Journal of Biochemistry and Molecular Biology, Vol. 37, No. 3, May 2004, pp. 269-274

JOURNAL OF Biochemistry and Molecular Biology

Review

New Performance from an Old Member: SNP Assay and *de Novo* Sequencing Mediated by Exo$^+$ DNA Polymerases

Jia Zhang[†,*] and Kai Li[†,‡]

[†]Genomapping Inc., Tianjin, China, [‡]SNP Institute, Nanhua University, Hengyang, China

Received 13 August 2003

DNA polymerases without the 3' exonuclease function (exo$^-$ pol) have been widely used in sequencing and SNP genotyping. As a major player that expedited the coming of the postgenomic era, exo$^-$ polymerases worked remarkably well in the Human Genome Sequencing Project. However, it has become a challenge for this class of polymerases to efficiently screen the large number of SNPs that are found in the human genome. For more than three decades it has been recognized that polymerase fidelity varied according to the presence of proofreading activity that is mediated by its internal 3' exonuclease. Polymerases with proofreading function are famous for their high fidelity in DNA replication both *in vivo* and *in vitro*, but this well-known class of polymerases has been almost completely neglected in genetic analysis in the postgenomic era. We speculate that exo$^+$ polymerases may exhibit higher nucleotide identification ability when compared to exo$^-$ polymerases for an *in vitro* genetic analysis. With the application of exo$^+$ polymerases in SNP assays, a novel mechanism for the maintenance of DNA replication, the on/off switch, was discovered. Two new SNP assays have been developed to carry out genome-wide genotyping, taking advantage of the enzymatic properties of exo$^+$ polymerases. Furthermore, the on/off switch mechanism embodies a powerful nucleotide identification ability, which can be used to discriminate the bases that are upstream of the 3' terminus, and thus defines a new concept in *de novo* sequencing technology. Application of exo$^+$ polymerases to genetic analysis, and especially SNP assays, will greatly accelerate the pace to personalized medicine.

Keywords: On/off switch, Proofreading, SNP, Sequencing, Exo$^+$ polymerase

*To whom correspondence should be addressed.
Tel: 858-812-1615; Fax: 858-812-1918
E-mail: jia34@yahoo.com or kaili34@yahoo.com

Introduction

The completion of the Human Genome Sequencing Project revealed that single nucleotide polymorphisms are the most common form of genetic variation. The efficient screening of known SNPs is of paramount importance since it can maximize the value of the sequence information in critical applications such as linkage disequlibrium analysis and personalized medicine. SNP-related study or Snpology consists largely of the following three parts: (i) development of SNP assay methodology, (ii) identification of new SNPs and screening of known SNPs, and (iii) studies related to the biological significance of SNPs (Zhang and Li, 2001; Zhang *et al.*, 2001). Efficient and high-throughput compatible SNP assays are the bottleneck in current SNP-related studies.

Although a variety of technologies have been developed for high-throughput SNP screening, a high rate of false positives remains one of the major obstacles to the effective application of available high throughput SNP assays, preventing them from being more widely used in clinical applications. Among the increasing number of SNP assays, many are variants of the allele-specific primer extension by exo$^-$ polymerases (Li *et al.*, 2003). In recent years, we have focused on the development of SNP assays using exo$^+$ polymerases with proofreading activity. This led to the successful development of two new types of SNP assays that are mediated by exo$^+$ polymerases: (i) the 3 terminal-labeled primer extension (Zhang and Li, 2001; Zhang *et al.*, 2003a, 2003b; Gu *et al.*, 2003a) and (ii) the SNP-operated on/off switch (Zhang and Li, 2003; Zhang *et al.*, 2003a, 2003b, 2003c). Furthermore, a proof-of-concept for *de novo* sequencing, based on the on/off switch, has been described (Zhang and Li, submitted). This short paper will briefly review our recent results using exo$^+$ polymerases in genetic analysis.

