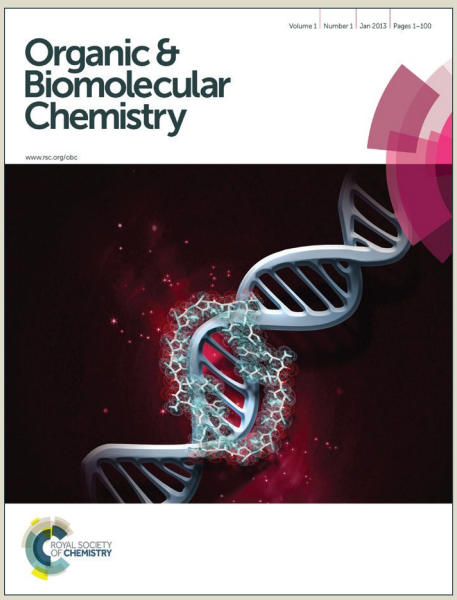
Organic & Biomolecular Chemistry

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/obc

Experimental design, modeling and optimization of polyplex formation between DNA oligonucleotide and branched polyethylenimine

Lilia Clima[a], Elena L. Ursu[a], Corneliu Cojocaru[b], Alexandru Rotaru*[a], Mihail Barboiu[c] and Mariana Pinteala[a]

[a] *Advanced Research Center for Bionanoconjugates and Biopolymers, "Petru Poni" Institute of Macromolecular Chemistry, Romanian Academy, Aleea Grigore Ghica Voda 41 A, 700487 Iasi, Romania.*

[b] *Department of Inorganic Polymers, "Petru Poni" Institute of Macromolecular Chemistry, Romanian Academy, Aleea Grigore Ghica Voda 41 A, 700487 Iasi, Romania.*

[c] *Adaptative Supramolecular Nanosystems Group, Institut Européen des Membranes, ENSCM/UMII/UMR-CNRS 5635, Pl. Euge`ne Bataillon, CC 047, 34095 Montpellier, Cedex 5, France.*

[*] *Corresponding author*: *Alexandru Rotaru*

Email: rotaru.alexandru@icmpp.ro

Tel: +40-232-217454

Fax: +40-232-211299

Abstract

The complexes formed by DNA and polycations have received a great attention owing to their potential application in gene therapy. In this study, the binding efficiency between double-stranded oligonucleotide (dsDNA) and branched polyethylenimine (B-PEI) has been quantified by processing of the images captured from the gel electrophoresis assays. The central composite experimental design has been employed to investigate the effects of controllable factors on the binding efficiency. On the basis of experimental data and response surface methodology, a multivariate regression model has been constructed and statistically validated. The model has enabled to predict the binding efficiency depending on experimental factors, such as concentrations of dsDNA and B-PEI as well as the initial pH of solution. The optimization of the binding process has been performed using simplex and gradient methods. The optimal conditions determined for polyplex formation have yield a maximal binding efficiency close to 100%. In order to reveal the mechanism of complex formation at atomic-scale, the molecular dynamic simulation has been carried out. According to the computation results, B-PEI amine hydrogen atoms have interacted with oxygen atoms from dsDNA phosphate groups. These interactions have conducted to the formation of hydrogen bonds between macromolecules, stabilizing the polyplex structure.

Keywords: DNA, PEI, polyplex, response surface methodology, molecular dynamic simulation

1. Introduction

Gene therapy, a method used to introduce genetic material into cells to treat various maladies requires specific therapeutic genes and efficient, yet non-toxic gene delivery system.[1] Cationic polymers as non-viral gene delivery systems (vectors) have great potential to create pharmaceuticals from nucleic acids.[2-4] They have less specific immune responses, are generally safer and easy to design with more flexible structures and chemical properties for various purposes.[5-9] Among

available cationic polymers, polyethylenimine (PEI), including branched (B-PEI) and linear polyethylenimine (L-PEI), is the most effective vector for various types of polymeric gene carrier, and has thus been extensively used for both in vitro and in vivo gene delivery.[10-12] One of the most important properties of PEI is its high cationic charge density. The theoretical ratio of primary, secondary and tertiary amino groups in B-PEI has been calculated as 1:2:1, respectively,[13] and there is a close relationship between the pH of PEI and the positive charge density on PEI.[14] Previous reports have revealed that the PEI protonation degree at physiological pH is 20%.[8,15] Later, other authors[16,17] have determined by both experiments and computations that for linear, star-like and comb-like PEI, the degree of protonation of amine groups was about 50% at pH=7.4. Recently, a cost-effective gene transfection by plasmid DNA compaction at pH 4 using PEI has been also reported.[18]

Besides plasmid DNA, PEI has also been widely used for the compacting and delivery of short, 20-25 double-stranded nucleotides for siRNA-mediated or oligonucleotide gene therapy.[13,19-22] Although plasmid DNA-based polyplexes are well characterized, it was recently shown that not all knowledge can be adapted from DNA-based polyplexes to short oligonucleotide-based polyplexes, as the synthetic sequence is around 250 times smaller and shows a higher conformational rigidity. Thus, Wagner and co-workers[23] have reported a detailed study on the influence of the various molar masses of B-PEI (0.6, 1.8, 10, 25 kDa) on the stability, formation, and uptake of siRNA-based polyplexes. It was found that, comparing to plasmid DNA, the optimal complexation of siRNA and the subsequent delivery into cells required B-PEI with much lower molecular masses.

Despite the broad interest, the systematic study on the conditions for polyplex formation of short oligonucleotides and PEI, their characteristics are only partially understood and not yet investigated in detail. To the best of our knowledge no studies have been performed to model and optimize the complexation of short oligonucleotides on PEI, including PEI/oligonucleotide ratio and pH. The design of experiments (DoE) and response surface methodology (RSM) are adequate tools

to address such issue. These statistical techniques have been widely accepted and applied for investigation, modeling and optimization of various biotechnological processes.[24-29]

In the current work, DoE and RSM has been applied for modeling and optimization of the complexation process of short 25 nucleotides double-stranded DNA (dsDNA) sequences on B-PEI. The experimental data for monitoring the complexation were obtained from the agarose gel electrophoresis assays by comparing unbound dsDNA band intensities determined by devoted software. In addition, the molecular dynamics simulation has been carried out to uncover the mechanism of polyplex formation.

2. Materials and methods

2.1. Materials

Branched PEI (B-PEI) with an average molecular weight of 2 kDa was purchased from Sigma-Aldrich (Munich, Germany). B-PEI was dissolved in mili-Q water and used as a 200 μM stock solution. All other chemicals (ethidium bromide and sucrose) were purchased from Sigma-Aldrich (Munich, Germany), agarose base, 10 xTAE were purchased from AppliChem GmbH, Germany. HPLC purified double-stranded DNA (dsDNA) was purchased from Metabion AG (Germany), diluted to the concentration of 100 μM and used as stock solution. The sense strand was: 5'-CAAGCCCTTAACGAACTTCAACGTA-3' and antisense strand was: 5'-TACGTTGAAGTTCGTTAAGGGCTTG-3'.

