

Molecular Modeling as a visualization tool in design of DNA crosslinked polyacrylamide

Karin Rafaels^{1,*}, John Kerrigan², Noshir Langrana^{1,3}, and David Lin³

¹ Department of Biomedical Engineering, Rutgers University, Piscataway, NJ

² IST-Academic Computing Services, University of Medicine and Dentistry of New Jersey, Piscataway, NJ

³ Department of Mechanical and Aerospace Engineering, Rutgers University, Piscataway, NJ

* Rutgers Undergraduate Research Fellow

Keywords: [DNA Crosslinking](#), [polyacrylamide gel](#), [molecular modeling](#)

Abstract

Since DNA-crosslinked gels are likely to find a range of applications it is important to know how to tailor the gel composition for a particular application. In this study, polyacrylamide gel crosslinked with DNA has been assayed with respect to conformational energy and linker size using AMBER 7.0 software [1]. The molecular models generated in AMBER make it possible to estimate the mechanical properties of the gel as a function of crosslinker density, polyacrylamide density, and crosslinker length. The structure of an equilibrium state is computed using an explicitly solvated model, in which water was the solvent. Visual inspection of the model determines other mechanical properties of the gel and helps predict chemical interactions. A long-term goal of this work is to use computer assisted modeling techniques to guide the experiments, to predict linker stiffness, and to examine other mechanical properties of the DNA crosslinker.

Introduction

DNA is a versatile material for the exploration of nanoscale structures because its hybridization chemistry is very specific. DNA crosslinked gels, for example, have properties that depend on the base sequence in the DNA crosslinks, an observation that suggests an approach for engineering the nanoscale structure of the gels. Previous experimental research performed by our group demonstrated that a critical concentration of crosslinking DNA strands can lead to gel formation in polyacrylamide and that the subsequent (reversible) addition of specially synthesized DNA strands can significantly increase the stiffness and strength of the gel. The process is shown schematically in Figure 1. In our studies, by using software designed in-house, DNA sequences were optimized against the formation of secondary structures and undesired hybridization between strands to prevent binding of complementary strands and off-alignment binding.

However, it is extremely difficult to visualize the detailed events that occur during the formation of polyacrylamide-DNA gels. Computer modeling is a tool that enables the researchers to study the structural aspects of the newly engineered DNA crosslinkers. Complete calculations are computationally demanding, but simplifications can be employed. In this study, polyacrylamide gel

crosslinked with DNA has been assayed with respect to conformational energy and size using AMBER 7.0 software [1].

Computer simulations are an attractive approach to supplementing experimental data. DNA-based computing has been done by Deaton et al [2]. The accuracy and precision of DNA computing relies on error-free binding of DNA strands. DNA-based computing uses the tendency of nucleotide bases to bind in preferred combinations. Using simple generic force field models [3] and global parameterization and validation [4], the molecular modeling can be used to test the efficiency of hybridization between complementary sequences.

Prior work

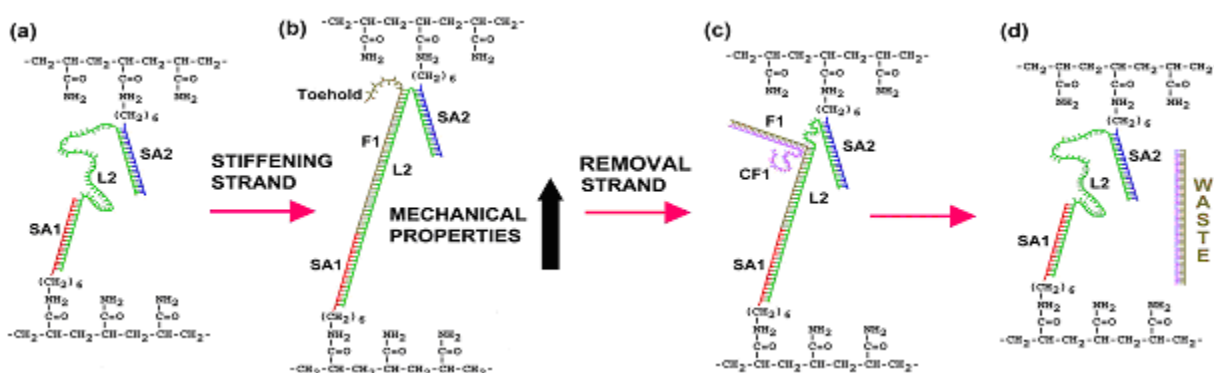


Figure 1.

