3.Introduction

3.1 Sequencing by Hybridisation (SBH)

The ability to sequence DNA accurately and rapidly has revolutionized biology and medicine. The current state-of-the-art technology for high-throughput DNA sequencing utilizes capillary array DNA sequencers with laser-induced fluorescence detection (Smith et al. 1986; Ju et al. 1995). Although capillary DNA sequencing technology addresses the throughput and read length requirements to some extent, the sequencing throughput and accuracy still need some significant improvement. For example, the first few bases after the priming site are often masked by the high fluorescence signal from excess dye-labeled primers or terminators and are difficult to determine. The requirement of electrophoresis for DNA sequencing is still a main limitation for high-throughput DNA-sequencing and mutation-detection projects. A variety of alternative methods have been exploited for DNA sequencing including sequencing by hybridisation (SBH) (Drmanac et al. 1998), mass spectrometry sequencing (Roskey et al. 1996; Fu et al. 1998), pyrosequencing (Ronaghi et al. 1998), massively parallel signature sequencing with enzymatic cleavage and ligation (Brenner et al. 2000), sequence-specific detection of single-stranded DNA using engineered nanopores (Kasianowicz et al. 1996), sequencing of single DNA molecules (Braslavsky et al. 2003) and polymerase colonies (Mitra et al. 2003).

3.1.1 Principle and application of SBH

SBH was first proposed and patented in 1987 as an alternative to gel-based sequencing (Drmanac & Crkvenjakov, 1987; Southern, 1988) and afterwards validated in arrays of 7 and 8 mers (Southern et al. 1992; Drmanac et al. 1998). The basic idea behind SBH is that longer sequences can be reconstructed by the hybridisation of their constituent oligoprobes. For example, the three octamers

ATCAGGTC,

TCAGGTCT, and

CAGGTCTG

define the decamer ATCAGGTCTG.

The intrinsic power of SBH is that many sequences are determined in parallel. Another powerful aspect is that sequence information obtained is quite redundant, particularly as the size of the oligoprobes grows. Thus, the method should be quite

resistant to experimental errors, and far fewer than all 4^n oligoprobes are required to obtain reliable sequence information (Broude et al. 1994).

Two modes of hybridisation have been proposed: 1) DNA bound to a surface and oligoprobes in solution (Bains et al. 1988) or 2) bound oligoprobes with free DNA (Drmanac et al. 1990). In either mode, hybridisations must discriminate between those samples containing duplexes with a perfect match and those having hybrids with the mismatched base pairs.

In SBH methods, no knowledge of the frequency or the position of the oligoprobes in the clone DNA (target) is needed; the knowledge of oligoprobe sequences and hybridisation results suffices. Numerous oligoprobes are tested for their ability to hybridise to a clone DNA. Each oligoprobe that hybridises to the target indicates the presence of a complementary sequence within the target, revealing a small piece of information about its sequence. SBH procedures (Drmanac et al. 1992) provide a foundation for the low cost and fast analysis of the complex cDNA and/or genomic libraries and the SBH of densely arrayed DNA samples is very promising and allows for a wide range of both genome resequencing or genotyping (Drmanac et al. 1998; Kruglyak & Nickerson, 2001) and *de novo* sequencing (Drmanac *et al*. 2001).

3.1.2 Drawbacks of current SBH method

There are still a number of potentially drawbacks preventing SBH from practical implementation in large-scale DNA sequencing efforts. 1). Unsatisfactory level of discrimination between a thermodynamically unstable but correct perfectly matched duplex and a stable mismatched duplex. 2). The effect of secondary structure in the target DNA. Occasionally intra-molecular pairing is more stable than the hybrid within the corresponding target region, which leads to false negative hybridisation, and as a result, sequence reconstruction is prevented. 3). The possibility that certain oligoprobe sequences will have anomalous behaviour and will be unavailable to hybridisation. 4). The ambiguities in reconstructing sequences longer than a few hundred base pairs due to the recurrences. When several sequences have the same spectrum, there is no way to determine the true sequence. 5). 4^n would be quite a large number if the whole set of the oligoprobes is contemplated. For example, for 8-mer and 7-mer oligoprobe, theoretically, the number of full set of oligoprobes for hybridisation is 65,536 and 16,384, respectively, which was very expensive for synthesis and subsequent hybridisation against dsDNA sequences.

