POLYELECTROLYTE MODELS AND MOLECULAR DYNAMICS STUDIES OF THE DNA POLYMER AND DNA−NCP COMPLEXES IN SALT SOLUTIONS

AHMAD ANGGRARIA JAYA AGUNG

DISSERTATION SUBMITTED IN FULFILLMENT
OF THE REQUIREMENT
FOR THE DEGREE OF MASTERS OF SCIENCE

CHEMISTRY DEPARTMENT
FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2011

Abstract

The effects of salt concentration on the flexibility and charge distribution of a DNA chain are probed using MD simulation. For this purpose, the DNA is modeled as 360 negatively charged monomer spheres with radius 10 Å and charge -12 (all in reduced units) linked linearly by a harmonic bonding potential. Each of these spheres represent 6 base pairs. The harmonic force constant k_h is determined by balancing the intramolecular and intermolecular forces between 2 monomers separated by an equilibrium bond length of 20.4 Å. This method resulted in good agreement with the experimental contour length and therefore guarantees the reasonableness of the k_h parameter. The bending angle force constant k_θ is determined by relating the chain bending modulus with the experimental DNA persistence length. In a cell, the DNA strand wraps around the so called nucleosome core particle (NCP). In this study, the NCP is represented by a large sphere with radius 35 Å and charge +150. The simulations were performed both with and without the NCP, which when present were 12 in number. The NaCl salt is represented as a charged sphere of radius 2 Å of either +1 or -1 charge modeling a single Na^+ or Cl^- ion respectively. The salt concentration was chosen to be in the $0.0-0.25$ mM range for a DNA concentration of 0.005 mg/ml and in the $0.0-100$ mM range for DNA concentration of 2.0 mg/ml. The upper limit corresponded to the maximum computational resources available here. The excluded volume effect was made possible by the application of a purely repulsive 6-12 Lennard-Jones potential and the interaction between charges was strictly Coulombic, where the Ewald summation was computed by the P3M method. The results show that the NaCl salt concentration within the cell play a significant role in determining the DNA conformation, as well as the strength of the DNA-NCP and DNA-ions interaction. Some theoretical equations for calculating the polyelectrolyte persistence length seem to disagree with the simulation results. A new concept that leads to the "ionic bridging effect" is introduced and shown to be important in determining the DNA topology. By the RDF data, it is shown that the classical PB cylindrical cell model neglects the counterion condensation at monomer neighbors. The above results all imply the importance of molecular simulations in obtaining information which cannot so readily be made available from experiments.

Abstrak

Simulasi dinamika molekular telah digunakan untuk menganalisa kesan kepekatan garam ke atas kelenturan and distribusi cas dalam rantai DNA. Untuk tujuan kajian ini, DNA telah dimodelkan sebagai gabungan 360 sfera monomer yang bercas negatif dengan radius 10 Å dan cas -12 dengan ikatan harmonik potential. Setiap sfera terdiri dari 6 pasangan bes. Pemalar tenaga harmonik, k_h, telah ditentukan dengan menyeimbangkan tenaga intramolekular dan intermolecular di antara 2 monomer dengan jarak ikatan keseimbangan 20.4 Å. Kaedah ini memberikan bacaan parameter k_h yang baik apabila dibandingkan dengan nilai eksperimentasi *contour length*. Pemalar tenaga pembengkokan sudut k_θ ditentukan dengan menghubungkan modulus pembengkokan rantai dengan data percubaan *persistence length* rantai DNA. Dalam sel, rantai DNA membungkus *nucleosome core particle* (NCP). Dalam kajian ini, NCP diwakili dengan sebuah sfera besar dengan radius 35 Å dan cas +150. Simulasi dilakukan dengan dan tanpa NCP, yang jika ada berjumlah 12. Garam NaCl diwakili sebagai sebuah sfera bercas dengan radius 2 Å dengan cas salah satu +1 atau -1 yang masing-masing memodelkan sebuah ion Na^+ and Cl^-. Kepekatan garam dipilih dalam jarak $0.0-0.25$ mM untuk kepekatan DNA 0.005 mg/ml and jarak $0.0-100$ mM untuk kepekatan DNA 2.0 mg/ml. Batas atas kepekatan garam sesuai dengan maksimum sumber pengkomputeran yang tersedia disini. Kesan pengecualian ruang dibuat mungkin dengan aplikasi penolakan Lennard-Jones potensi dan interaksi antara cas adalah Coulombic dimana penjumlahan Ewald dihitung dengan kaedah P3M. Keputusan menunjukkan bahawa kepekatan garam NaCl dalam sel bermain peranan penting dalam menentukan penglarasan DNA, maupun kekuatan interaksi DNA$-$NCP dan DNA$-$ion. Beberapa teori untuk menghitung *persistence length* polielektrolit kelihatan tidak bersetuju dengan keputusan simulasi. Sebuah konsep baru yang membawa kepada "kesan jambatan ionis" diperkenalkan dan ditunjukkan menjadi penting dalam menentukan topologi DNA. Dengan data RDF, ditunjukkan bahawa model klasik sel silinder PB mengabaikan kondensasi *counterion* pada monomer-monomer tetangga. Keputusan-keputusan diatas bererti pentingnya simulasi molekular dalam mendapat maklumat yang tidak mudah didapat dari percubaan.

