Nucleic acid – protein fingerprints. Novel protein classification based on nucleic acid – protein recognition

Alexander S. Krylov[1#], and Renad I. Zhdanov[1,2]

[1] Institute of General Pathology and Pathophysiology, Russian Academy of Medical Sciences, Russian Federation, Moscow, Baltiyskaya 8, E-mail: akrylo2@mail.ru

[2] Kazan Federal University, 18, Kremlin St., Kazan 420008 Russian Federation

Key words: protein classification, nucleic acid – protein recognition, oligonucleotides, microarray

Nucleic acid – protein fingerprints. Novel protein classification based on nucleic acid – protein recognition

Alexander S. Krylov[1], and Renad I. Zhdanov[2]

[1] Institute of General Pathology and Pathophysiology, Russian Academy of Medical Sciences, 8, Baltiyskaya St., , Moscow, Russian Federation, E-mail: akrylo2@mail.ru

2 Kazan Federal University, 18, Kremli St., Kazan 420008 Russian Federation

Abstract

Protein chemistry uses protein description and classification based on molecular mass and isoelectric point as general features. Enzymes are also compared by enzymatic reaction constants, namely K_m and k_{cat} values. Proteins are also studied by binding to different oligonucleotides. Here we suggest a simple experimental method for such a comparison of DNA binding proteins, which we call "nucleic acid-protein fingerprints". The experimental design of the method is based on an use of short oligonucleotides immobilized inside microarray of hydrogel cells - biochip. As a first stage, we solved a simple experimental task: what is the shortest single strand oligonucleotide to be recognized by protein? We tested binding of oligonucleotides from 2 to 12 bases, and we have obtained unexpected result that tetranucleotide one is long enough for specific protein binding. This 4-mer can contain two universal bases – 5-nitroindole nucleoside analogue (Ni) and only two meaningful bases, like A, G, T and C. The result obtained opens a way for constructing the simplest protein binding microarray. This microarray consists of 16 meaningful dinucleotides, like AA, AG, CT, GG etc. Physical sequences of all the nucleotides were NiNiAA, etc, where Ni is bound to gel through the amino linker. We prepared such an array and tested it for specific binding of several DNA/RNA binding proteins, labeled with fluorescent dyes like Texas Red of Bodipy. We tested RNase A and Binase for binding on the simplest microarray. It contains only 16 units, and there is a significant difference in the binding patterns. The microarray based on 3-mers must contains 64 units and must have much more specificity. The new principle of protein classification based on nucleic acid-protein recognition has been proposed and experimentally proved. Such an experimental approach must lead to a universal classification of specific DNA/RNA binding proteins.

Modern protein chemistry is based on molecular mass and isoelectric point as general features of protein description. Enzymes are also compared by K_m and k_{cat}. Currently, proteins are also studied by means of binding to oligonucleotides of different length and sequence. A whole large class of DNA/RNA binding proteins has no general scale for its reciprocal comparison. The process of nucleic acid – protein recognition is studied by different methods, like biochips, X-ray analysis, gel retardation, surface plasmon resonance, etc. [1 - 4]. The main aim of the study was to propose a simple experimental method for such a comparison of DNA binding proteins, which we call "nucleic acid-protein fingerprints". The method must be universal for all DNA/RNA binding proteins and based on a simple experimental procedure. We proposed binding to short single strand oligonucleotides as a universal procedure for identification of DNA binding specificity and patterns obtained can be named as "nucleic acid – protein fingerprints or nucleic acid – protein fingerprinting". Such oligonucleotides would be short, and therefore, the first stage of the study was to solve a basic technical and scientific problem: what is the shortest oligonucleotide, which binds proteins in a specific way? The most of the experimental methods on protein – DNA interaction are based on oligonucleotides with length from 10 to 50 bases or longer [2]. Our approach was to find the left side of this scale, i.e. to appreciate the shortest oligonucleotides, which can specifically bind to proteins. We have found the shortest possible oligonucleotides and designed a biochip (DNA microarray) on a base of them. The present paper describes hybridization of proteins with microarray of short oligonucleotides as a new and universal instrument for classification of DNA/RNA binding proteins.

