Biophysical Chemistry 133 (2008) 66 – 70

Biophysical Chemistry

http://www.elsevier.com/locate/biophyschem

Predicting ultraviolet spectrum of single stranded and double stranded deoxyribonucleic acids

Andrey V. Tataurov, Yong You, Richard Owczarzy *

Integrated DNA Technologies, 1710 Commercial Park, Coralville, IA-52241, USA

Received 10 December 2007; accepted 15 December 2007

Abstract

Synthetic oligodeoxynucleotides are widely used in many biological, biochemical and biophysical applications. The concentration, composition and structure of DNA are often determined from its ultraviolet spectrum. Although parameters for use with the nearest-neighbor model for prediction of extinction coefficients of single stranded DNAs at 260 nm were published some time ago, similar parameters for other wavelengths or for use with DNA duplexes have not been reported. Practical formulae and parameters for prediction of UV spectra, hypochromism and peak wavelengths were experimentally determined for both single stranded and double stranded oligodeoxynucleotides in the range from 215 to 310 nm. The accuracy of predictions made using the nearest-neighbor model and the base composition model was determined and compared. The spectrum of any DNA oligomer can be calculated using a Microsoft Excel® application that is available in the Appendix A.

Keywords: Extinction coefficient; Oligodeoxynucleotides; Hypochromism; Nearest-neighbor model; DNA; Ultraviolet spectrum

1. Introduction

Near ultraviolet spectra of DNAs and RNAs are often collected to determine concentrations [1–4], base composition [5,6] and secondary structure of nucleic acids [7,8]. Published models and parameters allow calculations of the oligonucleotide extinction coefficient, ε_{260}, at 260 nm [1,3]; a complete set of nearest-neighbor (n-n) parameters for other wavelengths has not been reported.

The extinction coefficient of double stranded DNA (dsDNA) is less than the sum of the extinction coefficients for the composing two single stranded DNAs (ssDNA) [9]. This hypochromic effect of nucleotides is attributed to dipole-induced dipole interactions between transition moments in stacked neighboring bases [10]. Empirical parameters for accurate predictions of oligonucleotide hypochromism have not been published [3] although this value is crucial for determination of concentration of double stranded DNA oligomers.

In this paper, the UV spectra of a large number (~200) of synthetic DNAs were analyzed. We report methods and empirical parameters for prediction of: (1) duplex hypochromism at 260 nm; (2) the wavelength of maximum absorbance; and (3) extinction coefficients for any single stranded or double stranded DNA oligomers in the near UV region with 1 nm resolution.

2. Experimental and theoretical methods

2.1. Base composition model

Several models have been proposed that relate an intensive physical property, *P*, of an oligonucleotide to its structure and base sequence [11,12]. The base composition model (zeroth-neighbor model) takes into account the number or frequency of occurrences of bases, but neglects interactions between neighbor bases or base pairs,

$$P = \sum_{i=\mathrm{A,C,G,T}} f_i P_i \tag{1}$$

---

* Corresponding author. Tel.: +1 319 626 8459; fax: +1 319 626 9621.
E-mail address: science@owczarzy.net (R. Owczarzy).
URL's: http://biophysics.idtdna.com, http://www.owczarzy.net (R. Owczarzy).

0301-4622/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.bpc.2007.12.004

This sum is computed over all kinds of bases present in single strands or over all types of base pairs in double stranded nucleic acids. Properties of each base or base pair, P_i, are determined from fit to experimental data. The fractions of bases, f_i, are calculated from the oligonucleotide sequence,

$$f_i = \frac{N_i}{\sum\limits_{i=\mathrm{A,C,G,T}} N_i} = \frac{N_i}{N_b} \tag{2}$$

N_i is the number of i type base in the oligonucleotide sequence. If the physical property P is extensive (e.g., hybridization free energy, extinction coefficient) the frequencies f_i are replaced with the number of bases in Eq. (1) [11].

