Elucidating Structure-Function Relationships of Poly(β-amino ester)s for Non-viral Gene Delivery via Principal Component Analysis

Corey J. Bishop, Budri Abubaker-Sharif, Toni-Rose Guiriba, Jordan J. Green
Department of Biomedical Engineering, The Johns Hopkins University School of Medicine

Statement of Purpose: Non-viral polymeric gene delivery methods are safer than viruses but are not as efficient in successfully delivering nucleic acid cargo. Understanding polymer structure-function relationships would allow for improved design and further optimized delivery systems. Principal Component Analysis (PCA) is a tool to reduce complex multivariate sets of data into linearly uncorrelated orthogonal components, termed principal components (PC). We endeavored to use PCA to analyze a poly(β-amino ester) (PBAE) library to elucidate polymer structure-function relationships for gene delivery to human primary glioblastoma multiforme cells. The correlative relationships between the variables and their ranking by the degree to which they drive transfection, uptake, and viability are reported.

Methods: 75 PBAEs which varied by their backbone, sidechains, and endcaps were included in the library. 24 normalized (0-1) physico-chemical properties of each of the PBAEs were determined via the Joback/ Crippen's fragmentation methods and included the backbone to sidechain (B:S) molar ratio during synthesis, the number of carbons in the backbone and sidechain (B and S), the total number of carbons in B and S (B+S), the number/weight-average molecular weights (Mn and Mw), polydispersity (PDI), degree of polymerization (DP), boiling and melting point, critical volume, Gibb's free energy (GFE), LogP (hydrophobicity), molar refractivity (MR), heat of formation (HtF), and the topological surface area (tPSA). The asterisk ('*') is used to denote parameters associated with the repeat monomer unit of the polymer. These polymers were previously assessed in vitro in primary human glioblastoma multiforme cells [1]. The 24 physico-chemical properties were compared against 3 cell-based functional variables: transfection, uptake, and viability. MatLab was used to perform PCA to assess the coefficients, scores, and variances of the variables. The variables were ranked by the degree to which they positively or negatively contributed to each PC according to their coefficients. Furthermore, the variables were ranked according to the degree to which they drive transfection, uptake and viability which was determined using Acos(Θ) obtained from the loading plot, where A is the magnitude of the vector being compared to transfection, uptake, or viability and Θ is the angle between the variable being compared.

Results: The first 5 PCs accounted for 96.6% of the variance in the data set (**Figure 1; top**). The top 4 variables (either positive or negative (-)) contributing to the 1st and 2nd PCs are listed above each PC (**Figure 1; top**). The correlative relationships between the variables are shown by the loading plot (**Figure 1; bottom**). Variables in the same quadrants are positively correlated, whereas variables in opposite quadrants indicate a negative correlation; adjacent quadrants suggest that the variables are positively correlated with one PC and not the other.

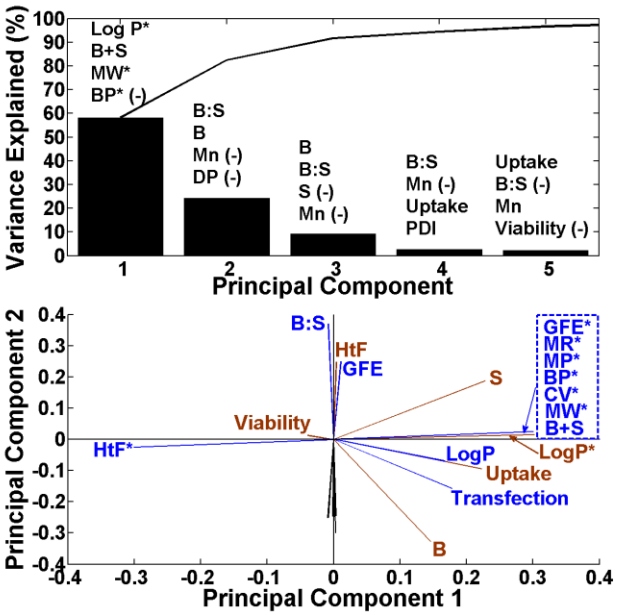

Figure 1. Top: Percentage of variance for each PC and the top 4 contributing variables. **Bottom:** Loading plot of variables showing correlations.

The top 4 variables positively driving transfection, uptake and viability were: B, uptake, LogP*, B+S, and LogP*, B+S, GFE*, MW*, and HtF*, B:S, HtF, and GFE, respectively. The top 4 variables negatively driving transfection, uptake and viability were: GFE, HtF, HtF*, B:S, and GFE, HtF, B:S, HtF*, and MP*, CV*, MR*, and LogP*, respectively. When the scores plot was plotted against transfection efficacy, it could be observed that there were 3 main clusters which formed along PC1 according to parameter B+S. As B+S increased, transfection increased and there was an optimal value for transfection along PC2.

Conclusions: We have successfully demonstrated that PCA is useful in determining which variables drive gene delivery. Some variables positively drive transfection and uptake while negatively driving viability and vice-versa (i.e., LogP* and B:S). Knowing the degree to which each variable drives gene delivery is helpful in focusing efforts for designing optimal gene delivery vectors.

References: [1] Tzeng SY. Adv Healthcare Mater. 2013;2:468-480.

Funding: NSF GRF to CB (DGE- 0707427); NIH (1R01EB016721).