Selection of aptamers by systematic evolution of ligands by exponential enrichment: Addressing the polymerase chain reaction issue

Michael U. Musheev, Sergey N. Krylov

Department of Chemistry, York University, Toronto, Ont., Canada M3J 1P3

Received 5 July 2005; received in revised form 26 September 2005; accepted 29 September 2005

Available online 2 November 2005

Abstract

Aptamers are DNA oligonucleotides capable of binding different classes of targets with high affinity and selectivity. They are particularly attractive as affinity probes in multiplexed quantitative analysis of proteins. Aptamers are typically selected from large libraries of random DNA sequences in a general approach termed systematic evolution of ligands by exponential enrichment (SELEX). SELEX involves repetitive rounds of two processes: (i) partitioning of aptamers from non-aptamers by an affinity method and (ii) amplification of aptamers by the polymerase chain reaction (PCR). New partitioning methods, which are characterized by exceptionally high efficiency of partitioning, have been recently introduced. For the overall SELEX procedure to be efficient, the high efficiency of new partitioning methods has to be matched by high efficiency of PCR. Here we present the first detailed study of PCR amplification of random DNA libraries used in aptamer selection. With capillary electrophoresis as an analytical tool, we found fundamental differences between PCR amplification of homogeneous DNA templates and that of large libraries of random DNA sequences. Product formation for a homogeneous DNA template proceeds until primers are exhausted. For a random DNA library as a template, product accumulation stops when PCR primers are still in excess of the products. The products then rapidly convert to by-products and virtually disappear after only 5 additional cycles of PCR. The yield of the products decreases with the increasing length of DNA molecules in the library. We also proved that the initial number of DNA molecules in PCR mixture has no effect on the by-products formation. While the increase of the Taq DNA polymerase concentration in PCR mixture selectively increases the yield of PCR products. Our findings suggest that standard procedures of PCR amplification of homogeneous DNA samples cannot be transferred to PCR amplification of random DNA libraries: to ensure efficient SELEX, PCR has to be optimized for the amplification of random DNA libraries.

Keywords: Aptamers; PCR; DNA library; Capillary electrophoresis; Protein Analysis; SELEX

1. Introduction

Aptamers are DNA (or RNA) molecules, which are capable of binding different classes of targets with high affinity and specificity [1–3]. Due to their unique properties, aptamers promise to revolutionize many areas of natural and life sciences ranging from affinity separation to diagnostics and treatment of diseases.

In particular, aptamers are very attractive as replacement for antibodies in affinity analysis of proteins. Aptamers have several advantages over antibodies. They are smaller, more stable, can be chemically synthesized, and can be fluorescently labeled without affecting their affinity [4–6]. Aptamers are typically selected from large libraries of random DNA sequences by a procedure termed systematic evolution of ligands by exponential enrichment (SELEX). SELEX involves repetitive rounds of two steps: (i) partitioning of aptamers from non-aptamers by an affinity method and (ii) PCR amplification of aptamers [7,8].