The set of oligodeoxyribonucleotides (ON) was synthesized to appreciate the minimum size of ON for stable and specific binding of proteins. The set consisted of the following seven ONs:

1. 5'-G-A-3'-NH_2 L = 2
2. 5'-G-A-G-3'-NH_2 L = 3
3. 5'-G-A-G-A-3'-NH_2 L = 4
4. 5'-G-A-G-A-G-A-3'-NH_2 L = 6
5. 5'-G-A-G-A-G-A-G-A-3'-NH_2 L = 8
6. 5'-G-A-G-A-G-A-G-A-G-A-3'-NH_2 L = 10
7. 5'-G-A-G-A-G-A-G-A-G-A-G-A-3'NH_2 L = 12

All the ONs were immobilized in gel pads of a microchip. Each ON was immobilized in a separate pad, three pads for one ON. The enzyme Binase, which is known to bind oligonucleotides specifically, was labeled with TR at low ratio TR / Protein. The microchip with immobilized ONs was subjected to hybridization with TR labeled Binase. Figure 1 demonstrates the results of hybridization. One can see that the intensity of hybridization increases sharply with the increase of

ON length from 2 to 6 with the slow increase from 6 to 12. The digital results of hybridization are presented in Figure 2. The dependence, shown in Figure 3 demonstrates a plateau for ONs of the length from 4 to 7. The further increase of the hybridization signal can be described by mechanical increase of the length of ONs. These results give us a possibility for manufacturing a microchip, based on short ONs – which are tetradeoxyribonucleotides.

The second step of our study was to design a simple chip for binding proteins, which is based on 4-mer ONs. We provided a special experiment to study possible introduction of 5-nitroindole-2'-deoxyribose nucleoside analogue in a chain of 4-mer ONs. 5-Nitroindole (Ni) is a hydrophobic aromatic compound, and it can be used as a universal nucleobase analogue in oligonucleotides, because it does not discriminate between the four natural nucleotide bases during duplex formation. The main idea of our experiment was to find an ON with the physical length of four bases, where only two bases are recognized by proteins. To solve this task we prepared a set of ONs with the following sequence:

1. 5'-G-A-3'-NH_2
2. 5'-G-A-Ni-3'-NH_2
3. 5'-G-A-Ni-Ni-3'-NH_2
4. 5'Ni-G-A-Ni-3'-NH_2

All the ONs were immobilized in gel pads of a microchip. The results of hybridization with Binase are presented in Figure 3. The main result of this experiment is that ON, consisting of four bases, where only two bases are used for recognition is a best base element for biochip. The ON of two bases has small signal of protein binding, the 3-mer has the same. The important result is that only 4-mer with GA at the 5'- end and NiNi at the 5'- end has a strong hybridization signal. The ON with the sequence NiGANi has a low signal. It means, that protein binds to dinucleotide GA, which is separated from the point of immobilization by dinucleotide NiNi. The ON GANiNi really has physical length of four bases and only two of them are meaningful. Such a model for recognition is easy to realize in practice.

We synthesized a set of 16 ONs with two different bases at the 5'- end and the tail of the structure NiNi-NH_2 at 3'- end. Typical structure of ONs was like 5'-G-T-Ni-Ni-3'-NH_2. All this set of ONs was immobilized in gel pads. We provide hybridization of this 16-mer biochip with fluorescently labeled proteins Binase and RNase A. The layout of the biochip is presented in Table 1. The visual results of hybridization are presented in Figure 4. One can see that visual signals of hybridization are very different for all dinucleotides and they form a specific pattern, which we call "fingerprints of protein – DNA interaction". This picture demonstrates that even on a visual level

