XVI IMEKO World Congress
Measurement - Supports Science - Improves Technology - Protects Environment ... and Provides Employment - Now and in the Future
Vienna, AUSTRIA, 2000, September 25-28

EVALUATION OF OPTICAL SIGNALS FOR DNA CHAIN ANALYSIS

A. Cysewska-Sobusiak and M. Boltrukiewicz
Institute of Electronics and Telecommunications
Poznan University of Technology, Piotrowo 3A, 60-965 Poznan, Poland

Abstract: The molecular complementarity of the two strands of the DNA double helix makes it possible to model changes in DNA sequences manifesting by related changes in their melting temperature, thermodynamic data and optical properties. The subject of the paper is focused on metrological problems connected with modeling of a virtual integrated oligonucleotide library, basing on the analysis of fluorescence signals to be obtained as a result of hybridization.

Keywords: oligonucleotide biooptics, modeling and simulation.

1 INTRODUCTION

The scale of the human genome is too large to be effectively analyzed using the traditional tools of molecular biology and genetics. Size of DNA content expressed in monomer units (i.e. basepairs) is equal to approximately 3 billion bp. for extremely complex human genome [1]. Modern tools for rapid and accurate genome analysis, which include DNA sequencing, physical as well as optical mapping, special software for acquiring and assembling sequence information, are still under development [2]. In order to improve the speed of DNA sequencing it is necessary to perform new methods such as e.g. sequencing by hybridization (SBH). Many of the techniques which are employed to analyze DNA sequences base on the molecular complementarity of the two strands of the double helix. Optical consequences of thermodynamic properties of gene molecules allow to use fluorescence techniques, easy to automated analysis. The questions considered in this paper are connected with modeling and evaluation of signals to be obtained with fluorescence detection. The metrological characteristics such as resolution and levels of signal intensity of a virtual library made by the testers for detection of complementary segments in the given strand of a denatured DNA chain to be studied are considered.

2 SPECIFICITY OF INTERACTIONS BETWEEN PAIRS OF DNA BASES

A gene DNA sequence is the linear order of nucleotides (called bases) derived from a four-letter biochemical alphabet, i.e. A, C, G and T. The molecular complementarity arising from the specific hydrogen-bonding interaction between pairs of DNA bases, A only with T and C only with G, and resulting in the complementarity of two strands of the double helix, is one great advantage of DNA from its analysis point of view (Figure 1). The double helix can be disrupted by heating or by exposure to low salt concentration – this process is called denaturation or melting. When one of the strands has been supplied from an exogenous source, two complementary single strands also can be renatured into the another duplex by hybridization process. Base pairing within the molecule can fix the location of one region relative to another. In a perfect duplex of DNA, the strands are precisely complementary.

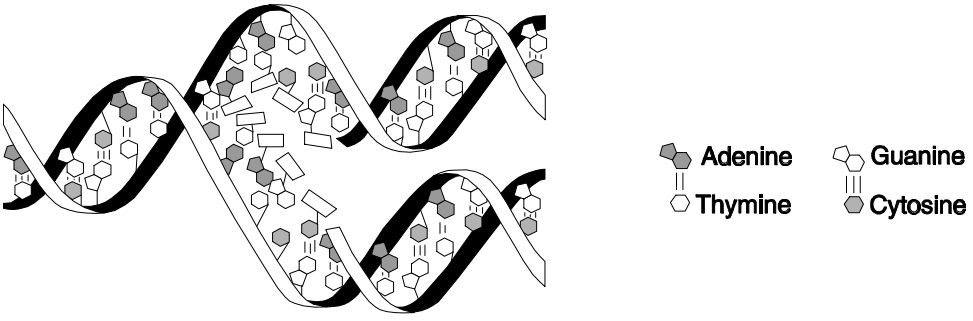

Figure 1. Illustration of the molecular complementarity for four nucleotides: base A (adenine) and base T (thymine), and base G (guanine) and base C (cytosine), respectively.