Acknowledgement

I deeply thank God, Alloh SWT, for directing me to come here.

I would like to thank my supervisor, Prof. Dr. Christopher Gunaseelan Jesudason for the opportunity to be his student. His teaching, supervising and training has absolutely developed my knowledge further. I also thank for the many meaningful discussions we have had over the many months.

I thank all ESPResSo contributors and volunteers over the world who directly or indirectly have supported open source and almost always free software that has made this study possible.

I thank my family and friends for their constant support. I save the rest of my gratitude for the beloved one who has always accompanied me over many difficult moments.

Contents

1 INTRODUCTION 1

1.1 DNA 1

1.2 Molecular Dynamics Simulation 3

2 RESEARCH METHODOLOGY 6

2.1 The DNA, NCP and NaCl Salt Models 6

2.2 Research Equipment 7

2.3 Research Method 7

2.4 Particles Properties 8

2.5 System Parameters 9

2.6 Force Fields 10

2.6.1 Intramolecular Force Fields 10

2.6.2 Intermolecular Force Fields 11

2.7 Particle Amounts in the Simulation 12

2.7.1 DNA−Salt System 13

2.7.2 DNA−NCP−Salt System 13

2.8 Determining the Harmonic Bonding Constant k_h and the Equilibrium Harmonic Distance r_h 14

3 DNA FLEXIBILITY AND PERSISTENCE LENGTH 20

3.1 Definition and Calculation of Persistence Length L_P 20

3.1.1 Calculating Persistence Length with Cosine Correlation Function 20

3.2 The Relationship between Persistence Length, Bending Modulus (B) and Angle Force Constant (k_θ) 22

3.2.1 Neutral Polymer Persistence Length 22
3.2.2 Polyelectrolyte Chain Persistence Length 24
3.3 A New Derivation of Polyelectrolyte Persistence Length 26
3.3.1 Details of the Fitting Procedures 34
3.3.2 Obtaining $f_1(n)$ 35
3.3.3 Obtaining $f_2(n)$ 36
3.3.4 Final Result of the Fitting Procedure 39
3.4 Discussion on The New Derived Persistence Length 39
3.5 Representing the Real DNA with Our Model. How Accurate is It? 42
3.6 Simulation Results 44
3.7 Experiment vs. Theory 49
3.8 Simulation vs. Theory 50
3.9 $L_{\mathrm{P}}^{\text{polyelectrolyte}} < L_{\mathrm{P}}^{\text{neutral polymer}}$. Is It Possible? 51
3.10 Ionic Bridging: a New Postulate? 53

