

# Contents

<b>Acknowledgements</b>	v
<b>Abstract (English/Français/Italiano)</b>	vii
<b>Introduction</b>	xiii
<b>List of figures</b>	xxii
<b>List of tables</b>	xxxii
<b>I Background</b>	<b>1</b>
<b>1 Introduction to DNA</b>	<b>3</b>
1.1 Torsional angles . . . . .	4
1.2 Sugar ring puckering . . . . .	5
1.3 Rigid-body configuration of ideal nucleic acid bases . . . . .	6
<b>2 Mathematics behind coarse-grain DNA models</b>	<b>9</b>
2.1 The groups $SO(3)$ and $SE(3)$ . . . . .	11
2.2 Polymer physics and persistence length . . . . .	14
2.3 From atoms to rigid-bodies . . . . .	16
2.4 Bichain interpretation of DNA . . . . .	19
2.4.1 Internal coordinates for $(\mathbf{g}, \mathcal{P})$ . . . . .	20
2.4.2 Internal Energy for $(\mathbf{g}, \mathcal{P})$ . . . . .	21
<b>3 On molecular dynamics simulation protocols and analysis</b>	<b>25</b>
3.1 Potential and force field . . . . .	25
3.2 The ABC collaboration and simulation protocol . . . . .	26
3.3 Analysis of trajectories . . . . .	27
3.3.1 Hydrogen bond filtering . . . . .	29
3.4 Estimation of banded stiffness matrix . . . . .	30

<b>4</b>	<b>cgDNA: a sequence-dependent rigid-base model for DNA</b>	<b>33</b>
4.1	Main assumptions underlying the cgDNA model . . . . .	33
4.2	The cgDNA parameter set . . . . .	36
4.3	Study of persistence length using the cgDNA model . . . . .	37
<b>5</b>	<b>Estimation of cgDNA parameter sets</b>	<b>41</b>
5.1	Kullback-Leibler divergence . . . . .	41
5.2	Estimation of parameters . . . . .	43
5.3	Positiveness of the best-fit parameter set . . . . .	45
<b>II</b>	<b>Comparison of cgDNA parameter sets</b>	<b>47</b>
<b>6</b>	<b>Sensitivity of cgDNA parameter sets to training data</b>	<b>49</b>
6.1	Introduction to training data comparison . . . . .	51
6.2	Sensitivity to simulation duration: ABC versus $\mu$ ABC . . . . .	53
6.3	Sensitivity to training library: $\mu$ ABC versus $MABC_0^K$ . . . . .	54
6.4	Sensitivity to force field: $MABC_0^K$ versus $MABC_1^K$ . . . . .	56
6.5	Sensitivity to ions: $MABC_1^K$ versus $MABC_1^{NaK}$ versus $MABC_1^{Na}$ . . . . .	59
6.6	Discussion and Conclusions . . . . .	59
<b>7</b>	<b>A Palindromic training library</b>	<b>63</b>
7.1	Designing palindromic libraries . . . . .	63
7.2	The palindromic training sets . . . . .	67
7.2.1	Assessing convergence of MD simulations . . . . .	67
7.2.2	Estimation of 1st and 2nd moments using palindromic symmetry . . . . .	72
7.3	Palindromic cgDNA parameter set . . . . .	74
7.3.1	Positiveness of the new format . . . . .	77
<b>III</b>	<b>cgDNA+</b>	<b>81</b>
<b>8</b>	<b>Coarse-grain configuration variables for double stranded DNA</b>	<b>83</b>
8.1	Double stranded DNA configurations with explicit backbone treatment . . . . .	83
8.1.1	Microstructure with explicit treatment of the phosphate group . . . . .	87
8.2	Internal energy for tetrachains . . . . .	88
8.2.1	Equilibrium configurations and variational principle . . . . .	89
8.3	Total force and torque acting on a single phosphate group . . . . .	90
8.4	Internal coordinates for tetrachain configurations . . . . .	91
8.4.1	Change of reading strand transformation . . . . .	94
<b>9</b>	<b>A sequence-dependent coarse-grain model of B-DNA with explicit treatment of the phosphate groups</b>	<b>95</b>
9.1	On the base-to-phosphate degrees of freedom . . . . .	96

9.1.1	Two possible definitions of internal coordinates . . . . .	96
9.1.2	On the cgDNA+ sparsity pattern . . . . .	102
9.2	Convergence of the phosphate degrees of freedom . . . . .	104
9.2.1	Convergence of oligomer-based Gaussian . . . . .	108
9.3	The cgDNA+ parameter set . . . . .	109
9.3.1	Fisher information matrix . . . . .	113
9.3.2	Computation of an admissible initial cgDNA+ parameter set . . . . .	115
9.3.3	Fisher informed gradient . . . . .	119
9.4	Proving positiveness of the best-fit parameter set . . . . .	121
9.5	Assessing the best-fit cgDNA+ parameter set . . . . .	124
9.5.1	Ground-state . . . . .	124
9.5.2	Tangent-tangent correlation . . . . .	125
9.6	Beyond nearest neighbour interactions . . . . .	130
<b>IV</b>	<b>Applications of cgDNA+</b>	<b>137</b>
	<b>Preliminary remarks</b>	<b>139</b>
<b>10</b>	<b>Study of sequence-dependent persistence lengths using cgDNA+</b>	<b>141</b>
<b>11</b>	<b>Crystal structure packing forces</b>	<b>149</b>
<b>12</b>	<b>Fine-graining of cgDNA+ ground-state backbone configurations</b>	<b>159</b>
12.1	Computation of sugar configurations and sugar puckering modes . . . . .	159
12.2	A sequence context study of BI-BII backbone conformations . . . . .	164
<b>13</b>	<b>Groove widths prediction</b>	<b>167</b>
	<b>Future development of the applications</b>	<b>177</b>
	<b>Conclusions and Outlook</b>	<b>179</b>
<b>V</b>	<b>Appendices</b>	<b>185</b>
<b>A</b>	<b>Ideal atom definitions</b>	<b>187</b>
<b>B</b>	<b>Training set libraries</b>	<b>191</b>
<b>C</b>	<b>The ABC molecular dynamic protocol</b>	<b>195</b>
<b>D</b>	<b>cgDNA+: complement to the simulation convergence discussion</b>	<b>197</b>

<b>E</b>	<b>Derivatives of KLd, Gradient, and Hessian matrix</b>	<b>201</b>
E.1	First derivative of the Kullback–Leibler divergence . . . . .	202
E.2	Gradient of the Kullback–Leibler divergence . . . . .	203
E.3	Second derivative of the Kullback–Leibler divergence . . . . .	203
E.4	Hessian matrix of the Kullback–Leibler divergence . . . . .	204
E.5	Derivation of KLd with respect to the parameter set . . . . .	205
<b>F</b>	<b>External load acting on a single base</b>	<b>209</b>
E1	The reading strand case . . . . .	209
E1.1	SE(3) perturbation of the matrix $\mathcal{P}^H$ . . . . .	210
E1.2	Closed form expression of force and torque on a base $\mathbf{g}^+$ . . . . .	211
E2	The complementary strand case . . . . .	212
E2.1	Closed form expression of force and torque on a base $\mathbf{g}^-$ . . . . .	212
	<b>Index</b>	<b>213</b>
	<b>Bibliography</b>	<b>216</b>
	<b>Curriculum Vitae</b>	<b>225</b>