Sequencing by Ligation with Double-Labeled Fluorescent Probes

Dan Pu, Jing Chen, and Pengfeng Xiao*

State Key Laboratory of Bioelectronics, School of Biological Science and Medical Engineering, Southeast University, Nanjing, 210096, China
xiaopf@seu.edu.cn

Abstract. We have synthesized a set of double-labeled fluorescent probes and evaluated the performance of these fluorescent probes in sequencing by ligation (SBL). The results indicated that the ligation efficiency and specificity of double-labeled fluorescent probes was as higher as that of singe-labeled fluorescent probes. Moreover, double-labeled fluorescent probes yielded a remarkable increase in signal intensities compared to single-labeled fluorescent probes. It could also reduce background and shorten exposition time compared to probes labeled with single FAM and FITC dyes. Thus, it could supplant singe-labeled fluorescent probes in SBL to increase sequencing accuracy.

Keywords: sequencing by ligation (SBL), double-labeled fluorescent probe, single-labeled fluorescent probe, high-throughput sequencing.

1 Introduction

Over the past few years, massively parallel DNA sequencing platforms have become widely available, reducing the cost of DNA sequencing by over two orders of magnitude. Sequencing by ligation (SBL), one of the massively parallel DNA sequencing platforms, uses DNA ligase and either one-base-encoded probes or two-base-encoded probes to determine template sequences [1]. In this method, the specificity and accuracy of probes hybridizing to their complementary sequences adjacent to the primed templates play an important role in the cyclic sequencing cycles. Many researchers have devoted to exploiting new probes for increasing sequencing specificity and accuracy. The SOLiD platform (Applied Biosystems; Foster City, CA, USA) uses specific fluorescent-labeled 8-mer probes, whose 4th and 5th bases are encoded by the attached fluorescent group [2]. Mir et al. have proposed a sequencing method, named cyclic ligation and cleavage (CycLiC), that uses oligonucleotide probes in which all but one nucleotide is degenerate [3]. Ho et al. have described SBL with deoxyinosine-containing query oligonucleotides, which could be digested by endonuclease V, leaving a ligatable end extended into the unknown sequence for further SBL cycles [4]. As an alternative, Li et al. have presented SBL with a kind of probe in which deoxynucleoside phosphorothioates are introduced into the probes to increase the cleavage accuracy of endonuclease V on double-stranded DNA templates [5]. We

have also proposed an SBL approach employing oligonucleotide probes with 3′-thiodeoxyinosine and the read length could reach up to 40 bp with high accuracy [6]. However, when this method was applied to sequence E. coli genome, sequencing error rate increased dramatically since the fourth cycle of ligation and the exposition time of FAM dye was almost the sum of the other three dyes. It was unable to improve the ratio of signal to background after the sixth cycle of ligation and cleavage. Herein, a set of double-labeled fluorescent probes for improving the background and exposition time was synthesized. The performance of double-labeled fluorescent probes in SBL was also evaluated. In addition, we exploited whether the single-labeled fluorescent probes could be replaced by double-labeled fluorescent probes for SBL.

2 Experimental Setup

2.1 Synthesis of Oligonucleotide Probes

All the ligonucleotide sequences used in this study were shown in Table 1. DNA templates and sequencing primers were friendly synthesized by Agene Bioinformative Company (Wuxi, China). T4 DNA ligase and 5×DNA ligase reaction buffer (250 mM Tris-HCl [pH 7.6], 50 mM $MgCl_2$, 5 mM ATP, 5 mM DTT, 25% [w/v] polyethylene glycol-8000) were purchased from Invitrogen^{TM} (Shanghai, China).