4 IONIC DISTRIBUTION IN A DNA SYSTEM 59

4.1 Theoretical Review of Ionic Distribution 59
4.2 Clarifying the Ambiguity of the Screening Parameter κ 61
4.3 Coordination number in various salt concentration 66
4.4 Determining the Electrostatic Free Energy of DNA$-$NCP interactions . . 67
4.4.1 The Poisson-Boltzmann Cylindrical Cell Model (PBCCM) 68
4.4.2 Modification of the Cylindrical to the Spherical Cell 73
4.5 Snapshots 78

5 Conclusion 82

A Reduced units 84

A.1 Length (L) 84
A.2 Temperature 84
A.3 Energy 85
A.4 Mass 85
A.5 Velocity 86

A.6 Time 86

A.7 Harmonic Bond Constant 87

A.8 Bending Angle Constant 88

A.9 Acceleration 88

A.10 Force 89

B Tool Command Language (TCL) Scripts 90

B.1 Equilibrium Run 90

B.2 Production Run 102

List of Figures

1.1 (a) The representation of the DNA double helix [1]. (b) The chemical structure of sugar (deoxiribose), bases (adenine, guanine, thymine and cytosine), and phosphate groups. (c) Dissociation of a DNA-NCP complex yields a DNA strand and eight histone proteins. [2, 3] 2

1.2 Periodic Boundary Condition. 4

2.1 An illustration of a monomer-monomer displacements (grey circle) from its equilibrium position (white circles). 10

2.2 An illustration of a bending angle atomic displacement (grey circle) from its equilibrium bending angle $\theta = \theta_0 = 0$ (white circles). 11

2.3 The energy profile of Lennard-Jones (LJ) potential (Eq. 2.3). The solid line is the original LJ model. The dashed line represents the shifted and truncated LJ potential at $r_{\mathrm{cut}} = 1.1225\sigma$. The vertical dotted line represents the sum of hard core radii of interacting particles. 12

2.4 k_{h}^{*} and $\bar{r}$ (written $\mathrm{r}_{\mathrm{bar}}$ in the graph) relationship obtained from the integration of Eq. 2.17. 17

2.5 Average contour length (a) and bond length (b), both with two standard deviation error bars, from simulations using the harmonic constant 2500 (in reduced unit) and equilibrium bond length set at 20.3418 Å. Simulation results are compared to the theoretical values. 19

3.1 A continuum chain showing some points that determine chain segments. The arrows indicate the orientation at each point segment. 21

3.2 A discrete chain. The angle θ_1, θ_2 and θ_3 are formed by two ends of segments with lengths b, $2b$ and $3b$. 21

3.3 L_{P} obtained by Eq. 3.10 (dashed line) and 3.9 (solid line) at different ratios of (B/k_BT) over s; s is fixed at $7 \times 20.4 = 142.8$ Å 27

3.4 Illustration of a polyelectrolyte segment with length s which bends in a circular form. The r_{ij} distance can be determined by the Cosines law. . . 29

3.5 The L_{P} and L_{P}^0 relationship (in two different Debye length κ^{-1}) obtained from numerical integration of Eq. 3.24. This figure confirms the $L_{\mathrm{P}} - L_{\mathrm{P}}^0$ relationship in Eq. 3.28 ($L_{\mathrm{P}} = L_{\mathrm{P}}^0 + L_{\mathrm{P}}^{\mathrm{el}}$). Each κ value correspond to a definite $L_{\mathrm{P}}^{\mathrm{el}}$. For all κ values, a linear dependency is observed but only two values are depicted. 33

3.6 The $L_{\mathrm{P}}^{\mathrm{el}}$ linear relationship to the Bjerrum length l_{B} (Fig. a) and quadratic relationship to the monomer charge q (Fig. b) obtained from numerical integration of Eq. 3.24 (for different Debye lengths κ^{-1}). These figures confirm the $L_{\mathrm{P}}^{\mathrm{el}} - l_{\mathrm{B}}$ and $L_{\mathrm{P}}^{\mathrm{el}} - q$ relationship in Eq. 3.34 33