2.2. Preparation of polyplexes

The buffers used in binding experiments contained 0.2xTAE (8 mM Tris, 0.4 mM acetic acid and 0.2 mM EDTA), 50 mM NaCl at correspondingly adjusted pH values. The total volume of an assay mixture was 15 μl. dsDNA was prepared by annealing of sense and antisense DNA strands (46,8 μM) in 2.4xTAE (96 mM Tris, 48 mM acetic acid and 2.4 mM EDTA) and 120 mM NaCl. The samples, containing corresponding dsDNA/B-PEI weight ratios (D/P ratios) and various pH

values were incubated at 25°C for 1 h prior to gel loading. The testing conditions are detailed in the experimental design section.

2.3. Agarose gel electrophoresis assay

Binding experiment involved mixing the appropriate amounts of dsDNA and B-PEI under the desired pH conditions and sample incubation for 1h at 25°C. Sucrose (5 µl with conc. 25% in water) was added and the samples were immediately loaded onto 1% agarose gel and run at 90 mV for 60 minutes at room temperature in 1xTAE buffer (40 mM Tris, 20 mM acetic acid and 1 mM EDTA). Subsequently, the gels were stained with ethidium bromide for 10 minutes at room temperature and then photographed. The dsDNA bands were quantified by analysing the gels at the wavelength of 254 nm using a DNR Bio-imaging system.

2.4. Quantification method

The investigation of the binding affinity of DNA by gel electrophoresis method is essential to quantify the formation of the polyplexes.[30] In this work, for the quantification of binding affinity of the samples, the gel captured images were analyzed by Gel Quant Express software (Fig. 1). Typically, a gel electrophoresis experiment at a certain pH value contained dsDNA as reference whose intensity was quantified by the software as 100%, and several (three or four) parallel samples of a given D/P ratio. To determine the maximum loading capacity of dsDNA on B-PEI, we have chosen to form a neutral dsDNA/B-PEI polyplex which did not migrate in gel but is only partially stained in gel pocket. The average signal intensity value of unbound dsDNA (Fig. 1) was determined comparing the band intensity of the reference signal and the band intensity of unbound dsDNA of the sample. If the average value represents the band intensity of unbound dsDNA, than, the binding efficiency Y (%) has been quantified by:

$$Y = 100 - I_{bs} \quad (1)$$

where 100% denotes the intensity of the reference signal (dsDNA); and I_{bs} (%) is the average intensity of the bottom spots from lanes in the assay image, representing the amount of the unbound dsDNA in the sample. The exact value of dsDNA/B-PEI binding efficiency was calculated for each experiment from an average of minimum of 3 loading samples.

3. Results and discussions

3.1. Experimental design and modeling of the polyplex formation process

Response surface methodology (RSM) is a collection of mathematical and statistical techniques used for modeling and optimization of the experimental processes.[31] This methodology combines efficiently the design of experiments (DoE) and multivariate regression modeling. The theoretical background of RSM can be found in the common textbooks[31-33] and in some review articles.[34,35]

The aim of RSM is to optimize the values of input variables (factors) to obtain the best response (output variable) of a process under investigation. Experimental design enables the studying of the process via simultaneous changing of the levels of factors, resulting in reduced number of experimental runs comparing with conventional one-variable-at-a-time approach.

In this work, the experimental design was applied to investigate quantitatively the binding (complexation) process between dsDNA and B-PEI employing three input variables (factors) for the design of the experiments, i.e. dsDNA concentration, B-PEI concentration and initial pH of solution. For modeling purpose, the input variables have been converted into the coded variables ranging from -1 (minimum level) to +1 (maximum level). Such codification enables the investigation of variables of different orders of magnitude using the same dimensionless scale for all of the factors.[31,32,34,35] The relation between actual values of these factors and their coded levels is summarized in Table 1. The central composite orthogonal design (CCD) was adopted for the investigation of the complexation process as presented in Table 2. For each run given in Table 2, the binding efficiency (response) was determined experimentally by gel electrophoresis. To

this end, agarose gel electrophoresis retardation assay was used to evaluate the binding between B-PEI and dsDNA sequence at different pH values and various dsDNA/B-PEI ratios (D/P ratios). The observed values of the response are reported in Table 2. The images of gel electrophoresis assays for each experimental run are given in the electronic supplementary information (ESI Fig.S1).

On the basis of experimental data (Table 2), a response surface model was constructed in terms of coded variables (x_1, x_2 and x_3) using the multivariate regression method[31,32,34] that can be written as:

$$\hat{Y} = 46.726 - 8.371x_1 + 5.869x_2 - 13.309x_3 - 6.662x_2^2 - 3.889x_3^2 - 2.769x_2x_3 \qquad (2)$$

$$\text{subject to: } -\alpha \le x_i \le +\alpha;\ \alpha=1.215;\ \forall i = \overline{1,\ 3}$$

The fitted model (Eq.2) involves only the significant coefficients that have been identified by *Student t-test.*[31,32] The importance of each factor was evaluated by calculation of the percentage effect of each term on the response. The percentage effect (p_i) of factors and their interaction was calculated according to Pareto analysis.[36]

$$p_i = \left(\frac{b_i^2}{\sum_{i=1}^{n} b_i^2} \right) \times 100 \quad i \neq 0 \qquad (3)$$

The results of the percentage effects are illustrated in Figure 2a. Thus, for the complexation process (between dsDNA and B-PEI), the pH factor had the most important effect (50.78%) on the binding efficiency, followed by the effect of dsDNA concentration (20.09%). The main effect of the B-PEI concentration is of 9.87%, and its quadratic effect (B-PEI × B-PEI) is of 12.72%. The quadratic effect of pH factor is about 4.34%. Likewise, a mutual effect of 2.20% is attributed to the interaction between B-PEI and pH.

The multivariate regression model has been validated statistically using the analysis of variance (ANOVA).[31,34] The results of this statistical test (ANOVA) are summarized in Table 3. Since, the *P*-value is quite low (i.e. *P*-value = 0.0007), the model is statistically significant. The value of the coefficient of determination R^2 indicates that the model can explain more than 89% of the data variation. The goodness-of-fits between predicted values and experimental observations are shown in Figure 2b. Since the data are scattered close to the bisector, the model is a good predictor for the binding efficiency *Y* (%), and the parity plot (Fig. 2b) corroborates the ANOVA results (Table 3). Finally, the multivariate regression model with actual (real) variables has been developed by the substitution technique and can be written as follows:

$$\hat{Y} = -411.534 - 5.023z_1 + 47.369z_2 + 60.519z_3 - 1.064z_2^2 - 3.889z_3^2 - 1.107z_2z_3 \qquad (4)$$

subject to:

$26.33 \le (z_1 = [\text{dsDNA}]) \le 30.33$ (μM); $14.51 \le (z_2 = [\text{B-PEI}]) \le 20.52$ (μM); $5.8 \le (z_3 = \text{pH}) \le 8.2$

On the basis of the multiple regression model (4), the response surfaces and contour-lines maps are plotted (Figs. 3-4) showing the couple effects of the input variables (factors) on the binding efficiency (response). According to Figure 3, the greater dsDNA concentration becomes, the less binding efficiency is, due to the extra amount of unbound dsDNA. The increment of B-PEI concentration up to 19 μM leads to the significant improvement to the binding efficiency. For greater values of B-PEI concentration ([B-PEI] >19 μM), this ascending trend is attenuated and slightly reversed owing to the quadratic effect produced by the model.