Table 1. Oligonucleotide sequences used in this study

Title		oligonucleotide Sequences (5′-3′)		
Template	T_1	NH_2-CTTTCCTCTCTATGGGCAGTCGGTGAT**AA**GCGTACCCCTA GCCCAAATGATCTGCTGTACCGTA		
	T_2	NH_2-CTTTCCTCTCTATGGGCAGTCGGTGAT**CC**TCCTTATCGGGC ATCTCACGCCCTGCTGTACCGTA		
	T_3	NH_2-CTTTCCTCTCTATGGGCAGTCGGTGAT**GG**CCATCTGTGCA CTGCCGAAACACTGCTGTACCGTA		
	T_4	NH_2-CTTTCCTCTCTATGGGCAGTCGGTGAT**TT**ACTTGGATCAG GGACGTAGAAGCTGCTGTACCGTA		
S-Primer	SP_1	PO_4^{3-}-ATCACCGACTGCCCATAGAGAGGAAAG		
S-Probe	P_1	6-FAM-IIINNNAA	P_5	$(6\text{-FAM})_2$-IIINNNAA
	P_2	6-FAM-IIINNNCC	P_6	$(6\text{-FAM})_2$-IIINNNCC
	P_3	6-FAM-IIINNNGG	P_7	$(6\text{-FAM})_2$-IIINNNGG
	P_4	6-FAM-IIINNNTT	P_8	$(6\text{-FAM})_2$-IIINNNTT

Underlined bases are areas of sequencing primers hybridization. The bold characters were the queried bases. S-Primer: sequencing primer. S-Probe: sequencing probe. 6-FAM: 6-carboxyfluorescein. $(6\text{-FAM})_2$: two 6-FAM dye labels. N: degenerate base. I: deoxyinosine.

2.2 Hybridization of the Sequencing Primers to the ssDNA Templates

The single-stranded DNA (ssDNA) microarrays were fabricated according to the previously published literature [7]. Amino-modified oligonucleotides, as single-stranded DNA templates, were first diluted in sodium carbonate buffer (0.1 M, pH 9.0) and then transferred to a 384-well plate. The ssDNA templates in the 384-well plate were spotted on the prepared glass slides using a PixSys 5500 microarray (Cartesian Technology). After spotting, the slides were incubated in a humid chamber at room temperature for 4 h and then at 37 ℃ for 2 h. Finally, the slides were washed by 2×SSC/0.5%SDS and 0.1×SSC/0.5%SDS, and then dried by nitrogen.

To hybridize the sequencing primers to ssDNA templates, a mixture containing sequencing primers (4 μM), 1×hybridization buffer (10 mM Tris -HCl, 0.9M NaCl, and 10% sodium dodecyl sulfate) was placed on the arrays by pipette, and then covered with coverslips. Afterwards, the slide was put in a humid chamber. Hybridization was firstly conducted at 80℃ for 5 min, then cooled at room temperature, and finally incubated at 37℃ for 1 h. After hybridization, the slide was washed and then dried by nitrogen.

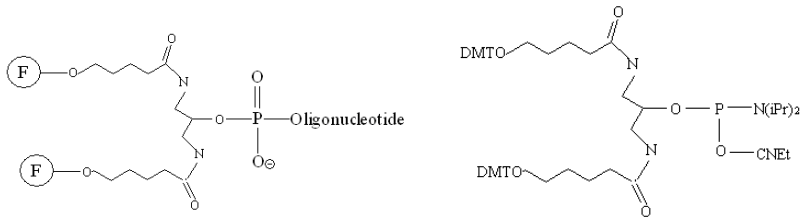
2.3 Ligation of Fluorescent-Labeled Probes

After hybridization, the templates on the microarrays were ligated with labeled probes. Ligation mixture containing the following was prepared and placed on the arrays: 1 ×DNA ligase reaction buffer (50 mM Tris-HCl (pH 7.5, 25 ℃), 10 mM $MgCl_2$, 10 mM DTT, 1 mM ATP), 0.025 U/μL T4 DNA ligase and a mixture of 2 μM fluorescent-labeled probes. Afterward, the microarrays with mix reagent were incubated for 30 minutes at 15℃ in a humid chamber. Finally, the slides were washed and dried by nitrogen. Fluorescent signals on the microarray were captured by using a homemade scanner (Luxscan-10K/A, Capital Biochip Corporation, China), and the data were analyzed with Spot Data Pro 3.0 software.