3.2 Oligonucleotide fingerprinting (OFP)

3.2.1 Principle of OFP

Since large-scale sequencing projects remain expensive, remarkable savings can be achieved by reducing the redundancy inherent in random approaches. One possible solution is the use of oligonucleotide fingerprinting (OFP) (Maier et al. 1994; Drmanac et al. 1996; Meier-Ewert et al. 1993, 1998; Milosavljevic et al. 1996). OFP represents a powerful method of cDNA and genomic DNA library characterization and normalisation. The purpose is to characterise a large number of clones with a well-defined set of organism specific hybridisation oligoprobes. The method is based on the sequence-specific hybridisation of a short number of synthetic oligoprobes to PCR products of individual clones (Fig. 1). The central principle of the OFP approach is that the fingerprint is characteristic for individual clone, providing partial sequence information about its DNA sequence. Obtained fingerprints are subjected to clustering analysis (Herwig et al. 1999) and thus clones sharing similar sequence are assigned to the same cluster whereas dissimilar ones are separated.

For cDNA library, the number and the size of the clusters present some information about the spectrum of expressed genes and their relative expression levels, respectively (Lennon and Lehrach, 1991). By comparing the computationally generated “in silico” fingerprints with database sequences, known genes are identified and potential candidate genes are discovered, thus reducing the sequencing efforts.

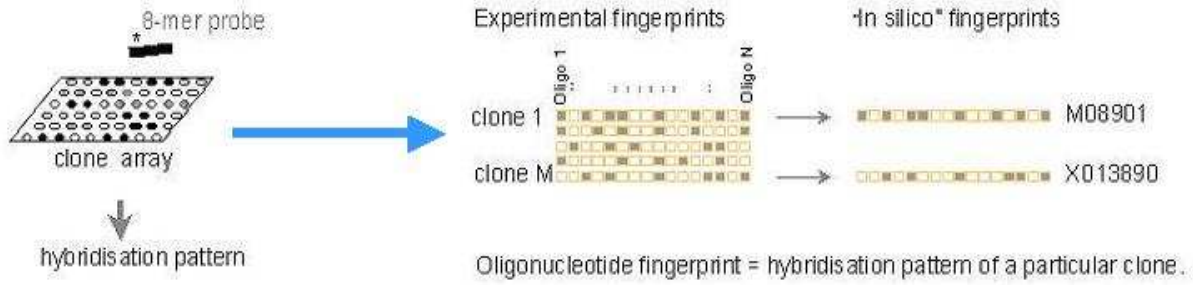

Figure 1. The principle of oligonucleotide fingerprinting

3.2.2 Applications for OFP

Since a large number of DNA samples (for example, 159,936 cDNA clones, Herwig, 2002) can be probed in parallel, sequence analysis by hybridisation is a high-

throughput method to gain biological information. OFP has been successfully employed in a wide range of applications, such as the establishment of clone maps (Hoheisel et al. 1991), analysis of genomic DNA (Radelof et al. 1998), expression profiling (Lennon and Lehrach, 1991; Meier-Ewert et al. 1993, 1998), as well as gene identification (Drmanac et al. 1996; Milosavljevic et al. 1996), and generation of non-redundant 'unigene' cDNA clone sets (Poustka et al. 1999; Clark et al. 2001; Herwig et al. 2002). The species investigated by oligonucleotide fingerprinting include E.coli (Milosavljevic et al. 1996), *Salmonella enterica* (Willse et al. 2004), mouse (Meier-Ewert et al. 1998), Sea Urchin (Poustka et al. 1999), zebrafish (Clark et al. 2001), sugar beet (Herwig et al. 2002) and human (Radelof et al. 1998; Herwig et al. 1999).