you can appreciate the specificity of DNA/RNA binding proteins. Digital results of hybridization for two different proteins are presented in Figure 4 (A, B). One can see that the enzymes have different patterns of DNA recognition. For example, both Binase and RNase A have strong signals of hybridization (near to 1300 a.u.) with G containing dinucleotides like GG, GA, GC and GT. But, the column, containing A at second position is very different for the two enzymes. Signals of GA, AA, CA and TA are equal to 1200, 170, 205 and 250 in case of Binase and they are very different in case of RNase A – 730, 80, 509 and 470. The differences are very striking and can be detected even on a visual level, like the picture in Figure S1. The method "fingerprints of protein – DNA Interaction" form specific image, which can be individual for all DNA binding proteins.

Thus, we propose a novel general approach of proteins classification based on specific binding of proteins to DNA. This approach has been realized on a simplest platform, containing gel pad biochip with short, tetranucleotides. We have found that even 4-mer ONs with two recognizable bases can bind proteins in a specific way. The main result is that biochip containing 16 dinucleotides can be used for visual images of many DNA/RNA binding proteins. The further improvement of this method includes creation of biochip with 4 recognizable bases. The 4-mer platform shall include 128 elements and could create more detailed image of proteins.

Methods

All the oligodeoxyribonucleotides (ON) used in this work were synthesized on an Applied Biosystems 394 DNA/RNA synthesizer, using standard phosphoramidite chemistry and 30-C(7) amino modifier CPG (Glen Research, Sterling, VA). All the ON used have the structure 5'-XX-3'-NH_2, where XX is ON with the length from 2 to 12 chains containing four bases (A, T, G, C) and Ni – 5-Nitroindole nucleobase analogue, NH_2 is an amino-linker used to immobilize the ON in the polyacrylamide gel pads of the microchips. The Texas Red sulphonyl chloride dye (TR) was from Molecular Probes, Eugene, OR.

The biochips were manufactured in two steps. First, the arrays of 144 (12x12) 5% polyacrylamide gel pads fixed on a hydrophobic glass slide (100x100x20 μm spaced at 200 μm) were prepared by photopolymerization [5]. Then 1 nl droplets of 1mM aqueous solutions of ON were applied to each gel pad and the ON were immobilized by reductive coupling of their aminogroups with the aldehyde groups of the gel [6]. Thus, the biochip was formed with single stranded oligonucleotides immobilized inside gel pad.

Binase from *Bacillus intermedius* of analytical grade was a gift of Dr. V.A.Mit'kevich [7] and Ribonuclease A (RNase A) was obtained from Sigma.

The fluorescence intensity measurements were performed by an automated 3.5 mm x 3.5 mm field epifluorescent microscope with mercury lamp as the excitation source and the filter set for TR dye (LOMO, Russia) (λ_{ex} = 580 nm, (λ_{em} = 630 nm). The microscope was equipped with a CCD-camera (Princeton Instruments, Trenton, NJ), and a computer supplied with a data acquisition board (National Instruments, Austin, TX). The fluorescence intensity measurements were carried out by scanning the microchip of a two-coordinate table, step motors, and a controller (Newport, Irvine, CA). Special software was designed for experimental control and data processing using the LabVIEW virtual instrument interface (National Instruments) [7,8].

