A General Shape Equation for Local Regular Structure of Biomolecular Chains

Liu Hong
Zhou Pei-Yuan Center for Applied Mathematics
Tsinghua University
Beijing, China, 100084
Email: hong-l04@mails.tsinghua.edu.cn

Jinzhi Lei
Zhou Pei-Yuan Center for Applied Mathematics
Tsinghua University
Beijing, China, 100084
Email: jzlei@mail.tsinghua.edu.cn

***Abstract*—A general shape equation for the local regular structure of biomolecular chains at the equilibrium state is established. It predicts a general relationship between the structural curvature and torsion, which only concerns about the elastic property of the molecular chain, and is independent of variable conformations of real biomolecules. Solutions corresponding to α-helix and β-hairpin in proteins, helical DNA, as well as spiral molecules are discussed, which show a fairly well agreement with experimental data.**

I. Introduction

It is generally believed that the native three-dimensional structure of biomolecules serves as a prerequisite for their biological functions[1]. Such as during the processes of replication and transcription of DNA, the translation of messenger RNA, and the binding and dissociation of proteins etc, the shapes and topological properties of molecular chains (such as proteins and DNA) all play a significant role[2]. Thus to determine the native structure of biomolecules becomes a central aim of modern structural biology. It will significantly enhance our understandings on the biological processes involved in life, and also has broad applications in medicine, food and materials etc.

During the past decades, plenty models have been suggested to describe the molecular chains. For example, the wormlike chain model[3] was established for DNA structure under small external forces; while the wormlike rod chain model[4] is more appropriate for a moderate force. And the equilibrium shape equations of vesicle membranes were derived by Ou-Yang and Helfrich[5], [6], by which some characters of membranes have been carefully examined. However, to the best of our knowledge, the general equilibrium shape equations of chain biomolecules have not been established yet, by which the native three-dimensional molecular structure can be determined precisely, and its dynamic behaviors can be understood well[7].

In this paper, we propose a general equation for the local regular structure of chain-like biological molecules at the equilibrium state. The equation is established based on the study of elastic energy density of chain molecules, which is expressed as a function of the curvature and torsion, and is independent to varied conformations of real biomolecules. From this shape equation, different solutions which correspond to various natural structures of biomolecules, like α-helix and β-hairpin in proteins, helical DNA and spiral molecules, could be found. When applied to real helical conformations of protein and DNA, two sets of parameters, $(3.26, -10.70)$ for protein and $(29.07, -135.69)$ for DNA, are obtained, which can effectively characterize the different elastic properties of protein and DNA. Same set of parameters has also been used to study the conformation of β-hairpin, which predicts a maximum loop length ≤ 10 monomers.

The following paper is organized as: In Section II, a general shape equation for the local regular structure of chain-like molecules at the equilibrium state is derived. In Section III, several representative solutions are discussed, which correspond to straight molecules, helical conformations of protein and DNA, β-hairpin and spiral molecules etc. Section IV is a brief conclusion.

II. Derivation of shape equation

Many important biological molecules, such as DNA, RNA and proteins, are linear (consecutive and unbranched), with the chain length much larger than the width[2]. Thus we can approximate them by one-dimensional smooth curves[8], and denote the centerline as $\mathbf{r} = \mathbf{r}(s)$, where s is the arc length. Then in general the elastic energy of a chain biomolecule can be written as:

$$F_e = \int_0^l g(\mathbf{r}(s))ds, \tag{1}$$

where $g(\mathbf{r}(s))$ is the elastic energy density, sensitively depending on the local geometric shape and elastic properties of the molecular chain; l is the chain length.