In 2002, we introduced a kinetic homogeneous affinity method termed non-equilibrium capillary electrophoresis of equilibrium mixtures (NECEEM) [9]. The power of NECEEM was demonstrated in a number of applications ranging from the determination of binding parameters to aptamer selection [9–17]. NECEEM-based aptamer selection is characterized by the efficiency, which is at least two orders of magnitude higher...
than that of the best conventional methods [16]. The high efficiency of NECEEM allows for aptamer selection in as few as one round of SELEX [15–17]. Very recently, we introduced an equilibrium technique for the selection of aptamers, termed equilibrium capillary electrophoresis of equilibrium mixtures (ECCEEM) [18]. ECCEEM uniquely facilitates selection of aptamers with pre-defined \( K_a \) values of target–aptamer interaction. The efficiency of ECCEEM is comparable to that of NECEEM. The exceptionally high efficiency of CE-based methods of aptamer selection implies a very low level of background of DNA molecules in collected fractions of aptamers. Because of the low background, the number of DNA molecules collected depends very much on the nature of the target and library used. This number can vary dramatically from selection to selection. In contrast to CE-based methods, conventional heterogeneous methods of aptamer partitioning have high levels of background and relatively constant initial template concentration, which is defined by the background and does not depend on the nature of the target or the library [19–26]. The present work was initially motivated by a suspicion that the variation in the number of DNA molecules collected by CE-based methods could potentially impose strict requirements on the quality of PCR amplification. PCR can yield undesired products due to properties of different DNA template sequences. For example, GC-rich regions can cause the formation of loop structures, which, in turn, causes polymerase jumping and production of shortened products [27–29]. A considerable amount of work was done for the optimization of PCR protocols to meet different specific objectives [30,31]. However, as far as new areas start utilizing PCR for their needs, the performance of PCR in new applications needs to be investigated. We thus studied in detail PCR amplification of random DNA libraries used in aptamer selection. The common methods for analysis of PCR products rely on agarose gel electrophoresis [32]. A number of studies demonstrated that CE is more accurate and reproducible method for quantification of PCR products [33,34]. Moreover, we recently reported on a gel-free analysis of PCR products, which can be used for quantitative analysis of PCR products, primers, and by-products [35]. This approach utilizes the discriminative power of single-stranded DNA-binding protein (SSB) to bind ssDNA but not to bind dsDNA. In the present work, however, we were able to separate PCR products by gel- and SSB-free CE and we applied this power for the quantitative analysis of products and by-products, which accumulate during PCR amplification of random DNA libraries and homogenous DNA samples. We show, for the first time, that PCR amplification of random DNA libraries differs significantly from that of homogeneous samples. If not properly controlled, PCR of DNA libraries could lead to a complete loss of products, which rapidly convert to by-products. We investigated different DNA libraries to see how the length and the initial concentration of DNA in the library, the primers design, and the concentration of Taq DNA polymerase affect the product formation. We showed that the dynamics of product formation depends on the design of the library and primes, but in contrast to our expectations, does not significantly depend on the variation in the initial template concentration. The increase in the length of template decreases the yield of the PCR product and favours its conversion to by-products. Finally, increasing the concentration of DNA polymerase within a large range of values increases the relative yield of PCR products. This work suggests that PCR amplification of random DNA libraries should be carefully optimized for optimum aptamer selection in SELEX.

2. Materials and methods

2.1. Chemicals and materials

Single-stranded DNA-binding protein from E. coli. Taq DNA polymerase, fluorescein, and buffer components were obtained from Sigma–Aldrich (Oakville, Canada). DNA template and PCR primers were custom-synthesized and PAGE purified by Integrated DNA Technologies (Coralville, IA, USA). All solutions were made, using Milli-Q-quality deionised water and filtered through a 0.22-\( \mu \)m filter (Millipore, Nepean, Canada).

2.2. DNA primers and library

6-Carboxyfluorescein (FAM)-labelled 5′-primers were used in PCR reactions for the synthesis of fluorescein-labelled dsDNA products. The libraries were dissolved in water to get appropriate template concentrations and stored at –20 °C. Two sets of primers were used. The first set consisted of 19- and 22-nucleotide primers: 5′-FAM-CTTCGGGCGCCCTCCCTCTTC and 5′-OTGGCCCGCCCTCCCTTCTTC. The second set consisted of 22- and 23-nucleotide primers: 5′-FAM-TTCCAA-CCTTATCAAGCCAC and 5′-GCCCATCCTCCCTCCCTC-TTACC. The homogeneous template had a length of 80 nucleotides: CTTCTGCGGCGCCCTTCCTCCCTGAAAGTCA- TTAATGGTGTTGGGTGGCGCGATGGCAGGACGAG- ATAGCCGCGACACT. Two libraries of random DNA sequences were used, which could produce PCR products of three lengths: 80-, 81-, and 122-base pairs, depending on the primers used. The first DNA library contained a central randomized sequence of 39 nucleotides flanked by 19- and 22-nucleotides constant regions. PCR amplification of this library resulted in 80-base-pairs-long PCR products. The second library contained a central randomized sequence of 36 nucleotides flanked by 19- and 22-nucleotides constant regions for “outer” primers and 22- and 23-nucleotide constant regions for “inner” primers. When outer primers were used, the PCR products had a total length of 122-base pairs. With the inner primers, PCR products had a length of 81-base pairs. Fig. 1 illustrates the structures of the two libraries, which lead to three types of PCR templates.