3.2.3 Drawbacks of classical OFP

Oligoprobes are widely used for hybridisation with support-bound DNA. When clone DNA is bound, oligoprobe lengths should be 6-10 nucleotides (Drmanac et al. 1989), which can provide high specificity of hybridisation (Drmanac et al. 1990). Pools of 16 decamers with a common octamer core (i.e. NXXXXXXXXN) (Meier-Ewert et al. 1998), 10-mer (Guerasimova et al. 1999; Poustka et al. 1999), 9-mer (Willse et al. 2004), 8-mer (Maier et al. 1994; Radelof et al. 1998; Herwig et al. 1999, 2000; Clark et al. 2001), and even 7-mer oligoprobes (Drmanac et al. 1996; Milosavljevic et al. 1996) have been applied for different organisms. Typically, the cDNAs or genomic DNA sequences of 1kb-2kb arrayed on filters are sequentially probed with 200-300 radioactively labeled 8-mer oligoprobes. The process has some inherent limitations to be overcome in order to further increase the degree of automation and throughput.

The bottleneck is primarily the high number, up to 250, of oligoprobes needed for analysis. 8-mer oligoprobes are required for hybridisation in order to achieve a sufficient partitioning and thus clustering of library (Herwig et al. 1999 and 2000). This high oligoprobe number leads to a laborious multi-step hybridisation procedure that is time- and resource-consuming. Additionally, the use of radioactive labeling renders the practical implementation hazardous and requires appropriate laboratory facilities for radioisotope handling. Furthermore, experiments with radioactively labeled heptamer DNA oligoprobes showed poor hybridisation stability (Herwig, 2000).

3.3 Locked nucleic acid (LNA)

Insufficient stability and poor mismatch discrimination of conventional DNA oligoprobes have resulted in growing interest in DNA analogues as tools for hybridisation. Among these are peptide nucleic acids (PNAs) (Nielson et al. 1991, 1997), 2'-fluoro N3-P5'-phosphoramidites (Schulz and Gryaznov, 1996), and 1', 5'-anhydrohexitol nucleic acids (HNAs) (Van Aerschot et al. 1996). Although the elevated thermal stabilities, however, they fail to provide enhanced target recognition.

3.3.1 Structure of LNA

Among the compounds introduced as DNA mimics in recent years, a new class of oligomers has become of particular interest: locked nucleic acid (LNA). The Wengel (1998) and Obika (1998) laboratories described a novel nucleotide termed LNA. LNA is a RNA analogue with high-affinity and biological stability containing nucleosides [2'-O, 4'-C-methylene-ß-D-ribofuranosyl monomers] whose major distinguishing characteristic is the presence of a methylene bridge that connects the 2′-oxygen of ribose with the 4′-carbon of the ribose ring as shown in Fig. 2. The structural studies show that LNA is an RNA mimic, fitting seamlessly into an A-type duplex geometry. This bridge results in a locked 3′-endo conformation, reducing the conformational flexibility of the ribose and increasing the local organization of the phosphate backbone as well as the strength of base stacking interactions. This in turn confers enhanced stability of LNA in pairing duplexes (Braasch and Corey, 2000). These molecular differences between conventional and LNA nucleosides allow increased stability of the nucleic acid duplexes formed between LNAs and other nucleic acids (Latorra et al. 2003).