According to classical elastic theory[9], [10], the elastic energy density can be further represented as a scalar function of the tangent vector $\mathbf{t} = d\mathbf{r}/s$ and its derivatives $d^n\mathbf{t}/ds^n (n = 1, 2, \cdots)$. Moreover, due to the rotational symmetry around $\mathbf{t}$, only those scalars that are invariant under simultaneous reversal of s and $\mathbf{t}$ can occur[9], [10]. Thus there is no first-order invariant. The only independent second-order invariant is $\frac{d\mathbf{t}}{ds} \cdot \frac{d\mathbf{t}}{ds}$. Among the third-order invariants, we choose two independent ones: $\mathbf{t} \cdot (\frac{d\mathbf{t}}{ds} \cdot \frac{d^2\mathbf{t}}{ds^2})$ and $\frac{d}{ds}(\frac{d\mathbf{t}}{ds} \cdot \frac{d\mathbf{t}}{ds})$.

Consequently the elastic energy density (up to third-order)

2011 IEEE International Conference on Systems Biology (ISB)
978-1-4577-1666-9/11/$26.00 ©2011 IEEE

can be expressed as

$$g = \lambda + c_0 \frac{d\mathbf{t}}{ds} \cdot \frac{d\mathbf{t}}{ds} + c_1 \mathbf{t} \cdot (\frac{d\mathbf{t}}{ds} \cdot \frac{d^2\mathbf{t}}{ds^2}) + c_2 \frac{d}{ds}(\frac{d\mathbf{t}}{ds} \cdot \frac{d\mathbf{t}}{ds}), \quad (2)$$

where λ is the spontaneous curvature; c_i are elastic coefficients. Using the formulas for curvature $\kappa = |d\mathbf{t}/ds|$ and torsion $\tau = \mathbf{t} \cdot (d\mathbf{t}/ds \times d^2\mathbf{t}/ds^2)/\kappa^2$ in differential geometry[11], we can eventually write Eq. 2 as

$$g = \lambda + (c_0 + c_1\tau)\kappa^2 + c_2(\kappa^2)' \quad (3)$$

The above formula coincides with the one suggested by Helfrich to third-order, except for the last term that he neglected[12]. According to Ou-Yang[13], it is "sufficient for a description of the centerline of uniform elastic Kirchhoff rods in equilibrium".

To determine the natural biomolecular structure, a popular traditional method is to minimize the elastic energy F_e with respect to the position vector $\mathbf{r}(s)$ by variational approach[5], [6], [14], as it is a common belief that the natural conformation of a biomolecule corresponds to the state with minimum free energy. However in many cases, this method turns out to be not so effective. One reason is the neglect of chemical interactions between nonlocal contact monomers, which is essential for the maintenance of biomolecular structure; and sole optimization of the elastic energy will not necessarily lead to a naturally optimized structure in principle (see Appendix A).

Another more fundamental reason is the neglect of stochastic force. In general, for local structures of biomolecules, due to the small system size and relatively weak chemical interactions between monomers ($\sim 1{-}10kJ/\mathrm{mol}$), the environmental stochastic forces ($\sim 1kJ/\mathrm{mol}$) can not be directly neglected in principle[15]. Furthermore the assumption of minimal free energy may not be applicable too.

In current paper, we will look for an alternative way. Generally speaking, structure bending and twisting are direct consequences of the chemical interactions between different monomers (F_c), such as hydrogen bonding, van der Waals interaction, hydrophobic effect etc[15], [16]. Although there is no strict relation between the elastic energy and chemical interaction energy, a general observation in our previous MD and MC simulations of protein folding[17] is that the stronger the monomer interaction is, the larger the molecular chain can be distorted. Therefore there is a positive correlation between the elastic energy and monomer interaction energy (neglect the negative sign) for a given molecular chain.

Due to its nonlocal nature, F_c can hardly be expressed through local position vector $\mathbf{r}(s)$. But since we are dealing with local regular structure of biomolecules, we can assume that it distributes uniformly within the structure. Furthermore if we suppose the elastic energy density is linearly proportional to the energy density of chemical interactions, a most simple form in applications, we get

$$g \propto F_c/l. \quad (4)$$

The validity of above formula can be checked through the comparison of its predictions with the experimental data, though a direct confirmation still requires plenty of future works. Substituting the formula for elastic energy density (Eq. 3), we reach a general shape equation that describes the local regular structure of chain biomolecules at the equilibrium state, i.e.

$$(c_0 + c_1\tau)\kappa^2 + c_2(\kappa^2)' = \gamma\frac{F_c}{l} - \lambda. \quad (5)$$

Here it should be noted that above equation can be misleading in the prediction of global conformations of biomolecules, since our present derivation is only valid for local regular structures.