Denaturation of DNA induces changes in its physical properties, including thermodynamic and optical features. The midpoint of a melting curve that expresses changes in optical density caused by

changes in temperature, indicates the melting temperature, denoted here as ϑ_m, over which the strands of DNA separate. According to the equation used by Breslauer et al. [3], a value of the duplex ϑ_m can be calculated as

$$\vartheta_m = \Delta H / [\Delta S + R \ln(c / 4)] + k, \quad (1)$$

where ΔH and ΔS are the sum of enthalpy and entropy, respectively, for all neighbor dimers, R means the gas constant, c is oligonucleotide concentration, and $k = 16.6 \log c_s$ is the experimental constant necessary to correct for a salt concentration c_s. On the other hand, the stability ΔG of a given pair at a given temperature ϑ can be evaluated by their standard thermodynamic relationship:

$$\Delta G = \Delta H - \vartheta \Delta S. \quad (2)$$

Considering different order of four possible bases, the identical base compositions in DNA molecules can exist at different thermodynamic profiles which result from nearest neighbor interactions. The Watson-Crick base pairing reduces the number of 16 possible dimers to 10 nearest neighbors AA/TT, AT/TA, TA/AT, CA/GT, GT/CA, CT/GA, GA/CT, CG/GC, GC/CG, GG/CC and the same number of thermodynamic parameters ΔH, ΔS and ΔG (Figure 2). Furthermore we have to use the relations that: TT/AA $\rightarrow$ AA/TT, AC/TG $\rightarrow$ GT/CA, AG/TC $\rightarrow$ CT/GA, TC/AG $\rightarrow$ GA/CT, TG/AC $\rightarrow$ CA/GT, CC/GG $\rightarrow$ GG/CC. This model allows to calculate a thermodynamic stability of nucleic acid duplexes with a sufficient precision to predict experimental conditions for the hybridization experiments of any oligonucleotide which length is few tenths of nucleotide units.

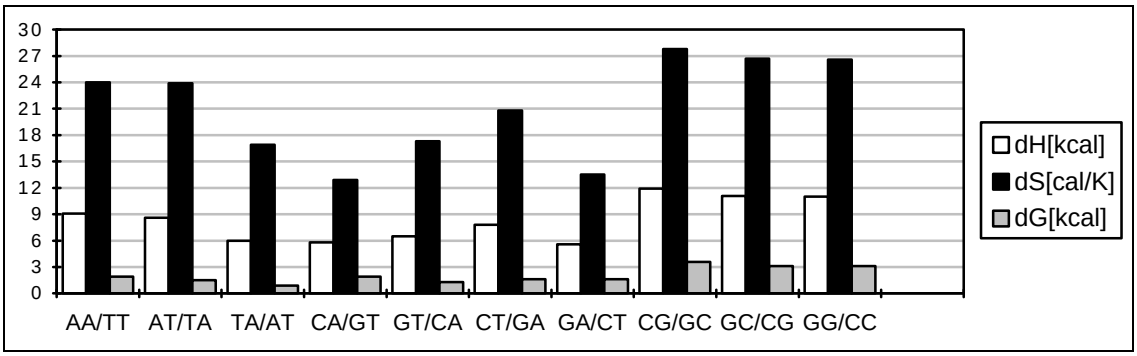

Figure 2. Values of quantities evaluating nearest neighbor thermodynamics.

The previous experience with utilization of software MELTEM [4] was taken here to study thermodynamic relationships for the nearest neighbor basepairs in a given DNA chain where sequence clustering depends on oligonucleotides melting temperature. The MELTEM allows to calculate values of ϑ_m for oligonucleotides of a given length and to generate subsets of oligonucleotides in a given DNA sequence. An illustrating example concerning a given short sequence is given below:

Upper sequence (12-mer): 5' AACGTTCAACCA 3'
Lower sequence (12-mer): 5' AACGTTCAACCA 3'
Upper/Lower dimer formation:

```
      5'  AACGTTCAACCA  3'
          | | | | | |
3'  ACCAACTTGCAA  5'
```

Complementary sequence for upper sequence:
5' TGGTTGAACGTT 3'

Upper sequence: 5' AACGTTCAACCA 3'
Complementary sequence (lower sequence): 5' TGGTTGAACGTT 3'
Melting temperature: ϑ_m = 26.5 °C,
Thermodynamic quantities: ΔH = 86.9 kcal/mol, ΔS = 217.6 cal/K·mol, ΔG = 21.7 kcal/mol.

The extent of hybridization between two-single-stranded nucleic acids can be taken to represent their degree of complementarity. Two sequences need not be perfectly complementary to hybridize; if they are closely related but not identical, an imperfect duplex is formed in which base pairing is interrupted at positions where the two single strands do not correspond.