(a) Polyacrylamide chains crosslinked by three DNA strands; (b) addition of the stiffening strand results in a swelling of the gel network and increased stiffness; (c) addition of the removal strand initiates strand displacement; (d) a waste product is formed.

Oligonucleotides were supplied by Integrated DNA Technologies (Coralville, IA). The two Acrydite™- modified strands (SA1 and SA2) of 20 nucleotides in length were separately polymerized with acrylamide monomer. Equal amounts of each mixture were mixed with the 80-mer crosslinking strand (L2) to form the crosslinked network shown in Figure 1a. The stiffening or fuel strand (F1) of 50- nucleotide length was designed with a 40-mer region of base sequence complementary to the single- stranded region of L2. The ten additional bases at the 3' end of strand F1 serve as a "toehold" where strand displacement through branch migration initiates. This process occurs with the introduction of the removal strand (the complement of F1), designated CF1. The time required for the F1 strands to migrate into the gel was approximately 60 minutes. Significant compression of the gel occurred during this time. The gel was observed for a subsequent period of 30 minutes to ascertain that the dye-labeled DNA were indeed attached to the crosslinks and not percolating the gel. Complete dissociation of the stiffening strands by strand displacement, through the introduction of CF1 by electrophoresis, required roughly four hours.

Table 1. Stiffness and modulus at each state

State	Stiffness (N/m)	Modulus (Pa)
Initial	0.9091	182.28
Stiffened	2.2211	445.35
Final	1.1644	233.47

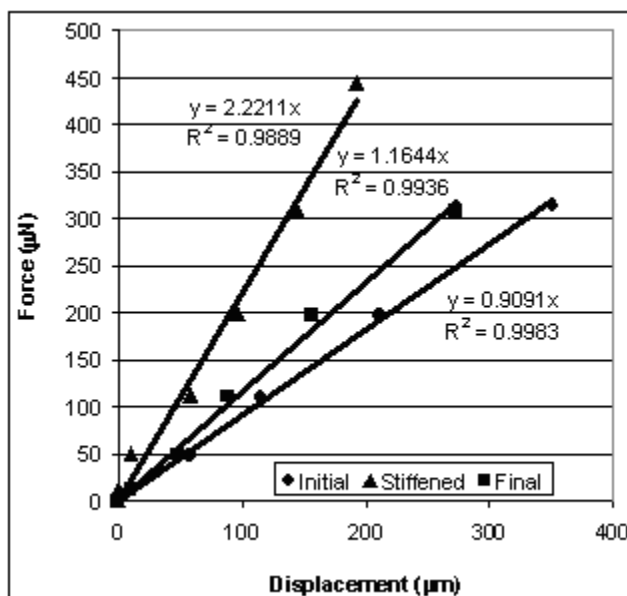


Figure 2. Force-displacement plots for the three states of the gel.

The force-displacement plots for the gel, before and after the introduction of F1, and after the re-annealing of the gel, are shown in Figure 2. All curves are linear with high R^2 values. The values for stiffness and Young's modulus at each state are presented in Table 1. In the stiffened state, stiffness and modulus increased by 144%. In the final state, stiffness and modulus were greater than their respective values in the initial state by 32%.

Objectives: Since DNA-crosslinked gels are likely to find a range of applications it is important to know how to tailor the gel composition for a particular application. It is also of interest to know what the composition is that would induce the greatest change in stiffness. Clearly, if the motor domain of the crosslink is very short, little prestress will build up and there will be little change in the elastic modulus of the material. On the other hand, if the motor domain of the crosslink is very long, the double stranded DNA resulting from the hybridization of F1 (see Figure 1) with the motor domain is likely to Euler buckle. The AMBER software has the ability to generate estimates of the mechanical properties of a gel as a function of crosslinker density, polyacrylamide density, and crosslinker length. These estimates can be checked against experimental results, which, in turn, allows the validation of theoretical models.