The 3' Terminally-labeled-primer Extension (TLPEx)

Exo$^+$ polymerases are not as frequently used as exo$^-$

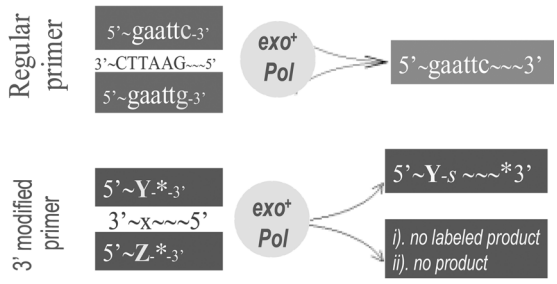

Fig. 1. Illustrations showing the difference between 3' modified and unmodified primers extended by high fidelity polymerase. Upper panel: Polymerases with 3' exonuclease activity discriminate allele-specific primers before their extension. However, the final product is the same irrespective of whether the primer-3'-termini are matched or mismatched. Lower panel: The 3' terminal labeled allele-specific primers are subjected to proofreading during primer extension using the high fidelity DNA polymerase with 3' to 5' exonuclease, depending on their match status. When primer-3'-terminus are matched with the template, then the extended products have a labeled signal incorporated since there is no proofreading involved. However, the labeled mismatched nucleotide is the target of exonuclease excision during proofreading, which removes the mismatched nucleotide before the primer is extended. The *represents the labeled signal.

polymerases in molecular biological and genetic studies. Part of the reason is their low efficiency in sequencing. Another, and possibly the more important reason, is that the proofreading activity decreases, instead of increasing, the nucleotide discrimination in SNP assays when unmodified allele-specific primers are used (Zhang *et al.*, 2003d). As shown in Fig. 1, whether the 3' terminal of an allele-specific primer is labeled or not makes a substantial difference. For the primers without labeling, identical products are generated from matched and mismatched primers by exo$^+$ polymerases (Zhang and Li, 2001). However, for the 3' terminally-labeled primers, the products can be easily differentiated by signal detection since perfect-match primers show positive results and 3' terminally-mismatched primers show negative results; either no product or a product without a label (Zhang *et al.*, 2003b; Cahill *et al.*, 2003).

The terminally-labeled primer extension is a SNP assay consisting of 3' terminally-labeled allele-specific primers and DNA polymerases with proofreading activity. This concept was first disclosed in 2000 to the United States Patent and Trademark Office (Zhang, 2000). As shown in Fig. 1 and Table 1, the 3' terminal modified primers make a substantial difference in the products that are amplified by high fidelity polymerases with 3' exonuclease activity. With unmodified primers, identical products are produced from both the matched and 3' terminal mismatched primers by exo$^+$ polymerases. There is an underlying proofreading process that occurs during primer extension for the 3' terminal mismatched primers, but this enzymatic excision is invisible at the product level. However, with 3' terminal labels, perfect-match primers generate labeled products; whereas, the 3' mismatched primers are expected to yield products without labels if the primers are extended. The high fidelity of the exo$^+$ polymerases ensures that 3' terminal mismatched primers will not be extended unless the mismatched nucleotide with a label is enzymatic removed by 3' exonucleases. Both the 3' terminal ^{3}[H] labeled and fluorescently-labeled primers have been successfully applied in a SNP assay (Cahill *et al.*, 2003; Gu *et al.*, 2003a; Zhang *et al.*, 2003b). The 3 terminally-mismatched nucleotide that bears the signal to be detected is removed by the proofreading function; whereas, the label is retained in the products when the primer and template are perfectly matched.

Interestingly, the difficulty in removing the Rox-labeled 3 terminal nucleotide occurred under standard PCR conditions in our experiments, suggesting the potential interference by this label on 3' exonuclease digestion (Zhang *et al.*, 2003a, 2003b). Recently, another group confirmed our strategy using fluorescently-labeled primers. They reported no difficulties in proofreading the labeled primers (Cahill *et al.*, 2003). It is possible that the experimental conditions, as well as the properties of the labels, may play a role in affecting exonuclease activity.

The terminal-labeled primer extension approach has several advantages over the current SNP assays. Its most significant advantage is that this methodology greatly decreases false positives. It is well known that a high rate of false positives is one of the main obstacles to the clinical application of current SNP assays. The advantage of the terminally-labeled primer

Table 1. Different products from 3 unmodified and modified primers extended by exo$^+$ polymerases

3' terminal of primers	3' base-pairing with template	results of the proofreading	sequence of extended products	features of extended products*
Unmodified	matched primers	no nucleotide excision	keep template-dependent	invisible themselves
	mismatched primers	3' terminal excision	revert to template-dependent	invisible themselves
3'-labeled	matched primers	no nucleotide excision	keep template-dependent	with detectable labels
	mismatched primers	3' labels removed by exo	revert to template-dependent	without detectable labels

*The labels on primer-3'-termini can be a variety of labels, such as isotopes or fluorophores. The proofreading process by 3' exonucleases during primer extension removes the mismatched nucleotide at primer-3'-termini. Therefore, with 3' labeled primers, products can be easily distinguishable, based on the labels in the primer-extended products: products with labels are from matched primer and products without labels are from mismatched primers.