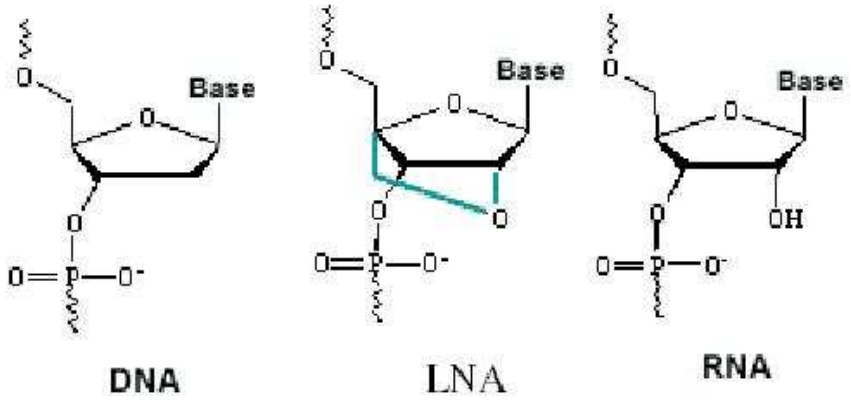

Figure 2. The chemical structures of DNA, LNA and RNA. The blue line highlights the methylene bridge. Base denotes nucleotise.

3.3.2 Characteristics of LNA

Hybridisation of LNA-modified oligoprobes with DNA or RNA targets has demonstrated unprecedented thermostabilities, shown by remarkable increases in melting points per LNA monomer introduced (ΔT_m values): ΔT_m=1~8℃ against DNA and ΔT_m=2~10℃ against RNA (Koshkin et al. 1998; Obika *et al.* 1998; Wengel, 1998; Braasch and Corey, 2001; Kurreck *et al.* 2002). The observed thermostabilities depend on oligomer length and its composition. The impact upon introduction of LNA monomers is most pronounced for single or multiple, but isolated modifications, and for short oligoprobes.

LNA bases are linked by the same phosphate backbone found in DNA or RNA, allowing LNA oligomers to be produced using standard reagents and automated synthesizers. The similarity of LNA to DNA/RNA synthesis also permits LNA bases to be interspersed among DNA and RNA, making it feasible for automated synthesis of both fully modified LNA and LNA/DNA or LNA/RNA chimers (Wengel, 1998). Another practical advantage is that LNAs are as soluble as DNA or RNA, facilitating their handling and simplifying experiments. Finally, LNA can be synthesised incorporating modifiers and labels such as biotin and various fluorescent dyes.

3.3.3 Applications for LNA

Enhanced nucleic acid recognition by LNA-containing oligoprobes made them desirable for many applications in molecular biology, including genotyping or single nucleotide polymorphism (SNP) analysis, antisense, decoy and fluorescence polarization (See review, Karkare and Bhatnagar, 2006), expression profiling or microarray, allele-specific PCR, triple helix formation, fluorescent in situ hybridisation (FISH) analysis, alteration of intron splicing and LNAzymes (See review, Jepsen et al. 2004), 5′ -nuclease assay (Letertre et al. 2003), real-time PCR (Costa et al. 2004), siRNA (Dahlgren et al. 2006) and microRNA (Orom et al. 2006).

3.4 Induced Fluorescence Resonance Energy Transfer (iFRET)

Induced Fluorescence Resonance Energy Transfer (iFRET) (Howell et al. 2002; Howell, 2006) is a variation of Fluorescence Resonance Energy Transfer (FRET) that is particularly well-suited for the detection of DNA hybridisation. FRET is a nonradioactive process whereby an excited fluorophore transfers energy onto a suitable fluorescent or nonfluorescent acceptor (Cardullo et al. 1988: Clegg, 1995). The efficiency of energy transfer varies inversely with the sixth power of the distance between 'donor' and 'acceptor'. Thus small changes in intermolecular spacing can produce large changes in transfer efficiency. Efficient transfer of the donor's excited state energy requires several spectroscopic and spatial criteria to be satisfied. These include significant overlap between the fluorescence emission spectrum of the donor and the absorption spectrum of the acceptor as illustrated in Fig. 3, approximately parallel alignment of the donor and acceptor transition dipole orientations, and the separation distance for a given donor–acceptor pair.