Materials

The mechanical properties of DNA crosslinked polyacrylamide gel with respect to energy and size are characterized using Assisted Model Building with Energy Refinement (AMBER) software. In this work, AMBER was run on a Silicon Graphics computer, as well as on a linux cluster. AMBER is one of the few molecular modeling programs that is commercially available, known, and well tested for DNA study [5]. Features of this software include:

- i. SANDER (Simulated Annealing with NMR-derived energy restraints), which is the basic energy minimizer and molecular dynamics program. In other words, SANDER searches for the most favorable physical configuration of a molecule by adjusting the positions of the constituent atoms until the molecule has the lowest possible energy.
- ii. PMEMD (Particle Mesh Ewald Molecular Dynamics) is a version of SANDER with improved performance for clusters of computers with the linux operating system; and
- iii. CARNAL is the coordinate analysis program.

In addition to AMBER, the design tools SYBYL (Tripos, Inc.) and MOE (Molecular Operating Environment, CCG, Inc.) are implemented. These molecule-building packages are employed to build the unknown structure of the polyacrylamide atom by atom as well as the known structure of DNA.

Approach

Most published studies that deal with modeling of biologically interesting molecules begin with a known structure, with perhaps a few unknown residues [6]. The structure of neither polyacrylamide nor the polyacrylamide/DNA adduct has not been studied much computationally, nor are they in the Cambridge Structural Database. The polyacrylamide solution and gel has been studied quite extensively, but not as the polyacrylamide-DNA hybrid. The reason in the latter case is presumably that the six-carbon chain connecting the DNA to the polyacrylamide is not normally found bonded to the polyacrylamide. A first estimate of the positions of atoms in the molecule, therefore, had to be made from scratch using SYBYL and MOE. Although both SYBYL and MOE use graphical interfaces, are user-friendly, and nominally allow identical molecules to be built problems encountered in SYBYL could only be circumvented by using MOE. The following table describes the different capabilities of SYBYL and MOE

Table 2: Comparison between SYBYL and MOE

Function	SYBYL	MOE
Minimization	Would not recognize DNA as DNA, just as atoms.	Able to minimize DNA and maintain DNA helical structure.
DNA Builder	Can build both single and double stranded DNA.	Text driven builder. Can modify the complimentary strand. Can only build double stranded DNA.
Molecule Movement	Cannot move two molecules in same screen separately.	Difficult to add molecules to pre-existing ones. Can move two molecules in same screen separately. Can obtain distances between atoms.

All the calculations on the DNA crosslinked polyacrylamide model built in SYBYL and MOE are performed under the AMBER umbrella. AMBER can calculate a total minimum energy that includes potential, kinetic, and bond energies and the structure of the molecule at that particular energy, However, there are a few preliminary steps necessary . AMBER requires that all files be parameterized in a special file format. This step is done using a program in AMBER called antechamber.

To stay within program limits, the choice was made to work only with a small piece of polyacrylamide seven monomer units per polymer chain from which a general trend would be inferred. The small piece of polyacrylamide was then parameterized. The parameterization first gives a "prep input" file as an output. This prep file is prepared using the AM1-BCC method, where BCC stands for bond charge corrections. This method generates high-quality atomic charges as the name indicates, but is best with a charge set for DNA. Parmchk, which means parameter check, is then run on the prep file to add any missing parameters, i.e. bond lengths, bond angles, torsional values. These two programs use information in the standard database that is in AMBER. Parameterization is not required for DNA because it is well known and part of the program's database.

Parameter files for polyacrylamide are not part of the standard AMBER data base. The relevant information is, however, available in parm99 and koll94, which are tools for AMBER users; these data were entered manually. After these two parameter files are set up successfully, the file can be read into xLEaP, a program within AMBER that provides a platform in which pdb files can be modified. The parameter file contains the properties of the atom, including charge and atom type. And the pdb file includes the atom names and positions. Here, xLEaP was used to include water in the structure and to include Na⁺ for charge balance. Even using an octahedral periodic box as the volume of water within which the calculations take place, the number of water molecules added is 10⁴. This number is large for the volume considered; the significance is that there is a large amount of solvent in comparison to the amount of DNA/polyacrylamide. The files created in xLEaP are then saved as prmtop, parameter topology files, and inpcrd, input coordinate files.