3.7 Numerical integration of Eq. 3.35 and its fitting 3.36, with parameters n=20, b= 20.4 Å; κ is varied between 1/5 and 1/2450 Å^{-1}. 35

3.8 $f_1(n)$ (from Eq. 3.35) versus n at different bond length b. These curves will be used to determine the best form of $f_1(n)$. 36

3.9 $f_2(n)$ (from Eq. 3.35) versus n at different bond lengths b. These curves will be used to determine the best form of $f_2(n)$. 37

3.10 The $f_2^*(b)$ and b relationship (dots) and the fitting function (curve) 38

3.11 Persistence length (L_{P}) versus Debye screening length $r_{\mathrm{D}} = 1/\kappa$ from numerical integration (Eq. 3.24), compared to the fitting equation (Eq. 3.41). For these curves, $L_{\mathrm{P}}^0 = 500$ Å, $l_{\mathrm{B}} = 7.13$ Å, $q = -12$, $b = 20.4$ Å, and $l_0 = 1$ Å. 40

3.12 $\langle\theta^2\rangle - n$ relationship based on the Eq. 3.6 ($\langle\theta^2\rangle \approx 2nb/L_{\mathrm{P}}(n)$). The $L_{\mathrm{P}}(n)$ is calculated by Eq. 3.41. The q and b parameters for Eq. 3.41 are given above each figure. $L_{\mathrm{P}}^0 = 500$Å, $l_{\mathrm{B}} = 7.13$Å, and $l_0 = 1$Å for all curves. . . 41

3.13 Some polyelectrolytes modelling a DNA chain. The bond length/charge ratio (1.7 Å/charge) is the same as the experimental DNA data. 42

3.14 Persistence length (L_P) of the neutral polymer possessing the same dimension with the polyelectrolyte DNA models, with 2 standard deviation error bars, at different bending modulus (B) parameters. The straight lines are the theoretical prediction. The simulations are performed with time-step 1×10^{-3} (Fig. (a)) and 1×10^{-5} (Fig. (b)). 45

3.15 Persistence length (L_P) of our DNA polyelectrolyte simulations, with 2 standard deviation error bars, at salt concentration $0.0-0.25$ mM. The simulation results are compared to the L_P prediction from OSF, Dobrynin and Manning theories (a, b, c). Comparison with our derivation is given in (d). The uncharged DNA bending modulus $B_0 = 500$ Å $k_\mathrm{B}T$ throughout. The horizontal dashed line is the persistence length of the neutral uncharged DNA ($L_\mathrm{P}^0 = 500$ Å) [see Eq. 3.28]. 46

3.16 Snapshots of DNA chains in different salt concentrations. Salt concentrations are written at the bottom of the pictures. 48

3.17 Snapshots of DNA conformation at different time intervals (reduced unit). DNA concentration 0.005 mg/ml and salt concentration 0.1 mM. 49

3.18 Experimental data and theoretical calculations of the DNA persistence length at different salt concentrations. The experimental data are taken from Baumann et al. [42] (rectangle); Nordmeier [41] (diamond); Rizzo and Schellman [43] (closed circle) ; Smith et al. [40] (cross). 50

3.19 Theoretical persistence length (L_P) calculated by the OSF, Dobrynin and Manning theory, and our derivation. 52

3.20 Figure (a) is the hypothetical condition. Fig. (b) is the real condition. The vertical dashed line is an imaginary boundary dividing the polymer environment at the equilibrium bending plane. 54

3.21 Snapshots from the DNA chain simulations in salt concentration 0.25 mM (a), and 0.30 mM (b). Area **A** gives the global ionic bridging, and region A_1 and A_2 denote the local ionic bridging. We run the simulation at 0.30 mM (which took one month) only to focus on the ionic bridging without sampling because this would too much time. 55