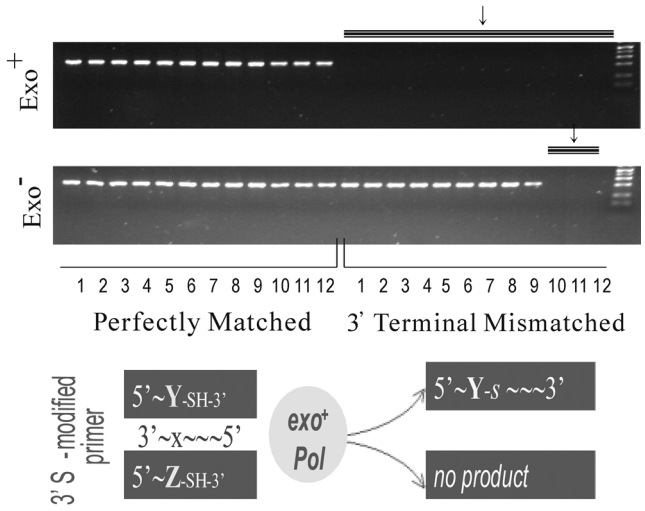

Fig. 2. The combination of 3' phosphorothioate-modified allele-specific primers and exo+ polymerase forms a molecular switch sensitive to single nucleotide differences. Upper panel: Substantial differences were observed between the results from exo– and exo+ polymerases mediated primer extension. The phosphorothioate-modification renders the primers resistant to exonuclease digestion. When the primer and template are complementary, the primers are extended; whereas, the 3' mismatched primers are subjected to exonuclease excision. The primer-3'-termini are trapped within the exo site which depletes the polymerization within the pol site. Therefore, DNA polymerization is turned off by the mismatched primer in this case. Numbers 1 to 12 represent different annealing temperatures from 46.0, 46.5, 47.5, 49.2, 51.5, 54.4, 57.8, 60.6, 62.8, 64.4, 65.5 and 66.0°C respectively. Consistent with literature, polymerases lacking proofreading function only discriminated 3' mismatched primers at specific annealing temperatures, 62.8 or higher in our study. Lower panel: A simplified model showing the on-switch and off-switch from matched and mismatched 3' phosphorothioate-modified primers, respectively.

extension is a consequence of the proofreading activity of exo+ polymerases. Further discrimination, by enzymatic or mechanical separation between the incorporated and unincorporated labels, would contribute significantly to the reduction of false positives. The second advantage of the terminally-labeled primer extension is its high sensitivity. Terminally-labeled primer extensions harness the power of PCR to improve the efficiency of genetic analysis.

SNP-operated On/off Switch

Exo+ polymerases, in combination with 3' phosphorothioate-modified mismatched primers, work as an off-switch in DNA polymerization (Fig. 2) for 3' allele-specific primers with phosphorothioate-modification, the perfect-match primer turns on and the mismatched primers turns off DNA polymerization (Zhang and Li, 2003; Zhang *et al.*, 2003b, 2003c). The result of the off-switch effect is a consequence of the exonuclease-resistant property of the phosphorothioate-modification, which blocks the mismatch excision. Phosphorothioate-modification renders the oligonucleotide nuclease-resistant, a strategy widely used in antisense technology as well as in single-base extension (Li and Zhang, 2001; Di Giusto and King *et al.*, 2003).

During the thermodynamic analysis of 3' mismatched primers, a 3' phosphorothioate-modified primer was used to elucidate the critical role that is played by a successful mismatch excision by 3' exonuclease in DNA polymerization. Exo– polymerases generate primer - dependent products from either unmodified or 3' phosphorothioate - modified mismatched primers in a broad range of annealing temperatures (Fig. 2, upper panel). Nucleotide discrimination by exo– polymerases occurred only at a specific annealing temperature or optimized experimental conditions, as reported in conventional SNP assays using an allele-specific primer extension in a SNP assay. A remarkable phenomenon was observed in the proofreading phosphorothioate-modified primer 3'-termini when there was a mismatch between the primer's 3'-terminus and the template. Instead of producing template-dependent products from exonuclease-digestible primers, the exo+ polymerases generated no products from the mismatched primers with 3' phosphorothioate-modification (Peng *et al.*, 2003; Zhang and Li, 2003; Zhang *et al.*, 2003b, 2003c). This breakthrough observation of an on/off switch action was repeatedly confirmed using either short artificial amplicons or genomic DNA templates. The potential physiological significance of the on/off switch mediated premature termination in DNA replication was confirmed by using unmodified primers with shifting the mismatched