Surface plots in Figure 4 show the couple effects of B-PEI concentration and pH upon the predicted response. As one can see, if pH value diminishes (from 8 to 6), the binding efficiency increases substantially. Additionally, an interaction effect can be observed between B-PEI concentration and solution pH underlying that the polyplex formation is more efficient at a larger amount of B-PEI and lower pH values. This can be attributed to the more intense protonation of

amine groups from B-PEI at lower pH values. The results of response surface analysis closely corroborate with the gel electrophoresis data (assays) presented as images in the supplementary material (Fig.S1).

3.2. Optimization of the polyplex formation process

The aim of the experimental design and response surface modeling is to optimize the polyplex formation process. To accomplish this goal, the model-based optimization has been performed to find the conditions that maximize the binding efficiency $\hat{Y}$ (%) between dsDNA and B-PEI, and it is given by:

$$\max\ \hat{Y}(x_1, x_2, x_3),\ \text{subject to:}\ x_i \in [-\alpha,\ \alpha],\ \ \forall i = \overline{1,\ 3} \tag{5}$$

The method of simplex algorithm[37] and an appropriate solver (*neldermead*) implemented in SciLab 5.4.1 open-source software has been used for the direct search optimization. The optimal solution given by simplex method, in terms of actual values of factors, is as follows: [dsDNA]=26.33 μM, [B-PEI]=19.26 μM and pH=5.8. Under these conditions, the predicted response is $\hat{Y}$ =70.52 %, and the confirmed experimental response is *Y*=70.79 %. The observed value of response (*Y*=70.79 %) is slightly greater than any binding efficiency value reported in the initial experimental design (Table 2). Therefore, these conditions are optimal for the investigated region of experimentation (valid region). However, we were motivated to figure out even higher values of binding efficiency (i.e. $Y > 90$ %). This required searching beyond the valid region where the model could not be applied. In this respect, the gradient method was employed for the additional optimization based on the experimentation and calculus.[38] According to this method the searching of optimum is accomplished in the gradient direction, which is given by the first order partial derivatives of the objective function. Thus, the one step displacement form the current point $z_i^{(k)}$ into the next one $z_i^{(k+1)}$ is given by the following relationship:

$$z_i^{(k+1)} = z_i^{(k)} + \lambda^{(k)} d_i^{(k)} \quad \forall i = 1,\ 2,\ \ldots\ n \tag{6}$$

where z_i is the real value of the variable, k is the iteration index, λ is the step-length, d_i is the gradient direction and n is the number of variables. For the problem of maximization, the gradient directions correspond to the steepest ascent and can be written as follows:

$$d_i^{(k)} = \frac{\left(\partial Y / \partial z_i^{(k)}\right)}{\sqrt{\sum_i \left(\partial Y / \partial z_i^{(k)}\right)^2}} \approx \frac{\left(\Delta Y / \Delta z_i^{(k)}\right)}{\sqrt{\sum_i \left(\Delta Y / \Delta z_i^{(k)}\right)^2}} \tag{7}$$

The gradient method can be applied even if the mathematical expression of the objective function (response) is unknown for a specific region of experimentation, but the experimental system is available. In our case, the gradient experimentation method was engaged to complement RSM and to improve dsDNA/B-PEI binding efficiency. The starting point for the gradient methodology was the optimal point given by Nelder-Mead simplex method, i.e. z_1=26.33 μM (dsDNA), z_2=19.26 μM (B-PEI) and z_3=5.8 (pH), which yielded the observed response of Y=70.79 %. Subsequently, only two variables z_1 (dsDNA) and z_2 (B-PEI) were selected for optimization, whereas the third variable z_3 (pH) was fixed at a constant level of pH 5.8 to avoid a more acidic pH that can damage dsDNA. Hence, the overall experimental optimization was reduced to the two-variable problem. The experimental design carried out according to the gradient method is summarized in Table 4. All values of binding efficiency Y (%) reported in Table 4 were determined experimentally by gel electrophoresis method (Fig.5, assays A-G1, A-G2 and A-G3). Thus, for the increment Δz_1=(28.00-26.33)=1.67 the corresponding variation of response is ΔY=(63.31-70.79)= -7.48; similarly for Δz_2=2.51 the variation of response is ΔY=29.12. Based on these values, the gradient directions have been calculated using Eq. (7). The obtained directions are as follows, d_1= -0.36004 and d_2=0.93294. Following Eq. (6) and using a fixed step-length of λ=3, the new values for the variables has been determined, i.e. z_1=25.25 and z_2=22.06. Thereof, the optimal conditions given by gradient method are [dsDNA]=25.25μM, [B-PEI]=22.06 μM and

pH=5.8. Following these conditions, a maximal value (*Y*=99.96%) of the binding efficiency was obtained experimentally (Fig.5, assay A-G4). Thus, under the calculated optimum conditions given by the gradient method a complete complexation occurred between dsDNA and B-PEI (binding efficiency close to 100%).

It is worth mentioning that RSM has enabled to study the relationship between input variables and response to identify the interaction effects between variables and to optimize the process of polyplex formation. However, this methodology belongs to the group of "black-box" models, which do not explain the mechanism of the process under study. In order to unveil the mechanism of dsDNA/B-PEI complexation, the additional modeling approach based on molecular dynamics has been carried out.

3.3. Molecular dynamics simulation of dsDNA / B-PEI *polyplex formation*

Molecular dynamics simulation is a computational tool useful for understanding the structures and functions of biological macromolecules as well as their interactions.[39] This kind of simulation provides details regarding the individual atomistic motions of macromolecules surrounded by explicit solvent molecules as a function of time. In the last years, several groups have addressed the molecular dynamics simulations of DNA/PEI complexes using a Drew-Dickerson dodecamer $d(CGCGAATTCGCG)_2$ as the model for short DNA helix.[40,41] Such modeling studies provided valuable information about the interaction between DNA and PEI, and helped to elucidate the mechanism of polyplexes formation.