3.22 A short DNA chain conformation in different time intervals. The number of monomer equals 21. NaCl salt concentration is 20 mM. This simulation was run to examine the availability of ionic bridging effect in different time intervals. The Na^+ and Cl^- particles are modeled by the purple and brown spheres respectively. 57

4.1 PLDNA$-Na^+$ and PLDNA$-Cl^-$ RDF obtained from simulation and PB model calculation. The PB model is calculated for three kinds of ionic screening parameters κ. PLDNA concentration 0.005 mg/ml and salt concentration 0.25 mM 62

4.2 PLDNA$-Na^+$ and PLDNA$-Cl^-$ RDF obtained from simulation and PB model calculation. The PB model is calculated for three kinds of ionic screening parameters κ. PLDNA concentration 2.0 mg/ml and salt concentration 0.25 mM. The only system difference from Fig. 4.1 is the PLDNA concentration 62

4.3 PLDNA$-Na^+$ RDF obtained from simulation, HNC and PYA. The HNC and PYA calculation are done for three kinds of ionic screening parameter κ. 65

4.4 The cylinder model of the charged system. The inner cylinder is an infinite charged rod with charge spacing b. R_c is the cylinder radius defined by the charged rod concentration. V_{slice} is a cylinder volume that is perpendicular to the rod with height b and radius R_c. 67

4.5 DNA$-Na^+$ and PLDNA$-Na^+$ RDF simulation compared to PB calculation. Simulation are done without (a) and with NCP (b) 69

4.6 Electrostatic profile of DNA$-$NCP interaction. The unit of E_{el}, TS_{el} and F_{el} is in k_BT 76

4.7 Coordination number of Na^+ ions about DNA in simulation with NCP and without NCP available in the system. The error estimation for each points is $\pm$ 0.5 76

4.8 Original figure from [62]. DNA-NCP configuration obtained by minimizing the free energy (Eq. 4.30) for a fixed NCP charge $Z = 40$ at increasing salt concentrations. Bars in the lower right are the respective screening lengths. (a) No added salt ($\kappa = 0$). (b) Salt concentration 0.6 mM ($\kappa = 0.08nm^{-1}$) and 1.6 mM ($\kappa = 0.13nm^{-1}$). (e) Salt concentration 16.6 mM ($\kappa = 0.418nm^{-1}$). (f) Salt concentration 10.6 M ($\kappa = 10.6nm^{-1}$) 77

4.9 Snapshots of DNA$-$NCP particles at different salt concentrations. DNA concentration 0.005 mg/ml. 79

4.10 Snapshots of DNA$-$NCP particles at different salt concentrations. DNA concentration 2.0 mg/ml. 80

4.11 DNA-NCP conformations in different time intervals. DNA concentration 0.005 mg/ml, salt concentration 0.075 mM. 81

List of Tables

2.1 The properties of DNA, NCP, Na^+ and Cl^- in simulation 8

2.2 The general parameter of the simulation system 9

2.3 Particle amounts in the DNA-Salt system (without NCP in simulation). . . 13

2.4 Particle amounts in the DNA$-$NCP$-$Salt system. 14

2.5 k_h^*, r_h and $\bar{r}$ relationship obtained from integration of Eq. 2.17. 18

3.1 Numerical result of fitting the function $f_1(n) = An + B$ in Fig. 3.8. The mean of parameter A is 0.00714, with standard deviation 0.000248. The mean of parameter B is 0.0234 with standard deviation 0.000819 36

3.2 Numerical result of fitting the function $f_2(n,b) = f_2^* n$ in Fig. 3.9. The mean of the f_b^* parameter is 0.177, with standard deviation 0.0123. 38

3.3 Persistence length determined by Eq. 3.41 at different salt concentrations for three models in Fig. 3.13 where $L_\mathrm{P}^0 = 500$ Å, $l_\mathrm{B} = 7.13$ Å, and $l_0 = 1$ Å for models (a), (b) and (c). The $\langle\theta^2(n)\rangle - n$ lines are plotted from $n = b$ to $n = s$ (see Sec. 3.4); r^2 is the correlation coefficient for the plots. 43