References

1. Krylov, A.S., Zasedateleva, O.A., Prokopenko, D.V., Rouviere-Yaniv, J., Mirzabekov A.D. Massive parallel analysis of the binding specificity of histone-like protein HU to single- and double-stranded DNA with generic oligodeoxyribonucleotide microchips. *Nucl. Acids. Res.* **29**, 2654-2660 (2001).
2. Bulyk, M.L., Huang, X., Choo,Y., Church, G.M. Exploring the DNA-binding specificities to zinc-fingers with DNA microarray. *Proc. Natl. Acad. Sci. USA* **98**, 7158–7163 (2001).
3. Garner, M.M., Revzin, A. A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the *Escherichia coli* lactose operon regulatory system. *Nucleic Acids Res.* **9**, 3047–3060 (1981).
4. Jost, J.P., Munch, O., Andersson, T. Study of protein-DNA interactions by surface plasmon resonance (real time kinetics). *Nucleic Acids Res.* **19**, 2788-2788 (1991).
5. Yersov, G., Barsky, V., Belgovskiy, A., Kirillov, E., Kreindlin, E., Ivanov, I., Parinov, S., Guschin, D., Drobyshev, A., Dubiley, S., Mirzabekov, A., DNA analysis and diagnostics on oligonycleotide microchips. *Proc. Natl. Acad. Sci. USA* **93**, 4913 – 4918 (1996).
6. Timofeev, E., Kochetkova, S.V., Mirzabekov, A.D., Florentiev, V.L. Regioselective immobilization of short oligonucleotides to acrylic copolymer gels. *Nucleic Acids Res.* **24**, 3142-3148 (1996).
7. Korchuganov, D.S, Shul'ga, AA, Ermoliuk, Ia.S, Mit'kevich, V.A, Reibarkh, M.Ia, Nol'de, S.B, Makarov, A.A, Arsen'ev, A.S, I87E mutation prevents barstar dimerization. *Rus. J. Bioorg. Chem*. **30**, 638-643 (2004).

Suplementary information

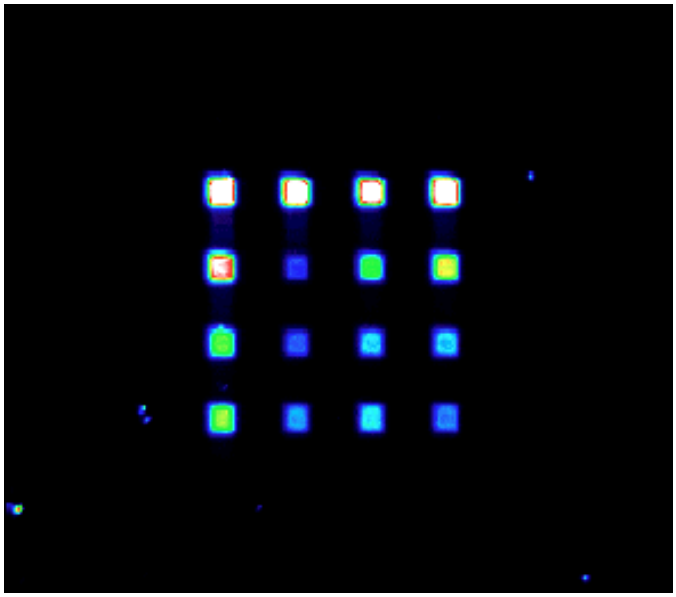
For section Methods:

For the experiments with ss DNA, proteins were labeled with sulfonyl chloride TR (30 mg/mg protein), in accordance with the standard protocol, in 0.15 M NaCl and 0.2 M sodium carbonate buffer (pH 9.0), at room temperature for 1.5 hour [1]. Excessive reagent was removed from the labeled protein by gel filtration on Sephadex G-25. The extent of modification was assessed using mass-spectroscopy at one or two dye residues per protein molecule (data not shown).

TR-labeled proteins binding to microchip was carried out in 200 ml hybridization chamber. Binding was performed at 0 ^{o}C in buffer A (0.2 M NaCl, 10 mM Na–phosphate buffer (pH 7.0), 1 mM EDTA, 0.1% (v/v) Tween-20) containing TR labeled protein at 0.2 mg/ml. After three hours, the solution was discarded and chip was washed with distilled water and dried.

For Results and Discussion section:

Figure S1: The visual image of hybridization pattern of Binase to 2-mer biochip. The intensity of signal increases from deep blue to red and then to white. The layout of biochip see at Table 1.