In this work, the simulated DNA helix has the same nucleotide sequences as dsDNA used for the experimentation, i.e. the sense strand 5'-CAAGCCCTTAACGAACTTCAACGTA-3' and the antisense strand 5'-TACGTTGAAGTTCGTTAAGGGCTTG-3'. The B-form of DNA was built using YASARA-Structure program.[42] Thus, the modeled dsDNA contains 50 nucleotides, which carries a total charge of -52 in the fully deprotonated state and has a molecular weight of 15.43 kDa. Note that, according to the modeled configuration of dsDNA built by YASARA program, there are

terminal phosphate groups at 5'-end positions having the negative charges of -2 (at O1P and O3P). The simulated PEI is a branched macromolecule (B-PEI) consisting of 32 amine groups with a molecular weight of 1.38 kDa. The B-PEI molecule was built in HyperChem software, optimized at the level of PM3 semiempirical model and finally exported to the YASARA-Structure program. Note that, the modeled B-PEI contains 16 amine groups in the backbone and 16 amine groups in the branched segments. As such, B-PEI involves 14 primary amine groups ($–NH_2$), 6 secondary amine groups (=NH) and 12 tertiary amine groups (=N–), latter being pinpointed only on the backbone chain. Hence, the macromolecule B-PEI carries a total charge of +32 in the fully protonated state. The chemical structure of the modeled B-PEI is given in detail in the ESI (Fig. S2).

The molecular dynamics simulation was performed by means of YASARA-Structure software package, version 14.12.2.[42-44] The YASARA program employs an automatic parameterization method (termed "AutoSMILES") for the unknown molecular structures. This algorithm was used to generate YASARA force field parameters for the investigated macromolecules.

In accordance with the simulation protocol, the macromolecules (dsDNA and B-PEI) were solvated in 32,373 TIP3P water molecules in a rectangular cell (box) with the size of 100 Å × 100 Å × 100 Å. The periodic boundary conditions were set for the simulating box, and the molecular system consisted of 99,201 atoms. First, the cell neutralization simulation was carried out to add monovalent counterions (Na^+ and Cl^-) attaining the mass fraction of 0.9 %. The simulation cell was also equilibrated by energy minimization using the steepest descent algorithm followed by the simulated annealing minimization and a short molecular dynamics simulation (2 ps). At the end of cell neutralization and equilibration simulations the configurations of macromolecules (dsDNA and B-PEI) were adopted as the initial structures for the molecular dynamics production run. In this study, a fully protonated state of B-PEI was considered at pH=5.8.

The molecular dynamics simulation has been performed using the self-parameterizing knowledge-based YASARA force field.[41,42] For the production run, the pressure control has been enabled by setting the solvent probe mode, i.e. water density of 0.997 g/cm^3, that corresponds to the

conditions of a constant pressure P = 1 bar and temperature of T = 298 K. A time step of 1 fs has been applied to integrate the equations of motion. For computation of the non-bonding interactions (van der Waals and electrostatic), a cut-off distance of 12 Å was applied. The electrostatic interactions have been computed by the particle mesh Ewald method (PME). For simulation run, a value of pH=5.8 has been fixed, representing the optimum value for polyplex formation according to the experimental study. Finally, 21 ns long molecular dynamics simulation has been started and the trajectories have been recorded periodically as snapshots every 10,000 steps. YASARA-Structure program has been used also for the visualization and trajectory analysis.

In Figure 6, the initial and final all-atom snapshots (including water molecules) from the molecular dynamics trajectory are presented. As shown in Figure 6a, at the initial time (t=0), dsDNA and polycation B-PEI are separated by a distance of 40 Å between their centers of geometries (COG distance). At the end of simulation (t = 21 ns), the formation of dsDNA/B-PEI polyplex structure can be clearly observed (Fig. 6b).

Figure 7 illustrates typical snapshots from the molecular dynamics trajectory showing the interaction between macromolecules at different simulation times. For clarity, water molecules are omitted. Thus, at the simulation time t=2 ns, the first contacts between macromolecules can be observed (Fig. 7a), which occur between B-PEI and the sense strand of dsDNA. At 5 ns simulation time (Fig. 7b), the contacts between B-PEI and both strands of dsDNA can be discerned in the vicinity of a minor groove, revealing an incomplete organized structure of the complex. For higher time values of 12 ns (Fig. 7c) and 20 ns (Fig. 7d), the well-defined polyplex structures can be observed. These polyplex structures involve the aligning of B-PEI with dsDNA phosphate groups of both strands, crossing over a minor groove of dsDNA. In accordance with molecular dynamics simulation data, the total and potential energy of the system decrease from the initial values and stabilize at lower values (ESI, Fig. S3). This suggests that the formation of the polyplex structure is an energetically favorable process.

Calculated plot in Figure 8a shows the distances between the centers of geometry (COG distance) of dsDNA and B-PEI over the course of the simulation. The COG distance between macromolecules decreases rapidly within the first 4 ns, from 40 Å to 25 Å (Fig. 8a). Afterwards (for $t > 4$ ns), the intermolecular distances fluctuate into the interval ranging from 23 Å to 29 Å. The COG distance hasn't decreased below 23 Å, since the significant part of B-PEI is aligned on a top minor groove of DNA helix, beyond the DNA center of geometry. Thus, the distance between centers of geometries is still distinguishable, even if the macromolecules are in the close contact. As the COG distance between dsDNA and B-PEI diminishes from 40 Å to 25 Å, the number of atoms in intermolecular contact (for a cutoff radius of 4 Å) increases significantly, from zero (at $t = 1$ ns) up to 500 (for $t > 12$ ns, see Fig. 8b). The increment of the number of atoms in intermolecular contact leads to the formation of hydrogen bonds (H-bonds) between B-PEI amine hydrogen atoms and backbone oxygen atoms of dsDNA.

The calculated number of hydrogen bonds formed against simulation time is depicted in Figure 9a. According to this plot, the first hydrogen bond is formed at the simulation time of t=1.2 ns, when the COG distance is about 36.8 Å and the number of atoms in intermolecular contact is equal to 8. The energy of this hydrogen bond is equal to 4.48 kcal/mol. For greater values of simulation time the number of H-bonds fluctuates (increases or decreases), generally following an upward trend. Figure 9b reports the total energy of the formed hydrogen bonds. As shown in this figure, for the simulation time smaller than 12 ns the number of hydrogen bonds is less than 7 and the total energy does not exceed the value 35 kcal/mol. For the simulation time greater than 12 ns, the number of hydrogen bonds is higher than 7, approaching a maximum value of 12 H-bonds and the total energy of H-bonds is greater than 35 kcal/mol, attaining a maximum value of 67 kcal/mol. Thus, for $t > 12$ ns the polyplex structure is stabilized by a higher number of hydrogen bonding between dsDNA and B-PEI.

The simulation maps showing dsDNA nucleotides (from both sense and antisense strands) involved in the formation of the hydrogen bonds via backbone oxygen atoms (O1P, O2P, O3* and

O5* from phosphate groups) are presented in Fig. 10. The details regarding the notation and positions of DNA backbone oxygen atoms are given schematically in the supplementary material (Fig.S4). According to simulation data, the hydrogen bonds are formed owing to DNA phosphate-polycation amine interactions. The involved DNA phosphate groups for H-bonds formation corresponds to the four bases T8, T9, A10 and A11 from the sense strand (Fig.10a), as well as to the six bases G20, G21, C22, T23, T24 and G25 form the antisense strand (Fig.10b). The hydrogen bonds are predominantly formed through phosphate oxygen atoms O1P and O2P and less frequently via oxygen atoms of type O3* and O5* (Fig.10).