III. Results and discussion

To examine the validity of Eq. 5, we discuss some possible solutions that may correspond to various natural conformations of biomolecules.

A. Straight line

Firstly when $F_c = \lambda = 0$, $\kappa' = 0$ and $\tau = 0$, we can easily obtain the vanishing curvature solution

$$\kappa = 0, \quad (6)$$

which gives rise to a conformation of straight biopolymer in the absence of external constraint.

B. Helical conformation

Helix is one of the most observed biological conformations in nature. As we know, DNA can adopt a variety of structures: A, B and Z, but crucially they are all helical. And the DNA double helix is stabilized primarily by two forces: hydrogen bonds between nucleotides and base-stacking interactions among the aromatic nucleobases[18]. While for proteins, right-handed α-helix, which is maintained by consecutive hydrogen bonds between i and $i+4$ residues, constitutes a common motif in the secondary structure[19].

Thus if we assume $\kappa' = 0,\ \lambda = 0$, Eq. 5 possess a solution in which the curvature and torsion satisfies

$$\frac{1}{\kappa^2} = \frac{c_0}{\gamma h_0} + \frac{c_1}{\gamma h_0}\tau, \quad (7)$$

where $h_0 = F_c/l = n\epsilon_h/l$ is the average interaction energy density (ϵ_h is the interaction energy for single monomer, n is the number of monomers within chain length l). So for helical structures, $1/\kappa^2$ and τ are expected to be linearly related through two dimensionless parameters $c_0/(\gamma h_0)$ and $c_1/(\gamma h_0)$, which should only depend on the elastic and chemical properties of a molecular chain, instead of real helical conformations.

To test above relationship, we studied the experimental data of several kinds of helical conformations in proteins and DNA. As listed in Table I, the helical conformation is described through

$$\mathbf{r}(s) = (r_0\cos\omega s, r_0\sin\omega s, \frac{md}{2\pi}\omega s),$$

where r_0 is the radius of a helix; m is the number of monomers per turn; d is the raise per monomer, positive for right-handed

978-1-4577-1666-9/11/$26.00 ©2011 IEEE

helix, and negative for left-handed helix. Let $p = md/2\pi$, $\omega = 1/\sqrt{(r_0^2 + p^2)}$, then these parameters associate with κ and τ through

$$\kappa = \frac{r_0}{r_0^2 + p^2}, \quad \tau = \frac{p}{r_0^2 + p^2}. \tag{8}$$

From Fig. 1(A) and 1(B), we can see that $1/\kappa^2$ and τ obey a strict linear relationship for various experimentally identified helical conformations of protein and DNA, as predicted by Eq. 7. And during the data fitting process, we can further determine the model parameter pair as $(c_0/(\gamma h_0), c_1/(\gamma h_0)) = (3.26, -10.70)$ for protein, and $(29.07, -135.69)$ for DNA. In both cases, the coefficients for twisting are negative. This means the left-handed helix will have higher elastic energy density than the right-handed one, which explains why right-handed helix is more often observed in nature[2]. Moreover, the elastic coefficient for DNA is about one order larger than that for protein. This hints that DNA is more rigid than protein[2], whose physical foundation lies on that DNA is made up of double-helical chains, while protein is single-stranded.

