Florescence Intensity	$ETBR = -0.712 - 0.522 \cdot CR + 0.077 \cdot CR^2 + 4.061 \cdot {}^{2}\chi/{}^{3}\kappa_N + 0.303 \cdot {}^{3}\kappa/{}^{3}\kappa_N$
	The data used for the **ETBR** correlation include 13 DNA complexes, constructed from DNA and 7 peptoids, 2 lipitoids, 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, $^{2}\zeta/_{N}$, and $^{3}\kappa/_{N}$ as parameters to correlate ζ – **potential** and ETBR which are variables. This combination of parameters creates three parameter phase diagrams which are shown in three dimensions below.

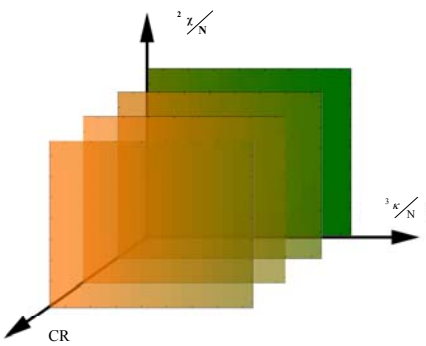


Figure 3. 3D Phase Diagram - Correlated

- 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 ($^{3}\kappa/_{N}$ > 0.405).
- Transparency in the slices is to help with viewing.

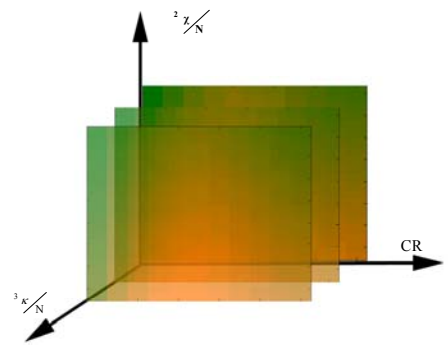


Figure 5. Phase Diagram - Correlated

- The 3D phase diagram in Fig. 4 shows a different view of the phase diagram with slices of $^{3}\kappa/_{N}$ = 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:

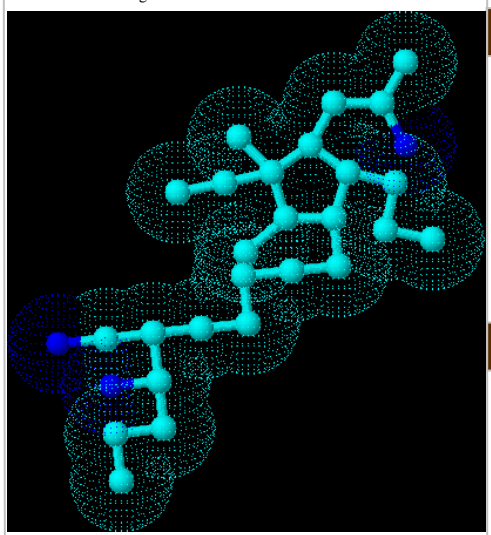


Figure 5. Generated Molecule from Light Green Region of Phase Diagram

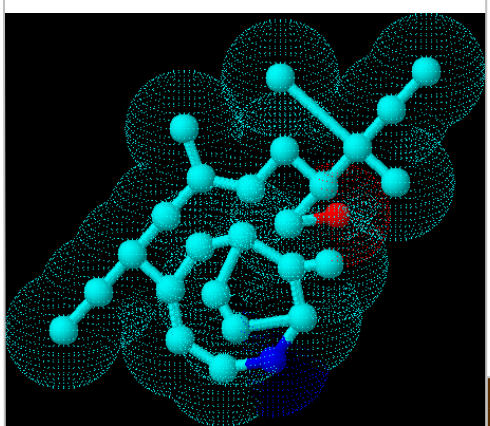


Figure 6. Generated Molecule from Light Orange Region of Phase Diagram

- The transfection efficiency of the molecules generated by our optimization approach will be tested experimentally. With those results, different regions of the phase diagram can be related to different transfection efficiencies
- 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

1. Gani, R., Tzouvars, N., Rasmussen,P. and Fredenslund, A., *Fluid Phase Equil.*, **47**, 133, 1989
2. Camarda, K.V., and Maranas, C.D., *Ind. Eng. Chem. Res.*, **28**, 1884, 1998
3. Venkatasubramanian, V., Chan K., and Caruthers, J.M., *Comp. Chem, Eng.*, **18**, 883, 1994
4. Siddhaye, S., Camarda, K.V., Topp, E., and Southard, M..Z., *Comp. Chem. Eng.*, **24,** 701, 2000
5. Lobo, Brian A., Ph.D. Thesis, University of Kansas, 2002
6. Glover, F. and Laguna, M., Tabu Search, Kluwer Academic Publishers, Boston, 1997
7. Kueltzo, L. A., Ersoy B., Ralston, J.P., and Middaugh, C.R.,, *J. Pharm. Sci.* **92**, 1805, 2003
8. Bicerano, J., Prediction of Polymer Properties, Marcel Dekker, New York, 2002
9. Raman, V. S. and Maranas C. D.. *Comp. Chem. Eng.*, **22**, 747, 1998
10. Wiethoff, C.M. and Middaugh, C.R., Nonviral Vectors for Gene Therapy, Humana Press, Totowa, New Jersey, 2002
11. Chavali, S., Lin, B., Miller, D. C. and Camarda, K.V., *Comp. Chem. Eng.*, **28**, 605, 2004
12. Zhao, H., J. P. Ralston, R. C. Middaugh and K. V. Camarda, *Application of Computational Molecular Design to Gene Delivery Polymers*, Proceedings of Foundations of Computer-Aided Process Design, 2004, 415-418, Computer Aids for Chemical Engineering Education, Austin, Texas, (2004)

Acknowledgements:

Dr. Brian Lobo, Dr. Sirirat Choosakoonkriang and Dr. Christopher Wiethoff provided experimental data on complexes.

Chad Braun, Haihong Fan and Dr. Joe Vetro provided consistent assistance for experiments and pharmaceuticals background.

Sarah Thompson and Sarah Hoelzle 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.