DNA-protein interactions are studied using a variety of methods such as gel-shift assays, footprinting, and transcriptional activation (Carey et al., *Cold Spring Harb Protoc*, **2012**(7): 733-57, 2012). While each of these methods may contribute distinct information about the location or effect of binding, they do not provide a simple way of quantitatively measuring specific binding. Fluorescence polarization/anisotropy provides a rapid, non-radioactive method for accurately measuring DNA-protein binding directly in solution without using filter binding, electrophoresis, or precipitation steps (Guest et al., 1991; Heyduk and Lee, 1990; LeTilly and Royer, 1993; Lundblad et al., 1996; Royer et al., 1992).

The molecular mechanisms by which genomic information directs the synthesis of different biomolecules has been the focus of much of molecular biology research over the last three decades. Previous studies have typically concentrated on individual genes, with the resulting general principles then providing insights into transcription, chromatin remodeling, messenger RNA splicing, DNA replication and numerous other genomic processes. Although many such principles seem valid as additional genes are investigated, they generally have not provided genome-wide insights about biological function. On the other hand, systematic analyses of transcripts and regulatory information are essential for the identification of genes and regulatory regions, and are an important resource for the study of human biology and disease. Such analyses can also provide comprehensive views of the organization and variability of genes and regulatory information across cellular contexts, species and individuals.

Genome-wide efforts such as the Encode project (Encyclopedia of DNA Elements) to identify e.g. all the transcription-factor-binding sites in the human genome have proven cumbersome and extremely labor-intensive (The ENCODE Project Consortium, *Nature*, **489**: 57-74, 2012).

There is thus still a need for a simple and reliable method for detecting protein/nucleic acid interactions.

Summary In one aspect, there is provided a method for detecting at least one modified base in a double-stranded nucleic acid molecule comprising a nucleic acid sequence, wherein the said double-stranded nucleic acid molecule is a

hairpin, and wherein one of the ends of the double-stranded nucleic acid is attached directly or indirectly to a support, and wherein the other end of the double-stranded nucleic acid is attached to a movable support, said method comprising the steps of:

a) denaturing said double-stranded nucleic acid molecule comprising a nucleic acid sequence by applying a physical force to said molecule;

b) providing a protein capable of binding said modified base;

c) renaturing said double stranded nucleic acid molecule in the presence of said protein;

d) detecting a blockage of the re-naturation of the double-stranded nucleic acid; and

e) determining the position of said blockage on the said double-stranded nucleic acid molecule, said determination comprising the steps of:

- measuring the distance (z) between the two ends of the double-stranded nucleic acid molecule which are attached to the support;
- measuring the distance (z_{high}) between the two ends of the double-stranded nucleic acid molecule which are attached to the support, when the said double-stranded nucleic acid molecule is denatured; and
- comparing z and z_{high}.

Detailed description of the invention

The present invention relates to a method for the determination of the binding of a protein to a nucleic acid molecule by physical manipulation.

Claims

1. A method for detecting at least one modified base in a double-stranded nucleic acid molecule comprising a nucleic acid sequence, wherein the said double-stranded nucleic acid molecule is a hairpin, and wherein one of the ends of the double-stranded nucleic acid is attached directly or indirectly to a support, and wherein the other end of the double-stranded nucleic acid is attached to a movable support, said method comprising the steps of:

   a) denaturing said double-stranded nucleic acid molecule comprising a nucleic acid sequence by applying a physical force to said molecule;
   b) providing a protein capable of binding said modified base;
   c) renaturing said double stranded nucleic acid molecule in the presence of said protein;
   d) detecting a blockage of the re-naturation of the double-stranded nucleic acid; and
   e) determining the position of said blockage on the said double-stranded nucleic acid molecule, said determination comprising the steps of:

      - measuring the distance (z) between the two ends of the double-stranded nucleic acid molecule which are attached to the support;
      - measuring the distance (z_{high}) between the two ends of the double-stranded nucleic acid molecule which are attached to the support, when the said double-stranded nucleic acid molecule is denatured; and
      - comparing z and z_{high}.

2. The method of claim 1, wherein a single-stranded nucleic acid molecule corresponding to said sequence is further provided in step b).

3. The method of any one of claim 1 or 2, wherein the double-stranded nucleic acid is denatured in step a) by moving away the supports.

4. The method of any one of claims 1 to 3, wherein a physical force above or equal to 15 pN, preferably above or equal to 17 pN, more preferably above or equal to 18 pN, is applied to the double-stranded molecule by moving away the supports.

5. The method of any one of claims 1 to 4, wherein the denatured double-stranded nucleic acid is re-natured in step c) by bringing the supports together.

6. The method of claim 5, wherein the force applied to the double-stranded molecule is reduced to less than or equal to 12 pN, preferably less than or equal to 11 pN, more preferably less than or equal to 10 pN, by bringing the supports together.

7. The method of any one of claims 1 to 6, wherein the steps a) to d) are repeated several times (so as to accumulate measurements and increase the signal/noise ratio).

8. The method of any one of claims 1 to 7, comprising a further step of measuring the duration of the blockage.

9. The method of claim 8, comprising a further step of comparing the duration of the blockage with a reference value.

10. The method of any one of claims 1 to 9, further comprising a step of sequencing the nucleic acid sequence bound by the said protein.

11. The method of any one of claims 1 to 10 wherein the at least one modified base is selected from an epigenetic modification, a mismatch, or an SNP.

12. The method of any one of claims 1 to 11, wherein the modified base is selected in the group consisting of 4-methylcytosine (4mC), 5-methylcytosine (5mC), 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) 5-carboxylcytosine (5caC), 5-hydroxymethyluracil (5hmU), and N6-methyladenosine (m6A).

13. The method of claim 12, wherein the modified base is 5mC.

14. The method of claim 12, wherein said protein is an antibody or an enzyme capable of binding said modified base.

15. The method of any one of claims 1 to 11, wherein said protein is selected from the group consisting of a MutS dimer, Msh2/Msh6 (MutSα), and Msh2/Msh3 (MutSβ).

16. The method of any one of claims 1 to 11 or 14, wherein the detection of a SNP in a sequence contained in a nucleic acid molecule, comprises the steps of:

a) hybridizing said nucleic acid with a single-stranded nucleic acid comprising the sequence found in the majority of the population; and

b) detecting the resulting mismatch by the method of any one of claims 1 to 11 or 14.