Table 2. Comparison of the two SNP assays mediated by exo+ polymerases

Method	Primers	Available switches	Mechanisms	Visualization	Genomic DNA*
TLPEx	3' labeled	prefer one, either label sense or antisense primer	remove 3' label from MM primer	upon to the label on the primers	applicable
SNP-operated on/off switch	3' PTO-modified	mono- or bi-switch: sense & antisense primers both can be 3' modified	resistant to 3' exo, MM primer is not extended	highly versatile: add labels before, during, or after PCR	applicable

MM, mismatch. *SNP genotyping directly applicable to genomic DNA is the common advantage of these two assays as compared to most other types of SNP assays. Particularly, the SNP-operated on/off switch will add an important impact on microarray-based SNP assay. SNP-operated on/off switch does not require prior PCR amplification, a bottleneck in the currently available SNP assays.

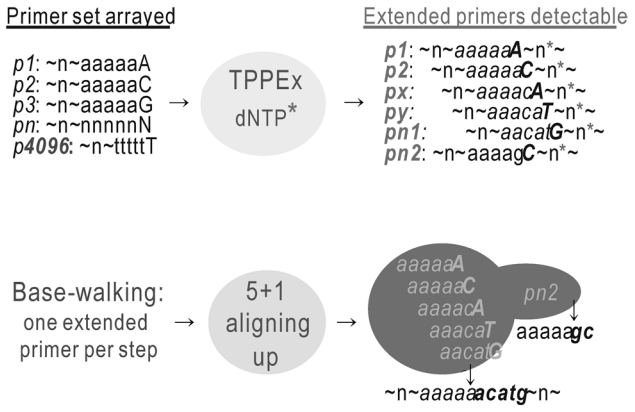

Fig. 3. A nucleotide-walking algorithm for array-based sequencing. In this model of terminal phosphorothioate-modified primer extension (TPPEx)-based nucleotide-walking assay, five matched base pairs are employed in bridging two extended primers. One unknown nucleotide is decoded by walking from the known sequences of one extended primer to another. The sequences of ***acatg*** represented the readout from one such bridged subset of extended primers, which are bolded in the lower panel.

nucleotides upstream from the primer-3'-termini (Zhang and Li, 2003c).

The data from this new on/off switch strongly recommends the potential of phosphorothioate-modified primers for practical SNP screening (Table 2). In the more homogenous reaction condition, the use of one on/off switch, either from the forward or reverse primers, reached a complete off action for exponential DNA amplification. In all of the experiments with short amplicons, only one of the primer pairs was allele-specific and phosphorothioate-modified; while another primer was not allele-specific and usually far away from the SNP in order to adjust to the different length of the amplified products when complex PCR was needed. With more heterogeneous genomic DNA, the application of one on/off switch is sensitive enough to discriminate the wild-type from the mutant-type. Occasionally, a feint-leaking band could be observed in the case of PCR over 30 cycles. One outstanding feature of the novel on/off switch is the possibility of using doubled switches; namely, both the forward and the reverse primers are 3'-phosphorothioate-modified and allele-specific, which is not comprised with lowering sensitivity, as seen in most of other primer extension-based SNP assays. A doubled on/off switch was tested in a SNP assay of a recently-identified GJB3 deafness mutation (Peng *et al.*, 2003; Zhang *et al.*, 2003c). Although the doubled on/off switches help to increase selectivity in some special experiments, such as the detection of mutation load, it is not necessary in SNP genotyping and is thus not recommended.