3 Results and Discussion

3.1 Synthesis of Double-Labeled Fluorescent Probes

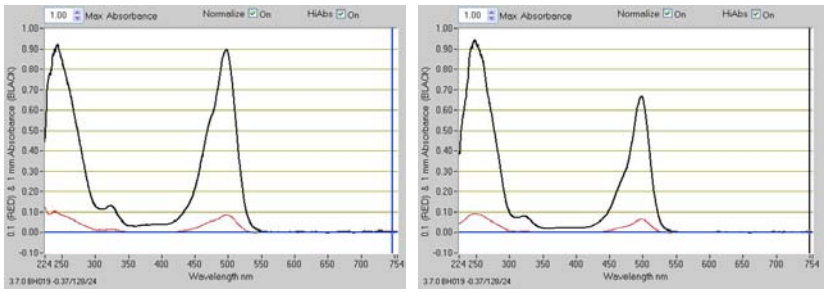
The double-labeled fluorescent probes were synthesized as previously described [8, 9]. To obtain double-labeled fluorescent probes (Fig. 1(a)), oligonucleotide sequence was synthesized in advance. A symmetric doubler phosphoramidite (Glen Research Corporation, Catalog No. 10-1920-90), which could introduce two hydroxyl radical groups to facilitate the conjunction of fluorescent dyes, was coupled to 5′ terminal of oligonucleotide sequence during synthesis (Fig. 1(b)). Subsequently, two fluorephores were linked to the symmetric doubler phosphoramidite. The resultant double-labeled fluorescent probes were characterized by ultraviolet (UV) spectra (Nanophotometer, Implen GmbH) and the results were shown in Fig. 2.



(a) (b)

Fig. 1. The structure of double-fluorescent labeled probe (a) and symmetric doubler phosphoramidite (b).

As shown in Fig. 2, the solid lines represented the normalized absorbance spectra (Groenzin and Mullins, 2000) for double-labeled fluorescent probe P_7 and single-labeled fluorescent probe P_3, respectively. Compared to single-labeled probes P_3, the absorbance peak of 6-FAM dye in double-labeled probes P_7 was much higher when their concentrations are equal. This confirmed that the label efficiency of double-labeled fluorescent probes and single-labeled probes were 120% and 89% when the absorbance of oligonucleotide sequences and 6-FAM dyes were 7.5×104 ($L.mol^{-1}.cm^{-1}$) and 6.2×104 ($L.mol^{-1}.cm^{-1}$), respectively. That was, the label ratio between single-labeled probes and double-labeled probes were 1:1.35 when their concentration were equivalent. Each double-labeled fluorescent probe had approximately labeled 1.2 fluorophore dyes at the 5′ terminus on average while each single-labeled probe had labeled about 0.89 fluorophore dyes.



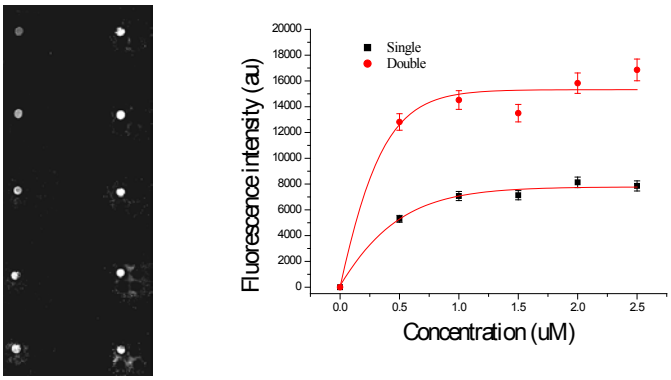
(a) (b)

Fig. 2. The absorption spectra for double-labeled fluorescent probes $(6\text{-FAM})_2$-IIXNNNGG (b) and single-labeled fluorescent probes 6-FAM-IIXNNNGG(a).