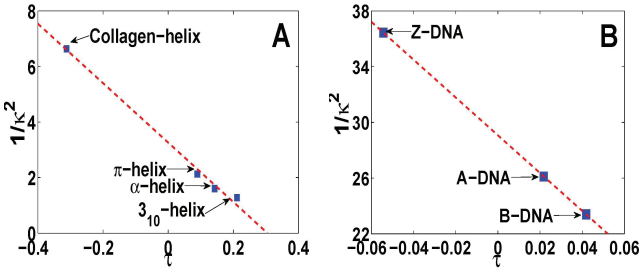

Fig. 1. Linear relations between $1/\kappa^2$ and τ for helical conformations of (A) protein and (B) DNA.

Therefore the model parameter pair (c_0, c_1) turns out to be an effective index to characterize the elastic properties of various biomolecules. Compared to the relationship predicted by traditional variational method[14], which fails to identity these characteristic parameters (see Appendix A), our prediction appears to be more reasonable.

C. β-hairpin

β-hairpin, one of the simplest supersecondary structures, are widespread in globular proteins, and have often been suggested as possible sites for nucleation[21]. It can be modeled by a combination of a planar circular loop region and two straight β-strands which are connected by consecutive hydrogen bonds. The solution for a straight line have been introduced in previous section; while for the loop region, we have $\tau = 0, \kappa' = 0$, $\lambda = 0$ and $F_c = \epsilon_h$ (the first pair of hydrogen bond makes the major contribution to maintain the loop region). Then Eq. 5 possess a special circle solution

$$\kappa = \sqrt{\frac{\gamma h_0}{(m+2)c_0}}, \tag{9}$$

where m is the number of monomers in loop region; $h_0 = \epsilon_h/l_0$ is the energy density for hydrogen bonds; $l_0 = 3.88\mathring{A}$ is the mean distance between neighbor residues.

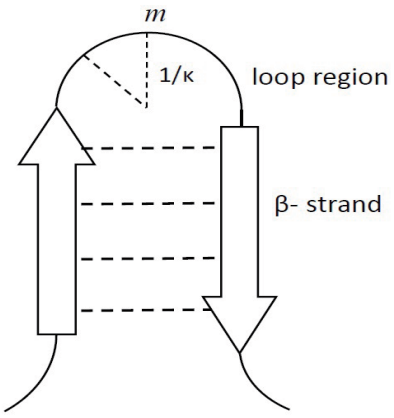

Fig. 2. Model for β-hairpin.

Substituting the value $c_0/(\gamma h_0) = 3.26$ obtained from fitting the helical conformation of proteins, we found that the length of loop region in β-hairpin can not exceed 10 monomers (a list of curvatures for all possible loops is given in Table II), otherwise it would be unable to form a circle (a part of circle to be exact). This result qualitatively agrees with the observation that most β-turns with loop length ≤ 7[22].

TABLE II
ALL POSSIBLE LOOPS FOR β-HAIRPIN.

Loop length (m)	Curvature of loop (κ)
1	0.320
2	0.277
3	0.248
4	0.226
5	0.209
6	0.196
7	0.185
8	0.175
9	0.167
10	0.160

Above discussions are also applicable to cDNA[23] and other circular molecules, although the interactions which induce bending may be different from hydrogen bonding as in β-hairpin.

D. Spiral structure

If we suppose τ as a const, a general solution of Eq. 5 is obtained as

$$\kappa(s) = \sqrt{\kappa_0 \left(1 - \alpha e^{-s(c_0 + c_1\tau)/c_2}\right)} \tag{10}$$

where α is a constant; s is the arc length; and $\kappa_0 = (\gamma h_0 - \lambda)/(c_0 + c_1\tau)$ is the initial curvature at $s = 0$. In general, Eq. 10 describes a spiral structure which have been widely observed in virus, DNA package and many other molecules[24].

TABLE I
PARAMETERS FOR HELICAL CONFORMATIONS IN PROTEINS AND DNA[20].