De novo Sequencing Strategy Using the On/off Switch

The introduction of ddNTP, a terminator, in primer extension paved the way for DNA sequencing (Sanger *et al.*, 1977). The scientific merit and technological value of ddNTP has been certified by the introduction of automated capillary-based sequencing to replace gel-based methods. Sequencing automation has led to the completion of the Human Genome Project. In return, the completion of the Human Genome Project has created a demand for faster and more powerful sequencing technologies. To meet the demand of sequence-based prevention, diagnostics, and therapeutics in personalized medicine as well as in other fields of biology, microarray-based sequencing is superior to electrophoresis and is thus the focus of technology development (Drmanac *et al.*, 1993, 2002). Although both enzyme-mediated and hybridization-based microarrays have been developed for single nucleotide polymorphism assays and for the resequencing of defined genes, the accuracy and cost-effectiveness of this new generation of sequencing technologies are far from being reliable enough to challenge conventional sequencing methods using terminator dideoxynucleotide. The obstacle is the lack of an effective microarray-oriented on/off switch in the development of array-based sequencing technologies.

While in search of new assays for the SNP screening, we discovered a novel on/off switch in DNA polymerization that is mediated by proofreading phosphorothioate-modified 3' allele-specific primers. Functionally, this new on/off switch works the same as that involved in terminator ddNTP. The structure of ddNTP prevents the formation of additional phosphodiester bonds once incorporated; whereas, mismatched-primer-3'-termini with phosphorothioate become physically trapped within the exo site blocking phosphodiester bonds formation in the pol site of the polymerase (Li *et al.*, 2003). With arrayed primer technologies available, the novel on/off switch is very promising in array-based sequencing, such a proof of concept is the nucleotide-walking assay. The following criteria should be met for the array set used in array-based sequencing: (i) each primer has two parts, the degenerate 5' end and the sequence-specific 3' terminus, (ii) all of the primers are 3' phosphorothioate-modified, and (iii) The primer set fully covers the combination of 4^n primers, depending on the number of sequence-specific nucleotides (i.e. n) at the 3' end, which works as base identifier.

Technically, the nucleotide-walking strategy relies on the primer extension using microtiter plates or microarray technology. Known or unknown sequences are used in an arrayed primer extension with exo^+ polymerases. Based on the powerful nucleotide discrimination ability of the SNP-operated on/off switch that accurately identifies a single nucleotide from the 3' terminal up to the - 6 upstream position, each of the extended primers correspondingly represents or decodes six nucleotides of the template according to the 6 nucleotides at the primer's 3' end. Using an algorithm for bridging the extended primers via a 5' base-pair match-up, the DNA template is thereby sequenced one base per step (extended primer) by walking from one extended primer to the

next (Fig. 3). This nucleotide-walking algorithm works for both re-sequencing and *de novo* sequencing through a simple alignment of the last few non-degenerate nucleotides at the 3' end of the extended primers.

When applied to practical array-based sequencing, a detectable signal could be introduced in the extended products by either isotopic or fluorescent labeling of the dNTPs. With the powerful fluorescent-labeling techniques, the suggested nucleotide-walking assay becomes quite feasible. As opposed to sequencing by a hybridization array, the nucleotide-walking assay provides a universal sequencing microarray format that can be manufactured with no consideration for the specific sequence of any individual gene or species. While the classical on/off switch of dNTP/ddNTP is mainly applicable for electrophoresis-based sequencing, our novel proofreading on/off switch using phosphorothioate-modified primers is fully applicable to microarray platforms. Accurately decoding a sequence from an arrayed primer extension is the current technical barrier. The novel on/off switch mechanism might be a solution for array-based sequencing. The novel on/off switch mechanism has the significant advantage that it can be applied to genome-wide searching for new SNPs as well as to a variety of clinical genetic diagnostics

Summary

The high fidelity of exo+ polymerases in DNA replication has been recognized for more than three decades. During our efforts to develop new SNP assays, we demonstrated the presence of a novel mechanism controlling premature termination of DNA chain elongation via a novel on/off switch of 3' exonuclease. The two distinct mechanisms of 3' exonuclease fidelity maintenance (nucleotide mismatch excision and off switch) were used in the development of two SNP assays: 3'-end labeled primer extension and SNP-operated on/off switch. Furthermore, the on/off switch of exo+ polymerases makes possible a novel strategy for *de novo* sequencing. We believe that great and immediate benefit for biomedical research and the healthcare community will be achieved by a wider application of the SNP assays that are mediated by exo+ polymerases and by further studies to introduce additional genetic assays using exo+ polymerases.

Acknowledgments The authors acknowledge the assistance of three Canadian laboratories: Dr. Newman's lab, Dr. Rouleau's lab, and Dr. Stewart's lab. Our exo+ pol-related studies would have been impossible without the excellent technical training that JZ received in Dr. Newmans lab or the inspirational training that KL received under Dr. Rouleau. We could not have envisioned exo+ pol-based technologies without our happy and enriching time in Dr. Stewart's lab in Toronto. Also, we particularly acknowledge Dr. Steve S. Sommer at the Department of Molecular Genetics and Human Genomics in the City of Hope where this manuscript was prepared.