A total overview of the history of the polycation amine groups interacted with dsDNA phosphate groups for the formation of hydrogen bonds are summarized in Figure 11. Two stages can be distinguished taking into account the number of amine groups that contributed to the formation of H-bonds. For the first stage (from 1 to 11 ns) a reduced number of amine groups has been involved, comprising only primary and secondary amine moieties. For the second stage (from 12 to 21 ns) a larger number of amine groups has participated for the intermolecular hydrogen bonding. Herein, the primary, secondary and tertiary amines moieties have been involved, unveiling a more stable polyplex structure in terms of total energy of H-bonds.

Since nitrogen atoms of B-PEI interact with backbone oxygen atoms from dsDNA, the radial distribution function (RDF) has been plotted (Fig.12). The radial distribution function of B-PEI nitrogen atoms around backbone oxygen atoms of dsDNA has two distinct peaks: one is pinpointed at ~ 3Å and the second is located at ~5 Å. The most intense peak at ~ 3Å indicates the main interactions occurred due to the direct interactions between B-PEI amine groups and dsDNA backbone oxygen atoms. Such direct interactions include the formation of hydrogen bonds between amine hydrogen atoms and the oxygen atoms from the phosphate groups. The less intense peak at ~ 5Å suggests the indirect interactions, which may be attributed to the water-mediated hydrogen bonding or other weaker interactions. Similar simulation outcomes, investigating the complexes

formation between PEI and a short DNA helix (Drew-Dickerson dodecamer), have been reported by other authors.[40,42]

A frequently used measure for the characterization of the conformations of a macromolecule is the radius of gyration (R_g), which measures the root-mean-square distance of a collection of segments from their common center of mass. Likewise, the root-mean-square-deviation (RMSD) of atomic positions is a widely employed parameter for the comparison of molecular conformers resulted from simulation and a target structure.[45,46] In this study, the RMSD values for dsDNA and B-PEI have been calculated with respect to their equilibrated structures (i.e. initial geometries). Figure 13 shows the changes in radii of gyration and RMSD values against the simulation time for dsDNA and B-PEI. The dsDNA radius of gyration rapidly drops down from its initial value of 28.4 Å, with the further fluctuations around the average value of 24.3 Å (Fig. 13a). This means that throughout molecular dynamic simulation, the DNA helix becomes slightly compacted comparing with its initial equilibrated structure. Similarly, the B-PEI radius of gyration decreases in the first moments from the initial value of 17.8 Å and subsequently fluctuates around the mean value of 12.8 Å. The survey of Figure 13b reveals that RMSD values exceed 3 Å, for both macromolecules. Note that, if RMSD value is higher than 3 Å then the molecular conformation is considered unlike to the target structure.[46] During the course of simulation process, RMSD values for dsDNA ranged between 4.70 Å and 8.07 Å; whereas for B-PEI, RMSD values varied between 5.23 Å and 8.85 Å. These findings suggest that the relaxed conformations of macromolecules are dissimilar to their initial equilibrated structures. Thus, both radii of gyration and RMSD values resulted from molecular dynamic simulation indicate that dsDNA and B-PEI are flexible macromolecules. On the basis of RMSD values (Fig. 13b), it can be stated that B-PEI is more flexible than dsDNA. For the flexible polymeric chains such as polyethylenimine is also useful to estimate the end-to-end distance, which is reported in ESI (Fig. S5). Thus, the end-to-end distance of B-PEI varied dramatically over the simulation from its initial value of 61.9 Å (stretched structure) up to 30.1 Å (compacted structure). The average end-to-end distance of B-PEI was 38.0 Å, as also given in Figure S5.

Concerning the flexibility of the investigated dsDNA, this may be attributed to the longer length of the strands (25 nucleotides) comparing with Drew-Dickerson DNA dodecamer (12 nucleotides) commonly used in molecular dynamic simulations. However, the examination of the spacing of the minor and major grooves of the initial and final dsDNA conformers has revealed that at the end of simulation the dsDNA helix has remained at the B-form, without any major structural changes. For instance, the minor grooves for the initial and final dsDNA structures are about 12 Å wide. The major grooves of the initial dsDNA helix are of 22.6 Å average wide. In case of the final dsDNA configuration, the major grooves are smaller, i.e. about 21.3 Å average wide, unveiling a slight condensation of dsDNA helix after the polyplex formation.

Subsequently, an additional experiment using circular dichroism (CD) technique has been carried out to confirm the B-form conformation of dsDNA helix from the polyplex structure. The experimental results have shown similar CD curves for both the dsDNA and complexed dsDNA with B-PEI at pH 5.8 (supplementary material Fig. S6). Thus, it has been proved experimentally that B-form of the investigated dsDNA sequence is stable during the course of the complexation process. Consequently, the MD simulation outcomes are in agreement with experimental observation regarding the stability of B-form for the considered dsDNA in the polyplex structure.

In conclusion, the outcomes of molecular dynamic simulation shed light on the molecular mechanism of complexation between dsDNA and B-PEI, which may assist forward the design of PEI-based gene delivery vectors.

4. Conclusions

In this work, a design of experiments was adopted to investigate the binding efficiency between double-stranded oligonucleotide (dsDNA) and branched polyethylenimine (B-PEI) under various conditions of D/P ratio (dsDNA / B-PEI) and pH of solution. The binding efficiency was quantified by processing of images captured from the gel electrophoresis assays. Based on experimental design data and response surface methodology, a multivariate regression model was

constructed and validated using ANOVA statistical test. The model enabled to find a functional relationship between the binding efficiency and experimental factors, such as concentrations of macromolecules and the initial pH of solution. The optimization of the binding process was carried out using simplex and gradient methods. The optimal conditions determined for polyplex formation were [dsDNA]=25.25 μM, [B-PEI]=22.06 μM, (D/P ratio 1.145) and pH=5.8. Under these conditions, a maximal value of the binding efficiency was obtained experimentally, i.e. 99.96%.

To unveil the mechanism of polycomplex formation at atomic-scale, the molecular dynamic simulation was performed. Starting from an initial separation of ~40 Å between macromolecules, dsDNA and B-PEI approached each other and formed a stabile complex after 12 ns. The simulation outcomes revealed that B-PEI amine hydrogen atoms mainly interacted with oxygen atoms from dsDNA phosphate groups. Such interactions conducted to the formation of hydrogen bonds between macromolecules, stabilizing the polyplex structure. Mostly, the hydrogen bonds were formed through dsDNA phosphate oxygen atoms O1P and O2P. The radial distribution function (RDF) corroborated the interactions between B-PEI and dsDNA. Regarding the polyplex structure, the polycation chain stayed close to the dsDNA helix making mostly contacts in the vicinity of a minor groove site. Both radii of gyration and RMSD values resulted from molecular dynamic simulation confirmed the flexibility of macromolecules. However, the oligonucleotide did not undergo any major structural changes on complexation and remained in slightly condensed B-form.