2.3. PCR

PCR was performed on a MasterCycler 5332 thermocycler (Eppendorf, Germany). In addition to a DNA template, PCR mixtures contained 50 mM KCl, 10 mM Tris–HCl (pH 8.6), 2.5 mM MgCl\(_2\), 200 \( \mu \)M of each deoxynucleoside triphosphate (dNTP), 300 nM of each primer, and 0.05 unit/\( \mu \)l Taq DNA polymerase. In the optimization experiments, the
concentration of Taq DNA polymerase ranged from 0.025 to 0.1 units/μL. The total volume of the PCR reaction mixture was 20 μL. A standard PCR cycle consisted of melting at 94°C for 10 s, annealing at 56°C for 10 s, and extension at 72°C for 10 s. The first cycle always had an extended melting step of 30 s. The final extension step was not conducted. After PCR was stopped, the reaction mixtures were analyzed by CE.

2.4. Capillary electrophoresis

A P/ACE MDQ apparatus with laser-induced fluorescence detection (Beckman–Coulter, Fullerton, CA) was used for CE analysis. Fluorescence was excited with a 488 nm line of an argon-ion laser. Fused-silica capillaries (75 μm inner diameter and 365 μm outer diameter) were purchased from Polymicro (Phoenix, AZ, USA). Capillary length was 50 cm with a distance to the detector of 39.8 cm. The electric field across the capillary was 600 V/cm. The temperature of the capillary was maintained at 15°C. The electrophoresis run buffer was 25 mM sodium tetraborate at pH 9.4. Prior to every CE run, the capillary was flushed with 0.1 M HCl, 0.1 M NaOH, water, and CE run buffer for 2 min each. Samples were injected into the capillary by a pressure pulse of 3.4 kPa for 5 s, which corresponded to a volume of 25 nL.

3. Results and discussion

3.1. PCR amplification of a DNA random library versus a homogeneous DNA sample

First, we compared the formation of products and by-products in PCR of the DNA library and the homogeneous DNA sample, both 80 nucleotides in length. PCR was carried out for 15, 20, and 25 cycles. PCR reaction mixtures were then analyzed by CE with a goal to separate dsDNA products from ssDNA primers and ss-dsDNA by-products[35]. One of the primers was labelled with fluorescein, so that all dsDNA product, the un-consumed primer, and ss-dsDNA by-products were fluorescently labelled after the PCR reaction. The label allowed us to detect the three types of DNA molecules with laser-induced fluorescence detection. Fig. 2 shows electropherograms of the PCR-amplified DNA library (panel A) and PCR-amplified homogeneous DNA template (panel B). The peak with a migration time of 5 min corresponds to the remaining primer, the peak with a migration time of 6.3 min corresponds to the PCR product, and the peak with the intermediate migration times correspond to ss-dsDNA by-products. Product formation for the library reached its maximum after 20 cycles. Five additional cycles completely converted the product to by-products, even though the primers...
We then compared PCR amplification of two DNA libraries with different lengths: 80 and 122 nucleotides (Fig. 1). The short library had a single set of priming regions. The longer library had two sets of priming regions for outer and inner primers. The outer primers of the long library were identical to the primers of the short library. For the long library, if outer primers were used, the resulting products had a length of 122-base pairs; if the inner primers were used, the resulting products had a length of 81-base pairs. Three types of products were thus formed: 80-base pairs, 81-base pairs, and 122-base pairs long. The 80- and 81-base-pair products were of similar lengths but had different primers, while the 122-base-pair products had primers identical to those in the 80-base-pair products and different from those in the 81-base-pair products. We performed PCR amplification of the three types of templates with a varying number of cycles of PCR. Every reaction mixture was analyzed by CE and amounts of (i) the ssDNA primer, (ii) the dsDNA products, and (iii) ss-dsDNA by-products were quantified by measuring areas of corresponding peaks. These data allowed us to plot dynamics of the consumption of primers and formation of products (Fig. 3). Experimental data shown in the three panels of Fig. 3 were obtained within the same PCR run, which makes them comparable. The experiment was run three times to confirm the qualitative trend in the dynamics of product formation (replicas within the same experiment could not be obtained due to the limiting number of wells in the thermocycler). Because of the “semi-quantitative” nature of PCR, the relative standard deviation of “y”-values between different runs was as high as 20%. The qualitative trend in product formation between the three panels within the same run was, however, perfectly reproducible.