3.4 End-to-end distance (R_e) and radii of gyration (R_g) of DNA models at different salt concentrations. The δ_{R_e} and δ_{R_g} are the standard deviations of R_e and R_g respectively. 47

4.1 Coordination number of Na^+ ions about DNA. DNA concentration 0.005 mg/ml. First valley distance of DNA$-$Na RDF= 26.6420 Å 66

4.2 Coordination number of Na^+ ions about DNA. DNA concentration 2.0 mg/ml. First valley distance of DNA$-$Na RDF= 26.4739 Å 66

4.3 Electrostatic profile (Spherical Cell) of DNA$-$NCP interaction and DNA-$-$Na$^+$ coordination number. DNA concentration 0.005 mg/ml 74

4.4 Electrostatic profile (Spherical Cell) of DNA$-$NCP interaction and DNA-$-$Na$^+$ coordination number. DNA concentration 2.0 mg/ml 75

List of Symbols and Abbreviations

DNA = Deoxyribonucleic acid
DHA = Debye-Huckel Approximation
FFA = Far-Field Approximation
NCP = Nucleosome Core Particle
OSF = Odijk-Skolnick-Fixman
PBC = Periodic Boundary Condition
PB = Poisson-Boltzmann
PBCCM = Poisson-Boltzmann Cylindrical Cell Model
PLDNA = Particle-like DNA monomer
RDF = radial distribution function
WLC = worm-like chain
a = acceleration
b = monomer bond length
B = bending modulus
c_i = concentration of ion i
c_{shift} = constant defining $\mathrm{U}_{\mathrm{LJ}}(r_{\mathrm{cut}}) = 0$
δx = standard deviation of the variable x
$\varepsilon_{\mathrm{LJ}}$ = Lennard-Jones energy unit
e_0, e = elementary proton charge
ε = dielectric constant of the medium
E_{el} = electrostatic energy
f' = random force
$f(r)$ = conservative force
$g(r)$ = radial distribution function
κ = ionic screening parameter (inverse of Debye length)
k_{B} = Boltzmann constant
k_h = harmonic bond constant
k_θ = bending constant
l_0 = additional parameter for our persistence length theory ($l_0 = 1\,\text{Å}$)
l_{B} = Bjerrum length
L = chain length
L_{P} = persistence length
L_{OSF} = electrostatic persistence length from OSF theory
L_{P}^0 = neutral/uncharged polymer persistence length

$L_{\mathrm{P}}^{\mathrm{el}}$ = electrostatic persistence length

m = mass

n = total number of bonds

n_{M}^{α} = coordination number of α type ions about M macroion

N = total number of monomers

N_{A} = Avogadro number

ϕ_{M} = reduced surface potential

ψ_{M} = surface potential

$\psi(r)$ = potential at position r from a macroion surface

q_i = charge of particle i

r = distance

$\overline{r}$ = average equilibrium bond distance from Boltzmann integration

r_{off} = sum of the hard core radii

$\mathbf{r}_{\mathrm{cg}}$ = center of gravity

r_{D} = Debye (screening) length

r_{h} = harmonic equilibrium bond distance

R_{c} = cylinder radius

R_{e} = end-to-end distance

R_{g} = radii of gyration

R_{n} = radius of the normalized RDF

ρ = density, charge density

s = segment length

S_{el} = electrostatic entropy

σ = sum of the soft core radii

σ_{M} = surface charge density

θ = bending angle

θ_0 = equilibrium bending angle

T = Temperature

$U_{\mathrm{C\text{-}P3M}}(r)$ = Coulombic potential energy calculated by P3M method

$U_{\mathrm{h}}(r)$ = harmonic stretching potential energy

$U_{\mathrm{LJ}}(r)$ = Lennard-Jones potential energy

U_{θ} = bending angle potential energy

ξ = frictional constant, line charge density

z_i, Z_i = valence