2.2. Nearest-neighbor model

The nearest-neighbor (first-neighbor) model broadens the base composition model and takes into account interactions between adjacent bases or base pairs. However, any interactions over a longer range (e.g., between next nearest-neighbor bases) are neglected. The intensive physical property P is obtained by summing properties of all types of nearest-neighbor subunits (doublets) present in the nucleic acid (e.g., AA, AC, AG),

$$P = \sum_{i,j=\mathrm{A,C,G,T}} f_{ij} P_{ij} \tag{3}$$

When DNA is composed from four native bases, there are 20 and 12 unique P_{ij} parameters for single stranded and double stranded oligomers, respectively [11]. The nearest-neighbor parameters include initiation (end) interactions, which are denoted using a fictitious base “E” for the end of an oligonucleotide [12].

We employed both the nearest-neighbor and base composition models and evaluated their ability to predict accurately various features of UV spectra for single stranded and double stranded oligodeoxynucleotides. The base composition model was preferred because of the lower number of parameters. However, if these parameters did not provide accurate predictions, the nearest-neighbor parameters are reported.

2.3. Determination of parameters

Once the property P is measured for a set of oligonucleotides, Eqs. (1) and (3) are constructed for each sequence. To decrease the impact of experimental errors on parameters there are usually more sequences than unknown parameters, i.e., the resulting system of linear equations is overdetermined,

$$\mathbf{F} \cdot \overrightarrow{\boldsymbol{P}}_{ij} = \overrightarrow{\boldsymbol{P}}_{\mathrm{exp}} \tag{4}$$

$\overrightarrow{\boldsymbol{P}}_{ij}$ is the column vector of parameters and $\overrightarrow{\boldsymbol{P}}_{\mathrm{exp}}$ is the column vector of the experimentally measured properties, one for each sequence. The design matrix **F** has a row for each oligonucleotide and a column for each parameter (base or nearest-neighbor unit). The overdetermined Eq. (4) was solved using the singular value decomposition method (SVD), so that the squares of residuals were minimized [13]. Least-squared fitted parameters were obtained. In our analysis, the design matrices had no singular values indicating that all parameters P_{ij} were unique and no parameter could be obtained from a linear combination of the remaining parameters [13]. The SVD was implemented using Excel Add-in, Matrix.xla package, version 2.3 (Foxes Team, Leonardo Volpi, http://digilander.libero.it/foxes).

The accuracy of predictions was statistically compared using reduced χ^2_r and the root mean square deviation (rmsd) between predicted and experimentally measured values [14].

2.4. Experimental methods

All spectra were recorded at 25 °C in 1.000±0.001 cm pathlength quartz cuvettes using a single-beam, diode-array Hewlett-Packard HP8453 spectrophotometer. Absorbance and wavelength calibration was regularly verified with reference dichromate and holmium oxide solutions (Starna Cells, Atascadereo, CA). Temperature was controlled by a Peltier cuvette holder. Absorbance data at 1 nm intervals were processed using UV–Visible ChemStation and Excel 2003. This work employed the sets of DNA oligomers from a previous study that explored the effects of sodium cations on T_m [15]. The experimental data set of single stranded oligodeoxynucleotides consisted of 202 sequences ranging in length from 6 to 30 bases and in GC content from 0 to 80%. The spectra of these single strands were measured in a low salt buffer (0.1 mM sodium phosphate, 0.5 mM NaCl, 0.01 mM $\mathrm{Na_2EDTA}$, pH=7.0) while strand concentrations ranged from 0.5 to 11 μM.

The data set of DNA duplex oligomers contained 86 unique sequences ranging in length from 10 to 60 base pairs and in GC content from 20 to 80%. The UV spectra of duplexes were collected in a high salt buffer (100 mM NaCl, 10 mM sodium phosphate, 1 mM $\mathrm{Na_2EDTA}$, pH=7.0, total $[\mathrm{Na^+}]=119$ mM) where all duplexes were stable. The helix–coil transitions and stabilities of all duplexes were previously determined using UV melting experiments [15].