After this step, SANDER is invoked. SANDER is the basic energy minimizer and molecular dynamics program in AMBER. It finds the optimal structure of a molecule by searching for the atomic coordinates that minimize the conformational energy. A determination of the absolute minimum energy in a polyatomic system would be prohibitively expensive in terms of computer time. As a compromise, the user sets a value that tells the computer to stop searching atomic coordinates when the improvements (decreases) in energy fall below a certain value. The molecular dynamics portion generates configurations of the system by integrating Newtonian equations of motion. The molecular dynamics approach allows small potential energy barriers to be crossed over. Because the calculations take place on the molecule and the added water, SANDER can take many hours, days, even weeks to run.

The molecular dynamics calculations provide a great deal of information about the conformation and structure of the crosslinked polymer, but it is hard to extrapolate engineering properties. Fortunately, AMBER has a program called CARNAL. CARNAL is the coordinate analysis program in AMBER. It uses a flexible command language. It analyzes trajectory measurements of each atom, as well as comparisons between multiple streams of coordinates. The output from CARNAL can be read in Excel as a comma delimited file. From Excel, plots are made to see at what times the structure had a minimum energy and where an equilibrium structure occurred. Equilibrium structures do not necessarily have to be at the minimum energy, they occur when the atoms in the molecule do not change much during the calculations. These graphs are used to create average structures from the equilibrium points determined from the root mean squared distance plot. These models give us an idea of what the molecular structure looks like.

Current Computational Design

The file of the model made in SYBYL and MOE must be converted to AMBER file formats. The polyacrylamide fragments were parameterized manually using the Amber 99 parameter set [3] and RHF/6-31G* derived partial atomic charges computed using the Gaussian 98 program [7].

Sander is the basic minimizer and molecular dynamics program in AMBER. Sander determines the minimum energies of the models, the position-restraint molecular dynamics, and molecular dynamics calculations.) In all these calculations, the pressure was held constant at 1 bar, the temperature was held constant at 300 K, and the volume of the solvent box was altered. Minimum energies were determined for both the long and short molecules in their unstiff and stiff states. The polyacrylamide gel model is solvated with approximately 10^4 water molecules, and charge is neutralized with sodium ions. All of these additions greatly increase the size of the model, which increase the calculation times. Because SANDER can take many hours, days, even weeks to run, PMEMD is used for faster calculations for the molecular dynamics calculations on a linux cluster.

CARNAL is the coordinate analysis program in AMBER that allows graphs to be plotted, showing the locations of equilibrium structures and distances between polyacrylamide chains. The root mean square deviation (RMSD) plots of the atoms are used to create average structures from the equilibrium points determined from the graph. These models provide an idea of what the molecular structure looks like. The distance between polyacrylamide chains and a Fourier curve fit to the RMSD data is performed to compare the stability of various DNA crosslinked polyacrylamide gel structures. To model our experimental research setup [8] unstiff and stiff models were developed, as seen in Figure 3.

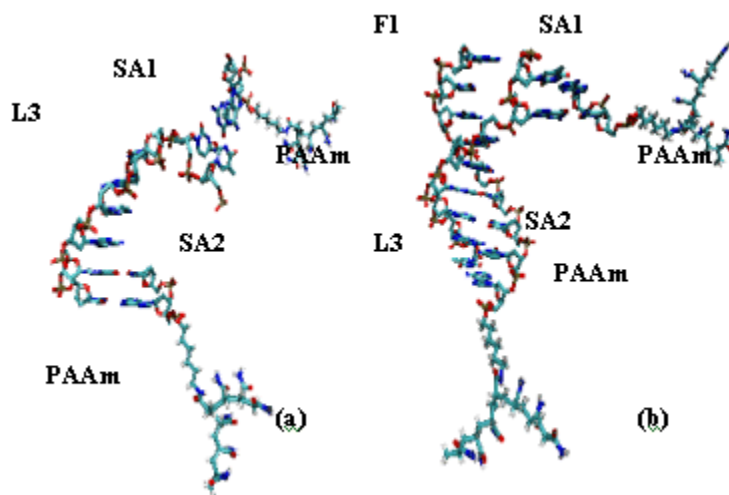


Figure 3.
(a) Polyacrylamide (PAAm) chains crosslinked by SA1, SA2 and L3 strands.
(b) The stiffening strands F1 were added