3 METROLOGICAL APPROACH TO THE NEAREST NEIGHBOR BIOOPTICS

3.1 Specificity of changes in optical density

The nearest-neighbor model [3] allows to predict an optical output resulting from changes in a DNA duplex status: melting temperature, thermodynamic data and optical density OD are quantities related each other [5]. The melting curve OD = f(ϑ_m) takes the form like a step function, however, its position on the temperature scale depends on the base composition. The C-G base pair is more stable than A-T base pair; if DNA is rich in C-G content, the greater energy is needed to separate the strands. What is specific that the optical density of DNA is lower than would be observed by a mixture of free nucleotides of the same composition. Thus, any disruption of the duplex has to cause increase in the optical density connected with the hyperchromicity. The higher stability in the same temperature means the lower optical density. Concluding: in the same conditions, the C-G pairs give lower components of total optical density than the A-T pairs.

Optical density is an useful measure of light attenuation to be caused by absorption and scattering. Optical media can be evaluated using transmission, reflection or emission kind of light-object interaction. The latter considered herein depends on utilization of optical signals to be acquired from a set of known oligonucleotides (making an integrated library) by their selective fluorescence in sites of hybridization. Using a given DNA sequence, which is labeled with an organic dye, makes possible to induce fluorescence as the effect of hybridization of this sequence with complementary oligonucleotides located in a library [6].

In the process of laser-induced fluorescence excitation, a molecule absorbs a photon of light, is excited to a higher molecular state, and decays from that state in such a way that fluorescent light of a longer wavelength is emitted. The reemission occurs in less than 10 ns. Organic dyes have specific absorption peaks at specific wavelengths. Using proper fluorophores, e.g. fluorescein or Texas Red, a large Stoke's shift can be realized. The frequency range of interest consists of the ultraviolet (180-390 nm) and the visible (390-780 nm) ranges. Nucleotides can absorb strongly ultraviolet and blue radiation (what is characteristic for each base that a maximum of absorption is close to 260 nm) – thus light in the red, orange and yellow regions may be emitted by fluorescence and then reliably detected. Selective absorption of UV and VIS radiation in DNA depends on absorption by molecular electrons. While an argon ion laser 488 nm can be used to excite a dye, a photomultiplier tube that is able to emit electrons in proportion to the intensity of light striking it may be a very sensitive photodetector [2]. In accordance to Beer's law, the total fluorescence intensity I_F is proportional to light absorbed I_{abs} and to the fluorescence efficiency η_F which expresses the ratio of quanta absorbed to the quanta emitted. The related relationship may written as:

$$I_F = I_{abs} \cdot \eta_F = (I_{IN} - I_T) \cdot \eta_F = I_{IN} \cdot (1 - 10^{-OD}) \cdot \eta_F, \tag{3}$$

where I_{IN} and I_T are intensities of incident and transmitted light, respectively, while OD = f(ϑ_m) is the optical density. Sensitivity can be improved by employing a fluorescent dye with the optimal excitation wavelength and a filter for optical conditioning. Because the photocurrent obtained from the photodetector has to depend on the signal-to-noise ratio, a dark current can be a significant factor influencing reliability of results.

3.2 Characteristics of fluorescence image from oligonucleotide library

The thermodynamic parameters predicted from the nearest neighbor model for duplexes are connected with a given molecule biooptics: melting temperature and optical density are related each other. The process of denaturation results in increase in the optical density and consequent decrease in light intensity emitted by fluorescence. On the other hand, effects of hybridization reflected by related intensity of light depend on several parameters, including concentration of dye used for labeling, temperature, salt concentration. The maximum fluorescence intensity of a given sequence of nucleotides depends very much on a length and base content, e.g. the 'strong' pairs C-G (equal to G-C) will shine more than the 'weak' pairs A-T (equal to T-A). Figure 3 illustrates general principles of modeling of an optical response from a determined library array.