Acknowledgement

The authors gratefully acknowledge the financial support of this work from the grant of Romanian National Authority for Scientific Research, CNCS-UEFISCDI; project number PN-II-ID-PCCE-2011-2-0028.

References

1. R. M. Blaese, K. W. Culver, A. D. Miller, C. S. Carter, T. Fleisher, M. Clerici, G. Shearer, L. Chang, Y. W. Chiang, P. Tolstoshev, J. J. Greenblatt, S. A. Rosenberg, H. Klein, M. Berger, C. A. Mullen, W. J. Ramsey, L. Muul, R. A. Morgan and W. F. Anderson, *Science*, 1995, **270**, 475-480.
2. M. E. Davis, *Curr Opin Biotech*, 2002, **13**, 128-131.
3. K. Wang, X. Yan, Y. Cui, Q. He and J. Li, *Bioconjug Chem*, 2007, **18**, 1735-1738.
4. X. Yan, J. Blacklock J., J. Li and H. Mohwald, *ACS Nano*, 2012, **6**, 111-117.
5. S. Han, R. I. Mahato, Y. K. Sung and S. W. Kim, *Mol Ther*, 2000, **2**, 302-317.
6. S. Li and L. Huang, *Gene Ther*, 2000, **7**, 31-34.
7. D. Luo and W. M. Saltzman, *Nat Biotechnol*, 2000, **18**, 33-37.
8. M. X. Tang and F. C. Szoka, *Gene Ther*, 1997, **4**, 823-832.
9. M. Thomas and A. M. Klibanov, *Appl Microbiol Biot*, 2003, **62**, 27-34.
10. B. Abdallah, A. Hassan, C. Benoist, D. Goula, J. P. Behr and B. A. Demeneix, *Hum Gene Ther*, 1996, **7**, 1947-1954.
11. W. T. Godbey, K. K. Wu and A. G. Mikos, *J Control Release*, 1999, **60**, 149-160.
12. H. Pollard, J. S. Remy, G. Loussouarn, S. Demolombe, J. P. Behr and D. Escande, *J Biol Chem*, 1998, **273**, 7507-7511.
13. A. von Harpe, H. Petersen, Y. X. Li and T. Kissel, *J Control Release*, 2000, **69**, 309-322.
14. M. Neu, D. Fischer and T. Kissel, *J Gene Med*, 2005, **7**, 992-1009.
15. J. Suh, H. J. Paik and B. K. Hwang, *Bioorg Chem*, 1994, **22**, 318-327.
16. M. Borkovec and G. J. M. Koper, *Macromolecules*, 1997, **30**, 2151-2158.
17. G. J. M. Koper, R. C. van Duijvenbode, D. D. P. W. Stam, U. Steuerle and A. Borkovec, *Macromolecules*, 2003, **36**, 2500-2507.
18. Y. Fukumoto, Y. Obata, K. Ishibashi, N. Tamura, I. Kikuchi, K. Aoyama, Y. Hattori, K. Tsuda, Y. Nakayama and N. Yamaguchi, *Cytotechnology*, 2010, **62**, 73-82.
19. M. E. Bonnet, P. Erbacher and A. L. Bolcato-Bellemin, *Pharm Res-Dord*, 2008, **25**, 2972-2982.
20. B. Brissault, A. Kichler, C. Guis, C. Leborgne, O. Danos and H. Cheradame, *Bioconjugate Chem*, 2003, **14**, 581-587.
21. Z. J. Dai and C. Wu, *Macromolecules*, 2012, **45**, 4346-4353.

22. J. J. Turner, S. W. Jones, S. A. Moschos, M. A. Lindsay and M. J. Gait, *Mol Biosyst*, 2007, **3**, 43-50.

23. M. Wagner, A. C. Rinkenauer, A. Schallon and U. S. Schubert, *Rsc Adv*, 2013, **3**, 12774-12785.

24. S. Inamdar, S. Joshi, V. Bapat and J. Jadhav, *J Biotechnol*, 2014, **170**, 28-34.

25. C. L. Liu, T. H. Lin and R. S. Juang, *Int J Biol Macromol*, 2013, **62**, 518-522.

26. A. Mehta, G. S. Prasad and A. R. Choudhury, *Int J Biol Macromol*, 2014, **64**, 252-256.

27. A. K. Nayak and D. Pal, *Int J Biol Macromol*, 2011, **49**, 784-793.

28. Z. Y. Wang, J. Li, S. Cheong, U. Bhaskar, O. Akihiro, F. M. Zhang, J. S. Dordick and R. J. Linhardt, *J Biotechnol*, 2011, **156**, 188-196.

29. W. B. Zhi, J. N. Song, F. Ouyang and J. X. Bi, *J Biotechnol*, 2005, **118**, 157-165.

30. M. M. Garner and A. Revzin, *Nucleic acids research*, 1981, **9**, 3047-3060.

31. D. C. Montgomery, *Design and Analysis of Experiments*, John Wiley & Sons, New York, 2001.

32. S. Akhnazarova and V. Kafarov, *Experiment Optimization in Chemistry and Chemical Engineering*, Mir Publisher, Moscow, second edn., 1982.

33. G. E. P. Box and N. R. Draper, *Response Surfaces, Mixtures, and Ridge Analyses*, John Wiley & Sons, New York, second edn., 2007.

34. M. A. Bezerra, R. E. Santelli, E. P. Oliveira, L. S. Villar and L. A. Escaleira, *Talanta*, 2008, **76**, 965-977.

35. A. Witek-Krowiak, K. Chojnacka, D. Podstawczyk, A. Dawiec and K. Pokomeda, *Bioresource Technol*, 2014, **160**, 150-160.

36. A. R. Amani-Ghadim, S. Aber, A. Olad and H. Ashassi-Sorkhabi, *Chem Eng Process*, 2013, **64**, 68-78.

37. A. Karimi and P. Siarry, *Eng Appl Artif Intel*, 2012, **25**, 48-55.

38. C. Cojocaru, G. Zakrzewska-Trznadel and A. Jaworska, *J Hazard Mater*, 2009, **169**, 599-609.

39. M. Karplus and J. A. McCammon, *Nat Struct Biol*, 2002, **9**, 646-652.

40. C. B. Sun, T. Tang, H. Uludag and J. E. Cuervo, *Biophys J*, 2011, **100**, 2754-2763.

41. J. Ziebarth and Y. M. Wang, *Biophys J*, 2009, **97**, 1971-1983.

42. http://www.yasara.org/ - YASARA official web-site, Yet another scientific artificial reality application: molecular graphics, modeling and simulation program.

43. E. Krieger, K. Joo, J. Lee, J. Lee, S. Raman, J. Thompson, M. Tyka, D. Baker and K. Karplus, *Proteins*, 2009, **77**, 114-122.