Acknowledgements

We thank Dr. V. Florentiev and Dr. E. Timofeev (BioChip Center, V.A. Engelgardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow) and Dr. A. Kubatiev (Institute of General Pathology Pathophysiology, Moscow) for help and discussions. We devote this paper to the 75th anniversary of birth of a brilliant scientist Professor Andrey Darievich Mirzabekov (1937-2003), whom ideas partly were used in this work. We acknowledge financial support from Fundamental research program of Russian Academy of Sciences, A. von Humboldt-Stiftung № V-8121/RUS/1032332 (Bonn, Germany) and Kazan Federal University (2011-2012 budget 12-26), Kazan, Russia.

Author Contributions

A.S.K. conceived the study and performed all experiments with DNA biochips. R.I.Z. performed experimental design, oligonucleotide synthesis. A.S.K. and R.I.Z. discussed the data. A.S.K. wrote the paper with significant contribution from R.I.Z.

Author Information

Reprints and permissions information is available at www.nature.com/reprints.

The paper is freely available to all readers. Readers are welcome to comment on the online version of this article. The authors declare no competing financial interests.

Correspondence and material requests should be addressed to: A.S.K. (akrylo2@mail.ru).

Table 1. The layout of biochip for protein – DNA fingerprints

GG	**GA**	**GC**	**GT**
AG	**AA**	**AC**	**AT**
CG	**CA**	**CC**	**CT**
TG	**TA**	**TC**	**TT**

Figure legends:

Figure 1. Visual image of Binase hybridization with oligonucleotides of increased length. The right column has the length 2, the left has length 12. See details in text. The intensity of signal increases from blue to red and, then to white.

Figure 2. The dependence of hybridization signal on the length of oligonucleotides.

Figure 3. The dependence of Binase hybridization signal upon length and Ni position.

Figure 4A. Hybridization pattern of Binase on biochip, based on 2-mer oligonucleotides. The sequences of the last row are GG, GA, GC, GT; the layout of all biochip is presented at Table 1.

Figure 4B. Hybridization pattern of RNase A on biochip, based on 2-mer oligonucleotides. The sequences of the last row are GG, GA, GC, GT; the layout of all biochip is presented at Table 1.