Two models were built to simulate both the unstiff, without binding, and stiff, when the fuel strand binds to L2, states of the gel. The actual number of bases in the DNA crosslink of the gel is eighty. Initially, in order to work within the computing capabilities, an unstiff and stiff short molecule were built with SA1 and SA2 lengths of 2 DNA bases, an L3 length of 8 bases, and an F1

length of 7 bases. The number of bases in the model was eight. However, with such a small number of bases, the crosslinks began to "unzip". To reduce the "unzipping" of the DNA crosslinks, four more bases were added to the crosslinks to increase the number of bonds holding the complementary strands of DNA together. This made the long unstiff and stiff molecule have SA1 and SA2 lengths of 4 DNA bases, an L3 length of 16 bases, and an F1 length of 11 bases. This longer model was used for the comparison with the shorter model. Computations on this model are performed on several computers, the Silicon Graphics computers, as well as a Linux cluster).

Initially, the models are built and parameterized on an SGI workstation at University of Medicine and Dentistry of New Jersey (UMDNJ). Then the minimizations and position-restraint molecular dynamics are calculated on a single Linux computer also at UMDNJ. The molecular dynamics calculations are run on at least 4 Linux workstations at the School of Engineering, running in parallel to reduce the amount of time needed to perform the calculations.

Results and Discussion

Average structures were created from the range of equilibrium points determined from the plot of root mean square deviations (RMSD) of the atoms. Changes in conformation occur over time, but at some point reach equilibrium. The model at equilibrium is the desired template for other models. The equilibrium point is found where the average root mean squared distance (RMSD) of the atoms levels off.

The unstiff molecule in Figure 4 does not show any areas of stability. Where stability is defined as a time period over which the energy of the system is lower than in adjacent regions. However, the stiff molecule shows some stability in the area within the circle. A Fourier curve fit to the RMSD is shown in Figure 4. As seen in Table 3, the coefficients a_0 , a_1 , a_2 , a_3 , and b_1 and b_2 have significant differences in the oscillations. These coefficients quantify the differences between the stiff and unstiff molecules.