3. Results and discussion

3.1. Hypochromism of DNA

When two single stranded oligodeoxynucleotides, S1 and S2, hybridize and form a DNA duplex, D, absorbance decreases [9]. The extinction coefficient of the duplex, ε_D, can be estimated from the extinction coefficients of both strands, ε_S1, ε_S2, and hypochromicity, h,

$$\varepsilon_\mathrm{D} = (1-h)\cdot(\varepsilon_\mathrm{S1}+\varepsilon_\mathrm{S2}) \tag{5}$$

The value of $1-h$ is the fraction of absorbance that remains when the single strands anneal to form the duplex,

$$1-h = \frac{A_\mathrm{D}}{(A_\mathrm{S1}V_\mathrm{S1}+A_\mathrm{S2}V_\mathrm{S2})/(V_\mathrm{S1}+V_\mathrm{S2})} \tag{6}$$

where A_D is the absorbance of the duplex solution. A_S1, A_S2 are absorbance values of the single strand S1 and S2 solutions, respectively. Volumes V_S1 and V_S2 of S1 and S2 strand solutions were mixed to make the duplex sample. Absorbance of 86 duplexes and composing single strand solutions were measured at 260 nm, neutral pH and at constant temperature of 25 °C. Hypochromicity was calculated from Eq. (6). Since hypochromism originates from interactions between neighbor bases, it is not surprising that hypochromism is sequence dependent. Fig. 1 shows that hypochromism decreases with increasing GC content. A similar dependence was reported earlier [16], however, their hypochromism was calculated from the difference between absorbance values at low and high temperatures [9]. We determine hypochromism at a constant temperature. The values of h at 260 nm were fit to base composition model, where hypochromicity is proportional to the fraction of GC and AT base pairs,

$$h(260\text{nm}) = f_{\text{AT}} \times 0.287 + f_{\text{GC}} \times 0.059 \tag{7}$$

The 29% hypochromicity for $\text{d}A-\text{d}T$ base pair is similar to 22–24% hypochromicity values observed for rA_nrU_n duplexes (n= 4−7) [17].

Since experimental errors of hypochromicity are about 0.02, the nearest-neighbor model did not significantly improve accuracy of hypochromicity predictions. The rmsd of 0.014 for the n-n model is comparable with rmsd of 0.02 obtained from Eq. (7). This result is consistent with measurements of genomic DNA samples where hypochromicity has been approximated reasonably well by a linear function of GC content [9].

3.2. Prediction of peak wavelength

The peak wavelength of single stranded DNA oligomers, λ_{peak}, where absorbance reaches local maximum is sequence dependent. Both the nearest-neighbor model [12] and the base composition model predicted the peak wavelength with similar accuracy. The peak wavelength can therefore be computed using the base composition model,

$$\lambda_{\text{peak}} = f_{\text{A}} \cdot \lambda_{\text{A}} + f_{\text{C}} \cdot \lambda_{\text{C}} + f_{\text{G}} \cdot \lambda_{\text{G}} + f_{\text{T}} \cdot \lambda_{\text{T}} \tag{8}$$

where f_{A}, f_{C}, f_{G} and f_{T} are the fractional contents of adenine, cytosine, guanine and thymine bases, respectively. Eq. (8) and the four parameters shown in Table 1 apply to any native single stranded DNA sequence. The rmsd value of 0.7 nm was observed for the set of 202 sequences, which is similar to the resolution of the spectrophotometer (1 nm). The peak wavelengths in Table 1 are from 2 to 6 nm smaller than the values observed for deoxynucleotide monophosphates [18]. A similar hypsochromic shift was reported for AAA trinucleoside diphosphate [19].