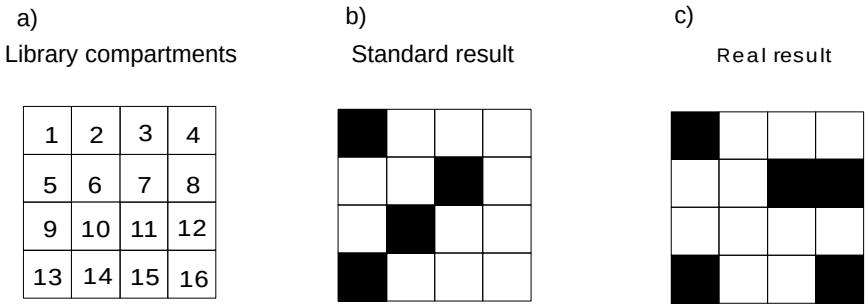

Figure 3. Illustration of modeling of optical effects occurring in a virtual library of oligonucleotides as a result of hybridization; a) example: a segment of library consisting of 16 integrated compartments, b) fluorescence image for a given standard DNA sequence, c) fluorescence image for a standard DNA sequence with induced changes.

The dark squares mean sites of fluorescence peaks – there are three differences in the image in Figure 3c when compared to a 'standard' image assumed in Figure 3b: positions 8 and 16 are two new sites of full complementarity while position 10 shows a lack of perfect hybridization. Concluding: induced fluorescence can perform an optical detection of changes in a given DNA chain when compared with its standard configuration.

Difficult interdisciplinary problems are connected with design of a virtual library of oligonucleotides (which are the known short L-mers) in order to estimate effects of fluorescence induced by light-directed hybridization with corresponding fragments of a labeled sequence to be studied (which one is an unknown long N-mer). From the metrological point of view, such a library must meet requirements to perform reliable results by sufficient resolution and sensitivity of detection and accuracy of optical responses. Each L-mer should be located in its own known position to detect its contribution to hybridization. Detection of hybridization status bases on analysis of fluorescence signals to be obtained from particular sites. Thus, the optical consequences of differences in ϑ_m appear in the library as micro-images corresponding to the places where process of hybridization is performed. Major problems connected with interpretation of a given optical outcome from the library achieved for a given N-mer to be studied concern the questions:

1) capacity, density and arrangement of the library volume,

2) number of levels of light intensity which should be distinguished to estimate or predict changes in the micro-images.

The whole menu of the high-density library can be divided into M integrated compartments. In particular compartments, the determined oligonucleotides of the same length L but different composition are placed as testing sequences TSs. A number of TSs makes a library volume and is equal to 4^L, however, content of compartments may decide about variants of the library. Two useful variants are as follows: 1) $M = 4^L$, i.e. each compartment contains only one TS, and 2) $M = 4^L/4$, i.e. a given compartment consists of four oligonucleotides which differ each other only by one base of a given position. Depending of composition of bases included in a given TS, a given light signal can be expected. TSs will be included in the optical spectrum of a labeled N-mer when the levels of emitted light allow to accept these testers as fully complementary. If any TS in a given compartment is able for full hybridization, only this one will be taken as the oligonucleotide detected because the obtained image is most bright when compared to all others images. Figure 4 illustrates an approach, based on the nearest-neighbor model, used here to assume number levels of light intensity to be detected.

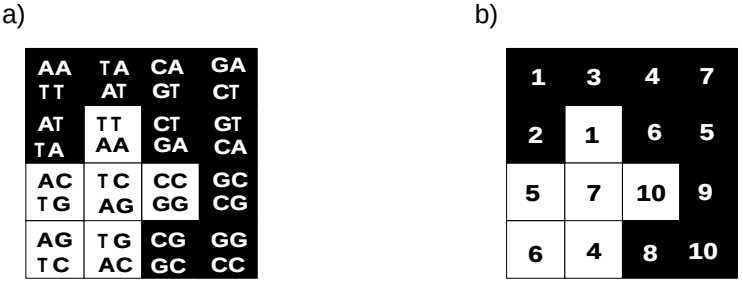

Figure 4. Illustration of relationship between a) 16 possible dimers and 10 different nearest neighbors and b) 10 (against 16) levels of light intensity to be detected from a matrix of 4^2 neighbors.