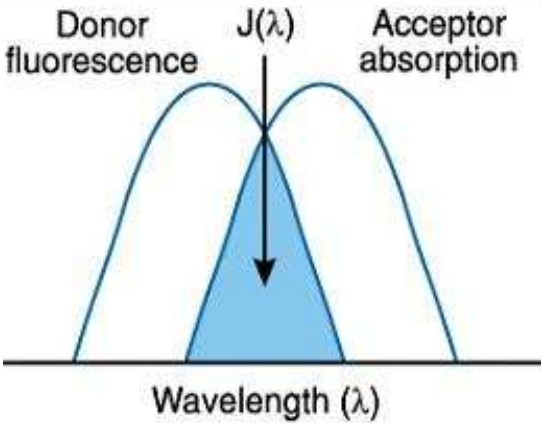

Figure 3. Overlap between the fluorescence emission spectrum of the donor and the absorption spectrum of the acceptor.

In iFRET, as showed in Fig. 4, the donor is a double-stranded (ds) DNA intercalating dye (e.g., SYBR Green I) that only fluoresces while interacting with double-stranded DNA, and the acceptor dye (denoted as a or b) is covalently linked to an oligoprobe. Hybridisation of the oligoprobe to its complement induces fluorescent emission of the donor dye and subsequent energy transfer to the acceptor dye, which can be detected as iFRET signal. When the duplex is denatured, the donor dye will not intercalate with single-stranded DNA, therefore no energy will be transferred from the donor dye to the acceptor dye, with no iFRET signal measurable. The hybridisation status can easily be measured by monitoring the

fluorescence output of the acceptor dye. Because the interaction of the donor dye is reversible and dependent on the presence of double-stranded DNA, iFRET is useful in the generation of DNA melting curves (Howell et al. 2002) and SNP genotyping (Jobs et al. 2003).

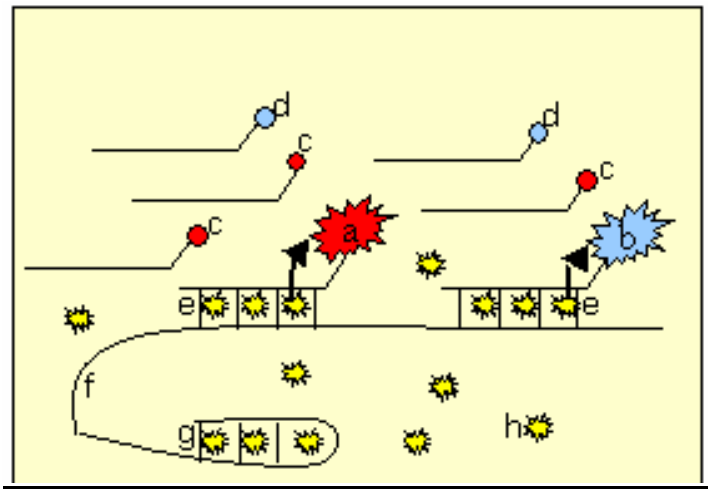

Figure 4. The strategy for detecting hybridisation with iFRET technology. **'a'** and **'b'**, oligoprobes labeled by different acceptor dyes (e.g., Cy-5 and Rox) binding with the complementary part of the target. **'c' and 'd'**, free oligoprobes labeled by different acceptor dyes (e.g., Cy-5 and Rox). **'e'**, intercalating dye (e.g., Sybr Green I) with the target-oligoprobe duplex from where the iFRET signal can be detected. **'f'**, single-stranded part of the target. **'g'**, intercalating dye (e.g., Sybr Green I) with the self-paired duplex of the target. **'h'**, free intercalating dye (e.g., Sybr Green I).

The donor signal is "induced" by the dye intercalation, hence the use of an intercalating dye as a donor for tracking DNA hybridisation and denaturation is fundamental for the iFRET concept. The iFRET configuration combines the advantages of intercalating dyes fluorescence (Howell et al. 1999), such as high signal strengths, much reduced background and low cost, with high specificity and multiplexing potential offered by the traditional FRET techniques..