References

Cahill, P., Bakis, M., Hurley, J., Kamath, V., Nielsen, W., Weymouth, D., Dupuis, J., Doucette-Stamm, L. and Smith, D. R. (2003) Exo-proofreading, a versatile SNP scoring technology. *Genome Research* **13**, 925-931.

Di Giusto, D. and King, G. C. (2003) Single base extension (SBE) with proofreading polymerases and phosphorothioate primers: improved fidelity in single-substrate assays. *Nucleic Acids Res.* **31**, e7.

Drmanac, R., Drmanac, S., Chui, G., Diaz, R., Hou, A., Jin, H., Jin, P., Kwon, S., Lacy, S., Moeur, B., Shafto, J., Swanson, D., Ukrainczyk, T., Xu, C. and Little, D. (2002) Sequencing by hybridization (SBH) advantages, achievements, and opportunities. *Adv. Biochem. Eng. Biotechnol.* **77**, 75-101.

Drmanac, R., Drmanac, S., Strezoska, Z., Paunesku, T., Labat, I., Zeremski, M., Snoddy, J., Funkhouser, W. K., Koop, B., Hood, L. *et al.* (1993) DNA sequence determination by hybridization: a strategy for efficient large-scale sequencing. *Science* **260**, 1649-1652.

Guo, Z., Zhang, J., Chen, L. and Liao, D. (2003a) Single base discrimination using terminal labeled primer extension. *J. Nanhua Univ.* **31**, 135-136.

Guo, Z., Zhang, J., Liao, D., Yang, X., He, S. and Li, K. (2003b) Effect of 3' exonuclease activity of polymerase on extension of phosphorothioate-modified primers. *Chin. Med. Genet. J.* **20**, 328-330.

Li, K. and Zhang, J. (2001) ISIS-3521. Isis phamaceuticals. *Curr. Opinion Invest. Drugs.* **2**, 1454-1461.

Li, K. and Zhang, J. (2003) New SNP assays from an old concept of proofreading. *Curr. Drug Disc.* **11**, 37-39.

Peng, Z., Zhang, J., Chen, L. L. and Liao, D. F. (2003) Discrimination of C to T point mutation in GJB3 senorineural deafness by SNP-operated on/off switch. *J. Nanhua Univ.* **31**, 132-134.

Sanger, F., Nicklen, S. and Coulson, A. R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467.

Shamir, R. and Tsur, D. (2002) Large scale sequencing by hybridization. *J. Comput. Biol.* **9**, 413-428.

Zhang, J. (2000) Method for high throughput assay of genetic polymorphism. Disclosure number 486671 to United States Patent and Trademark Office.

Zhang, J. and Li, K. (2001) The 3' terminal labeled primer extension: A new method of high throughput screening for SNP analysis. *Current Drug Discovery.* **9**, 21-23.

Zhang, J. and Li, K. (2003) On/off regulation of 3' exonuclease excision to DNA polymerization by exo+ polymerase. *J. Biochem. Mol. Biol.* **36**, 525-528.

Zhang, J., Li, K., Deng, Z., Liao, D., Fang, W. and Zhang, X. (2003d) Efficient mutagenesis method for producing the templates of single nucleotide polymorphisms. *Mol. Biotechnol.* **24**, 105-110.

Zhang, J., Li, K., Liao, D., Pardinas, J. R. and Zhang, X. (2003b) Different applications of polymerases with and without proofreading activity in SNP analysis. *Lab. Invest.* **83**, 1147-1154.

Zhang, J., Liao, D., Chen, L., Guo, Z., Peng, Z. and Li, K. (2003c) On/off switch mediated by exo+ polymerases: Experimental analysis for its physiological and technological

implications. *J. Biochem. Mol. Biol.* **36**, 529-532.

Zhang, J., Liao, D., Pardinas, J. R., Li, H. and Li, K. (2003a) SNP discrimination through proofreading and OFF-switch of exo^+ polymerase. *Mol. Biotechnol.* **27**, 75-80.

Zhang, J., Liao, D., Yan, F., Zhang, X. and Li, K. (2001) Development of new generation of biochips for direct genetic analysis. *Am. J. Chin. Clin. Med.* **3**, 312-315.