44. E. Krieger, G. Koraimann and G. Vriend, *Proteins*, 2002, **47**, 393-402.

45. O. M. Becker, *Conformational analysis*, in: O. M. Becker, A. D. MacKerell, B. Roux and M. Watanabe (Eds.), *Computational Biochemistry and Biophysics*, Marcel Dekker, New York, 2001, pp. 69-90.

46. C. S. Tsai, *An Introduction to Computational Biochemistry*, Wiley-Liss, New York, first edn., 2002.

Experimental design, modeling and optimization of polyplex formation between DNA oligonucleotide and branched polyethylenimine

Lilia Clima, Elena L. Ursu, Corneliu Cojocaru, Alexandru Rotaru*, Mihail Barboiu and Mariana Pinteala

TABLES

Table 1

Design variables and their coded and actual values used for studying of dsDNA/B-PEI complexation process.

Design variables (factors)	Coded variables	Real values of coded levels				
		- α	-1	0	+1	+α
dsDNA concentration, [dsDNA] (μM)	x_1	26.33	26.67	28.33	30.00	30.33
B-PEI concentration, [B-PEI] (μM)	x_2	14.51	15.01	17.51	20.02	20.52
Initial *pH* of solution	x_3	5.8	6.0	7.0	8.0	8.2

Note: α = 1.215 (value of axial point for orthogonal CCD in case of three factors)

Table 2

Central composite orthogonal design used for investigating dsDNA/B-PEI condensation process and experimental response (binding efficiency) determined for each run.

Run	Design variables						Response (experimental)
	DNA Concentration, (μM)		B-PEI Concentration, (μM)		Solution pH		
	[dsDNA]	x_1	[B-PEI]	x_2	pH	x_3	Y (%)
1	30.00	+ 1	20.02	+ 1	8.0	+ 1	21.87
2	26.67	- 1	20.02	+ 1	8.0	+ 1	39.07
3	30.00	+ 1	15.01	- 1	8.0	+ 1	13.30
4	26.67	- 1	15.01	- 1	8.0	+ 1	25.53
5	30.00	+ 1	20.02	+ 1	6.0	- 1	46.55
6	26.67	- 1	20.02	+ 1	6.0	- 1	70.60
7	30.00	+ 1	15.01	- 1	6.0	- 1	28.15
8	26.67	- 1	15.01	- 1	6.0	- 1	44.74
9	30.33	+ α	17.51	0	7.0	0	34.95
10	26.33	- α	17.51	0	7.0	0	52.74
11	28.33	0	20.52	+ α	7.0	0	35.89
12	28.33	0	14.51	- α	7.0	0	37.61
13	28.33	0	17.51	0	8.2	+ α	18.00
14	28.33	0	17.51	0	5.8	- α	63.68
15	28.33	0	17.51	0	7.0	0	52.65
16	28.33	0	17.51	0	7.0	0	53.29

Table 3

Analysis of Variance (ANOVA) for the significance of the multivariate regression model.

Source	*DF*[A]	*SS*[B]	*MS*[C]	*F-value*	*P-value*	R^2	R_{adj}^2
Model	6	3.554×10^3	592.319	12.347	0.0007	0.892	0.819
Residual	9	431.756	47.973				
Total	15	3.986×10^3					

[A] – degree of freedom; [B]- sum of squares; [C] - mean square

Table 4

Experimental design carried out according to the gradient method, pH=5.8.

No.	Assay code	z_1 : [dsDNA], μM	z_2 : [B-PEI], μM	Y (%), experimental values
1	A-G1	26.33	19.26	70.79
2	A-G2	28.00	19.26	63.31
3	A-G3	26.33	21.77	99.91

CAPTIONS FOR FIGURES

Fig. 1. An example of gel electrophoresis experiment performed at pH 8.0 and analyzed by Gel Quant Express software. Lanes 1, 2 and 4 represent loaded samples with D/P ratio of 1.498 with the bright bands in the well (top) corresponding to the formed polyplex, and the lower migrated bands (bottom) correspond to the unbound dsDNA. Lane 3 represents a reference dsDNA whose signal intensity was attributed as 100%.

Fig. 2. Effects of factors and goodness-of-fit analysis: (**a**) percentage effect of factors on the predicted response; (**b**) agreement between experimental observations and predictions.

Fig. 3. Response surface plot and contour-line map showing the effects of [dsDNA] and [B-PEI] factors on the binding efficiency $\hat{Y}$ (%), for the fixed level of pH=7.0.

Fig. 4. Response surface plot and contour-line map showing the effects of [B-PEI] and pH factors on the binding efficiency $\hat{Y}$ (%), for the fixed level of [dsDNA]=28.33 μM.

Fig. 5. Gel electrophoresis assays carried out according to the experimental conditions given by gradient methodology. All assays have been performed at pH 5.8, and band intensities have been analyzed by Gel Quant Express software. Assay **A-G1**: lane 1, 2 – samples with D/P=1.367, average unbound dsDNA band intensity 29.21%; lane 3 – reference dsDNA, 26.33 μm. Assay **A-G2**: lane 1, 2 – samples with D/P=1.453, average unbound dsDNA band intensity 36.69%; lane 3 – reference dsDNA , 28.00 μm. Assay **A-G3**: lane 1 – reference dsDNA, 26.33 μm; lane 2, 3 – sample with D/P=1.209, average unbound dsDNA band intensity 0.09%. Assay **A-G4**: lane 1 – reference dsDNA, 25.55 μm; lane 2, 3 – sample with D/P=1.145, average unbound dsDNA band intensity 0.04%.

Fig. 6. Rendering of macromolecules dsDNA and B-PEI in a simulation cell with explicit water molecules (solvent): (**a**) initial equilibrated structures; (**b**) polyplex structure at t = 21 ns.

Fig. 7. Snapshots from trajectory showing the interactions between dsDNA and B-PEI with formation of the polyplex at pH 5.8 and different simulation times: (**a**) t=2 ns; (**b**) t=5 ns; (**c**) t=12 ns; (**d**) t=20 ns; (water molecules are omitted).

Fig. 8. Plots as functions of simulation time: (**a**) the distance between the centers of geometry of dsDNA and B-PEI; (**b**) the number of atoms in intermolecular contact for a cutoff radius of 4 Å.

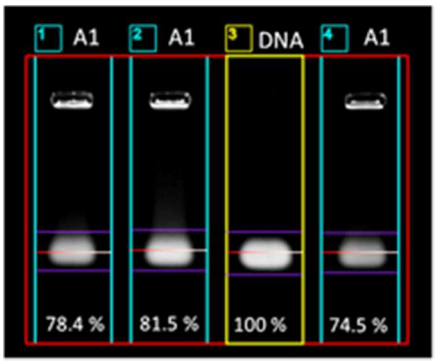
Fig. 9. Formation of hydrogen bonds between dsDNA and B-PEI as a function of simulation time: a) total number of hydrogen bonds; b) total energy of hydrogen bonds.