3.2 SBL with Double-Labeled Fluorescent Probes

To investigate whether SBL with double-labeled fluorescent probes could be performed successfully, we spotted synthetic templates T_1, T_2, T_3 and T_4 onto a microarray. As shown in the left subarrays in Fig. 3(a), SBL with double-labeled fluorescent probes could successfully performed and each probe could specifically ligated with

their complementary template. To compare the ligation efficiency of SBL with single-labeled probes to that of SBL with double-labeled probes, single-labeled probe P_4 (6-FAM-IIXNNNTT) and double-labeled fluorescent probe P_8 ((6-FAM)$_2$-IIXNNNTT) were ligated with those templates (Fig. 3). As shown in Fig. 3(a), fluorescence intensity of double-labeled fluorescent probe P_8 was nearly two times stronger than that of the single-labeled fluorescent probe P_4 at the same concentration. It indicated that double-labeled fluorescent probes might not quench the fluorephore dyes during ligation reactions, and had nearly no impact on the ligation reactions. The relationship between fluorescence intensity and ligation concentration was shown in Fig. 3 (b), fluorescence intensities of probes P_4 and P_8 increased almost linearly from 0 to 0.5 μM, and then reached a plateau with the highest fluorescence intensity at a concentration of 1 μM. The highest fluorescence intensity kept stable as the probe concentration continued to increase. It indicated that two fluorophore dyes in probe P_8 might slightly affect the hybridization kinetics, but this unobvious effect might be negligible during the half-hour long reactions. That was, probes P_8 and P_4 had almost the same kinetic during the ligation reactions. In addition, probe P_8 and P_4 had the same ligation specificity during ligation reactions. Thus, single-labeled probes could be supplanted by double-labeled fluorescent probes in SBL.



(a) (b)

Fig. 3. The fluorescent intensities of probes P_4 and P_8 at the concentration of 0.5, 1.0, 1.5, 2.0, 2.5 μM, respectively (a) The microarray contained ten subarrays, and each subarray included 4 spots of the templates T_1, T_2, T_3 and T_4. From top to bottom, the left subarrays were ligated with probe P_4 at the concentration of 0.5, 1, 1.5, 2, 2.5 μM, respectively while the right subarrays were ligated with probe P8 at the concentration of 0.5, 1, 1.5, 2, 2.5 μM, respectively. (b) Correlation between the concentration of ligation probe and fluorescence intensity. Single: single-labeled fluorescent probe. Double: double-labeled fluorescent probe.

We also evaluated the performance of iterative cycles of ligation and cleavage with four kinds of double 6-FAM labeled fluorescent probes. The results were compared with that of SBL using single 6-FAM labeled fluorescent probes (Table 2). At each given concentration, the fluorescence intensities of SBL with double-labeled fluorescence probes were nearly one times stronger than that of SBL with single-labeled fluorescence probes. That was, the ratio of fluorescence intensities of SBL with dou-

ble-labeled to single-labeled fluorescence probes was about 2:1. Therefore, SBL with double-labeled fluorescence probes could obtain stronger fluorescence intensity and could improve sequencing accuracy. In addition, SBL with 6-FAM labeled fluorescent probes could avoid longer exposition time and higher background existing in SBL with one FAM labeled, or FITC dyes labeled probes. Although the method has advantage in the improvement of signal intensity and the extension of exposition time, some disadvantages of this method needed to be emphasized. First, the synthesis of fluorescent dyes might increase the cost of SBL. Second, labeling two FAM fluorephore dyes at the 5′ terminus might make the ligation more difficult due to the steric hindrance. However, the increase of signal intensity might make up the higher cost and steric hindrance, since the higher signal intensity might reduce the sequencing error rate and the exposition time of FAM dye was almost the sum of the other three dyes from the fourth cycle of ligation. Thus, double-labeled fluorescence probes could replace single-labeled probes for SBL. It would be useful for improving sequencing accuracy.