The highest amount of the product was obtained for the 80-base-pair long product (Fig. 3, panel A). The 81-base-pair long product formed only to half of the amount of the 80-base-pair product (Fig. 3, panel B). The lowest amount of PCR product was observed for the 122-base-pair long product. The maximum yield of the longest product was achieved after 15 cycles, while shorter products continued to accumulate during 2–3 additional cycles. These results confirm that both primer design and the length of the library influence the dynamics of product formation. We have no rationale for explaining how the primer design influences the conversion of products to by-products. The correlation between the library length and product yield, however, is clear: longer libraries result in a lower yield of the product as the onset of by-products starts earlier and with higher rate. Results obtained here confirm the hypothesis that by-products are formed from product–product hybridization.

### 3.3. Effect of initial template concentration

We then studied the influence of the initial concentration of the template on the yield of the products and by-products in PCR amplification of a random DNA library. Experiments similar to those described in the previous section were performed for the 80-base-pair long DNA library with the initial number of template molecules in PCR ranging from $10^5$ to $10^9$. As we expected, different numbers of PCR cycles were required for reaching the maximum yield of the products: $29$, $25$, $22$, $18$, and $15$ cycles for $10^5$, $10^6$, $10^7$, $10^8$, and $10^9$ template molecules, respectively. Interestingly, however, the maximum yield of products and corresponding yield of by-products did not change significantly with the changing of the initial number of template molecules (Fig. 4). When a homogeneous DNA template is PCR ampli-
The influence of the number of initial template molecules on the maximum yield of products and corresponding yield of by-products for PCR amplification of the 80-nucleotide long DNA library. The concentration of Taq DNA polymerase was 0.05 activity units per microliter. The data are presented for the number of PCR cycles at which the yield of the products reached its maximum.

The explanation lies in different mechanisms of by-product formation in PCR amplification of random DNA libraries and non-specific product formation in PCR amplification of homogeneous DNA templates. The major mechanism of non-specific product formation for homogeneous templates is primer–primer hybridization. Non-specific products are dsDNA, which accumulate slowly with the number of cycles. The number of cycles required for the formation of the same amount of the product decreases with increasing initial concentration of the template; thus, the amount of non-specific products decreases with increasing initial concentration of the template. The mechanism is different for by-product formation in PCR amplification of random DNA libraries. By-products are ss-dsDNA, which are formed through product–product hybridization. By-product formation thus starts only when a threshold amount of the products is formed. From our CE analysis of PCR mixtures, we observed, that by-product accumulation started when dsDNA products reached the concentration of 20–50 nM (total of $10^{11}–10^{12}$ molecules); the primers were still in about 10-times excess to the products (see Fig. 3). The threshold amount of dsDNA product slightly varied with changing the template length.

It should be noted that primer–primer hybridization does not contribute to by-product formation in PCR of random DNA libraries because the formation of by-products through the product–product hybridization starts much later and develops much faster.

Finally, we studied the influence of Taq DNA polymerase concentration on the yield of products and by-products. Experiments were similar to those described in the previous section with the only difference that the initial amount of template molecules was fixed while the concentration of polymerase was varied between 0.025 and 0.1 activity units per microliter. Note that the same number of PCR cycles, 18, were required to reach the maximum of product yield for all polymerase concentrations used. We found that the maximum yield of the product increased with increasing concentration of polymerase in this range (Fig. 5). Concentrations of the enzyme higher than 0.1 activity units per microliter had no effect on the yield of products, while slightly increasing the yield of by-products. This result is different from PCR amplification of homogeneous DNA samples, for which exceeding 0.025 activity units per microliter led to decreasing yield of the products and increasing yield of non-specific products [37,38].

4. Conclusions

In this work, we showed for the first time significant differences between amplification of homogeneous DNA samples and libraries of random DNA sequences. We used highly informative CE analysis of PCR reaction mixtures to quantify accumulated products and by-products as well as remaining primers. We proved that the heterogeneous nature of the library leads to rapid product conversion to by-products. The most likely mechanism of this conversion is product–product hybridization. All evidence, which we introduced here, supports this hypothesis, however to prove it, a more detailed analysis of by-products is...
required. We further demonstrated that the dynamics of product and by-product formation depends on the design of the library and primers. On the other hand, the dynamic of products formation does not depend on the initial concentration of the library but on the concentration of DNA polymerase. This work suggests that PCR amplification of random DNA libraries should be carefully optimized for efficient aptamer selection in SELEX. In particular, it is crucial to avoid over-amplification, which can lead to a complete loss of the products. This work also demonstrates that CE is a powerful tool for optimizing PCR of random DNA libraries.

Acknowledgment

This work was supported by the Natural Sciences and Engineering Research Council of Canada.

References