Helix	Monomers per turn (m)	Rise per monomer (d)	Radius (r_0)	Energy density $(h_0)^{(a)}$	Curvature (κ)	Torsion (τ)
α-helix	3.6	$+1.5$	2.3	0.23	0.38	0.14
3_{10}-helix	3.0	$+2.0$	1.9	0.22	0.42	0.21
π-helix	4.3	$+1.1$	2.8	0.24	0.33	0.089
collagen-helix	3.3	$-2.9^{(c)}$	1.6	$0.71^{(b)}$	0.33	$-0.31^{(c)}$
A-DNA	11	$+2.3$	13	0.13	0.070	0.022
B-DNA	10	$+3.4$	10	0.14	0.077	0.042
Z-DNA	12	$-3.8^{(c)}$	9	0.17	0.067	$-0.054^{(c)}$

$^{(a)}$ The average interaction energy density is estimated as $h_0 = \frac{m\epsilon_h}{2\pi\sqrt{r_0^2+(md/2\pi)^2}}$.

$^{(b)}$ Since collagen-helix is formed by three proteins, its energy density is three times of normal value.

$^{(c)}$ Left-handed.

Particularly, if $c_0 + c_1\tau = 0$, $\kappa(s) = \sqrt{\alpha + s(\gamma h_0 - \lambda)/c_2}$; if $c_2 = 0$ or $\alpha = 0$, we reobtain Eq. 7 for the helical conformation.

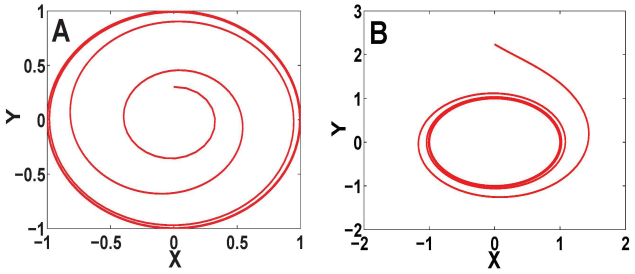

Fig. 3. Typical planar spiral solutions according to Eq. 9. (a) Outer spiral with $c_0/(\gamma h_0) = 1,\ c_0/c_2 = 1/3,\ \alpha = -10$. (b) Inner spiral with $c_0/(\gamma h_0) = 1,\ c_0/c_2 = 1/5,\ \alpha = 0.8$.

IV. CONCLUSION

In this paper, we have derived a general shape equation (Eq. 5) for the local regular structure of chain biomolecules at the equilibrium, based on the study of elastic energy density and chemical interactions for chain molecules. Different solutions which correspond to various natural structures of biomolecules, like α-helix and β-hairpin in proteins, helical DNA and spiral molecules, were discussed. When applied to real helical conformations of protein and DNA, we found that parameters (c_0, c_1) have provided fairly good description for the elastic properties of protein and DNA, by which various types of helical conformation were linked through a general relation between curvature and torsion (Eq. 7). Meanwhile traditional variational approach that was based on the assumption of minimum elastic energy failed to obtain such a relationship (see Appendix A).

Therefore a major application of our current study is to facilitate the structure prediction of biomolecules, which is a very hot topic in modern molecule biology, and have not been well solved yet. By providing a first-step rough geometrical description of the native structures of biomolecules with our current model, following computer-aided refinements could be relatively easily made to achieve high-resolution atomic-level structural details. Another potential application is that our model provides a new way to characterize the elastic property of different molecular chains by studying their native geometrical conformations (just as what we have done in Fig. 1). This would be interesting in AFM studies.

APPENDIX A
COMPARISON WITH VARIATIONAL APPROACH

A general variational equation for the elastic energy density, which is formulated as a function of curvature κ and torsion τ, is given by Zhang[14], i.e.