Since the intensity of fluorescence from free acceptor oligoprobes (non-hybridised) is negligible, it is possible to detect DNA/RNA binding reactions in homogeneous solutions without the need for separation step or removing the free oligoprobes. Therefore FRET is suitable for many high-throughput screening (HTS) applications in homogeneous screening assays. Homogeneous DNA hybridisation assays based on FRET are attractive because of their simplicity of operation and use of standard optical equipment (Didenko, 2001; Takakusa et al. 2002).

3.5 Microarray

Microarrays are defined as a high density of arrays consisting of nucleic acid targets or oligoprobes immobilized on a solid substrate (e.g. glass or a nitrocellulose or nylon membrane). The power of DNA microarray lies on the specific molecular interaction via complementary base pairing in combination with their miniaturized scale suitable for massively parallel analyses.

Microarrays are extensions of hybridisation-based techniques which has been used for identification and quantitation of nucleic acids, such as in Southern and Northen blots. Samples under investigation are labeled and allowed to hybridise with the array. After sufficient hybridisation time and following appropriate washing steps, an image of the microarray is acquired and the representation of individual nucleic acids in the sample is fished out by the amount of hybridisation to the complementary DNAs placed at known positions on the array.

3.5.1 Formats of microarray

DNA microarrays are produced in two ways: 1). An array of DNA in form of ‘short’ oligoprobes is synthesized in situ (Maskos and Southern, 1992; Fodor et al. 1993). The array is exposed to labeled sample DNA, hybridised, and complementary sequences are determined. 2). Pre-amplified clone DNA is immobilized to a solid surface such as nylon membrane or glass, and exposed to a set of labeled oligoprobes separately (Drmanac et al. 1992) or in a mixture (Schena et al. 1995).

3.5.2 Supports for microarray

Currently, two support media are being used most for printed DNA arrays: nylon filter and glass, both with beneficial and disadvantageous features.

3.5.2.1 Membrane slides

In case of membrane microarrays, nylon or nitrocellulose membranes is utilized to immobilize DNA. Because membranes cannot accommodate as many spots as glass slides, this method provides more limited but less labor-intensive analysis of gene expression. Main advantage is that filter arrays can be re-used frequently, because the DNA attaches to the nylon surface without significant loss of DNA material after stripped for several times (Hauser et al. 1998). On the other hand,

more DNA is required for array production, since spot sizes cannot be reduced to a level which is possible for glass or other non-porous media.

Several positively charged commercial membranes are currently available in market, such as Amersham Hybond N+, Schleider & Schuell Nitran supercharged, Boehringer Nylon membranes + charged, and Pall Biodyne B (Alberola et al. 2004). A general disadvantage of these membranes is high auto-fluorescence background and inflexibility for automation.

3.5.2.2 Glass slides

Potential for miniaturisation, chemical inertness and low intrinsic fluorescence, are the main advantages of glass and polypropylene permitting high oligoprobe density. Because of the planar surface structure, however, the loading capacity could appear as a limiting factor. A negative aspect of plain glass support is also the limitation in the number of experiments that can be done using a single microarray. Most systems permit only single usage, thus even preventing proper quality control on the very chip that is to be used in the actual experiment; re-usability of microchips also eliminates from experimentation the variance between presumably identical chips, which significantly affects the experimental reliability of chip-based analyses.

3.5.3 Applications for microarray

Microarrays have become the method of choice for many hybridisation-based assays on a genomic scale. Microarray technology was initially developed for DNA-mapping (Poustka et al. 1986; Craig et al. 1990), sequencing (Cantor et al. 1992) and sequencing-by-hybridisation (SBH) applications (Bains & Smith, 1988; Drmanac et al. 1989, 1998). Few years later, microarrays have applied to differential gene expression analysis (Schena et al. 1995, 1996; Bowtell et al. 1999), polymorphism analysis (Gunderson et al. 1998), micro-organism identification (Wang *et al.* 2003), tumor classification (van't Veer *et al.* 2002) and analysis of the endocrine system (Jiang & Wang, 2003).