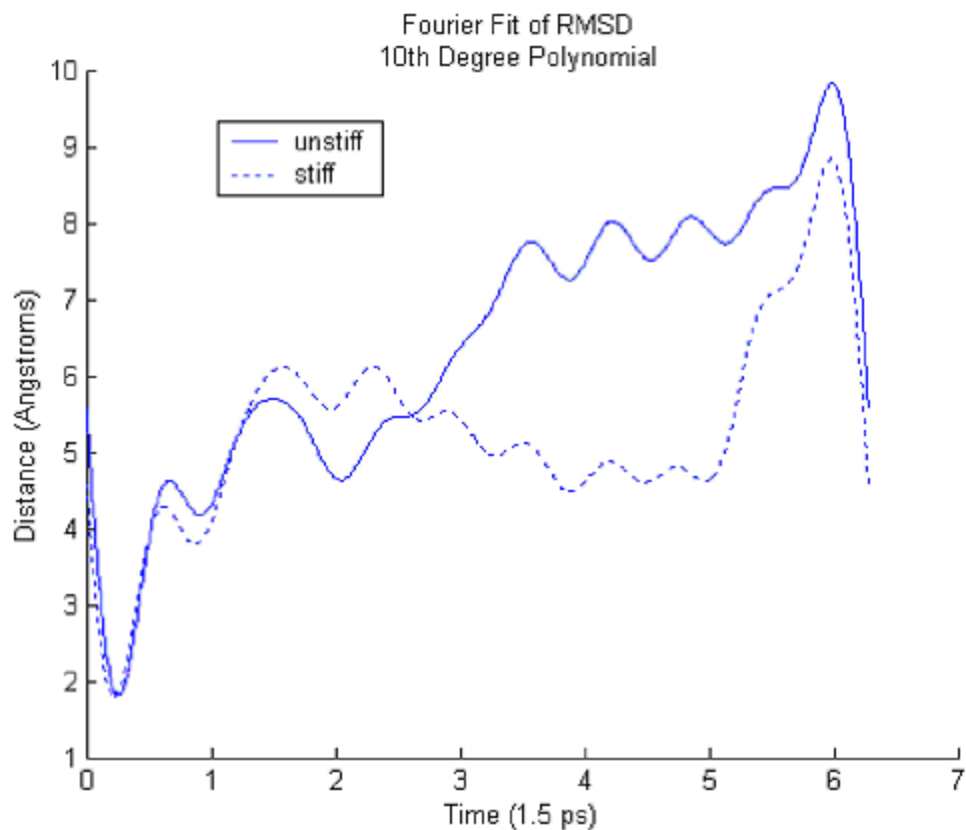


Figure 4.
Fourier analysis on unstiff and stiff molecule RMSD data

Table 3: Fourier coefficients for the Long Unstiff and Stiff Molecules

	a_0	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}
Unstiff	6.3891	0.2484	0.1481	0.2290	0.0329	0.0711	0.1887	0.0875	0.0087	0.1084	0.0088
Stiff	5.2784	0.0528	0.0020	0.0140	-0.1221	0.1972	-0.2147	0.1303	-0.0053	0.0170	-0.0908
	b_0	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_{10}
Unstiff	0.0	1.9084	0.6367	1.0062	-0.6647	0.4853	-0.3728	0.5201	-0.3015	0.2186	-0.2590
Stiff	0.0	0.1229	1.1813	0.9608	-0.6591	0.3516	-0.2285	0.4028	-0.3903	0.2274	-0.2365

Distances between the polyacrylamide chains were also calculated and plotted as seen in the Figure 5 to indicate stiffness. The distance was determined by measuring the distance between the center of gravity of each polymer using CARNAL and VMD (Visual Molecular Dynamics). The average distance for the long unstiff molecule was 55.7Å and 45.7Å for the long stiff molecule. This

implies that the stiffness increased by 10%, when the complementary DNA strand was added. Shorter distances imply stiffness because the molecule does not have as much freedom to move about.