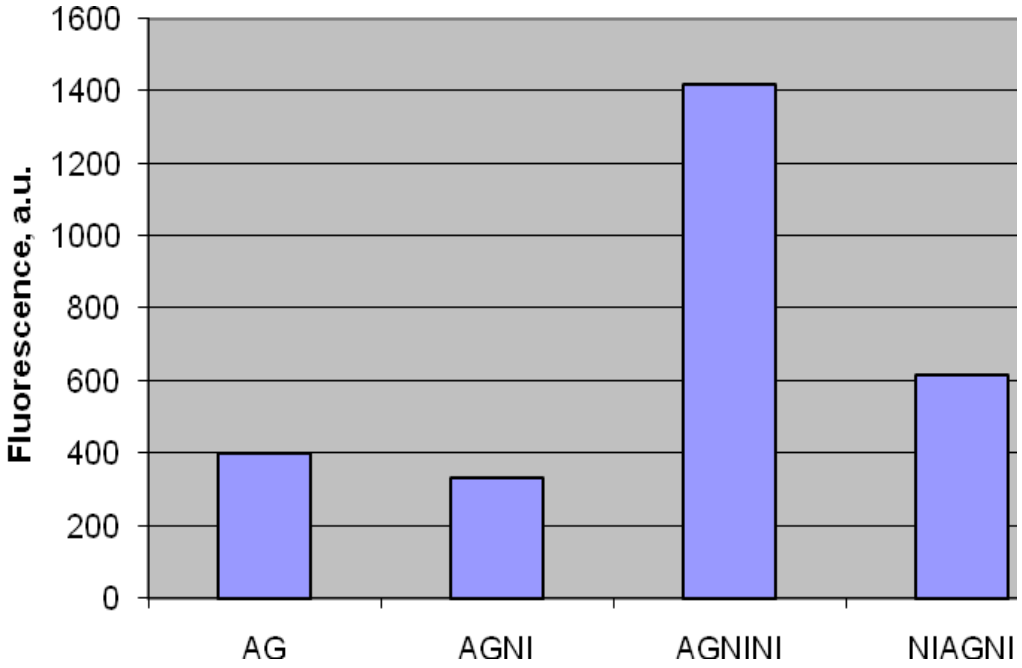


Figure 3

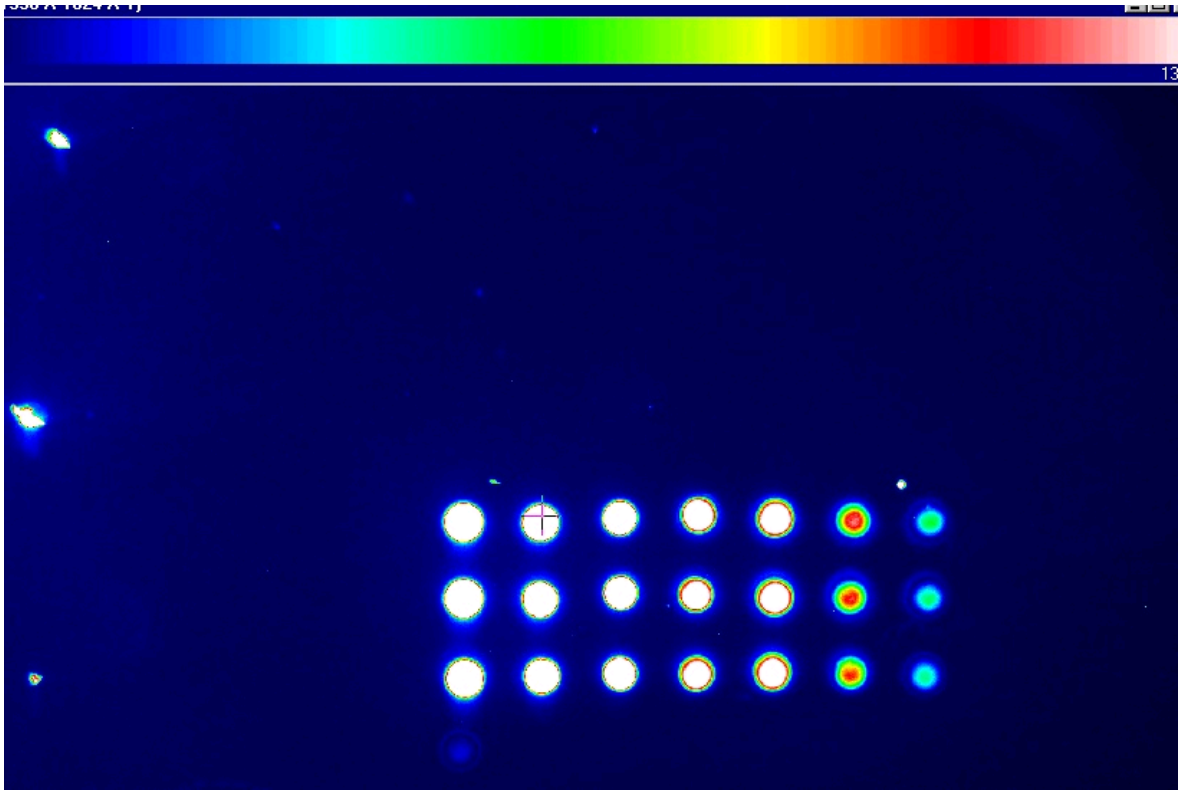


Figure 1

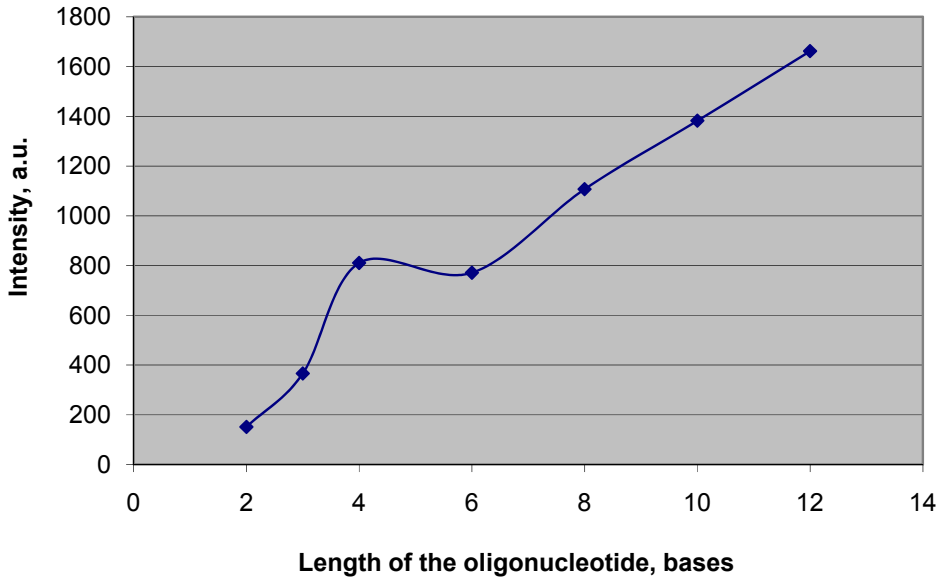