The example of analysis concerning the sequence from Section 2 (N =12), when to assume L = 6 and data from Figure 4, is given below. In such a case the whole library consists of 4096 oligonucleotides, however, in process of regeneration of the studied DNA sequence only 7 fully complementary 6-mers and related sites of full hybridization in the library are employed. The relationships derived to predict differences in optical response from the considered sites in the library are as follows:

Sites of full hybridization for a single labeled sequence to be studied: 5' AACGTTCAACCA 3'		Related nearest neighbors	Levels of light intensity
1. TTGCAA	pairs: 4 A-T, 2 G-C	AA/TT, AC/TG, CG/GC, GT/CA, TT/AA	1, 5, 8, 5, 1
2. TGCAAG	pairs: 3 A-T, 3 G-C	AC/TG, CG/GC, GT/CA, TT/AA, TC/AG	5, 8, 5, 1, 7
3. GCAAGT	pairs: 3 A-T, 3 G-C	CG/GC, GT/CA, TT/AA, TC/AG, CA/GT	8, 5, 1, 7, 4
4. CAAGTT	pairs: 4 A-T, 2 G-C	GT/CA, TT/AA, TC/AG, CA/GT, AA/TT	5, 1, 7, 4, 1
5. AAGTTG	pairs: 4 A-T, 2 G-C	TT/AA, TC/AG, CA/GT, AA/TT, AC/TG	1, 7, 4, 1, 5
6. AGTTGG	pairs: 3 A-T, 3 G-C	TC/AG, CA/GT, AA/TT, AC/TG, CC/GG	7, 4, 1, 5, 10
7. GTTGGT	pairs: 3 A-T, 3 G-C	CA/GT, AA/TT, AC/TG, CC/GG, CA/GT	4, 1, 5, 10, 4

Referring to Figures 2 and 4, the higher stability in the same temperature means the lower optical density and higher light intensity. In such an approach, the level 1 is weakest, the level 10 is most bright. Taking into account overlapping intensities, we can predict output images, consisting of shining points. For example, for the above site 7, there are 6 points shining at 6 pairs with resulting levels of intensity such as: C-G (4), A-T (4+1), A-T (1+5), C-G (5+10), C-G (10+4), A-T (4).

4 CONCLUSION

Optical and optoelectronic methods of measurements are of growing importance in current metrology including modern biomedical measurements. Problems with modeling and simulation of fluorescence signals for development of a virtual integrated oligonucleotide library are considered in this work. The fact that DNA can be denatured and renatured has been taken as a base to predict optical effects of hybridization, depending on base pairing between independent complementary single strands. If to consider a set of oligonucleotides which all have their own locations in a library, after hybridization we may obtain a fluorescence result creating a matrix of micro-images; each an image corresponds to the optical answer of a given oligonucleotide included in the investigated sequence spectrum. The fluorescence signal intensity is proportional to the amount of dye, and very depends on the sequence order as well as the extent of hybridization – when the degree of complementarity and a number of G-C pairs increase, the emitted signal also increases. In measurements, qualitative changes in a given sequence will manifest by quantitative changes in micro-images order when to compare standard and real optical outputs.

ACKNOWLEDGMENT

This work was supported by the KBN grant 8 T11F 018 15.

REFERENCES

[1] M.V. Olson, The human genome project, *Proc. Natl. Acad. Sci. USA* **90** (1993) 4338.
[2] J.D. Bronzino (Editor-in-Chief), *The Biomedical Engineering Handbook*, CRC Press, USA, 1995, 2862p.
[3] K.J. Breslauer, R. Frank, H. Blöcker and L.A. Marky, Predicting DNA duplex stability from the base sequence, *Proc. Natl. Acad. Sci. USA*, *Biochemistry* **83** (1986) 3746-3750.
[4] J. Blazewicz, A. Cysewska-Sobusiak, A. Lerczak, M. Kasprzak and W.T. Markiewicz, Parameter analysis of clusters of melting temperatures of DNA chains, *Computational Methods in Science and Technology* **3** (1997) 7-17.
[5] C.E.Longfellow, R.Kierzek, D.H.Turner, Thermodynamic and spectroscopic study of bulge loops in oligoribonucleotides, *Biochemistry* **29** (1990) 278-285.
[6] L.M.Smith, J.Z.Sanders, R.J.Kaiser P.Hughes, C.Dodd, C.R.Connell, C.Heiner, S.B.H.Kent and L.E.Hood, Fluorescence detection in automated DNA sequence analysis, *Nature* **321** (1986) 674-679.

AUTHORS: Univ. Prof. Dr. Anna CYSEWSKA-SOBUSIAK, M.Sc. Michal BOLTRUKIEWICZ, Institute of Electronics and Telecommunications, Poznan University of Technology, Piotrowo 3A, 60-965 Poznan, Poland, Phone Int. +48 61 6652633, Fax Int. +48 61 6652572
E-mail: cysewska@et.put.poznan.pl boltruk@et.put.poznan.pl