$$\frac{d^2}{ds^2}(\frac{2f_2'\tau}{\kappa} + f_1') + \frac{d}{ds}(\frac{2f_2'\kappa'\tau}{\kappa^2} + \frac{3f_2'\tau'}{\kappa}) + f_1'(\kappa^2 - \tau^2)$$
$$-f_2'(2\kappa\tau - \frac{\kappa'\tau'}{\kappa^2} + \frac{\tau''}{\kappa}) - f\kappa + f_3'(3\kappa\kappa' - 2\tau\tau')$$
$$-\frac{d}{ds}[f_3'(\kappa^2 - \tau^2)] - \frac{d^3}{ds^3}f_3' = 0, \quad (11)$$
$$\frac{d^3}{ds^3}(\frac{f_2'}{\kappa}) + \frac{d^2}{ds^2}(\frac{f_2'\kappa'}{\kappa^2}) + \frac{d}{ds}[\frac{f_2'}{\kappa}(\kappa^2 - \tau^2) - 2\tau f_2'] + f_1'\tau'$$
$$-f_2'\kappa' + \frac{f_2'\tau'\tau}{\kappa} + \frac{d^2}{ds^2}(2f_3'\tau) - \frac{d}{ds}(3f_3'\tau') + f_3'\tau'' = 0, \quad (12)$$

where $f = f(\kappa(s), \tau(s), \kappa'(s))$ is the elastic energy density, $f_1' = \partial f/\partial\kappa,\ f_2' = \partial f/\partial\tau,\ f_3' = \partial F/\partial\kappa'$. Substituting the elastic energy density we derived in the methodology section $g = \lambda + (c_0 + c_1\tau)\kappa^2 + c_2\kappa\kappa'$, we get following shape equations as:

$$c_0(\kappa^3 - 2\kappa\tau^2 + 2k'') + c_1(6\kappa''\tau + 6\kappa\tau'' + 14\kappa'\tau',$$
$$-2\kappa\tau^3 - \kappa^3\tau) - \lambda\kappa = 0 \quad (13)$$
$$2c_0\kappa\tau' + 2c_2(\kappa''\tau + \kappa'\tau') + c_1(2\kappa''' + 2\kappa^2\kappa'$$
$$-4\kappa\kappa'\tau - \kappa'\tau^2 - 2\kappa^2\tau' + \kappa\tau\tau') = 0. \quad (14)$$

For simple solutions for straight line and circle, Eqs. 13-14 give same results as those obtained from Eq. 5. For helical conformations, we can find κ and τ must satisfy following relationship:

$$c_0(\kappa^2 - 2\tau^2) - c_1(\kappa^2\tau + 2\tau^3) - \lambda = 0. \quad (15)$$

As parameters c_0, c_1 and spontaneous curvature λ are only associated with the elastic property of a given molecule, and should be independent of real conformations of helical conformations, we expect there is a linear relationship between $\kappa^2 - 2\tau^2$ and $\kappa^2\tau + 2\tau^3$.

978-1-4577-1666-9/11/$26.00 ©2011 IEEE

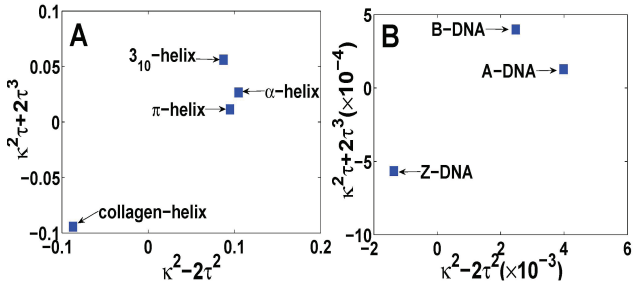

Fig. 4. Relations of $\kappa^2 - 2\tau^2$ and $\kappa^2\tau + 2\tau^3$ for different helical conformations of (A) protein and (B) DNA.

However we can not convict such a linear relationships when substituting parameters given in Table II (see Fig. 4), which hints sole optimization of the elastic energy may not be proper for the determination of natural helical conformations. As we can clearly see that the strong chemical interaction between neighboring contacted monomers, which are completely neglected in Zhang's variational approach, are the major forces that maintain the stability of whole helical chain. Thus, we argue that our current approach, at least in this example, is more realistic than the minimal elastic energy assumption in the description of local regular structures of chain biomolecules.