Figure 2.

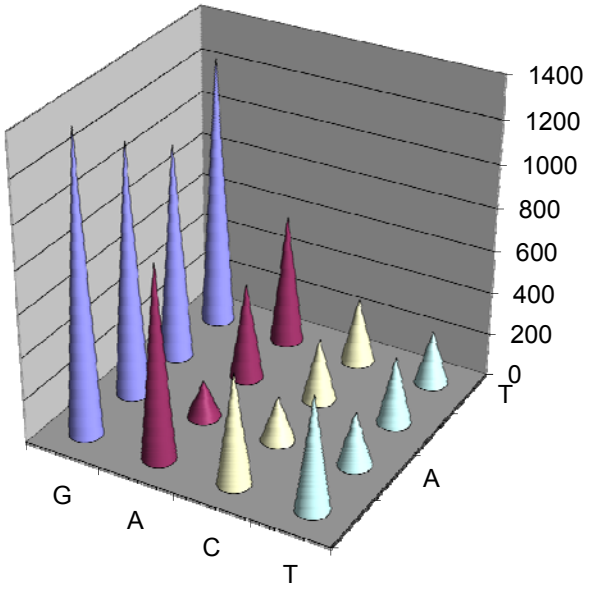


Figure 4 A.

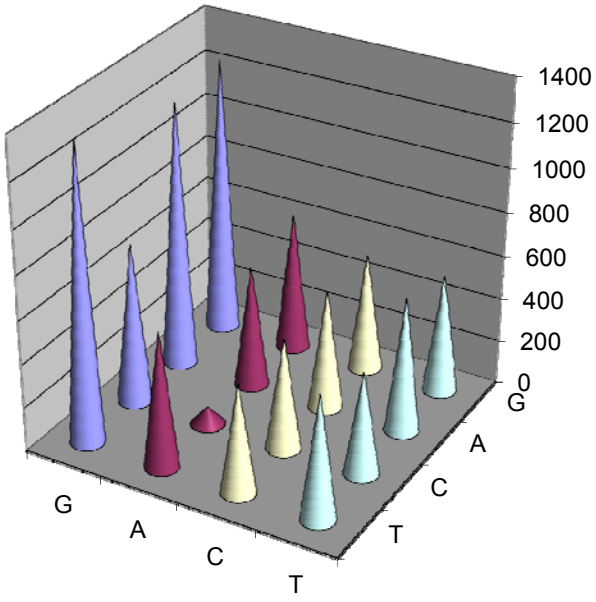


Figure 4 B.