Table 1
Parameters used in Eq. (8)

Parameter	λ_{A}	λ_{C}	λ_{G}	λ_{T}
Peak wavelength (nm)	256.7	265.7	247.3	265.1

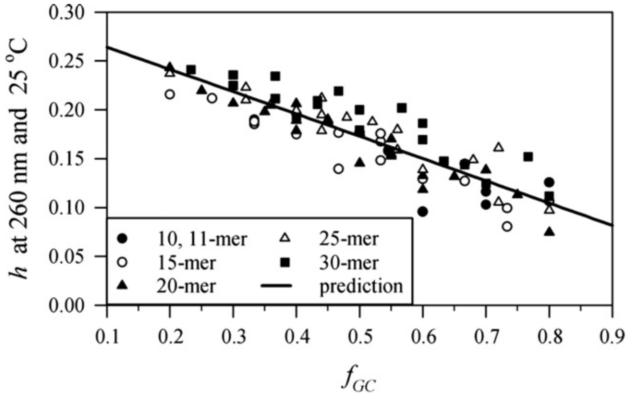

Fig. 1. Hypochromicity values decrease with increasing GC base pair content.

3.3. Prediction of ultraviolet spectrum

It has been well established that both absorbance and extinction coefficients, ε, of DNA oligomers vary with wavelength [1]. These properties are predicted accurately from the nearest-neighbor model. The n-n ε parameters have been previously published for single stranded oligonucleotides at 260 nm [1,3]. Using these parameters, the extinction coefficient is traditionally calculated from the following formula [1–3],

$$\varepsilon_{260} = \sum_{i=1}^{N_b-1} \varepsilon_{i,i+1} - \sum_{i=2}^{N_b-1} \varepsilon_i \tag{9}$$

The $\varepsilon_{i,i+1}$ are the extinction coefficients for doublets of nucleotides, the ε_i is the extinction coefficient for the nucleotide i, and N_b is the number of nucleotides in the oligomer sequence. The second term of Eq. (9) subtracts extinction coefficients of internal bases, which are counted twice in the first term. The parameters are listed in the first two columns of Table 2. The format of Eq. (9) differs from the usual format of the nearest-neighbor model, Eq. (3) [11],

$$\varepsilon_{260} = \sum_{i,j=\text{A,C,G,T,E}} N_{ij}\varepsilon_{ij}^{\text{n-n}} \tag{10}$$

The initiation (end) interactions (EA, AE, EG, etc.) present at each oligonucleotide terminus are included in Eq. (10). The parameters of Eq. (9) can be converted to parameters of the simpler Eq. (10). The converted parameters are listed in the last two columns of Table 2. Eqs. (9) and (10) are mathematically equivalent and provide the same extinction coefficient for any oligodeoxynucleotide sequence. However, Eq. (10) better demonstrates that the total number of nearest-neighbor units (doublets) is N_b+1 for an oligonucleotide that is N_b bases long,

$$\sum_{i,j=\text{A,C,G,T,E}} N_{ij} = N_b + 1 \tag{11}$$

The extinction coefficient, ε_λ, at any wavelength λ can be calculated from ε_{260} and the ratio of absorbance at wavelength λ to the absorbance at 260 nm,

$$R_\lambda^{260} = \frac{A_\lambda}{A_{260}} \tag{12}$$

$$\varepsilon_\lambda = R_\lambda^{260} \cdot \varepsilon_{260} \tag{13}$$

This method allows easy updating of ε_λ values whenever more accurate parameters for ε_{260} are introduced in the literature [3].

We have measured experimentally the ratios R_λ^{260} for both single stranded and double stranded DNA oligomers. The nearest-neighbor model was found to provide a substantially better fit of these ratios (rmsd of $R_\lambda^{260}=0.007$ for dsDNA) than the base composition model (rmsd of $R_\lambda^{260}=0.011$) for both ssDNAs and dsDNAs. The ratio R_λ^{260} is an intensive physical variable and depends therefore on the fraction of specific nearest-neighbor doublets,

$$R_\lambda^{260} = \sum_{i,j=\mathrm{A,C,G,T,E}} f_{ij} R_{ij}^{\mathrm{n-n}} \tag{14}$$

where the fraction of a particular nearest-neighbor doublet is defined as,

$$f_{ij} = \frac{N_{ij}}{\sum\limits_{i,j=\mathrm{A,C,G,T,E}} N_{ij}} = \frac{N_{ij}}{N_b + 1} \tag{15}$$