Table 2. Fluorescence intensities of single-labeled and doubled-labeled fluorescence probes

Probes	AFI		Probes	AFI		The ratio of AFI	
	1 μM	1.5 μM		1 μM	1.5 μM	1 μM	1.5 μM
P_1	10801	12553	P_5	20845	24359	P_5/P_1=1.93	P_5/P_1=1.94
P_2	10881	12022	P_6	21936	23804	P_6/P_2=2.02	P_6/P_2=1.98
P_3	11964	15606	P_7	28613	29729	P_7/P_3=2.39	P_7/P_3=1.90
P_4	11663	11883	P_8	21106	23172	P_8/P_4=1.81	P_8/P_4=1.95

AFI: the average fluorescence intensities from four iterative experiments. In this table, oligonucleotide sequences P_1~P_4 were labeled with single fluorescence dyes while oligonucleotide sequences P_5~P_8 were labeled with double fluorescence dyes at the 5′ terminus, respectively.

4 Summary

In this report, we have synthesized a set of double-labeled fluorescent probes and have evaluated the performance of double-labeled fluorescent probes in SBL. The results demonstrated that the ligation efficiency and specificity of double-labeled fluorescent probes were as higher as that of singe-labeled fluorescent probes. Moreover, SBL with double-labeled fluorescent probes yielded a remarkable increase in signal intensities compared to SBL with single-labeled fluorescent probes, making SBL reactions more accurate. It also could reduce background and shortened exposition time. Thus, it could supplant singe-labeled fluorescent probes in SBL to increase sequencing accuracy and specificity.

Acknowledgments

This work was funded by the Major State Basic Research Development Program of China (2012CB517706), the National Natural Science Foundation of China

(60971018) a nd the F undamental R esearch F unds f or t he C entral U niversities (CXLX13_112).

References

1. Metzker, M. L. : Sequencing t echnologies-the n ext g eneration,Nat. Rev. G enet. 11, 31 – 46(2010).
2. Mardis, E. R.: Next-generation DNA s equencing m ethods. Annu. R ev. G enom, H uman Genet. 9, 387–402 (2008).
3. Mir, K. U., Qi, H., Salata, O., Scozzafava, G.: Sequencing by Cyclic Ligation and Cleavage (CycLiC) directly on a microarray captured template. Nucleic Acids Res. 37, e5 (2009).
4. Li, Y. Q., Pan, Z. Q., Tang, J., Pu, D., Xiao, P. F., Lu, Z. H.: Accurate cleavage position controlled by deoxyinosine oligonucleotide based on deoxynucleoside phosphorothioates by endonuclease V for sequencing-by-ligation. Analyst, 137, 4421–4424 (2012).
5. Ho, A., Murphy, M., Wilson, S., Atlas, S. R., Edwards, J. S.: Sequencing by ligation variation w ith e ndonuclease V d igestion and deoxyinosine-containing que ry ol igonucleotides. BMC Genomics. 12, 598–606 (2011).
6. Pu, D., Chen, J., Bai, Y. F., Tu, J., Xie, H. M., Wang, W. J., Xiao, P. F., Lu, Z. H.: Sequencing-by-Ligation us ing oligonucleotide probes w ith 3’ -thio-deoxyinosine. J. Biomedical. Nanotechnol. 10, 751–759 (2014).
7. Li, Y. Q., Tang, J., Pan, Z. Q., Xiao, P. F., Zhou, D., Jin, L., Pan, M., Lu, Z. H.: Single nucleotide polymorphism genotyping and point mutation detection by ligation on microarrays. J. Nanosci. Nanotechnol. 11, 994–1033 (2011).
8. Shchepinov, M. S., Udalova, I. A., Bridgman, A. J., Southern, E. M.: Oligonucleotide dendrimers: synthesis and use as polylabelled DNA probes. Nucleic Acids Res. 25, 4447–4454 (1997).
9. Shchepinov, M. S., Mir, K. U., Elder, J. K., Frank-Kamenetskii, M. D., Southern, E. M.: Oligonucleotide dendrimers: s table na no-structures. Nucleic A cids R es. 27,.3035–3041(1999).