ACKNOWLEDGMENT

This work is supported in part by the Tsinghua University Initiative Scientific Research Program (20101081751).

REFERENCES

[1] Branden C, Tooze J, *Introduction to protein structure*, Garland Pub, 1998.
[2] Lehninger AL, Nelson DL, Cox MM, *Lehninger Principles of Biochemistry 4th Edition*, W.H. Freeman, 2005.
[3] Goldstein RE, Langer SA, Nonlinear dynamics of stiff polymers, *Phys. Rev. Lett.*, **75**, 1094, 1995.
[4] Bouchiat C, Mezard M, Elasticity model of a supercoiled DNA molecule, *Phys. Rev. Lett.*, **80**, 1556, 1998.
[5] Ou-Yang Zhong-can, Helfrich W, Instability and deformation of a spherical vesicle by pressure, *Phys. Rev. Lett.*, **59**, 2486, 1987.
[6] Ou-Yang Zhong-can, Helfrich W, Bending energy of vesicle membranes: General expressions for the first, second, and third variation of the shape energy and applications to spheres and cylinders, *Phys. Rev. A*, **39**, 5280, 1989.
[7] Zaki M, Bystroff C, *Protein Structure Prediction*, Humana Press, 2010.
[8] Kamien RD, The geometry of soft materials: a primer, *Reviews of Modern physics*, **74**, 953-970, 2002.
[9] Frank FC, On the theory of liquid crystals, *J. Chem. Soc. Faraday Trans., II.* **25**, 19-28, 1958.
[10] W. Flugge. *Tensor analysis and continuum mechanics*, Springer-Verlag, 1972.
[11] M.P. do Carmo, *Differential geometry of curves and surfaces,* Prentice Hall, 1976.
[12] W. Helfrich, Elastic Theory of Helical Fibers, *Langmuir*, **7**, 567-568, 1991.
[13] Q.H. Liu, Y.F. Liu, B. Hu and Z.C. Ou-Yang, Fourth order elastic chiral filament model and the centerline of uniform Kirchhoff elastic rod, *Physics Letters A*, **352**, 358-361, 2006.
[14] Zhang, S., Zou, X., Xia, M., Zhao, S., Zhang, E., General equilibrium shape equations of polymer chains, *Phy. Rev. E*, **70**, 051902 (2004).
[15] Dill KA, Bromberg S, *Molecular driving forces: statistical thermodynamics in chemistry and biology*, Garland Science, 2003.
[16] Marko JF, Siggia ED, Bending and twisting elasticity of DNA, *Macromolecules*, **27**, 981-988, 1994.
[17] Lei JZ, Huang K, Elastic energy of proteins and the stages of protein folding, *Europhys. Lett.*, **88**, 68004, 2009.
[18] Watson JD, Crick FHC, Molecular structure of nucleic acids, *Naute*, **171**, 737C738, 1953.
[19] Pauling L, Corey RB, The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain, *Proc. Natl. Acad. Sci. USA*, **37**, 205-211, 1951.
[20] http://biomedapps.curtin.edu.au/biochem/tutorials/prottute/helices.htm; http://www.uic.edu/classes/phys/phys461/phys450/ANJUM04/.
[21] Finkelstein AV, Ptitsyn OB, *Protein physics: a course of lectures*, Academic Press, 2002.
[22] Sibanda BL, Thornton JM, Beta-hairpin families in globular proteins, *Nature*, 316, 170-174, 1985.
[23] W. Zhao, H.J. Zhou and Z.C. Ou-Yang, Kink instability in circular DNA studied as Helfrich chiral chains, *Physical Review E*, **58**, 8040-8043, 1998.
[24] Marko JF, Siggia ED, Fluctuation and supercoiling of DNA, *Science*, **265**, 506-508, 1994.
[25] Doi M, Edwards SF, *The theory of polymer dynamics*, Clarendon Press, 1988.

978-1-4577-1666-9/11/$26.00 ©2011 IEEE