The $R_{ij}^{\mathrm{n\text{-}n}}$ are sequence dependent nearest-neighbor parameters for doublets of nucleotides, e.g. AA, TG, EA. Twenty and twelve unique parameters were determined for ssDNA and dsDNA, respectively, at each wavelength. These $R_{ij}^{\mathrm{n\text{-}n}}$ values are reported in the Appendix A. Some bases (e.g., *d*A) do not absorb above 295 nm. Under these conditions their $R_{ij}^{\mathrm{n\text{-}n}}$ parameters and Eq. (14) may predict small negative R_λ^{260} values due to experimental errors. Since negative R_λ^{260} are not observed experimentally, these negative values should be set to zero. Fig. 2 demonstrates the accuracy of prediction for two oligodeoxynucleotides of very different base composition. Excellent agreement between experimentally measured and predicted spectra is seen. When R_λ^{260} values were predicted for our dataset using the n-n parameters, overall root mean square deviations of 0.013 and 0.007 were observed for single stranded and double stranded DNAs, respectively.

Table 2
Two equivalent sets of parameters for the extinction coefficient of oligodeoxynucleotides at 260 nm

Traditional format (Eq. (9))		New format (Eq. (10))	
i,j	$\varepsilon_{i,j}$ (260 nm)	ij	$\varepsilon_{ij}^{\mathrm{n\text{-}n}}$ (260 nm)
(5′ to 3′)	(10^3 M^{-1} cm^{-1})	(5′ to 3′)	(10^3 M^{-1} cm^{-1})
A	15.4	EA, AE	7.70
C	7.4	EC, CE	3.70
G	11.5	EG, GE	5.75
T	8.7	ET, TE	4.35
AA	27.4	AA	12.00
AC	21.2	AC	9.80
AG	25.0	AG	11.55
AT	22.8	AT	10.75
CA	21.2	CA	9.80
CC	14.6	CC	7.20
CG	18.0	CG	8.55
CT	15.2	CT	7.15
GA	25.2	GA	11.75
GC	17.6	GC	8.15
GG	21.6	GG	10.10
GT	20.0	GT	9.90
TA	23.4	TA	11.35
TC	16.2	TC	8.15
TG	19.0	TG	8.90
TT	16.8	TT	8.10

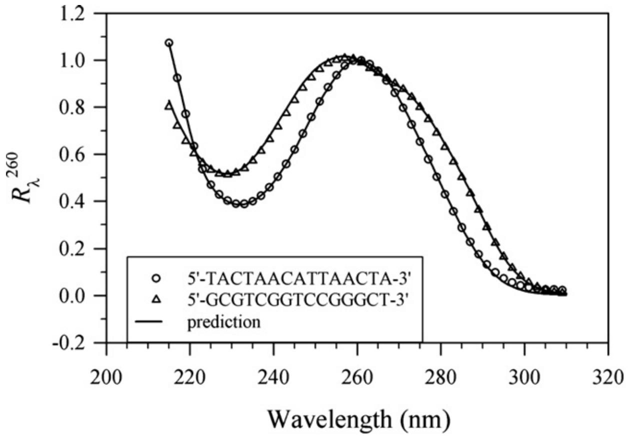

Fig. 2. Comparison of predicted (solid lines) and experimentally measured (symbols) spectra for two single stranded DNA oligomers. The GC base pair content of sequences is 20% (open circle) and 80% (open triangle).

3.4. Software overview

The Excel spreadsheet provided in the Appendix A can be used to calculate the UV spectrum of any oligodeoxynucleotide. Visual Basic for Applications macros must be enabled in order to use these functions (set security to medium level in the Excel submenu Tools\Macro\Security and restart the Excel program). Sequences should be entered in column B. The top part of the sheet calculates the peak wavelength and the extinction coefficients at 260 nm. The bottom part of the sheet predicts the entire UV spectrum for a single sequence entered in the cell B27. Sheets “SSData” and “DuplexData” contain tables of nearest-neighbor parameters for single stranded and double stranded DNA oligomers, respectively. Available custom functions are described in the sheet “Help”.