Fig. 10. Simulation map showing dsDNA nucleotides involved in the formation of the hydrogen bonds via backbone oxygen atoms (O1P, O2P, O3* and O5*): (**a**) nucleotides from sense strand; (**b**) nucleotides from antisense strand.

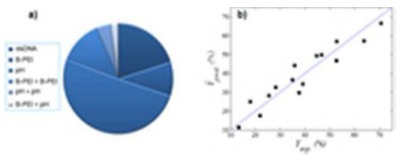
Fig. 11. Number and type of B-PEI nitrogens (amine groups) involved in the formation of hydrogen bonds versus simulation time.

Fig. 12. Radial distribution function (RDF) of B-PEI nitrogens around dsDNA backbone oxygens based on the last 21 ns trajectory of simulation.

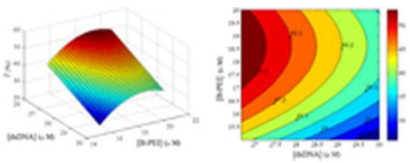
Fig. 13. Plots of conformational parameters for dsDNA and B-PEI versus simulation time: (**a**) radius of gyration (R_g); (**b**) root-mean-square deviation (RMSD).



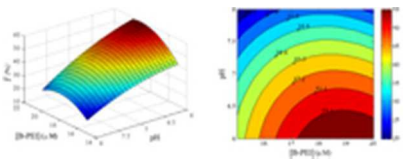
18x14mm (300 x 300 DPI)



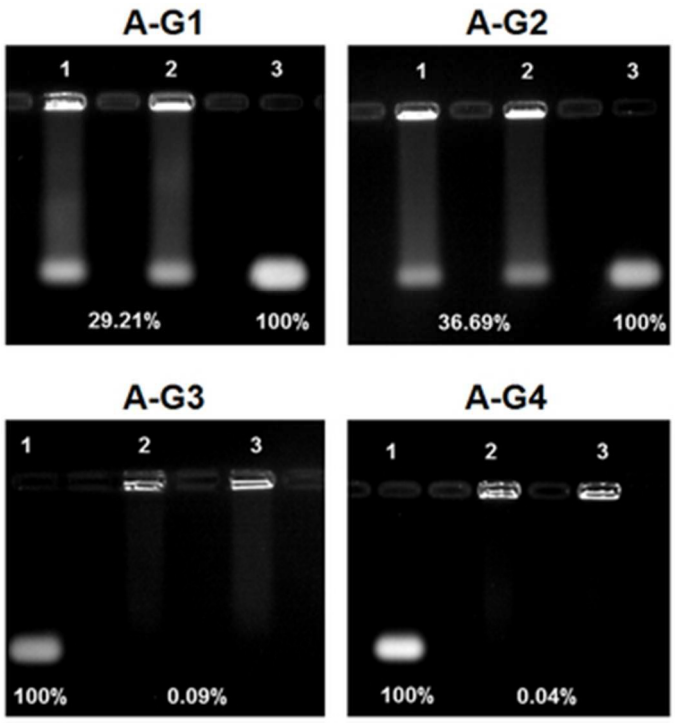
16x6mm (300 x 300 DPI)



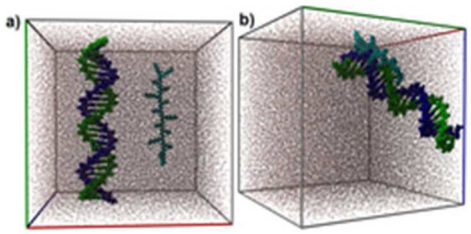
17x6mm (300 x 300 DPI)



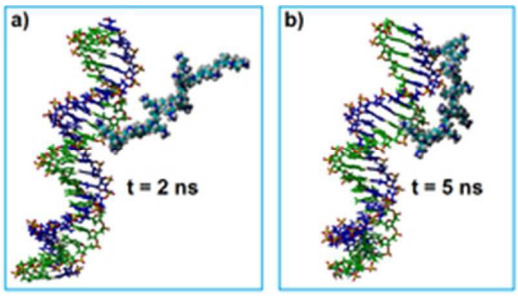
16x6mm (300 x 300 DPI)



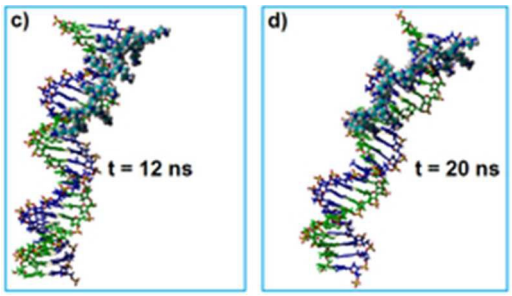
28x30mm (300 x 300 DPI)



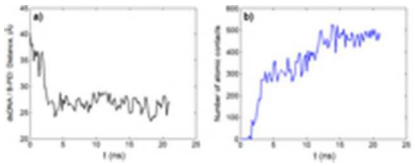
19x9mm (300 x 300 DPI)



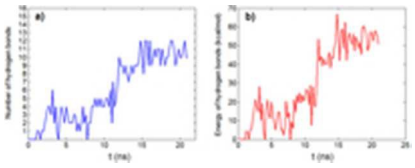
21x12mm (300 x 300 DPI)



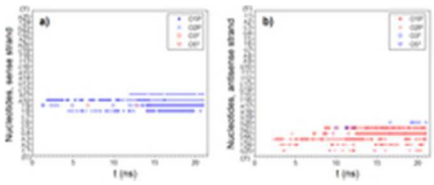
21x12mm (300 x 300 DPI)



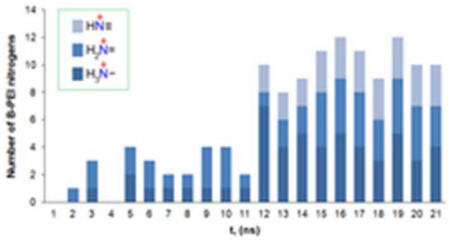
17x6mm (300 x 300 DPI)



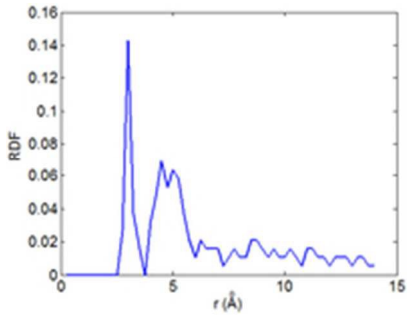
17x6mm (300 x 300 DPI)



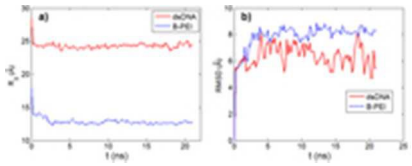
18x7mm (300 x 300 DPI)



18x9mm (300 x 300 DPI)



17x13mm (300 x 300 DPI)



17x6mm (300 x 300 DPI)