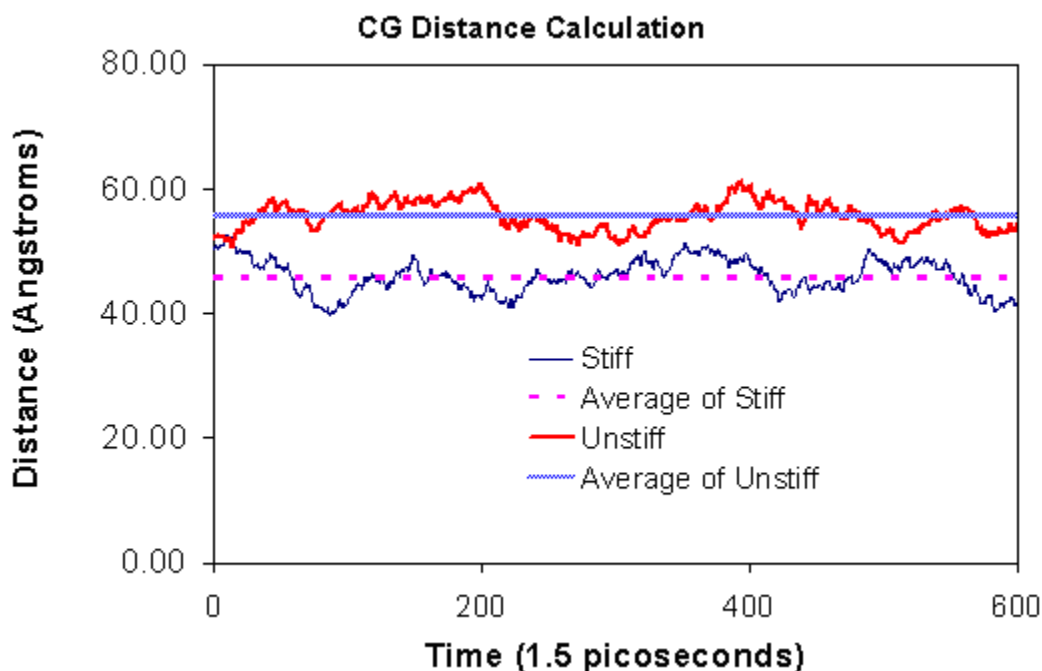


Figure 5.

The minimum energies were determined for each of the four models. The minimum energy for the long unstiff molecule was -1.6361×10^5 kcal/mol. The minimum energy for the long stiff molecule was -1.9125×10^5 kcal/mol. The lower energy for the stiff molecule is another indication of the stiff molecule having more stability, hence a stiffer conformation.

Conclusions

For the first time the structure and certain properties of plain polyacrylamide gel and of polyacrylamide gel cross linked with DNA have been modeled on a molecular level. The modeling saves both time and money and the results agree with experiment. In the laboratory, experimental studies of the polyacrylamide/DNA gels take two days to run with more time needed for preparation. Now that initial parameter files for building the molecular models are in hand, computer calculations take three to four days to complete. These results show that the simulations can provide a valid, predictive tool for future experiments.

Although molecular modeling has been used for a two decades, it is still evolving and problems involving solvation and position restraint are still being refined. Many assumptions were necessary to obtain the simulations, and there were inherent limitations in the computer programs, analysis tools, and in the computers, themselves. Therefore the values obtained are better indicators of trends than of absolute values.

Future Work

Future work will model molecules that vary crosslinker length, change DNA base sequences and change the ratios between SA1 or SA2 with L3 (see figure 1). Testing different crosslink densities, DNA base sequences, crosslink lengths, and so on, can be used to obtain an optimized model of the gel.

The Gibbs module with in AMBER may be used to obtain free energies. The free energies determine the mechanical properties of the constituent components of the model. Thus, the material properties as a whole can be better understood.

NMODE, also within AMBER, maybe used to find the transition states between conformation states) of the molecule. Comparing the transition states between molecules indicates their relative stability.

Molecular modeling provides an inexpensive design tool that validates theoretical models and provides a predictive tool to guide additional experiments.

Acknowledgments

This study was performed in the partial fulfillment of the Slade Scholar program of the first author. We would like to thank Alexei Kotelnikov for all his assistance in this project. The facility support from Rutgers University and the University of Medicine and Dentistry of New Jersey is highly appreciated.

References

1. Case, D.A., et al., *AMBER 7*. 2002, San Francisco, CA.: University of California San Francisco.
2. Deaton, R., et al., *Reliability and efficiency of a DNA-based computation*. *Phys Rev Lett*, 1998. **80**: p. 417-20.
3. Wang, J., P. Cieplak, and P.A. Kollman, *How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules?* *J Comput Chem*, 2000. 21(12): p. 1049-10174.
4. Jakalian, A., D.B. Jack, and C.I. Bayly, *Fast efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation*. *J Comput Chem*, 2002. **23**: p. 1623-1641.
5. Cheatham III, T.E. and M.A. Young, *Molecular dynamics simulation of nucleic acids: Successes, limitations and promise*. *Biopolymers*, 2001. **56**: p. 232-256.
6. Kalra, P., A. Das, and B. Jayaram, *Free-energy analysis of enzyme-inhibitor binding: aspartic proteinase-pepstatin complexes*. *Appl Biochem Biotechnol*, 2001. **96**: p. 93-108.
7. Frisch, M., et al., *Gaussian 98*. 1998, Pittsburgh, PA: Gaussian, Inc.
8. Lin, D.C., B. Yurke, and N.A. Langrana, *Mechanical properties of a reversible, DNA-crosslinked polyacrylamide hydrogel*. *J Biomech Eng*, 2004. **126**(1): p. 104-10.

Copyright 2004 by Rafaels.