4. Conclusion

This paper present methods, experimentally determined parameters and easy-to-use software that calculates the ultraviolet spectrum of deoxyribonucleic acids. Spectra of both single stranded and double stranded DNA oligomers were shown to be accurately predicted.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bpc.2007.12.004.

References

[1] C.R. Cantor, M.M. Warshaw, H. Shapiro, Oligonucleotide interactions. III. Circular dichroism studies of the conformation of deoxyoligonucleotides, Biopolymers 9 (1970) 1059–1077.

[2] G. Kallansrud, B. Ward, A comparison of measured and calculated single- and double-stranded oligodeoxynucleotide extinction coefficients, Anal. Biochem. 236 (1996) 134–138.

[3] M.J. Cavaluzzi, P.N. Borer, Revised UV extinction coefficients for nucleoside-5′-monophosphates and unpaired DNA and RNA, Nucleic Acids Res. 32 (2004) e13.

[4] G.D. Fasman, Handbook of Biochemistry and Molecular Biology, vol. I, CRC Press, Cleveland, 1975, p. 589.

[5] J.T. Henderson, A.S. Benight, S. Hanlon, A semi-micromethod for the determination of the extinction coefficients of duplex and single-stranded DNA, Anal. Biochem. 201 (1992) 17–29.

[6] R.D. Blake, T.G. Hydorn, Spectral analysis for base composition of DNA undergoing melting, J. Biochem. Biophys. Methods 11 (1985) 307–316.

[7] J.-L. Mergny, J. Li, L. Lacroix, S. Amrane, J.B. Chaires, Thermal difference spectra: a specific signature for nucleic acid structures, Nucleic Acids Res. 33 (2005) e138.

[8] J. Jaumot, N. Escaja, R. Gargallo, C. Gonzales, E. Pedroso, R. Tauler, Multivariate curve resolution: a powerful tool for the analysis of conformational transitions in nucleic acids, Nucleic Acids Res. 30 (2002) e92.

[9] G. Felsenfeld, S.Z. Hirschman, A neighbor-interaction analysis of the hypochromism and spectra of DNA, J. Mol. Biol. 13 (1965) 407–427.

[10] I. Tinoco Jr., Hypochromism in polynucleotides, J. Am. Chem. Soc. 82 (1960) 4785–4790.

[11] D.M. Gray, Derivation of nearest-neighbor properties from data on nucleic acid oligomers. I. Simple sets of independent sequences and the influence of absent nearest neighbors, Biopolymers 42 (1997) 783–793.

[12] R.F. Goldstein, A.S. Benight, How many numbers are required to specify sequence-dependent properties of polynucleotides? Biopolymers 32 (1992) 1679–1693.

[13] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, Numerical Recipes in Fortran 77: The Art of Scientific Computing, Cambridge University Press, Cambridge, UK, 1992.

[14] P.R. Bevington, D.K. Robinson, Data Reduction and Error Analysis for the Physical Sciences, WCB/McGraw-Hill, Boston, 1992.

[15] R. Owczarzy, Y. You, B.G. Moreira, J.A. Manthey, L. Huang, M.A. Behlke, J.A. Walder, Effects of sodium ions on DNA duplex oligomers: improved predictions of melting temperatures, Biochemistry 43 (2004) 3537–3554.

[16] V.I. Danilov, On the origin of the hypochromic effect in double-stranded polynucleotides, FEBS Lett. 47 (1974) 155–157.

[17] K.J. Breslauer, J.M. Sturtevant, I. Tinoco Jr., Calorimetric and spectroscopic investigation of the helix-to-coil transition of a ribo-oligonucleotide: rA_7U_7, J. Mol. Biol. 99 (1975) 549–565.

[18] C.A. Sprecher, W.C. Johnson Jr., Circular dichroism of the nucleic acid monomers, Biopolymers 16 (1977) 2243–2264.

[19] C.R. Cantor, I. Tinoco Jr., Absorption and optical rotary dispersion of seven trinucleoside diphosphates, J. Mol. Biol. 13 (1965) 65–77.